

---

# Pyro Documentation

Uber AI Labs

Dec 07, 2019



<b>1</b>	<b>Installation</b>	<b>1</b>
1.1	Install from Source . . . . .	1
<b>2</b>	<b>Getting Started</b>	<b>3</b>
<b>3</b>	<b>Primitives</b>	<b>5</b>
<b>4</b>	<b>Inference</b>	<b>11</b>
4.1	SVI . . . . .	11
4.2	ELBO . . . . .	12
4.3	Importance . . . . .	19
4.4	Reweighted Wake-Sleep . . . . .	20
4.5	Sequential Monte Carlo . . . . .	22
4.6	Stein Methods . . . . .	22
4.7	Likelihood free methods . . . . .	24
4.8	Discrete Inference . . . . .	25
4.9	Inference Utilities . . . . .	26
4.10	MCMC . . . . .	29
4.11	Automatic Guide Generation . . . . .	35
<b>5</b>	<b>Distributions</b>	<b>43</b>
5.1	PyTorch Distributions . . . . .	43
5.2	Pyro Distributions . . . . .	47
5.3	Transforms . . . . .	70
5.4	TransformModules . . . . .	72
5.5	Transform Factories . . . . .	84
<b>6</b>	<b>Parameters</b>	<b>89</b>
6.1	ParamStore . . . . .	89
<b>7</b>	<b>Neural Networks</b>	<b>93</b>
7.1	Pyro Modules . . . . .	93
7.2	AutoRegressiveNN . . . . .	97
7.3	ConditionalAutoRegressiveNN . . . . .	97
<b>8</b>	<b>Optimization</b>	<b>99</b>
8.1	Pyro Optimizers . . . . .	99

8.2	PyTorch Optimizers . . . . .	101
8.3	Higher-Order Optimizers . . . . .	102
<b>9</b>	<b>Poutine (Effect handlers)</b>	<b>105</b>
9.1	Handlers . . . . .	105
9.2	Trace . . . . .	113
9.3	Runtime . . . . .	115
9.4	Utilities . . . . .	116
9.5	Messengers . . . . .	117
<b>10</b>	<b>Miscellaneous Ops</b>	<b>127</b>
10.1	Utilities for HMC . . . . .	127
10.2	Newton Optimizers . . . . .	129
10.3	Tensor Indexing . . . . .	131
10.4	Tensor Contraction . . . . .	132
10.5	Gaussian Contraction . . . . .	134
10.6	Statistical Utilities . . . . .	136
10.7	State Space Model and GP Utilities . . . . .	139
<b>11</b>	<b>Automatic Name Generation</b>	<b>141</b>
11.1	Named Data Structures . . . . .	143
11.2	Scoping . . . . .	145
<b>12</b>	<b>Bayesian Neural Networks</b>	<b>149</b>
12.1	HiddenLayer . . . . .	149
<b>13</b>	<b>Causal Effect VAE</b>	<b>151</b>
13.1	CEVAE Class . . . . .	151
13.2	CEVAE Components . . . . .	153
13.3	Utilities . . . . .	154
<b>14</b>	<b>Easy Custom Guides</b>	<b>157</b>
14.1	EasyGuide . . . . .	157
14.2	easy_guide . . . . .	158
14.3	Group . . . . .	158
<b>15</b>	<b>Pyro Examples</b>	<b>161</b>
15.1	Datasets . . . . .	161
15.2	Utilities . . . . .	161
<b>16</b>	<b>Gaussian Processes</b>	<b>163</b>
16.1	Models . . . . .	164
16.2	Kernels . . . . .	175
16.3	Likelihoods . . . . .	181
16.4	Parameterized . . . . .	184
16.5	Util . . . . .	185
<b>17</b>	<b>Mini Pyro</b>	<b>187</b>
<b>18</b>	<b>Optimal Experiment Design</b>	<b>189</b>
18.1	Expected Information Gain . . . . .	190
18.2	Generalised Linear Mixed Models . . . . .	196
<b>19</b>	<b>Time Series</b>	<b>199</b>
19.1	Abstract Models . . . . .	199
19.2	Gaussian Processes . . . . .	200

19.3	Linear Gaussian State Space Models . . . . .	202
<b>20</b>	<b>Tracking</b>	<b>205</b>
20.1	Data Association . . . . .	205
20.2	Distributions . . . . .	208
20.3	Dynamic Models . . . . .	208
20.4	Extended Kalman Filter . . . . .	212
20.5	Hashing . . . . .	214
20.6	Measurements . . . . .	216
<b>21</b>	<b>Indices and tables</b>	<b>219</b>
	<b>Python Module Index</b>	<b>221</b>
	<b>Index</b>	<b>223</b>



### 1.1 Install from Source

Pyro supports 3.4+. To setup, install [PyTorch](#) then run:

```
pip install pyro-ppl
```

or install from source:

```
git clone https://github.com/pyro-ppl/pyro.git
cd pyro
python setup.py install
```





## CHAPTER 2

---

### Getting Started

---

- [Install Pyro.](#)
- Learn the basic concepts of Pyro: [models](#) and [inference](#).
- Dive in to other [tutorials](#) and [examples](#).



**sample** (*name*, *fn*, \**args*, \*\**kwargs*)

Calls the stochastic function *fn* with additional side-effects depending on *name* and the enclosing context (e.g. an inference algorithm). See [Intro I](#) and [Intro II](#) for a discussion.

### Parameters

- **name** – name of sample
- **fn** – distribution class or function
- **obs** – observed datum (optional; should only be used in context of inference) optionally specified in *kwargs*
- **infer** (*dict*) – Optional dictionary of inference parameters specified in *kwargs*. See inference documentation for details.

**Returns** sample

**param** (*name*, \**args*, \*\**kwargs*)

Saves the variable as a parameter in the param store. To interact with the param store or write to disk, see [Parameters](#).

### Parameters

- **name** (*str*) – name of parameter
- **init\_tensor** (*torch.Tensor* or *callable*) – initial tensor or lazy callable that returns a tensor. For large tensors, it may be cheaper to write e.g. `lambda: torch.randn(100000)`, which will only be evaluated on the initial statement.
- **constraint** (*torch.distributions.constraints.Constraint*) – torch constraint, defaults to `constraints.real`.
- **event\_dim** (*int*) – (optional) number of rightmost dimensions unrelated to batching. Dimension to the left of this will be considered batch dimensions; if the param statement is inside a subsampled plate, then corresponding batch dimensions of the parameter will be correspondingly subsampled. If unspecified, all dimensions will be considered event dims and no subsampling will be performed.

**Returns** parameter

**Return type** `torch.Tensor`

**module** (*name*, *nn\_module*, *update\_module\_params=False*)

Takes a `torch.nn.Module` and registers its parameters with the `ParamStore`. In conjunction with the `ParamStore` `save()` and `load()` functionality, this allows the user to save and load modules.

**Parameters**

- **name** (*str*) – name of module
- **nn\_module** (*torch.nn.Module*) – the module to be registered with Pyro
- **update\_module\_params** – determines whether Parameters in the PyTorch module get overridden with the values found in the `ParamStore` (if any). Defaults to *False*

**Returns** `torch.nn.Module`

**random\_module** (*name*, *nn\_module*, *prior*, *\*args*, *\*\*kwargs*)

**Warning:** The `random_module` primitive is deprecated, and will be removed in a future release. Use `PyroModule` instead to create Bayesian modules from `torch.nn.Module` instances. See the [Bayesian Regression tutorial](#) for an example.

Places a prior over the parameters of the module *nn\_module*. Returns a distribution (callable) over *nn.Modules*, which upon calling returns a sampled *nn.Module*.

**Parameters**

- **name** (*str*) – name of pyro module
- **nn\_module** (*torch.nn.Module*) – the module to be registered with pyro
- **prior** – pyro distribution, stochastic function, or python dict with parameter names as keys and respective distributions/stochastic functions as values.

**Returns** a callable which returns a sampled module

**factor** (*name*, *log\_factor*)

Factor statement to add arbitrary log probability factor to a probabilistic model.

**Parameters**

- **name** (*str*) – Name of the trivial sample
- **log\_factor** (*torch.Tensor*) – A possibly batched log probability factor.

**deterministic** (*name*, *value*, *event\_dim=None*)

EXPERIMENTAL Deterministic statement to add a *Delta* site with name *name* and value *value* to the trace. This is useful when we want to record values which are completely determined by their parents. For example:

```
x = sample("x", dist.Normal(0, 1))
x2 = deterministic("x2", x ** 2)
```

---

**Note:** The site does not affect the model density. This currently converts to a `sample()` statement, but may change in the future.

---

**Parameters**

- **name** (*str*) – Name of the site.
- **value** (*torch.Tensor*) – Value of the site.
- **event\_dim** (*int*) – Optional event dimension, defaults to *value.ndim*.

**class plate** (*name*, *size=None*, *subsample\_size=None*, *subsample=None*, *dim=None*, *use\_cuda=None*, *device=None*)

Construct for conditionally independent sequences of variables.

`plate` can be used either sequentially as a generator or in parallel as a context manager (formerly `irange` and `iarange`, respectively).

Sequential `plate` is similar to `range()` in that it generates a sequence of values.

Vectorized `plate` is similar to `torch.arange()` in that it yields an array of indices by which other tensors can be indexed. `plate` differs from `torch.arange()` in that it also informs inference algorithms that the variables being indexed are conditionally independent. To do this, `plate` is provided as context manager rather than a function, and users must guarantee that all computation within an `plate` context is conditionally independent:

```
with plate("name", size) as ind:
    # ...do conditionally independent stuff with ind...
```

Additionally, `plate` can take advantage of the conditional independence assumptions by subsampling the indices and informing inference algorithms to scale various computed values. This is typically used to subsample minibatches of data:

```
with plate("data", len(data), subsample_size=100) as ind:
    batch = data[ind]
    assert len(batch) == 100
```

By default `subsample_size=False` and this simply yields a `torch.arange(0, size)`. If  $0 < \text{subsample\_size} \leq \text{size}$  this yields a single random batch of indices of size `subsample_size` and scales all log likelihood terms by `size/batch_size`, within this context.

**Warning:** This is only correct if all computation is conditionally independent within the context.

### Parameters

- **name** (*str*) – A unique name to help inference algorithms match `plate` sites between models and guides.
- **size** (*int*) – Optional size of the collection being subsampled (like `stop` in builtin `range`).
- **subsample\_size** (*int*) – Size of minibatches used in subsampling. Defaults to *size*.
- **subsample** (Anything supporting `len()`) – Optional custom subsample for user-defined subsampling schemes. If specified, then `subsample_size` will be set to `len(subsample)`.
- **dim** (*int*) – An optional dimension to use for this independence index. If specified, `dim` should be negative, i.e. should index from the right. If not specified, `dim` is set to the rightmost dim that is left of all enclosing `plate` contexts.
- **use\_cuda** (*bool*) – DEPRECATED, use the `device` arg instead. Optional bool specifying whether to use cuda tensors for `subsample` and `log_prob`. Defaults to `torch.Tensor.is_cuda`.

- **device** (*str*) – Optional keyword specifying which device to place the results of *subsample* and *log\_prob* on. By default, results are placed on the same device as the default tensor.

**Returns** A reusable context manager yielding a single 1-dimensional `torch.Tensor` of indices.

Examples:

```
>>> # This version declares sequential independence and subsamples data:
>>> for i in plate('data', 100, subsample_size=10):
...     if z[i]: # Control flow in this example prevents vectorization.
...         obs = sample('obs_{}'.format(i), dist.Normal(loc, scale),
↪ obs=data[i])
```

```
>>> # This version declares vectorized independence:
>>> with plate('data'):
...     obs = sample('obs', dist.Normal(loc, scale), obs=data)
```

```
>>> # This version subsamples data in vectorized way:
>>> with plate('data', 100, subsample_size=10) as ind:
...     obs = sample('obs', dist.Normal(loc, scale), obs=data[ind])
```

```
>>> # This wraps a user-defined subsampling method for use in pyro:
>>> ind = torch.randint(0, 100, (10,)).long() # custom subsample
>>> with plate('data', 100, subsample=ind):
...     obs = sample('obs', dist.Normal(loc, scale), obs=data[ind])
```

```
>>> # This reuses two different independence contexts.
>>> x_axis = plate('outer', 320, dim=-1)
>>> y_axis = plate('inner', 200, dim=-2)
>>> with x_axis:
...     x_noise = sample("x_noise", dist.Normal(loc, scale))
...     assert x_noise.shape == (320,)
>>> with y_axis:
...     y_noise = sample("y_noise", dist.Normal(loc, scale))
...     assert y_noise.shape == (200, 1)
>>> with x_axis, y_axis:
...     xy_noise = sample("xy_noise", dist.Normal(loc, scale))
...     assert xy_noise.shape == (200, 320)
```

See [SVI Part II](#) for an extended discussion.

**get\_param\_store()**

Returns the ParamStore

**clear\_param\_store()**

Clears the ParamStore. This is especially useful if you're working in a REPL.

**validation\_enabled** (*is\_validate=True*)

Context manager that is useful when temporarily enabling/disabling validation checks.

**Parameters** *is\_validate* (*bool*) – (optional; defaults to True) temporary validation check override.

**enable\_validation** (*is\_validate=True*)

Enable or disable validation checks in Pyro. Validation checks provide useful warnings and errors, e.g. NaN checks, validating distribution arguments and support values, etc. which is useful for debugging. Since some of these checks may be expensive, we recommend turning this off for mature models.

**Parameters** `is_validate` (*bool*) – (optional; defaults to True) whether to enable validation checks.

**trace** (*fn=None, ignore\_warnings=False, jit\_options=None*)

Lazy replacement for `torch.jit.trace()` that works with Pyro functions that call `pyro.param()`.

The actual compilation artifact is stored in the `compiled` attribute of the output. Call diagnostic methods on this attribute.

Example:

```
def model(x):
    scale = pyro.param("scale", torch.tensor(0.5), constraint=constraints.
↪positive)
    return pyro.sample("y", dist.Normal(x, scale))

@pyro.ops.jit.trace
def model_log_prob_fn(x, y):
    cond_model = pyro.condition(model, data={"y": y})
    tr = pyro.poutine.trace(cond_model).get_trace(x)
    return tr.log_prob_sum()
```

### Parameters

- **fn** (*callable*) – The function to be traced.
- **ignore\_warnings** (*bool*) – Whether to ignore jit warnings.
- **jit\_options** (*dict*) – Optional dict of options to pass to `torch.jit.trace()` , e.g. `{"optimize": False}`.

**set\_rng\_seed** (*rng\_seed*)

Sets seeds of `torch` and `torch.cuda` (if available).

**Parameters** `rng_seed` (*int*) – The seed value.





In the context of probabilistic modeling, learning is usually called inference. In the particular case of Bayesian inference, this often involves computing (approximate) posterior distributions. In the case of parameterized models, this usually involves some sort of optimization. Pyro supports multiple inference algorithms, with support for stochastic variational inference (SVI) being the most extensive. Look here for more inference algorithms in future versions of Pyro.

See [Intro II](#) for a discussion of inference in Pyro.

## 4.1 SVI

**class SVI** (*model, guide, optim, loss, loss\_and\_grads=None, num\_samples=0, num\_steps=0, \*\*kwargs*)  
 Bases: `pyro.infer.abstract_infer.TracePosterior`

### Parameters

- **model** – the model (callable containing Pyro primitives)
- **guide** – the guide (callable containing Pyro primitives)
- **optim** (`pyro.optim.PyroOptim`) – a wrapper for a PyTorch optimizer
- **loss** (`pyro.infer.elbo.ELBO`) – an instance of a subclass of `ELBO`. Pyro provides three built-in losses: `Trace_ELBO`, `TraceGraph_ELBO`, and `TraceEnum_ELBO`. See the `ELBO` docs to learn how to implement a custom loss.
- **num\_samples** – (DEPRECATED) the number of samples for Monte Carlo posterior approximation
- **num\_steps** – (DEPRECATED) the number of optimization steps to take in `run()`

A unified interface for stochastic variational inference in Pyro. The most commonly used loss is `loss=Trace_ELBO()`. See the tutorial [SVI Part I](#) for a discussion.

**evaluate\_loss** (*\*args, \*\*kwargs*)

**Returns** estimate of the loss

**Return type** `float`

Evaluate the loss function. Any args or kwargs are passed to the model and guide.

`run(*args, **kwargs)`

**Warning:** This method is deprecated, and will be removed in a future release. For inference, use `step()` directly, and for predictions, use the `Predictive` class.

`step(*args, **kwargs)`

**Returns** estimate of the loss

**Return type** `float`

Take a gradient step on the loss function (and any auxiliary loss functions generated under the hood by `loss_and_grads`). Any args or kwargs are passed to the model and guide

## 4.2 ELBO

```
class ELBO(num_particles=1, max_plate_nesting=inf, max_iarange_nesting=None, vectorize_particles=False, strict_enumeration_warning=True, ignore_jit_warnings=False, jit_options=None, retain_graph=None, tail_adaptive_beta=-1.0)
```

Bases: `object`

`ELBO` is the top-level interface for stochastic variational inference via optimization of the evidence lower bound.

Most users will not interact with this base class `ELBO` directly; instead they will create instances of derived classes: `Trace_ELBO`, `TraceGraph_ELBO`, or `TraceEnum_ELBO`.

### Parameters

- **num\_particles** – The number of particles/samples used to form the ELBO (gradient) estimators.
- **max\_plate\_nesting** (`int`) – Optional bound on max number of nested `pyro.plate()` contexts. This is only required when enumerating over sample sites in parallel, e.g. if a site sets `infer={"enumerate": "parallel"}`. If omitted, ELBO may guess a valid value by running the (model,guide) pair once, however this guess may be incorrect if model or guide structure is dynamic.
- **vectorize\_particles** (`bool`) – Whether to vectorize the ELBO computation over `num_particles`. Defaults to False. This requires static structure in model and guide.
- **strict\_enumeration\_warning** (`bool`) – Whether to warn about possible misuse of enumeration, i.e. that `pyro.infer.traceenum_elbo.TraceEnum_ELBO` is used iff there are enumerated sample sites.
- **ignore\_jit\_warnings** (`bool`) – Flag to ignore warnings from the JIT tracer. When this is True, all `torch.jit.TracerWarning` will be ignored. Defaults to False.
- **jit\_options** (`bool`) – Optional dict of options to pass to `torch.jit.trace()`, e.g. `{"check_trace": True}`.
- **retain\_graph** (`bool`) – Whether to retain autograd graph during an SVI step. Defaults to None (False).

- **tail\_adaptive\_beta** (*float*) – Exponent beta with  $-1.0 \leq \text{beta} < 0.0$  for use with *TraceTailAdaptive\_ELBO*.

## References

- [1] *Automated Variational Inference in Probabilistic Programming* David Wingate, Theo Weber
- [2] *Black Box Variational Inference*, Rajesh Ranganath, Sean Gerrish, David M. Blei

```
class Trace_ELBO(num_particles=1, max_plate_nesting=inf, max_iarange_nesting=None, vectorize_particles=False, strict_enumeration_warning=True, ignore_jit_warnings=False, jit_options=None, retain_graph=None, tail_adaptive_beta=-1.0)
```

Bases: *pyro.infer.elbo.ELBO*

A trace implementation of ELBO-based SVI. The estimator is constructed along the lines of references [1] and [2]. There are no restrictions on the dependency structure of the model or the guide. The gradient estimator includes partial Rao-Blackwellization for reducing the variance of the estimator when non-reparameterizable random variables are present. The Rao-Blackwellization is partial in that it only uses conditional independence information that is marked by *plate* contexts. For more fine-grained Rao-Blackwellization, see *TraceGraph\_ELBO*.

## References

- [1] **Automated Variational Inference in Probabilistic Programming**, David Wingate, Theo Weber
- [2] **Black Box Variational Inference**, Rajesh Ranganath, Sean Gerrish, David M. Blei

```
loss (model, guide, *args, **kwargs)
```

**Returns** returns an estimate of the ELBO

**Return type** *float*

Evaluates the ELBO with an estimator that uses num\_particles many samples/particles.

```
differentiable_loss (model, guide, *args, **kwargs)
```

Computes the surrogate loss that can be differentiated with autograd to produce gradient estimates for the model and guide parameters

```
loss_and_grads (model, guide, *args, **kwargs)
```

**Returns** returns an estimate of the ELBO

**Return type** *float*

Computes the ELBO as well as the surrogate ELBO that is used to form the gradient estimator. Performs backward on the latter. Num\_particle many samples are used to form the estimators.

```
class JitTrace_ELBO(num_particles=1, max_plate_nesting=inf, max_iarange_nesting=None, vectorize_particles=False, strict_enumeration_warning=True, ignore_jit_warnings=False, jit_options=None, retain_graph=None, tail_adaptive_beta=-1.0)
```

Bases: *pyro.infer.trace\_elbo.Trace\_ELBO*

Like *Trace\_ELBO* but uses `pyro.ops.jit.compile()` to compile *loss\_and\_grads()*.

This works only for a limited set of models:

- Models must have static structure.
- Models must not depend on any global data (except the param store).
- All model inputs that are tensors must be passed in via *\*args*.
- All model inputs that are *not* tensors must be passed in via *\*\*kwargs*, and compilation will be triggered once per unique *\*\*kwargs*.

```
loss_and_surrogate_loss(model, guide, *args, **kwargs)
```

```
differentiable_loss(model, guide, *args, **kwargs)
```

```
loss_and_grads(model, guide, *args, **kwargs)
```

```
class TraceGraph_ELBO(num_particles=1, max_plate_nesting=inf, max_iarange_nesting=None,
                      vectorize_particles=False, strict_enumeration_warning=True, ignore_jit_warnings=False,
                      jit_options=None, retain_graph=None, tail_adaptive_beta=-1.0)
```

Bases: `pyro.infer.elbo.ELBO`

A `TraceGraph` implementation of ELBO-based SVI. The gradient estimator is constructed along the lines of reference [1] specialized to the case of the ELBO. It supports arbitrary dependency structure for the model and guide as well as baselines for non-reparameterizable random variables. Where possible, conditional dependency information as recorded in the `Trace` is used to reduce the variance of the gradient estimator. In particular two kinds of conditional dependency information are used to reduce variance:

- the sequential order of samples (z is sampled after y => y does not depend on z)
- `plate` generators

References

[1] *Gradient Estimation Using Stochastic Computation Graphs*, John Schulman, Nicolas Heess, Theophane Weber, Pieter Abbeel

[2] *Neural Variational Inference and Learning in Belief Networks* Andriy Mnih, Karol Gregor

```
loss(model, guide, *args, **kwargs)
```

**Returns** returns an estimate of the ELBO

**Return type** `float`

Evaluates the ELBO with an estimator that uses `num_particles` many samples/particles.

```
loss_and_grads(model, guide, *args, **kwargs)
```

**Returns** returns an estimate of the ELBO

**Return type** `float`

Computes the ELBO as well as the surrogate ELBO that is used to form the gradient estimator. Performs backward on the latter. `Num_particle` many samples are used to form the estimators. If baselines are present, a baseline loss is also constructed and differentiated.

```
class JitTraceGraph_ELBO(num_particles=1, max_plate_nesting=inf, max_iarange_nesting=None,
                        vectorize_particles=False, strict_enumeration_warning=True, ignore_jit_warnings=False,
                        jit_options=None, retain_graph=None, tail_adaptive_beta=-1.0)
```

Bases: `pyro.infer.tracegraph_elbo.TraceGraph_ELBO`

Like `TraceGraph_ELBO` but uses `torch.jit.trace()` to compile `loss_and_grads()`.

This works only for a limited set of models:

- Models must have static structure.
- Models must not depend on any global data (except the param store).
- All model inputs that are tensors must be passed in via `*args`.
- All model inputs that are *not* tensors must be passed in via `**kwargs`, and compilation will be triggered once per unique `**kwargs`.

```
loss_and_grads(model, guide, *args, **kwargs)
```

```
class BackwardSampleMessenger (enum_trace, guide_trace)
```

Bases: `pyro.poutine.messenger.Messenger`

Implements forward filtering / backward sampling for sampling from the joint posterior distribution

```
class TraceEnum_ELBO (num_particles=1,    max_plate_nesting=inf,    max_iarange_nesting=None,
                      vectorize_particles=False,    strict_enumeration_warning=True,    ignore_jit_warnings=False,    jit_options=None,    retain_graph=None,
                      tail_adaptive_beta=-1.0)
```

Bases: `pyro.infer.elbo.ELBO`

A trace implementation of ELBO-based SVI that supports - exhaustive enumeration over discrete sample sites, and - local parallel sampling over any sample site in the guide.

To enumerate over a sample site in the guide, mark the site with either `infer={'enumerate': 'sequential'}` or `infer={'enumerate': 'parallel'}`. To configure all guide sites at once, use `config_enumerate()`. To enumerate over a sample site in the model, mark the site `infer={'enumerate': 'parallel'}` and ensure the site does not appear in the guide.

This assumes restricted dependency structure on the model and guide: variables outside of an `plate` can never depend on variables inside that `plate`.

```
loss (model, guide, *args, **kwargs)
```

**Returns** an estimate of the ELBO

**Return type** `float`

Estimates the ELBO using `num_particles` many samples (particles).

```
differentiable_loss (model, guide, *args, **kwargs)
```

**Returns** a differentiable estimate of the ELBO

**Return type** `torch.Tensor`

**Raises** `ValueError` – if the ELBO is not differentiable (e.g. is identically zero)

Estimates a differentiable ELBO using `num_particles` many samples (particles). The result should be infinitely differentiable (as long as underlying derivatives have been implemented).

```
loss_and_grads (model, guide, *args, **kwargs)
```

**Returns** an estimate of the ELBO

**Return type** `float`

Estimates the ELBO using `num_particles` many samples (particles). Performs backward on the ELBO of each particle.

```
compute_marginals (model, guide, *args, **kwargs)
```

Computes marginal distributions at each model-enumerated sample site.

**Returns** a dict mapping site name to marginal `Distribution` object

**Return type** `OrderedDict`

```
sample_posterior (model, guide, *args, **kwargs)
```

Sample from the joint posterior distribution of all model-enumerated sites given all observations

```
class JitTraceEnum_ELBO (num_particles=1,    max_plate_nesting=inf,    max_iarange_nesting=None,
                        vectorize_particles=False,    strict_enumeration_warning=True,    ignore_jit_warnings=False,    jit_options=None,    retain_graph=None,
                        tail_adaptive_beta=-1.0)
```

Bases: `pyro.infer.traceenum_elbo.TraceEnum_ELBO`

Like `TraceEnum_ELBO` but uses `pyro.ops.jit.compile()` to compile `loss_and_grads()`.

This works only for a limited set of models:

- Models must have static structure.
- Models must not depend on any global data (except the param store).
- All model inputs that are tensors must be passed in via `*args`.
- All model inputs that are *not* tensors must be passed in via `**kwargs`, and compilation will be triggered once per unique `**kwargs`.

**differentiable\_loss** (*model*, *guide*, *\*args*, *\*\*kwargs*)

**loss\_and\_grads** (*model*, *guide*, *\*args*, *\*\*kwargs*)

```
class TraceMeanField_ELBO (num_particles=1, max_plate_nesting=inf, max_iarange_nesting=None,
                           vectorize_particles=False, strict_enumeration_warning=True, ignore_jit_warnings=False,
                           jit_options=None, retain_graph=None, tail_adaptive_beta=-1.0)
```

Bases: `pyro.infer.trace_elbo.Trace_ELBO`

A trace implementation of ELBO-based SVI. This is currently the only ELBO estimator in Pyro that uses analytic KL divergences when those are available.

In contrast to, e.g., `TraceGraph_ELBO` and `Trace_ELBO` this estimator places restrictions on the dependency structure of the model and guide. In particular it assumes that the guide has a mean-field structure, i.e. that it factorizes across the different latent variables present in the guide. It also assumes that all of the latent variables in the guide are reparameterized. This latter condition is satisfied for, e.g., the Normal distribution but is not satisfied for, e.g., the Categorical distribution.

**Warning:** This estimator may give incorrect results if the mean-field condition is not satisfied.

Note for advanced users:

The mean field condition is a sufficient but not necessary condition for this estimator to be correct. The precise condition is that for every latent variable  $z$  in the guide, its parents in the model must not include any latent variables that are descendants of  $z$  in the guide. Here ‘parents in the model’ and ‘descendants in the guide’ is with respect to the corresponding (statistical) dependency structure. For example, this condition is always satisfied if the model and guide have identical dependency structures.

**loss** (*model*, *guide*, *\*args*, *\*\*kwargs*)

**Returns** returns an estimate of the ELBO

**Return type** `float`

Evaluates the ELBO with an estimator that uses `num_particles` many samples/particles.

```
class JitTraceMeanField_ELBO (num_particles=1, max_plate_nesting=inf,
                              max_iarange_nesting=None, vectorize_particles=False,
                              strict_enumeration_warning=True, ignore_jit_warnings=False,
                              jit_options=None, retain_graph=None, tail_adaptive_beta=-1.0)
```

Bases: `pyro.infer.trace_mean_field_elbo.TraceMeanField_ELBO`

Like `TraceMeanField_ELBO` but uses `pyro.ops.jit.trace()` to compile `loss_and_grads()`.

This works only for a limited set of models:

- Models must have static structure.
- Models must not depend on any global data (except the param store).

- All model inputs that are tensors must be passed in via `*args`.
- All model inputs that are *not* tensors must be passed in via `**kwargs`, and compilation will be triggered once per unique `**kwargs`.

`differentiable_loss(model, guide, *args, **kwargs)`

`loss_and_grads(model, guide, *args, **kwargs)`

```
class TraceTailAdaptive_ELBO(num_particles=1, max_plate_nesting=inf,
                             max_iarange_nesting=None, vectorize_particles=False,
                             strict_enumeration_warning=True, ignore_jit_warnings=False,
                             jit_options=None, retain_graph=None, tail_adaptive_beta=-1.0)
```

Bases: `pyro.infer.trace_elbo.Trace_ELBO`

Interface for Stochastic Variational Inference with an adaptive f-divergence as described in ref. [1]. Users should specify `num_particles > 1` and `vectorize_particles==True`. The argument `tail_adaptive_beta` can be specified to modify how the adaptive f-divergence is constructed. See reference for details.

Note that this interface does not support computing the variational objective itself; rather it only supports computing gradients of the variational objective. Consequently, one might want to use another SVI interface (e.g. `RenyiELBO`) in order to monitor convergence.

Note that this interface only supports models in which all the latent variables are fully reparameterized. It also does not support data subsampling.

References [1] “Variational Inference with Tail-adaptive f-Divergence”, Dilin Wang, Hao Liu, Qiang Liu, NeurIPS 2018 <https://papers.nips.cc/paper/7816-variational-inference-with-tail-adaptive-f-divergence>

`loss(model, guide, *args, **kwargs)`

It is not necessary to estimate the tail-adaptive f-divergence itself in order to compute the corresponding gradients. Consequently the loss method is left unimplemented.

```
class RenyiELBO(alpha=0, num_particles=2, max_plate_nesting=inf, max_iarange_nesting=None, vectorize_particles=False, strict_enumeration_warning=True)
```

Bases: `pyro.infer.elbo.ELBO`

An implementation of Renyi’s  $\alpha$ -divergence variational inference following reference [1].

In order for the objective to be a strict lower bound, we require  $\alpha \geq 0$ . Note, however, that according to reference [1], depending on the dataset  $\alpha < 0$  might give better results. In the special case  $\alpha = 0$ , the objective function is that of the important weighted autoencoder derived in reference [2].

---

**Note:** Setting  $\alpha < 1$  gives a better bound than the usual ELBO. For  $\alpha = 1$ , it is better to use `Trace_ELBO` class because it helps reduce variances of gradient estimations.

---

### Parameters

- **alpha** (`float`) – The order of  $\alpha$ -divergence. Here  $\alpha \neq 1$ . Default is 0.
- **num\_particles** – The number of particles/samples used to form the objective (gradient) estimator. Default is 2.
- **max\_plate\_nesting** (`int`) – Bound on max number of nested `pyro.plate()` contexts. Default is infinity.
- **strict\_enumeration\_warning** (`bool`) – Whether to warn about possible misuse of enumeration, i.e. that `TraceEnum_ELBO` is used iff there are enumerated sample sites.

References:

[1] *Renyi Divergence Variational Inference*, Yingzhen Li, Richard E. Turner

[2] *Importance Weighted Autoencoders*, Yuri Burda, Roger Grosse, Ruslan Salakhutdinov

`loss(model, guide, *args, **kwargs)`

**Returns** returns an estimate of the ELBO

**Return type** `float`

Evaluates the ELBO with an estimator that uses `num_particles` many samples/particles.

`loss_and_grads(model, guide, *args, **kwargs)`

**Returns** returns an estimate of the ELBO

**Return type** `float`

Computes the ELBO as well as the surrogate ELBO that is used to form the gradient estimator. Performs backward on the latter. `Num_particle` many samples are used to form the estimators.

```
class TraceTMC_ELBO(num_particles=1, max_plate_nesting=inf, max_iarange_nesting=None,
                    vectorize_particles=False, strict_enumeration_warning=True, ignore_jit_warnings=False,
                    jit_options=None, retain_graph=None, tail_adaptive_beta=-1.0)
```

Bases: `pyro.infer.elbo.ELBO`

A trace-based implementation of Tensor Monte Carlo [1] by way of Tensor Variable Elimination [2] that supports: - local parallel sampling over any sample site in the model or guide - exhaustive enumeration over any sample site in the model or guide

To take multiple samples, mark the site with `infer={'enumerate': 'parallel', 'num_samples': N}`. To configure all sites in a model or guide at once, use `config_enumerate()`. To enumerate or sample a sample site in the model, mark the site and ensure the site does not appear in the guide.

This assumes restricted dependency structure on the model and guide: variables outside of an `plate` can never depend on variables inside that `plate`.

References

[1] *Tensor Monte Carlo: Particle Methods for the GPU Era*, Laurence Aitchison (2018)

[2] *Tensor Variable Elimination for Plated Factor Graphs*, Fritz Obermeyer, Eli Bingham, Martin Jankowiak, Justin Chiu, Neeraj Pradhan, Alexander Rush, Noah Goodman (2019)

`differentiable_loss(model, guide, *args, **kwargs)`

**Returns** a differentiable estimate of the marginal log-likelihood

**Return type** `torch.Tensor`

**Raises** `ValueError` – if the ELBO is not differentiable (e.g. is identically zero)

Computes a differentiable TMC estimate using `num_particles` many samples (particles). The result should be infinitely differentiable (as long as underlying derivatives have been implemented).

`loss(model, guide, *args, **kwargs)`

`loss_and_grads(model, guide, *args, **kwargs)`



## 4.3 Importance

**class Importance** (*model*, *guide*=None, *num\_samples*=None)

Bases: `pyro.infer.abstract_infer.TracePosterior`

### Parameters

- **model** – probabilistic model defined as a function
- **guide** – guide used for sampling defined as a function
- **num\_samples** – number of samples to draw from the guide (default 10)

This method performs posterior inference by importance sampling using the guide as the proposal distribution. If no guide is provided, it defaults to proposing from the model’s prior.

**get\_ESS** ()

Compute (Importance Sampling) Effective Sample Size (ESS).

**get\_log\_normalizer** ()

Estimator of the normalizing constant of the target distribution. (mean of the unnormalized weights)

**get\_normalized\_weights** (*log\_scale*=False)

Compute the normalized importance weights.

**psis\_diagnostic** (*model*, *guide*, \**args*, \*\**kwargs*)

Computes the Pareto tail index  $k$  for a model/guide pair using the technique described in [1], which builds on previous work in [2]. If  $0 < k < 0.5$  the guide is a good approximation to the model posterior, in the sense described in [1]. If  $0.5 \leq k \leq 0.7$ , the guide provides a suboptimal approximation to the posterior, but may still be useful in practice. If  $k > 0.7$  the guide program provides a poor approximation to the full posterior, and caution should be used when using the guide. Note, however, that a guide may be a poor fit to the full posterior while still yielding reasonable model predictions. If  $k < 0.0$  the importance weights corresponding to the model and guide appear to be bounded from above; this would be a bizarre outcome for a guide trained via ELBO maximization. Please see [1] for a more complete discussion of how the tail index  $k$  should be interpreted.

Please be advised that a large number of samples may be required for an accurate estimate of  $k$ .

Note that we assume that the model and guide are both vectorized and have static structure. As is canonical in Pyro, the args and kwargs are passed to the model and guide.

References [1] ‘Yes, but Did It Work?: Evaluating Variational Inference.’ Yuling Yao, Aki Vehtari, Daniel Simpson, Andrew Gelman [2] ‘Pareto Smoothed Importance Sampling.’ Aki Vehtari, Andrew Gelman, Jonah Gabry

### Parameters

- **model** (*callable*) – the model program.
- **guide** (*callable*) – the guide program.
- **num\_particles** (*int*) – the total number of times we run the model and guide in order to compute the diagnostic. defaults to 1000.
- **max\_simultaneous\_particles** – the maximum number of simultaneous samples drawn from the model and guide. defaults to *num\_particles*. *num\_particles* must be divisible by *max\_simultaneous\_particles*. compute the diagnostic. defaults to 1000.
- **max\_plate\_nesting** (*int*) – optional bound on max number of nested `pyro.plate()` contexts in the model/guide. defaults to 7.

**Returns float** the PSIS diagnostic  $k$

**vectorized\_importance\_weights** (*model*, *guide*, \**args*, \*\**kwargs*)

**Parameters**

- **model** – probabilistic model defined as a function
- **guide** – guide used for sampling defined as a function
- **num\_samples** – number of samples to draw from the guide (default 1)
- **max\_plate\_nesting** (*int*) – Bound on max number of nested `pyro.plate()` contexts.
- **normalized** (*bool*) – set to True to return self-normalized importance weights

**Returns** returns a `(num_samples,)`-shaped tensor of importance weights and the model and guide traces that produced them

Vectorized computation of importance weights for models with static structure:

```
log_weights, model_trace, guide_trace = \
    vectorized_importance_weights(model, guide, *args,
                                  num_samples=1000,
                                  max_plate_nesting=4,
                                  normalized=False)
```

## 4.4 Reweighted Wake-Sleep

```
class ReweightedWakeSleep(num_particles=2,      insomnia=1.0,      model_has_params=True,
                           num_sleep_particles=None,      vectorize_particles=True,
                           max_plate_nesting=inf, strict_enumeration_warning=True)
```

Bases: `pyro.infer.elbo.ELBO`

An implementation of Reweighted Wake Sleep following reference [1].

---

**Note:** Sampling and `log_prob` evaluation asymptotic complexity:

---

- 1) **Using wake-theta and/or wake-phi**  $O(\text{num\_particles})$  samples from guide,  $O(\text{num\_particles})$  `log_prob` evaluations of model and guide
  - 2) **Using sleep-phi**  $O(\text{num\_sleep\_particles})$  samples from model,  $O(\text{num\_sleep\_particles})$  `log_prob` evaluations of guide
- if 1) and 2) are combined**,  $O(\text{num\_particles})$  samples from the guide,  $O(\text{num\_sleep\_particles})$  from the model,  $O(\text{num\_particles} + \text{num\_sleep\_particles})$  `log_prob` evaluations of the guide, and  $O(\text{num\_particles})$  evaluations of the model

---

**Note:** This is particularly useful for models with stochastic branching, as described in [2].

---

---

**Note:** This returns `_two_` losses, one each for (a) the model parameters (*theta*), computed using the *iwae* objective, and (b) the guide parameters (*phi*), computed using (a combination of) the *csis* objective and a self-normalized importance-sampled version of the *csis* objective.

---

**Note:** In order to enable computing the sleep-phi terms, the guide program must have its observations explicitly passed in through the keyworded argument *observations*. Where the value of the observations is unknown during definition, such as for amortized variational inference, it may be given a default argument as *observations=None*, and the correct value supplied during learning through *svi.step(observations=...)*.

**Warning:** Mini-batch training is not supported yet.

### Parameters

- **num\_particles** (*int*) – The number of particles/samples used to form the objective (gradient) estimator. Default is 2.
- **insomnia** – The scaling between the wake-phi and sleep-phi terms. Default is 1.0 [wake-phi]
- **model\_has\_params** (*bool*) – Indicate if model has learnable params. Useful in avoiding extra computation when running in pure sleep mode [csis]. Default is True.
- **num\_sleep\_particles** (*int*) – The number of particles used to form the sleep-phi estimator. Matches *num\_particles* by default.
- **vectorize\_particles** (*bool*) – Whether the traces should be vectorised across *num\_particles*. Default is True.
- **max\_plate\_nesting** (*int*) – Bound on max number of nested *pyro.plate()* contexts. Default is infinity.
- **strict\_enumeration\_warning** (*bool*) – Whether to warn about possible misuse of enumeration, i.e. that *TraceEnum\_ELBO* is used iff there are enumerated sample sites.

References:

[1] *Reweighted Wake-Sleep*, Jörg Bornschein, Yoshua Bengio

[2] *Revisiting Reweighted Wake-Sleep for Models with Stochastic Control Flow*, Tuan Anh Le, Adam R. Kosior, N. Siddharth, Yee Whye Teh, Frank Wood

**loss** (*model*, *guide*, \**args*, \*\**kwargs*)

**Returns** returns model loss and guide loss

**Return type** *float*, *float*

Computes the re-weighted wake-sleep estimators for the model (wake-theta) and the guide (insomnia \* wake-phi + (1 - insomnia) \* sleep-phi).

**loss\_and\_grads** (*model*, *guide*, \**args*, \*\**kwargs*)

**Returns** returns model loss and guide loss

**Return type** *float*

Computes the RWS estimators for the model (wake-theta) and the guide (wake-phi). Performs backward as appropriate on both, using num\_particle many samples/particles.

## 4.5 Sequential Monte Carlo

**class** `SMCFilter` (*model*, *guide*, *num\_particles*, *max\_plate\_nesting*)

Bases: `object`

`SMCFilter` is the top-level interface for filtering via sequential monte carlo.

The model and guide should be objects with two methods: `.init(state, ...)` and `.step(state, ...)`, intended to be called first with `init()`, then with `step()` repeatedly. These two methods should have the same signature as `SMCFilter`'s `init()` and `step()` of this class, but with an extra first argument `state` that should be used to store all tensors that depend on sampled variables. The `state` will be a dict-like object, `SMCState`, with arbitrary keys and `torch.Tensor` values. Models can read and write `state` but guides can only read from it.

Inference complexity is  $O(\text{len}(\text{state}) * \text{num\_time\_steps})$ , so to avoid quadratic complexity in Markov models, ensure that `state` has fixed size.

### Parameters

- **model** (*object*) – probabilistic model with `init` and `step` methods
- **guide** (*object*) – guide used for sampling, with `init` and `step` methods
- **num\_particles** (*int*) – The number of particles used to form the distribution.
- **max\_plate\_nesting** (*int*) – Bound on max number of nested `pyro.plate()` contexts.

**get\_empirical** ()

**Returns** a marginal distribution over all state tensors.

**Return type** a dictionary with keys which are latent variables and values which are `Empirical` objects.

**init** (\*args, \*\*kwargs)

Perform any initialization for sequential importance resampling. Any args or kwargs are passed to the model and guide

**step** (\*args, \*\*kwargs)

Take a filtering step using sequential importance resampling updating the particle weights and values while resampling if desired. Any args or kwargs are passed to the model and guide

**class** `SMCState` (*num\_particles*)

Bases: `dict`

Dictionary-like object to hold a vectorized collection of tensors to represent all state during inference with `SMCFilter`. During inference, the `SMCFilter` resample these tensors.

Keys may have arbitrary hashable type. Values must be `torch.Tensor`s.

**Parameters** `num_particles` (*int*) –

## 4.6 Stein Methods

**class** `IMQSteinKernel` (*alpha=0.5*, *beta=-0.5*, *bandwidth\_factor=None*)

Bases: `pyro.infer.svgd.SteinKernel`

An IMQ (inverse multi-quadratic) kernel for use in the SVGD inference algorithm [1]. The bandwidth of the kernel is chosen from the particles using a simple heuristic as in reference [2]. The kernel takes the form

$$K(x, y) = (\alpha + \|x - y\|^2/h)^{\beta}$$

where  $\alpha$  and  $\beta$  are user-specified parameters and  $h$  is the bandwidth.

#### Parameters

- **alpha** (*float*) – Kernel hyperparameter, defaults to 0.5.
- **beta** (*float*) – Kernel hyperparameter, defaults to -0.5.
- **bandwidth\_factor** (*float*) – Optional factor by which to scale the bandwidth, defaults to 1.0.

**Variables** **bandwidth\_factor** (*float*) – Property that controls the factor by which to scale the bandwidth at each iteration.

#### References

- [1] “Stein Points,” Wilson Ye Chen, Lester Mackey, Jackson Gorham, Francois-Xavier Briol, Chris. J. Oates.
- [2] “Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm,” Qiang Liu, Dilin Wang

#### **bandwidth\_factor**

**log\_kernel\_and\_grad** (*particles*)

See `pyro.infer.svgd.SteinKernel.log_kernel_and_grad()`

**class** **RBFSteinKernel** (*bandwidth\_factor=None*)

Bases: `pyro.infer.svgd.SteinKernel`

A RBF kernel for use in the SVGD inference algorithm. The bandwidth of the kernel is chosen from the particles using a simple heuristic as in reference [1].

**Parameters** **bandwidth\_factor** (*float*) – Optional factor by which to scale the bandwidth, defaults to 1.0.

**Variables** **bandwidth\_factor** (*float*) – Property that controls the factor by which to scale the bandwidth at each iteration.

#### References

- [1] “Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm,” Qiang Liu, Dilin Wang

#### **bandwidth\_factor**

**log\_kernel\_and\_grad** (*particles*)

See `pyro.infer.svgd.SteinKernel.log_kernel_and_grad()`

**class** **SVGD** (*model, kernel, optim, num\_particles, max\_plate\_nesting, mode='univariate'*)

Bases: `object`

A basic implementation of Stein Variational Gradient Descent as described in reference [1].

#### Parameters

- **model** – The model (callable containing Pyro primitives). Model must be fully vectorized and may only contain continuous latent variables.
- **kernel** – a SVGD compatible kernel like `RBFSteinKernel`.
- **optim** (`pyro.optim.PyroOptim`) – A wrapper for a PyTorch optimizer.
- **num\_particles** (*int*) – The number of particles used in SVGD.
- **max\_plate\_nesting** (*int*) – The max number of nested `pyro.plate()` contexts in the model.

- **mode** (*str*) – Whether to use a Kernelized Stein Discrepancy that makes use of *multivariate* test functions (as in [1]) or *univariate* test functions (as in [2]). Defaults to *univariate*.

Example usage:

```
from pyro.infer import SVGD, RBFSteinKernel
from pyro.optim import Adam

kernel = RBFSteinKernel()
adam = Adam({"lr": 0.1})
svgd = SVGD(model, kernel, adam, num_particles=50, max_plate_nesting=0)

for step in range(500):
    svgd.step(model_arg1, model_arg2)

final_particles = svgd.get_named_particles()
```

References

- [1] “Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm,” Qiang Liu, Dilin Wang
- [2] “Kernelized Complete Conditional Stein Discrepancy,” Raghav Singhal, Saad Lahlou, Rajesh Ranganath

**get\_named\_particles()**

Create a dictionary mapping name to vectorized value, of the form {name: tensor}. The leading dimension of each tensor corresponds to particles, i.e. this creates a struct of arrays.

**step** (\*args, \*\*kwargs)

Computes the SVGD gradient, passing args and kwargs to the model, and takes a gradient step.

**Return dict** A dictionary of the form {name: float}, where each float is a mean squared gradient. This can be used to monitor the convergence of SVGD.

**class SteinKernel**

Bases: `object`

Abstract class for kernels used in the `SVGD` inference algorithm.

**log\_kernel\_and\_grad** (particles)

Compute the component kernels and their gradients.

**Parameters** **particles** – a tensor with shape (N, D)

**Returns** A pair (*log\_kernel*, *kernel\_grad*) where *log\_kernel* is a (N, N, D)-shaped tensor equal to the logarithm of the kernel and *kernel\_grad* is a (N, N, D)-shaped tensor where the entry (n, m, d) represents the derivative of *log\_kernel* w.r.t.  $x_{\{m,d\}}$ , where  $x_{\{m,d\}}$  is the  $d^{\text{th}}$  dimension of particle m.

**vectorize** (fn, num\_particles, max\_plate\_nesting)

## 4.7 Likelihood free methods

**class EnergyDistance** (beta=1.0, prior\_scale=0.0, num\_particles=2, max\_plate\_nesting=inf)

Bases: `object`

Posterior predictive energy distance [1,2] with optional Bayesian regularization by the prior.

Let  $p(x,z)=p(z)p(x|z)$  be the model,  $q(z|x)$  be the guide. Then given data  $x$  and drawing an iid pair of samples  $(Z, X)$  and  $(Z', X')$  (where  $Z$  is latent and  $X$  is the posterior predictive),

$$\begin{aligned} Z &\sim q(z|x); & X &\sim p(x|Z) \\ Z' &\sim q(z|x); & X' &\sim p(x|Z') \\ \text{loss} &= \mathbb{E}_X \|X - x\|^\beta - \frac{1}{2} \mathbb{E}_{X, X'} \|X - X'\|^\beta - \lambda \mathbb{E}_Z \log p(Z) \end{aligned}$$

This is a likelihood-free inference algorithm, and can be used for likelihoods without tractable density functions. The  $\beta$  energy distance is a robust loss functions, and is well defined for any distribution with finite fractional moment  $\mathbb{E}[\|X\|^\beta]$ .

This requires static model structure, a fully reparametrized guide, and reparametrized likelihood distributions in the model. Model latent distributions may be non-reparametrized.

## References

- [1] **Gabor J. Szekely, Maria L. Rizzo (2003)** Energy Statistics: A Class of Statistics Based on Distances.
- [2] **Tilmann Gneiting, Adrian E. Raftery (2007)** Strictly Proper Scoring Rules, Prediction, and Estimation.  
<https://www.stat.washington.edu/raftery/Research/PDF/Gneiting2007jasa.pdf>

## Parameters

- **beta** (*float*) – Exponent  $\beta$  from  $[1,2]$ . The loss function is strictly proper for distributions with finite *beta*-absolute moment  $E[\|X\|^\beta]$ . Thus for heavy tailed distributions *beta* should be small, e.g. for *Cauchy* distributions,  $\beta < 1$  is strictly proper. Defaults to 1. Must be in the open interval  $(0,2)$ .
- **prior\_scale** (*float*) – Nonnegative scale for prior regularization. Model parameters are trained only if this is positive. If zero (default), then model log densities will not be computed (guide log densities are never computed).
- **num\_particles** (*int*) – The number of particles/samples used to form the gradient estimators. Must be at least 2.
- **max\_plate\_nesting** (*int*) – Optional bound on max number of nested *pyro.plate()* contexts. If omitted, this will guess a valid value by running the (model,guide) pair once.

**\_\_call\_\_** (*model*, *guide*, *\*args*, *\*\*kwargs*)

Computes the surrogate loss that can be differentiated with autograd to produce gradient estimates for the model and guide parameters.

**loss** (*\*args*, *\*\*kwargs*)

Not implemented. Added for compatibility with unit tests only.

## 4.8 Discrete Inference

**infer\_discrete** (*fn=None*, *first\_available\_dim=None*, *temperature=1*)

A routine that samples discrete sites marked with `site["infer"]["enumerate"] = "parallel"` from the posterior, conditioned on observations.

Example:

```

@infer_discrete(first_available_dim=-1, temperature=0)
@config_enumerate
def viterbi_decoder(data, hidden_dim=10):
    transition = 0.3 / hidden_dim + 0.7 * torch.eye(hidden_dim)
    means = torch.arange(float(hidden_dim))
    states = [0]
    for t in pyro.markov(range(len(data))):
        states.append(pyro.sample("states_{}".format(t),
                                dist.Categorical(transition[states[-1]])))
        pyro.sample("obs_{}".format(t),
                    dist.Normal(means[states[-1]], 1.),
                    obs=data[t])
    return states # returns maximum likelihood states

```

### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **first\_available\_dim** (*int*) – The first tensor dimension (counting from the right) that is available for parallel enumeration. This dimension and all dimensions left may be used internally by Pyro. This should be a negative integer.
- **temperature** (*int*) – Either 1 (sample via forward-filter backward-sample) or 0 (optimize via Viterbi-like MAP inference). Defaults to 1 (sample).

```

class TraceEnumSample_ELBO (num_particles=1,                                max_plate_nesting=inf,
                           max_iarange_nesting=None,                    vectorize_particles=False,
                           strict_enumeration_warning=True,              ignore_jit_warnings=False,
                           jit_options=None, retain_graph=None, tail_adaptive_beta=-1.0)

```

Bases: `pyro.infer.traceenum_elbo.TraceEnum_ELBO`

This extends `TraceEnum_ELBO` to make it cheaper to sample from discrete latent states during SVI.

The following are equivalent but the first is cheaper, sharing work between the computations of `loss` and `z`:

```

# Version 1.
elbo = TraceEnumSample_ELBO(max_plate_nesting=1)
loss = elbo.loss(*args, **kwargs)
z = elbo.sample_saved()

# Version 2.
elbo = TraceEnum_ELBO(max_plate_nesting=1)
loss = elbo.loss(*args, **kwargs)
guide_trace = poutine.trace(guide).get_trace(*args, **kwargs)
z = infer_discrete(poutine.replay(model, guide_trace),
                  first_available_dim=-2) (*args, **kwargs)

```

**sample\_saved()**

Generate latent samples while reusing work from `SVI.step()`.

## 4.9 Inference Utilities

```

class Predictive (model, posterior_samples=None, guide=None, num_samples=None, return_sites=(),
                 parallel=False)
Bases: torch.nn.modules.module.Module

```



EXPERIMENTAL class used to construct predictive distribution. The predictive distribution is obtained by running the *model* conditioned on latent samples from *posterior\_samples*. If a *guide* is provided, then posterior samples from all the latent sites are also returned.

**Warning:** The interface for the *Predictive* class is experimental, and might change in the future.

#### Parameters

- **model** – Python callable containing Pyro primitives.
- **posterior\_samples** (*dict*) – dictionary of samples from the posterior.
- **guide** (*callable*) – optional guide to get posterior samples of sites not present in *posterior\_samples*.
- **num\_samples** (*int*) – number of samples to draw from the predictive distribution. This argument has no effect if *posterior\_samples* is non-empty, in which case, the leading dimension size of samples in *posterior\_samples* is used.
- **return\_sites** (*list, tuple, or set*) – sites to return; by default only sample sites not present in *posterior\_samples* are returned.
- **parallel** (*bool*) – predict in parallel by wrapping the existing model in an outermost *plate* messenger. Note that this requires that the model has all batch dims correctly annotated via *plate*. Default is *False*.

**call** (\*args, \*\*kwargs)

Method that calls *forward()* and returns parameter values of the guide as a *tuple* instead of a *dict*, which is a requirement for JIT tracing. Unlike *forward()*, this method can be traced by `torch.jit.trace_module()`.

**Warning:** This method may be removed once PyTorch JIT tracer starts accepting *dict* as valid return types. See [issue](#).

**forward** (\*args, \*\*kwargs)

Returns dict of samples from the predictive distribution. By default, only sample sites not contained in *posterior\_samples* are returned. This can be modified by changing the *return\_sites* keyword argument of this *Predictive* instance.

#### Parameters

- **args** – model arguments.
- **kwargs** – model keyword arguments.

**get\_samples** (\*args, \*\*kwargs)

**get\_vectorized\_trace** (\*args, \*\*kwargs)

Returns a single vectorized *trace* from the predictive distribution. Note that this requires that the model has all batch dims correctly annotated via *plate*.

#### Parameters

- **args** – model arguments.
- **kwargs** – model keyword arguments.

```
class EmpiricalMarginal (trace_posterior, sites=None, validate_args=None)
```

Bases: `pyro.distributions.empirical.Empirical`

Marginal distribution over a single site (or multiple, provided they have the same shape) from the `TracePosterior`'s model.

---

**Note:** If multiple sites are specified, they must have the same tensor shape. Samples from each site will be stacked and stored within a single tensor. See [Empirical](#). To hold the marginal distribution of sites having different shapes, use [Marginals](#) instead.

---

#### Parameters

- **trace\_posterior** (`TracePosterior`) – a `TracePosterior` instance representing a Monte Carlo posterior.
- **sites** (`list`) – optional list of sites for which we need to generate the marginal distribution.

```
class Marginals (trace_posterior, sites=None, validate_args=None)
```

Bases: `object`

Holds the marginal distribution over one or more sites from the `TracePosterior`'s model. This is a convenience container class, which can be extended by `TracePosterior` subclasses. e.g. for implementing diagnostics.

#### Parameters

- **trace\_posterior** (`TracePosterior`) – a `TracePosterior` instance representing a Monte Carlo posterior.
- **sites** (`list`) – optional list of sites for which we need to generate the marginal distribution.

#### `empirical`

A dictionary of sites' names and their corresponding [EmpiricalMarginal](#) distribution.

**Type** `OrderedDict`

**support** (`flatten=False`)

Gets support of this marginal distribution.

**Parameters** **flatten** (`bool`) – A flag to decide if we want to flatten `batch_shape` when the marginal distribution is collected from the posterior with `num_chains > 1`. Defaults to `False`.

**Returns** a dict with keys are sites' names and values are sites' supports.

**Return type** `OrderedDict`

```
class TracePosterior (num_chains=1)
```

Bases: `object`

Abstract `TracePosterior` object from which posterior inference algorithms inherit. When run, collects a bag of execution traces from the approximate posterior. This is designed to be used by other utility classes like [EmpiricalMarginal](#), that need access to the collected execution traces.

**information\_criterion** (`pointwise=False`)

Computes information criterion of the model. Currently, returns only “Widely Applicable/Watanabe-Akaike Information Criterion” (WAIC) and the corresponding effective number of parameters.

Reference:

[1] *Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC*, Aki Vehtari, Andrew Gelman, and Jonah Gabry

**Parameters** `pointwise` (*bool*) – a flag to decide if we want to get a vectorized WAIC or not. When `pointwise=False`, returns the sum.

**Returns** a dictionary containing values of WAIC and its effective number of parameters.

**Return type** `OrderedDict`

**marginal** (*sites=None*)

Generates the marginal distribution of this posterior.

**Parameters** `sites` (*list*) – optional list of sites for which we need to generate the marginal distribution.

**Returns** A *Marginals* class instance.

**Return type** *Marginals*

**run** (*\*args, \*\*kwargs*)

Calls `self._traces` to populate execution traces from a stochastic Pyro model.

**Parameters**

- **args** – optional args taken by `self._traces`.
- **kwargs** – optional keywords args taken by `self._traces`.

**class** `TracePredictive` (*model, posterior, num\_samples, keep\_sites=None*)

Bases: `pyro.infer.abstract_infer.TracePosterior`

**Warning:** This class is deprecated and will be removed in a future release. Use the *Predictive* class instead.

Generates and holds traces from the posterior predictive distribution, given model execution traces from the approximate posterior. This is achieved by constraining latent sites to randomly sampled parameter values from the model execution traces and running the model forward to generate traces with new response (“\_RETURN”) sites. :param model: arbitrary Python callable containing Pyro primitives. :param TracePosterior posterior: trace posterior instance holding samples from the model’s approximate posterior. :param int num\_samples: number of samples to generate. :param keep\_sites: The sites which should be sampled from posterior distribution (default: all)

**marginal** (*sites=None*)

Gets marginal distribution for this predictive posterior distribution.

## 4.10 MCMC

### 4.10.1 MCMC

**class** `MCMC` (*kernel, num\_samples, warmup\_steps=None, initial\_params=None, num\_chains=1, hook\_fn=None, mp\_context=None, disable\_progbar=False, disable\_validation=True, transforms=None*)

Bases: `object`

Wrapper class for Markov Chain Monte Carlo algorithms. Specific MCMC algorithms are `TraceKernel` instances and need to be supplied as a `kernel` argument to the constructor.

---

**Note:** The case of `num_chains > 1` uses python multiprocessing to run parallel chains in multiple processes. This goes with the usual caveats around multiprocessing in python, e.g. the model used to initialize the `kernel` must be serializable via `pickle`, and the performance / constraints will be platform dependent (e.g. only the “spawn” context is available in Windows). This has also not been extensively tested on the Windows platform.

---

### Parameters

- **kernel** – An instance of the `TraceKernel` class, which when given an execution trace returns another sample trace from the target (posterior) distribution.
- **num\_samples** (*int*) – The number of samples that need to be generated, excluding the samples discarded during the warmup phase.
- **warmup\_steps** (*int*) – Number of warmup iterations. The samples generated during the warmup phase are discarded. If not provided, default is half of `num_samples`.
- **num\_chains** (*int*) – Number of MCMC chains to run in parallel. Depending on whether `num_chains` is 1 or more than 1, this class internally dispatches to either `_UnarySampler` or `_MultiSampler`.
- **initial\_params** (*dict*) – dict containing initial tensors in unconstrained space to initiate the markov chain. The leading dimension’s size must match that of `num_chains`. If not specified, parameter values will be sampled from the prior.
- **hook\_fn** – Python callable that takes in (*kernel*, *samples*, *stage*, *i*) as arguments. *stage* is either *sample* or *warmup* and *i* refers to the *i*’th sample for the given stage. This can be used to implement additional logging, or more generally, run arbitrary code per generated sample.
- **mp\_context** (*str*) – Multiprocessing context to use when `num_chains > 1`. Only applicable for Python 3.5 and above. Use `mp_context=“spawn”` for CUDA.
- **disable\_progbar** (*bool*) – Disable progress bar and diagnostics update.
- **disable\_validation** (*bool*) – Disables distribution validation check. This is disabled by default, since divergent transitions will lead to exceptions. Switch to `True` for debugging purposes.
- **transforms** (*dict*) – dictionary that specifies a transform for a sample site with constrained support to unconstrained space.

### `diagnostics()`

Gets some diagnostics statistics such as effective sample size, split Gelman-Rubin, or divergent transitions from the sampler.

### `get_samples(num_samples=None, group_by_chain=False)`

Get samples from the MCMC run, potentially resampling with replacement.

### Parameters

- **num\_samples** (*int*) – Number of samples to return. If `None`, all the samples from an MCMC chain are returned in their original ordering.
- **group\_by\_chain** (*bool*) – Whether to preserve the chain dimension. If `True`, all samples will have `num_chains` as the size of their leading dimension.

**Returns** dictionary of samples keyed by site name.

**run**

**summary** (*prob=0.9*)

Prints a summary table displaying diagnostics of samples obtained from posterior. The diagnostics displayed are mean, standard deviation, median, the 90% Credibility Interval, `effective_sample_size()`, `split_gelman_rubin()`.

**Parameters** `prob` (*float*) – the probability mass of samples within the credibility interval.

## 4.10.2 HMC

**class** `HMC` (*model=None, potential\_fn=None, step\_size=1, trajectory\_length=None, num\_steps=None, adapt\_step\_size=True, adapt\_mass\_matrix=True, full\_mass=False, transforms=None, max\_plate\_nesting=None, jit\_compile=False, jit\_options=None, ignore\_jit\_warnings=False, target\_accept\_prob=0.8*)

Bases: `pyro.infer.mcmc.mcmc_kernel.MCMCKernel`

Simple Hamiltonian Monte Carlo kernel, where `step_size` and `num_steps` need to be explicitly specified by the user.

### References

[1] *MCMC Using Hamiltonian Dynamics*, Radford M. Neal

### Parameters

- **model** – Python callable containing Pyro primitives.
- **potential\_fn** – Python callable calculating potential energy with input is a dict of real support parameters.
- **step\_size** (*float*) – Determines the size of a single step taken by the verlet integrator while computing the trajectory using Hamiltonian dynamics. If not specified, it will be set to 1.
- **trajectory\_length** (*float*) – Length of a MCMC trajectory. If not specified, it will be set to `step_size x num_steps`. In case `num_steps` is not specified, it will be set to  $2\pi$ .
- **num\_steps** (*int*) – The number of discrete steps over which to simulate Hamiltonian dynamics. The state at the end of the trajectory is returned as the proposal. This value is always equal to `int(trajectory_length / step_size)`.
- **adapt\_step\_size** (*bool*) – A flag to decide if we want to adapt `step_size` during warm-up phase using Dual Averaging scheme.
- **adapt\_mass\_matrix** (*bool*) – A flag to decide if we want to adapt mass matrix during warm-up phase using Welford scheme.
- **full\_mass** (*bool*) – A flag to decide if mass matrix is dense or diagonal.
- **transforms** (*dict*) – Optional dictionary that specifies a transform for a sample site with constrained support to unconstrained space. The transform should be invertible, and implement `log_abs_det_jacobian`. If not specified and the model has sites with constrained support, automatic transformations will be applied, as specified in `torch.distributions.constraint_registry`.
- **max\_plate\_nesting** (*int*) – Optional bound on max number of nested `pyro.plate()` contexts. This is required if model contains discrete sample sites that can be enumerated over in parallel.
- **jit\_compile** (*bool*) – Optional parameter denoting whether to use the PyTorch JIT to trace the log density computation, and use this optimized executable trace in the integrator.

- **jit\_options** (*dict*) – A dictionary contains optional arguments for `torch.jit.trace()` function.
- **ignore\_jit\_warnings** (*bool*) – Flag to ignore warnings from the JIT tracer when `jit_compile=True`. Default is False.
- **target\_accept\_prob** (*float*) – Increasing this value will lead to a smaller step size, hence the sampling will be slower and more robust. Default to 0.8.

---

**Note:** Internally, the mass matrix will be ordered according to the order of the names of latent variables, not the order of their appearance in the model.

---

Example:

```
>>> true_coefs = torch.tensor([1., 2., 3.])
>>> data = torch.randn(2000, 3)
>>> dim = 3
>>> labels = dist.Bernoulli(logits=(true_coefs * data).sum(-1)).sample()
>>>
>>> def model(data):
...     coefs_mean = torch.zeros(dim)
...     coefs = pyro.sample('beta', dist.Normal(coefs_mean, torch.ones(3)))
...     y = pyro.sample('y', dist.Bernoulli(logits=(coefs * data).sum(-1)),
... ↪obs=labels)
...     return y
>>>
>>> hmc_kernel = HMC(model, step_size=0.0855, num_steps=4)
>>> mcmc = MCMC(hmc_kernel, num_samples=500, warmup_steps=100)
>>> mcmc.run(data)
>>> mcmc.get_samples()['beta'].mean(0) # doctest: +SKIP
tensor([ 0.9819,  1.9258,  2.9737])
```

**cleanup()**

**clear\_cache()**

**diagnostics()**

**initial\_params**

**inverse\_mass\_matrix**

**logging()**

**num\_steps**

**sample(params)**

**setup(warmup\_steps, \*args, \*\*kwargs)**

**step\_size**

### 4.10.3 NUTS

```
class NUTS(model=None, potential_fn=None, step_size=1, adapt_step_size=True,
            adapt_mass_matrix=True, full_mass=False, use_multinomial_sampling=True, trans-
            forms=None, max_plate_nesting=None, jit_compile=False, jit_options=None, ig-
            nore_jit_warnings=False, target_accept_prob=0.8, max_tree_depth=10)
Bases: pyro.infer.mcmc.hmc.HMC
```

No-U-Turn Sampler kernel, which provides an efficient and convenient way to run Hamiltonian Monte Carlo. The number of steps taken by the integrator is dynamically adjusted on each call to `sample` to ensure an optimal length for the Hamiltonian trajectory [1]. As such, the samples generated will typically have lower autocorrelation than those generated by the `HMC` kernel. Optionally, the NUTS kernel also provides the ability to adapt step size during the warmup phase.

Refer to the [baseball example](#) to see how to do Bayesian inference in Pyro using NUTS.

## References

- [1] *The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo*, Matthew D. Hoffman, and Andrew Gelman.
- [2] *A Conceptual Introduction to Hamiltonian Monte Carlo*, Michael Betancourt
- [3] *Slice Sampling*, Radford M. Neal

## Parameters

- **model** – Python callable containing Pyro primitives.
- **potential\_fn** – Python callable calculating potential energy with input is a dict of real support parameters.
- **step\_size** (*float*) – Determines the size of a single step taken by the verlet integrator while computing the trajectory using Hamiltonian dynamics. If not specified, it will be set to 1.
- **adapt\_step\_size** (*bool*) – A flag to decide if we want to adapt step\_size during warm-up phase using Dual Averaging scheme.
- **adapt\_mass\_matrix** (*bool*) – A flag to decide if we want to adapt mass matrix during warm-up phase using Welford scheme.
- **full\_mass** (*bool*) – A flag to decide if mass matrix is dense or diagonal.
- **use\_multinomial\_sampling** (*bool*) – A flag to decide if we want to sample candidates along its trajectory using “multinomial sampling” or using “slice sampling”. Slice sampling is used in the original NUTS paper [1], while multinomial sampling is suggested in [2]. By default, this flag is set to True. If it is set to *False*, NUTS uses slice sampling.
- **transforms** (*dict*) – Optional dictionary that specifies a transform for a sample site with constrained support to unconstrained space. The transform should be invertible, and implement *log\_abs\_det\_jacobian*. If not specified and the model has sites with constrained support, automatic transformations will be applied, as specified in `torch.distributions.constraint_registry`.
- **max\_plate\_nesting** (*int*) – Optional bound on max number of nested `pyro.plate()` contexts. This is required if model contains discrete sample sites that can be enumerated over in parallel.
- **jit\_compile** (*bool*) – Optional parameter denoting whether to use the PyTorch JIT to trace the log density computation, and use this optimized executable trace in the integrator.
- **jit\_options** (*dict*) – A dictionary contains optional arguments for `torch.jit.trace()` function.
- **ignore\_jit\_warnings** (*bool*) – Flag to ignore warnings from the JIT tracer when `jit_compile=True`. Default is False.
- **target\_accept\_prob** (*float*) – Target acceptance probability of step size adaptation scheme. Increasing this value will lead to a smaller step size, so the sampling will be slower but more robust. Default to 0.8.

- **max\_tree\_depth** (*int*) – Max depth of the binary tree created during the doubling scheme of NUTS sampler. Default to 10.

Example:

```
>>> true_coefs = torch.tensor([1., 2., 3.])
>>> data = torch.randn(2000, 3)
>>> dim = 3
>>> labels = dist.Bernoulli(logits=(true_coefs * data).sum(-1)).sample()
>>>
>>> def model(data):
...     coefs_mean = torch.zeros(dim)
...     coefs = pyro.sample('beta', dist.Normal(coefs_mean, torch.ones(3)))
...     y = pyro.sample('y', dist.Bernoulli(logits=(coefs * data).sum(-1)),
... ↪obs=labels)
...     return y
>>>
>>> nuts_kernel = NUTS(model, adapt_step_size=True)
>>> mcmc = MCMC(nuts_kernel, num_samples=500, warmup_steps=300)
>>> mcmc.run(data)
>>> mcmc.get_samples()['beta'].mean(0) # doctest: +SKIP
tensor([ 0.9221,  1.9464,  2.9228])
```

**sample** (*params*)

#### 4.10.4 Utilities

**initialize\_model** (*model*, *model\_args*=(), *model\_kwargs*={}, *transforms*=None, *max\_plate\_nesting*=None, *jit\_compile*=False, *jit\_options*=None, *skip\_jit\_warnings*=False, *num\_chains*=1)

Given a Python callable with Pyro primitives, generates the following model-specific properties needed for inference using HMC/NUTS kernels:

- initial parameters to be sampled using a HMC kernel,
- a potential function whose input is a dict of parameters in unconstrained space,
- transforms to transform latent sites of *model* to unconstrained space,
- a prototype trace to be used in MCMC to consume traces from sampled parameters.

##### Parameters

- **model** – a Pyro model which contains Pyro primitives.
- **model\_args** (*tuple*) – optional args taken by *model*.
- **model\_kwargs** (*dict*) – optional kwargs taken by *model*.
- **transforms** (*dict*) – Optional dictionary that specifies a transform for a sample site with constrained support to unconstrained space. The transform should be invertible, and implement *log\_abs\_det\_jacobian*. If not specified and the model has sites with constrained support, automatic transformations will be applied, as specified in `torch.distributions.constraint_registry`.
- **max\_plate\_nesting** (*int*) – Optional bound on max number of nested *pyro.plate()* contexts. This is required if model contains discrete sample sites that can be enumerated over in parallel.



- **jit\_compile** (*bool*) – Optional parameter denoting whether to use the PyTorch JIT to trace the log density computation, and use this optimized executable trace in the integrator.
- **jit\_options** (*dict*) – A dictionary contains optional arguments for `torch.jit.trace()` function.
- **ignore\_jit\_warnings** (*bool*) – Flag to ignore warnings from the JIT tracer when `jit_compile=True`. Default is False.
- **num\_chains** (*int*) – Number of parallel chains. If `num_chains > 1`, the returned *initial\_params* will be a list with `num_chains` elements.

**Returns** a tuple of (*initial\_params*, *potential\_fn*, *transforms*, *prototype\_trace*)

**diagnostics** (*samples*, *group\_by\_chain=True*)

Gets diagnostics statistics such as effective sample size and split Gelman-Rubin using the samples drawn from the posterior distribution.

**Parameters**

- **samples** (*dict*) – dictionary of samples keyed by site name.
- **group\_by\_chain** (*bool*) – If True, each variable in *samples* will be treated as having shape `num_chains x num_samples x sample_shape`. Otherwise, the corresponding shape will be `num_samples x sample_shape` (i.e. without chain dimension).

**Returns** dictionary of diagnostic stats for each sample site.

## 4.11 Automatic Guide Generation

### 4.11.1 AutoGuide

**class** `AutoGuide` (*model*)

Bases: `pyro.nn.module.PyroModule`

Base class for automatic guides.

Derived classes must implement the `forward()` method, with the same *\*args*, *\*\*kwargs* as the base model.

Auto guides can be used individually or combined in an `AutoGuideList` object.

**Parameters** **model** (*callable*) – a pyro model

**call** (*\*args*, *\*\*kwargs*)

Method that calls `forward()` and returns parameter values of the guide as a *tuple* instead of a *dict*, which is a requirement for JIT tracing. Unlike `forward()`, this method can be traced by `torch.jit.trace_module()`.

**Warning:** This method may be removed once PyTorch JIT tracer starts accepting *dict* as valid return types. See *issue* <<https://github.com/pytorch/pytorch/issues/27743>>\_.

**median** (*\*args*, *\*\*kwargs*)

Returns the posterior median value of each latent variable.

**Returns** A dict mapping sample site name to median tensor.

**Return type** `dict`

**model**

**sample\_latent** (*\*\*kwargs*)

Samples an encoded latent given the same *\*args*, *\*\*kwargs* as the base model.

### 4.11.2 AutoGuideList

**class AutoGuideList** (*model*)

Bases: `pyro.infer.autoguide.guides.AutoGuide`, `torch.nn.modules.container.ModuleList`

Container class to combine multiple automatic guides.

Example usage:

```
guide = AutoGuideList(my_model)
guide.add(AutoDiagonalNormal(poutine.block(model, hide=["assignment"])))
guide.add(AutoDiscreteParallel(poutine.block(model, expose=["assignment"])))
svi = SVI(model, guide, optim, Trace_ELBO())
```

**Parameters** **model** (*callable*) – a Pyro model

**add** (*part*)

Deprecated alias for `append()`.

**append** (*part*)

Add an automatic guide for part of the model. The guide should have been created by blocking the model to restrict to a subset of sample sites. No two parts should operate on any one sample site.

**Parameters** **part** (`AutoGuide` or *callable*) – a partial guide to add

**forward** (*\*args*, *\*\*kwargs*)

A composite guide with the same *\*args*, *\*\*kwargs* as the base model.

**Returns** A dict mapping sample site name to sampled value.

**Return type** `dict`

**median** (*\*args*, *\*\*kwargs*)

Returns the posterior median value of each latent variable.

**Returns** A dict mapping sample site name to median tensor.

**Return type** `dict`

### 4.11.3 AutoCallable

**class AutoCallable** (*model*, *guide*, *median*=`<function AutoCallable.<lambda>>`)

Bases: `pyro.infer.autoguide.guides.AutoGuide`

`AutoGuide` wrapper for simple callable guides.

This is used internally for composing autoguides with custom user-defined guides that are simple callables, e.g.:

```
def my_local_guide(*args, **kwargs):
    ...

guide = AutoGuideList(model)
```

(continues on next page)

(continued from previous page)

```
guide.add(AutoDelta(poutine.block(model, expose=['my_global_param'])))
guide.add(my_local_guide) # automatically wrapped in an AutoCallable
```

To specify a median callable, you can instead:

```
def my_local_median(*args, **kwargs)
    ...

guide.add(AutoCallable(model, my_local_guide, my_local_median))
```

For more complex guides that need e.g. access to plates, users should instead subclass `AutoGuide`.

#### Parameters

- **model** (*callable*) – a Pyro model
- **guide** (*callable*) – a Pyro guide (typically over only part of the model)
- **median** (*callable*) – an optional callable returning a dict mapping sample site name to computed median tensor.

**forward** (\*args, \*\*kwargs)

### 4.11.4 AutoDelta

**class** `AutoDelta` (*model*, *init\_loc\_fn*=<function *init\_to\_median*>)

Bases: `pyro.infer.autoguide.guides.AutoGuide`

This implementation of `AutoGuide` uses Delta distributions to construct a MAP guide over the entire latent space. The guide does not depend on the model's `*args`, `**kwargs`.

---

**Note:** This class does MAP inference in constrained space.

---

Usage:

```
guide = AutoDelta(model)
svi = SVI(model, guide, ...)
```

Latent variables are initialized using `init_loc_fn()`. To change the default behavior, create a custom `init_loc_fn()` as described in [Initialization](#), for example:

```
def my_init_fn(site):
    if site["name"] == "level":
        return torch.tensor([-1., 0., 1.])
    if site["name"] == "concentration":
        return torch.ones(k)
    return init_to_sample(site)
```

#### Parameters

- **model** (*callable*) – A Pyro model.
- **init\_loc\_fn** (*callable*) – A per-site initialization function. See [Initialization](#) section for available functions.

**forward** (\*args, \*\*kwargs)

An automatic guide with the same \*args, \*\*kwargs as the base model.

**Returns** A dict mapping sample site name to sampled value.

**Return type** dict

**median** (\*args, \*\*kwargs)

Returns the posterior median value of each latent variable.

**Returns** A dict mapping sample site name to median tensor.

**Return type** dict

### 4.11.5 AutoContinuous

**class AutoContinuous** (model, init\_loc\_fn=<function init\_to\_median>)

Bases: pyro.infer.autoguide.guides.AutoGuide

Base class for implementations of continuous-valued Automatic Differentiation Variational Inference [1].

This uses `torch.distributions.transforms` to transform each constrained latent variable to an unconstrained space, then concatenate all variables into a single unconstrained latent variable. Each derived class implements a `get_posterior()` method returning a distribution over this single unconstrained latent variable.

Assumes model structure and latent dimension are fixed, and all latent variables are continuous.

**Parameters** **model** (*callable*) – a Pyro model

Reference:

[1] *Automatic Differentiation Variational Inference*, Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, David M. Blei

**Parameters**

- **model** (*callable*) – A Pyro model.
- **init\_loc\_fn** (*callable*) – A per-site initialization function. See [Initialization](#) section for available functions.

**forward** (\*args, \*\*kwargs)

An automatic guide with the same \*args, \*\*kwargs as the base model.

**Returns** A dict mapping sample site name to sampled value.

**Return type** dict

**get\_posterior** (\*args, \*\*kwargs)

Returns the posterior distribution.

**median** (\*args, \*\*kwargs)

Returns the posterior median value of each latent variable.

**Returns** A dict mapping sample site name to median tensor.

**Return type** dict

**quantiles** (quantiles, \*args, \*\*kwargs)

Returns posterior quantiles each latent variable. Example:

```
print(guide.quantiles([0.05, 0.5, 0.95]))
```

**Parameters** `quantiles` (*torch.Tensor* or *list*) – A list of requested quantiles between 0 and 1.

**Returns** A dict mapping sample site name to a list of quantile values.

**Return type** `dict`

**sample\_latent** (*\*args, \*\*kwargs*)

Samples an encoded latent given the same *\*args, \*\*kwargs* as the base model.

#### 4.11.6 AutoMultivariateNormal

**class** `AutoMultivariateNormal` (*model, init\_loc\_fn=<function init\_to\_median>, init\_scale=0.1*)

Bases: `pyro.infer.autoguide.guides.AutoContinuous`

This implementation of `AutoContinuous` uses a Cholesky factorization of a Multivariate Normal distribution to construct a guide over the entire latent space. The guide does not depend on the model's *\*args, \*\*kwargs*.

Usage:

```
guide = AutoMultivariateNormal(model)
svi = SVI(model, guide, ...)
```

By default the mean vector is initialized by `init_loc_fn()` and the Cholesky factor is initialized to the identity times a small factor.

##### Parameters

- **model** (*callable*) – A generative model.
- **init\_loc\_fn** (*callable*) – A per-site initialization function. See [Initialization](#) section for available functions.
- **init\_scale** (*float*) – Initial scale for the standard deviation of each (unconstrained transformed) latent variable.

**get\_posterior** (*\*args, \*\*kwargs*)

Returns a MultivariateNormal posterior distribution.

#### 4.11.7 AutoDiagonalNormal

**class** `AutoDiagonalNormal` (*model, init\_loc\_fn=<function init\_to\_median>, init\_scale=0.1*)

Bases: `pyro.infer.autoguide.guides.AutoContinuous`

This implementation of `AutoContinuous` uses a Normal distribution with a diagonal covariance matrix to construct a guide over the entire latent space. The guide does not depend on the model's *\*args, \*\*kwargs*.

Usage:

```
guide = AutoDiagonalNormal(model)
svi = SVI(model, guide, ...)
```

By default the mean vector is initialized to zero and the scale is initialized to the identity times a small factor.

##### Parameters

- **model** (*callable*) – A generative model.
- **init\_loc\_fn** (*callable*) – A per-site initialization function. See [Initialization](#) section for available functions.
- **init\_scale** (*float*) – Initial scale for the standard deviation of each (unconstrained transformed) latent variable.

**get\_posterior** (\*args, \*\*kwargs)  
Returns a diagonal Normal posterior distribution.

#### 4.11.8 AutoLowRankMultivariateNormal

**class AutoLowRankMultivariateNormal** (model, init\_loc\_fn=<function init\_to\_median>, init\_scale=0.1, rank=None)  
Bases: `pyro.infer.autoguide.guides.AutoContinuous`

This implementation of `AutoContinuous` uses a low rank plus diagonal Multivariate Normal distribution to construct a guide over the entire latent space. The guide does not depend on the model's `*args`, `**kwargs`.

Usage:

```
guide = AutoLowRankMultivariateNormal(model, rank=10)
svi = SVI(model, guide, ...)
```

By default the `cov_diag` is initialized to a small constant and the `cov_factor` is initialized randomly such that on average `cov_factor.matmul(cov_factor.t())` has the same scale as `cov_diag`.

##### Parameters

- **model** (*callable*) – A generative model.
- **rank** (*int or None*) – The rank of the low-rank part of the covariance matrix. Defaults to approximately `sqrt(latent dim)`.
- **init\_loc\_fn** (*callable*) – A per-site initialization function. See [Initialization](#) section for available functions.
- **init\_scale** (*float*) – Approximate initial scale for the standard deviation of each (unconstrained transformed) latent variable.

**get\_posterior** (\*args, \*\*kwargs)  
Returns a LowRankMultivariateNormal posterior distribution.

#### 4.11.9 AutoIAFNormal

**class AutoIAFNormal** (model, hidden\_dim=None, init\_loc\_fn=<function init\_to\_median>)  
Bases: `pyro.infer.autoguide.guides.AutoContinuous`

This implementation of `AutoContinuous` uses a Diagonal Normal distribution transformed via a `AffineAutoregressive` to construct a guide over the entire latent space. The guide does not depend on the model's `*args`, `**kwargs`.

Usage:

```
guide = AutoIAFNormal(model, hidden_dim=latent_dim)
svi = SVI(model, guide, ...)
```

##### Parameters

- **model** (*callable*) – a generative model
- **hidden\_dim** (*int*) – number of hidden dimensions in the IAF
- **init\_loc\_fn** (*callable*) – A per-site initialization function. See [Initialization](#) section for available functions.

**get\_posterior** (\*args, \*\*kwargs)

Returns a diagonal Normal posterior distribution transformed by InverseAutoregressiveFlow.

#### 4.11.10 AutoLaplaceApproximation

**class AutoLaplaceApproximation** (model, init\_loc\_fn=<function init\_to\_median>)

Bases: `pyro.infer.autoguide.guides.AutoContinuous`

Laplace approximation (quadratic approximation) approximates the posterior  $\log p(z|x)$  by a multivariate normal distribution in the unconstrained space. Under the hood, it uses Delta distributions to construct a MAP guide over the entire (unconstrained) latent space. Its covariance is given by the inverse of the hessian of  $-\log p(x, z)$  at the MAP point of  $z$ .

Usage:

```
delta_guide = AutoLaplaceApproximation(model)
svi = SVI(model, delta_guide, ...)
# ...then train the delta_guide...
guide = delta_guide.laplace_approximation()
```

By default the mean vector is initialized to an empirical prior median.

##### Parameters

- **model** (*callable*) – a generative model
- **init\_loc\_fn** (*callable*) – A per-site initialization function. See [Initialization](#) section for available functions.

**get\_posterior** (\*args, \*\*kwargs)

Returns a Delta posterior distribution for MAP inference.

**laplace\_approximation** (\*args, \*\*kwargs)

Returns a `AutoMultivariateNormal` instance whose posterior's *loc* and *scale\_tril* are given by Laplace approximation.

#### 4.11.11 AutoDiscreteParallel

**class AutoDiscreteParallel** (model)

Bases: `pyro.infer.autoguide.guides.AutoGuide`

A discrete mean-field guide that learns a latent discrete distribution for each discrete site in the model.

**forward** (\*args, \*\*kwargs)

An automatic guide with the same \*args, \*\*kwargs as the base model.

**Returns** A dict mapping sample site name to sampled value.

**Return type** `dict`

### 4.11.12 Initialization

The `pyro.infer.autoguide.initialization` module contains initialization functions for automatic guides.

The standard interface for initialization is a function that inputs a Pyro trace `site` dict and returns an appropriately sized `value` to serve as an initial constrained value for a guide estimate.

**`init_to_feasible`** (*site*)

Initialize to an arbitrary feasible point, ignoring distribution parameters.

**`init_to_sample`** (*site*)

Initialize to a random sample from the prior.

**`init_to_median`** (*site*, *num\_samples=15*)

Initialize to the prior median; fallback to a feasible point if median is undefined.

**`init_to_mean`** (*site*)

Initialize to the prior mean; fallback to median if mean is undefined.

**`class InitMessenger`** (*init\_fn*)

Bases: `pyro.poutine.messenger.Messenger`

Initializes a site by replacing `.sample()` calls with values drawn from an initialization strategy. This is mainly for internal use by autoguide classes.

**Parameters** `init_fn` (*callable*) – An initialization function.



## 5.1 PyTorch Distributions

Most distributions in Pyro are thin wrappers around PyTorch distributions. For details on the PyTorch distribution interface, see `torch.distributions.distribution.Distribution`. For differences between the Pyro and PyTorch interfaces, see *TorchDistributionMixin*.

### 5.1.1 Bernoulli

```
class Bernoulli (probs=None, logits=None, validate_args=None)  
    Wraps torch.distributions.bernoulli.Bernoulli with TorchDistributionMixin.
```

### 5.1.2 Beta

```
class Beta (concentration1, concentration0, validate_args=None)  
    Wraps torch.distributions.beta.Beta with TorchDistributionMixin.
```

### 5.1.3 Binomial

```
class Binomial (total_count=1, probs=None, logits=None, validate_args=None)  
    Wraps torch.distributions.binomial.Binomial with TorchDistributionMixin.
```

### 5.1.4 Categorical

```
class Categorical (probs=None, logits=None, validate_args=None)  
    Wraps torch.distributions.categorical.Categorical with TorchDistributionMixin.
```

### 5.1.5 Cauchy

```
class Cauchy (loc, scale, validate_args=None)  
    Wraps torch.distributions.cauchy.Cauchy with TorchDistributionMixin.
```

### 5.1.6 Chi2

```
class Chi2 (df, validate_args=None)  
    Wraps torch.distributions.chi2.Chi2 with TorchDistributionMixin.
```

### 5.1.7 Dirichlet

```
class Dirichlet (concentration, validate_args=None)  
    Wraps torch.distributions.dirichlet.Dirichlet with TorchDistributionMixin.
```

### 5.1.8 Exponential

```
class Exponential (rate, validate_args=None)  
    Wraps torch.distributions.exponential.Exponential with TorchDistributionMixin.
```

### 5.1.9 ExponentialFamily

```
class ExponentialFamily (batch_shape=torch.Size([]), event_shape=torch.Size([]),  
                          validate_args=None)  
    Wraps torch.distributions.exp_family.ExponentialFamily with  
    TorchDistributionMixin.
```

### 5.1.10 FisherSnedecor

```
class FisherSnedecor (df1, df2, validate_args=None)  
    Wraps torch.distributions.fishersnedecor.FisherSnedecor with  
    TorchDistributionMixin.
```

### 5.1.11 Gamma

```
class Gamma (concentration, rate, validate_args=None)  
    Wraps torch.distributions.gamma.Gamma with TorchDistributionMixin.
```

### 5.1.12 Geometric

```
class Geometric (probs=None, logits=None, validate_args=None)  
    Wraps torch.distributions.geometric.Geometric with TorchDistributionMixin.
```

### 5.1.13 Gumbel

```
class Gumbel (loc, scale, validate_args=None)  
    Wraps torch.distributions.gumbel.Gumbel with TorchDistributionMixin.
```

### 5.1.14 HalfCauchy

**class HalfCauchy** (*scale, validate\_args=None*)  
Wraps `torch.distributions.half_cauchy.HalfCauchy` with *TorchDistributionMixin*.

### 5.1.15 HalfNormal

**class HalfNormal** (*scale, validate\_args=None*)  
Wraps `torch.distributions.half_normal.HalfNormal` with *TorchDistributionMixin*.

### 5.1.16 Independent

**class Independent** (*base\_distribution, reinterpreted\_batch\_ndims, validate\_args=None*)  
Wraps `torch.distributions.independent.Independent` with *TorchDistributionMixin*.

### 5.1.17 Laplace

**class Laplace** (*loc, scale, validate\_args=None*)  
Wraps `torch.distributions.laplace.Laplace` with *TorchDistributionMixin*.

### 5.1.18 LogNormal

**class LogNormal** (*loc, scale, validate\_args=None*)  
Wraps `torch.distributions.log_normal.LogNormal` with *TorchDistributionMixin*.

### 5.1.19 LogisticNormal

**class LogisticNormal** (*loc, scale, validate\_args=None*)  
Wraps `torch.distributions.logistic_normal.LogisticNormal` with *TorchDistributionMixin*.

### 5.1.20 LowRankMultivariateNormal

**class LowRankMultivariateNormal** (*loc, cov\_factor, cov\_diag, validate\_args=None*)  
Wraps `torch.distributions.lowrank_multivariate_normal.LowRankMultivariateNormal` with *TorchDistributionMixin*.

### 5.1.21 Multinomial

**class Multinomial** (*total\_count=1, probs=None, logits=None, validate\_args=None*)  
Wraps `torch.distributions.multinomial.Multinomial` with *TorchDistributionMixin*.

### 5.1.22 MultivariateNormal

**class MultivariateNormal** (*loc, covariance\_matrix=None, precision\_matrix=None, scale\_tril=None, validate\_args=None*)  
Wraps `torch.distributions.multivariate_normal.MultivariateNormal` with *TorchDistributionMixin*.

### 5.1.23 NegativeBinomial

```
class NegativeBinomial (total_count, probs=None, logits=None, validate_args=None)
    Wraps      torch.distributions.negative_binomial.NegativeBinomial      with
    TorchDistributionMixin.
```

### 5.1.24 Normal

```
class Normal (loc, scale, validate_args=None)
    Wraps torch.distributions.normal.Normal with TorchDistributionMixin.
```

### 5.1.25 OneHotCategorical

```
class OneHotCategorical (probs=None, logits=None, validate_args=None)
    Wraps      torch.distributions.one_hot_categorical.OneHotCategorical      with
    TorchDistributionMixin.
```

### 5.1.26 Pareto

```
class Pareto (scale, alpha, validate_args=None)
    Wraps torch.distributions.pareto.Pareto with TorchDistributionMixin.
```

### 5.1.27 Poisson

```
class Poisson (rate, validate_args=None)
    Wraps torch.distributions.poisson.Poisson with TorchDistributionMixin.
```

### 5.1.28 RelaxedBernoulli

```
class RelaxedBernoulli (temperature, probs=None, logits=None, validate_args=None)
    Wraps      torch.distributions.relaxed_bernoulli.RelaxedBernoulli      with
    TorchDistributionMixin.
```

### 5.1.29 RelaxedOneHotCategorical

```
class RelaxedOneHotCategorical (temperature, probs=None, logits=None, validate_args=None)
    Wraps torch.distributions.relaxed_categorical.RelaxedOneHotCategorical with
    TorchDistributionMixin.
```

### 5.1.30 StudentT

```
class StudentT (df, loc=0.0, scale=1.0, validate_args=None)
    Wraps torch.distributions.studentT.StudentT with TorchDistributionMixin.
```

### 5.1.31 TransformedDistribution

```
class TransformedDistribution (base_distribution, transforms, validate_args=None)
    Wraps torch.distributions.transformed_distribution.TransformedDistribution
    with TorchDistributionMixin.
```

### 5.1.32 Uniform

```
class Uniform (low, high, validate_args=None)
    Wraps torch.distributions.uniform.Uniform with TorchDistributionMixin.
```

### 5.1.33 Weibull

```
class Weibull (scale, concentration, validate_args=None)
    Wraps torch.distributions.weibull.Weibull with TorchDistributionMixin.
```

## 5.2 Pyro Distributions

### 5.2.1 Abstract Distribution

```
class Distribution
```

Bases: `object`

Base class for parameterized probability distributions.

Distributions in Pyro are stochastic function objects with `sample()` and `log_prob()` methods. Distribution are stochastic functions with fixed parameters:

```
d = dist.Bernoulli(param)
x = d()                                # Draws a random sample.
p = d.log_prob(x)                      # Evaluates log probability of x.
```

#### Implementing New Distributions:

Derived classes must implement the methods: `sample()`, `log_prob()`.

#### Examples:

Take a look at the [examples](#) to see how they interact with inference algorithms.

```
__call__ (*args, **kwargs)
```

Samples a random value (just an alias for `.sample(*args, **kwargs)`).

For tensor distributions, the returned tensor should have the same `.shape` as the parameters.

**Returns** A random value.

**Return type** `torch.Tensor`

```
enumerate_support (expand=True)
```

Returns a representation of the parametrized distribution's support, along the first dimension. This is implemented only by discrete distributions.

Note that this returns support values of all the batched RVs in lock-step, rather than the full cartesian product.

**Parameters** `expand (bool)` – whether to expand the result to a tensor of shape  $(n,) + \text{batch\_shape} + \text{event\_shape}$ . If false, the return value has unexpanded shape  $(n,) + (1,) * \text{len}(\text{batch\_shape}) + \text{event\_shape}$  which can be broadcasted to the full shape.

**Returns** An iterator over the distribution’s discrete support.

**Return type** iterator

`has_enumerate_support = False`

`has_rsample = False`

`log_prob (x, *args, **kwargs)`

Evaluates log probability densities for each of a batch of samples.

**Parameters** `x (torch.Tensor)` – A single value or a batch of values batched along axis 0.

**Returns** log probability densities as a one-dimensional `Tensor` with same batch size as value and params. The shape of the result should be `self.batch_size`.

**Return type** `torch.Tensor`

`sample (*args, **kwargs)`

Samples a random value.

For tensor distributions, the returned tensor should have the same `.shape` as the parameters, unless otherwise noted.

**Parameters** `sample_shape (torch.Size)` – the size of the iid batch to be drawn from the distribution.

**Returns** A random value or batch of random values (if parameters are batched). The shape of the result should be `self.shape()`.

**Return type** `torch.Tensor`

`score_parts (x, *args, **kwargs)`

Computes ingredients for stochastic gradient estimators of ELBO.

The default implementation is correct both for non-reparameterized and for fully reparameterized distributions. Partially reparameterized distributions should override this method to compute correct `.score_function` and `.entropy_term` parts.

**Parameters** `x (torch.Tensor)` – A single value or batch of values.

**Returns** A `ScoreParts` object containing parts of the ELBO estimator.

**Return type** `ScoreParts`

## 5.2.2 TorchDistributionMixin

**class** `TorchDistributionMixin`

Bases: `pyro.distributions.distribution.Distribution`

Mixin to provide Pyro compatibility for PyTorch distributions.

You should instead use *TorchDistribution* for new distribution classes.

This is mainly useful for wrapping existing PyTorch distributions for use in Pyro. Derived classes must first inherit from `torch.distributions.distribution.Distribution` and then inherit from *TorchDistributionMixin*.

**\_\_call\_\_** (*sample\_shape=torch.Size([])*)

Samples a random value.

This is reparameterized whenever possible, calling `rsample()` for reparameterized distributions and `sample()` for non-reparameterized distributions.

**Parameters** **sample\_shape** (*torch.Size*) – the size of the iid batch to be drawn from the distribution.

**Returns** A random value or batch of random values (if parameters are batched). The shape of the result should be *self.shape()*.

**Return type** `torch.Tensor`

**event\_dim**

**Returns** Number of dimensions of individual events.

**Return type** `int`

**shape** (*sample\_shape=torch.Size([])*)

The tensor shape of samples from this distribution.

Samples are of shape:

```
d.shape(sample_shape) == sample_shape + d.batch_shape + d.event_shape
```

**Parameters** **sample\_shape** (*torch.Size*) – the size of the iid batch to be drawn from the distribution.

**Returns** Tensor shape of samples.

**Return type** `torch.Size`

**expand** (*batch\_shape, \_instance=None*)

Returns a new `ExpandedDistribution` instance with batch dimensions expanded to *batch\_shape*.

**Parameters**

- **batch\_shape** (*tuple*) – batch shape to expand to.
- **\_instance** – unused argument for compatibility with `torch.distributions.Distribution.expand()`

**Returns** an instance of `ExpandedDistribution`.

**Return type** `ExpandedDistribution`

**expand\_by** (*sample\_shape*)

Expands a distribution by adding *sample\_shape* to the left side of its *batch\_shape*.

To expand internal dims of *self.batch\_shape* from 1 to something larger, use `expand()` instead.

**Parameters** **sample\_shape** (*torch.Size*) – The size of the iid batch to be drawn from the distribution.

**Returns** An expanded version of this distribution.

**Return type** `ExpandedDistribution`

**reshape** (*sample\_shape=None, extra\_event\_dims=None*)

**to\_event** (*reinterpreted\_batch\_ndims=None*)

Reinterprets the *n* rightmost dimensions of this distributions *batch\_shape* as event dims, adding them to the left side of *event\_shape*.

Example:

```
>>> [d1.batch_shape, d1.event_shape]
[torch.Size([2, 3]), torch.Size([4, 5])]
>>> d2 = d1.to_event(1)
>>> [d2.batch_shape, d2.event_shape]
[torch.Size([2]), torch.Size([3, 4, 5])]
>>> d3 = d1.to_event(2)
>>> [d3.batch_shape, d3.event_shape]
[torch.Size([]), torch.Size([2, 3, 4, 5])]
```

**Parameters** `reinterpreted_batch_ndims` (*int*) – The number of batch dimensions to reinterpret as event dimensions.

**Returns** A reshaped version of this distribution.

**Return type** `pyro.distributions.torch.Independent`

**independent** (*reinterpreted\_batch\_ndims=None*)

**mask** (*mask*)

Masks a distribution by a boolean or boolean-valued tensor that is broadcastable to the distributions `batch_shape`.

**Parameters** `mask` (*bool or torch.Tensor*) – A boolean or boolean valued tensor.

**Returns** A masked copy of this distribution.

**Return type** `MaskedDistribution`

### 5.2.3 TorchDistribution

**class** `TorchDistribution` (*batch\_shape=torch.Size([]), event\_shape=torch.Size([]), validate\_args=None*)

Bases: `torch.distributions.distribution.Distribution`, `pyro.distributions.torch_distribution.TorchDistributionMixin`

Base class for PyTorch-compatible distributions with Pyro support.

This should be the base class for almost all new Pyro distributions.

---

**Note:** Parameters and data should be of type `Tensor` and all methods return type `Tensor` unless otherwise noted.

---

#### Tensor Shapes:

TorchDistributions provide a method `.shape()` for the tensor shape of samples:

```
x = d.sample(sample_shape)
assert x.shape == d.shape(sample_shape)
```

Pyro follows the same distribution shape semantics as PyTorch. It distinguishes between three different roles for tensor shapes of samples:

- *sample shape* corresponds to the shape of the iid samples drawn from the distribution. This is taken as an argument by the distribution's *sample* method.
- *batch shape* corresponds to non-identical (independent) parameterizations of the distribution, inferred from the distribution's parameter shapes. This is fixed for a distribution instance.



- *event shape* corresponds to the event dimensions of the distribution, which is fixed for a distribution class. These are collapsed when we try to score a sample from the distribution via *d.log\_prob(x)*.

These shapes are related by the equation:

```
assert d.shape(sample_shape) == sample_shape + d.batch_shape + d.event_shape
```

Distributions provide a vectorized `log_prob()` method that evaluates the log probability density of each event in a batch independently, returning a tensor of shape `sample_shape + d.batch_shape`:

```
x = d.sample(sample_shape)
assert x.shape == d.shape(sample_shape)
log_p = d.log_prob(x)
assert log_p.shape == sample_shape + d.batch_shape
```

### Implementing New Distributions:

Derived classes must implement the methods `sample()` (or `rsample()` if `.has_rsample == True`) and `log_prob()`, and must implement the properties `batch_shape`, and `event_shape`. Discrete classes may also implement the `enumerate_support()` method to improve gradient estimates and set `.has_enumerate_support = True`.

**expand** (*batch\_shape*, *\_instance=None*)

Returns a new `ExpandedDistribution` instance with batch dimensions expanded to *batch\_shape*.

#### Parameters

- **batch\_shape** (*tuple*) – batch shape to expand to.
- **\_instance** – unused argument for compatibility with `torch.distributions.Distribution.expand()`

**Returns** an instance of *ExpandedDistribution*.

**Return type** `ExpandedDistribution`

## 5.2.4 AVFMultivariateNormal

**class** `AVFMultivariateNormal` (*loc*, *scale\_tril*, *control\_var*)

Bases: `pyro.distributions.torch.MultivariateNormal`

Multivariate normal (Gaussian) distribution with transport equation inspired control variates (adaptive velocity fields).

A distribution over vectors in which all the elements have a joint Gaussian density.

#### Parameters

- **loc** (*torch.Tensor*) – D-dimensional mean vector.
- **scale\_tril** (*torch.Tensor*) – Cholesky of Covariance matrix; D x D matrix.
- **control\_var** (*torch.Tensor*) – 2 x L x D tensor that parameterizes the control variate; L is an arbitrary positive integer. This parameter needs to be learned (i.e. adapted) to achieve lower variance gradients. In a typical use case this parameter will be adapted concurrently with the *loc* and *scale\_tril* that define the distribution.

Example usage:

```
control_var = torch.tensor(0.1 * torch.ones(2, 1, D), requires_grad=True)
opt_cv = torch.optim.Adam([control_var], lr=0.1, betas=(0.5, 0.999))

for _ in range(1000):
    d = AVFMultivariateNormal(loc, scale_tril, control_var)
    z = d.rsample()
    cost = torch.pow(z, 2.0).sum()
    cost.backward()
    opt_cv.step()
    opt_cv.zero_grad()
```

```
arg_constraints = {'control_var': Real(), 'loc': Real(), 'scale_tril': LowerTriangular()
rsample (sample_shape=torch.Size([]))
```

### 5.2.5 BetaBinomial

```
class BetaBinomial (concentration1, concentration0, total_count=1, validate_args=None)
```

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Compound distribution comprising of a beta-binomial pair. The probability of success (`probs` for the [Binomial](#) distribution) is unknown and randomly drawn from a [Beta](#) distribution prior to a certain number of Bernoulli trials given by `total_count`.

#### Parameters

- or `torch.Tensor concentration1 (float)` – 1st concentration parameter (alpha) for the Beta distribution.
- or `torch.Tensor concentration0 (float)` – 2nd concentration parameter (beta) for the Beta distribution.
- or `torch.Tensor total_count (int)` – number of Bernoulli trials.

```
arg_constraints = {'concentration0': GreaterThan(lower_bound=0.0), 'concentration1':
concentration0
concentration1
enumerate_support (expand=True)
expand (batch_shape, _instance=None)
has_enumerate_support = True
log_prob (value)
mean
sample (sample_shape=())
support
variance
```

### 5.2.6 ConditionalDistribution

```
class ConditionalDistribution
```

Bases: `abc.ABC`

**condition** (*context*)

**Return type** torch.distributions.Distribution

## 5.2.7 ConditionalTransformedDistribution

**class** ConditionalTransformedDistribution (*base\_dist, transforms*)

Bases: pyro.distributions.conditional.ConditionalDistribution

**condition** (*context*)

## 5.2.8 Delta

**class** Delta (*v, log\_density=0.0, event\_dim=0, validate\_args=None*)

Bases: pyro.distributions.torch\_distribution.TorchDistribution

Degenerate discrete distribution (a single point).

Discrete distribution that assigns probability one to the single element in its support. Delta distribution parameterized by a random choice should not be used with MCMC based inference, as doing so produces incorrect results.

### Parameters

- **v** (*torch.Tensor*) – The single support element.
- **log\_density** (*torch.Tensor*) – An optional density for this Delta. This is useful to keep the class of *Delta* distributions closed under differentiable transformation.
- **event\_dim** (*int*) – Optional event dimension, defaults to zero.

**arg\_constraints** = {'log\_density': Real(), 'v': Real() }

**expand** (*batch\_shape, \_instance=None*)

**has\_rsample** = True

**log\_prob** (*x*)

**mean**

**rsample** (*sample\_shape=torch.Size([])*)

**support** = Real()

**variance**

## 5.2.9 DirichletMultinomial

**class** DirichletMultinomial (*concentration, total\_count=1, is\_sparse=False, validate\_args=None*)

Bases: pyro.distributions.torch\_distribution.TorchDistribution

Compound distribution comprising of a dirichlet-multinomial pair. The probability of classes (probs for the *Multinomial* distribution) is unknown and randomly drawn from a *Dirichlet* distribution prior to a certain number of Categorical trials given by *total\_count*.

### Parameters

- **or torch.Tensor concentration** (*float*) – concentration parameter (alpha) for the Dirichlet distribution.

- **or** `torch.Tensor total_count` (*int*) – number of Categorical trials.
- **is\_sparse** (*bool*) – Whether to assume value is mostly zero when computing `log_prob()`, which can speed up computation when data is sparse.

```
arg_constraints = {'concentration': GreaterThan(lower_bound=0.0), 'total_count': IntsOf(
    concentration
expand(batch_shape, _instance=None)
log_prob(value)
mean
sample(sample_shape=())
support
variance
```

### 5.2.10 DiscreteHMM

**class DiscreteHMM**(*initial\_logits, transition\_logits, observation\_dist, validate\_args=None*)  
Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Hidden Markov Model with discrete latent state and arbitrary observation distribution. This uses [1] to parallelize over time, achieving  $O(\log(\text{time}))$  parallel complexity.

The `event_shape` of this distribution includes time on the left:

```
event_shape = (num_steps,) + observation_dist.event_shape
```

This distribution supports any combination of homogeneous/heterogeneous time dependency of `transition_logits` and `observation_dist`. However, because time is included in this distribution's `event_shape`, the homogeneous+homogeneous case will have a broadcastable `event_shape` with `num_steps = 1`, allowing `log_prob()` to work with arbitrary length data:

```
# homogeneous + homogeneous case:
event_shape = (1,) + observation_dist.event_shape
```

#### References:

[1] Simo Sarkka, Angel F. Garcia-Fernandez (2019) “Temporal Parallelization of Bayesian Filters and Smoothers” <https://arxiv.org/pdf/1905.13002.pdf>

#### Parameters

- **initial\_logits** (*Tensor*) – A logits tensor for an initial categorical distribution over latent states. Should have rightmost size `state_dim` and be broadcastable to `batch_shape + (state_dim,)`.
- **transition\_logits** (*Tensor*) – A logits tensor for transition conditional distributions between latent states. Should have rightmost shape `(state_dim, state_dim)` (old, new), and be broadcastable to `batch_shape + (num_steps, state_dim, state_dim)`.
- **observation\_dist** (*Distribution*) – A conditional distribution of observed data conditioned on latent state. The `.batch_shape` should have rightmost size `state_dim` and be broadcastable to `batch_shape + (num_steps, state_dim)`. The `.event_shape` may be arbitrary.

```

arg_constraints = {'initial_logits': Real(), 'transition_logits': Real()}
expand (batch_shape, _instance=None)
filter (value)
    Compute posterior over final state given a sequence of observations.

    Parameters value (Tensor) – A sequence of observations.

    Returns A posterior distribution over latent states at the final time step. result.logits can
    then be used as initial_logits in a sequential Pyro model for prediction.

    Return type Categorical

log_prob (value)

```

### 5.2.11 EmpiricalDistribution

```

class Empirical (samples, log_weights, validate_args=None)
    Bases: pyro.distributions.torch_distribution.TorchDistribution

```

Empirical distribution associated with the sampled data. Note that the shape requirement for `log_weights` is that its shape must match the leftmost shape of `samples`. Samples are aggregated along the `aggregation_dim`, which is the rightmost dim of `log_weights`.

Example:

```

>>> emp_dist = Empirical(torch.randn(2, 3, 10), torch.ones(2, 3))
>>> emp_dist.batch_shape
torch.Size([2])
>>> emp_dist.event_shape
torch.Size([10])

```

```

>>> single_sample = emp_dist.sample()
>>> single_sample.shape
torch.Size([2, 10])
>>> batch_sample = emp_dist.sample((100,))
>>> batch_sample.shape
torch.Size([100, 2, 10])

```

```

>>> emp_dist.log_prob(single_sample).shape
torch.Size([2])
>>> # Vectorized samples cannot be scored by log_prob.
>>> with pyro.validation_enabled():
...     emp_dist.log_prob(batch_sample).shape
Traceback (most recent call last):
...
ValueError: ``value.shape`` must be torch.Size([2, 10])

```

#### Parameters

- **samples** (*torch.Tensor*) – samples from the empirical distribution.
- **log\_weights** (*torch.Tensor*) – log weights (optional) corresponding to the samples.

```

arg_constraints = {}

```

**enumerate\_support** (*expand=True*)  
See `pyro.distributions.torch_distribution.TorchDistribution.enumerate_support()`

**event\_shape**  
See `pyro.distributions.torch_distribution.TorchDistribution.event_shape()`

**has\_enumerate\_support = True**

**log\_prob** (*value*)  
Returns the log of the probability mass function evaluated at *value*. Note that this currently only supports scoring values with empty *sample\_shape*.  
  
Parameters **value** (*torch.Tensor*) – scalar or tensor value to be scored.

**log\_weights**

**mean**  
See `pyro.distributions.torch_distribution.TorchDistribution.mean()`

**sample** (*sample\_shape=torch.Size([])*)  
See `pyro.distributions.torch_distribution.TorchDistribution.sample()`

**sample\_size**  
Number of samples that constitute the empirical distribution.  
  
Return **int** number of samples collected.

**support = Real()**

**variance**  
See `pyro.distributions.torch_distribution.TorchDistribution.variance()`

### 5.2.12 FoldedDistribution

**class FoldedDistribution** (*base\_dist, validate\_args=None*)  
Bases: `pyro.distributions.torch.TransformedDistribution`  
  
Equivalent to `TransformedDistribution(base_dist, AbsTransform())`, but additionally supports `log_prob()`.  
  
Parameters **base\_dist** (*Distribution*) – The distribution to reflect.  
  
**expand** (*batch\_shape, \_instance=None*)  
  
**log\_prob** (*value*)  
  
**support = GreaterThan(lower\_bound=0.0)**

### 5.2.13 GammaPoisson

**class GammaPoisson** (*concentration, rate, validate\_args=None*)  
Bases: `pyro.distributions.torch_distribution.TorchDistribution`  
  
Compound distribution comprising of a gamma-poisson pair, also referred to as a gamma-poisson mixture. The *rate* parameter for the *Poisson* distribution is unknown and randomly drawn from a *Gamma* distribution.

---

**Note:** This can be treated as an alternate parametrization of the [NegativeBinomial](#) (`total_count`, `probs`) distribution, with  $concentration = total\_count$  and  $rate = (1 - probs) / probs$ .

---

### Parameters

- or **torch.Tensor concentration** (*float*) – shape parameter (alpha) of the Gamma distribution.
- or **torch.Tensor rate** (*float*) – rate parameter (beta) for the Gamma distribution.

```
arg_constraints = {'concentration': GreaterThan(lower_bound=0.0), 'rate': GreaterThan(
concentration
expand(batch_shape, _instance=None)
log_prob(value)
mean
rate
sample(sample_shape=())
support = IntegerGreaterThan(lower_bound=0)
variance
```

## 5.2.14 GaussianHMM

```
class GaussianHMM(initial_dist, transition_matrix, transition_dist, observation_matrix, observa-
tion_dist, validate_args=None)
Bases: pyro.distributions.torch_distribution.TorchDistribution
```

Hidden Markov Model with Gaussians for initial, transition, and observation distributions. This adapts [1] to parallelize over time to achieve  $O(\log(\text{time}))$  parallel complexity, however it differs in that it tracks the log normalizer to ensure `log_prob()` is differentiable.

This corresponds to the generative model:

```
z = initial_distribution.sample()
x = []
for t in range(num_events):
    z = z @ transition_matrix + transition_dist.sample()
    x.append(z @ observation_matrix + observation_dist.sample())
```

The `event_shape` of this distribution includes time on the left:

```
event_shape = (num_steps,) + observation_dist.event_shape
```

This distribution supports any combination of homogeneous/heterogeneous time dependency of `transition_dist` and `observation_dist`. However, because time is included in this distribution's `event_shape`, the homogeneous+homogeneous case will have a broadcastable `event_shape` with `num_steps = 1`, allowing `log_prob()` to work with arbitrary length data:

```
event_shape = (1, obs_dim) # homogeneous + homogeneous case
```

**References:**

[1] Simo Sarkka, Angel F. Garcia-Fernandez (2019) “Temporal Parallelization of Bayesian Filters and Smoothers” <https://arxiv.org/pdf/1905.13002.pdf>

#### Variables

- **hidden\_dim** (*int*) – The dimension of the hidden state.
- **obs\_dim** (*int*) – The dimension of the observed state.

#### Parameters

- **initial\_dist** (*MultivariateNormal*) – A distribution over initial states. This should have `batch_shape` broadcastable to `self.batch_shape`. This should have `event_shape` (`hidden_dim,`).
- **transition\_matrix** (*Tensor*) – A linear transformation of hidden state. This should have shape broadcastable to `self.batch_shape + (num_steps, hidden_dim, hidden_dim)` where the rightmost dims are ordered (`old, new`).
- **transition\_dist** (*MultivariateNormal*) – A process noise distribution. This should have `batch_shape` broadcastable to `self.batch_shape + (num_steps,)`. This should have `event_shape` (`hidden_dim,`).
- **observation\_matrix** (*Tensor*) – A linear transformation from hidden to observed state. This should have shape broadcastable to `self.batch_shape + (num_steps, hidden_dim, obs_dim)`.
- **observation\_dist** (*MultivariateNormal or Normal*) – An observation noise distribution. This should have `batch_shape` broadcastable to `self.batch_shape + (num_steps,)`. This should have `event_shape` (`obs_dim,`).

**arg\_constraints** = {}

**expand** (*batch\_shape, \_instance=None*)

**filter** (*value*)

Compute posterior over final state given a sequence of observations.

**Parameters** **value** (*Tensor*) – A sequence of observations.

**Returns** A posterior distribution over latent states at the final time step. `result` can then be used as `initial_dist` in a sequential Pyro model for prediction.

**Return type** *MultivariateNormal*

**log\_prob** (*value*)

### 5.2.15 GaussianMRF

**class GaussianMRF** (*initial\_dist, transition\_dist, observation\_dist, validate\_args=None*)

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Temporal Markov Random Field with Gaussian factors for initial, transition, and observation distributions. This adapts [1] to parallelize over time to achieve  $O(\log(\text{time}))$  parallel complexity, however it differs in that it tracks the log normalizer to ensure `log_prob()` is differentiable.

The `event_shape` of this distribution includes time on the left:

```
event_shape = (num_steps,) + observation_dist.event_shape
```



This distribution supports any combination of homogeneous/heterogeneous time dependency of `transition_dist` and `observation_dist`. However, because time is included in this distribution's `event_shape`, the homogeneous+homogeneous case will have a broadcastable `event_shape` with `num_steps = 1`, allowing `log_prob()` to work with arbitrary length data:

```
event_shape = (1, obs_dim) # homogeneous + homogeneous case
```

#### References:

[1] Simo Sarkka, Angel F. Garcia-Fernandez (2019) “Temporal Parallelization of Bayesian Filters and Smoothers” <https://arxiv.org/pdf/1905.13002.pdf>

#### Variables

- **hidden\_dim**(*int*) – The dimension of the hidden state.
- **obs\_dim**(*int*) – The dimension of the observed state.

#### Parameters

- **initial\_dist** (*MultivariateNormal*) – A distribution over initial states. This should have `batch_shape` broadcastable to `self.batch_shape`. This should have `event_shape (hidden_dim,)`.
- **transition\_dist** (*MultivariateNormal*) – A joint distribution factor over a pair of successive time steps. This should have `batch_shape` broadcastable to `self.batch_shape + (num_steps,)`. This should have `event_shape (hidden_dim + hidden_dim,)` (old+new).
- **observation\_dist** (*MultivariateNormal*) – A joint distribution factor over a hidden and an observed state. This should have `batch_shape` broadcastable to `self.batch_shape + (num_steps,)`. This should have `event_shape (hidden_dim + obs_dim,)`.

```
arg_constraints = {}
```

```
expand (batch_shape, _instance=None)
```

```
log_prob (value)
```

## 5.2.16 GaussianScaleMixture

```
class GaussianScaleMixture (coord_scale, component_logits, component_scale)
```

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Mixture of Normal distributions with zero mean and diagonal covariance matrices.

That is, this distribution is a mixture with  $K$  components, where each component distribution is a  $D$ -dimensional Normal distribution with zero mean and a  $D$ -dimensional diagonal covariance matrix. The  $K$  different covariance matrices are controlled by the parameters `coord_scale` and `component_scale`. That is, the covariance matrix of the  $k$ 'th component is given by

$$\text{Sigma}_{ii} = (\text{component\_scale}_k * \text{coord\_scale}_i) ** 2 \quad (i = 1, \dots, D)$$

where `component_scale_k` is a positive scale factor and `coord_scale_i` are positive scale parameters shared between all  $K$  components. The mixture weights are controlled by a  $K$ -dimensional vector of softmax logits, `component_logits`. This distribution implements pathwise derivatives for samples from the distribution. This distribution does not currently support batched parameters.

See reference [1] for details on the implementations of the pathwise derivative. Please consider citing this reference if you use the pathwise derivative in your research.

[1] Pathwise Derivatives for Multivariate Distributions, Martin Jankowiak & Theofanis Karaletsos. arXiv:1806.01856

Note that this distribution supports both even and odd dimensions, but the former should be more a bit higher precision, since it doesn't use any erfs in the backward call. Also note that this distribution does not support  $D = 1$ .

#### Parameters

- **coord\_scale** (*torch.tensor*) – D-dimensional vector of scales
- **component\_logits** (*torch.tensor*) – K-dimensional vector of logits
- **component\_scale** (*torch.tensor*) – K-dimensional vector of scale multipliers

```
arg_constraints = {'component_logits': Real(), 'component_scale': GreaterThan(lower_bound=0.0)}
has_rsample = True
log_prob(value)
rsample(sample_shape=torch.Size([]))
```

### 5.2.17 InverseGamma

```
class InverseGamma(concentration, rate, validate_args=None)
```

Bases: `pyro.distributions.torch.TransformedDistribution`

Creates an inverse-gamma distribution parameterized by *concentration* and *rate*.

$X \sim \text{Gamma}(\text{concentration}, \text{rate})$   $Y = 1/X \sim \text{InverseGamma}(\text{concentration}, \text{rate})$

#### Parameters

- **concentration** (*torch.Tensor*) – the concentration parameter (i.e. alpha).
- **rate** (*torch.Tensor*) – the rate parameter (i.e. beta).

```
arg_constraints = {'concentration': GreaterThan(lower_bound=0.0), 'rate': GreaterThan(lower_bound=0.0)}
concentration
expand(batch_shape, _instance=None)
has_rsample = True
rate
support = GreaterThan(lower_bound=0.0)
```

### 5.2.18 LKJCorrCholesky

```
class LKJCorrCholesky(d, eta, validate_args=None)
```

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Generates cholesky factors of correlation matrices using an LKJ prior.

The expected use is to combine it with a vector of variances and pass it to the `scale_tril` parameter of a multivariate distribution such as `MultivariateNormal`.

E.g., if  $\theta$  is a (positive) vector of covariances with the same dimensionality as this distribution, and  $\Omega$  is sampled from this distribution, `scale_tril=torch.mm(torch.diag(sqrt( $\theta$ )),  $\Omega$ )`

Note that the `event_shape` of this distribution is  $[d, d]$

---

**Note:** When using this distribution with HMC/NUTS, it is important to use a `step_size` such as `1e-4`. If not, you are likely to experience LAPACK errors regarding positive-definiteness.

---

For example usage, refer to [pyro/examples/lkj.py](#).

#### Parameters

- `d` (`int`) – Dimensionality of the matrix
- `eta` (`torch.Tensor`) – A single positive number parameterizing the distribution.

`arg_constraints = {'eta': GreaterThan(lower_bound=0.0)}`

`expand(batch_shape, _instance=None)`

`has_rsample = False`

`lkj_constant(eta, K)`

`log_prob(x)`

`sample(sample_shape=torch.Size([]))`

`support = CorrCholesky()`

### 5.2.19 MaskedMixture

**class MaskedMixture** (*mask, component0, component1, validate\_args=None*)

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

A masked deterministic mixture of two distributions.

This is useful when the mask is sampled from another distribution, possibly correlated across the batch. Often the mask can be marginalized out via enumeration.

Example:

```
change_point = pyro.sample("change_point",
                           dist.Categorical(torch.ones(len(data) + 1)),
                           infer={'enumerate': 'parallel'})
mask = torch.arange(len(data), dtype=torch.long) >= change_point
with pyro.plate("data", len(data)):
    pyro.sample("obs", MaskedMixture(mask, dist1, dist2), obs=data)
```

#### Parameters

- `mask` (`torch.Tensor`) – A byte tensor toggling between `component0` and `component1`.
- `component0` (`pyro.distributions.TorchDistribution`) – a distribution for batch elements `mask == 0`.
- `component1` (`pyro.distributions.TorchDistribution`) – a distribution for batch elements `mask == 1`.

`arg_constraints = {}`

```
expand(batch_shape)
has_rsample
log_prob(value)
mean
rsample(sample_shape=torch.Size([]))
sample(sample_shape=torch.Size([]))
support
variance
```

### 5.2.20 MixtureOfDiagNormals

```
class MixtureOfDiagNormals(locs, coord_scale, component_logits)
    Bases: pyro.distributions.torch_distribution.TorchDistribution
```

Mixture of Normal distributions with arbitrary means and arbitrary diagonal covariance matrices.

That is, this distribution is a mixture with  $K$  components, where each component distribution is a  $D$ -dimensional Normal distribution with a  $D$ -dimensional mean parameter and a  $D$ -dimensional diagonal covariance matrix. The  $K$  different component means are gathered into the  $K \times D$  dimensional parameter *locs* and the  $K$  different scale parameters are gathered into the  $K \times D$  dimensional parameter *coord\_scale*. The mixture weights are controlled by a  $K$ -dimensional vector of softmax logits, *component\_logits*. This distribution implements pathwise derivatives for samples from the distribution.

See reference [1] for details on the implementations of the pathwise derivative. Please consider citing this reference if you use the pathwise derivative in your research. Note that this distribution does not support dimension  $D = 1$ .

[1] Pathwise Derivatives for Multivariate Distributions, Martin Jankowiak & Theofanis Karaletsos. arXiv:1806.01856

#### Parameters

- **locs** (*torch.Tensor*) –  $K \times D$  mean matrix
- **coord\_scale** (*torch.Tensor*) –  $K \times D$  scale matrix
- **component\_logits** (*torch.Tensor*) –  $K$ -dimensional vector of softmax logits

```
arg_constraints = {'component_logits': Real(), 'coord_scale': GreaterThan(lower_bound)}
expand(batch_shape, _instance=None)
has_rsample = True
log_prob(value)
rsample(sample_shape=torch.Size([]))
```

### 5.2.21 MixtureOfDiagNormalsSharedCovariance

```
class MixtureOfDiagNormalsSharedCovariance(locs, coord_scale, component_logits)
    Bases: pyro.distributions.torch_distribution.TorchDistribution
```

Mixture of Normal distributions with diagonal covariance matrices.

That is, this distribution is a mixture with  $K$  components, where each component distribution is a  $D$ -dimensional Normal distribution with a  $D$ -dimensional mean parameter `loc` and a  $D$ -dimensional diagonal covariance matrix specified by a scale parameter `coord_scale`. The  $K$  different component means are gathered into the parameter `locs` and the scale parameter is shared between all  $K$  components. The mixture weights are controlled by a  $K$ -dimensional vector of softmax logits, `component_logits`. This distribution implements pathwise derivatives for samples from the distribution.

See reference [1] for details on the implementations of the pathwise derivative. Please consider citing this reference if you use the pathwise derivative in your research. Note that this distribution does not support dimension  $D = 1$ .

[1] Pathwise Derivatives for Multivariate Distributions, Martin Jankowiak & Theofanis Karaletsos. arXiv:1806.01856

#### Parameters

- `locs` (*torch.Tensor*) –  $K \times D$  mean matrix
- `coord_scale` (*torch.Tensor*) – shared  $D$ -dimensional scale vector
- `component_logits` (*torch.Tensor*) –  $K$ -dimensional vector of softmax logits

```
arg_constraints = {'component_logits': Real(), 'coord_scale': GreaterThan(lower_bound=0.0)}
expand(batch_shape, _instance=None)
has_rsample = True
log_prob(value)
rsample(sample_shape=torch.Size([]))
```

## 5.2.22 MultivariateStudentT

```
class MultivariateStudentT(df, loc, scale_tril, validate_args=None)
```

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Creates a multivariate Student's  $t$ -distribution parameterized by degree of freedom `df`, mean `loc` and scale `scale_tril`.

#### Parameters

- `df` (*Tensor*) – degrees of freedom
- `loc` (*Tensor*) – mean of the distribution
- `scale_tril` (*Tensor*) – scale of the distribution, which is a lower triangular matrix with positive diagonal entries

```
arg_constraints = {'df': GreaterThan(lower_bound=0.0), 'loc': RealVector(), 'scale_tril': PositiveDefiniteMatrix()}
covariance_matrix
expand(batch_shape, _instance=None)
has_rsample = True
log_prob(value)
mean
precision_matrix
rsample(sample_shape=torch.Size([]))
support = RealVector()
```

variance

### 5.2.23 OMTMultivariateNormal

**class OMTMultivariateNormal** (*loc, scale\_tril*)

Bases: `pyro.distributions.torch.MultivariateNormal`

Multivariate normal (Gaussian) distribution with OMT gradients w.r.t. both parameters. Note the gradient computation w.r.t. the Cholesky factor has cost  $O(D^3)$ , although the resulting gradient variance is generally expected to be lower.

A distribution over vectors in which all the elements have a joint Gaussian density.

#### Parameters

- **loc** (*torch.Tensor*) – Mean.
- **scale\_tril** (*torch.Tensor*) – Cholesky of Covariance matrix.

**arg\_constraints** = {'loc': `Real()`, 'scale\_tril': `LowerTriangular()`}

**rsample** (*sample\_shape=*`torch.Size([])`)

### 5.2.24 RelaxedBernoulliStraightThrough

**class RelaxedBernoulliStraightThrough** (*temperature, probs=None, logits=None, validate\_args=None*)

Bases: `pyro.distributions.torch.RelaxedBernoulli`

An implementation of `RelaxedBernoulli` with a straight-through gradient estimator.

This distribution has the following properties:

- The samples returned by the `rsample()` method are discrete/quantized.
- The `log_prob()` method returns the log probability of the relaxed/unquantized sample using the GumbelSoftmax distribution.
- In the backward pass the gradient of the sample with respect to the parameters of the distribution uses the relaxed/unquantized sample.

References:

[1] **The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables**, Chris J. Maddison, Andriy Mnih, Yee Whye Teh

[2] **Categorical Reparameterization with Gumbel-Softmax**, Eric Jang, Shixiang Gu, Ben Poole

**log\_prob** (*value*)

See `pyro.distributions.torch.RelaxedBernoulli.log_prob()`

**rsample** (*sample\_shape=*`torch.Size([])`)

See `pyro.distributions.torch.RelaxedBernoulli.rsample()`

### 5.2.25 RelaxedOneHotCategoricalStraightThrough

**class RelaxedOneHotCategoricalStraightThrough** (*temperature, probs=None, logits=None, validate\_args=None*)

Bases: `pyro.distributions.torch.RelaxedOneHotCategorical`

An implementation of `RelaxedOneHotCategorical` with a straight-through gradient estimator.

This distribution has the following properties:

- The samples returned by the `rsample()` method are discrete/quantized.
- The `log_prob()` method returns the log probability of the relaxed/unquantized sample using the GumbelSoftmax distribution.
- In the backward pass the gradient of the sample with respect to the parameters of the distribution uses the relaxed/unquantized sample.

References:

[1] **The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables**, Chris J. Maddison, Andriy Mnih, Yee Whye Teh

[2] **Categorical Reparameterization with Gumbel-Softmax**, Eric Jang, Shixiang Gu, Ben Poole

`log_prob(value)`

See `pyro.distributions.torch.RelaxedOneHotCategorical.log_prob()`

`rsample(sample_shape=torch.Size([]))`

See `pyro.distributions.torch.RelaxedOneHotCategorical.rsample()`

## 5.2.26 Rejector

**class Rejector** (*propose, log\_prob\_accept, log\_scale*)

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Rejection sampled distribution given an acceptance rate function.

### Parameters

- **propose** (*Distribution*) – A proposal distribution that samples batched proposals via `propose().rsample()` supports a `sample_shape` arg only if `propose()` supports a `sample_shape` arg.
- **log\_prob\_accept** (*callable*) – A callable that inputs a batch of proposals and returns a batch of log acceptance probabilities.
- **log\_scale** – Total log probability of acceptance.

`has_rsample = True`

`log_prob(x)`

`rsample(sample_shape=torch.Size([]))`

`score_parts(x)`

## 5.2.27 SpanningTree

**class SpanningTree** (*edge\_logits, sampler\_options=None, validate\_args=None*)

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Distribution over spanning trees on a fixed number  $V$  of vertices.

A tree is represented as `torch.LongTensor` edges of shape  $(V-1, 2)$  satisfying the following properties:

1. The edges constitute a tree, i.e. are connected and cycle free.
2. Each edge  $(v1, v2) = \text{edges}[e]$  is sorted, i.e.  $v1 < v2$ .
3. The entire tensor is sorted in colexicographic order.

Use `validate_edges()` to verify *edges* are correctly formed.

The `edge_logits` tensor has one entry for each of the  $V * (V-1) // 2$  edges in the complete graph on  $V$  vertices, where edges are each sorted and the edge order is colexicographic:

```
(0,1), (0,2), (1,2), (0,3), (1,3), (2,3), (0,4), (1,4), (2,4), ...
```

This ordering corresponds to the size-independent pairing function:

```
k = v1 + v2 * (v2 - 1) // 2
```

where  $k$  is the rank of the edge  $(v1, v2)$  in the complete graph. To convert a matrix of edge logits to the linear representation used here:

```
assert my_matrix.shape == (V, V)
i, j = make_complete_graph(V)
edge_logits = my_matrix[i, j]
```

### Parameters

- **edge\_logits** (*torch.Tensor*) – A tensor of length  $V * (V-1) // 2$  containing logits (aka negative energies) of all edges in the complete graph on  $V$  vertices. See above comment for edge ordering.
- **sampler\_options** (*dict*) – An optional dict of sampler options including: `mcmc_steps` defaulting to a single MCMC step (which is pretty good); `initial_edges` defaulting to a cheap approximate sample; `backend` one of “python” or “cpp”, defaulting to “python”.

```
arg_constraints = {'edge_logits': Real() }
```

```
enumerate_support (expand=True)
```

This is implemented for trees with up to 6 vertices (and 5 edges).

```
has_enumerate_support = True
```

```
log_partition_function
```

```
log_prob (edges)
```

```
sample (sample_shape=torch.Size([]))
```

This sampler is implemented using MCMC run for a small number of steps after being initialized by a cheap approximate sampler. This sampler is approximate and cubic time. This is faster than the classic Aldous-Broder sampler [1,2], especially for graphs with large mixing time. Recent research [3,4] proposes samplers that run in sub-matrix-multiply time but are more complex to implement.

### References

- [1] *Generating random spanning trees* Andrei Broder (1989)
- [2] *The Random Walk Construction of Uniform Spanning Trees and Uniform Labelled Trees*, David J. Aldous (1990)
- [3] *Sampling Random Spanning Trees Faster than Matrix Multiplication*, David Durfee, Rasmus Kyng, John Peebles, Anup B. Rao, Sushant Sachdeva (2017) <https://arxiv.org/abs/1611.07451>
- [4] *An almost-linear time algorithm for uniform random spanning tree generation*, Aaron Schild (2017) <https://arxiv.org/abs/1711.06455>

```
support = IntegerGreaterThan(lower_bound=0)
```



**validate\_edges** (*edges*)

Validates a batch of edges tensors, as returned by `sample()` or `enumerate_support()` or as input to `log_prob()`.

**Parameters** *edges* (`torch.LongTensor`) – A batch of edges.

**Raises** `ValueError`

**Returns** `None`

## 5.2.28 Stable

**class Stable** (*stability, skew, scale=1.0, loc=0.0, validate\_args=None*)

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Levy  $\alpha$ -stable distribution. See [1] for a review.

This uses Nolan’s parametrization [2] of the `loc` parameter, which is required for continuity and differentiability. This corresponds to the notation  $S_{\alpha}^0(\beta, \sigma, \mu_0)$  of [1], where  $\alpha$  = stability,  $\beta$  = skew,  $\sigma$  = scale, and  $\mu_0$  = loc.

This implements a reparametrized sampler `rsample()`, but does not implement `log_prob()`. Use in inference is thus limited to likelihood-free algorithms such as `EnergyDistance`.

[1] S. Borak, W. Hardle, R. Weron (2005). Stable distributions. <https://edoc.hu-berlin.de/bitstream/handle/18452/4526/8.pdf>

[2] J.P. Nolan (1997). Numerical calculation of stable densities and distribution functions.

[3] Rafal Weron (1996). On the Chambers-Mallows-Stuck Method for Simulating Skewed Stable Random Variables.

[4] J.P. Nolan (2017). Stable Distributions: Models for Heavy Tailed Data. <http://fs2.american.edu/jpnolan/www/stable/chap1.pdf>

**Parameters**

- **stability** (*Tensor*) – Levy stability parameter  $\alpha \in (0, 2]$ .
- **skew** (*Tensor*) – Skewness  $\beta \in [-1, 1]$ .
- **scale** (*Tensor*) – Scale  $\sigma > 0$ . Defaults to 1.
- **loc** (*Tensor*) – Location  $\mu_0$  in Nolan’s parametrization [2]. Defaults to 0.

**arg\_constraints** = {'loc': `Real()`, 'scale': `GreaterThan(lower_bound=0.0)`, 'skew': `In`

**expand** (*batch\_shape, \_instance=None*)

**has\_rsample** = `True`

**log\_prob** (*value*)

**rsample** (*sample\_shape=torch.Size([])*)

**support** = `Real()`

## 5.2.29 Unit

**class Unit** (*log\_factor, validate\_args=None*)

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Trivial nonnormalized distribution representing the unit type.

The unit type has a single value with no data, i.e. `value.numel() == 0`.

This is used for `pyro.factor()` statements.

```
arg_constraints = {'log_factor': Real()}
expand(batch_shape, _instance=None)
log_prob(value)
sample(sample_shape=torch.Size([]))
support = Real()
```

### 5.2.30 VonMises

```
class VonMises(loc, concentration, validate_args=None)
```

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

A circular von Mises distribution.

This implementation uses polar coordinates. The `loc` and `value` args can be any real number (to facilitate unconstrained optimization), but are interpreted as angles modulo 2 pi.

See [VonMises3D](#) for a 3D cartesian coordinate cousin of this distribution.

#### Parameters

- `loc` (`torch.Tensor`) – an angle in radians.
- `concentration` (`torch.Tensor`) – concentration parameter

```
arg_constraints = {'concentration': GreaterThan(lower_bound=0.0), 'loc': Real() }
```

```
expand(batch_shape)
```

```
has_rsample = False
```

```
log_prob(value)
```

**mean**

The provided mean is the circular one.

```
sample(sample_shape=torch.Size([]))
```

The sampling algorithm for the von Mises distribution is based on the following paper: Best, D. J., and Nicholas I. Fisher. “Efficient simulation of the von Mises distribution.” *Applied Statistics* (1979): 152-157.

```
support = Real()
```

**variance**

The provided variance is the circular one.

### 5.2.31 VonMises3D

```
class VonMises3D(concentration, validate_args=None)
```

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Spherical von Mises distribution.

This implementation combines the direction parameter and concentration parameter into a single combined parameter that contains both direction and magnitude. The `value` arg is represented in cartesian coordinates: it must be a normalized 3-vector that lies on the 2-sphere.

See *VonMises* for a 2D polar coordinate cousin of this distribution.

Currently only `log_prob()` is implemented.

**Parameters** `concentration` (`torch.Tensor`) – A combined location-and-concentration vector. The direction of this vector is the location, and its magnitude is the concentration.

`arg_constraints` = {'concentration': `Real()`}

`expand` (`batch_shape`)

`log_prob` (`value`)

`support` = `Real()`

### 5.2.32 ZeroInflatedPoisson

**class** `ZeroInflatedPoisson` (`gate`, `rate`, `validate_args=None`)

Bases: `pyro.distributions.zero_inflated.ZeroInflatedDistribution`

A Zero Inflated Poisson distribution.

**Parameters**

- `gate` (`torch.Tensor`) – probability of extra zeros.
- `rate` (`torch.Tensor`) – rate of poisson distribution.

`support` = `IntegerGreaterThan(lower_bound=0)`

### 5.2.33 ZeroInflatedNegativeBinomial

**class** `ZeroInflatedNegativeBinomial` (`gate`, `total_count`, `probs=None`, `logits=None`, `validate_args=None`)

Bases: `pyro.distributions.zero_inflated.ZeroInflatedDistribution`

A Zero Inflated Negative Binomial distribution.

**Parameters**

- `gate` (`torch.Tensor`) – probability of extra zeros.
- `(float or Tensor)` (`total_count`) – non-negative number of negative Bernoulli trials
- `(Tensor)` (`logits`) – Event probabilities of success in the half open interval [0, 1)
- `(Tensor)` – Event log-odds for probabilities of success

`support` = `IntegerGreaterThan(lower_bound=0)`

### 5.2.34 ZeroInflatedDistribution

**class** `ZeroInflatedDistribution` (`gate`, `base_dist`, `validate_args=None`)

Bases: `pyro.distributions.torch_distribution.TorchDistribution`

Base class for a Zero Inflated distribution.

**Parameters**

- `gate` (`torch.Tensor`) – probability of extra zeros given via a Bernoulli distribution.
- `base_dist` (`TorchDistribution`) – the base distribution.

```
arg_constraints = {'gate': Interval(lower_bound=0.0, upper_bound=1.0)}
expand(batch_shape, _instance=None)
log_prob(value)
mean
sample(sample_shape=torch.Size([]))
variance
```

## 5.3 Transforms

### 5.3.1 ConditionalTransform

```
class ConditionalTransform
    Bases: abc.ABC
    condition(context)

    Return type torch.distributions.Transform
```

### 5.3.2 CorrLCholeskyTransform

```
class CorrLCholeskyTransform(cache_size=0)
    Bases: torch.distributions.transforms.Transform

    Transforms a vector into the cholesky factor of a correlation matrix.

    The input should have shape  $[batch\_shape] + [d * (d-1)/2]$ . The output will have shape  $[batch\_shape] + [d, d]$ .

    Reference:

    [1] Cholesky Factors of Correlation Matrices, Stan Reference Manual v2.18, Section 10.12

    bijective = True
    codomain = CorrCholesky()
    domain = Real()
    event_dim = 1
    log_abs_det_jacobian(x, y)
    sign = 1
```

### 5.3.3 ELUTransform

```
class ELUTransform(cache_size=0)
    Bases: torch.distributions.transforms.Transform

    Bijective transform via the mapping  $y = \text{ELU}(x)$ .

    bijective = True
    codomain = GreaterThan(lower_bound=0.0)
    domain = Real()
```

```
log_abs_det_jacobian(x, y)
sign = 1
```

### 5.3.4 LeakyReLUTransform

```
class LeakyReLUTransform(cache_size=0)
    Bases: torch.distributions.transforms.Transform
    Bijective transform via the mapping  $y = \text{LeakyReLU}(x)$ .
    bijective = True
    codomain = GreaterThan(lower_bound=0.0)
    domain = Real()
    log_abs_det_jacobian(x, y)
    sign = 1
```

### 5.3.5 LowerCholeskyAffine

```
class LowerCholeskyAffine(loc, scale_tril)
    Bases: torch.distributions.transforms.Transform
    A bijection of the form  $y = Lx + r$  where  $\text{mathbf{L}}$  is a lower triangular matrix and  $\text{mathbf{r}}$  is a vector.
```

**Parameters**

- **loc** (*torch.tensor*) – the fixed D-dimensional vector to shift the input by.
- **scale\_tril** (*torch.tensor*) – the D x D lower triangular matrix used in the transformation.

```
bijective = True
codomain = RealVector()
event_dim = 1
log_abs_det_jacobian(x, y)
    Calculates the elementwise determinant of the log Jacobian, i.e.  $\log(\text{abs}(\text{dy}/\text{dx}))$ .
volume_preserving = False
```

### 5.3.6 Permute

```
class Permute(permutation)
    Bases: torch.distributions.transforms.Transform
```

A bijection that reorders the input dimensions, that is, multiplies the input by a permutation matrix. This is useful in between [AffineAutoregressive](#) transforms to increase the flexibility of the resulting distribution and stabilize learning. Whilst not being an autoregressive transform, the log absolute determinate of the Jacobian is easily calculable as 0. Note that reordering the input dimension between two layers of [AffineAutoregressive](#) is not equivalent to reordering the dimension inside the MADE networks that those IAFs use; using a [Permute](#) transform results in a distribution with more flexibility.

Example usage:

```
>>> from pyro.nn import AutoRegressiveNN
>>> from pyro.distributions.transforms import AffineAutoregressive, Permute
>>> base_dist = dist.Normal(torch.zeros(10), torch.ones(10))
>>> iaf1 = AffineAutoregressive(AutoRegressiveNN(10, [40]))
>>> ff = Permute(torch.randperm(10, dtype=torch.long))
>>> iaf2 = AffineAutoregressive(AutoRegressiveNN(10, [40]))
>>> flow_dist = dist.TransformedDistribution(base_dist, [iaf1, ff, iaf2])
>>> flow_dist.sample() # doctest: +SKIP
tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
        0.1389, -0.4629,  0.0986])
```

Parameters **permutation** (*torch.LongTensor*) – a permutation ordering that is applied to the inputs.

**bijective** = True

**codomain** = *Real()*

**event\_dim** = 1

**inv\_permutation**

**log\_abs\_det\_jacobian** (*x, y*)

Calculates the elementwise determinant of the log Jacobian, i.e.  $\log(\text{abs}([\text{dy}_0/\text{dx}_0, \dots, \text{dy}_{\{N-1\}}/\text{dx}_{\{N-1\}}]))$ . Note that this type of transform is not autoregressive, so the log Jacobian is not the sum of the previous expression. However, it turns out it's always 0 (since the determinant is -1 or +1), and so returning a vector of zeros works.

**volume\_preserving** = True

## 5.3.7 TanhTransform

**class TanhTransform** (*cache\_size=0*)

Bases: *torch.distributions.transforms.Transform*

Bijection transform via the mapping  $y = \tanh(x)$ .

**static atanh** (*x*)

**bijective** = True

**codomain** = *Interval(lower\_bound=-1.0, upper\_bound=1.0)*

**domain** = *Real()*

**log\_abs\_det\_jacobian** (*x, y*)

**sign** = 1

## 5.4 TransformModules

### 5.4.1 AffineAutoregressive

**class AffineAutoregressive** (*autoregressive\_nn*, *log\_scale\_min\_clip=-5.0*,

*log\_scale\_max\_clip=3.0, sigmoid\_bias=2.0, stable=False*)

Bases: *pyro.distributions.torch\_transform.TransformModule*

An implementation of the bijective transform of Inverse Autoregressive Flow (IAF), using by default Eq (10) from Kingma Et Al., 2016,

$$\mathbf{y} = \mu_t + \sigma_t \odot \mathbf{x}$$

where  $\mathbf{x}$  are the inputs,  $\mathbf{y}$  are the outputs,  $\mu_t, \sigma_t$  are calculated from an autoregressive network on  $\mathbf{x}$ , and  $\sigma_t > 0$ .

If the `stable` keyword argument is set to `True` then the transformation used is,

$$\mathbf{y} = \sigma_t \odot \mathbf{x} + (1 - \sigma_t) \odot \mu_t$$

where  $\sigma_t$  is restricted to  $(0, 1)$ . This variant of IAF is claimed by the authors to be more numerically stable than one using Eq (10), although in practice it leads to a restriction on the distributions that can be represented, presumably since the input is restricted to rescaling by a number on  $(0, 1)$ .

Together with `TransformedDistribution` this provides a way to create richer variational approximations.

Example usage:

```
>>> from pyro.nn import AutoRegressiveNN
>>> base_dist = dist.Normal(torch.zeros(10), torch.ones(10))
>>> transform = AffineAutoregressive(AutoRegressiveNN(10, [40]))
>>> pyro.module("my_transform", transform) # doctest: +SKIP
>>> flow_dist = dist.TransformedDistribution(base_dist, [transform])
>>> flow_dist.sample() # doctest: +SKIP
tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
        0.1389, -0.4629,  0.0986])
```

The inverse of the Bijector is required when, e.g., scoring the log density of a sample with `TransformedDistribution`. This implementation caches the inverse of the Bijector when its forward operation is called, e.g., when sampling from `TransformedDistribution`. However, if the cached value isn't available, either because it was overwritten during sampling a new value or an arbitrary value is being scored, it will calculate it manually. Note that this is an operation that scales as  $O(D)$  where  $D$  is the input dimension, and so should be avoided for large dimensional uses. So in general, it is cheap to sample from IAF and score a value that was sampled by IAF, but expensive to score an arbitrary value.

#### Parameters

- **autoregressive\_nn** (`nn.Module`) – an autoregressive neural network whose forward call returns a real-valued mean and logit-scale as a tuple
- **log\_scale\_min\_clip** (`float`) – The minimum value for clipping the log(scale) from the autoregressive NN
- **log\_scale\_max\_clip** (`float`) – The maximum value for clipping the log(scale) from the autoregressive NN
- **sigmoid\_bias** (`float`) – A term to add the logit of the input when using the stable transform.
- **stable** (`bool`) – When true, uses the alternative “stable” version of the transform (see above).

#### References:

1. Improving Variational Inference with Inverse Autoregressive Flow [arXiv:1606.04934] Diederik P. Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, Max Welling
2. Variational Inference with Normalizing Flows [arXiv:1505.05770] Danilo Jimenez Rezende, Shakir Mohamed
3. MADE: Masked Autoencoder for Distribution Estimation [arXiv:1502.03509] Mathieu Germain, Karol Gregor, Iain Murray, Hugo Larochelle

```

autoregressive = True
bijective = True
codomain = Real()
domain = Real()
event_dim = 1
log_abs_det_jacobian(x, y)
    Calculates the elementwise determinant of the log Jacobian

```

## 5.4.2 AffineCoupling

**class AffineCoupling** (*split\_dim*, *hypernet*, *log\_scale\_min\_clip*=-5.0, *log\_scale\_max\_clip*=3.0)  
 Bases: `pyro.distributions.torch_transform.TransformModule`

An implementation of the affine coupling layer of RealNVP (Dinh et al., 2017) that uses the bijective transform,

$$\mathbf{y}_{1:d} = \mathbf{x}_{1:d} \quad \mathbf{y}_{(d+1):D} = \mu + \sigma \odot \mathbf{x}_{(d+1):D}$$

where  $\mathbf{x}$  are the inputs,  $\mathbf{y}$  are the outputs, e.g.  $\mathbf{x}_{1:d}$  represents the first  $d$  elements of the inputs, and  $\mu, \sigma$  are shift and translation parameters calculated as the output of a function inputting only  $\mathbf{x}_{1:d}$ .

That is, the first  $d$  components remain unchanged, and the subsequent  $D - d$  are shifted and translated by a function of the previous components.

Together with [TransformedDistribution](#) this provides a way to create richer variational approximations.

Example usage:

```

>>> from pyro.nn import DenseNN
>>> input_dim = 10
>>> split_dim = 6
>>> base_dist = dist.Normal(torch.zeros(input_dim), torch.ones(input_dim))
>>> hypernet = DenseNN(split_dim, [10*input_dim], [input_dim-split_dim, input_dim-
    ↳split_dim])
>>> transform = AffineCoupling(split_dim, hypernet)
>>> pyro.module("my_transform", transform) # doctest: +SKIP
>>> flow_dist = dist.TransformedDistribution(base_dist, [transform])
>>> flow_dist.sample() # doctest: +SKIP
    tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
            0.1389, -0.4629,  0.0986])

```

The inverse of the Bijector is required when, e.g., scoring the log density of a sample with [TransformedDistribution](#). This implementation caches the inverse of the Bijector when its forward operation is called, e.g., when sampling from [TransformedDistribution](#). However, if the cached value isn't available, either because it was overwritten during sampling a new value or an arbitrary value is being scored, it will calculate it manually.

This is an operation that scales as  $O(1)$ , i.e. constant in the input dimension. So in general, it is cheap to sample and score (an arbitrary value) from [AffineCoupling](#).

### Parameters

- **split\_dim** (*int*) – Zero-indexed dimension  $d$  upon which to perform input/output split for transformation.
- **hypernet** (*callable*) – an autoregressive neural network whose forward call returns a real-valued mean and logit-scale as a tuple. The input should have final dimension `split_dim` and the output final dimension `input_dim-split_dim` for each member of the tuple.



- `log_scale_min_clip(float)` – The minimum value for clipping the  $\log(\text{scale})$  from the autoregressive NN
- `log_scale_max_clip(float)` – The maximum value for clipping the  $\log(\text{scale})$  from the autoregressive NN

References:

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using Real NVP. ICLR 2017.

`bijective = True`

`codomain = Real()`

`domain = Real()`

`event_dim = 1`

`log_abs_det_jacobian(x, y)`

Calculates the elementwise determinant of the log jacobian

### 5.4.3 BatchNorm

**class** `BatchNorm(input_dim, momentum=0.1, epsilon=1e-05)`

Bases: `pyro.distributions.torch_transform.TransformModule`

A type of batch normalization that can be used to stabilize training in normalizing flows. The inverse operation is defined as

$$x = (y - \hat{\mu}) \oslash \sqrt{\hat{\sigma}^2} \otimes \gamma + \beta$$

that is, the standard batch norm equation, where  $x$  is the input,  $y$  is the output,  $\gamma, \beta$  are learnable parameters, and  $\hat{\mu}/\hat{\sigma}^2$  are smoothed running averages of the sample mean and variance, respectively. The constraint  $\gamma > 0$  is enforced to ease calculation of the log-det-Jacobian term.

This is an element-wise transform, and when applied to a vector, learns two parameters  $(\gamma, \beta)$  for each dimension of the input.

When the module is set to training mode, the moving averages of the sample mean and variance are updated every time the inverse operator is called, e.g., when a normalizing flow scores a minibatch with the `log_prob` method.

Also, when the module is set to training mode, the sample mean and variance on the current minibatch are used in place of the smoothed averages,  $\hat{\mu}$  and  $\hat{\sigma}^2$ , for the inverse operator. For this reason it is not the case that  $x = g(g^{-1}(x))$  during training, i.e., that the inverse operation is the inverse of the forward one.

Example usage:

```
>>> from pyro.nn import AutoRegressiveNN
>>> from pyro.distributions.transforms import AffineAutoregressive
>>> base_dist = dist.Normal(torch.zeros(10), torch.ones(10))
>>> iaifs = [AffineAutoregressive(AutoRegressiveNN(10, [40])) for _ in range(2)]
>>> bn = BatchNorm(10)
>>> flow_dist = dist.TransformedDistribution(base_dist, [iaifs[0], bn, iaifs[1]])
>>> flow_dist.sample() # doctest: +SKIP
tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
        0.1389, -0.4629,  0.0986])
```

#### Parameters

- `input_dim(int)` – the dimension of the input

- **momentum** (*float*) – momentum parameter for updating moving averages
- **epsilon** (*float*) – small number to add to variances to ensure numerical stability

References:

[1] Sergey Ioffe and Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In International Conference on Machine Learning, 2015. <https://arxiv.org/abs/1502.03167>

[2] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density Estimation using Real NVP. In International Conference on Learning Representations, 2017. <https://arxiv.org/abs/1605.08803>

[3] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked Autoregressive Flow for Density Estimation. In Neural Information Processing Systems, 2017. <https://arxiv.org/abs/1705.07057>

**bijective** = True

**codomain** = Real()

**constrained\_gamma**

**domain** = Real()

**event\_dim** = 0

**log\_abs\_det\_jacobian** (*x*, *y*)

Calculates the elementwise determinant of the log Jacobian, dx/dy

## 5.4.4 BlockAutoregressive

**class BlockAutoregressive** (*input\_dim*, *hidden\_factors*=[8, 8], *activation*='tanh', *residual*=None)

Bases: `pyro.distributions.torch_transform.TransformModule`

An implementation of Block Neural Autoregressive Flow (block-NAF) (De Cao et al., 2019) bijective transform. Block-NAF uses a similar transformation to deep dense NAF, building the autoregressive NN into the structure of the transform, in a sense.

Together with *TransformedDistribution* this provides a way to create richer variational approximations.

Example usage:

```
>>> base_dist = dist.Normal(torch.zeros(10), torch.ones(10))
>>> naf = BlockAutoregressive(input_dim=10)
>>> pyro.module("my_naf", naf) # doctest: +SKIP
>>> naf_dist = dist.TransformedDistribution(base_dist, [naf])
>>> naf_dist.sample() # doctest: +SKIP
tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
         0.1389, -0.4629,  0.0986])
```

The inverse operation is not implemented. This would require numerical inversion, e.g., using a root finding method - a possibility for a future implementation.

### Parameters

- **input\_dim** (*int*) – The dimensionality of the input and output variables.
- **hidden\_factors** (*list*) – Hidden layer *i* has `hidden_factors[i]` hidden units per input dimension. This corresponds to both *a* and *b* in De Cao et al. (2019). The elements of `hidden_factors` must be integers.

- **activation** (*string*) – Activation function to use. One of ‘ELU’, ‘LeakyReLU’, ‘sigmoid’, or ‘tanh’.
- **residual** (*string*) – Type of residual connections to use. Choices are “None”, “normal” for  $\mathbf{y} + f(\mathbf{y})$ , and “gated” for  $\alpha\mathbf{y} + (1 - \alpha)\mathbf{y}$  for learnable parameter  $\alpha$ .

References:

Block Neural Autoregressive Flow [arXiv:1904.04676] Nicola De Cao, Ivan Titov, Wilker Aziz

**autoregressive** = True

**bijective** = True

**codomain** = Real()

**domain** = Real()

**event\_dim** = 1

**log\_abs\_det\_jacobian** (*x*, *y*)

Calculates the elementwise determinant of the log jacobian

### 5.4.5 ConditionalPlanar

**class ConditionalPlanar** (*nn*)

Bases: `pyro.distributions.conditional.ConditionalTransformModule`

A conditional ‘planar’ bijective transform using the equation,

$$\mathbf{y} = \mathbf{x} + \mathbf{u} \tanh(\mathbf{w}^T \mathbf{z} + b)$$

where  $\mathbf{x}$  are the inputs with dimension  $D$ ,  $\mathbf{y}$  are the outputs, and the pseudo-parameters  $b \in \mathbb{R}$ ,  $\mathbf{u} \in \mathbb{R}^D$ , and  $\mathbf{w} \in \mathbb{R}^D$  are the output of a function, e.g. a NN, with input  $\mathbf{z} \in \mathbb{R}^M$  representing the context variable to condition on. For this to be an invertible transformation, the condition  $\mathbf{w}^T \mathbf{u} > -1$  is enforced.

Together with `ConditionalTransformedDistribution` this provides a way to create richer variational approximations.

Example usage:

```
>>> from pyro.nn.dense_nn import DenseNN
>>> input_dim = 10
>>> context_dim = 5
>>> batch_size = 3
>>> base_dist = dist.Normal(torch.zeros(input_dim), torch.ones(input_dim))
>>> hypernet = DenseNN(context_dim, [50, 50], param_dims=[1, input_dim, input_
↳ dim])
>>> transform = ConditionalPlanar(hypernet)
>>> z = torch.rand(batch_size, context_dim)
>>> flow_dist = dist.ConditionalTransformedDistribution(base_dist, [transform]).
↳ condition(z)
>>> flow_dist.sample(sample_shape=torch.Size([batch_size])) # doctest: +SKIP
```

The inverse of this transform does not possess an analytical solution and is left unimplemented. However, the inverse is cached when the forward operation is called during sampling, and so samples drawn using the planar transform can be scored.

**Parameters** *nn* (*callable*) – a function inputting the context variable and outputting a triplet of real-valued parameters of dimensions  $(1, D, D)$ .

References: Variational Inference with Normalizing Flows [arXiv:1505.05770] Danilo Jimenez Rezende, Shakir Mohamed

```
bijective = True
codomain = Real()
condition(context)
    See pyro.distributions.conditional.ConditionalTransformModule.
    condition()
domain = Real()
event_dim = 1
```

### 5.4.6 ConditionalTransformModule

```
class ConditionalTransformModule(*args, **kwargs)
    Bases: pyro.distributions.conditional.ConditionalTransform, torch.nn.modules.
    module.Module
```

Conditional transforms with learnable parameters such as normalizing flows should inherit from this class rather than `ConditionalTransform` so they are also a subclass of `Module` and inherit all the useful methods of that class.

### 5.4.7 Householder

```
class Householder(input_dim, count_transforms=1)
    Bases: pyro.distributions.torch_transform.TransformModule
```

Represents multiple applications of the Householder bijective transformation. A single Householder transformation takes the form,

$$\mathbf{y} = (I - 2 * \frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2})\mathbf{x}$$

where  $\mathbf{x}$  are the inputs,  $\mathbf{y}$  are the outputs, and the learnable parameters are  $\mathbf{u} \in \mathbb{R}^D$  for input dimension  $D$ .

The transformation represents the reflection of  $\mathbf{x}$  through the plane passing through the origin with normal  $\mathbf{u}$ .

$D$  applications of this transformation are able to transform standard i.i.d. standard Gaussian noise into a Gaussian variable with an arbitrary covariance matrix. With  $K < D$  transformations, one is able to approximate a full-rank Gaussian distribution using a linear transformation of rank  $K$ .

Together with `TransformedDistribution` this provides a way to create richer variational approximations.

Example usage:

```
>>> base_dist = dist.Normal(torch.zeros(10), torch.ones(10))
>>> transform = Householder(10, count_transforms=5)
>>> pyro.module("my_transform", p) # doctest: +SKIP
>>> flow_dist = dist.TransformDistribution(base_dist, [transform])
>>> flow_dist.sample() # doctest: +SKIP
tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
        0.1389, -0.4629,  0.0986])
```

#### Parameters

- **input\_dim** (*int*) – the dimension of the input (and output) variable.

- **count\_transforms** (*int*) – number of applications of Householder transformation to apply.

References:

Improving Variational Auto-Encoders using Householder Flow, [arXiv:1611.09630] Tomczak, J. M., & Welling, M.

**bijective** = True

**codomain** = Real()

**domain** = Real()

**event\_dim** = 1

**log\_abs\_det\_jacobian** (*x*, *y*)

Calculates the elementwise determinant of the log jacobian. Householder flow is measure preserving, so  $\log(|\det J|) = 0$

**reset\_parameters** ()

**u** ()

**volume\_preserving** = True

### 5.4.8 NeuralAutoregressive

**class NeuralAutoregressive** (*autoregressive\_nn*, *hidden\_units=16*, *activation='sigmoid'*)

Bases: `pyro.distributions.torch_transform.TransformModule`

An implementation of the deep Neural Autoregressive Flow (NAF) bijective transform of the “IAF flavour” that can be used for sampling and scoring samples drawn from it (but not arbitrary ones).

Example usage:

```
>>> from pyro.nn import AutoRegressiveNN
>>> base_dist = dist.Normal(torch.zeros(10), torch.ones(10))
>>> arn = AutoRegressiveNN(10, [40], param_dims=[16]*3)
>>> transform = NeuralAutoregressive(arn, hidden_units=16)
>>> pyro.module("my_transform", transform) # doctest: +SKIP
>>> flow_dist = dist.TransformedDistribution(base_dist, [transform])
>>> flow_dist.sample() # doctest: +SKIP
tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
        0.1389, -0.4629,  0.0986])
```

The inverse operation is not implemented. This would require numerical inversion, e.g., using a root finding method - a possibility for a future implementation.

#### Parameters

- **autoregressive\_nn** (*nn.Module*) – an autoregressive neural network whose forward call returns a tuple of three real-valued tensors, whose last dimension is the input dimension, and whose penultimate dimension is equal to *hidden\_units*.
- **hidden\_units** (*int*) – the number of hidden units to use in the NAF transformation (see Eq (8) in reference)
- **activation** (*string*) – Activation function to use. One of ‘ELU’, ‘LeakyReLU’, ‘sigmoid’, or ‘tanh’.

Reference:

Neural Autoregressive Flows [arXiv:1804.00779] Chin-Wei Huang, David Krueger, Alexandre Lacoste, Aaron Courville

```
bijective = True
codomain = Real()
domain = Real()
event_dim = 1
log_abs_det_jacobian(x, y)
    Calculates the elementwise determinant of the log Jacobian
```

## 5.4.9 Planar

```
class Planar(input_dim)
    Bases: pyro.distributions.transforms.planar.ConditionedPlanar, pyro.
            distributions.torch_transform.TransformModule
```

A ‘planar’ bijective transform with equation,

$$\mathbf{y} = \mathbf{x} + \mathbf{u} \tanh(\mathbf{w}^T \mathbf{z} + b)$$

where  $\mathbf{x}$  are the inputs,  $\mathbf{y}$  are the outputs, and the learnable parameters are  $b \in \mathbb{R}$ ,  $\mathbf{u} \in \mathbb{R}^D$ ,  $\mathbf{w} \in \mathbb{R}^D$  for input dimension  $D$ . For this to be an invertible transformation, the condition  $\mathbf{w}^T \mathbf{u} > -1$  is enforced.

Together with [TransformedDistribution](#) this provides a way to create richer variational approximations.

Example usage:

```
>>> base_dist = dist.Normal(torch.zeros(10), torch.ones(10))
>>> transform = Planar(10)
>>> pyro.module("my_transform", transform) # doctest: +SKIP
>>> flow_dist = dist.TransformedDistribution(base_dist, [transform])
>>> flow_dist.sample() # doctest: +SKIP
    tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
            0.1389, -0.4629,  0.0986])
```

The inverse of this transform does not possess an analytical solution and is left unimplemented. However, the inverse is cached when the forward operation is called during sampling, and so samples drawn using the planar transform can be scored.

**Parameters** `input_dim` (*int*) – the dimension of the input (and output) variable.

References:

Variational Inference with Normalizing Flows [arXiv:1505.05770] Danilo Jimenez Rezende, Shakir Mohamed

```
bijective = True
codomain = Real()
domain = Real()
event_dim = 1
reset_parameters()
```

### 5.4.10 Polynomial

**class Polynomial** (*autoregressive\_nn*, *input\_dim*, *count\_degree*, *count\_sum*)

Bases: `pyro.distributions.torch_transform.TransformModule`

An autoregressive bijective transform as described in Jaini et al. (2019) applying following equation element-wise,

$$y_n = c_n + \int_0^{x_n} \sum_{k=1}^K \left( \sum_{r=0}^R a_{r,k}^{(n)} u^r \right) du$$

where  $x_n$  is the  $n$  is the  $n$ ,  $\{a_{r,k}^{(n)} \in \mathbb{R}\}$  are learnable parameters that are the output of an autoregressive NN inputting  $x_{<n} = x_1, x_2, \dots, x_{n-1}$ .

Together with [TransformedDistribution](#) this provides a way to create richer variational approximations.

Example usage:

```
>>> from pyro.nn import AutoRegressiveNN
>>> input_dim = 10
>>> count_degree = 4
>>> count_sum = 3
>>> base_dist = dist.Normal(torch.zeros(input_dim), torch.ones(input_dim))
>>> arn = AutoRegressiveNN(input_dim, [input_dim*10], param_dims=[(count_degree + 1)*count_sum])
>>> transform = Polynomial(arn, input_dim=input_dim, count_degree=count_degree, count_sum=count_sum)
>>> pyro.module("my_transform", transform) # doctest: +SKIP
>>> flow_dist = dist.TransformedDistribution(base_dist, [transform])
>>> flow_dist.sample() # doctest: +SKIP
tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
        0.1389, -0.4629,  0.0986])
```

The inverse of this transform does not possess an analytical solution and is left unimplemented. However, the inverse is cached when the forward operation is called during sampling, and so samples drawn using a polynomial transform can be scored.

#### Parameters

- **autoregressive\_nn** (*nn.Module*) – an autoregressive neural network whose forward call returns a tensor of real-valued numbers of size (batch\_size, (count\_degree+1)\*count\_sum, input\_dim)
- **count\_degree** (*int*) – The degree of the polynomial to use for each element-wise transformation.
- **count\_sum** (*int*) – The number of polynomials to sum in each element-wise transformation.

References:

Sum-of-squares polynomial flow. [arXiv:1905.02325] Priyank Jaini, Kira A. Shelby, Yaoliang Yu

**autoregressive = True**

**bijective = True**

**codomain = Real()**

**domain = Real()**

**event\_dim = 1**

```
log_abs_det_jacobian(x, y)
    Calculates the elementwise determinant of the log Jacobian

reset_parameters()
```

### 5.4.11 Radial

```
class Radial(input_dim)
    Bases: pyro.distributions.torch_transform.TransformModule

    A 'radial' bijective transform using the equation,
```

$$\mathbf{y} = \mathbf{x} + \beta h(\alpha, r)(\mathbf{x} - \mathbf{x}_0)$$

where  $\mathbf{x}$  are the inputs,  $\mathbf{y}$  are the outputs, and the learnable parameters are  $\alpha \in \mathbb{R}^+$ ,  $\beta \in \mathbb{R}$ ,  $\mathbf{x}_0 \in \mathbb{R}^D$ , for input dimension  $D$ ,  $r = \|\mathbf{x} - \mathbf{x}_0\|_2$ ,  $h(\alpha, r) = 1/(\alpha + r)$ . For this to be an invertible transformation, the condition  $\beta > -\alpha$  is enforced.

Together with [TransformedDistribution](#) this provides a way to create richer variational approximations.

Example usage:

```
>>> base_dist = dist.Normal(torch.zeros(10), torch.ones(10))
>>> transform = Radial(10)
>>> pyro.module("my_transform", transform) # doctest: +SKIP
>>> flow_dist = dist.TransformedDistribution(base_dist, [transform])
>>> flow_dist.sample() # doctest: +SKIP
    tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
            0.1389, -0.4629,  0.0986])
```

The inverse of this transform does not possess an analytical solution and is left unimplemented. However, the inverse is cached when the forward operation is called during sampling, and so samples drawn using the radial transform can be scored.

**Parameters** `input_dim` (*int*) – the dimension of the input (and output) variable.

References:

Variational Inference with Normalizing Flows [arXiv:1505.05770] Danilo Jimenez Rezende, Shakir Mohamed

```
bijective = True
codomain = Real()
domain = Real()
event_dim = 1
log_abs_det_jacobian(x, y)
    Calculates the elementwise determinant of the log Jacobian

reset_parameters()
```

### 5.4.12 Sylvester

```
class Sylvester(input_dim, count_transforms=1)
    Bases: pyro.distributions.transforms.householder.Householder
```

An implementation of the Sylvester bijective transform of the Householder variety (Van den Berg Et Al., 2018),

$$\mathbf{y} = \mathbf{x} + QR \tanh(SQ^T \mathbf{x} + \mathbf{b})$$



where  $\mathbf{x}$  are the inputs,  $\mathbf{y}$  are the outputs,  $R, S \sim D \times D$  are upper triangular matrices for input dimension  $D$ ,  $Q \sim D \times D$  is an orthogonal matrix, and  $\mathbf{b} \sim D$  is learnable bias term.

The Sylvester transform is a generalization of [Planar](#). In the Householder type of the Sylvester transform, the orthogonality of  $Q$  is enforced by representing it as the product of Householder transformations.

Together with [TransformedDistribution](#) it provides a way to create richer variational approximations.

Example usage:

```
>>> base_dist = dist.Normal(torch.zeros(10), torch.ones(10))
>>> transform = Sylvester(10, count_transforms=4)
>>> pyro.module("my_transform", transform) # doctest: +SKIP
>>> flow_dist = dist.TransformedDistribution(base_dist, [transform])
>>> flow_dist.sample() # doctest: +SKIP
tensor([-0.4071, -0.5030,  0.7924, -0.2366, -0.2387, -0.1417,  0.0868,
        0.1389, -0.4629,  0.0986])
```

The inverse of this transform does not possess an analytical solution and is left unimplemented. However, the inverse is cached when the forward operation is called during sampling, and so samples drawn using the Sylvester transform can be scored.

References:

Rianne van den Berg, Leonard Hasenclever, Jakub M. Tomczak, Max Welling. Sylvester Normalizing Flows for Variational Inference. In proceedings of The 34th Conference on Uncertainty in Artificial Intelligence (UAI 2018).

$Q(x)$

$R()$

$S()$

**bijective** = True

**codomain** = Real()

**domain** = Real()

**dtanh\_dx**( $x$ )

**event\_dim** = 1

**log\_abs\_det\_jacobian**( $x, y$ )

Calculates the elementwise determinant of the log Jacobian

**reset\_parameters2**()

### 5.4.13 TransformModule

**class TransformModule** (\*args, \*\*kwargs)

Bases: `torch.distributions.transforms.Transform`, `torch.nn.modules.module.Module`

Transforms with learnable parameters such as normalizing flows should inherit from this class rather than *Transform* so they are also a subclass of *nn.Module* and inherit all the useful methods of that class.

## 5.5 Transform Factories

Each `Transform` and `TransformModule` includes a corresponding helper function in lower case that inputs, at minimum, the input dimensions of the transform, and possibly additional arguments to customize the transform in an intuitive way. The purpose of these helper functions is to hide from the user whether or not the transform requires the construction of a hypernet, and if so, the input and output dimensions of that hypernet.

### 5.5.1 affine\_autoregressive

**affine\_autoregressive** (*input\_dim*, *hidden\_dims=None*, *\*\*kwargs*)

A helper function to create an `AffineAutoregressive` object that takes care of constructing an autoregressive network with the correct input/output dimensions.

#### Parameters

- **input\_dim** (*int*) – Dimension of input variable
- **hidden\_dims** (*list[int]*) – The desired hidden dimensions of the autoregressive network. Defaults to using  $[3 * \text{input\_dim} + 1]$
- **log\_scale\_min\_clip** (*float*) – The minimum value for clipping the  $\log(\text{scale})$  from the autoregressive NN
- **log\_scale\_max\_clip** (*float*) – The maximum value for clipping the  $\log(\text{scale})$  from the autoregressive NN
- **sigmoid\_bias** (*float*) – A term to add the logit of the input when using the stable transform.
- **stable** (*bool*) – When true, uses the alternative “stable” version of the transform (see above).

### 5.5.2 affine\_coupling

**affine\_coupling** (*input\_dim*, *hidden\_dims=None*, *split\_dim=None*, *\*\*kwargs*)

A helper function to create an `AffineCoupling` object that takes care of constructing a dense network with the correct input/output dimensions.

#### Parameters

- **input\_dim** (*int*) – Dimension of input variable
- **hidden\_dims** (*list[int]*) – The desired hidden dimensions of the dense network. Defaults to using  $[10 * \text{input\_dim}]$
- **split\_dim** (*int*) – The dimension to split the input on for the coupling transform. Defaults to using  $\text{input\_dim} // 2$
- **log\_scale\_min\_clip** (*float*) – The minimum value for clipping the  $\log(\text{scale})$  from the autoregressive NN
- **log\_scale\_max\_clip** (*float*) – The maximum value for clipping the  $\log(\text{scale})$  from the autoregressive NN

### 5.5.3 batchnorm

**batchnorm** (*input\_dim*, *\*\*kwargs*)

A helper function to create a *BatchNorm* object for consistency with other helpers.

**Parameters**

- **input\_dim** (*int*) – Dimension of input variable
- **momentum** (*float*) – momentum parameter for updating moving averages
- **epsilon** (*float*) – small number to add to variances to ensure numerical stability

### 5.5.4 block\_autoregressive

**block\_autoregressive** (*input\_dim*, *\*\*kwargs*)

A helper function to create a *BlockAutoregressive* object for consistency with other helpers.

**Parameters**

- **input\_dim** (*int*) – Dimension of input variable
- **hidden\_factors** (*list*) – Hidden layer *i* has `hidden_factors[i]` hidden units per input dimension. This corresponds to both *a* and *b* in De Cao et al. (2019). The elements of `hidden_factors` must be integers.
- **activation** (*string*) – Activation function to use. One of ‘ELU’, ‘LeakyReLU’, ‘sigmoid’, or ‘tanh’.
- **residual** (*string*) – Type of residual connections to use. Choices are “None”, “normal” for  $y + f(y)$ , and “gated” for  $\alpha y + (1 - \alpha)y$  for learnable parameter  $\alpha$ .

### 5.5.5 conditional\_planar

**conditional\_planar** (*input\_dim*, *context\_dim*, *hidden\_dims=None*)

A helper function to create a *ConditionalPlanar* object that takes care of constructing a dense network with the correct input/output dimensions.

**Parameters**

- **input\_dim** (*int*) – Dimension of input variable
- **context\_dim** (*int*) – Dimension of context variable
- **hidden\_dims** (*list[int]*) – The desired hidden dimensions of the dense network. Defaults to using `[input_dim * 10, input_dim * 10]`

### 5.5.6 elu

**elu** ()

A helper function to create an *ELUTransform* object for consistency with other helpers.

### 5.5.7 householder

**householder** (*input\_dim*, *count\_transforms=None*)

A helper function to create a *Householder* object for consistency with other helpers.

**Parameters**

- **input\_dim** (*int*) – Dimension of input variable
- **count\_transforms** (*int*) – number of applications of Householder transformation to apply.

### 5.5.8 leaky\_relu

**leaky\_relu** ()

A helper function to create a *LeakyReLUTransform* object for consistency with other helpers.

### 5.5.9 neural\_autoregressive

**neural\_autoregressive** (*input\_dim*, *hidden\_dims=None*, *activation='sigmoid'*, *width=16*)

A helper function to create a *NeuralAutoregressive* object that takes care of constructing an autoregressive network with the correct input/output dimensions.

**Parameters**

- **input\_dim** (*int*) – Dimension of input variable
- **hidden\_dims** (*list[int]*) – The desired hidden dimensions of the autoregressive network. Defaults to using  $[3 * \text{input\_dim} + 1]$
- **activation** (*string*) – Activation function to use. One of 'ELU', 'LeakyReLU', 'sigmoid', or 'tanh'.
- **width** (*int*) – The width of the “multilayer perceptron” in the transform (see paper). Defaults to 16

### 5.5.10 permute

**permute** (*input\_dim*, *permutation=None*)

A helper function to create a *Permute* object for consistency with other helpers.

**Parameters**

- **input\_dim** (*int*) – Dimension of input variable
- **permutation** (*torch.LongTensor*) – Torch tensor of integer indices representing permutation. Defaults to a random permutation.

### 5.5.11 planar

**planar** (*input\_dim*)

A helper function to create a *Planar* object for consistency with other helpers.

**Parameters** **input\_dim** (*int*) – Dimension of input variable

### 5.5.12 polynomial

**polynomial** (*input\_dim*, *hidden\_dims=None*)

A helper function to create a *Polynomial* object that takes care of constructing an autoregressive network with the correct input/output dimensions.

**Parameters**

- **input\_dim** (*int*) – Dimension of input variable
- **hidden\_dims** – The desired hidden dimensions of the autoregressive network. Defaults to using `[input_dim * 10]`

### 5.5.13 radial

**radial** (*input\_dim*)

A helper function to create a *Radial* object for consistency with other helpers.

**Parameters** **input\_dim** (*int*) – Dimension of input variable

### 5.5.14 sylvester

**sylvester** (*input\_dim*, *count\_transforms=None*)

A helper function to create a *Sylvester* object for consistency with other helpers.

**Parameters**

- **input\_dim** (*int*) – Dimension of input variable
- **count\_transforms** (*int*) – Number of Sylvester operations to apply. Defaults to `input_dim // 2 + 1`.

### 5.5.15 tanh

**tanh** ()

A helper function to create a *TanhTransform* object for consistency with other helpers.



Parameters in Pyro are basically thin wrappers around PyTorch Tensors that carry unique names. As such Parameters are the primary stateful objects in Pyro. Users typically interact with parameters via the Pyro primitive *pyro.param*. Parameters play a central role in stochastic variational inference, where they are used to represent point estimates for the parameters in parameterized families of models and guides.

## 6.1 ParamStore

**class ParamStoreDict**

Bases: `object`

Global store for parameters in Pyro. This is basically a key-value store. The typical user interacts with the ParamStore primarily through the primitive *pyro.param*.

See [Intro Part II](#) for further discussion and [SVI Part I](#) for some examples.

Some things to bear in mind when using parameters in Pyro:

- parameters must be assigned unique names
- the *init\_tensor* argument to *pyro.param* is only used the first time that a given (named) parameter is registered with Pyro.
- for this reason, a user may need to use the *clear()* method if working in a REPL in order to get the desired behavior. this method can also be invoked with *pyro.clear\_param\_store()*.
- the internal name of a parameter within a PyTorch *nn.Module* that has been registered with Pyro is prepended with the Pyro name of the module. so nothing prevents the user from having two different modules each of which contains a parameter named *weight*. by contrast, a user can only have one top-level parameter named *weight* (outside of any module).
- parameters can be saved and loaded from disk using *save* and *load*.

**clear()**

Clear the ParamStore

**items()**

Iterate over (name, constrained\_param) pairs.

**keys()**

Iterate over param names.

**values()**

Iterate over constrained parameter values.

**setdefault** (name, init\_constrained\_value, constraint=Real())

Retrieve a constrained parameter value from the if it exists, otherwise set the initial value. Note that this is a little fancier than `dict.setdefault()`.

If the parameter already exists, `init_constrained_tensor` will be ignored. To avoid expensive creation of `init_constrained_tensor` you can wrap it in a `lambda` that will only be evaluated if the parameter does not already exist:

```
param_store.get("foo", lambda: (0.001 * torch.randn(1000, 1000)).exp(),
                 constraint=constraints.positive)
```

### Parameters

- **name** (*str*) – parameter name
- **init\_constrained\_value** (*torch.Tensor* or callable returning a *torch.Tensor*) – initial constrained value
- **constraint** (*torch.distributions.constraints.Constraint*) – torch constraint object

**Returns** constrained parameter value

**Return type** `torch.Tensor`

**named\_parameters()**

Returns an iterator over (name, unconstrained\_value) tuples for each parameter in the ParamStore.

**get\_all\_param\_names()**

**replace\_param** (param\_name, new\_param, old\_param)

**get\_param** (name, init\_tensor=None, constraint=Real(), event\_dim=None)

Get parameter from its name. If it does not yet exist in the ParamStore, it will be created and stored. The Pyro primitive *pyro.param* dispatches to this method.

### Parameters

- **name** (*str*) – parameter name
- **init\_tensor** (*torch.Tensor*) – initial tensor
- **constraint** (*torch.distributions.constraints.Constraint*) – torch constraint
- **event\_dim** (*int*) – (ignored)

**Returns** parameter

**Return type** `torch.Tensor`

**match** (name)

Get all parameters that match regex. The parameter must exist.



**Parameters** `name` (*str*) – regular expression

**Returns** dict with key param name and value torch Tensor

**param\_name** (*p*)

Get parameter name from parameter

**Parameters** `p` – parameter

**Returns** parameter name

**get\_state** ()

Get the ParamStore state.

**set\_state** (*state*)

Set the ParamStore state using state from a previous `get_state()` call

**save** (*filename*)

Save parameters to disk

**Parameters** `filename` (*str*) – file name to save to

**load** (*filename*, *map\_location=None*)

Loads parameters from disk

---

**Note:** If using `pyro.module()` on parameters loaded from disk, be sure to set the `update_module_params` flag:

```
pyro.get_param_store().load('saved_params.save')
pyro.module('module', nn, update_module_params=True)
```

---

## Parameters

- **filename** (*str*) – file name to load from
- **map\_location** (*function, torch.device, string or a dict*) – specifies how to remap storage locations

**param\_with\_module\_name** (*pyro\_name, param\_name*)

**module\_from\_param\_with\_module\_name** (*param\_name*)

**user\_param\_name** (*param\_name*)



The module *pyro.nn* provides implementations of neural network modules that are useful in the context of deep probabilistic programming.

## 7.1 Pyro Modules

Pyro includes an experimental class *PyroModule*, a subclass of `torch.nn.Module`, whose attributes can be modified by Pyro effects. To create a poutine-aware attribute, use either the *PyroParam* struct or the *PyroSample* struct:

```
my_module = PyroModule()
my_module.x = PyroParam(torch.tensor(1.), constraint=constraints.positive)
my_module.y = PyroSample(dist.Normal(0, 1))
```

### **class PyroParam**

Bases: `tuple`

Structure to declare a Pyro-managed learnable parameter of a *PyroModule*.

#### **constraint**

Alias for field number 1

#### **event\_dim**

Alias for field number 2

#### **init\_value**

Alias for field number 0

### **class PyroSample**

Bases: `tuple`

Structure to declare a Pyro-managed random parameter of a *PyroModule*.

#### **prior**

Alias for field number 0

```
class PyroModule(name="")
```

```
Bases: torch.nn.modules.module.Module
```

EXPERIMENTAL Subclass of `torch.nn.Module` whose attributes can be modified by Pyro effects. Attributes can be set using helpers `PyroParam` and `PyroSample`, and methods can be decorated by `pyro_method()`.

### Parameters

To create a Pyro-managed parameter attribute, set that attribute using either `torch.nn.Parameter` (for unconstrained parameters) or `PyroParam` (for constrained parameters). Reading that attribute will then trigger a `pyro.param()` statement. For example:

```
# Create Pyro-managed parameter attributes.
my_module = PyroModule()
my_module.loc = nn.Parameter(torch.tensor(0.))
my_module.scale = PyroParam(torch.tensor(1.),
                             constraint=constraints.positive)

# Read the attributes.
loc = my_module.loc # Triggers a pyro.param statement.
scale = my_module.scale # Triggers another pyro.param statement.
```

Note that, unlike normal `torch.nn.Module`s, `PyroModule`s should not be registered with `pyro.module()` statements. `PyroModule`s can contain other `PyroModule`s and normal `torch.nn.Module`s. Accessing a normal `torch.nn.Module` attribute of a `PyroModule` triggers a `pyro.module()` statement. If multiple `PyroModule`s appear in a single Pyro model or guide, they should be included in a single root `PyroModule` for that model.

`PyroModule`s synchronize data with the param store at each `setattr`, `getattr`, and `delattr` event, based on the nested name of an attribute:

- Setting `mod.x = x_init` tries to read `x` from the param store. If a value is found in the param store, that value is copied into `mod` and `x_init` is ignored; otherwise `x_init` is copied into both `mod` and the param store.
- Reading `mod.x` tries to read `x` from the param store. If a value is found in the param store, that value is copied into `mod`; otherwise `mod`'s value is copied into the param store. Finally `mod` and the param store agree on a single value to return.
- Deleting `del mod.x` removes a value from both `mod` and the param store.

Note two `PyroModule` of the same name will both synchronize with the global param store and thus contain the same data. When creating a `PyroModule`, then deleting it, then creating another with the same name, the latter will be populated with the former's data from the param store. To avoid this persistence, either `pyro.clear_param_store()` or call `clear()` before deleting a `PyroModule`.

`PyroModule`s can be saved and loaded either directly using `torch.save()` / `torch.load()` or indirectly using the param store's `save()` / `load()`. Note that `torch.load()` will be overridden by any values in the param store, so it is safest to `pyro.clear_param_store()` before loading.

### Samples

To create a Pyro-managed random attribute, set that attribute using the `PyroSample` helper, specifying a prior distribution. Reading that attribute will then trigger a `pyro.sample()` statement. For example:

```
# Create Pyro-managed random attributes.
my_module.x = PyroSample(dist.Normal(0, 1))
my_module.y = PyroSample(lambda self: dist.Normal(self.loc, self.scale))

# Sample the attributes.
```

(continues on next page)

(continued from previous page)

```
x = my_module.x  # Triggers a pyro.sample statement.
y = my_module.y  # Triggers one pyro.sample + two pyro.param statements.
```

Sampling is cached within each invocation of `__call__()` or method decorated by `pyro_method()`. Because sample statements can appear only once in a Pyro trace, you should ensure that traced access to sample attributes is wrapped in a single invocation of `__call__()` or method decorated by `pyro_method()`.

To make an existing module probabilistic, you can create a subclass and overwrite some parameters with `PyroSample`s:

```
class RandomLinear(nn.Linear, PyroModule): # used as a mixin
    def __init__(self, in_features, out_features):
        super().__init__(in_features, out_features)
        self.weight = PyroSample(
            lambda self: dist.Normal(0, 1)
                           .expand([self.out_features,
                                   self.in_features])
                           .to_event(2))
```

### Mixin classes

`PyroModule` can be used as a mixin class, and supports simple syntax for dynamically creating mixins, for example the following are equivalent:

```
# Version 1. create a named mixin class
class PyroLinear(nn.Linear, PyroModule):
    pass

m.linear = PyroLinear(m, n)

# Version 2. create a dynamic mixin class
m.linear = PyroModule[nn.Linear](m, n)
```

This notation can be used recursively to create Bayesian modules, e.g.:

```
model = PyroModule[nn.Sequential](
    PyroModule[nn.Linear](28 * 28, 100),
    PyroModule[nn.Sigmoid](),
    PyroModule[nn.Linear](100, 100),
    PyroModule[nn.Sigmoid](),
    PyroModule[nn.Linear](100, 10),
)
assert isinstance(model, nn.Sequential)
assert isinstance(model, PyroModule)

# Now we can be Bayesian about weights in the first layer.
model[0].weight = PyroSample(
    prior=dist.Normal(0, 1).expand([28 * 28, 100]).to_event(2))
guide = AutoDiagonalNormal(model)
```

Note that `PyroModule[...]` does not recursively mix in `PyroModule` to submodules of the input Module; hence we needed to wrap each submodule of the `nn.Sequential` above.

**Parameters** `name` (*str*) – Optional name for a root `PyroModule`. This is ignored in sub-`PyroModules` of another `PyroModule`.

**add\_module** (*name, module*)

Adds a child module to the current module.

**named\_pyro\_params** (*prefix=""*, *recurse=True*)

Returns an iterator over PyroModule parameters, yielding both the name of the parameter as well as the parameter itself.

**Parameters**

- **prefix** (*str*) – prefix to prepend to all parameter names.
- **recurse** (*bool*) – if True, then yields parameters of this module and all submodules. Otherwise, yields only parameters that are direct members of this module.

**Returns** a generator which yields tuples containing the name and parameter

**pyro\_method** (*fn*)

Decorator for top-level methods of a *PyroModule* to enable pyro effects and cache `pyro.sample` statements.

This should be applied to all public methods that read Pyro-managed attributes, but is not needed for `.forward()`.

**clear** (*mod*)

Removes data from both a *PyroModule* and the param store.

**Parameters** *mod* (*PyroModule*) – A module to clear.

**to\_pyro\_module** (*m*, *recurse=True*)

Converts an ordinary `torch.nn.Module` instance to a *PyroModule* **in-place**.

This is useful for adding Pyro effects to third-party modules: no third-party code needs to be modified. For example:

```
model = nn.Sequential(
    nn.Linear(28 * 28, 100),
    nn.Sigmoid(),
    nn.Linear(100, 100),
    nn.Sigmoid(),
    nn.Linear(100, 10),
)
to_pyro_module_(model)
assert isinstance(model, PyroModule[nn.Sequential])
assert isinstance(model[0], PyroModule[nn.Linear])

# Now we can attempt to be fully Bayesian:
for m in model.modules():
    for name, value in list(m.named_parameters(recurse=False)):
        setattr(m, name, PyroSample(prior=dist.Normal(0, 1)
                                     .expand(value.shape)
                                     .to_event(value.dim()))))

guide = AutoDiagonalNormal(model)
```

**Parameters**

- *m* (*torch.nn.Module*) – A module instance.
- **recurse** (*bool*) – Whether to convert submodules to *PyroModules*.

## 7.2 AutoRegressiveNN

```
class AutoRegressiveNN(input_dim, hidden_dims, param_dims=[1, 1], permutation=None,
                        skip_connections=False, nonlinearity=ReLU())
```

Bases: `pyro.nn.auto_reg_nn.ConditionalAutoRegressiveNN`

An implementation of a MADE-like auto-regressive neural network.

Example usage:

```
>>> x = torch.randn(100, 10)
>>> arn = AutoRegressiveNN(10, [50], param_dims=[1])
>>> p = arn(x) # 1 parameters of size (100, 10)
>>> arn = AutoRegressiveNN(10, [50], param_dims=[1, 1])
>>> m, s = arn(x) # 2 parameters of size (100, 10)
>>> arn = AutoRegressiveNN(10, [50], param_dims=[1, 5, 3])
>>> a, b, c = arn(x) # 3 parameters of sizes, (100, 1, 10), (100, 5, 10), (100, 3,
↪ 10)
```

### Parameters

- **input\_dim** (*int*) – the dimensionality of the input variable
- **hidden\_dims** (*list[int]*) – the dimensionality of the hidden units per layer
- **param\_dims** (*list[int]*) – shape the output into parameters of dimension (p\_n, input\_dim) for p\_n in param\_dims when p\_n > 1 and dimension (input\_dim) when p\_n == 1. The default is [1, 1], i.e. output two parameters of dimension (input\_dim), which is useful for inverse autoregressive flow.
- **permutation** (*torch.LongTensor*) – an optional permutation that is applied to the inputs and controls the order of the autoregressive factorization. in particular for the identity permutation the autoregressive structure is such that the Jacobian is upper triangular. By default this is chosen at random.
- **skip\_connections** (*bool*) – Whether to add skip connections from the input to the output.
- **nonlinearity** (*torch.nn.module*) – The nonlinearity to use in the feedforward network such as `torch.nn.ReLU()`. Note that no nonlinearity is applied to the final network output, so the output is an unbounded real number.

Reference:

MADE: Masked Autoencoder for Distribution Estimation [arXiv:1502.03509] Mathieu Germain, Karol Gregor, Iain Murray, Hugo Larochelle

**forward** (*x*)

The forward method

## 7.3 ConditionalAutoRegressiveNN

```
class ConditionalAutoRegressiveNN(input_dim, context_dim, hidden_dims, param_dims=[1,
                                                                              1], permutation=None, skip_connections=False, nonlinearity=ReLU())
```

Bases: `torch.nn.modules.module.Module`

An implementation of a MADE-like auto-regressive neural network that can input an additional context variable. (See Reference [2] Section 3.3 for an explanation of how the conditional MADE architecture works.)

Example usage:

```
>>> x = torch.randn(100, 10)
>>> y = torch.randn(100, 5)
>>> arn = ConditionalAutoRegressiveNN(10, 5, [50], param_dims=[1])
>>> p = arn(x, context=y) # 1 parameters of size (100, 10)
>>> arn = ConditionalAutoRegressiveNN(10, 5, [50], param_dims=[1, 1])
>>> m, s = arn(x, context=y) # 2 parameters of size (100, 10)
>>> arn = ConditionalAutoRegressiveNN(10, 5, [50], param_dims=[1, 5, 3])
>>> a, b, c = arn(x, context=y) # 3 parameters of sizes, (100, 1, 10), (100, 5, 10), (100, 3, 10)
```

### Parameters

- **input\_dim** (*int*) – the dimensionality of the input variable
- **context\_dim** (*int*) – the dimensionality of the context variable
- **hidden\_dims** (*list[int]*) – the dimensionality of the hidden units per layer
- **param\_dims** (*list[int]*) – shape the output into parameters of dimension (p\_n, input\_dim) for p\_n in param\_dims when p\_n > 1 and dimension (input\_dim) when p\_n == 1. The default is [1, 1], i.e. output two parameters of dimension (input\_dim), which is useful for inverse autoregressive flow.
- **permutation** (*torch.LongTensor*) – an optional permutation that is applied to the inputs and controls the order of the autoregressive factorization. in particular for the identity permutation the autoregressive structure is such that the Jacobian is upper triangular. By default this is chosen at random.
- **skip\_connections** (*bool*) – Whether to add skip connections from the input to the output.
- **nonlinearity** (*torch.nn.module*) – The nonlinearity to use in the feedforward network such as torch.nn.ReLU(). Note that no nonlinearity is applied to the final network output, so the output is an unbounded real number.

Reference:

1. MADE: Masked Autoencoder for Distribution Estimation [arXiv:1502.03509] Mathieu Germain, Karol Gregor, Iain Murray, Hugo Larochelle
2. Inference Networks for Sequential Monte Carlo in Graphical Models [arXiv:1602.06701] Brooks Paige, Frank Wood

**forward** (*x, context*)

The forward method

**get\_permutation** ()

Get the permutation applied to the inputs (by default this is chosen at random)



The module `pyro.optim` provides support for optimization in Pyro. In particular it provides `PyroOptim`, which is used to wrap PyTorch optimizers and manage optimizers for dynamically generated parameters (see the tutorial [SVI Part I](#) for a discussion). Any custom optimization algorithms are also to be found here.

## 8.1 Pyro Optimizers

**class** `PyroOptim`(*optim\_constructor*, *optim\_args*, *clip\_args=None*)

Bases: `object`

A wrapper for `torch.optim.Optimizer` objects that helps with managing dynamically generated parameters.

### Parameters

- **`optim_constructor`** – a `torch.optim.Optimizer`
- **`optim_args`** – a dictionary of learning arguments for the optimizer or a callable that returns such dictionaries
- **`clip_args`** – a dictionary of `clip_norm` and/or `clip_value` args or a callable that returns such dictionaries

**`__call__`**(*params*, *\*args*, *\*\*kwargs*)

**Parameters** **`params`** (*an iterable of strings*) – a list of parameters

Do an optimization step for each param in `params`. If a given param has never been seen before, initialize an optimizer for it.

**`get_state`**()

Get state associated with all the optimizers in the form of a dictionary with key-value pairs (parameter name, optim state dicts)

**`set_state`**(*state\_dict*)

Set the state associated with all the optimizers using the state obtained from a previous call to `get_state()`

**`save`**(*filename*)

**Parameters** `filename` (*str*) – file name to save to

Save optimizer state to disk

`load` (*filename*)

**Parameters** `filename` (*str*) – file name to load from

Load optimizer state from disk

**AdagradRMSPProp** (*optim\_args*)

Wraps `pyro.optim.adagrad_rmsprop.AdagradRMSPProp` with `PyroOptim`.

**ClippedAdam** (*optim\_args*)

Wraps `pyro.optim.clipped_adam.ClippedAdam` with `PyroOptim`.

**class PyroLRScheduler** (*scheduler\_constructor*, *optim\_args*, *clip\_args=None*)

Bases: `pyro.optim.optim.PyroOptim`

A wrapper for `lr_scheduler` objects that adjusts learning rates for dynamically generated parameters.

#### Parameters

- **scheduler\_constructor** – a `lr_scheduler`
- **optim\_args** – a dictionary of learning arguments for the optimizer or a callable that returns such dictionaries. must contain the key ‘optimizer’ with pytorch optimizer value
- **clip\_args** – a dictionary of `clip_norm` and/or `clip_value` args or a callable that returns such dictionaries.

Example:

```
optimizer = torch.optim.SGD
scheduler = pyro.optim.ExponentialLR({'optimizer': optimizer, 'optim_args': {'lr
→': 0.01}, 'gamma': 0.1})
svi = SVI(model, guide, scheduler, loss=TraceGraph_ELBO())
for i in range(epochs):
    for minibatch in DataLoader(dataset, batch_size):
        svi.step(minibatch)
    scheduler.step(epoch=i)
```

`__call__` (*params*, *\*args*, *\*\*kwargs*)

**step** (*\*args*, *\*\*kwargs*)

Takes the same arguments as the PyTorch scheduler (optional `epoch` kwarg or `loss` in for `ReduceLROnPlateau`)

**class AdagradRMSPProp** (*params*, *eta=1.0*, *delta=1e-16*, *t=0.1*)

Bases: `torch.optim.optimizer.Optimizer`

Implements a mash-up of the Adagrad algorithm and RMSProp. For the precise update equation see equations 10 and 11 in reference [1].

References: [1] ‘Automatic Differentiation Variational Inference’, Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, David M. Blei URL: <https://arxiv.org/abs/1603.00788> [2] ‘Lecture 6.5 RmsProp: Divide the gradient by a running average of its recent magnitude’, Tieleman, T. and Hinton, G., COURSERA: Neural Networks for Machine Learning. [3] ‘Adaptive subgradient methods for online learning and stochastic optimization’, Duchi, John, Hazan, E and Singer, Y.

Arguments:

#### Parameters

- **params** – iterable of parameters to optimize or dicts defining parameter groups

- **eta** (*float*) – sets the step size scale (optional; default: 1.0)
- **t** (*float*) – t, optional): momentum parameter (optional; default: 0.1)
- **delta** (*float*) – modulates the exponent that controls how the step size scales (optional; default: 1e-16)

**share\_memory** ()

**step** (*closure=None*)

Performs a single optimization step.

**Parameters** **closure** – A (optional) closure that reevaluates the model and returns the loss.

**class ClippedAdam** (*params, lr=0.001, betas=(0.9, 0.999), eps=1e-08, weight\_decay=0, clip\_norm=10.0, lrd=1.0*)

Bases: `torch.optim.optimizer.Optimizer`

**Parameters**

- **params** – iterable of parameters to optimize or dicts defining parameter groups
- **lr** – learning rate (default: 1e-3)
- **betas** (*Tuple*) – coefficients used for computing running averages of gradient and its square (default: (0.9, 0.999))
- **eps** – term added to the denominator to improve numerical stability (default: 1e-8)
- **weight\_decay** – weight decay (L2 penalty) (default: 0)
- **clip\_norm** – magnitude of norm to which gradients are clipped (default: 10.0)
- **lrd** – rate at which learning rate decays (default: 1.0)

Small modification to the Adam algorithm implemented in `torch.optim.Adam` to include gradient clipping and learning rate decay.

Reference

*A Method for Stochastic Optimization*, Diederik P. Kingma, Jimmy Ba <https://arxiv.org/abs/1412.6980>

**step** (*closure=None*)

**Parameters** **closure** – An optional closure that reevaluates the model and returns the loss.

Performs a single optimization step.

## 8.2 PyTorch Optimizers

**Adadelta** (*optim\_args, clip\_args=None*)

Wraps `torch.optim.Adadelta` with *PyroOptim*.

**Adagrad** (*optim\_args, clip\_args=None*)

Wraps `torch.optim.Adagrad` with *PyroOptim*.

**Adam** (*optim\_args, clip\_args=None*)

Wraps `torch.optim.Adam` with *PyroOptim*.

**AdamW** (*optim\_args, clip\_args=None*)

Wraps `torch.optim.AdamW` with *PyroOptim*.

**SparseAdam** (*optim\_args, clip\_args=None*)

Wraps `torch.optim.SparseAdam` with *PyroOptim*.

**Adamax** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.Adamax` with *PyroOptim*.

**ASGD** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.ASGD` with *PyroOptim*.

**SGD** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.SGD` with *PyroOptim*.

**Rprop** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.Rprop` with *PyroOptim*.

**RMSprop** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.RMSprop` with *PyroOptim*.

**LambdaLR** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.LambdaLR` with *PyroLRScheduler*.

**StepLR** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.StepLR` with *PyroLRScheduler*.

**MultiStepLR** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.MultiStepLR` with *PyroLRScheduler*.

**ExponentialLR** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.ExponentialLR` with *PyroLRScheduler*.

**CosineAnnealingLR** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.CosineAnnealingLR` with *PyroLRScheduler*.

**ReduceLROnPlateau** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.ReduceLROnPlateau` with *PyroLRScheduler*.

**CyclicLR** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.CyclicLR` with *PyroLRScheduler*.

**CosineAnnealingWarmRestarts** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.CosineAnnealingWarmRestarts` with *PyroLRScheduler*.

**OneCycleLR** (*optim\_args*, *clip\_args=None*)

Wraps `torch.optim.OneCycleLR` with *PyroLRScheduler*.

## 8.3 Higher-Order Optimizers

**class MultiOptimizer**

Bases: `object`

Base class of optimizers that make use of higher-order derivatives.

Higher-order optimizers generally use `torch.autograd.grad()` rather than `torch.Tensor.backward()`, and therefore require a different interface from usual Pyro and PyTorch optimizers. In this interface, the `step()` method inputs a loss tensor to be differentiated, and backpropagation is triggered one or more times inside the optimizer.

Derived classes must implement `step()` to compute derivatives and update parameters in-place.

Example:

```
tr = poutine.trace(model).get_trace(*args, **kwargs)
loss = -tr.log_prob_sum()
params = {name: site['value'].unconstrained()
          for name, site in tr.nodes.items()
          if site['type'] == 'param'}
optim.step(loss, params)
```

**step** (*loss*, *params*)

Performs an in-place optimization step on parameters given a differentiable `loss` tensor.

Note that this detaches the updated tensors.

#### Parameters

- **loss** (*torch.Tensor*) – A differentiable tensor to be minimized. Some optimizers require this to be differentiable multiple times.
- **params** (*dict*) – A dictionary mapping param name to unconstrained value as stored in the param store.

**get\_step** (*loss*, *params*)

Computes an optimization step of parameters given a differentiable `loss` tensor, returning the updated values.

Note that this preserves derivatives on the updated tensors.

#### Parameters

- **loss** (*torch.Tensor*) – A differentiable tensor to be minimized. Some optimizers require this to be differentiable multiple times.
- **params** (*dict*) – A dictionary mapping param name to unconstrained value as stored in the param store.

**Returns** A dictionary mapping param name to updated unconstrained value.

**Return type** *dict*

**class PyroMultiOptimizer** (*optim*)

Bases: *pyro.optim.multi.MultiOptimizer*

Facade to wrap *PyroOptim* objects in a *MultiOptimizer* interface.

**step** (*loss*, *params*)

**class TorchMultiOptimizer** (*optim\_constructor*, *optim\_args*)

Bases: *pyro.optim.multi.PyroMultiOptimizer*

Facade to wrap *Optimizer* objects in a *MultiOptimizer* interface.

**class MixedMultiOptimizer** (*parts*)

Bases: *pyro.optim.multi.MultiOptimizer*

Container class to combine different *MultiOptimizer* instances for different parameters.

**Parameters** *parts* (*list*) – A list of (*names*, *optim*) pairs, where each *names* is a list of parameter names, and each *optim* is a *MultiOptimizer* or *PyroOptim* object to be used for the named parameters. Together the *names* should partition up all desired parameters to optimize.

**Raises** **ValueError** – if any name is optimized by multiple optimizers.

**step** (*loss*, *params*)

**get\_step** (*loss*, *params*)

**class** `Newton` (*trust\_radii*={})

Bases: `pyro.optim.multi.MultiOptimizer`

Implementation of `MultiOptimizer` that performs a Newton update on batched low-dimensional variables, optionally regularizing via a per-parameter `trust_radius`. See `newton_step()` for details.

The result of `get_step()` will be differentiable, however the updated values from `step()` will be detached.

**Parameters** `trust_radii` (*dict*) – a dict mapping parameter name to radius of trust region.

Missing names will use unregularized Newton update, equivalent to infinite trust radius.

**get\_step** (*loss*, *params*)

---

## Poutine (Effect handlers)

---

Beneath the built-in inference algorithms, Pyro has a library of composable effect handlers for creating new inference algorithms and working with probabilistic programs. Pyro's inference algorithms are all built by applying these handlers to stochastic functions.

### 9.1 Handlers

Poutine is a library of composable effect handlers for recording and modifying the behavior of Pyro programs. These lower-level ingredients simplify the implementation of new inference algorithms and behavior.

Handlers can be used as higher-order functions, decorators, or context managers to modify the behavior of functions or blocks of code:

For example, consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

We can mark sample sites as observed using `condition`, which returns a callable with the same input and output signatures as `model`:

```
>>> conditioned_model = poutine.condition(model, data={"z": 1.0})
```

We can also use handlers as decorators:

```
>>> @pyro.condition(data={"z": 1.0})
... def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

Or as context managers:

```
>>> with pyro.condition(data={"z": 1.0}):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(0., s))
...     y = z ** 2
```

Handlers compose freely:

```
>>> conditioned_model = poutine.condition(model, data={"z": 1.0})
>>> traced_model = poutine.trace(conditioned_model)
```

Many inference algorithms or algorithmic components can be implemented in just a few lines of code:

```
guide_tr = poutine.trace(guide).get_trace(...)
model_tr = poutine.trace(poutine.replay(conditioned_model, trace=guide_tr)).get_
↳ trace(...)
monte_carlo_elbo = model_tr.log_prob_sum() - guide_tr.log_prob_sum()
```

**block** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of [BlockMessenger](#)

This handler selectively hides Pyro primitive sites from the outside world. Default behavior: block everything.

A site is hidden if at least one of the following holds:

0. `hide_fn(msg)` is `True` or `(not expose_fn(msg))` is `True`
1. `msg["name"]` in `hide`
2. `msg["type"]` in `hide_types`
3. `msg["name"]` not in `expose` and `msg["type"]` not in `expose_types`
4. `hide`, `hide_types`, and `expose_types` are all `None`

For example, suppose the stochastic function `fn` has two sample sites “a” and “b”. Then any effect outside of `BlockMessenger(fn, hide=["a"])` will not be applied to site “a” and will only see site “b”:

```
>>> def fn():
...     a = pyro.sample("a", dist.Normal(0., 1.))
...     return pyro.sample("b", dist.Normal(a, 1.))
>>> fn_inner = pyro.poutine.trace(fn)
>>> fn_outer = pyro.poutine.trace(pyro.poutine.block(fn_inner, hide=["a"]))
>>> trace_inner = fn_inner.get_trace()
>>> trace_outer = fn_outer.get_trace()
>>> "a" in trace_inner
True
>>> "a" in trace_outer
False
>>> "b" in trace_inner
True
>>> "b" in trace_outer
True
```

### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **hide\_fn** – function that takes a site and returns `True` to hide the site or `False/None` to expose it. If specified, all other parameters are ignored. Only specify one of `hide_fn` or `expose_fn`, not both.



- **expose\_fn** – function that takes a site and returns True to expose the site or False/None to hide it. If specified, all other parameters are ignored. Only specify one of `hide_fn` or `expose_fn`, not both.
- **hide\_all** (*bool*) – hide all sites
- **expose\_all** (*bool*) – expose all sites normally
- **hide** (*list*) – list of site names to hide
- **expose** (*list*) – list of site names to be exposed while all others hidden
- **hide\_types** (*list*) – list of site types to be hidden
- **expose\_types** (*list*) – list of site types to be exposed while all others hidden

**Returns** stochastic function decorated with a *BlockMessenger*

**broadcast** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of *BroadcastMessenger*

Automatically broadcasts the batch shape of the stochastic function at a sample site when inside a single or nested plate context. The existing *batch\_shape* must be broadcastable with the size of the *plate* contexts installed in the *cond\_indep\_stack*.

Notice how *model\_automatic\_broadcast* below automates expanding of distribution batch shapes. This makes it easy to modularize a Pyro model as the sub-components are agnostic of the wrapping *plate* contexts.

```
>>> def model_broadcast_by_hand():
...     with IndepMessenger("batch", 100, dim=-2):
...         with IndepMessenger("components", 3, dim=-1):
...             sample = pyro.sample("sample", dist.Bernoulli(torch.ones(3) * 0.5)
...                                     .expand_by(100))
...             assert sample.shape == torch.Size((100, 3))
...     return sample
```

```
>>> @poutine.broadcast
... def model_automatic_broadcast():
...     with IndepMessenger("batch", 100, dim=-2):
...         with IndepMessenger("components", 3, dim=-1):
...             sample = pyro.sample("sample", dist.Bernoulli(torch.tensor(0.5)))
...             assert sample.shape == torch.Size((100, 3))
...     return sample
```

**condition** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of *ConditionMessenger*

Given a stochastic function with some sample statements and a dictionary of observations at names, change the sample statements at those names into observes with those values.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

To observe a value for site *z*, we can write

```
>>> conditioned_model = pyro.poutine.condition(model, data={"z": torch.tensor(1.)})
↪
```

This is equivalent to adding *obs=value* as a keyword argument to *pyro.sample("z", ...)* in *model*.

#### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **data** – a dict or a *Trace*

**Returns** stochastic function decorated with a *ConditionMessenger*

**do** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of *DoMessenger*

Given a stochastic function with some sample statements and a dictionary of values at names, set the return values of those sites equal to the values as if they were hard-coded to those values and introduce fresh sample sites with the same names whose values do not propagate.

Composes freely with *condition()* to represent counterfactual distributions over potential outcomes. See Single World Intervention Graphs [1] for additional details and theory.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

To intervene with a value for site *z*, we can write

```
>>> intervened_model = pyro.poutine.do(model, data={"z": torch.tensor(1.)})
```

This is equivalent to replacing *z = pyro.sample("z", ...)* with *z = torch.tensor(1.)* and introducing a fresh sample site *pyro.sample("z", ...)* whose value is not used elsewhere.

#### References

[1] *Single World Intervention Graphs: A Primer*, Thomas Richardson, James Robins

#### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **data** – a dict mapping sample site names to interventions

**Returns** stochastic function decorated with a *DoMessenger*

**enum** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of *EnumMessenger*

Enumerates in parallel over discrete sample sites marked *infer={"enumerate": "parallel"}*.

**Parameters** **first\_available\_dim** (*int*) – The first tensor dimension (counting from the right) that is available for parallel enumeration. This dimension and all dimensions left may be used internally by Pyro. This should be a negative integer or None.

**escape** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of *EscapeMessenger*

Messenger that does a nonlocal exit by raising a *util.NonlocalExit* exception

**infer\_config** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of *InferConfigMessenger*

Given a callable *fn* that contains Pyro primitive calls and a callable *config\_fn* taking a trace site and returning a dictionary, updates the value of the infer kwarg at a sample site to *config\_fn(site)*.

#### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **config\_fn** – a callable taking a site and returning an infer dict

**Returns** stochastic function decorated with *InferConfigMessenger*

**lift** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of *LiftMessenger*

Given a stochastic function with param calls and a prior distribution, create a stochastic function where all param calls are replaced by sampling from prior. Prior should be a callable or a dict of names to callables.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
>>> lifted_model = pyro.poutine.lift(model, prior={"s": dist.Exponential(0.3)})
```

*lift* makes param statements behave like sample statements using the distributions in prior. In this example, site *s* will now behave as if it was replaced with `s = pyro.sample("s", dist.Exponential(0.3))`:

```
>>> tr = pyro.poutine.trace(lifted_model).get_trace(0.0)
>>> tr.nodes["s"]["type"] == "sample"
True
>>> tr2 = pyro.poutine.trace(lifted_model).get_trace(0.0)
>>> bool((tr2.nodes["s"]["value"] == tr.nodes["s"]["value"]).all())
False
```

#### Parameters

- **fn** – function whose parameters will be lifted to random values
- **prior** – prior function in the form of a Distribution or a dict of stochastic fns

**Returns** *fn* decorated with a *LiftMessenger*

**markov** (*fn=None, history=1, keep=False, dim=None, name=None*)

Markov dependency declaration.

This can be used in a variety of ways: - as a context manager - as a decorator for recursive functions - as an iterator for markov chains

#### Parameters

- **history** (*int*) – The number of previous contexts visible from the current context. Defaults to 1. If zero, this is similar to *pyro.plate*.
- **keep** (*bool*) – If true, frames are replayable. This is important when branching: if *keep=True*, neighboring branches at the same level can depend on each other; if *keep=False*, neighboring branches are independent (conditioned on their share)
- **dim** (*int*) – An optional dimension to use for this independence index. Interface stub, behavior not yet implemented.

- **name** (*str*) – An optional unique name to help inference algorithms match `pyro.markov()` sites between models and guides. Interface stub, behavior not yet implemented.

**mask** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of [MaskMessenger](#)

Given a stochastic function with some batched sample statements and masking tensor, mask out some of the sample statements elementwise.

#### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **mask** (*torch.BoolTensor*) – a  $\{0, 1\}$ -valued masking tensor (1 includes a site, 0 excludes a site)

**Returns** stochastic function decorated with a `MaskMessenger`

**queue** (*fn=None, queue=None, max\_tries=None, extend\_fn=None, escape\_fn=None, num\_samples=None*)

Used in sequential enumeration over discrete variables.

Given a stochastic function and a queue, return a return value from a complete trace in the queue.

#### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **queue** – a queue data structure like `multiprocessing.Queue` to hold partial traces
- **max\_tries** – maximum number of attempts to compute a single complete trace
- **extend\_fn** – function (possibly stochastic) that takes a partial trace and a site, and returns a list of extended traces
- **escape\_fn** – function (possibly stochastic) that takes a partial trace and a site, and returns a boolean value to decide whether to exit
- **num\_samples** – optional number of extended traces for `extend_fn` to return

**Returns** stochastic function decorated with poutine logic

**replay** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of [ReplayMessenger](#)

Given a callable that contains Pyro primitive calls, return a callable that runs the original, reusing the values at sites in trace at those sites in the new trace

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

`replay` makes sample statements behave as if they had sampled the values at the corresponding sites in the trace:

```
>>> old_trace = pyro.poutine.trace(model).get_trace(1.0)
>>> replayed_model = pyro.poutine.replay(model, trace=old_trace)
>>> bool(replayed_model(0.0) == old_trace.nodes["_RETURN"]["value"])
True
```

#### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **trace** – a [Trace](#) data structure to replay against
- **params** – dict of names of param sites and constrained values in fn to replay against

**Returns** a stochastic function decorated with a [ReplayMessenger](#)

**scale** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of [ScaleMessenger](#)

Given a stochastic function with some sample statements and a positive scale factor, scale the score of all sample and observe sites in the function.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s), obs=1.0)
...     return z ** 2
```

`scale` multiplicatively scales the log-probabilities of sample sites:

```
>>> scaled_model = pyro.poutine.scale(model, scale=0.5)
>>> scaled_tr = pyro.poutine.trace(scaled_model).get_trace(0.0)
>>> unscaled_tr = pyro.poutine.trace(model).get_trace(0.0)
>>> bool((scaled_tr.log_prob_sum() == 0.5 * unscaled_tr.log_prob_sum()).all())
True
```

### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **scale** – a positive scaling factor

**Returns** stochastic function decorated with a [ScaleMessenger](#)

**seed** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of [SeedMessenger](#)

Handler to set the random number generator to a pre-defined state by setting its seed. This is the same as calling `pyro.set_rng_seed()` before the call to *fn*. This handler has no additional effect on primitive statements on the standard Pyro backend, but it might intercept `pyro.sample` calls in other backends. e.g. the NumPy backend.

### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls).
- **rng\_seed** (*int*) – rng seed.

**trace** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of [TraceMessenger](#)

Return a handler that records the inputs and outputs of primitive calls and their dependencies.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

We can record its execution using `trace` and use the resulting data structure to compute the log-joint probability of all of the sample sites in the execution or extract all parameters.

```
>>> trace = pyro.poutine.trace(model).get_trace(0.0)
>>> logp = trace.log_prob_sum()
>>> params = [trace.nodes[name]["value"].unconstrained() for name in trace.param_
↳ nodes]
```

### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **graph\_type** – string that specifies the kind of graph to construct
- **param\_only** – if true, only records params and not samples

**Returns** stochastic function decorated with a *TraceMessenger*

**uncondition** (*fn=None, \*args, \*\*kwargs*)

Convenient wrapper of *UnconditionMessenger*

Messenger to force the value of observed nodes to be sampled from their distribution, ignoring observations.

**config\_enumerate** (*guide=None, default='parallel', expand=False, num\_samples=None, tmc='diagonal'*)

Configures enumeration for all relevant sites in a guide. This is mainly used in conjunction with *TraceEnum\_ELBO*.

When configuring for exhaustive enumeration of discrete variables, this configures all sample sites whose distribution satisfies `.has_enumerate_support == True`. When configuring for local parallel Monte Carlo sampling via `default="parallel"`, `num_samples=n`, this configures all sample sites. This does not overwrite existing annotations `infer={"enumerate": ...}`.

This can be used as either a function:

```
guide = config_enumerate(guide)
```

or as a decorator:

```
@config_enumerate
def guide1(*args, **kwargs):
    ...

@config_enumerate(default="sequential", expand=True)
def guide2(*args, **kwargs):
    ...
```

### Parameters

- **guide** (*callable*) – a pyro model that will be used as a guide in *SVI*.
- **default** (*str*) – Which enumerate strategy to use, one of “sequential”, “parallel”, or None. Defaults to “parallel”.
- **expand** (*bool*) – Whether to expand enumerated sample values. See *enumerate\_support()* for details. This only applies to exhaustive enumeration, where `num_samples=None`. If `num_samples` is not None, then this samples will always be expanded.

- **num\_samples** (*int* or *None*) – if not *None*, use local Monte Carlo sampling rather than exhaustive enumeration. This makes sense for both continuous and discrete distributions.
- **tmc** (*string* or *None*) – “mixture” or “diagonal” strategies to use in Tensor Monte Carlo

**Returns** an annotated guide

**Return type** callable

## 9.2 Trace

**class Trace** (*graph\_type='flat'*)

Bases: `object`

Graph data structure denoting the relationships amongst different pyro primitives in the execution trace.

An execution trace of a Pyro program is a record of every call to `pyro.sample()` and `pyro.param()` in a single execution of that program. Traces are directed graphs whose nodes represent primitive calls or input/output, and whose edges represent conditional dependence relationships between those primitive calls. They are created and populated by `poutine.trace`.

Each node (or site) in a trace contains the name, input and output value of the site, as well as additional metadata added by inference algorithms or user annotation. In the case of `pyro.sample`, the trace also includes the stochastic function at the site, and any observed data added by users.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

We can record its execution using `pyro.poutine.trace` and use the resulting data structure to compute the log-joint probability of all of the sample sites in the execution or extract all parameters.

```
>>> trace = pyro.poutine.trace(model).get_trace(0.0)
>>> logp = trace.log_prob_sum()
>>> params = [trace.nodes[name]["value"].unconstrained() for name in trace.param_
↳nodes]
```

We can also inspect or manipulate individual nodes in the trace. `trace.nodes` contains a `collections.OrderedDict` of site names and metadata corresponding to `x`, `s`, `z`, and the return value:

```
>>> list(name for name in trace.nodes.keys()) # doctest: +SKIP
["_INPUT", "s", "z", "_RETURN"]
```

Values of `trace.nodes` are dictionaries of node metadata:

```
>>> trace.nodes["z"] # doctest: +SKIP
{'type': 'sample', 'name': 'z', 'is_observed': False,
 'fn': Normal(), 'value': tensor(0.6480), 'args': (), 'kwargs': {},
 'infer': {}, 'scale': 1.0, 'cond_indep_stack': (),
 'done': True, 'stop': False, 'continuation': None}
```

'infer' is a dictionary of user- or algorithm-specified metadata. 'args' and 'kwargs' are the arguments passed via `pyro.sample` to `fn.__call__` or `fn.log_prob`. 'scale' is used to scale the log-probability of the site when computing the log-joint. 'cond\_indep\_stack' contains data structures corresponding to `pyro.plate` contexts appearing in the execution. 'done', 'stop', and 'continuation' are only used by Pyro's internals.

**Parameters** `graph_type` (*string*) – string specifying the kind of trace graph to construct

**add\_edge** (*site1*, *site2*)

**add\_node** (*site\_name*, *\*\*kwargs*)

**Parameters** `site_name` (*string*) – the name of the site to be added

Adds a site to the trace.

Raises an error when attempting to add a duplicate node instead of silently overwriting.

**compute\_log\_prob** (*site\_filter=<function Trace.<lambda>>>*)

Compute the site-wise log probabilities of the trace. Each `log_prob` has shape equal to the corresponding `batch_shape`. Each `log_prob_sum` is a scalar. Both computations are memoized.

**compute\_score\_parts** ()

Compute the batched local score parts at each site of the trace. Each `log_prob` has shape equal to the corresponding `batch_shape`. Each `log_prob_sum` is a scalar. All computations are memoized.

**copy** ()

Makes a shallow copy of self with nodes and edges preserved.

**detach\_** ()

Detach values (in-place) at each sample site of the trace.

**edges**

**format\_shapes** (*title='Trace Shapes:', last\_site=None*)

Returns a string showing a table of the shapes of all sites in the trace.

**iter\_stochastic\_nodes** ()

**Returns** an iterator over stochastic nodes in the trace.

**log\_prob\_sum** (*site\_filter=<function Trace.<lambda>>>*)

Compute the site-wise log probabilities of the trace. Each `log_prob` has shape equal to the corresponding `batch_shape`. Each `log_prob_sum` is a scalar. The computation of `log_prob_sum` is memoized.

**Returns** total log probability.

**Return type** `torch.Tensor`

**nonreparam\_stochastic\_nodes**

**Returns** a list of names of sample sites whose stochastic functions are not reparameterizable primitive distributions

**observation\_nodes**

**Returns** a list of names of observe sites

**pack\_tensors** (*plate\_to\_symbol=None*)

Computes packed representations of tensors in the trace. This should be called after `compute_log_prob()` or `compute_score_parts()`.

**param\_nodes**

**Returns** a list of names of param sites



**predecessors** (*site\_name*)

**remove\_node** (*site\_name*)

**reparameterized\_nodes**

**Returns** a list of names of sample sites whose stochastic functions are reparameterizable primitive distributions

**stochastic\_nodes**

**Returns** a list of names of sample sites

**successors** (*site\_name*)

**symbolize\_dims** (*plate\_to\_symbol=None*)

Assign unique symbols to all tensor dimensions.

**topological\_sort** (*reverse=False*)

Return a list of nodes (site names) in topologically sorted order.

**Parameters** **reverse** (*bool*) – Return the list in reverse order.

**Returns** list of topologically sorted nodes (site names).

## 9.3 Runtime

**exception NonlocalExit** (*site, \*args, \*\*kwargs*)

Bases: `Exception`

Exception for exiting nonlocally from poutine execution.

Used by `poutine.EscapeMessenger` to return site information.

**reset\_stack** ()

Reset the state of the frames remaining in the stack. Necessary for multiple re-executions in `poutine.queue`.

**am\_i\_wrapped** ()

Checks whether the current computation is wrapped in a poutine. :returns: `bool`

**apply\_stack** (*initial\_msg*)

Execute the effect stack at a single site according to the following scheme:

1. For each `Messenger` in the stack from bottom to top, execute `Messenger._process_message` with the message; if the message field “stop” is `True`, stop; otherwise, continue
2. Apply default behavior (`default_process_message`) to finish remaining site execution
3. For each `Messenger` in the stack from top to bottom, execute `_postprocess_message` to update the message and internal messenger state with the site results
4. If the message field “continuation” is not `None`, call it with the message

**Parameters** **initial\_msg** (*dict*) – the starting version of the trace site

**Returns** `None`

**default\_process\_message** (*msg*)

Default method for processing messages in inference.

**Parameters** **msg** – a message to be processed

**Returns** `None`

**effectful** (*fn=None, type=None*)

**Parameters**

- **fn** – function or callable that performs an effectful computation
- **type** (*str*) – the type label of the operation, e.g. “*sample*”

Wrapper for calling `apply_stack()` to apply any active effects.

## 9.4 Utilities

**all\_escape** (*trace, msg*)

**Parameters**

- **trace** – a partial trace
- **msg** – the message at a Pyro primitive site

**Returns** boolean decision value

Utility function that checks if a site is not already in a trace.

Used by EscapeMessenger to decide whether to do a nonlocal exit at a site. Subroutine for approximately integrating out variables for variance reduction.

**discrete\_escape** (*trace, msg*)

**Parameters**

- **trace** – a partial trace
- **msg** – the message at a Pyro primitive site

**Returns** boolean decision value

Utility function that checks if a sample site is discrete and not already in a trace.

Used by EscapeMessenger to decide whether to do a nonlocal exit at a site. Subroutine for integrating out discrete variables for variance reduction.

**enable\_validation** (*is\_validate*)

**enum\_extend** (*trace, msg, num\_samples=None*)

**Parameters**

- **trace** – a partial trace
- **msg** – the message at a Pyro primitive site
- **num\_samples** – maximum number of extended traces to return.

**Returns** a list of traces, copies of input trace with one extra site

Utility function to copy and extend a trace with sites based on the input site whose values are enumerated from the support of the input site’s distribution.

Used for exact inference and integrating out discrete variables.

**is\_validation\_enabled** ()

**mc\_extend** (*trace, msg, num\_samples=None*)

**Parameters**

- **trace** – a partial trace
- **msg** – the message at a Pyro primitive site
- **num\_samples** – maximum number of extended traces to return.

**Returns** a list of traces, copies of input trace with one extra site

Utility function to copy and extend a trace with sites based on the input site whose values are sampled from the input site's function.

Used for Monte Carlo marginalization of individual sample sites.

**prune\_subsample\_sites**(*trace*)

Copies and removes all subsample sites from a trace.

**site\_is\_subsample**(*site*)

Determines whether a trace site originated from a subsample statement inside an *plate*.

## 9.5 Messengers

Messenger objects contain the implementations of the effects exposed by handlers. Advanced users may modify the implementations of messengers behind existing handlers or write new messengers that implement new effects and compose correctly with the rest of the library.

### 9.5.1 Messenger

**class Messenger**

Bases: `object`

Context manager class that modifies behavior and adds side effects to stochastic functions i.e. callables containing Pyro primitive statements.

This is the base Messenger class. It implements the default behavior for all Pyro primitives, so that the joint distribution induced by a stochastic function `fn` is identical to the joint distribution induced by `Messenger()(fn)`.

Class of transformers for messages passed during inference. Most inference operations are implemented in subclasses of this.

**classmethod register**(*fn=None, type=None, post=None*)

**Parameters**

- **fn** – function implementing operation
- **type** (*str*) – name of the operation (also passed to `effectful()`)
- **post** (*bool*) – if *True*, use this operation as postprocess

Dynamically add operations to an effect. Useful for generating wrappers for libraries.

Example:

```
@SomeMessengerClass.register
def some_function(msg):
    ...do_something...
    return msg
```

**classmethod unregister**(*fn=None, type=None*)

### Parameters

- **fn** – function implementing operation
- **type** (*str*) – name of the operation (also passed to *effective()*)

Dynamically remove operations from an effect. Useful for removing wrappers from libraries.

Example:

```
SomeMessengerClass.unregister(some_function, "name")
```

## 9.5.2 BlockMessenger

**class BlockMessenger** (*hide\_fn=None, expose\_fn=None, hide\_all=True, expose\_all=False, hide=None, expose=None, hide\_types=None, expose\_types=None*)  
Bases: *pyro.poutine.messenger.Messenger*

This handler selectively hides Pyro primitive sites from the outside world. Default behavior: block everything.

A site is hidden if at least one of the following holds:

0. `hide_fn(msg)` is `True` or `(not expose_fn(msg))` is `True`
1. `msg["name"]` in `hide`
2. `msg["type"]` in `hide_types`
3. `msg["name"]` not in `expose` and `msg["type"]` not in `expose_types`
4. `hide`, `hide_types`, and `expose_types` are all `None`

For example, suppose the stochastic function `fn` has two sample sites “a” and “b”. Then any effect outside of `BlockMessenger(fn, hide=["a"])` will not be applied to site “a” and will only see site “b”:

```
>>> def fn():
...     a = pyro.sample("a", dist.Normal(0., 1.))
...     return pyro.sample("b", dist.Normal(a, 1.))
>>> fn_inner = pyro.poutine.trace(fn)
>>> fn_outer = pyro.poutine.trace(pyro.poutine.block(fn_inner, hide=["a"]))
>>> trace_inner = fn_inner.get_trace()
>>> trace_outer = fn_outer.get_trace()
>>> "a" in trace_inner
True
>>> "a" in trace_outer
False
>>> "b" in trace_inner
True
>>> "b" in trace_outer
True
```

### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **hide\_fn** – function that takes a site and returns `True` to hide the site or `False/None` to expose it. If specified, all other parameters are ignored. Only specify one of `hide_fn` or `expose_fn`, not both.
- **expose\_fn** – function that takes a site and returns `True` to expose the site or `False/None` to hide it. If specified, all other parameters are ignored. Only specify one of `hide_fn` or `expose_fn`, not both.

- `hide_all (bool)` – hide all sites
- `expose_all (bool)` – expose all sites normally
- `hide (list)` – list of site names to hide
- `expose (list)` – list of site names to be exposed while all others hidden
- `hide_types (list)` – list of site types to be hidden
- `expose_types (list)` – list of site types to be exposed while all others hidden

**Returns** stochastic function decorated with a *BlockMessenger*

### 9.5.3 BroadcastMessenger

**class BroadcastMessenger**

Bases: *pyro.poutine.messenger.Messenger*

Automatically broadcasts the batch shape of the stochastic function at a sample site when inside a single or nested plate context. The existing *batch\_shape* must be broadcastable with the size of the *plate* contexts installed in the *cond\_indep\_stack*.

Notice how *model\_automatic\_broadcast* below automates expanding of distribution batch shapes. This makes it easy to modularize a Pyro model as the sub-components are agnostic of the wrapping *plate* contexts.

```
>>> def model_broadcast_by_hand():
...     with IndepMessenger("batch", 100, dim=-2):
...         with IndepMessenger("components", 3, dim=-1):
...             sample = pyro.sample("sample", dist.Bernoulli(torch.ones(3) * 0.5)
...                                     .expand_by(100))
...             assert sample.shape == torch.Size((100, 3))
...     return sample
```

```
>>> @poutine.broadcast
... def model_automatic_broadcast():
...     with IndepMessenger("batch", 100, dim=-2):
...         with IndepMessenger("components", 3, dim=-1):
...             sample = pyro.sample("sample", dist.Bernoulli(torch.tensor(0.5)))
...             assert sample.shape == torch.Size((100, 3))
...     return sample
```

### 9.5.4 ConditionMessenger

**class ConditionMessenger (data)**

Bases: *pyro.poutine.messenger.Messenger*

Given a stochastic function with some sample statements and a dictionary of observations at names, change the sample statements at those names into observes with those values.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

To observe a value for site *z*, we can write

```
>>> conditioned_model = pyro.poutine.condition(model, data={"z": torch.tensor(1.)})
↪)
```

This is equivalent to adding *obs=value* as a keyword argument to *pyro.sample("z", ...)* in *model*.

#### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **data** – a dict or a *Trace*

**Returns** stochastic function decorated with a *ConditionMessenger*

## 9.5.5 DoMessenger

**class DoMessenger** (*data*)

Bases: *pyro.poutine.messenger.Messenger*

Given a stochastic function with some sample statements and a dictionary of values at names, set the return values of those sites equal to the values as if they were hard-coded to those values and introduce fresh sample sites with the same names whose values do not propagate.

Composes freely with *condition()* to represent counterfactual distributions over potential outcomes. See Single World Intervention Graphs [1] for additional details and theory.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

To intervene with a value for site *z*, we can write

```
>>> intervened_model = pyro.poutine.do(model, data={"z": torch.tensor(1.)})
```

This is equivalent to replacing *z = pyro.sample("z", ...)* with *z = torch.tensor(1.)* and introducing a fresh sample site *pyro.sample("z", ...)* whose value is not used elsewhere.

#### References

[1] *Single World Intervention Graphs: A Primer*, Thomas Richardson, James Robins

#### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **data** – a dict mapping sample site names to interventions

**Returns** stochastic function decorated with a *DoMessenger*

## 9.5.6 EnumMessenger

**class EnumMessenger** (*first\_available\_dim=None*)

Bases: *pyro.poutine.messenger.Messenger*

Enumerates in parallel over discrete sample sites marked *infer={"enumerate": "parallel"}*.

**Parameters** `first_available_dim` (*int*) – The first tensor dimension (counting from the right) that is available for parallel enumeration. This dimension and all dimensions left may be used internally by Pyro. This should be a negative integer or None.

`enumerate_site` (*msg*)

### 9.5.7 EscapeMessenger

**class** `EscapeMessenger` (*escape\_fn*)

Bases: `pyro.poutine.messenger.Messenger`

Messenger that does a nonlocal exit by raising a `util.NonlocalExit` exception

### 9.5.8 IndepMessenger

**class** `CondIndepStackFrame`

Bases: `pyro.poutine.indep_messenger.CondIndepStackFrame`

**vectorized**

**class** `IndepMessenger` (*name=None, size=None, dim=None, device=None*)

Bases: `pyro.poutine.messenger.Messenger`

This messenger keeps track of stack of independence information declared by nested `plate` contexts. This information is stored in a `cond_indep_stack` at each sample/observe site for consumption by `TraceMessenger`.

Example:

```
x_axis = IndepMessenger('outer', 320, dim=-1)
y_axis = IndepMessenger('inner', 200, dim=-2)
with x_axis:
    x_noise = sample("x_noise", dist.Normal(loc, scale).expand_by([320]))
with y_axis:
    y_noise = sample("y_noise", dist.Normal(loc, scale).expand_by([200, 1]))
with x_axis, y_axis:
    xy_noise = sample("xy_noise", dist.Normal(loc, scale).expand_by([200, 320]))
```

**indices**

`next_context()`

Increments the counter.

### 9.5.9 InferConfigMessenger

**class** `InferConfigMessenger` (*config\_fn*)

Bases: `pyro.poutine.messenger.Messenger`

Given a callable *fn* that contains Pyro primitive calls and a callable *config\_fn* taking a trace site and returning a dictionary, updates the value of the infer kwarg at a sample site to *config\_fn(site)*.

**Parameters**

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **config\_fn** – a callable taking a site and returning an infer dict

**Returns** stochastic function decorated with `InferConfigMessenger`

### 9.5.10 LiftMessenger

**class LiftMessenger** (*prior*)

Bases: `pyro.poutine.messenger.Messenger`

Given a stochastic function with param calls and a prior distribution, create a stochastic function where all param calls are replaced by sampling from prior. Prior should be a callable or a dict of names to callables.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
>>> lifted_model = pyro.poutine.lift(model, prior={"s": dist.Exponential(0.3)})
```

`lift` makes param statements behave like sample statements using the distributions in `prior`. In this example, site `s` will now behave as if it was replaced with `s = pyro.sample("s", dist.Exponential(0.3))`:

```
>>> tr = pyro.poutine.trace(lifted_model).get_trace(0.0)
>>> tr.nodes["s"]["type"] == "sample"
True
>>> tr2 = pyro.poutine.trace(lifted_model).get_trace(0.0)
>>> bool((tr2.nodes["s"]["value"] == tr.nodes["s"]["value"]).all())
False
```

#### Parameters

- **fn** – function whose parameters will be lifted to random values
- **prior** – prior function in the form of a Distribution or a dict of stochastic fns

**Returns** `fn` decorated with a `LiftMessenger`

### 9.5.11 MarkovMessenger

**class MarkovMessenger** (*history=1, keep=False, dim=None, name=None*)

Bases: `pyro.poutine.reentrant_messenger.ReentrantMessenger`

Markov dependency declaration.

This is a statistical equivalent of a memory management arena.

#### Parameters

- **history** (*int*) – The number of previous contexts visible from the current context. Defaults to 1. If zero, this is similar to `pyro.plate`.
- **keep** (*bool*) – If true, frames are replayable. This is important when branching: if `keep=True`, neighboring branches at the same level can depend on each other; if `keep=False`, neighboring branches are independent (conditioned on their shared ancestors).
- **dim** (*int*) – An optional dimension to use for this independence index. Interface stub, behavior not yet implemented.
- **name** (*str*) – An optional unique name to help inference algorithms match `pyro.markov()` sites between models and guides. Interface stub, behavior not yet implemented.



**generator** (*iterable*)

### 9.5.12 MaskMessenger

**class MaskMessenger** (*mask*)

Bases: `pyro.poutine.messenger.Messenger`

Given a stochastic function with some batched sample statements and masking tensor, mask out some of the sample statements elementwise.

#### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **mask** (`torch.BoolTensor`) – a  $\{0, 1\}$ -valued masking tensor (1 includes a site, 0 excludes a site)

**Returns** stochastic function decorated with a MaskMessenger

### 9.5.13 PlateMessenger

**class PlateMessenger** (*name, size=None, subsample\_size=None, subsample=None, dim=None, use\_cuda=None, device=None*)

Bases: `pyro.poutine.subsample_messenger.SubsampleMessenger`

Swiss army knife of broadcasting amazingness: combines shape inference, independence annotation, and sub-sampling

### 9.5.14 ReentrantMessenger

**class ReentrantMessenger**

Bases: `pyro.poutine.messenger.Messenger`

### 9.5.15 ReplayMessenger

**class ReplayMessenger** (*trace=None, params=None*)

Bases: `pyro.poutine.messenger.Messenger`

Given a callable that contains Pyro primitive calls, return a callable that runs the original, reusing the values at sites in trace at those sites in the new trace

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

`replay` makes sample statements behave as if they had sampled the values at the corresponding sites in the trace:

```
>>> old_trace = pyro.poutine.trace(model).get_trace(1.0)
>>> replayed_model = pyro.poutine.replay(model, trace=old_trace)
>>> bool(replayed_model(0.0) == old_trace.nodes["_RETURN"]["value"])
True
```

**Parameters**

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **trace** – a *Trace* data structure to replay against
- **params** – dict of names of param sites and constrained values in fn to replay against

**Returns** a stochastic function decorated with a *ReplayMessenger*

## 9.5.16 ScaleMessenger

**class ScaleMessenger** (*scale*)

Bases: *pyro.poutine.messenger.Messenger*

Given a stochastic function with some sample statements and a positive scale factor, scale the score of all sample and observe sites in the function.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s), obs=1.0)
...     return z ** 2
```

scale multiplicatively scales the log-probabilities of sample sites:

```
>>> scaled_model = pyro.poutine.scale(model, scale=0.5)
>>> scaled_tr = pyro.poutine.trace(scaled_model).get_trace(0.0)
>>> unscaled_tr = pyro.poutine.trace(model).get_trace(0.0)
>>> bool((scaled_tr.log_prob_sum() == 0.5 * unscaled_tr.log_prob_sum()).all())
True
```

**Parameters**

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **scale** – a positive scaling factor

**Returns** stochastic function decorated with a *ScaleMessenger*

## 9.5.17 SeedMessenger

**class SeedMessenger** (*rng\_seed*)

Bases: *pyro.poutine.messenger.Messenger*

Handler to set the random number generator to a pre-defined state by setting its seed. This is the same as calling *pyro.set\_rng\_seed()* before the call to *fn*. This handler has no additional effect on primitive statements on the standard Pyro backend, but it might intercept *pyro.sample* calls in other backends. e.g. the NumPy backend.

**Parameters**

- **fn** – a stochastic function (callable containing Pyro primitive calls).
- **rng\_seed** (*int*) – rng seed.

### 9.5.18 SubsampleMessenger

**class SubsampleMessenger** (*name, size=None, subsample\_size=None, subsample=None, dim=None, use\_cuda=None, device=None*)

Bases: `pyro.poutine.indep_messenger.IndepMessenger`

Extension of `IndepMessenger` that includes subsampling.

### 9.5.19 TraceMessenger

**class TraceHandler** (*msngr, fn*)

Bases: `object`

Execution trace poutine.

A `TraceHandler` records the input and output to every Pyro primitive and stores them as a site in a `Trace()`. This should, in theory, be sufficient information for every inference algorithm (along with the implicit computational graph in the `Variables?`)

We can also use this for visualization.

**get\_trace** (*\*args, \*\*kwargs*)

**Returns** data structure

**Return type** `pyro.poutine.Trace`

Helper method for a very common use case. Calls this poutine and returns its trace instead of the function's return value.

**trace**

**class TraceMessenger** (*graph\_type=None, param\_only=None*)

Bases: `pyro.poutine.messenger.Messenger`

Return a handler that records the inputs and outputs of primitive calls and their dependencies.

Consider the following Pyro program:

```
>>> def model(x):
...     s = pyro.param("s", torch.tensor(0.5))
...     z = pyro.sample("z", dist.Normal(x, s))
...     return z ** 2
```

We can record its execution using `trace` and use the resulting data structure to compute the log-joint probability of all of the sample sites in the execution or extract all parameters.

```
>>> trace = pyro.poutine.trace(model).get_trace(0.0)
>>> logp = trace.log_prob_sum()
>>> params = [trace.nodes[name]["value"].unconstrained() for name in trace.param_
↳ nodes]
```

#### Parameters

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **graph\_type** – string that specifies the kind of graph to construct
- **param\_only** – if true, only records params and not samples

**Returns** stochastic function decorated with a `TraceMessenger`

`get_trace()`

**Returns** data structure

**Return type** `pyro.poutine.Trace`

Helper method for a very common use case. Returns a shallow copy of `self.trace`.

`identify_dense_edges(trace)`

Modifies a trace in-place by adding all edges based on the `cond_indep_stack` information stored at each site.

## 9.5.20 UnconditionMessenger

**class** `UnconditionMessenger`

Bases: `pyro.poutine.messenger.Messenger`

Messenger to force the value of observed nodes to be sampled from their distribution, ignoring observations.

The `pyro.ops` module implements tensor utilities that are mostly independent of the rest of Pyro.

## 10.1 Utilities for HMC

**class** `DualAveraging` (`prox_center=0`, `t0=10`, `kappa=0.75`, `gamma=0.05`)

Bases: `object`

Dual Averaging is a scheme to solve convex optimization problems. It belongs to a class of subgradient methods which uses subgradients to update parameters (in primal space) of a model. Under some conditions, the averages of generated parameters during the scheme are guaranteed to converge to an optimal value. However, a counter-intuitive aspect of traditional subgradient methods is “new subgradients enter the model with decreasing weights” (see [1]). Dual Averaging scheme solves that phenomenon by updating parameters using weights equally for subgradients (which lie in a dual space), hence we have the name “dual averaging”.

This class implements a dual averaging scheme which is adapted for Markov chain Monte Carlo (MCMC) algorithms. To be more precise, we will replace subgradients by some statistics calculated during an MCMC trajectory. In addition, introducing some free parameters such as `t0` and `kappa` is helpful and still guarantees the convergence of the scheme.

### References

[1] *Primal-dual subgradient methods for convex problems*, Yurii Nesterov

[2] *The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo*, Matthew D. Hoffman, Andrew Gelman

### Parameters

- **`prox_center`** (`float`) – A “prox-center” parameter introduced in [1] which pulls the primal sequence towards it.
- **`t0`** (`float`) – A free parameter introduced in [2] that stabilizes the initial steps of the scheme.

- **kappa** (*float*) – A free parameter introduced in [2] that controls the weights of steps of the scheme. For a small **kappa**, the scheme will quickly forget states from early steps. This should be a number in (0.5, 1].
- **gamma** (*float*) – A free parameter which controls the speed of the convergence of the scheme.

**reset** ()

**step** (*g*)

Updates states of the scheme given a new statistic/subgradient *g*.

**Parameters** *g* (*float*) – A statistic calculated during an MCMC trajectory or subgradient.

**get\_state** ()

Returns the latest  $x_t$  and average of  $\{x_i\}_{i=1}^t$  in primal space.

**velocity\_verlet** (*z*, *r*, *potential\_fn*, *inverse\_mass\_matrix*, *step\_size*, *num\_steps=1*, *z\_grads=None*)

Second order symplectic integrator that uses the velocity verlet algorithm.

**Parameters**

- **z** (*dict*) – dictionary of sample site names and their current values (type *Tensor*).
- **r** (*dict*) – dictionary of sample site names and corresponding momenta (type *Tensor*).
- **potential\_fn** (*callable*) – function that returns potential energy given *z* for each sample site. The negative gradient of the function with respect to *z* determines the rate of change of the corresponding sites' momenta *r*.
- **inverse\_mass\_matrix** (*torch.Tensor*) – a tensor  $M^{-1}$  which is used to calculate kinetic energy:  $E_{kinetic} = \frac{1}{2}z^T M^{-1}z$ . Here *M* can be a 1D tensor (diagonal matrix) or a 2D tensor (dense matrix).
- **step\_size** (*float*) – step size for each time step iteration.
- **num\_steps** (*int*) – number of discrete time steps over which to integrate.
- **z\_grads** (*torch.Tensor*) – optional gradients of potential energy at current *z*.

**Return tuple** (*z\_next*, *r\_next*, *z\_grads*, *potential\_energy*) next position and momenta, together with the potential energy and its gradient w.r.t. *z\_next*.

**potential\_grad** (*potential\_fn*, *z*)

Gradient of *potential\_fn* w.r.t. parameters *z*.

**Parameters**

- **potential\_fn** – python callable that takes in a dictionary of parameters and returns the potential energy.
- **z** (*dict*) – dictionary of parameter values keyed by site name.

**Returns** tuple of (*z\_grads*, *potential\_energy*), where *z\_grads* is a dictionary with the same keys as *z* containing gradients and *potential\_energy* is a torch scalar.

**class WelfordCovariance** (*diagonal=True*)

Bases: *object*

Implements Welford's online scheme for estimating (co)variance (see [1]). Useful for adapting diagonal and dense mass structures for HMC.

**References**

[1] *The Art of Computer Programming*, Donald E. Knuth

```

reset ()
update (sample)
get_covariance (regularize=True)

```

## 10.2 Newton Optimizers

**newton\_step** (*loss, x, trust\_radius=None*)

Performs a Newton update step to minimize loss on a batch of variables, optionally constraining to a trust region [1].

This is especially useful because the final solution of newton iteration is differentiable wrt the inputs, even when all but the final `x` is detached, due to this method's quadratic convergence [2]. `loss` must be twice-differentiable as a function of `x`. If `loss` is 2+d-times differentiable, then the return value of this function is d-times differentiable.

When `loss` is interpreted as a negative log probability density, then the return values `mode, cov` of this function can be used to construct a Laplace approximation `MultivariateNormal(mode, cov)`.

**Warning:** Take care to detach the result of this function when used in an optimization loop. If you forget to detach the result of this function during optimization, then backprop will propagate through the entire iteration process, and worse will compute two extra derivatives for each step.

Example use inside a loop:

```

x = torch.zeros(1000, 2) # arbitrary initial value
for step in range(100):
    x = x.detach() # block gradients through previous steps
    x.requires_grad = True # ensure loss is differentiable wrt x
    loss = my_loss_function(x)
    x = newton_step(loss, x, trust_radius=1.0)
# the final x is still differentiable

```

[1] Yuan, Ya-xiang. ICIAM. Vol. 99. 2000. "A review of trust region algorithms for optimization." <ftp://ftp.cc.ac.cn/pub/yyx/papers/p995.pdf>

[2] Christianson, Bruce. Optimization Methods and Software 3.4 (1994) "Reverse accumulation and attractive fixed points." <http://uhra.herts.ac.uk/bitstream/handle/2299/4338/903839.pdf>

### Parameters

- **loss** (*torch.Tensor*) – A scalar function of `x` to be minimized.
- **x** (*torch.Tensor*) – A dependent variable of shape  $(N, D)$  where  $N$  is the batch size and  $D$  is a small number.
- **trust\_radius** (*float*) – An optional trust region `trust_radius`. The updated value `mode` of this function will be within `trust_radius` of the input `x`.

**Returns** A pair (`mode, cov`) where `mode` is an updated tensor of the same shape as the original value `x`, and `cov` is an estimate of the covariance  $D \times D$  matrix with `cov.shape == x.shape[:-1] + (D, D)`.

**Return type** tuple

**newton\_step\_1d** (*loss*, *x*, *trust\_radius=None*)

Performs a Newton update step to minimize loss on a batch of 1-dimensional variables, optionally regularizing to constrain to a trust region.

See [newton\\_step\(\)](#) for details.

**Parameters**

- **loss** (*torch.Tensor*) – A scalar function of *x* to be minimized.
- **x** (*torch.Tensor*) – A dependent variable with rightmost size of 1.
- **trust\_radius** (*float*) – An optional trust region *trust\_radius*. The updated value mode of this function will be within *trust\_radius* of the input *x*.

**Returns** A pair (*mode*, *cov*) where *mode* is an updated tensor of the same shape as the original value *x*, and *cov* is an estimate of the covariance 1x1 matrix with *cov.shape == x.shape[:-1] + (1,1)*.

**Return type** *tuple*

**newton\_step\_2d** (*loss*, *x*, *trust\_radius=None*)

Performs a Newton update step to minimize loss on a batch of 2-dimensional variables, optionally regularizing to constrain to a trust region.

See [newton\\_step\(\)](#) for details.

**Parameters**

- **loss** (*torch.Tensor*) – A scalar function of *x* to be minimized.
- **x** (*torch.Tensor*) – A dependent variable with rightmost size of 2.
- **trust\_radius** (*float*) – An optional trust region *trust\_radius*. The updated value mode of this function will be within *trust\_radius* of the input *x*.

**Returns** A pair (*mode*, *cov*) where *mode* is an updated tensor of the same shape as the original value *x*, and *cov* is an estimate of the covariance 2x2 matrix with *cov.shape == x.shape[:-1] + (2,2)*.

**Return type** *tuple*

**newton\_step\_3d** (*loss*, *x*, *trust\_radius=None*)

Performs a Newton update step to minimize loss on a batch of 3-dimensional variables, optionally regularizing to constrain to a trust region.

See [newton\\_step\(\)](#) for details.

**Parameters**

- **loss** (*torch.Tensor*) – A scalar function of *x* to be minimized.
- **x** (*torch.Tensor*) – A dependent variable with rightmost size of 2.
- **trust\_radius** (*float*) – An optional trust region *trust\_radius*. The updated value mode of this function will be within *trust\_radius* of the input *x*.

**Returns** A pair (*mode*, *cov*) where *mode* is an updated tensor of the same shape as the original value *x*, and *cov* is an estimate of the covariance 3x3 matrix with *cov.shape == x.shape[:-1] + (3,3)*.

**Return type** *tuple*



## 10.3 Tensor Indexing

**vindex**(*tensor*, *args*)

Vectorized advanced indexing with broadcasting semantics.

See also the convenience wrapper *Vindex*.

This is useful for writing indexing code that is compatible with batching and enumeration, especially for selecting mixture components with discrete random variables.

For example suppose *x* is a parameter with `x.dim() == 3` and we wish to generalize the expression `x[i, :, j]` from integer *i*, *j* to tensors *i*, *j* with batch dims and enum dims (but no event dims). Then we can write the generalize version using *Vindex*

```
xij = Vindex(x)[i, :, j]

batch_shape = broadcast_shape(i.shape, j.shape)
event_shape = (x.size(1),)
assert xij.shape == batch_shape + event_shape
```

To handle the case when *x* may also contain batch dimensions (e.g. if *x* was sampled in a plated context as when using vectorized particles), *vindex()* uses the special convention that `Ellipsis` denotes batch dimensions (hence `...` can appear only on the left, never in the middle or in the right). Suppose *x* has event dim 3. Then we can write:

```
old_batch_shape = x.shape[:-3]
old_event_shape = x.shape[-3:]

xij = Vindex(x)[..., i, :, j]  # The ... denotes unknown batch shape.

new_batch_shape = broadcast_shape(old_batch_shape, i.shape, j.shape)
new_event_shape = (x.size(1),)
assert xij.shape == new_batch_shape + new_event_shape
```

Note that this special handling of `Ellipsis` differs from the NEP [1].

Formally, this function assumes:

1. Each *arg* is either `Ellipsis`, `slice(None)`, an integer, or a batched `torch.LongTensor` (i.e. with empty event shape). This function does not support Nontrivial slices or `torch.BoolTensor` masks. `Ellipsis` can only appear on the left as `args[0]`.
2. If `args[0]` is not `Ellipsis` then *tensor* is not batched, and its event dim is equal to `len(args)`.
3. If `args[0]` is `Ellipsis` then *tensor* is batched and its event dim is equal to `len(args[1:])`. Dims of *tensor* to the left of the event dims are considered batch dims and will be broadcasted with dims of *tensor* *args*.

Note that if none of the *args* is a tensor with `.dim() > 0`, then this function behaves like standard indexing:

```
if not any(isinstance(a, torch.Tensor) and a.dim() for a in args):
    assert Vindex(x)[args] == x[args]
```

### References

- [1] <https://www.numpy.org/neps/nep-0021-advanced-indexing.html> introduces *vindex* as a helper for vectorized indexing. The Pyro implementation is similar to the proposed notation `x.vindex[]` except for slightly different handling of `Ellipsis`.

**Parameters**

- **tensor** (*torch.Tensor*) – A tensor to be indexed.
- **args** (*tuple*) – An index, as args to `__getitem__`.

**Returns** A nonstandard interpretation of `tensor[args]`.

**Return type** `torch.Tensor`

**class** `Vindex` (*tensor*)

Bases: `object`

Convenience wrapper around `vindex()`.

The following are equivalent:

```
Vindex(x)[..., i, j, :]  
vindex(x, (Ellipsis, i, j, slice(None)))
```

**Parameters** **tensor** (*torch.Tensor*) – A tensor to be indexed.

**Returns** An object with a special `__getitem__()` method.

## 10.4 Tensor Contraction

**contract\_expression** (*equation, \*shapes, \*\*kwargs*)

Wrapper around `opt_einsum.contract_expression()` that optionally uses Pyro’s cheap optimizer and optionally caches contraction paths.

**Parameters** **cache\_path** (*bool*) – whether to cache the contraction path. Defaults to True.

**contract** (*equation, \*operands, \*\*kwargs*)

Wrapper around `opt_einsum.contract()` that optionally uses Pyro’s cheap optimizer and optionally caches contraction paths.

**Parameters** **cache\_path** (*bool*) – whether to cache the contraction path. Defaults to True.

**einsum** (*equation, \*operands, \*\*kwargs*)

Generalized plated sum-product algorithm via tensor variable elimination.

This generalizes `contract()` in two ways:

1. Multiple outputs are allowed, and intermediate results can be shared.
2. Inputs and outputs can be plated along symbols given in `plates`; reductions along `plates` are product reductions.

The best way to understand this function is to try the examples below, which show how `einsum()` calls can be implemented as multiple calls to `contract()` (which is generally more expensive).

To illustrate multiple outputs, note that the following are equivalent:

```
z1, z2, z3 = einsum('ab,bc->a,b,c', x, y) # multiple outputs  
  
z1 = contract('ab,bc->a', x, y)  
z2 = contract('ab,bc->b', x, y)  
z3 = contract('ab,bc->c', x, y)
```

To illustrate plated inputs, note that the following are equivalent:

```
assert len(x) == 3 and len(y) == 3
z = einsum('ab,ai,bi->b', w, x, y, plates='i')

z = contract('ab,a,a,a,b,b,b->b', w, *x, *y)
```

When a sum dimension  $a$  always appears with a plate dimension  $i$ , then  $a$  corresponds to a distinct symbol for each slice of  $a$ . Thus the following are equivalent:

```
assert len(x) == 3 and len(y) == 3
z = einsum('ai,ai->', x, y, plates='i')

z = contract('a,b,c,a,b,c->', *x, *y)
```

When such a sum dimension appears in the output, it must be accompanied by all of its plate dimensions, e.g. the following are equivalent:

```
assert len(x) == 3 and len(y) == 3
z = einsum('abi,abi->bi', x, y, plates='i')

z0 = contract('ab,ac,ad,ab,ac,ad->b', *x, *y)
z1 = contract('ab,ac,ad,ab,ac,ad->c', *x, *y)
z2 = contract('ab,ac,ad,ab,ac,ad->d', *x, *y)
z = torch.stack([z0, z1, z2])
```

Note that each plate slice through the output is multilinear in all plate slices through all inputs, thus e.g. batch matrix multiply would be implemented *without* plates, so the following are all equivalent:

```
xy = einsum('abc,acd->abd', x, y, plates='')
xy = torch.stack([xa.mm(ya) for xa, ya in zip(x, y)])
xy = torch.bmm(x, y)
```

Among all valid equations, some computations are polynomial in the sizes of the input tensors and other computations are exponential in the sizes of the input tensors. This function raises `NotImplementedError` whenever the computation is exponential.

#### Parameters

- **equation** (*str*) – An einsum equation, optionally with multiple outputs.
- **operands** (*torch.Tensor*) – A collection of tensors.
- **plates** (*str*) – An optional string of plate symbols.
- **backend** (*str*) – An optional einsum backend, defaults to ‘torch’.
- **cache** (*dict*) – An optional `shared_intermediates()` cache.
- **modulo\_total** (*bool*) – Optionally allow einsum to arbitrarily scale each result plate, which can significantly reduce computation. This is safe to set whenever each result plate denotes a nonnormalized probability distribution whose total is not of interest.

**Returns** a tuple of tensors of requested shape, one entry per output.

**Return type** *tuple*

#### Raises

- **ValueError** – if tensor sizes mismatch or an output requests a plated dim without that dim’s plates.
- **NotImplementedError** – if contraction would have cost exponential in the size of any input tensor.

**ubersum** (*equation*, \**operands*, \*\**kwargs*)

Deprecated, use `einsum()` instead.

## 10.5 Gaussian Contraction

**class Gaussian** (*log\_normalizer*, *info\_vec*, *precision*)

Bases: `object`

Non-normalized Gaussian distribution.

This represents an arbitrary semidefinite quadratic function, which can be interpreted as a rank-deficient scaled Gaussian distribution. The precision matrix may have zero eigenvalues, thus it may be impossible to work directly with the covariance matrix.

### Parameters

- **log\_normalizer** (`torch.Tensor`) – a normalization constant, which is mainly used to keep track of normalization terms during contractions.
- **info\_vec** (`torch.Tensor`) – information vector, which is a scaled version of the mean  $\text{info\_vec} = \text{precision} @ \text{mean}$ . We use this representation to make gaussian contraction fast and stable.
- **precision** (`torch.Tensor`) – precision matrix of this gaussian.

**dim**()

**batch\_shape**

**expand** (*batch\_shape*)

**reshape** (*batch\_shape*)

**\_\_getitem\_\_** (*index*)

Index into the batch\_shape of a Gaussian.

**static cat** (*parts*, *dim=0*)

Concatenate a list of Gaussians along a given batch dimension.

**event\_pad** (*left=0*, *right=0*)

Pad along event dimension.

**event\_permute** (*perm*)

Permute along event dimension.

**\_\_add\_\_** (*other*)

Adds two Gaussians in log-density space.

**log\_density** (*value*)

Evaluate the log density of this Gaussian at a point value:

```
-0.5 * value.T @ precision @ value + value.T @ info_vec + log_normalizer
```

This is mainly used for testing.

**condition** (*value*)

Condition this Gaussian on a trailing subset of its state. This should satisfy:

```
g.condition(y).dim() == g.dim() - y.size(-1)
```

Note that since this is a non-normalized Gaussian, we include the density of  $y$  in the result. Thus `condition()` is similar to a `functools.partial` binding of arguments:

```
left = x[..., :n]
right = x[..., n:]
g.log_density(x) == g.condition(right).log_density(left)
```

**marginalize** (*left=0, right=0*)

Marginalizing out variables on either side of the event dimension:

```
g.marginalize(left=n).event_logsumexp() = g.logsumexp()
g.marginalize(right=n).event_logsumexp() = g.logsumexp()
```

and for data  $x$ :

```
g.condition(x).event_logsumexp() = g.marginalize(left=g.dim() - x.size(-1)).log_density(x)
```

**event\_logsumexp()**

Integrates out all latent state (i.e. operating on event dimensions).

**class AffineNormal** (*matrix, loc, scale*)

Bases: `object`

Represents a conditional diagonal normal distribution over a random variable  $Y$  whose mean is an affine function of a random variable  $X$ . The likelihood of  $X$  is thus:

```
AffineNormal(matrix, loc, scale).condition(y).log_density(x)
```

which is equivalent to:

```
Normal(x @ matrix + loc, scale).to_event(1).log_prob(y)
```

### Parameters

- **matrix** (*torch.Tensor*) – A transformation from  $X$  to  $Y$ . Should have rightmost shape  $(x\_dim, y\_dim)$ .
- **loc** (*torch.Tensor*) – A constant offset for  $Y$ 's mean. Should have rightmost shape  $(y\_dim,)$ .
- **scale** (*torch.Tensor*) – Standard deviation for  $Y$ . Should have rightmost shape  $(y\_dim,)$ .

**condition** (*value*)

Condition on a  $Y$  value.

**Parameters** *value* (*torch.Tensor*) – A value of  $Y$ .

**Return Gaussian** A gaussian likelihood over  $X$ .

**mvn\_to\_gaussian** (*mvn*)

Convert a MultivariateNormal distribution to a Gaussian.

**Parameters** *mvn* (*MultivariateNormal*) – A multivariate normal distribution.

**Returns** An equivalent Gaussian object.

**Return type** *Gaussian*

**matrix\_and\_mvn\_to\_gaussian** (*matrix, mvn*)

Convert a noisy affine function to a Gaussian. The noisy affine function is defined as:

```
y = x @ matrix + mvn.sample()
```

**Parameters**

- **matrix** (*Tensor*) – A matrix with rightmost shape  $(x\_dim, y\_dim)$ .
- **mvn** (*MultivariateNormal*) – A multivariate normal distribution.

**Returns** A Gaussian with broadcasted batch shape and `.dim() == x_dim + y_dim`.

**Return type** *Gaussian*

**gaussian\_tensordot** (*x, y, dims=0*)

Computes the integral over two gaussians:

$$(x @ y)(a,c) = \log(\text{integral}(\exp(x(a,b) + y(b,c)), b)),$$

where  $x$  is a gaussian over variables  $(a,b)$ ,  $y$  is a gaussian over variables  $(b,c)$ ,  $(a,b,c)$  can each be sets of zero or more variables, and  $dims$  is the size of  $b$ .

**Parameters**

- **x** – a Gaussian instance
- **y** – a Gaussian instance
- **dims** – number of variables to contract

## 10.6 Statistical Utilities

**gelman\_rubin** (*input, chain\_dim=0, sample\_dim=1*)

Computes R-hat over chains of samples. It is required that `input.size(sample_dim) >= 2` and `input.size(chain_dim) >= 2`.

**Parameters**

- **input** (*torch.Tensor*) – the input tensor.
- **chain\_dim** (*int*) – the chain dimension.
- **sample\_dim** (*int*) – the sample dimension.

**Returns** *torch.Tensor* R-hat of `input`.

**split\_gelman\_rubin** (*input, chain\_dim=0, sample\_dim=1*)

Computes R-hat over chains of samples. It is required that `input.size(sample_dim) >= 4`.

**Parameters**

- **input** (*torch.Tensor*) – the input tensor.
- **chain\_dim** (*int*) – the chain dimension.
- **sample\_dim** (*int*) – the sample dimension.

**Returns** *torch.Tensor* split R-hat of `input`.

**autocorrelation** (*input, dim=0*)

Computes the autocorrelation of samples at dimension `dim`.

Reference: [https://en.wikipedia.org/wiki/Autocorrelation#Efficient\\_computation](https://en.wikipedia.org/wiki/Autocorrelation#Efficient_computation)

**Parameters**

- **input** (`torch.Tensor`) – the input tensor.
- **dim** (`int`) – the dimension to calculate autocorrelation.

Returns `torch.Tensor` autocorrelation of `input`.

**autocovariance** (`input, dim=0`)

Computes the autocovariance of samples at dimension `dim`.

#### Parameters

- **input** (`torch.Tensor`) – the input tensor.
- **dim** (`int`) – the dimension to calculate autocorrelation.

Returns `torch.Tensor` autocorrelation of `input`.

**effective\_sample\_size** (`input, chain_dim=0, sample_dim=1`)

Computes effective sample size of `input`.

Reference:

[1] *Introduction to Markov Chain Monte Carlo*, Charles J. Geyer

[2] *Stan Reference Manual version 2.18*, Stan Development Team

#### Parameters

- **input** (`torch.Tensor`) – the input tensor.
- **chain\_dim** (`int`) – the chain dimension.
- **sample\_dim** (`int`) – the sample dimension.

Returns `torch.Tensor` effective sample size of `input`.

**resample** (`input, num_samples, dim=0, replacement=False`)

Draws `num_samples` samples from `input` at dimension `dim`.

#### Parameters

- **input** (`torch.Tensor`) – the input tensor.
- **num\_samples** (`int`) – the number of samples to draw from `input`.
- **dim** (`int`) – dimension to draw from `input`.

Returns `torch.Tensor` samples drawn randomly from `input`.

**quantile** (`input, probs, dim=0`)

Computes quantiles of `input` at `probs`. If `probs` is a scalar, the output will be squeezed at `dim`.

#### Parameters

- **input** (`torch.Tensor`) – the input tensor.
- **probs** (`list`) – quantile positions.
- **dim** (`int`) – dimension to take quantiles from `input`.

Returns `torch.Tensor` quantiles of `input` at `probs`.

**pi** (`input, prob, dim=0`)

Computes percentile interval which assigns equal probability mass to each tail of the interval.

#### Parameters

- **input** (`torch.Tensor`) – the input tensor.

- **prob** (*float*) – the probability mass of samples within the interval.
- **dim** (*int*) – dimension to calculate percentile interval from `input`.

Returns **torch.Tensor** quantiles of `input` at `probs`.

**hpdi** (*input, prob, dim=0*)

Computes “highest posterior density interval” which is the narrowest interval with probability mass `prob`.

**Parameters**

- **input** (*torch.Tensor*) – the input tensor.
- **prob** (*float*) – the probability mass of samples within the interval.
- **dim** (*int*) – dimension to calculate percentile interval from `input`.

Returns **torch.Tensor** quantiles of `input` at `probs`.

**waic** (*input, log\_weights=None, pointwise=False, dim=0*)

Computes “Widely Applicable/Watanabe-Akaike Information Criterion” (WAIC) and its corresponding effective number of parameters.

Reference:

[1] *WAIC and cross-validation in Stan*, Aki Vehtari, Andrew Gelman

**Parameters**

- **input** (*torch.Tensor*) – the input tensor, which is log likelihood of a model.
- **log\_weights** (*torch.Tensor*) – weights of samples along `dim`.
- **dim** (*int*) – the sample dimension of `input`.

Returns **tuple** tuple of WAIC and effective number of parameters.

**fit\_generalized\_pareto** (*X*)

Given a dataset `X` assumed to be drawn from the Generalized Pareto Distribution, estimate the distributional parameters `k`, `sigma` using a variant of the technique described in reference [1], as described in reference [2].

References [1] ‘A new and efficient estimation method for the generalized Pareto distribution.’ Zhang, J. and Stephens, M.A. (2009). [2] ‘Pareto Smoothed Importance Sampling.’ Aki Vehtari, Andrew Gelman, Jonah Gabry

**Parameters** **torch.Tensor** – the input data `X`

**Returns** **tuple** tuple of floats (`k`, `sigma`) corresponding to the fit parameters

**crps\_empirical** (*pred, truth*)

Computes negative Continuous Ranked Probability Score CRPS\* [1] between a set of samples `pred` and true data `truth`. This uses an  $n \log(n)$  time algorithm to compute a quantity equal that would naively have complexity quadratic in the number of samples `n`:

```
CRPS* = E|pred - truth| - 1/2 E|pred - pred'|  
      = (pred - truth).abs().mean(0)  
      - (pred - pred.unsqueeze(1)).abs().mean([0, 1]) / 2
```

Note that for a single sample this reduces to absolute error.

References [1] *Strictly Proper Scoring Rules, Prediction, and Estimation* Tilmann Gneiting, Adrian E. Raftery (2007) <https://www.stat.washington.edu/raftery/Research/PDF/Gneiting2007jasa.pdf>

**Parameters**



- **pred** (`torch.Tensor`) – A set of sample predictions batched on rightmost dim. This should have shape `(num_samples,) + truth.shape`.
- **truth** (`torch.Tensor`) – A tensor of true observations.

**Returns** A tensor of shape `truth.shape`.

**Return type** `torch.Tensor`

## 10.7 State Space Model and GP Utilities

**class** `MaternKernel` (`nu=1.5`, `num_gps=1`, `length_scale_init=None`, `kernel_scale_init=None`)

Bases: `pyro.nn.module.PyroModule`

Provides the building blocks for representing univariate Gaussian Processes (GPs) with Matern kernels as state space models.

### Parameters

- **nu** (`float`) – The order of the Matern kernel (one of 0.5, 1.5 or 2.5)
- **num\_gps** (`int`) – the number of GPs
- **length\_scale\_init** (`torch.Tensor`) – optional `num_gps`-dimensional vector of initializers for the length scale
- **kernel\_scale\_init** (`torch.Tensor`) – optional `num_gps`-dimensional vector of initializers for the kernel scale

### References

- [1] *Kalman Filtering and Smoothing Solutions to Temporal Gaussian Process Regression Models*, Jouni Hartikainen and Simo Sarkka.
- [2] *Stochastic Differential Equation Methods for Spatio-Temporal Gaussian Process Regression*, Arno Solin.

**transition\_matrix** (`dt`)

Compute the (exponentiated) transition matrix of the GP latent space. The resulting matrix has layout `(num_gps, old_state, new_state)`, i.e. this matrix multiplies states from the right.

See section 5 in reference [1] for details.

**Parameters** `dt` (`float`) – the time interval over which the GP latent space evolves.

**Returns** `torch.Tensor` a 3-dimensional tensor of transition matrices of shape `(num_gps, state_dim, state_dim)`.

**stationary\_covariance** ()

Compute the stationary state covariance. See Eqn. 3.26 in reference [2].

**Returns** `torch.Tensor` a 3-dimensional tensor of covariance matrices of shape `(num_gps, state_dim, state_dim)`.

**process\_covariance** (`A`)

Given a transition matrix `A` computed with `transition_matrix` compute the the process covariance as described in Eqn. 3.11 in reference [2].

**Returns** `torch.Tensor` a batched covariance matrix of shape `(num_gps, state_dim, state_dim)`

**transition\_matrix\_and\_covariance** (`dt`)

Get the transition matrix and process covariance corresponding to a time interval `dt`.

**Parameters** `dt` (*float*) – the time interval over which the GP latent space evolves.

**Returns tuple** (*transition\_matrix*, *process\_covariance*) both 3-dimensional tensors of shape (num\_gps, state\_dim, state\_dim)

---

Automatic Name Generation

---

The `pyro.contrib.autoname` module provides tools for automatically generating unique, semantically meaningful names for sample sites.

**scope** (*fn=None, prefix=None, inner=None*)

**Parameters**

- **fn** – a stochastic function (callable containing Pyro primitive calls)
- **prefix** – a string to prepend to sample names (optional if *fn* is provided)
- **inner** – switch to determine where duplicate name counters appear

**Returns** *fn* decorated with a `ScopeMessenger`

`scope` prepends a prefix followed by a `/` to the name at a Pyro sample site. It works much like TensorFlow's `name_scope` and `variable_scope`, and can be used as a context manager, a decorator, or a higher-order function.

`scope` is very useful for aligning compositional models with guides or data.

Example:

```
>>> @scope(prefix="a")
... def model():
...     return pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "a/x" in poutine.trace(model).get_trace()
```

Example:

```
>>> def model():
...     with scope(prefix="a"):
...         return pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "a/x" in poutine.trace(model).get_trace()
```

Scopes compose as expected, with outer scopes appearing before inner scopes in names:

```
>>> @scope(prefix="b")
... def model():
...     with scope(prefix="a"):
...         return pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "b/a/x" in poutine.trace(model).get_trace()
```

When used as a decorator or higher-order function, `scope` will use the name of the input function as the prefix if no user-specified prefix is provided.

Example:

```
>>> @scope
... def model():
...     return pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "model/x" in poutine.trace(model).get_trace()
```

**name\_count** (*fn=None*)

`name_count` is a very simple autonaming scheme that simply appends a suffix “\_\_” plus a counter to any name that appears multiple times in an execution. Only duplicate instances of a name get a suffix; the first instance is not modified.

Example:

```
>>> @name_count
... def model():
...     for i in range(3):
...         pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "x" in poutine.trace(model).get_trace()
>>> assert "x__1" in poutine.trace(model).get_trace()
>>> assert "x__2" in poutine.trace(model).get_trace()
```

`name_count` also composes with `scope()` by adding a suffix to duplicate scope entrances:

Example:

```
>>> @name_count
... def model():
...     for i in range(3):
...         with pyro.contrib.autoname.scope(prefix="a"):
...             pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "a/x" in poutine.trace(model).get_trace()
>>> assert "a__1/x" in poutine.trace(model).get_trace()
>>> assert "a__2/x" in poutine.trace(model).get_trace()
```

Example:

```
>>> @name_count
... def model():
...     with pyro.contrib.autoname.scope(prefix="a"):
...         for i in range(3):
...             pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "a/x" in poutine.trace(model).get_trace()
```

(continues on next page)

(continued from previous page)

```
>>> assert "a/x__1" in poutine.trace(model).get_trace()
>>> assert "a/x__2" in poutine.trace(model).get_trace()
```

## 11.1 Named Data Structures

The `pyro.contrib.named` module is a thin syntactic layer on top of Pyro. It allows Pyro models to be written to look like programs with operating on Python data structures like `latent.x.sample(...)`, rather than programs with string-labeled statements like `x = pyro.sample("x", ...)`.

This module provides three container data structures `named.Object`, `named.List`, and `named.Dict`. These data structures are intended to be nested in each other. Together they track the address of each piece of data in each data structure, so that this address can be used as a Pyro site. For example:

```
>>> state = named.Object("state")
>>> print(str(state))
state

>>> z = state.x.y.z # z is just a placeholder.
>>> print(str(z))
state.x.y.z

>>> state.xs = named.List() # Create a contained list.
>>> x0 = state.xs.add()
>>> print(str(x0))
state.xs[0]

>>> state.ys = named.Dict()
>>> foo = state.ys['foo']
>>> print(str(foo))
state.ys['foo']
```

These addresses can now be used inside `sample`, `observe` and `param` statements. These named data structures even provide in-place methods that alias Pyro statements. For example:

```
>>> state = named.Object("state")
>>> loc = state.loc.param_(torch.zeros(1, requires_grad=True))
>>> scale = state.scale.param_(torch.ones(1, requires_grad=True))
>>> z = state.z.sample_(dist.Normal(loc, scale))
>>> obs = state.x.sample_(dist.Normal(loc, scale), obs=z)
```

For deeper examples of how these can be used in model code, see the [Tree Data](#) and [Mixture](#) examples.

Authors: Fritz Obermeyer, Alexander Rush

**class** `Object` (*name*)

Bases: `object`

Object to hold immutable latent state.

This object can serve either as a container for nested latent state or as a placeholder to be replaced by a tensor via a `named.sample`, `named.observe`, or `named.param` statement. When used as a placeholder, `Object` objects take the place of strings in normal `pyro.sample` statements.

**Parameters** `name` (*str*) – The name of the object.

Example:

```
state = named.Object("state")
state.x = 0
state.ys = named.List()
state.zs = named.Dict()
state.a.b.c.d.e.f.g = 0 # Creates a chain of named.Objects.
```

**Warning:** This data structure is write-once: data may be added but may not be mutated or removed. Trying to mutate this data structure may result in silent errors.

**sample\_**(*fn*, \**args*, \*\**kwargs*)

Calls the stochastic function *fn* with additional side-effects depending on *name* and the enclosing context (e.g. an inference algorithm). See [Intro I](#) and [Intro II](#) for a discussion.

**Parameters**

- **name** – name of sample
- **fn** – distribution class or function
- **obs** – observed datum (optional; should only be used in context of inference) optionally specified in *kwargs*
- **infer** (*dict*) – Optional dictionary of inference parameters specified in *kwargs*. See inference documentation for details.

**Returns** sample

**param\_**( \**args*, \*\**kwargs*)

Saves the variable as a parameter in the param store. To interact with the param store or write to disk, see [Parameters](#).

**Parameters**

- **name** (*str*) – name of parameter
- **init\_tensor** (*torch.Tensor* or *callable*) – initial tensor or lazy callable that returns a tensor. For large tensors, it may be cheaper to write e.g. `lambda: torch.randn(100000)`, which will only be evaluated on the initial statement.
- **constraint** (*torch.distributions.constraints.Constraint*) – torch constraint, defaults to `constraints.real`.
- **event\_dim** (*int*) – (optional) number of rightmost dimensions unrelated to batching. Dimension to the left of this will be considered batch dimensions; if the param statement is inside a subsampled plate, then corresponding batch dimensions of the parameter will be correspondingly subsampled. If unspecified, all dimensions will be considered event dims and no subsampling will be performed.

**Returns** parameter

**Return type** `torch.Tensor`

**class List** (*name=None*)

Bases: `list`

List-like object to hold immutable latent state.

This must either be given a name when constructed:

```
latent = named.List("root")
```

or must be immediately stored in a `named.Object`:

```
latent = named.Object("root")
latent.xs = named.List() # Must be bound to a Object before use.
```

**Warning:** This data structure is write-once: data may be added but may not be mutated or removed. Trying to mutate this data structure may result in silent errors.

**add()**

Append one new `named.Object`.

**Returns** a new latent object at the end

**Return type** *named.Object*

**class Dict** (*name=None*)

Bases: `dict`

Dict-like object to hold immutable latent state.

This must either be given a name when constructed:

```
latent = named.Dict("root")
```

or must be immediately stored in a `named.Object`:

```
latent = named.Object("root")
latent.xs = named.Dict() # Must be bound to a Object before use.
```

**Warning:** This data structure is write-once: data may be added but may not be mutated or removed. Trying to mutate this data structure may result in silent errors.

## 11.2 Scoping

`pyro.contrib.autoname.scoping` contains the implementation of `pyro.contrib.autoname.scope()`, a tool for automatically appending a semantically meaningful prefix to names of sample sites.

**class NameCountMessenger**

Bases: `pyro.poutine.messenger.Messenger`

`NameCountMessenger` is the implementation of `pyro.contrib.autoname.name_count()`

**class ScopeMessenger** (*prefix=None, inner=None*)

Bases: `pyro.poutine.messenger.Messenger`

`ScopeMessenger` is the implementation of `pyro.contrib.autoname.scope()`

**scope** (*fn=None, prefix=None, inner=None*)

**Parameters**

- **fn** – a stochastic function (callable containing Pyro primitive calls)

- **prefix** – a string to prepend to sample names (optional if `fn` is provided)
- **inner** – switch to determine where duplicate name counters appear

**Returns** fn decorated with a *ScopeMessenger*

`scope` prepends a prefix followed by a `/` to the name at a Pyro sample site. It works much like TensorFlow's `name_scope` and `variable_scope`, and can be used as a context manager, a decorator, or a higher-order function.

scope is very useful for aligning compositional models with guides or data.

Example:

```
>>> @scope(prefix="a")
... def model():
...     return pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "a/x" in poutine.trace(model).get_trace()
```

Example:

```
>>> def model():
...     with scope(prefix="a"):
...         return pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "a/x" in poutine.trace(model).get_trace()
```

Scopes compose as expected, with outer scopes appearing before inner scopes in names:

```
>>> @scope(prefix="b")
... def model():
...     with scope(prefix="a"):
...         return pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "b/a/x" in poutine.trace(model).get_trace()
```

When used as a decorator or higher-order function, `scope` will use the name of the input function as the prefix if no user-specified prefix is provided.

Example:

```
>>> @scope
... def model():
...     return pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "model/x" in poutine.trace(model).get_trace()
```

**name\_count** (*fn=None*)

name\_count is a very simple autonaming scheme that simply appends a suffix “\_\_” plus a counter to any name that appears multiple times in an execution. Only duplicate instances of a name get a suffix; the first instance is not modified.

Example:

```
>>> @name_count
... def model():
...     for i in range(3):
...         pyro.sample("x", dist.Bernoulli(0.5))
... 
```

(continues on next page)



(continued from previous page)

```
>>> assert "x" in poutine.trace(model).get_trace()
>>> assert "x__1" in poutine.trace(model).get_trace()
>>> assert "x__2" in poutine.trace(model).get_trace()
```

`name_count` also composes with `scope()` by adding a suffix to duplicate scope entrances:

Example:

```
>>> @name_count
... def model():
...     for i in range(3):
...         with pyro.contrib.autoname.scope(prefix="a"):
...             pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "a/x" in poutine.trace(model).get_trace()
>>> assert "a__1/x" in poutine.trace(model).get_trace()
>>> assert "a__2/x" in poutine.trace(model).get_trace()
```

Example:

```
>>> @name_count
... def model():
...     with pyro.contrib.autoname.scope(prefix="a"):
...         for i in range(3):
...             pyro.sample("x", dist.Bernoulli(0.5))
...
>>> assert "a/x" in poutine.trace(model).get_trace()
>>> assert "a/x__1" in poutine.trace(model).get_trace()
>>> assert "a/x__2" in poutine.trace(model).get_trace()
```



## 12.1 HiddenLayer

```
class HiddenLayer (X=None, A_mean=None, A_scale=None, non_linearity=<function  

    relu>, KL_factor=1.0, A_prior_scale=1.0, include_hidden_bias=True,  

    weight_space_sampling=False)
```

This distribution is a basic building block in a Bayesian neural network. It represents a single hidden layer, i.e. an affine transformation applied to a set of inputs  $X$  followed by a non-linearity. The uncertainty in the weights is encoded in a Normal variational distribution specified by the parameters  $A\_scale$  and  $A\_mean$ . The so-called ‘local reparameterization trick’ is used to reduce variance (see reference below). In effect, this means the weights are never sampled directly; instead one samples in pre-activation space (i.e. before the non-linearity is applied). Since the weights are never directly sampled, when this distribution is used within the context of variational inference, care must be taken to correctly scale the KL divergence term that corresponds to the weight matrix. This term is folded into the *log\_prob* method of this distributions.

In effect, this distribution encodes the following generative process:

$A \sim \text{Normal}(A\_mean, A\_scale)$  output  $\sim \text{non\_linearity}(AX)$

### Parameters

- **X** (*torch.Tensor*) – B x D dimensional mini-batch of inputs
- **A\_mean** (*torch.Tensor*) – D x H dimensional specifying weight mean
- **A\_scale** (*torch.Tensor*) – D x H dimensional (diagonal covariance matrix) specifying weight uncertainty
- **non\_linearity** (*callable*) – a callable that specifies the non-linearity used. defaults to ReLU.
- **KL\_factor** (*float*) – scaling factor for the KL divergence. prototypically this is equal to the size of the mini-batch divided by the size of the whole dataset. defaults to 1.0.
- **A\_prior** (*float or torch.Tensor*) – the prior over the weights is assumed to be normal with mean zero and scale factor  $A\_prior$ . default value is 1.0.

- **include\_hidden\_bias** (*bool*) – controls whether the activations should be augmented with a 1, which can be used to incorporate bias terms. defaults to *True*.
- **weight\_space\_sampling** (*bool*) – controls whether the local reparameterization trick is used. this is only intended to be used for internal testing. defaults to *False*.

Reference:

Kingma, Diederik P., Tim Salimans, and Max Welling. “Variational dropout and the local reparameterization trick.” *Advances in Neural Information Processing Systems*. 2015.

This module implements the Causal Effect Variational Autoencoder [1], which demonstrates a number of innovations including:

- a generative model for causal effect inference with hidden confounders;
- a model and guide with twin neural nets to allow imbalanced treatment; and
- a custom training loss that includes both ELBO terms and extra terms needed to train the guide to be able to answer counterfactual queries.

The main interface is the `CEVAE` class, but users may customize by using components `Model`, `Guide`, `TraceCausalEffect_ELBO` and utilities.

### References

[1] C. Louizos, U. Shalit, J. Mooij, D. Sontag, R. Zemel, M. Welling (2017).

Causal Effect Inference with Deep Latent-Variable Models.

<http://papers.nips.cc/paper/7223-causal-effect-inference-with-deep-latent-variable-models.pdf>

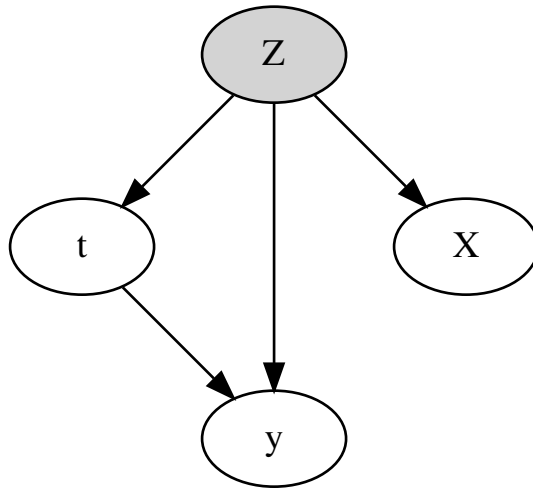
<https://github.com/AMLab-Amsterdam/CEVAE>

## 13.1 CEVAE Class

```
class CEVAE (feature_dim, outcome_dist='bernoulli', latent_dim=20, hidden_dim=200, num_layers=3,  
             num_samples=100)
```

```
Bases: torch.nn.modules.module.Module
```

Main class implementing a Causal Effect VAE [1]. This assumes a graphical model



where  $t$  is a binary treatment variable,  $y$  is an outcome,  $Z$  is an unobserved confounder, and  $X$  is a noisy function of the hidden confounder  $Z$ .

Example:

```
cevae = CEVAE(feature_dim=5)
cevae.fit(x_train, t_train, y_train)
ite = cevae.ite(x_test) # individual treatment effect
ate = ite.mean()       # average treatment effect
```

### Variables

- **model** (`Model`) – Generative model.
- **guide** (`Guide`) – Inference model.

### Parameters

- **feature\_dim** (`int`) – Dimension of the feature space  $x$ .
- **outcome\_dist** (`str`) – One of: “bernoulli” (default), “exponential”, “laplace”, “normal”, “studentt”.
- **latent\_dim** (`int`) – Dimension of the latent variable  $z$ . Defaults to 20.
- **hidden\_dim** (`int`) – Dimension of hidden layers of fully connected networks. Defaults to 200.
- **num\_layers** (`int`) – Number of hidden layers in fully connected networks.
- **num\_samples** (`int`) – Default number of samples for the `ite()` method. Defaults to 100.

**fit** ( $x$ ,  $t$ ,  $y$ , `num_epochs=100`, `batch_size=100`, `learning_rate=0.001`, `learning_rate_decay=0.1`, `weight_decay=0.0001`)  
Train using *SVI* with the *TraceCausalEffect\_ELBO* loss.

**Parameters**

- **x** (*Tensor*) –
- **t** (*Tensor*) –
- **y** (*Tensor*) –
- **num\_epochs** (*int*) – Number of training epochs. Defaults to 100.
- **batch\_size** (*int*) – Batch size. Defaults to 100.
- **learning\_rate** (*float*) – Learning rate. Defaults to 1e-3.
- **learning\_rate\_decay** (*float*) – Learning rate decay over all epochs; the per-step decay rate will depend on batch size and number of epochs such that the initial learning rate will be `learning_rate` and the final learning rate will be `learning_rate * learning_rate_decay`. Defaults to 0.1.
- **weight\_decay** (*float*) – Weight decay. Defaults to 1e-4.

**Returns** list of epoch losses

**ite** (*x*, *num\_samples=None*, *batch\_size=None*)

Computes Individual Treatment Effect for a batch of data *x*.

$$ITE(x) = \mathbb{E}[y \mid \mathbf{X} = x, do(t = 1)] - \mathbb{E}[y \mid \mathbf{X} = x, do(t = 0)]$$

This has complexity  $O(\text{len}(x) * \text{num\_samples} ** 2)$ .

**Parameters**

- **x** (*Tensor*) – A batch of data.
- **num\_samples** (*int*) – The number of monte carlo samples. Defaults to `self.num_samples` which defaults to 100.
- **batch\_size** (*int*) – Batch size. Defaults to `len(x)`.

**Returns** A `len(x)`-sized tensor of estimated effects.

**Return type** *Tensor*

**to\_script\_module** ()

Compile this module using `torch.jit.trace_module()`, assuming `self` has already been fit to data.

**Returns** A traced version of `self` with an `ite()` method.

**Return type** `torch.jit.ScriptModule`

## 13.2 CEVAE Components

**class Model** (*config*)

Bases: `pyro.nn.module.PyroModule`

Generative model for a causal model with latent confounder *z* and binary treatment *t*:

```
z ~ p(z)           # latent confounder
x ~ p(x|z)         # partial noisy observation of z
t ~ p(t|z)         # treatment, whose application is biased by z
y ~ p(y|t, z)      # outcome
```

Each of these distributions is defined by a neural network. The  $y$  distribution is defined by a disjoint pair of neural networks defining  $p(y|t=0, z)$  and  $p(y|t=1, z)$ ; this allows highly imbalanced treatment.

**Parameters** `config` (*dict*) – A dict specifying `feature_dim`, `latent_dim`, `hidden_dim`, `num_layers`, and `outcome_dist`.

**forward** ( $x, t=None, y=None, size=None$ )

**y\_mean** ( $x, t=None$ )

**z\_dist** ()

**x\_dist** ( $z$ )

**y\_dist** ( $t, z$ )

**t\_dist** ( $z$ )

**class** `Guide` (*config*)

Bases: `pyro.nn.module.PyroModule`

Inference model for causal effect estimation with latent confounder  $z$  and binary treatment  $t$ :

```
t ~ p(t|x)      # treatment
y ~ p(y|t,x)    # outcome
z ~ p(t|y,t,x)  # latent confounder, an embedding
```

Each of these distributions is defined by a neural network. The  $y$  and  $z$  distributions are defined by disjoint pairs of neural networks defining  $p(-|t=0, \dots)$  and  $p(-|t=1, \dots)$ ; this allows highly imbalanced treatment.

**Parameters** `config` (*dict*) – A dict specifying `feature_dim`, `latent_dim`, `hidden_dim`, `num_layers`, and `outcome_dist`.

**forward** ( $x, t=None, y=None, size=None$ )

**t\_dist** ( $x$ )

**y\_dist** ( $t, x$ )

**z\_dist** ( $y, t, x$ )

**class** `TraceCausalEffect_ELBO` (*num\_particles=1*, *max\_plate\_nesting=inf*,  
*max\_iarange\_nesting=None*, *vectorize\_particles=False*,  
*strict\_enumeration\_warning=True*, *ignore\_jit\_warnings=False*,  
*jit\_options=None*, *retain\_graph=None*, *tail\_adaptive\_beta=-1.0*)

Bases: `pyro.infer.trace_elbo.Trace_ELBO`

Loss function for training a *CEVAE*. From [1], the CEVAE objective (to maximize) is:

```
-loss = ELBO + log q(t|x) + log q(y|t,x)
```

**loss** (*model*, *guide*, *\*args*, *\*\*kwargs*)

## 13.3 Utilities

**class** `FullyConnected` (*sizes*, *final\_activation=None*)

Bases: `torch.nn.modules.container.Sequential`

Fully connected multi-layer network with ELU activations.

**append** (*layer*)



**class DistributionNet**Bases: `torch.nn.modules.module.Module`

Base class for distribution nets.

**static get\_class** (*dtype*)

Get a subclass by a prefix of its name, e.g.:

```
assert DistributionNet.get_class("bernoulli") is BernoulliNet
```

**class BernoulliNet** (*sizes*)Bases: `pyro.contrib.cevae.DistributionNet`*FullyConnected* network outputting a single logits value.

This is used to represent a conditional probability distribution of a single Bernoulli random variable conditioned on a `sizes[0]`-sized real value, for example:

```
net = BernoulliNet([3, 4])
z = torch.randn(3)
logits, = net(z)
t = net.make_dist(logits).sample()
```

**forward** (*x*)**static make\_dist** (*logits*)**class ExponentialNet** (*sizes*)Bases: `pyro.contrib.cevae.DistributionNet`*FullyConnected* network outputting a constrained rate.

This is used to represent a conditional probability distribution of a single Normal random variable conditioned on a `sizes[0]`-size real value, for example:

```
net = ExponentialNet([3, 4])
x = torch.randn(3)
rate, = net(x)
y = net.make_dist(rate).sample()
```

**forward** (*x*)**static make\_dist** (*rate*)**class LaplaceNet** (*sizes*)Bases: `pyro.contrib.cevae.DistributionNet`*FullyConnected* network outputting a constrained loc, scale pair.

This is used to represent a conditional probability distribution of a single Laplace random variable conditioned on a `sizes[0]`-size real value, for example:

```
net = LaplaceNet([3, 4])
x = torch.randn(3)
loc, scale = net(x)
y = net.make_dist(loc, scale).sample()
```

**forward** (*x*)**static make\_dist** (*loc, scale*)**class NormalNet** (*sizes*)Bases: `pyro.contrib.cevae.DistributionNet`

*FullyConnected* network outputting a constrained `loc`, `scale` pair.

This is used to represent a conditional probability distribution of a single Normal random variable conditioned on a `sizes[0]`-size real value, for example:

```
net = NormalNet([3, 4])
x = torch.randn(3)
loc, scale = net(x)
y = net.make_dist(loc, scale).sample()
```

**forward** (*x*)

**static make\_dist** (*loc*, *scale*)

**class StudentTNet** (*sizes*)

Bases: *pyro.contrib.cevae.DistributionNet*

*FullyConnected* network outputting a constrained `df`, `loc`, `scale` triple, with shared `df > 1`.

This is used to represent a conditional probability distribution of a single Student's t random variable conditioned on a `sizes[0]`-size real value, for example:

```
net = StudentTNet([3, 4])
x = torch.randn(3)
df, loc, scale = net(x)
y = net.make_dist(df, loc, scale).sample()
```

**forward** (*x*)

**static make\_dist** (*df*, *loc*, *scale*)

**class DiagNormalNet** (*sizes*)

Bases: `torch.nn.modules.module.Module`

*FullyConnected* network outputting a constrained `loc`, `scale` pair.

This is used to represent a conditional probability distribution of a `sizes[-1]`-sized diagonal Normal random variable conditioned on a `sizes[0]`-size real value, for example:

```
net = DiagNormalNet([3, 4, 5])
z = torch.randn(3)
loc, scale = net(z)
x = dist.Normal(loc, scale).sample()
```

This is intended for the latent `z` distribution and the prewhitened `x` features, and conservatively clips `loc` and `scale` values.

**forward** (*x*)

## 14.1 EasyGuide

**class EasyGuide** (*model*)

Bases: *pyro.nn.module.PyroModule*

Base class for “easy guides”, which are more flexible than `AutoGuide`s, but are easier to write than raw Pyro guides.

Derived classes should define a *guide()* method. This *guide()* method can combine ordinary guide statements (e.g. `pyro.sample` and `pyro.param`) with the following special statements:

- `group = self.group(...)` selects multiple `pyro.sample` sites in the model. See `Group` for subsequent methods.
- `with self.plate(...):` ... should be used instead of `pyro.plate`.
- `self.map_estimate(...)` uses a `Delta` guide for a single site.

Derived classes may also override the *init()* method to provide custom initialization for models sites.

**Parameters** *model* (*callable*) – A Pyro model.

**model**

**guide** (*\*args, \*\*kwargs*)

Guide implementation, to be overridden by user.

**init** (*site*)

Model initialization method, may be overridden by user.

This should input a site and output a valid sample from that site. The default behavior is to draw a random sample:

```
return site["fn"]()
```

For other possible initialization functions see <http://docs.pyro.ai/en/stable/infer.autoguide.html#module-pyro.infer.autoguide.initialization>

**forward** (\*args, \*\*kwargs)

Runs the guide. This is typically used by inference algorithms.

**plate** (name, size=None, subsample\_size=None, subsample=None, \*args, \*\*kwargs)

A wrapper around `pyro.plate` to allow *EasyGuide* to automatically construct plates. You should use this rather than `pyro.plate` inside your `guide()` implementation.

**group** (match='.\*')

Select a *Group* of model sites for joint guidance.

**Parameters** **match** (*str*) – A regex string matching names of model sample sites.

**Returns** A group of model sites.

**Return type** *Group*

**map\_estimate** (name)

Construct a maximum a posteriori (MAP) guide using Delta distributions.

**Parameters** **name** (*str*) – The name of a model sample site.

**Returns** A sampled value.

**Return type** `torch.Tensor`

## 14.2 easy\_guide

**easy\_guide** (model)

Convenience decorator to create an *EasyGuide*. The following are equivalent:

```
# Version 1. Decorate a function.
@easy_guide(model)
def guide(self, foo, bar):
    return my_guide(foo, bar)

# Version 2. Create and instantiate a subclass of EasyGuide.
class Guide(EasyGuide):
    def guide(self, foo, bar):
        return my_guide(foo, bar)
guide = Guide(model)
```

Note `@easy_guide` wrappers cannot be pickled; to build a guide that can be pickled, instead subclass from *EasyGuide*.

**Parameters** **model** (*callable*) – a Pyro model.

## 14.3 Group

**class Group** (guide, sites)

Bases: `object`

An autoguide helper to match a group of model sites.

**Variables**

- **event\_shape** (`torch.Size`) – The total flattened concatenated shape of all matching sample sites in the model.

- **prototype\_sites** (*list*) – A list of all matching sample sites in a prototype trace of the model.

**Parameters**

- **guide** (*EasyGuide*) – An easyguide instance.
- **sites** (*list*) – A list of model sites.

**guide**

**sample** (*guide\_name*, *fn*, *infer=None*)

Wrapper around `pyro.sample()` to create a single auxiliary sample site and then unpack to multiple sample sites for model replay.

**Parameters**

- **guide\_name** (*str*) – The name of the auxiliary guide site.
- **fn** (*callable*) – A distribution with shape `self.event_shape`.
- **infer** (*dict*) – Optional inference configuration dict.

**Returns** A pair (`guide_z`, `model_zs`) where `guide_z` is the single concatenated blob and `model_zs` is a dict mapping site name to constrained model sample.

**Return type** *tuple*

**map\_estimate** ()

Construct a maximum a posteriori (MAP) guide using Delta distributions.

**Returns** A dict mapping model site name to sampled value.

**Return type** *dict*



### **15.1 Datasets**

#### **15.1.1 Multi MNIST**

#### **15.1.2 BART Ridership**

### **15.2 Utilities**





See the [Gaussian Processes tutorial](#) for an introduction.

### **class Parameterized**

Bases: *pyro.nn.module.PyroModule*

A wrapper of *PyroModule* whose parameters can be set constraints, set priors.

By default, when we set a prior to a parameter, an auto Delta guide will be created. We can use the method *autoguide()* to setup other auto guides.

Example:

```
>>> class Linear(Parameterized):
...     def __init__(self, a, b):
...         super(Linear, self).__init__()
...         self.a = Parameter(a)
...         self.b = Parameter(b)
...
...     def forward(self, x):
...         return self.a * x + self.b
...
>>> linear = Linear(torch.tensor(1.), torch.tensor(0.))
>>> linear.a = PyroParam(torch.tensor(1.), constraints.positive)
>>> linear.b = PyroSample(dist.Normal(0, 1))
>>> linear.autoguide("b", dist.Normal)
>>> assert "a_unconstrained" in dict(linear.named_parameters())
>>> assert "b_loc" in dict(linear.named_parameters())
>>> assert "b_scale_unconstrained" in dict(linear.named_parameters())
```

Note that by default, data of a parameter is a float `torch.Tensor` (unless we use `torch.set_default_tensor_type()` to change default tensor type). To cast these parameters to a correct data type or GPU device, we can call methods such as `double()` or `cuda()`. See `torch.nn.Module` for more information.

**set\_prior** (*name, prior*)

Sets prior for a parameter.

**Parameters**

- **name** (*str*) – Name of the parameter.
- **prior** (*Distribution*) – A Pyro prior distribution.

**autoguide** (*name, dist\_constructor*)

Sets an autoguide for an existing parameter with name *name* (mimic the behavior of module `pyro.infer.autoguide`).

---

**Note:** *dist\_constructor* should be one of `Delta`, `Normal`, and `MultivariateNormal`. More distribution constructor will be supported in the future if needed.

---

**Parameters**

- **name** (*str*) – Name of the parameter.
- **dist\_constructor** – A `Distribution` constructor.

**set\_mode** (*mode*)

Sets *mode* of this object to be able to use its parameters in stochastic functions. If *mode*="model", a parameter will get its value from its prior. If *mode*="guide", the value will be drawn from its guide.

---

**Note:** This method automatically sets *mode* for submodules which belong to `Parameterized` class.

---

**Parameters mode** (*str*) – Either “model” or “guide”.

*mode*

## 16.1 Models

### 16.1.1 GPModel

**class GPModel** (*X, y, kernel, mean\_function=None, jitter=1e-06*)

Bases: `pyro.contrib.gp.parameterized.Parameterized`

Base class for Gaussian Process models.

The core of a Gaussian Process is a covariance function *k* which governs the similarity between input points. Given *k*, we can establish a distribution over functions *f* by a multivariate normal distribution

$$p(f(X)) = \mathcal{N}(0, k(X, X)),$$

where *X* is any set of input points and *k(X, X)* is a covariance matrix whose entries are outputs *k(x, z)* of *k* over input pairs (*x, z*). This distribution is usually denoted by

$$f \sim \mathcal{GP}(0, k).$$

---

**Note:** Generally, beside a covariance matrix *k*, a Gaussian Process can also be specified by a mean function *m* (which is a zero-value function by default). In that case, its distribution will be

$$p(f(X)) = \mathcal{N}(m(X), k(X, X)).$$

Gaussian Process models are `Parameterized` subclasses. So its parameters can be learned, set priors, or fixed by using corresponding methods from `Parameterized`. A typical way to define a Gaussian Process model is

```
>>> X = torch.tensor([[1., 5, 3], [4, 3, 7]])
>>> y = torch.tensor([2., 1])
>>> kernel = gp.kernels.RBF(input_dim=3)
>>> kernel.variance = pyro.nn.PyroSample(dist.Uniform(torch.tensor(0.5), torch.
↪tensor(1.5)))
>>> kernel.lengthscale = pyro.nn.PyroSample(dist.Uniform(torch.tensor(1.0), torch.
↪tensor(3.0)))
>>> gpr = gp.models.GPRegression(X, y, kernel)
```

There are two ways to train a Gaussian Process model:

- Using an MCMC algorithm (in module `pyro.infer.mcmc`) on `model()` to get posterior samples for the Gaussian Process's parameters. For example:

```
>>> hmc_kernel = HMC(gpr.model)
>>> mcmc = MCMC(hmc_kernel, num_samples=10)
>>> mcmc.run()
>>> ls_name = "kernel.lengthscale"
>>> posterior_ls = mcmc.get_samples()[ls_name]
```

- Using a variational inference on the pair `model()`, `guide()`:

```
>>> optimizer = torch.optim.Adam(gpr.parameters(), lr=0.01)
>>> loss_fn = pyro.infer.TraceMeanField_ELBO().differentiable_loss
>>>
>>> for i in range(1000):
...     svi.step() # doctest: +SKIP
...     optimizer.zero_grad()
...     loss = loss_fn(gpr.model, gpr.guide) # doctest: +SKIP
...     loss.backward() # doctest: +SKIP
...     optimizer.step()
```

To give a prediction on new dataset, simply use `forward()` like any PyTorch `torch.nn.Module`:

```
>>> Xnew = torch.tensor([[2., 3, 1]])
>>> f_loc, f_cov = gpr(Xnew, full_cov=True)
```

Reference:

[1] *Gaussian Processes for Machine Learning*, Carl E. Rasmussen, Christopher K. I. Williams

#### Parameters

- **X** (`torch.Tensor`) – A input data for training. Its first dimension is the number of data points.
- **y** (`torch.Tensor`) – An output data for training. Its last dimension is the number of data points.
- **kernel** (`Kernel`) – A Pyro kernel object, which is the covariance function  $k$ .
- **mean\_function** (`callable`) – An optional mean function  $m$  of this Gaussian process. By default, we use zero mean.

- **jitter** (*float*) – A small positive term which is added into the diagonal part of a covariance matrix to help stabilize its Cholesky decomposition.

**model** ()

A “model” stochastic function. If `self.y` is `None`, this method returns mean and variance of the Gaussian Process prior.

**guide** ()

A “guide” stochastic function to be used in variational inference methods. It also gives posterior information to the method `forward()` for prediction.

**forward** (*Xnew*, *full\_cov=False*)

Computes the mean and covariance matrix (or variance) of Gaussian Process posterior on a test input data  $X_{new}$ :

$$p(f^* \mid X_{new}, X, y, k, \theta),$$

where  $\theta$  are parameters of this model.

---

**Note:** Model’s parameters  $\theta$  together with kernel’s parameters have been learned from a training procedure (MCMC or SVI).

---

### Parameters

- **Xnew** (*torch.Tensor*) – A input data for testing. Note that `Xnew.shape[1:]` must be the same as `X.shape[1:]`.
- **full\_cov** (*bool*) – A flag to decide if we want to predict full covariance matrix or just variance.

**Returns** loc and covariance matrix (or variance) of  $p(f^*(X_{new}))$

**Return type** `tuple(torch.Tensor, torch.Tensor)`

**set\_data** (*X*, *y=None*)

Sets data for Gaussian Process models.

Some examples to utilize this method are:

- Batch training on a sparse variational model:

```
>>> Xu = torch.tensor([[1., 0, 2]]) # inducing input
>>> likelihood = gp.likelihoods.Gaussian()
>>> vsgp = gp.models.VariationalSparseGP(X, y, kernel, Xu, likelihood)
>>> optimizer = torch.optim.Adam(vsgp.parameters(), lr=0.01)
>>> loss_fn = pyro.infer.TraceMeanField_ELBO().differentiable_loss
>>> batched_X, batched_y = X.split(split_size=10), y.split(split_size=10)
>>> for Xi, yi in zip(batched_X, batched_y):
...     optimizer.zero_grad()
...     vsgp.set_data(Xi, yi)
...     svi.step() # doctest: +SKIP
...     loss = loss_fn(vsgp.model, vsgp.guide) # doctest: +SKIP
...     loss.backward() # doctest: +SKIP
...     optimizer.step()
```

- Making a two-layer Gaussian Process stochastic function:

```

>>> gpr1 = gp.models.GPRegression(X, None, kernel)
>>> Z, _ = gpr1.model()
>>> gpr2 = gp.models.GPRegression(Z, y, kernel)
>>> def two_layer_model():
...     Z, _ = gpr1.model()
...     gpr2.set_data(Z, y)
...     return gpr2.model()

```

References:

[1] *Scalable Variational Gaussian Process Classification*, James Hensman, Alexander G. de G. Matthews, Zoubin Ghahramani

[2] *Deep Gaussian Processes*, Andreas C. Damianou, Neil D. Lawrence

#### Parameters

- **X** (`torch.Tensor`) – A input data for training. Its first dimension is the number of data points.
- **y** (`torch.Tensor`) – An output data for training. Its last dimension is the number of data points.

## 16.1.2 GPRegression

**class** `GPRegression` (*X*, *y*, *kernel*, *noise=None*, *mean\_function=None*, *jitter=1e-06*)

Bases: `pyro.contrib.gp.models.model.GPModel`

Gaussian Process Regression model.

The core of a Gaussian Process is a covariance function  $k$  which governs the similarity between input points. Given  $k$ , we can establish a distribution over functions  $f$  by a multivariate normal distribution

$$p(f(X)) = \mathcal{N}(0, k(X, X)),$$

where  $X$  is any set of input points and  $k(X, X)$  is a covariance matrix whose entries are outputs  $k(x, z)$  of  $k$  over input pairs  $(x, z)$ . This distribution is usually denoted by

$$f \sim \mathcal{GP}(0, k).$$

**Note:** Generally, beside a covariance matrix  $k$ , a Gaussian Process can also be specified by a mean function  $m$  (which is a zero-value function by default). In that case, its distribution will be

$$p(f(X)) = \mathcal{N}(m(X), k(X, X)).$$

Given inputs  $X$  and their noisy observations  $y$ , the Gaussian Process Regression model takes the form

$$\begin{aligned}
 f &\sim \mathcal{GP}(0, k(X, X)), \\
 y &\sim f + \epsilon,
 \end{aligned}$$

where  $\epsilon$  is Gaussian noise.

**Note:** This model has  $\mathcal{O}(N^3)$  complexity for training,  $\mathcal{O}(N^3)$  complexity for testing. Here,  $N$  is the number of train inputs.

Reference:

[1] *Gaussian Processes for Machine Learning*, Carl E. Rasmussen, Christopher K. I. Williams

#### Parameters

- **x** (`torch.Tensor`) – A input data for training. Its first dimension is the number of data points.
- **y** (`torch.Tensor`) – An output data for training. Its last dimension is the number of data points.
- **kernel** (`Kernel`) – A Pyro kernel object, which is the covariance function  $k$ .
- **noise** (`torch.Tensor`) – Variance of Gaussian noise of this model.
- **mean\_function** (`callable`) – An optional mean function  $m$  of this Gaussian process. By default, we use zero mean.
- **jitter** (`float`) – A small positive term which is added into the diagonal part of a covariance matrix to help stabilize its Cholesky decomposition.

**model** ()

**guide** ()

**forward** ( $X_{new}$ , `full_cov=False`, `noiseless=True`)

Computes the mean and covariance matrix (or variance) of Gaussian Process posterior on a test input data  $X_{new}$ :

$$p(f^* | X_{new}, X, y, k, \epsilon) = \mathcal{N}(loc, cov).$$

---

**Note:** The noise parameter `noise` ( $\epsilon$ ) together with `kernel`'s parameters have been learned from a training procedure (MCMC or SVI).

---

#### Parameters

- **Xnew** (`torch.Tensor`) – A input data for testing. Note that `Xnew.shape[1:]` must be the same as `self.X.shape[1:]`.
- **full\_cov** (`bool`) – A flag to decide if we want to predict full covariance matrix or just variance.
- **noiseless** (`bool`) – A flag to decide if we want to include noise in the prediction output or not.

**Returns** `loc` and covariance matrix (or variance) of  $p(f^*(X_{new}))$

**Return type** `tuple(torch.Tensor, torch.Tensor)`

**iter\_sample** (`noiseless=True`)

Iteratively constructs a sample from the Gaussian Process posterior.

Recall that at test input points  $X_{new}$ , the posterior is multivariate Gaussian distributed with mean and covariance matrix given by `forward()`.

This method samples lazily from this multivariate Gaussian. The advantage of this approach is that later query points can depend upon earlier ones. Particularly useful when the querying is to be done by an optimisation routine.

---

**Note:** The noise parameter `noise` ( $\epsilon$ ) together with kernel’s parameters have been learned from a training procedure (MCMC or SVI).

---

**Parameters** `noiseless` (*bool*) – A flag to decide if we want to add sampling noise to the samples beyond the noise inherent in the GP posterior.

**Returns** sampler

**Return type** function

### 16.1.3 SparseGPRegression

**class** `SparseGPRegression` (*X*, *y*, *kernel*, *Xu*, *noise=None*, *mean\_function=None*, *approx=None*, *jitter=1e-06*)

Bases: `pyro.contrib.gp.models.model.GPModel`

Sparse Gaussian Process Regression model.

In `GPRegression` model, when the number of input data  $X$  is large, the covariance matrix  $k(X, X)$  will require a lot of computational steps to compute its inverse (for log likelihood and for prediction). By introducing an additional inducing-input parameter  $X_u$ , we can reduce computational cost by approximate  $k(X, X)$  by a low-rank Nymström approximation  $Q$  (see reference [1]), where

$$Q = k(X, X_u)k(X, X)^{-1}k(X_u, X).$$

Given inputs  $X$ , their noisy observations  $y$ , and the inducing-input parameters  $X_u$ , the model takes the form:

$$\begin{aligned} u &\sim \mathcal{GP}(0, k(X_u, X_u)), \\ f &\sim q(f \mid X, X_u) = \mathbb{E}_{p(u)} q(f \mid X, X_u, u), \\ y &\sim f + \epsilon, \end{aligned}$$

where  $\epsilon$  is Gaussian noise and the conditional distribution  $q(f \mid X, X_u, u)$  is an approximation of

$$p(f \mid X, X_u, u) = \mathcal{N}(m, k(X, X) - Q),$$

whose terms  $m$  and  $k(X, X) - Q$  is derived from the joint multivariate normal distribution:

$$[f, u] \sim \mathcal{GP}(0, k([X, X_u], [X, X_u])).$$

This class implements three approximation methods:

- Deterministic Training Conditional (DTC):

$$q(f \mid X, X_u, u) = \mathcal{N}(m, 0),$$

which in turns will imply

$$f \sim \mathcal{N}(0, Q).$$

- Fully Independent Training Conditional (FITC):

$$q(f \mid X, X_u, u) = \mathcal{N}(m, \text{diag}(k(X, X) - Q)),$$

which in turns will correct the diagonal part of the approximation in DTC:

$$f \sim \mathcal{N}(0, Q + \text{diag}(k(X, X) - Q)).$$

- Variational Free Energy (VFE), which is similar to DTC but has an additional *trace\_term* in the model’s log likelihood. This additional term makes “VFE” equivalent to the variational approach in SparseVariationalGP (see reference [2]).

---

**Note:** This model has  $\mathcal{O}(NM^2)$  complexity for training,  $\mathcal{O}(NM^2)$  complexity for testing. Here,  $N$  is the number of train inputs,  $M$  is the number of inducing inputs.

---

References:

- [1] *A Unifying View of Sparse Approximate Gaussian Process Regression*, Joaquin Quiñonero-Candela, Carl E. Rasmussen
- [2] *Variational learning of inducing variables in sparse Gaussian processes*, Michalis Titsias

#### Parameters

- **X** (*torch.Tensor*) – A input data for training. Its first dimension is the number of data points.
- **y** (*torch.Tensor*) – An output data for training. Its last dimension is the number of data points.
- **kernel** (*Kernel*) – A Pyro kernel object, which is the covariance function  $k$ .
- **Xu** (*torch.Tensor*) – Initial values for inducing points, which are parameters of our model.
- **noise** (*torch.Tensor*) – Variance of Gaussian noise of this model.
- **mean\_function** (*callable*) – An optional mean function  $m$  of this Gaussian process. By default, we use zero mean.
- **approx** (*str*) – One of approximation methods: “DTC”, “FITC”, and “VFE” (default).
- **jitter** (*float*) – A small positive term which is added into the diagonal part of a covariance matrix to help stabilize its Cholesky decomposition.
- **name** (*str*) – Name of this model.

**model** ()

**guide** ()

**forward** ( $X_{\text{new}}$ ,  $\text{full\_cov}=\text{False}$ ,  $\text{noiseless}=\text{True}$ )

Computes the mean and covariance matrix (or variance) of Gaussian Process posterior on a test input data  $X_{\text{new}}$ :

$$p(f^* \mid X_{\text{new}}, X, y, k, X_u, \epsilon) = \mathcal{N}(\text{loc}, \text{cov}).$$



---

**Note:** The noise parameter `noise` ( $\epsilon$ ), the inducing-point parameter  $X_u$ , together with kernel’s parameters have been learned from a training procedure (MCMC or SVI).

---

### Parameters

- **Xnew** (`torch.Tensor`) – A input data for testing. Note that `Xnew.shape[1:]` must be the same as `self.X.shape[1:]`.
- **full\_cov** (`bool`) – A flag to decide if we want to predict full covariance matrix or just variance.
- **noiseless** (`bool`) – A flag to decide if we want to include noise in the prediction output or not.

**Returns** loc and covariance matrix (or variance) of  $p(f^*(X_{new}))$

**Return type** `tuple(torch.Tensor, torch.Tensor)`

## 16.1.4 VariationalGP

**class VariationalGP** (`X`, `y`, `kernel`, `likelihood`, `mean_function=None`, `latent_shape=None`, `whiten=False`, `jitter=1e-06`)

Bases: `pyro.contrib.gp.models.model.GPModel`

Variational Gaussian Process model.

This model deals with both Gaussian and non-Gaussian likelihoods. Given inputs  $X$  and their noisy observations  $y$ , the model takes the form

$$\begin{aligned} f &\sim \mathcal{GP}(0, k(X, X)), \\ y &\sim p(y) = p(y | f)p(f), \end{aligned}$$

where  $p(y | f)$  is the likelihood.

We will use a variational approach in this model by approximating  $q(f)$  to the posterior  $p(f | y)$ . Precisely,  $q(f)$  will be a multivariate normal distribution with two parameters `f_loc` and `f_scale_tril`, which will be learned during a variational inference process.

---

**Note:** This model can be seen as a special version of `SparseVariationalGP` model with  $X_u = X$ .

---



---

**Note:** This model has  $\mathcal{O}(N^3)$  complexity for training,  $\mathcal{O}(N^3)$  complexity for testing. Here,  $N$  is the number of train inputs. Size of variational parameters is  $\mathcal{O}(N^2)$ .

---

### Parameters

- **X** (`torch.Tensor`) – A input data for training. Its first dimension is the number of data points.
- **y** (`torch.Tensor`) – An output data for training. Its last dimension is the number of data points.
- **kernel** (`Kernel`) – A Pyro kernel object, which is the covariance function  $k$ .
- **Likelihood likelihood** (`likelihood`) – A likelihood object.

- **mean\_function** (*callable*) – An optional mean function  $m$  of this Gaussian process. By default, we use zero mean.
- **latent\_shape** (*torch.Size*) – Shape for latent processes (*batch\_shape* of  $q(f)$ ). By default, it equals to output batch shape  $y.shape[-1]$ . For the multi-class classification problems, `latent_shape[-1]` should correspond to the number of classes.
- **whiten** (*bool*) – A flag to tell if variational parameters `f_loc` and `f_scale_tril` are transformed by the inverse of `Lff`, where `Lff` is the lower triangular decomposition of  $kernel(X, X)$ . Enable this flag will help optimization.
- **jitter** (*float*) – A small positive term which is added into the diagonal part of a covariance matrix to help stabilize its Cholesky decomposition.

**model** ()

**guide** ()

**forward** ( $X_{new}$ , *full\_cov=False*)

Computes the mean and covariance matrix (or variance) of Gaussian Process posterior on a test input data  $X_{new}$ :

$$p(f^* | X_{new}, X, y, k, f_{loc}, f_{scale\_tril}) = \mathcal{N}(loc, cov).$$

---

**Note:** Variational parameters `f_loc`, `f_scale_tril`, together with kernel's parameters have been learned from a training procedure (MCMC or SVI).

---

#### Parameters

- **Xnew** (*torch.Tensor*) – A input data for testing. Note that `Xnew.shape[1:]` must be the same as `self.X.shape[1:]`.
- **full\_cov** (*bool*) – A flag to decide if we want to predict full covariance matrix or just variance.

**Returns** loc and covariance matrix (or variance) of  $p(f^*(X_{new}))$

**Return type** `tuple(torch.Tensor, torch.Tensor)`

### 16.1.5 VariationalSparseGP

**class VariationalSparseGP** ( $X, y, kernel, X_u, likelihood, mean\_function=None, latent\_shape=None, num\_data=None, whiten=False, jitter=1e-06$ )

Bases: `pyro.contrib.gp.models.model.GPModel`

Variational Sparse Gaussian Process model.

In *VariationalGP* model, when the number of input data  $X$  is large, the covariance matrix  $k(X, X)$  will require a lot of computational steps to compute its inverse (for log likelihood and for prediction). This model introduces an additional inducing-input parameter  $X_u$  to solve that problem. Given inputs  $X$ , their noisy observations  $y$ , and the inducing-input parameters  $X_u$ , the model takes the form:

$$\begin{aligned} [f, u] &\sim \mathcal{GP}(0, k([X, X_u], [X, X_u])), \\ y &\sim p(y) = p(y | f)p(f), \end{aligned}$$

where  $p(y | f)$  is the likelihood.

We will use a variational approach in this model by approximating  $q(f, u)$  to the posterior  $p(f, u \mid y)$ . Precisely,  $q(f) = p(f \mid u)q(u)$ , where  $q(u)$  is a multivariate normal distribution with two parameters `u_loc` and `u_scale_tril`, which will be learned during a variational inference process.

---

**Note:** This model can be learned using MCMC method as in reference [2]. See also `GPMModel`.

---



---

**Note:** This model has  $\mathcal{O}(NM^2)$  complexity for training,  $\mathcal{O}(M^3)$  complexity for testing. Here,  $N$  is the number of train inputs,  $M$  is the number of inducing inputs. Size of variational parameters is  $\mathcal{O}(M^2)$ .

---

References:

[1] *Scalable variational Gaussian process classification*, James Hensman, Alexander G. de G. Matthews, Zoubin Ghahramani

[2] *MCMC for Variationally Sparse Gaussian Processes*, James Hensman, Alexander G. de G. Matthews, Maurizio Filippone, Zoubin Ghahramani

#### Parameters

- **x** (`torch.Tensor`) – A input data for training. Its first dimension is the number of data points.
- **y** (`torch.Tensor`) – An output data for training. Its last dimension is the number of data points.
- **kernel** (`Kernel`) – A Pyro kernel object, which is the covariance function  $k$ .
- **Xu** (`torch.Tensor`) – Initial values for inducing points, which are parameters of our model.
- **Likelihood likelihood** (`likelihood`) – A likelihood object.
- **mean\_function** (`callable`) – An optional mean function  $m$  of this Gaussian process. By default, we use zero mean.
- **latent\_shape** (`torch.Size`) – Shape for latent processes (`batch_shape` of  $q(u)$ ). By default, it equals to output batch shape `y.shape[-1]`. For the multi-class classification problems, `latent_shape[-1]` should correspond to the number of classes.
- **num\_data** (`int`) – The size of full training dataset. It is useful for training this model with mini-batch.
- **whiten** (`bool`) – A flag to tell if variational parameters `u_loc` and `u_scale_tril` are transformed by the inverse of `Luu`, where `Luu` is the lower triangular decomposition of  $\text{kernel}(X_u, X_u)$ . Enable this flag will help optimization.
- **jitter** (`float`) – A small positive term which is added into the diagonal part of a covariance matrix to help stabilize its Cholesky decomposition.

**model** ()

**guide** ()

**forward** (`Xnew`, `full_cov=False`)

Computes the mean and covariance matrix (or variance) of Gaussian Process posterior on a test input data  $X_{new}$ :

$$p(f^* \mid X_{new}, X, y, k, X_u, u_{loc}, u_{scale\_tril}) = \mathcal{N}(loc, cov).$$

---

**Note:** Variational parameters `u_loc`, `u_scale_tril`, the inducing-point parameter `Xu`, together with kernel's parameters have been learned from a training procedure (MCMC or SVI).

---

### Parameters

- **Xnew** (`torch.Tensor`) – A input data for testing. Note that `Xnew.shape[1:]` must be the same as `self.X.shape[1:]`.
- **full\_cov** (`bool`) – A flag to decide if we want to predict full covariance matrix or just variance.

**Returns** loc and covariance matrix (or variance) of  $p(f^*(X_{new}))$

**Return type** `tuple(torch.Tensor, torch.Tensor)`

## 16.1.6 GPLVM

**class GPLVM** (`base_model`)

Bases: `pyro.contrib.gp.parameterized.Parameterized`

Gaussian Process Latent Variable Model (GPLVM) model.

GPLVM is a Gaussian Process model with its train input data is a latent variable. This model is useful for dimensional reduction of high dimensional data. Assume the mapping from low dimensional latent variable to is a Gaussian Process instance. Then the high dimensional data will play the role of train output `y` and our target is to learn latent inputs which best explain `y`. For the purpose of dimensional reduction, latent inputs should have lower dimensions than `y`.

We follows reference [1] to put a unit Gaussian prior to the input and approximate its posterior by a multivariate normal distribution with two variational parameters: `X_loc` and `X_scale_tril`.

For example, we can do dimensional reduction on Iris dataset as follows:

```
>>> # With y as the 2D Iris data of shape 150x4 and we want to reduce its
↪dimension
>>> # to a tensor X of shape 150x2, we will use GPLVM.
```

```
>>> # First, define the initial values for X parameter:
>>> X_init = torch.zeros(150, 2)
>>> # Then, define a Gaussian Process model with input X_init and output
↪y:
>>> kernel = gp.kernels.RBF(input_dim=2, lengthscale=torch.ones(2))
>>> Xu = torch.zeros(20, 2) # initial inducing inputs of sparse model
>>> gpmodule = gp.models.SparseGPRegression(X_init, y, kernel, Xu)
>>> # Finally, wrap gpmodule by GPLVM, optimize, and get the "learned"
↪mean of X:
>>> gplvm = gp.models.GPLVM(gpmodule)
>>> gp.util.train(gplvm) # doctest: +SKIP
>>> X = gplvm.X
```

Reference:

[1] Bayesian Gaussian Process Latent Variable Model Michalis K. Titsias, Neil D. Lawrence

**Parameters** `base_model` (`GPMModel`) – A Pyro Gaussian Process model object. Note that `base_model.X` will be the initial value for the variational parameter `X_loc`.

**model** ()

`guide()`

`forward(**kwargs)`

Forward method has the same signal as its `base_model`. Note that the train input data of `base_model` is sampled from GPLVM.

## 16.2 Kernels

### 16.2.1 Kernel

**class** `Kernel` (`input_dim`, `active_dims=None`)

Bases: `pyro.contrib.gp.parameterized.Parameterized`

Base class for kernels used in this Gaussian Process module.

Every inherited class should implement a `forward()` pass which takes inputs  $X$ ,  $Z$  and returns their covariance matrix.

To construct a new kernel from the old ones, we can use methods `add()`, `mul()`, `exp()`, `warp()`, `vertical_scale()`.

References:

[1] *Gaussian Processes for Machine Learning*, Carl E. Rasmussen, Christopher K. I. Williams

#### Parameters

- **`input_dim`** (`int`) – Number of feature dimensions of inputs.
- **`variance`** (`torch.Tensor`) – Variance parameter of this kernel.
- **`active_dims`** (`list`) – List of feature dimensions of the input which the kernel acts on.

**`forward`** ( $X$ ,  $Z=None$ , `diag=False`)

Calculates covariance matrix of inputs on active dimensionals.

#### Parameters

- **`X`** (`torch.Tensor`) – A 2D tensor with shape  $N \times input\_dim$ .
- **`Z`** (`torch.Tensor`) – An (optional) 2D tensor with shape  $M \times input\_dim$ .
- **`diag`** (`bool`) – A flag to decide if we want to return full covariance matrix or just its diagonal part.

**Returns** covariance matrix of  $X$  and  $Z$  with shape  $N \times M$

**Return type** `torch.Tensor`

### 16.2.2 Brownian

**class** `Brownian` (`input_dim`, `variance=None`, `active_dims=None`)

Bases: `pyro.contrib.gp.kernels.kernel.Kernel`

This kernel corresponds to a two-sided Brownian motion (Wiener process):

$$k(x, z) = \begin{cases} \sigma^2 \min(|x|, |z|), & \text{if } x \cdot z \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Note that the input dimension of this kernel must be 1.

Reference:

[1] *Theory and Statistical Applications of Stochastic Processes*, Yuliya Mishura, Georgiy Shevchenko

**forward** (*X*, *Z=None*, *diag=False*)

### 16.2.3 Combination

**class** **Combination** (*kern0*, *kern1*)

Bases: `pyro.contrib.gp.kernels.kernel.Kernel`

Base class for kernels derived from a combination of kernels.

#### Parameters

- **kern0** (`Kernel`) – First kernel to combine.
- **kern1** (`Kernel` or `numbers.Number`) – Second kernel to combine.

### 16.2.4 Constant

**class** **Constant** (*input\_dim*, *variance=None*, *active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.kernel.Kernel`

Implementation of Constant kernel:

$$k(x, z) = \sigma^2.$$

**forward** (*X*, *Z=None*, *diag=False*)

### 16.2.5 Coregionalize

**class** **Coregionalize** (*input\_dim*, *rank=None*, *components=None*, *diagonal=None*, *active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.kernel.Kernel`

A kernel for the linear model of coregionalization  $k(x, z) = x^T(WW^T + D)z$  where  $W$  is an `input_dim`-by-`rank` matrix and typically `rank < input_dim`, and  $D$  is a diagonal matrix.

This generalizes the `Linear` kernel to multiple features with a low-rank-plus-diagonal weight matrix. The typical use case is for modeling correlations among outputs of a multi-output GP, where outputs are coded as distinct data points with one-hot coded features denoting which output each datapoint represents.

If only `rank` is specified, the kernel  $(W W^T + D)$  will be randomly initialized to a matrix with expected value the identity matrix.

References:

[1] **Mauricio A. Alvarez, Lorenzo Rosasco, Neil D. Lawrence (2012)** *Kernels for Vector-Valued Functions: a Review*

#### Parameters

- **input\_dim** (`int`) – Number of feature dimensions of inputs.
- **rank** (`int`) – Optional rank. This is only used if `components` is unspecified. If neither `rank` nor `components` is specified, then `rank` defaults to `input_dim`.

- **components** (*torch.Tensor*) – An optional (input\_dim, rank) shaped matrix that maps features to rank-many components. If unspecified, this will be randomly initialized.
- **diagonal** (*torch.Tensor*) – An optional vector of length input\_dim. If unspecified, this will be set to constant 0.5.
- **active\_dims** (*list*) – List of feature dimensions of the input which the kernel acts on.
- **name** (*str*) – Name of the kernel.

**forward** (*X*, *Z=None*, *diag=False*)

### 16.2.6 Cosine

**class Cosine** (*input\_dim*, *variance=None*, *lengthscale=None*, *active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.isotropic.Isotropy`

Implementation of Cosine kernel:

$$k(x, z) = \sigma^2 \cos\left(\frac{|x-z|}{l}\right).$$

**Parameters** **lengthscale** (*torch.Tensor*) – Length-scale parameter of this kernel.

**forward** (*X*, *Z=None*, *diag=False*)

### 16.2.7 DotProduct

**class DotProduct** (*input\_dim*, *variance=None*, *active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.kernel.Kernel`

Base class for kernels which are functions of  $x \cdot z$ .

### 16.2.8 Exponent

**class Exponent** (*kern*)

Bases: `pyro.contrib.gp.kernels.kernel.Transforming`

Creates a new kernel according to

$$k_{new}(x, z) = \exp(k(x, z)).$$

**forward** (*X*, *Z=None*, *diag=False*)

### 16.2.9 Exponential

**class Exponential** (*input\_dim*, *variance=None*, *lengthscale=None*, *active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.isotropic.Isotropy`

Implementation of Exponential kernel:

$$k(x, z) = \sigma^2 \exp\left(-\frac{|x-z|}{l}\right).$$

**forward** (*X*, *Z=None*, *diag=False*)

### 16.2.10 Isotropy

**class Isotropy** (*input\_dim, variance=None, lengthscale=None, active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.kernel.Kernel`

Base class for a family of isotropic covariance kernels which are functions of the distance  $|x - z|/l$ , where  $l$  is the length-scale parameter.

By default, the parameter `lengthscale` has size 1. To use the isotropic version (different lengthscale for each dimension), make sure that `lengthscale` has size equal to `input_dim`.

**Parameters** `lengthscale` (*torch.Tensor*) – Length-scale parameter of this kernel.

### 16.2.11 Linear

**class Linear** (*input\_dim, variance=None, active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.dot_product.DotProduct`

Implementation of Linear kernel:

$$k(x, z) = \sigma^2 x \cdot z.$$

Doing Gaussian Process regression with linear kernel is equivalent to doing a linear regression.

---

**Note:** Here we implement the homogeneous version. To use the inhomogeneous version, consider using `Polynomial` kernel with `degree=1` or making a `Sum` with a `Constant` kernel.

---

**forward** (*X, Z=None, diag=False*)

### 16.2.12 Matern32

**class Matern32** (*input\_dim, variance=None, lengthscale=None, active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.isotropic.Isotropy`

Implementation of Matern32 kernel:

$$k(x, z) = \sigma^2 \left( 1 + \sqrt{3} \times \frac{|x-z|}{l} \right) \exp \left( -\sqrt{3} \times \frac{|x-z|}{l} \right).$$

**forward** (*X, Z=None, diag=False*)

### 16.2.13 Matern52

**class Matern52** (*input\_dim, variance=None, lengthscale=None, active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.isotropic.Isotropy`

Implementation of Matern52 kernel:

$$k(x, z) = \sigma^2 \left( 1 + \sqrt{5} \times \frac{|x-z|}{l} + \frac{5}{3} \times \frac{|x-z|^2}{l^2} \right) \exp \left( -\sqrt{5} \times \frac{|x-z|}{l} \right).$$

**forward** (*X, Z=None, diag=False*)



### 16.2.14 Periodic

**class Periodic** (*input\_dim, variance=None, lengthscale=None, period=None, active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.kernel.Kernel`

Implementation of Periodic kernel:

$$k(x, z) = \sigma^2 \exp \left( -2 \times \frac{\sin^2(\pi(x-z)/p)}{l^2} \right),$$

where  $p$  is the period parameter.

References:

[1] *Introduction to Gaussian processes*, David J.C. MacKay

#### Parameters

- **lengthscale** (`torch.Tensor`) – Length scale parameter of this kernel.
- **period** (`torch.Tensor`) – Period parameter of this kernel.

**forward** ( $X, Z=None, diag=False$ )

### 16.2.15 Polynomial

**class Polynomial** (*input\_dim, variance=None, bias=None, degree=1, active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.dot_product.DotProduct`

Implementation of Polynomial kernel:

$$k(x, z) = \sigma^2 (\text{bias} + x \cdot z)^d.$$

#### Parameters

- **bias** (`torch.Tensor`) – Bias parameter of this kernel. Should be positive.
- **degree** (`int`) – Degree  $d$  of the polynomial.

**forward** ( $X, Z=None, diag=False$ )

### 16.2.16 Product

**class Product** (*kern0, kern1*)

Bases: `pyro.contrib.gp.kernels.kernel.Combination`

Returns a new kernel which acts like a product/tensor product of two kernels. The second kernel can be a constant.

**forward** ( $X, Z=None, diag=False$ )

### 16.2.17 RBF

**class RBF** (*input\_dim, variance=None, lengthscale=None, active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.isotropic.Isotropy`

Implementation of Radial Basis Function kernel:

$$k(x, z) = \sigma^2 \exp \left( -0.5 \times \frac{|x-z|^2}{l^2} \right).$$

---

**Note:** This kernel also has name *Squared Exponential* in literature.

---

**forward** ( $X, Z=None, diag=False$ )

### 16.2.18 RationalQuadratic

**class RationalQuadratic** ( $input\_dim, variance=None, lengthscale=None, scale\_mixture=None, active\_dims=None$ )

Bases: `pyro.contrib.gp.kernels.isotropic.Isotropy`

Implementation of RationalQuadratic kernel:

$$k(x, z) = \sigma^2 \left( 1 + 0.5 \times \frac{|x-z|^2}{\alpha l^2} \right)^{-\alpha}.$$

**Parameters** `scale_mixture` (`torch.Tensor`) – Scale mixture ( $\alpha$ ) parameter of this kernel. Should have size 1.

**forward** ( $X, Z=None, diag=False$ )

### 16.2.19 Sum

**class Sum** ( $kern0, kern1$ )

Bases: `pyro.contrib.gp.kernels.kernel.Combination`

Returns a new kernel which acts like a sum/direct sum of two kernels. The second kernel can be a constant.

**forward** ( $X, Z=None, diag=False$ )

### 16.2.20 Transforming

**class Transforming** ( $kern$ )

Bases: `pyro.contrib.gp.kernels.kernel.Kernel`

Base class for kernels derived from a kernel by some transforms such as warping, exponent, vertical scaling.

**Parameters** `kern` (`Kernel`) – The original kernel.

### 16.2.21 VerticalScaling

**class VerticalScaling** ( $kern, vscaling\_fn$ )

Bases: `pyro.contrib.gp.kernels.kernel.Transforming`

Creates a new kernel according to

$$k_{new}(x, z) = f(x)k(x, z)f(z),$$

where  $f$  is a function.

**Parameters** `vscaling_fn` (`callable`) – A vertical scaling function  $f$ .

**forward** ( $X, Z=None, diag=False$ )

## 16.2.22 Warping

**class Warping** (*kern, iwarping\_fn=None, owarping\_coef=None*)

Bases: `pyro.contrib.gp.kernels.kernel.Transforming`

Creates a new kernel according to

$$k_{new}(x, z) = q(k(f(x), f(z))),$$

where  $f$  is an function and  $q$  is a polynomial with non-negative coefficients `owarping_coef`.

We can take advantage of  $f$  to combine a Gaussian Process kernel with a deep learning architecture. For example:

```
>>> linear = torch.nn.Linear(10, 3)
>>> # register its parameters to Pyro's ParamStore and wrap it by lambda
>>> # to call the primitive pyro.module each time we use the linear function
>>> pyro_linear_fn = lambda x: pyro.module("linear", linear)(x)
>>> kernel = gp.kernels.Matern52(input_dim=3, lengthscale=torch.ones(3))
>>> warped_kernel = gp.kernels.Warping(kernel, pyro_linear_fn)
```

Reference:

[1] *Deep Kernel Learning*, Andrew G. Wilson, Zhiting Hu, Ruslan Salakhutdinov, Eric P. Xing

### Parameters

- **iwarping\_fn** (*callable*) – An input warping function  $f$ .
- **owarping\_coef** (*list*) – A list of coefficients of the output warping polynomial. These coefficients must be non-negative.

**forward** ( $X, Z=None, diag=False$ )

## 16.2.23 WhiteNoise

**class WhiteNoise** (*input\_dim, variance=None, active\_dims=None*)

Bases: `pyro.contrib.gp.kernels.kernel.Kernel`

Implementation of WhiteNoise kernel:

$$k(x, z) = \sigma^2 \delta(x, z),$$

where  $\delta$  is a Dirac delta function.

**forward** ( $X, Z=None, diag=False$ )

## 16.3 Likelihoods

### 16.3.1 Likelihood

**class Likelihood**

Bases: `pyro.contrib.gp.parameterized.Parameterized`

Base class for likelihoods used in Gaussian Process.

Every inherited class should implement a forward pass which takes an input  $f$  and returns a sample  $y$ .

**forward** ( $f_{loc}, f_{var}, y=None$ )

Samples  $y$  given  $f_{loc}, f_{var}$ .

**Parameters**

- **f\_loc** (*torch.Tensor*) – Mean of latent function output.
- **f\_var** (*torch.Tensor*) – Variance of latent function output.
- **y** (*torch.Tensor*) – Training output tensor.

**Returns** a tensor sampled from likelihood

**Return type** *torch.Tensor*

### 16.3.2 Binary

**class Binary** (*response\_function=None*)

Bases: *pyro.contrib.gp.likelihoods.likelihood.Likelihood*

Implementation of Binary likelihood, which is used for binary classification problems.

Binary likelihood uses *Bernoulli* distribution, so the output of *response\_function* should be in range (0, 1). By default, we use *sigmoid* function.

**Parameters** **response\_function** (*callable*) – A mapping to correct domain for Binary likelihood.

**forward** (*f\_loc, f\_var, y=None*)

Samples *y* given *f\_loc*, *f\_var* according to

$$\begin{aligned}f &\sim \mathcal{N}(f_{loc}, f_{var}), \\ y &\sim \mathcal{B}(f).\end{aligned}$$

---

**Note:** The log likelihood is estimated using Monte Carlo with 1 sample of *f*.

---

**Parameters**

- **f\_loc** (*torch.Tensor*) – Mean of latent function output.
- **f\_var** (*torch.Tensor*) – Variance of latent function output.
- **y** (*torch.Tensor*) – Training output tensor.

**Returns** a tensor sampled from likelihood

**Return type** *torch.Tensor*

### 16.3.3 Gaussian

**class Gaussian** (*variance=None*)

Bases: *pyro.contrib.gp.likelihoods.likelihood.Likelihood*

Implementation of Gaussian likelihood, which is used for regression problems.

Gaussian likelihood uses *Normal* distribution.

**Parameters** **variance** (*torch.Tensor*) – A variance parameter, which plays the role of noise in regression problems.

**forward** ( $f\_loc, f\_var, y=None$ )  
 Samples  $y$  given  $f\_loc, f\_var$  according to

$$y \sim \mathcal{N}(f\_loc, f\_var + \epsilon),$$

where  $\epsilon$  is the variance parameter of this likelihood.

#### Parameters

- **f\_loc** (`torch.Tensor`) – Mean of latent function output.
- **f\_var** (`torch.Tensor`) – Variance of latent function output.
- **y** (`torch.Tensor`) – Training output tensor.

**Returns** a tensor sampled from likelihood

**Return type** `torch.Tensor`

### 16.3.4 MultiClass

**class MultiClass** ( $num\_classes, response\_function=None$ )

Bases: `pyro.contrib.gp.likelihoods.likelihood.Likelihood`

Implementation of MultiClass likelihood, which is used for multi-class classification problems.

MultiClass likelihood uses [Categorical](#) distribution, so `response_function` should normalize its input's rightmost axis. By default, we use *softmax* function.

#### Parameters

- **num\_classes** (`int`) – Number of classes for prediction.
- **response\_function** (`callable`) – A mapping to correct domain for MultiClass likelihood.

**forward** ( $f\_loc, f\_var, y=None$ )  
 Samples  $y$  given  $f\_loc, f\_var$  according to

$$\begin{aligned} f &\sim \mathcal{N}(f\_loc, f\_var), \\ y &\sim \mathcal{C}(\text{softmax}(f)). \end{aligned}$$

---

**Note:** The log likelihood is estimated using Monte Carlo with 1 sample of  $f$ .

---

#### Parameters

- **f\_loc** (`torch.Tensor`) – Mean of latent function output.
- **f\_var** (`torch.Tensor`) – Variance of latent function output.
- **y** (`torch.Tensor`) – Training output tensor.

**Returns** a tensor sampled from likelihood

**Return type** `torch.Tensor`

### 16.3.5 Poisson

**class** `Poisson` (*response\_function=None*)

Bases: `pyro.contrib.gp.likelihoods.likelihood.Likelihood`

Implementation of Poisson likelihood, which is used for count data.

Poisson likelihood uses the *Poisson* distribution, so the output of `response_function` should be positive. By default, we use `torch.exp()` as response function, corresponding to a log-Gaussian Cox process.

**Parameters** `response_function` (*callable*) – A mapping to positive real numbers.

**forward** (*f\_loc, f\_var, y=None*)

Samples  $y$  given  $f_{loc}$ ,  $f_{var}$  according to

$$\begin{aligned} f &\sim \mathcal{N}(\mu, \Sigma)(f_{loc}, f_{var}), \\ y &\sim \mathbb{P}(\cdot | f). \end{aligned}$$

---

**Note:** The log likelihood is estimated using Monte Carlo with 1 sample of  $f$ .

---

#### Parameters

- **f\_loc** (`torch.Tensor`) – Mean of latent function output.
- **f\_var** (`torch.Tensor`) – Variance of latent function output.
- **y** (`torch.Tensor`) – Training output tensor.

**Returns** a tensor sampled from likelihood

**Return type** `torch.Tensor`

## 16.4 Parameterized

**class** `Parameterized`

Bases: `pyro.nn.module.PyroModule`

A wrapper of *PyroModule* whose parameters can be set constraints, set priors.

By default, when we set a prior to a parameter, an auto Delta guide will be created. We can use the method `autoguide()` to setup other auto guides.

Example:

```
>>> class Linear(Parameterized):
...     def __init__(self, a, b):
...         super(Linear, self).__init__()
...         self.a = Parameter(a)
...         self.b = Parameter(b)
...
...     def forward(self, x):
...         return self.a * x + self.b
...
>>> linear = Linear(torch.tensor(1.), torch.tensor(0.))
>>> linear.a = PyroParam(torch.tensor(1.), constraints.positive)
>>> linear.b = PyroSample(dist.Normal(0, 1))
```

(continues on next page)

(continued from previous page)

```
>>> linear.autoguide("b", dist.Normal)
>>> assert "a_unconstrained" in dict(linear.named_parameters())
>>> assert "b_loc" in dict(linear.named_parameters())
>>> assert "b_scale_unconstrained" in dict(linear.named_parameters())
```

Note that by default, data of a parameter is a float `torch.Tensor` (unless we use `torch.set_default_tensor_type()` to change default tensor type). To cast these parameters to a correct data type or GPU device, we can call methods such as `double()` or `cuda()`. See `torch.nn.Module` for more information.

**set\_prior** (*name*, *prior*)  
Sets prior for a parameter.

#### Parameters

- **name** (*str*) – Name of the parameter.
- **prior** (*Distribution*) – A Pyro prior distribution.

**autoguide** (*name*, *dist\_constructor*)  
Sets an autoguide for an existing parameter with name *name* (mimic the behavior of module `pyro.infer.autoguide`).

---

**Note:** *dist\_constructor* should be one of `Delta`, `Normal`, and `MultivariateNormal`. More distribution constructor will be supported in the future if needed.

---

#### Parameters

- **name** (*str*) – Name of the parameter.
- **dist\_constructor** – A *Distribution* constructor.

**set\_mode** (*mode*)  
Sets mode of this object to be able to use its parameters in stochastic functions. If *mode*="model", a parameter will get its value from its prior. If *mode*="guide", the value will be drawn from its guide.

---

**Note:** This method automatically sets *mode* for submodules which belong to `Parameterized` class.

---

**Parameters** **mode** (*str*) – Either “model” or “guide”.

**mode**

## 16.5 Util

**conditional** (*Xnew*, *X*, *kernel*, *f\_loc*, *f\_scale\_tril*=None, *Lff*=None, *full\_cov*=False, *whiten*=False, *jitter*=1e-06)

Given  $X_{new}$ , predicts loc and covariance matrix of the conditional multivariate normal distribution

$$p(f^*(X_{new}) \mid X, k, f_{loc}, f_{scale\_tril}).$$

Here *f\_loc* and *f\_scale\_tril* are variation parameters of the variational distribution

$$q(f \mid f_{loc}, f_{scale\_tril}) \sim p(f \mid X, y),$$

where  $f$  is the function value of the Gaussian Process given input  $X$

$$p(f(X)) \sim \mathcal{N}(0, k(X, X))$$

and  $y$  is computed from  $f$  by some likelihood function  $p(y|f)$ .

In case `f_scale_tril=None`, we consider  $f = f_{loc}$  and computes

$$p(f^*(X_{new}) \mid X, k, f).$$

In case `f_scale_tril` is not `None`, we follow the derivation from reference [1]. For the case `f_scale_tril=None`, we follow the popular reference [2].

References:

[1] [Sparse GPs: approximate the posterior, not the model](#)

[2] *Gaussian Processes for Machine Learning*, Carl E. Rasmussen, Christopher K. I. Williams

### Parameters

- **Xnew** (`torch.Tensor`) – A new input data.
- **X** (`torch.Tensor`) – An input data to be conditioned on.
- **kernel** (`Kernel`) – A Pyro kernel object.
- **f\_loc** (`torch.Tensor`) – Mean of  $q(f)$ . In case `f_scale_tril=None`,  $f_{loc} = f$ .
- **f\_scale\_tril** (`torch.Tensor`) – Lower triangular decomposition of covariance matrix of  $q(f)$ 's.
- **Lff** (`torch.Tensor`) – Lower triangular decomposition of  $kernel(X, X)$  (optional).
- **full\_cov** (`bool`) – A flag to decide if we want to return full covariance matrix or just variance.
- **whiten** (`bool`) – A flag to tell if `f_loc` and `f_scale_tril` are already transformed by the inverse of `Lff`.
- **jitter** (`float`) – A small positive term which is added into the diagonal part of a covariance matrix to help stabilize its Cholesky decomposition.

**Returns** loc and covariance matrix (or variance) of  $p(f^*(X_{new}))$

**Return type** `tuple(torch.Tensor, torch.Tensor)`

**train** (`gpmodule`, `optimizer=None`, `loss_fn=None`, `retain_graph=None`, `num_steps=1000`)

A helper to optimize parameters for a GP module.

### Parameters

- **gpmodule** (`GPMoDel`) – A GP module.
- **optimizer** (`Optimizer`) – A PyTorch optimizer instance. By default, we use Adam with `lr=0.01`.
- **loss\_fn** (`callable`) – A loss function which takes inputs are `gpmodule`, `model`, `gpmodule.guide`, and returns ELBO loss. By default, `loss_fn=TraceMeanField_ELBO().differentiable_loss`.
- **retain\_graph** (`bool`) – An optional flag of `torch.autograd.backward`.
- **num\_steps** (`int`) – Number of steps to run SVI.

**Returns** a list of losses during the training procedure

**Return type** `list`



# CHAPTER 17

---

## Mini Pyro

---

This file contains a minimal implementation of the Pyro Probabilistic Programming Language. The API (method signatures, etc.) match that of the full implementation as closely as possible. This file is independent of the rest of Pyro, with the exception of the `pyro.distributions` module.

An accompanying example that makes use of this implementation can be found at `examples/minipyro.py`.

```
class Adam(optim_args)
    Bases: object

    __call__(params)

class JitTrace_ELBO(**kwargs)
    Bases: object

    __call__(model, guide, *args)

class Messenger(fn=None)
    Bases: object

    __call__(*args, **kwargs)

    postprocess_message(msg)

    process_message(msg)

class PlateMessenger(fn, size, dim)
    Bases: pyro.contrib.minipyro.Messenger

    process_message(msg)

class SVI(model, guide, optim, loss)
    Bases: object

    step(*args, **kwargs)

Trace_ELBO(**kwargs)

apply_stack(msg)
```

```
class block (fn=None, hide_fn=<function block.<lambda>>)
    Bases: pyro.contrib.minipyro.Messenger

    process_message (msg)

elbo (model, guide, *args, **kwargs)

get_param_store ()

param (name, init_value=None, constraint=Real(), event_dim=None)

plate (name, size, dim=None)

class replay (fn, guide_trace)
    Bases: pyro.contrib.minipyro.Messenger

    process_message (msg)

sample (name, fn, *args, **kwargs)

class seed (fn=None, rng_seed=None)
    Bases: pyro.contrib.minipyro.Messenger

class trace (fn=None)
    Bases: pyro.contrib.minipyro.Messenger

    get_trace (*args, **kwargs)

    postprocess_message (msg)
```

---

## Optimal Experiment Design

---

Tasks such as choosing the next question to ask in a psychology study, designing an election polling strategy, and deciding which compounds to synthesize and test in biological sciences are all fundamentally asking the same question: how do we design an experiment to maximize the information gathered? Pyro is designed to support automated optimal experiment design: specifying a model and guide is enough to obtain optimal designs for many different kinds of experiment scenarios. Check out our experimental design tutorials that use Pyro to [design an adaptive psychology study]([https://pyro.ai/examples/working\\_memory.html](https://pyro.ai/examples/working_memory.html)) that uses past data to select the next question, and [design an election polling strategy](<https://pyro.ai/examples/elections.html>) that aims to give the strongest prediction about the eventual winner of the election.

Bayesian optimal experimental design (BOED) is a powerful methodology for tackling experimental design problems and is the framework adopted by Pyro. In the BOED framework, we begin with a Bayesian model with a likelihood  $p(y|\theta, d)$  and a prior  $p(\theta)$  on the target latent variables. In Pyro, any fully Bayesian model can be used in the BOED framework. The sample sites corresponding to experimental outcomes are the *observation* sites, those corresponding to latent variables of interest are the *target* sites. The design  $d$  is the argument to the model, and is not a random variable.

In the BOED framework, we choose the design that optimizes the expected information gain (EIG) on the targets  $\theta$  from running the experiment

$$\text{EIG}(d) = \mathbf{E}_{p(y|d)}[H[p(\theta)]H[p(\theta|y, d)]] ,$$

where  $H[\cdot]$  represents the entropy and  $p(\theta|y, d) \propto p(\theta)p(y|\theta, d)$  is the posterior we get from running the experiment with design  $d$  and observing  $y$ . In other words, the optimal design is the one that, in expectation over possible future observations, most reduces posterior entropy over the target latent variables. If the predictive model is correct, this forms a design strategy that is (one-step) optimal from an information-theoretic viewpoint. For further details, see [1, 2].

The `pyro.contrib.oed` module provides tools to create optimal experimental designs for Pyro models. In particular, it provides estimators for the expected information gain (EIG).

To estimate the EIG for a particular design, we first set up our Pyro model. For example:

```
def model(design):
    # This line allows batching of designs, treating all batch dimensions as
    # independent
```

(continues on next page)

(continued from previous page)

```

with pyro.plate_stack("plate_stack", design.shape):

    # We use a Normal prior for theta
    theta = pyro.sample("theta", dist.Normal(torch.tensor(0.0), torch.tensor(1.
→0)))

    # We use a simple logistic regression model for the likelihood
    logit_p = theta - design
    y = pyro.sample("y", dist.Bernoulli(logits=logit_p))

    return y

```

We then select an appropriate EIG estimator, such as:

```

eig = nmc_eig(model, design, observation_labels=["y"], target_labels=["theta"],
→N=2500, M=50)

```

It is possible to estimate the EIG across a grid of designs:

```

designs = torch.stack([design1, design2], dim=0)

```

to find the best design from a number of options.

[1] Chaloner, Kathryn, and Isabella Verdinelli. “Bayesian experimental design: A review.” *Statistical Science* (1995): 273-304.

[2] Foster, Adam, et al. “Variational Bayesian Optimal Experimental Design.” *arXiv preprint arXiv:1903.05480* (2019).

## 18.1 Expected Information Gain

**laplace\_eig**(*model*, *design*, *observation\_labels*, *target\_labels*, *guide*, *loss*, *optim*, *num\_steps*, *final\_num\_samples*, *y\_dist=None*, *eig=True*, *\*\*prior\_entropy\_kwargs*)  
 Estimates the expected information gain (EIG) by making repeated Laplace approximations to the posterior.

### Parameters

- **model** (*function*) – Pyro stochastic function taking *design* as only argument.
- **design** (*torch.Tensor*) – Tensor of possible designs.
- **observation\_labels** (*list*) – labels of sample sites to be regarded as observables.
- **target\_labels** (*list*) – labels of sample sites to be regarded as latent variables of interest, i.e. the sites that we wish to gain information about.
- **guide** (*function*) – Pyro stochastic function corresponding to *model*.
- **loss** – a Pyro loss such as *pyro.infer.Trace\_ELBO().differentiable\_loss*.
- **optim** – optimizer for the loss
- **num\_steps** (*int*) – Number of gradient steps to take per sampled pseudo-observation.
- **final\_num\_samples** (*int*) – Number of *y* samples (pseudo-observations) to take.
- **y\_dist** – Distribution to sample *y* from- if *None* we use the Bayesian marginal distribution.

- **eig** (*bool*) – Whether to compute the EIG or the average posterior entropy (APE). The EIG is given by  $EIG = \text{prior entropy} - APE$ . If *True*, the prior entropy will be estimated analytically, or by Monte Carlo as appropriate for the *model*. If *False* the APE is returned.
- **prior\_entropy\_kwargs** (*dict*) – parameters for estimating the prior entropy: *num\_prior\_samples* indicating the number of samples for a MC estimate of prior entropy, and *mean\_field* indicating if an analytic form for a mean-field prior should be tried.

**Returns** EIG estimate, optionally includes full optimization history

**Return type** `torch.Tensor`

**vi\_eig** (*model*, *design*, *observation\_labels*, *target\_labels*, *vi\_parameters*, *is\_parameters*, *y\_dist=None*, *eig=True*, *\*\*prior\_entropy\_kwargs*)

Deprecated since version 0.4.1: Use *posterior\_eig* instead.

Estimates the expected information gain (EIG) using variational inference (VI).

The APE is defined as

$$APE(d) = E_{Y \sim p(y|\theta, d)}[H(p(\theta|Y, d))]$$

where  $H[p(x)]$  is the [differential entropy](#). The APE is related to expected information gain (EIG) by the equation

$$EIG(d) = H[p(\theta)] - APE(d)$$

in particular, minimising the APE is equivalent to maximising EIG.

#### Parameters

- **model** (*function*) – A pyro model accepting *design* as only argument.
- **design** (`torch.Tensor`) – Tensor representation of design
- **observation\_labels** (*list*) – A subset of the sample sites present in *model*. These sites are regarded as future observations and other sites are regarded as latent variables over which a posterior is to be inferred.
- **target\_labels** (*list*) – A subset of the sample sites over which the posterior entropy is to be measured.
- **vi\_parameters** (*dict*) – Variational inference parameters which should include: *optim*: an instance of `pyro.Optim`, *guide*: a guide function compatible with *model*, *num\_steps*: the number of VI steps to make, and *loss*: the loss function to use for VI
- **is\_parameters** (*dict*) – Importance sampling parameters for the marginal distribution of *Y*. May include *num\_samples*: the number of samples to draw from the marginal.
- **y\_dist** (`pyro.distributions.Distribution`) – (optional) the distribution assumed for the response variable *Y*
- **eig** (*bool*) – Whether to compute the EIG or the average posterior entropy (APE). The EIG is given by  $EIG = \text{prior entropy} - APE$ . If *True*, the prior entropy will be estimated analytically, or by Monte Carlo as appropriate for the *model*. If *False* the APE is returned.
- **prior\_entropy\_kwargs** (*dict*) – parameters for estimating the prior entropy: *num\_prior\_samples* indicating the number of samples for a MC estimate of prior entropy, and *mean\_field* indicating if an analytic form for a mean-field prior should be tried.

**Returns** EIG estimate, optionally includes full optimization history

**Return type** `torch.Tensor`

**nmc\_eig** (*model*, *design*, *observation\_labels*, *target\_labels=None*, *N=100*, *M=10*, *M\_prime=None*, *independent\_priors=False*)

Nested Monte Carlo estimate of the expected information gain (EIG). The estimate is, when there are not any random effects,

$$\frac{1}{N} \sum_{n=1}^N \log p(y_n | \theta_n, d) - \frac{1}{N} \sum_{n=1}^N \log \left( \frac{1}{M} \sum_{m=1}^M p(y_n | \theta_m, d) \right)$$

where  $\theta_n, y_n \sim p(\theta, y | d)$  and  $\theta_m \sim p(\theta)$ . The estimate in the presence of random effects is

$$\frac{1}{N} \sum_{n=1}^N \log \left( \frac{1}{M'} \sum_{m=1}^{M'} p(y_n | \theta_n, \tilde{\theta}_{nm}, d) \right) - \frac{1}{N} \sum_{n=1}^N \log \left( \frac{1}{M} \sum_{m=1}^M p(y_n | \theta_m, \tilde{\theta}_m, d) \right)$$

where  $\tilde{\theta}$  are the random effects with  $\tilde{\theta}_{nm} \sim p(\tilde{\theta} | \theta = \theta_n)$  and  $\theta_m, \tilde{\theta}_m \sim p(\theta, \tilde{\theta})$ . The latter form is used when  $M\_prime \neq None$ .

#### Parameters

- **model** (*function*) – A pyro model accepting *design* as only argument.
- **design** (*torch.Tensor*) – Tensor representation of design
- **observation\_labels** (*list*) – A subset of the sample sites present in *model*. These sites are regarded as future observations and other sites are regarded as latent variables over which a posterior is to be inferred.
- **target\_labels** (*list*) – A subset of the sample sites over which the posterior entropy is to be measured.
- **N** (*int*) – Number of outer expectation samples.
- **M** (*int*) – Number of inner expectation samples for  $p(y | d)$ .
- **M\_prime** (*int*) – Number of samples for  $p(y | \theta, d)$  if required.
- **independent\_priors** (*bool*) – Only used when *M\_prime* is not *None*. Indicates whether the prior distributions for the target variables and the nuisance variables are independent. In this case, it is not necessary to sample the targets conditional on the nuisance variables.

**Returns** EIG estimate, optionally includes full optimization history

**Return type** *torch.Tensor*

**donsker\_varadhan\_eig** (*model*, *design*, *observation\_labels*, *target\_labels*, *num\_samples*, *num\_steps*, *T*, *optim*, *return\_history=False*, *final\_design=None*, *final\_num\_samples=None*)

Donsker-Varadhan estimate of the expected information gain (EIG).

The Donsker-Varadhan representation of EIG is

$$\sup_T E_{p(y, \theta | d)} [T(y, \theta)] - \log E_{p(y | d) p(\theta)} [\exp(T(\bar{y}, \bar{\theta}))]$$

where  $T$  is any (measurable) function.

This methods optimises the loss function over a pre-specified class of functions  $T$ .

#### Parameters

- **model** (*function*) – A pyro model accepting *design* as only argument.
- **design** (*torch.Tensor*) – Tensor representation of design

- **observation\_labels** (*list*) – A subset of the sample sites present in *model*. These sites are regarded as future observations and other sites are regarded as latent variables over which a posterior is to be inferred.
- **target\_labels** (*list*) – A subset of the sample sites over which the posterior entropy is to be measured.
- **num\_samples** (*int*) – Number of samples per iteration.
- **num\_steps** (*int*) – Number of optimization steps.
- **or torch.nn.Module T** (*function*) – optimisable function *T* for use in the Donsker-Varadhan loss function.
- **optim** (*pyro.optim.Optim*) – Optimiser to use.
- **return\_history** (*bool*) – If *True*, also returns a tensor giving the loss function at each step of the optimization.
- **final\_design** (*torch.Tensor*) – The final design tensor to evaluate at. If *None*, uses *design*.
- **final\_num\_samples** (*int*) – The number of samples to use at the final evaluation, If *None*, uses 'num\_samples'.

**Returns** EIG estimate, optionally includes full optimization history

**Return type** `torch.Tensor` or `tuple`

**posterior\_eig** (*model*, *design*, *observation\_labels*, *target\_labels*, *num\_samples*, *num\_steps*, *guide*, *optim*, *return\_history=False*, *final\_design=None*, *final\_num\_samples=None*, *eig=True*, *prior\_entropy\_kwargs={}*, *\*args*, *\*\*kwargs*)

Posterior estimate of expected information gain (EIG) computed from the average posterior entropy (APE) using  $EIG(d) = H[p(\theta)] - APE(d)$ . See [1] for full details.

The posterior representation of APE is

$$\sup_q E_{p(y, \theta|d)} [\log q(\theta|y, d)]$$

where *q* is any distribution on  $\theta$ .

This method optimises the loss over a given *guide* family representing *q*.

[1] Foster, Adam, et al. “Variational Bayesian Optimal Experimental Design.” arXiv preprint arXiv:1903.05480 (2019).

#### Parameters

- **model** (*function*) – A pyro model accepting *design* as only argument.
- **design** (*torch.Tensor*) – Tensor representation of design
- **observation\_labels** (*list*) – A subset of the sample sites present in *model*. These sites are regarded as future observations and other sites are regarded as latent variables over which a posterior is to be inferred.
- **target\_labels** (*list*) – A subset of the sample sites over which the posterior entropy is to be measured.
- **num\_samples** (*int*) – Number of samples per iteration.
- **num\_steps** (*int*) – Number of optimization steps.
- **guide** (*function*) – guide family for use in the (implicit) posterior estimation. The parameters of *guide* are optimised to maximise the posterior objective.
- **optim** (*pyro.optim.Optim*) – Optimiser to use.

- **return\_history** (*bool*) – If *True*, also returns a tensor giving the loss function at each step of the optimization.
- **final\_design** (*torch.Tensor*) – The final design tensor to evaluate at. If *None*, uses *design*.
- **final\_num\_samples** (*int*) – The number of samples to use at the final evaluation, If *None*, uses *num\_samples*.
- **eig** (*bool*) – Whether to compute the EIG or the average posterior entropy (APE). The EIG is given by  $EIG = \text{prior entropy} - APE$ . If *True*, the prior entropy will be estimated analytically, or by Monte Carlo as appropriate for the *model*. If *False* the APE is returned.
- **prior\_entropy\_kwargs** (*dict*) – parameters for estimating the prior entropy: *num\_prior\_samples* indicating the number of samples for a MC estimate of prior entropy, and *mean\_field* indicating if an analytic form for a mean-field prior should be tried.

**Returns** EIG estimate, optionally includes full optimization history

**Return type** *torch.Tensor* or *tuple*

**marginal\_eig** (*model*, *design*, *observation\_labels*, *target\_labels*, *num\_samples*, *num\_steps*, *guide*, *optim*, *return\_history=False*, *final\_design=None*, *final\_num\_samples=None*)

Estimate EIG by estimating the marginal entropy  $p(y|d)$ . See [1] for full details.

The marginal representation of EIG is

$$\inf_q E_{p(y,\theta|d)} \left[ \log \frac{p(y|\theta,d)}{q(y|d)} \right]$$

where  $q$  is any distribution on  $y$ . A variational family for  $q$  is specified in the *guide*.

**Warning:** This method does **not** estimate the correct quantity in the presence of random effects.

[1] Foster, Adam, et al. “Variational Bayesian Optimal Experimental Design.” arXiv preprint arXiv:1903.05480 (2019).

#### Parameters

- **model** (*function*) – A pyro model accepting *design* as only argument.
- **design** (*torch.Tensor*) – Tensor representation of design
- **observation\_labels** (*list*) – A subset of the sample sites present in *model*. These sites are regarded as future observations and other sites are regarded as latent variables over which a posterior is to be inferred.
- **target\_labels** (*list*) – A subset of the sample sites over which the posterior entropy is to be measured.
- **num\_samples** (*int*) – Number of samples per iteration.
- **num\_steps** (*int*) – Number of optimization steps.
- **guide** (*function*) – guide family for use in the marginal estimation. The parameters of *guide* are optimised to maximise the log-likelihood objective.
- **optim** (*pyro.optim.Optim*) – Optimiser to use.
- **return\_history** (*bool*) – If *True*, also returns a tensor giving the loss function at each step of the optimization.
- **final\_design** (*torch.Tensor*) – The final design tensor to evaluate at. If *None*, uses *design*.



- **final\_num\_samples** (*int*) – The number of samples to use at the final evaluation, If *None*, uses ‘num\_samples’.

**Returns** EIG estimate, optionally includes full optimization history

**Return type** `torch.Tensor` or `tuple`

**lfire\_eig** (*model*, *design*, *observation\_labels*, *target\_labels*, *num\_y\_samples*, *num\_theta\_samples*, *num\_steps*, *classifier*, *optim*, *return\_history=False*, *final\_design=None*, *final\_num\_samples=None*)

Estimates the EIG using the method of Likelihood-Free Inference by Ratio Estimation (LFIRE) as in [1]. LFIRE is run separately for several samples of  $\theta$ .

[1] Kleinegesse, Steven, and Michael Gutmann. “Efficient Bayesian Experimental Design for Implicit Models.” arXiv preprint arXiv:1810.09912 (2018).

#### Parameters

- **model** (*function*) – A pyro model accepting *design* as only argument.
- **design** (`torch.Tensor`) – Tensor representation of design
- **observation\_labels** (*list*) – A subset of the sample sites present in *model*. These sites are regarded as future observations and other sites are regarded as latent variables over which a posterior is to be inferred.
- **target\_labels** (*list*) – A subset of the sample sites over which the posterior entropy is to be measured.
- **num\_y\_samples** (*int*) – Number of samples to take in  $y$  for each  $\theta$ .
- **num\_steps** (*int*) – Number of optimization steps.
- **classifier** (*function*) – a Pytorch or Pyro classifier used to distinguish between samples of  $y$  under  $p(y|d)$  and samples under  $p(y|\theta, d)$  for some  $\theta$ .
- **optim** (`pyro.optim.Optim`) – Optimiser to use.
- **return\_history** (*bool*) – If *True*, also returns a tensor giving the loss function at each step of the optimization.
- **final\_design** (`torch.Tensor`) – The final design tensor to evaluate at. If *None*, uses *design*.
- **final\_num\_samples** (*int*) – The number of samples to use at the final evaluation, If *None*, uses ‘num\_samples’.

**Param** `int num_theta_samples`: Number of initial samples in  $\theta$  to take. The likelihood ratio is estimated by LFIRE for each sample.

**Returns** EIG estimate, optionally includes full optimization history

**Return type** `torch.Tensor` or `tuple`

**vnmc\_eig** (*model*, *design*, *observation\_labels*, *target\_labels*, *num\_samples*, *num\_steps*, *guide*, *optim*, *return\_history=False*, *final\_design=None*, *final\_num\_samples=None*)

Estimates the EIG using Variational Nested Monte Carlo (VNMC). The VNMC estimate [1] is

$$\frac{1}{N} \sum_{n=1}^N \left[ \log p(y_n | \theta_n, d) - \log \left( \frac{1}{M} \sum_{m=1}^M \frac{p(\theta_{mn}) p(y_n | \theta_{mn}, d)}{q(\theta_{mn} | y_n)} \right) \right]$$

where  $q(\theta|y)$  is the learned variational posterior approximation and  $\theta_n, y_n \sim p(\theta, y|d)$ ,  $\theta_{mn} \sim q(\theta|y = y_n)$ .

As  $N \rightarrow \infty$  this is an upper bound on EIG. We minimise this upper bound by stochastic gradient descent.

**Warning:** This method cannot be used in the presence of random effects.

[1] Foster, Adam, et al. “Variational Bayesian Optimal Experimental Design.” arXiv preprint arXiv:1903.05480 (2019).

#### Parameters

- **model** (*function*) – A pyro model accepting *design* as only argument.
- **design** (*torch.Tensor*) – Tensor representation of design
- **observation\_labels** (*list*) – A subset of the sample sites present in *model*. These sites are regarded as future observations and other sites are regarded as latent variables over which a posterior is to be inferred.
- **target\_labels** (*list*) – A subset of the sample sites over which the posterior entropy is to be measured.
- **num\_samples** (*tuple*) – Number of  $(N, M)$  samples per iteration.
- **num\_steps** (*int*) – Number of optimization steps.
- **guide** (*function*) – guide family for use in the posterior estimation. The parameters of *guide* are optimised to minimise the VNMC upper bound.
- **optim** (*pyro.optim.Optim*) – Optimiser to use.
- **return\_history** (*bool*) – If *True*, also returns a tensor giving the loss function at each step of the optimization.
- **final\_design** (*torch.Tensor*) – The final design tensor to evaluate at. If *None*, uses *design*.
- **final\_num\_samples** (*tuple*) – The number of  $(N, M)$  samples to use at the final evaluation, If *None*, uses ‘*num\_samples*’.

**Returns** EIG estimate, optionally includes full optimization history

**Return type** *torch.Tensor* or *tuple*

## 18.2 Generalised Linear Mixed Models

**Warning:** This module will eventually be deprecated in favor of *brmp*

The *pyro.contrib.oed.glmm* module provides models and guides for generalised linear mixed models (GLMM). It also includes the Normal-inverse-gamma family.

To create a classical Bayesian linear model, use:

```
from pyro.contrib.oed.glmm import known_covariance_linear_model

# Note: coef is a p-vector, observation_sd is a scalar
# Here, p=1 (one feature)
model = known_covariance_linear_model(coef_mean=torch.tensor([0.]),
                                     coef_sd=torch.tensor([10.]),
                                     observation_sd=torch.tensor(2.))
```

(continues on next page)

(continued from previous page)

```
# An n x p design tensor
# Here, n=2 (two observations)
design = torch.tensor(torch.tensor([[1.], [-1.])))

model(design)
```

A non-linear link function may be introduced, for instance:

```
from pyro.contrib.oed.glmm import logistic_regression_model

# No observation_sd is needed for logistic models
model = logistic_regression_model(coef_mean=torch.tensor([0.]),
                                  coef_sd=torch.tensor([10.]))
```

Random effects may be incorporated as regular Bayesian regression coefficients. For random effects with a shared covariance matrix, see `pyro.contrib.oed.glmm.lmer_model()`.



See the [GP example](#) for example usage.

## 19.1 Abstract Models

**class** `TimeSeriesModel` (*name*=")

Bases: `pyro.nn.module.PyroModule`

Base class for univariate and multivariate time series models.

**log\_prob** (*targets*)

Log probability function.

**Parameters** **targets** (`torch.Tensor`) – A 2-dimensional tensor of real-valued targets of shape  $(T, \text{obs\_dim})$ , where  $T$  is the length of the time series and  $\text{obs\_dim}$  is the dimension of the real-valued targets at each time step

**Returns** **torch.Tensor** A 0-dimensional log probability for the case of properly multivariate time series models in which the output dimensions are correlated; otherwise returns a 1-dimensional tensor of log probabilities for batched univariate time series models.

**forecast** (*targets*, *dts*)

**Parameters**

- **targets** (`torch.Tensor`) – A 2-dimensional tensor of real-valued targets of shape  $(T, \text{obs\_dim})$ , where  $T$  is the length of the time series and  $\text{obs\_dim}$  is the dimension of the real-valued targets at each time step. These represent the training data that are conditioned on for the purpose of making forecasts.
- **dts** (`torch.Tensor`) – A 1-dimensional tensor of times to forecast into the future, with zero corresponding to the time of the final target `targets[-1]`.

**Returns** **torch.distributions.Distribution** Returns a predictive distribution with batch shape  $(S,)$  and event shape  $(\text{obs\_dim},)$ , where  $S$  is the size of `dts`. That is, the resulting predictive distributions do not encode correlations between distinct times in `dts`.

## 19.2 Gaussian Processes

**class IndependentMaternGP** (*nu=1.5, dt=1.0, obs\_dim=1, length\_scale\_init=None, kernel\_scale\_init=None, obs\_noise\_scale\_init=None*)  
Bases: `pyro.contrib.timeseries.base.TimeSeriesModel`

A time series model in which each output dimension is modeled independently with a univariate Gaussian Process with a Matern kernel. The targets are assumed to be evenly spaced in time. Training and inference are logarithmic in the length of the time series  $T$ .

### Parameters

- **nu** (*float*) – The order of the Matern kernel; one of 0.5, 1.5 or 2.5.
- **dt** (*float*) – The time spacing between neighboring observations of the time series.
- **obs\_dim** (*int*) – The dimension of the targets at each time step.
- **length\_scale\_init** (*torch.Tensor*) – optional initial values for the kernel length scale given as a `obs_dim`-dimensional tensor
- **kernel\_scale\_init** (*torch.Tensor*) – optional initial values for the kernel scale given as a `obs_dim`-dimensional tensor
- **obs\_noise\_scale\_init** (*torch.Tensor*) – optional initial values for the observation noise scale given as a `obs_dim`-dimensional tensor

**log\_prob** (*targets*)

**Parameters** *targets* (*torch.Tensor*) – A 2-dimensional tensor of real-valued targets of shape  $(T, \text{obs\_dim})$ , where  $T$  is the length of the time series and `obs_dim` is the dimension of the real-valued targets at each time step

**Returns** *torch.Tensor* A 1-dimensional tensor of log probabilities of shape  $(\text{obs\_dim},)$

**forecast** (*targets, dts*)

### Parameters

- **targets** (*torch.Tensor*) – A 2-dimensional tensor of real-valued targets of shape  $(T, \text{obs\_dim})$ , where  $T$  is the length of the time series and `obs_dim` is the dimension of the real-valued targets at each time step. These represent the training data that are conditioned on for the purpose of making forecasts.
- **dts** (*torch.Tensor*) – A 1-dimensional tensor of times to forecast into the future, with zero corresponding to the time of the final target `targets[-1]`.

**Returns** *torch.distributions.Normal* Returns a predictive Normal distribution with batch shape  $(S,)$  and event shape  $(\text{obs\_dim},)$ , where  $S$  is the size of `dts`.

**class LinearlyCoupledMaternGP** (*nu=1.5, dt=1.0, obs\_dim=2, num\_gps=1, length\_scale\_init=None, kernel\_scale\_init=None, obs\_noise\_scale\_init=None*)  
Bases: `pyro.contrib.timeseries.base.TimeSeriesModel`

A time series model in which each output dimension is modeled as a linear combination of shared univariate Gaussian Processes with Matern kernels.

In more detail, the generative process is:

$$y_i(t) = \sum_j A_{ij} f_j(t) + \epsilon_i(t)$$

The targets  $y_i$  are assumed to be evenly spaced in time. Training and inference are logarithmic in the length of the time series  $T$ .

**Parameters**

- **nu** (*float*) – The order of the Matern kernel; one of 0.5, 1.5 or 2.5.
- **dt** (*float*) – The time spacing between neighboring observations of the time series.
- **obs\_dim** (*int*) – The dimension of the targets at each time step.
- **num\_gps** (*int*) – The number of independent GPs that are mixed to model the time series. Typical values might be  $\text{gp} \in [\text{obs}/2, \text{obs}]$
- **length\_scale\_init** (*torch.Tensor*) – optional initial values for the kernel length scale given as a num\_gps-dimensional tensor
- **kernel\_scale\_init** (*torch.Tensor*) – optional initial values for the kernel scale given as a num\_gps-dimensional tensor
- **obs\_noise\_scale\_init** (*torch.Tensor*) – optional initial values for the observation noise scale given as a obs\_dim-dimensional tensor

**log\_prob** (*targets*)

**Parameters** **targets** (*torch.Tensor*) – A 2-dimensional tensor of real-valued targets of shape (T, obs\_dim), where T is the length of the time series and obs\_dim is the dimension of the real-valued targets at each time step

**Returns** **torch.Tensor** a (scalar) log probability

**forecast** (*targets, dts*)

**Parameters**

- **targets** (*torch.Tensor*) – A 2-dimensional tensor of real-valued targets of shape (T, obs\_dim), where T is the length of the time series and obs\_dim is the dimension of the real-valued targets at each time step. These represent the training data that are conditioned on for the purpose of making forecasts.
- **dts** (*torch.Tensor*) – A 1-dimensional tensor of times to forecast into the future, with zero corresponding to the time of the final target targets[-1].

**Returns** **torch.distributions.MultivariateNormal** Returns a predictive MultivariateNormal distribution with batch shape (S,) and event shape (obs\_dim,), where S is the size of dts.

```
class DependentMaternGP (nu=1.5, dt=1.0, obs_dim=1, linearly_coupled=False,
                        length_scale_init=None, obs_noise_scale_init=None)
```

Bases: *pyro.contrib.timeseries.base.TimeSeriesModel*

A time series model in which each output dimension is modeled as a univariate Gaussian Process with a Matern kernel. The different output dimensions become correlated because the Gaussian Processes are driven by a correlated Wiener process; see reference [1] for details. If, in addition, *linearly\_coupled* is True, additional correlation is achieved through linear mixing as in *LinearlyCoupledMaternGP*. The targets are assumed to be evenly spaced in time. Training and inference are logarithmic in the length of the time series T.

**Parameters**

- **nu** (*float*) – The order of the Matern kernel; must be 1.5.
- **dt** (*float*) – The time spacing between neighboring observations of the time series.
- **obs\_dim** (*int*) – The dimension of the targets at each time step.
- **linearly\_coupled** (*bool*) – Whether to linearly mix the various gaussian processes in the likelihood. Defaults to False.

- **length\_scale\_init** (*torch.Tensor*) – optional initial values for the kernel length scale given as a `obs_dim`-dimensional tensor
- **obs\_noise\_scale\_init** (*torch.Tensor*) – optional initial values for the observation noise scale given as a `obs_dim`-dimensional tensor

References [1] “Dependent Matern Processes for Multivariate Time Series,” Alexander Vandenberg-Rodes, Babak Shahbaba.

**log\_prob** (*targets*)

**Parameters** **targets** (*torch.Tensor*) – A 2-dimensional tensor of real-valued targets of shape  $(T, \text{obs\_dim})$ , where  $T$  is the length of the time series and `obs_dim` is the dimension of the real-valued `targets` at each time step

**Returns** **torch.Tensor** A (scalar) log probability

**forecast** (*targets, dts*)

**Parameters**

- **targets** (*torch.Tensor*) – A 2-dimensional tensor of real-valued targets of shape  $(T, \text{obs\_dim})$ , where  $T$  is the length of the time series and `obs_dim` is the dimension of the real-valued targets at each time step. These represent the training data that are conditioned on for the purpose of making forecasts.
- **dts** (*torch.Tensor*) – A 1-dimensional tensor of times to forecast into the future, with zero corresponding to the time of the final target `targets[-1]`.

**Returns** **torch.distributions.MultivariateNormal** Returns a predictive MultivariateNormal distribution with batch shape  $(S,)$  and event shape  $(\text{obs\_dim},)$ , where  $S$  is the size of `dts`.

## 19.3 Linear Gaussian State Space Models

```
class GenericLGSSM(obs_dim=1, state_dim=2, obs_noise_scale_init=None, learnable_observation_loc=False)
Bases: pyro.contrib.timeseries.base.TimeSeriesModel
```

A generic Linear Gaussian State Space Model parameterized with arbitrary time invariant transition and observation dynamics. The targets are (implicitly) assumed to be evenly spaced in time. Training and inference are logarithmic in the length of the time series  $T$ .

**Parameters**

- **obs\_dim** (*int*) – The dimension of the targets at each time step.
- **state\_dim** (*int*) – The dimension of latent state at each time step.
- **learnable\_observation\_loc** (*bool*) – whether the mean of the observation model should be learned or not; defaults to False.

**log\_prob** (*targets*)

**Parameters** **targets** (*torch.Tensor*) – A 2-dimensional tensor of real-valued targets of shape  $(T, \text{obs\_dim})$ , where  $T$  is the length of the time series and `obs_dim` is the dimension of the real-valued `targets` at each time step

**Returns** **torch.Tensor** A (scalar) log probability.

**forecast** (*targets, N\_timesteps*)



**Parameters**

- **targets** (*torch.Tensor*) – A 2-dimensional tensor of real-valued targets of shape  $(T, \text{obs\_dim})$ , where  $T$  is the length of the time series and  $\text{obs\_dim}$  is the dimension of the real-valued targets at each time step. These represent the training data that are conditioned on for the purpose of making forecasts.
- **N\_timesteps** (*int*) – The number of timesteps to forecast into the future from the final target `targets[-1]`.

**Returns** `torch.distributions.MultivariateNormal` Returns a predictive MultivariateNormal distribution with batch shape  $(N\_timesteps,)$  and event shape  $(\text{obs\_dim},)$

```
class GenericLGSSMWithGPNoiseModel (obs_dim=1, state_dim=2, nu=1.5,  
                                     obs_noise_scale_init=None, length_scale_init=None, ker-  
                                     nel_scale_init=None, learnable_observation_loc=False)
```

Bases: `pyro.contrib.timeseries.base.TimeSeriesModel`

A generic Linear Gaussian State Space Model parameterized with arbitrary time invariant transition and observation dynamics together with separate Gaussian Process noise models for each output dimension. In more detail, the generative process is:

$$y_i(t) = \sum_j A_{ij} z_j(t) + f_i(t) + \epsilon_i(t)$$

where the latent variables  $\mathbf{z}(t)$  follow generic time invariant Linear Gaussian dynamics and the  $f_i(t)$  are Gaussian Processes with Matern kernels.

The targets are (implicitly) assumed to be evenly spaced in time. In particular a timestep of  $dt = 1.0$  for the continuous-time GP dynamics corresponds to a single discrete step of the  $\mathbf{z}$ -space dynamics. Training and inference are logarithmic in the length of the time series  $T$ .

**Parameters**

- **obs\_dim** (*int*) – The dimension of the targets at each time step.
- **state\_dim** (*int*) – The dimension of the  $\mathbf{z}$  latent state at each time step.
- **nu** (*float*) – The order of the Matern kernel; one of 0.5, 1.5 or 2.5.
- **length\_scale\_init** (*torch.Tensor*) – optional initial values for the kernel length scale given as a  $\text{obs\_dim}$ -dimensional tensor
- **kernel\_scale\_init** (*torch.Tensor*) – optional initial values for the kernel scale given as a  $\text{obs\_dim}$ -dimensional tensor
- **obs\_noise\_scale\_init** (*torch.Tensor*) – optional initial values for the observation noise scale given as a  $\text{obs\_dim}$ -dimensional tensor
- **learnable\_observation\_loc** (*bool*) – whether the mean of the observation model should be learned or not; defaults to False.

**log\_prob** (*targets*)

**Parameters** **targets** (*torch.Tensor*) – A 2-dimensional tensor of real-valued targets of shape  $(T, \text{obs\_dim})$ , where  $T$  is the length of the time series and  $\text{obs\_dim}$  is the dimension of the real-valued targets at each time step

**Returns** `torch.Tensor` A (scalar) log probability.

**forecast** (*targets, N\_timesteps*)

**Parameters**

- **targets** (*torch.Tensor*) – A 2-dimensional tensor of real-valued targets of shape  $(T, \text{obs\_dim})$ , where  $T$  is the length of the time series and  $\text{obs\_dim}$  is the dimension

of the real-valued targets at each time step. These represent the training data that are conditioned on for the purpose of making forecasts.

- **N\_timesteps** (*int*) – The number of timesteps to forecast into the future from the final target `targets[-1]`.

**Returns** `torch.distributions.MultivariateNormal` Returns a predictive `MultivariateNormal` distribution with batch shape `(N_timesteps,)` and event shape `(obs_dim,)`

## 20.1 Data Association

**class** `MarginalAssignment` (*exists\_logits*, *assign\_logits*, *bp\_iters=None*)

Computes marginal data associations between objects and detections.

This assumes that each detection corresponds to zero or one object, and each object corresponds to zero or more detections. Specifically this does not assume detections have been partitioned into frames of mutual exclusion as is common in 2-D assignment problems.

### Parameters

- **exists\_logits** (*torch.Tensor*) – a tensor of shape `[num_objects]` representing per-object factors for existence of each potential object.
- **assign\_logits** (*torch.Tensor*) – a tensor of shape `[num_detections, num_objects]` representing per-edge factors of assignment probability, where each edge denotes that a given detection associates with a single object.
- **bp\_iters** (*int*) – optional number of belief propagation iterations. If unspecified or `None` an expensive exact algorithm will be used.

### Variables

- **num\_detections** (*int*) – the number of detections
- **num\_objects** (*int*) – the number of (potentially existing) objects
- **exists\_dist** (*pyro.distributions.Bernoulli*) – a mean field posterior distribution over object existence.
- **assign\_dist** (*pyro.distributions.Categorical*) – a mean field posterior distribution over the object (or `None`) to which each detection associates. This has `.event_shape == (num_objects + 1,)` where the final element denotes spurious detection, and `.batch_shape == (num_frames, num_detections)`.

```
class MarginalAssignmentSparse (num_objects, num_detections, edges, exists_logits, assign_logits,
                                bp_iters)
```

A cheap sparse version of *MarginalAssignment*.

#### Parameters

- **num\_detections** (*int*) – the number of detections
- **num\_objects** (*int*) – the number of (potentially existing) objects
- **edges** (*torch.LongTensor*) – a  $[2, \text{num\_edges}]$ -shaped tensor of (detection, object) index pairs specifying feasible associations.
- **exists\_logits** (*torch.Tensor*) – a tensor of shape  $[\text{num\_objects}]$  representing per-object factors for existence of each potential object.
- **assign\_logits** (*torch.Tensor*) – a tensor of shape  $[\text{num\_edges}]$  representing per-edge factors of assignment probability, where each edge denotes that a given detection associates with a single object.
- **bp\_iters** (*int*) – optional number of belief propagation iterations. If unspecified or None an expensive exact algorithm will be used.

#### Variables

- **num\_detections** (*int*) – the number of detections
- **num\_objects** (*int*) – the number of (potentially existing) objects
- **exists\_dist** (*pyro.distributions.Bernoulli*) – a mean field posterior distribution over object existence.
- **assign\_dist** (*pyro.distributions.Categorical*) – a mean field posterior distribution over the object (or None) to which each detection associates. This has `.event_shape == (num_objects + 1,)` where the final element denotes spurious detection, and `.batch_shape == (num_frames, num_detections)`.

```
class MarginalAssignmentPersistent (exists_logits, assign_logits, bp_iters=None,
                                    bp_momentum=0.5)
```

This computes marginal distributions of a multi-frame multi-object data association problem with an unknown number of persistent objects.

The inputs are factors in a factor graph (existence probabilities for each potential object and assignment probabilities for each object-detection pair), and the outputs are marginal distributions of posterior existence probability of each potential object and posterior assignment probabilities of each object-detection pair.

This assumes a shared (maximum) number of detections per frame; to handle variable number of detections, simply set corresponding elements of `assign_logits` to `-float('inf')`.

#### Parameters

- **exists\_logits** (*torch.Tensor*) – a tensor of shape  $[\text{num\_objects}]$  representing per-object factors for existence of each potential object.
- **assign\_logits** (*torch.Tensor*) – a tensor of shape  $[\text{num\_frames}, \text{num\_detections}, \text{num\_objects}]$  representing per-edge factors of assignment probability, where each edge denotes that at a given time frame a given detection associates with a single object.
- **bp\_iters** (*int*) – optional number of belief propagation iterations. If unspecified or None an expensive exact algorithm will be used.
- **bp\_momentum** (*float*) – optional momentum to use for belief propagation. Should be in the interval  $[0, 1)$ .

## Variables

- **num\_frames** (*int*) – the number of time frames
- **num\_detections** (*int*) – the (maximum) number of detections per frame
- **num\_objects** (*int*) – the number of (potentially existing) objects
- **exists\_dist** (`pyro.distributions.Bernoulli`) – a mean field posterior distribution over object existence.
- **assign\_dist** (`pyro.distributions.Categorical`) – a mean field posterior distribution over the object (or None) to which each detection associates. This has `.event_shape == (num_objects + 1,)` where the final element denotes spurious detection, and `.batch_shape == (num_frames, num_detections)`.

**compute\_marginals** (*exists\_logits, assign\_logits*)

This implements exact inference of pairwise marginals via enumeration. This is very expensive and is only useful for testing.

See *MarginalAssignment* for args and problem description.

**compute\_marginals\_bp** (*exists\_logits, assign\_logits, bp\_iters*)

This implements approximate inference of pairwise marginals via loopy belief propagation, adapting the approach of [1].

See *MarginalAssignment* for args and problem description.

[1] Jason L. Williams, Roslyn A. Lau (2014) Approximate evaluation of marginal association probabilities with belief propagation <https://arxiv.org/abs/1209.6299>

**compute\_marginals\_sparse\_bp** (*num\_objects, num\_detections, edges, exists\_logits, assign\_logits, bp\_iters*)

This implements approximate inference of pairwise marginals via loopy belief propagation, adapting the approach of [1].

See *MarginalAssignmentSparse* for args and problem description.

[1] Jason L. Williams, Roslyn A. Lau (2014) Approximate evaluation of marginal association probabilities with belief propagation <https://arxiv.org/abs/1209.6299>

**compute\_marginals\_persistent** (*exists\_logits, assign\_logits*)

This implements exact inference of pairwise marginals via enumeration. This is very expensive and is only useful for testing.

See *MarginalAssignmentPersistent* for args and problem description.

**compute\_marginals\_persistent\_bp** (*exists\_logits, assign\_logits, bp\_iters, bp\_momentum=0.5*)

This implements approximate inference of pairwise marginals via loopy belief propagation, adapting the approach of [1], [2].

See *MarginalAssignmentPersistent* for args and problem description.

[1] Jason L. Williams, Roslyn A. Lau (2014) Approximate evaluation of marginal association probabilities with belief propagation <https://arxiv.org/abs/1209.6299>

[2] Ryan Turner, Steven Bottone, Bhargav Avasarala (2014) A Complete Variational Tracker <https://papers.nips.cc/paper/5572-a-complete-variational-tracker.pdf>

## 20.2 Distributions

**class** `EKFDistribution` (*x0*, *P0*, *dynamic\_model*, *measurement\_cov*, *time\_steps=1*, *dt=1.0*, *validate\_args=None*)

Distribution over EKF states. See [EKState](#). Currently only supports *log\_prob*.

### Parameters

- **x0** (*torch.Tensor*) – PV tensor (mean)
- **P0** (*torch.Tensor*) – covariance
- **dynamic\_model** – *DynamicModel* object
- **measurement\_cov** (*torch.Tensor*) – measurement covariance
- **time\_steps** (*int*) – number time step
- **dt** (*torch.Tensor*) – time step

**filter\_states** (*value*)

Returns the ekf states given measurements

**Parameters** **value** (*torch.Tensor*) – measurement means of shape (*time\_steps*, *event\_shape*)

**log\_prob** (*value*)

Returns the joint log probability of the innovations of a tensor of measurements

**Parameters** **value** (*torch.Tensor*) – measurement means of shape (*time\_steps*, *event\_shape*)

## 20.3 Dynamic Models

**class** `DynamicModel` (*dimension*, *dimension\_pv*, *num\_process\_noise\_parameters=None*)

Dynamic model interface.

### Parameters

- **dimension** – native state dimension.
- **dimension\_pv** – PV state dimension.
- **num\_process\_noise\_parameters** – process noise parameter space dimension. This for UKF applications. Can be left as `None` for EKF and most other filters.

**dimension**

Native state dimension access.

**dimension\_pv**

PV state dimension access.

**num\_process\_noise\_parameters**

Process noise parameters space dimension access.

**forward** (*x*, *dt*, *do\_normalization=True*)

Integrate native state *x* over time interval *dt*.

### Parameters

- **x** – current native state. If the *DynamicModel* is non-differentiable, be sure to handle the case of *x* being augmented with process noise parameters.

- **dt** – time interval to integrate over.
- **do\_normalization** – whether to perform normalization on output, e.g., mod'ing angles into an interval.

**Returns** Native state  $x$  integrated  $dt$  into the future.

**geodesic\_difference** ( $x1, x0$ )

Compute and return the geodesic difference between 2 native states. This is a generalization of the Euclidean operation  $x1 - x0$ .

**Parameters**

- **x1** – native state.
- **x0** – native state.

**Returns** Geodesic difference between native states  $x1$  and  $x2$ .

**mean2pv** ( $x$ )

Compute and return PV state from native state. Useful for combining state estimates of different types in IMM (Interacting Multiple Model) filtering.

**Parameters** **x** – native state estimate mean.

**Returns** PV state estimate mean.

**cov2pv** ( $P$ )

Compute and return PV covariance from native covariance. Useful for combining state estimates of different types in IMM (Interacting Multiple Model) filtering.

**Parameters** **P** – native state estimate covariance.

**Returns** PV state estimate covariance.

**process\_noise\_cov** ( $dt=0.0$ )

Compute and return process noise covariance ( $Q$ ).

**Parameters** **dt** – time interval to integrate over.

**Returns** Read-only covariance ( $Q$ ). For a `DifferentiableDynamicModel`, this is the covariance of the native state  $x$  resulting from stochastic integration (for use with EKF). Otherwise, it is the covariance directly of the process noise parameters (for use with UKF).

**process\_noise\_dist** ( $dt=0.0$ )

Return a distribution object of state displacement from the process noise distribution over a time interval.

**Parameters** **dt** – time interval that process noise accumulates over.

**Returns** `MultivariateNormal`.

**class DifferentiableDynamicModel** (*dimension, dimension\_pv, num\_process\_noise\_parameters=None*)

DynamicModel for which state transition Jacobians can be efficiently calculated, usu. analytically or by automatic differentiation.

**jacobian** ( $dt$ )

Compute and return native state transition Jacobian ( $F$ ) over time interval  $dt$ .

**Parameters** **dt** – time interval to integrate over.

**Returns** Read-only Jacobian ( $F$ ) of integration map ( $f$ ).

**class Ncp** (*dimension, sv2*)

NCP (Nearly-Constant Position) dynamic model. May be subclassed, e.g., with CWNV (Continuous White Noise Velocity) or DWNV (Discrete White Noise Velocity).

**Parameters**

- **dimension** – native state dimension.
- **sv2** – variance of velocity. Usually chosen so that the standard deviation is roughly half of the max velocity one would ever expect to observe.

**forward** (*x*, *dt*, *do\_normalization=True*)Integrate native state *x* over time interval *dt*.**Parameters**

- **x** – current native state. If the DynamicModel is non-differentiable, be sure to handle the case of *x* being augmented with process noise parameters.
- **dt** – time interval to integrate over. *do\_normalization*: whether to perform normalization on output, e.g., mod'ing angles into an interval. Has no effect for this subclass.

**Returns** Native state *x* integrated *dt* into the future.**mean2pv** (*x*)

Compute and return PV state from native state. Useful for combining state estimates of different types in IMM (Interacting Multiple Model) filtering.

**Parameters** **x** – native state estimate mean.**Returns** PV state estimate mean.**cov2pv** (*P*)

Compute and return PV covariance from native covariance. Useful for combining state estimates of different types in IMM (Interacting Multiple Model) filtering.

**Parameters** **P** – native state estimate covariance.**Returns** PV state estimate covariance.**jacobian** (*dt*)Compute and return cached native state transition Jacobian (F) over time interval *dt*.**Parameters** **dt** – time interval to integrate over.**Returns** Read-only Jacobian (F) of integration map (f).**process\_noise\_cov** (*dt=0.0*)

Compute and return cached process noise covariance (Q).

**Parameters** **dt** – time interval to integrate over.**Returns** Read-only covariance (Q) of the native state *x* resulting from stochastic integration (for use with EKF).**class Ncv** (*dimension*, *sa2*)

NCV (Nearly-Constant Velocity) dynamic model. May be subclassed, e.g., with CWNA (Continuous White Noise Acceleration) or DWNA (Discrete White Noise Acceleration).

**Parameters**

- **dimension** – native state dimension.
- **sa2** – variance of acceleration. Usually chosen so that the standard deviation is roughly half of the max acceleration one would ever expect to observe.

**forward** (*x*, *dt*, *do\_normalization=True*)Integrate native state *x* over time interval *dt*.**Parameters**



- **x** – current native state. If the DynamicModel is non-differentiable, be sure to handle the case of **x** being augmented with process noise parameters.
- **dt** – time interval to integrate over.
- **do\_normalization** – whether to perform normalization on output, e.g., mod'ing angles into an interval. Has no effect for this subclass.

**Returns** Native state **x** integrated **dt** into the future.

**mean2pv** (*x*)

Compute and return PV state from native state. Useful for combining state estimates of different types in IMM (Interacting Multiple Model) filtering.

**Parameters** **x** – native state estimate mean.

**Returns** PV state estimate mean.

**cov2pv** (*P*)

Compute and return PV covariance from native covariance. Useful for combining state estimates of different types in IMM (Interacting Multiple Model) filtering.

**Parameters** **P** – native state estimate covariance.

**Returns** PV state estimate covariance.

**jacobian** (*dt*)

Compute and return cached native state transition Jacobian (**F**) over time interval **dt**.

**Parameters** **dt** – time interval to integrate over.

**Returns** Read-only Jacobian (**F**) of integration map (**f**).

**process\_noise\_cov** (*dt=0.0*)

Compute and return cached process noise covariance (**Q**).

**Parameters** **dt** – time interval to integrate over.

**Returns** Read-only covariance (**Q**) of the native state **x** resulting from stochastic integration (for use with EKF).

**class NcpContinuous** (*dimension, sv2*)

NCP (Nearly-Constant Position) dynamic model with CWNV (Continuous White Noise Velocity).

**References:** “Estimation with Applications to Tracking and Navigation” by Y. Bar- Shalom et al, 2001, p.269.

**Parameters**

- **dimension** – native state dimension.
- **sv2** – variance of velocity. Usually chosen so that the standard deviation is roughly half of the max velocity one would ever expect to observe.

**process\_noise\_cov** (*dt=0.0*)

Compute and return cached process noise covariance (**Q**).

**Parameters** **dt** – time interval to integrate over.

**Returns** Read-only covariance (**Q**) of the native state **x** resulting from stochastic integration (for use with EKF).

**class NcvContinuous** (*dimension, sa2*)

NCV (Nearly-Constant Velocity) dynamic model with CWNA (Continuous White Noise Acceleration).

**References:** “Estimation with Applications to Tracking and Navigation” by Y. Bar- Shalom et al, 2001, p.269.

**Parameters**

- **dimension** – native state dimension.
- **sa2** – variance of acceleration. Usually chosen so that the standard deviation is roughly half of the max acceleration one would ever expect to observe.

**process\_noise\_cov** (*dt=0.0*)

Compute and return cached process noise covariance (Q).

**Parameters** **dt** – time interval to integrate over.

**Returns** Read-only covariance (Q) of the native state  $x$  resulting from stochastic integration (for use with EKF).

**class NcpDiscrete** (*dimension, sv2*)

NCP (Nearly-Constant Position) dynamic model with DWNV (Discrete White Noise Velocity).

**Parameters**

- **dimension** – native state dimension.
- **sv2** – variance of velocity. Usually chosen so that the standard deviation is roughly half of the max velocity one would ever expect to observe.

**References:** “Estimation with Applications to Tracking and Navigation” by Y. Bar- Shalom et al, 2001, p.273.

**process\_noise\_cov** (*dt=0.0*)

Compute and return cached process noise covariance (Q).

**Parameters** **dt** – time interval to integrate over.

**Returns** Read-only covariance (Q) of the native state  $x$  resulting from stochastic integration (for use with EKF).

**class NcvDiscrete** (*dimension, sa2*)

NCV (Nearly-Constant Velocity) dynamic model with DWNA (Discrete White Noise Acceleration).

**Parameters**

- **dimension** – native state dimension.
- **sa2** – variance of acceleration. Usually chosen so that the standard deviation is roughly half of the max acceleration one would ever expect to observe.

**References:** “Estimation with Applications to Tracking and Navigation” by Y. Bar- Shalom et al, 2001, p.273.

**process\_noise\_cov** (*dt=0.0*)

Compute and return cached process noise covariance (Q).

**Parameters** **dt** – time interval to integrate over.

**Returns** Read-only covariance (Q) of the native state  $x$  resulting from stochastic integration (for use with EKF). (Note that this Q, modulo numerical error, has rank  $dimension/2$ . So, it is only positive semi-definite.)

## 20.4 Extended Kalman Filter

**class EKFState** (*dynamic\_model, mean, cov, time=None, frame\_num=None*)

State-Centric EKF (Extended Kalman Filter) for use with either an NCP (Nearly-Constant Position) or NCV

(Nearly-Constant Velocity) target dynamic model. Stores a target dynamic model, state estimate, and state time. Incoming `Measurement` provide sensor information for updates.

**Warning:** For efficiency, the dynamic model is only shallow-copied. Make a deep copy outside as necessary to protect against unexpected changes.

### Parameters

- **dynamic\_model** – target dynamic model.
- **mean** – mean of target state estimate.
- **cov** – covariance of target state estimate.
- **time** – time of state estimate.

### **dynamic\_model**

Dynamic model access.

### **dimension**

Native state dimension access.

### **mean**

Native state estimate mean access.

### **cov**

Native state estimate covariance access.

### **dimension\_pv**

PV state dimension access.

### **mean\_pv**

Compute and return cached PV state estimate mean.

### **cov\_pv**

Compute and return cached PV state estimate covariance.

### **time**

Continuous State time access.

### **frame\_num**

Discrete State time access.

### **predict** (*dt=None, destination\_time=None, destination\_frame\_num=None*)

Use dynamic model to predict (aka propagate aka integrate) state estimate in-place.

### Parameters

- **dt** – time to integrate over. The state time will be automatically incremented this amount unless you provide `destination_time`. Using `destination_time` may be preferable for prevention of roundoff error accumulation.
- **destination\_time** – optional value to set continuous state time to after integration. If this is not provided, then `destination_frame_num` must be.
- **destination\_frame\_num** – optional value to set discrete state time to after integration. If this is not provided, then `destination_frame_num` must be.

### **innovation** (*measurement*)

Compute and return the innovation that a measurement would induce if it were used for an update, but

don't actually perform the update. Assumes state and measurement are time-aligned. Useful for computing  $\text{Chi}^2$  stats and likelihoods.

**Parameters** `measurement` – measurement

**Returns** Innovation mean and covariance of hypothetical update.

**Return type** `tuple(torch.Tensor, torch.Tensor)`

**`log_likelihood_of_update`** (*measurement*)

Compute and return the likelihood of a potential update, but don't actually perform the update. Assumes state and measurement are time-aligned. Useful for gating and calculating costs in assignment problems for data association.

**Param** `measurement`.

**Returns** Likelihood of hypothetical update.

**`update`** (*measurement*)

Use measurement to update state estimate in-place and return innovation. The innovation is useful, e.g., for evaluating filter consistency or updating model likelihoods when the `EKFState` is part of an `IMMState`.

**Param** `measurement`.

**Returns** EKF State, Innovation mean and covariance.

## 20.5 Hashing

**class** `LSH` (*radius*)

Implements locality-sensitive hashing for low-dimensional euclidean space.

Allows to efficiently find neighbours of a point. Provides 2 guarantees:

- Difference between coordinates of points not returned by `nearby()` and input point is larger than `radius`.
- Difference between coordinates of points returned by `nearby()` and input point is smaller than `2 * radius`.

Example:

```
>>> radius = 1
>>> lsh = LSH(radius)
>>> a = torch.tensor([-0.51, -0.51]) # hash(a)=(-1,-1)
>>> b = torch.tensor([-0.49, -0.49]) # hash(a)=(0,0)
>>> c = torch.tensor([1.0, 1.0]) # hash(b)=(1,1)
>>> lsh.add('a', a)
>>> lsh.add('b', b)
>>> lsh.add('c', c)
>>> # even though c is within 2radius of a
>>> lsh.nearby('a') # doctest: +SKIP
{'b'}
>>> lsh.nearby('b') # doctest: +SKIP
{'a', 'c'}
>>> lsh.remove('b')
>>> lsh.nearby('a') # doctest: +SKIP
set()
```

**Parameters** **radius** (*float*) – Scaling parameter used in hash function. Determines the size of the neighbourhood.

**add** (*key*, *point*)

Adds (*key*, *point*) pair to the hash.

**Parameters**

- **key** – Key used identify point.
- **point** (*torch.Tensor*) – data, should be detached and on cpu.

**remove** (*key*)

Removes *key* and corresponding point from the hash.

Raises `KeyError` if *key* is not in hash.

**Parameters** **key** – key used to identify point.

**nearby** (*key*)

Returns a set of keys which are neighbours of the point identified by *key*.

Two points are nearby if difference of each element of their hashes is smaller than 2. In euclidean space, this corresponds to all points  $\mathbf{p}$  where  $|\mathbf{p}_k - (\mathbf{p}_{\text{key}})_k| < r$ , and some points (all points not guaranteed) where  $|\mathbf{p}_k - (\mathbf{p}_{\text{key}})_k| < 2r$ .

**Parameters** **key** – key used to identify input point.

**Returns** a set of keys identifying neighbours of the input point.

**Return type** `set`

**class** **ApproxSet** (*radius*)

Queries low-dimensional euclidean space for approximate occupancy.

**Parameters** **radius** (*float*) – scaling parameter used in hash function. Determines the size of the bin. See [LSH](#) for details.

**try\_add** (*point*)

Attempts to add *point* to set. Only adds there are no points in the *point*'s bin.

**Parameters** **point** (*torch.Tensor*) – Point to be queried, should be detached and on cpu.

**Returns** `True` if point is successfully added, `False` if there is already a point in *point*'s bin.

**Return type** `bool`

**merge\_points** (*points*, *radius*)

Greedy merge points that are closer than given radius.

This uses [LSH](#) to achieve complexity that is linear in the number of merged clusters and quadratic in the size of the largest merged cluster.

**Parameters**

- **points** (*torch.Tensor*) – A tensor of shape  $(K, D)$  where  $K$  is the number of points and  $D$  is the number of dimensions.
- **radius** (*float*) – The minimum distance nearer than which points will be merged.

**Returns** A tuple (`merged_points`, `groups`) where `merged_points` is a tensor of shape  $(J, D)$  where  $J \leq K$ , and `groups` is a list of tuples of indices mapping merged points to original points. Note that `len(groups) == J` and `sum(len(group) for group in groups) == K`.

**Return type** `tuple`

## 20.6 Measurements

**class Measurement** (*mean, cov, time=None, frame\_num=None*)

Gaussian measurement interface.

**Parameters**

- **mean** – mean of measurement distribution.
- **cov** – covariance of measurement distribution.
- **time** – continuous time of measurement. If this is not provided, *frame\_num* must be.
- **frame\_num** – discrete time of measurement. If this is not provided, *time* must be.

**dimension**

Measurement space dimension access.

**mean**

Measurement mean ( $z$  in most Kalman Filtering literature).

**cov**

Noise covariance ( $R$  in most Kalman Filtering literature).

**time**

Continuous time of measurement.

**frame\_num**

Discrete time of measurement.

**geodesic\_difference** (*z1, z0*)

Compute and return the geodesic difference between 2 measurements. This is a generalization of the Euclidean operation  $z1 - z0$ .

**Parameters**

- **z1** – measurement.
- **z0** – measurement.

**Returns** Geodesic difference between  $z1$  and  $z2$ .

**class DifferentiableMeasurement** (*mean, cov, time=None, frame\_num=None*)

Interface for Gaussian measurement for which Jacobians can be efficiently calculated, usu. analytically or by automatic differentiation.

**jacobian** (*x=None*)

Compute and return Jacobian ( $H$ ) of measurement map ( $h$ ) at target PV state  $x$ .

**Parameters** **x** – PV state. Use default argument `None` when the Jacobian is not state-dependent.

**Returns** Read-only Jacobian ( $H$ ) of measurement map ( $h$ ).

**class PositionMeasurement** (*mean, cov, time=None, frame\_num=None*)

Full-rank Gaussian position measurement in Euclidean space.

**Parameters**

- **mean** – mean of measurement distribution.
- **cov** – covariance of measurement distribution.
- **time** – time of measurement.

**jacobian** (*x=None*)

Compute and return Jacobian ( $H$ ) of measurement map ( $h$ ) at target PV state  $x$ .

**Parameters**  $\mathbf{x}$  – PV state. The default argument `None` may be used in this subclass since the Jacobian is not state-dependent.

**Returns** Read-only Jacobian (H) of measurement map (h).





## CHAPTER 21

---

### Indices and tables

---

- `genindex`
- `search`



### p

`pyro.contrib.autoname`, 141  
`pyro.contrib.autoname.named`, 143  
`pyro.contrib.autoname.scoping`, 145  
`pyro.contrib.bnn`, 149  
`pyro.contrib.bnn.hidden_layer`, 149  
`pyro.contrib.cevae`, 151  
`pyro.contrib.easyguide`, 157  
`pyro.contrib.gp`, 163  
`pyro.contrib.gp.kernels`, 175  
`pyro.contrib.gp.likelihoods`, 181  
`pyro.contrib.gp.models.gplvm`, 174  
`pyro.contrib.gp.models.gpr`, 167  
`pyro.contrib.gp.models.model`, 164  
`pyro.contrib.gp.models.sgpr`, 169  
`pyro.contrib.gp.models.vgp`, 171  
`pyro.contrib.gp.models.vsgp`, 172  
`pyro.contrib.gp.parameterized`, 184  
`pyro.contrib.gp.util`, 185  
`pyro.contrib.minipyro`, 186  
`pyro.contrib.oed`, 189  
`pyro.contrib.oed.eig`, 190  
`pyro.contrib.oed.glmm`, 196  
`pyro.contrib.timeseries.base`, 199  
`pyro.contrib.timeseries.gp`, 200  
`pyro.contrib.timeseries.lgssm`, 202  
`pyro.contrib.timeseries.lgssmgp`, 203  
`pyro.contrib.tracking`, 205  
`pyro.contrib.tracking.assignment`, 205  
`pyro.contrib.tracking.distributions`, 208  
`pyro.contrib.tracking.dynamic_models`, 208  
`pyro.contrib.tracking.extended_kalman_filter`, 212  
`pyro.contrib.tracking.hashing`, 214  
`pyro.contrib.tracking.measurements`, 216  
`pyro.distributions.torch`, 43  
`pyro.infer.abstract_infer`, 27  
`pyro.infer.autoguide`, 35  
`pyro.infer.autoguide.initialization`, 42  
`pyro.infer.discrete`, 25  
`pyro.infer.elbo`, 12  
`pyro.infer.energy_distance`, 24  
`pyro.infer.importance`, 19  
`pyro.infer.predictive`, 26  
`pyro.infer.renyi_elbo`, 17  
`pyro.infer.rws`, 20  
`pyro.infer.smcfilter`, 22  
`pyro.infer.svgd`, 22  
`pyro.infer.svi`, 11  
`pyro.infer.trace_elbo`, 13  
`pyro.infer.trace_mean_field_elbo`, 16  
`pyro.infer.trace_tail_adaptive_elbo`, 17  
`pyro.infer.traceenum_elbo`, 14  
`pyro.infer.tracegraph_elbo`, 14  
`pyro.infer.tracetmc_elbo`, 18  
`pyro.nn.module`, 93  
`pyro.ops.dual_averaging`, 127  
`pyro.ops.einsum`, 132  
`pyro.ops.gaussian`, 134  
`pyro.ops.indexing`, 131  
`pyro.ops.integrator`, 128  
`pyro.ops.newton`, 129  
`pyro.ops.ssm_gp`, 139  
`pyro.ops.stats`, 136  
`pyro.ops.welford`, 128  
`pyro.optim.adagrad_rmsprop`, 100  
`pyro.optim.clipped_adam`, 101  
`pyro.optim.lr_scheduler`, 100  
`pyro.optim.multi`, 102  
`pyro.optim.optim`, 99  
`pyro.optim.pytorch_optimizers`, 101  
`pyro.params.param_store`, 89  
`pyro.poutine.block_messenger`, 118  
`pyro.poutine.broadcast_messenger`, 119  
`pyro.poutine.condition_messenger`, 119  
`pyro.poutine.do_messenger`, 120  
`pyro.poutine.enum_messenger`, 120  
`pyro.poutine.escape_messenger`, 121

`pyro.poutine.handlers`, [105](#)  
`pyro.poutine.indep_messenger`, [121](#)  
`pyro.poutine.infer_config_messenger`, [121](#)  
`pyro.poutine.lift_messenger`, [122](#)  
`pyro.poutine.markov_messenger`, [122](#)  
`pyro.poutine.mask_messenger`, [123](#)  
`pyro.poutine.messenger`, [117](#)  
`pyro.poutine.plate_messenger`, [123](#)  
`pyro.poutine.reentrant_messenger`, [123](#)  
`pyro.poutine.replay_messenger`, [123](#)  
`pyro.poutine.runtime`, [115](#)  
`pyro.poutine.scale_messenger`, [124](#)  
`pyro.poutine.seed_messenger`, [124](#)  
`pyro.poutine.subsample_messenger`, [125](#)  
`pyro.poutine.trace_messenger`, [125](#)  
`pyro.poutine.uncondition_messenger`, [126](#)  
`pyro.poutine.util`, [116](#)

## Symbols

[\\_\\_add\\_\\_\(\) \(Gaussian method\)](#), 134  
[\\_\\_call\\_\\_\(\) \(Adam method\)](#), 187  
[\\_\\_call\\_\\_\(\) \(Distribution method\)](#), 47  
[\\_\\_call\\_\\_\(\) \(EnergyDistance method\)](#), 25  
[\\_\\_call\\_\\_\(\) \(JitTrace\\_ELBO method\)](#), 187  
[\\_\\_call\\_\\_\(\) \(Messenger method\)](#), 187  
[\\_\\_call\\_\\_\(\) \(PyroLRScheduler method\)](#), 100  
[\\_\\_call\\_\\_\(\) \(PyroOptim method\)](#), 99  
[\\_\\_call\\_\\_\(\) \(TorchDistributionMixin method\)](#), 48  
[\\_\\_getitem\\_\\_\(\) \(Gaussian method\)](#), 134

## A

[Adadelta\(\)](#) (in module [pyro.optim.pytorch\\_optimizers](#)), 101  
[Adagrad\(\)](#) (in module [pyro.optim.pytorch\\_optimizers](#)), 101  
[AdagradRMSProp](#) (class in [pyro.optim.adagrad\\_rmsprop](#)), 100  
[AdagradRMSProp\(\)](#) (in module [pyro.optim.optim](#)), 100  
[Adam](#) (class in [pyro.contrib.minipyro](#)), 187  
[Adam\(\)](#) (in module [pyro.optim.pytorch\\_optimizers](#)), 101  
[Adamax\(\)](#) (in module [pyro.optim.pytorch\\_optimizers](#)), 101  
[AdamW\(\)](#) (in module [pyro.optim.pytorch\\_optimizers](#)), 101  
[add\(\)](#) ([AutoGuideList](#) method), 36  
[add\(\)](#) ([List](#) method), 145  
[add\(\)](#) ([LSH](#) method), 215  
[add\\_edge\(\)](#) ([Trace](#) method), 114  
[add\\_module\(\)](#) ([PyroModule](#) method), 95  
[add\\_node\(\)](#) ([Trace](#) method), 114  
[affine\\_autoregressive\(\)](#) (in module [pyro.distributions.transforms](#)), 84  
[affine\\_coupling\(\)](#) (in module [pyro.distributions.transforms](#)), 84  
[AffineAutoregressive](#) (class in [pyro.distributions.transforms](#)), 72  
[AffineCoupling](#) (class in [pyro.distributions.transforms](#)), 74  
[AffineNormal](#) (class in [pyro.ops.gaussian](#)), 135  
[all\\_escape\(\)](#) (in module [pyro.poutine.util](#)), 116  
[am\\_i\\_wrapped\(\)](#) (in module [pyro.poutine.runtime](#)), 115  
[append\(\)](#) ([AutoGuideList](#) method), 36  
[append\(\)](#) ([FullyConnected](#) method), 154  
[apply\\_stack\(\)](#) (in module [pyro.contrib.minipyro](#)), 187  
[apply\\_stack\(\)](#) (in module [pyro.poutine.runtime](#)), 115  
[ApproxSet](#) (class in [pyro.contrib.tracking.hashing](#)), 215  
[arg\\_constraints](#) ([AVFMultivariateNormal](#) attribute), 52  
[arg\\_constraints](#) ([BetaBinomial](#) attribute), 52  
[arg\\_constraints](#) ([Delta](#) attribute), 53  
[arg\\_constraints](#) ([DirichletMultinomial](#) attribute), 54  
[arg\\_constraints](#) ([DiscreteHMM](#) attribute), 55  
[arg\\_constraints](#) ([Empirical](#) attribute), 55  
[arg\\_constraints](#) ([GammaPoisson](#) attribute), 57  
[arg\\_constraints](#) ([GaussianHMM](#) attribute), 58  
[arg\\_constraints](#) ([GaussianMRF](#) attribute), 59  
[arg\\_constraints](#) ([GaussianScaleMixture](#) attribute), 60  
[arg\\_constraints](#) ([InverseGamma](#) attribute), 60  
[arg\\_constraints](#) ([LKJCorrCholesky](#) attribute), 61  
[arg\\_constraints](#) ([MaskedMixture](#) attribute), 61  
[arg\\_constraints](#) ([MixtureOfDiagNormals](#) attribute), 62  
[arg\\_constraints](#) ([MixtureOfDiagNormalsSharedCovariance](#) attribute), 63  
[arg\\_constraints](#) ([MultivariateStudentT](#) attribute), 63  
[arg\\_constraints](#) ([OMTMultivariateNormal](#) attribute), 64  
[arg\\_constraints](#) ([SpanningTree](#) attribute), 66  
[arg\\_constraints](#) ([Stable](#) attribute), 67

- arg\_constraints (*Unit attribute*), 68  
 arg\_constraints (*VonMises attribute*), 68  
 arg\_constraints (*VonMises3D attribute*), 69  
 arg\_constraints (*ZeroInflatedDistribution attribute*), 70  
 ASGD () (*in module pyro.optim.pytorch\_optimizers*), 102  
 atanh () (*TanhTransform static method*), 72  
 AutoCallable (*class in pyro.infer.autoguide*), 36  
 AutoContinuous (*class in pyro.infer.autoguide*), 38  
 autocorrelation () (*in module pyro.ops.stats*), 136  
 autocovariance () (*in module pyro.ops.stats*), 137  
 AutoDelta (*class in pyro.infer.autoguide*), 37  
 AutoDiagonalNormal (*class in pyro.infer.autoguide*), 39  
 AutoDiscreteParallel (*class in pyro.infer.autoguide*), 41  
 AutoGuide (*class in pyro.infer.autoguide*), 35  
 autoguide () (*Parameterized method*), 164, 185  
 AutoGuideList (*class in pyro.infer.autoguide*), 36  
 AutoIAFNormal (*class in pyro.infer.autoguide*), 40  
 AutoLaplaceApproximation (*class in pyro.infer.autoguide*), 41  
 AutoLowRankMultivariateNormal (*class in pyro.infer.autoguide*), 40  
 AutoMultivariateNormal (*class in pyro.infer.autoguide*), 39  
 autoregressive (*AffineAutoregressive attribute*), 73  
 autoregressive (*BlockAutoregressive attribute*), 77  
 autoregressive (*Polynomial attribute*), 81  
 AutoRegressiveNN (*class in pyro.nn.auto\_reg\_nn*), 97  
 AVFMultivariateNormal (*class in pyro.distributions*), 51
- ## B
- BackwardSampleMessenger (*class in pyro.infer.traceenum\_elbo*), 14  
 bandwidth\_factor (*IMQSteinKernel attribute*), 23  
 bandwidth\_factor (*RBFSteinKernel attribute*), 23  
 batch\_shape (*Gaussian attribute*), 134  
 BatchNorm (*class in pyro.distributions.transforms*), 75  
 batchnorm () (*in module pyro.distributions.transforms*), 85  
 Bernoulli (*class in pyro.distributions*), 43  
 BernoulliNet (*class in pyro.contrib.cevae*), 155  
 Beta (*class in pyro.distributions*), 43  
 BetaBinomial (*class in pyro.distributions*), 52  
 bijective (*AffineAutoregressive attribute*), 74  
 bijective (*AffineCoupling attribute*), 75  
 bijective (*BatchNorm attribute*), 76  
 bijective (*BlockAutoregressive attribute*), 77  
 bijective (*ConditionalPlanar attribute*), 78  
 bijective (*CorrLCholeskyTransform attribute*), 70  
 bijective (*ELUTransform attribute*), 70  
 bijective (*Householder attribute*), 79  
 bijective (*LeakyReLUTransform attribute*), 71  
 bijective (*LowerCholeskyAffine attribute*), 71  
 bijective (*NeuralAutoregressive attribute*), 80  
 bijective (*Permute attribute*), 72  
 bijective (*Planar attribute*), 80  
 bijective (*Polynomial attribute*), 81  
 bijective (*Sylvester attribute*), 83  
 bijective (*TanhTransform attribute*), 72  
 Binary (*class in pyro.contrib.gp.likelihoods*), 182  
 Binomial (*class in pyro.distributions*), 43  
 block (*class in pyro.contrib.minipyro*), 187  
 block () (*in module pyro.poutine.handlers*), 106  
 block\_autoregressive () (*in module pyro.distributions.transforms*), 85  
 BlockAutoregressive (*class in pyro.distributions.transforms*), 76  
 BlockMessenger (*class in pyro.poutine.block\_messenger*), 118  
 broadcast () (*in module pyro.poutine.handlers*), 107  
 BroadcastMessenger (*class in pyro.poutine.broadcast\_messenger*), 119  
 Brownian (*class in pyro.contrib.gp.kernels*), 175
- ## C
- call () (*AutoGuide method*), 35  
 call () (*Predictive method*), 27  
 cat () (*Gaussian static method*), 134  
 Categorical (*class in pyro.distributions*), 43  
 Cauchy (*class in pyro.distributions*), 44  
 CEVAE (*class in pyro.contrib.cevae*), 151  
 Chi2 (*class in pyro.distributions*), 44  
 cleanup () (*HMC method*), 32  
 clear () (*in module pyro.nn.module*), 96  
 clear () (*ParamStoreDict method*), 89  
 clear\_cache () (*HMC method*), 32  
 clear\_param\_store () (*in module pyro*), 8  
 ClippedAdam (*class in pyro.optim.clipped\_adam*), 101  
 ClippedAdam () (*in module pyro.optim.optim*), 100  
 codomain (*AffineAutoregressive attribute*), 74  
 codomain (*AffineCoupling attribute*), 75  
 codomain (*BatchNorm attribute*), 76  
 codomain (*BlockAutoregressive attribute*), 77  
 codomain (*ConditionalPlanar attribute*), 78  
 codomain (*CorrLCholeskyTransform attribute*), 70  
 codomain (*ELUTransform attribute*), 70  
 codomain (*Householder attribute*), 79  
 codomain (*LeakyReLUTransform attribute*), 71  
 codomain (*LowerCholeskyAffine attribute*), 71  
 codomain (*NeuralAutoregressive attribute*), 80  
 codomain (*Permute attribute*), 72  
 codomain (*Planar attribute*), 80  
 codomain (*Polynomial attribute*), 81

- codomain (*Radial attribute*), 82  
 codomain (*Sylvester attribute*), 83  
 codomain (*TanhTransform attribute*), 72  
 Combination (*class in pyro.contrib.gp.kernels*), 176  
 compute\_log\_prob() (*Trace method*), 114  
 compute\_marginals() (*in module pyro.contrib.tracking.assignment*), 207  
 compute\_marginals() (*TraceEnum\_ELBO method*), 15  
 compute\_marginals\_bp() (*in module pyro.contrib.tracking.assignment*), 207  
 compute\_marginals\_persistent() (*in module pyro.contrib.tracking.assignment*), 207  
 compute\_marginals\_persistent\_bp() (*in module pyro.contrib.tracking.assignment*), 207  
 compute\_marginals\_sparse\_bp() (*in module pyro.contrib.tracking.assignment*), 207  
 compute\_score\_parts() (*Trace method*), 114  
 concentration (*DirichletMultinomial attribute*), 54  
 concentration (*GammaPoisson attribute*), 57  
 concentration (*InverseGamma attribute*), 60  
 concentration0 (*BetaBinomial attribute*), 52  
 concentration1 (*BetaBinomial attribute*), 52  
 CondIndepStackFrame (*class in pyro.poutine.indep\_messenger*), 121  
 condition() (*AffineNormal method*), 135  
 condition() (*ConditionalDistribution method*), 52  
 condition() (*ConditionalPlanar method*), 78  
 condition() (*ConditionalTransform method*), 70  
 condition() (*ConditionalTransformedDistribution method*), 53  
 condition() (*Gaussian method*), 134  
 condition() (*in module pyro.poutine.handlers*), 107  
 conditional() (*in module pyro.contrib.gp.util*), 185  
 conditional\_planar() (*in module pyro.distributions.transforms*), 85  
 ConditionalAutoRegressiveNN (*class in pyro.nn.auto\_reg\_nn*), 97  
 ConditionalDistribution (*class in pyro.distributions*), 52  
 ConditionalPlanar (*class in pyro.distributions.transforms*), 77  
 ConditionalTransform (*class in pyro.distributions*), 70  
 ConditionalTransformedDistribution (*class in pyro.distributions*), 53  
 ConditionalTransformModule (*class in pyro.distributions*), 78  
 ConditionMessenger (*class in pyro.poutine.condition\_messenger*), 119  
 config\_enumerate() (*in module pyro.infer.enum*), 112  
 Constant (*class in pyro.contrib.gp.kernels*), 176  
 constrained\_gamma (*BatchNorm attribute*), 76  
 constraint (*PyroParam attribute*), 93  
 contract() (*in module pyro.ops.einsum*), 132  
 contract\_expression() (*in module pyro.ops.einsum*), 132  
 copy() (*Trace method*), 114  
 Coregionalize (*class in pyro.contrib.gp.kernels*), 176  
 CorrLCholeskyTransform (*class in pyro.distributions.transforms*), 70  
 Cosine (*class in pyro.contrib.gp.kernels*), 177  
 CosineAnnealingLR() (*in module pyro.optim.pytorch\_optimizers*), 102  
 CosineAnnealingWarmRestarts() (*in module pyro.optim.pytorch\_optimizers*), 102  
 cov (*EKFState attribute*), 213  
 cov (*Measurement attribute*), 216  
 cov2pv() (*DynamicModel method*), 209  
 cov2pv() (*Ncp method*), 210  
 cov2pv() (*Ncv method*), 211  
 cov\_pv (*EKFState attribute*), 213  
 covariance\_matrix (*MultivariateStudentT attribute*), 63  
 crps\_empirical() (*in module pyro.ops.stats*), 138  
 CyclicLR() (*in module pyro.optim.pytorch\_optimizers*), 102
- ## D
- default\_process\_message() (*in module pyro.poutine.runtime*), 115  
 Delta (*class in pyro.distributions*), 53  
 DependentMaternGP (*class in pyro.contrib.timeseries.gp*), 201  
 detach\_() (*Trace method*), 114  
 deterministic() (*in module pyro*), 6  
 DiagNormalNet (*class in pyro.contrib.cevae*), 156  
 diagnostics() (*HMC method*), 32  
 diagnostics() (*in module pyro.infer.mcmc.util*), 35  
 diagnostics() (*MCMC method*), 30  
 Dict (*class in pyro.contrib.autoname.named*), 145  
 differentiable\_loss() (*JitTrace\_ELBO method*), 14  
 differentiable\_loss() (*JitTraceEnum\_ELBO method*), 16  
 differentiable\_loss() (*JitTraceMean-Field\_ELBO method*), 17  
 differentiable\_loss() (*Trace\_ELBO method*), 13  
 differentiable\_loss() (*TraceEnum\_ELBO method*), 15  
 differentiable\_loss() (*TraceTMC\_ELBO method*), 18  
 DifferentiableDynamicModel (*class in pyro.contrib.tracking.dynamic\_models*), 209



- DifferentiableMeasurement (class in *pyro.contrib.tracking.measurements*), 216  
 dim() (Gaussian method), 134  
 dimension (DynamicModel attribute), 208  
 dimension (EKFState attribute), 213  
 dimension (Measurement attribute), 216  
 dimension\_pv (DynamicModel attribute), 208  
 dimension\_pv (EKFState attribute), 213  
 Dirichlet (class in *pyro.distributions*), 44  
 DirichletMultinomial (class in *pyro.distributions*), 53  
 discrete\_escape() (in module *pyro.poutine.util*), 116  
 DiscreteHMM (class in *pyro.distributions*), 54  
 Distribution (class in *pyro.distributions*), 47  
 DistributionNet (class in *pyro.contrib.cevae*), 154  
 do() (in module *pyro.poutine.handlers*), 108  
 domain (AffineAutoregressive attribute), 74  
 domain (AffineCoupling attribute), 75  
 domain (BatchNorm attribute), 76  
 domain (BlockAutoregressive attribute), 77  
 domain (ConditionalPlanar attribute), 78  
 domain (CorrLCholeskyTransform attribute), 70  
 domain (ELUTransform attribute), 70  
 domain (Householder attribute), 79  
 domain (LeakyReLUTransform attribute), 71  
 domain (NeuralAutoregressive attribute), 80  
 domain (Planar attribute), 80  
 domain (Polynomial attribute), 81  
 domain (Radial attribute), 82  
 domain (Sylvester attribute), 83  
 domain (TanhTransform attribute), 72  
 DoMessenger (class in *pyro.poutine.do\_messenger*), 120  
 donsker\_varadhan\_eig() (in module *pyro.contrib.oed.eig*), 192  
 DotProduct (class in *pyro.contrib.gp.kernels*), 177  
 dtanh\_dx() (Sylvester method), 83  
 DualAveraging (class in *pyro.ops.dual\_averaging*), 127  
 dynamic\_model (EKFState attribute), 213  
 DynamicModel (class in *pyro.contrib.tracking.dynamic\_models*), 208
- ## E
- easy\_guide() (in module *pyro.contrib.easyguide*), 158  
 EasyGuide (class in *pyro.contrib.easyguide*), 157  
 edges (Trace attribute), 114  
 effectful() (in module *pyro.poutine.runtime*), 115  
 effective\_sample\_size() (in module *pyro.ops.stats*), 137  
 einsum() (in module *pyro.ops.contract*), 132  
 EKFDistribution (class in *pyro.contrib.tracking.distributions*), 208  
 EKFState (class in *pyro.contrib.tracking.extended\_kalman\_filter*), 212  
 ELBO (class in *pyro.infer.elbo*), 12  
 elbo() (in module *pyro.contrib.minipyro*), 188  
 elu() (in module *pyro.distributions.transforms*), 85  
 ELUTransform (class in *pyro.distributions.transforms*), 70  
 Empirical (class in *pyro.distributions*), 55  
 empirical (Marginals attribute), 28  
 EmpiricalMarginal (class in *pyro.infer.abstract\_infer*), 27  
 enable\_validation() (in module *pyro*), 8  
 enable\_validation() (in module *pyro.poutine.util*), 116  
 EnergyDistance (class in *pyro.infer.energy\_distance*), 24  
 enum() (in module *pyro.poutine.handlers*), 108  
 enum\_extend() (in module *pyro.poutine.util*), 116  
 enumerate\_site() (in module *pyro.poutine.enum\_messenger*), 121  
 enumerate\_support() (BetaBinomial method), 52  
 enumerate\_support() (Distribution method), 47  
 enumerate\_support() (Empirical method), 55  
 enumerate\_support() (SpanningTree method), 66  
 EnumMessenger (class in *pyro.poutine.enum\_messenger*), 120  
 escape() (in module *pyro.poutine.handlers*), 108  
 EscapeMessenger (class in *pyro.poutine.escape\_messenger*), 121  
 evaluate\_loss() (SVI method), 11  
 event\_dim (AffineAutoregressive attribute), 74  
 event\_dim (AffineCoupling attribute), 75  
 event\_dim (BatchNorm attribute), 76  
 event\_dim (BlockAutoregressive attribute), 77  
 event\_dim (ConditionalPlanar attribute), 78  
 event\_dim (CorrLCholeskyTransform attribute), 70  
 event\_dim (Householder attribute), 79  
 event\_dim (LowerCholeskyAffine attribute), 71  
 event\_dim (NeuralAutoregressive attribute), 80  
 event\_dim (Permute attribute), 72  
 event\_dim (Planar attribute), 80  
 event\_dim (Polynomial attribute), 81  
 event\_dim (PyroParam attribute), 93  
 event\_dim (Radial attribute), 82  
 event\_dim (Sylvester attribute), 83  
 event\_dim (TorchDistributionMixin attribute), 49  
 event\_logsumexp() (Gaussian method), 135  
 event\_pad() (Gaussian method), 134  
 event\_permute() (Gaussian method), 134  
 event\_shape (Empirical attribute), 56  
 expand() (BetaBinomial method), 52  
 expand() (Delta method), 53



`expand()` (*DirichletMultinomial method*), 54  
`expand()` (*DiscreteHMM method*), 55  
`expand()` (*FoldedDistribution method*), 56  
`expand()` (*GammaPoisson method*), 57  
`expand()` (*Gaussian method*), 134  
`expand()` (*GaussianHMM method*), 58  
`expand()` (*GaussianMRF method*), 59  
`expand()` (*InverseGamma method*), 60  
`expand()` (*LKJCorrCholesky method*), 61  
`expand()` (*MaskedMixture method*), 61  
`expand()` (*MixtureOfDiagNormals method*), 62  
`expand()` (*MixtureOfDiagNormalsSharedCovariance method*), 63  
`expand()` (*MultivariateStudentT method*), 63  
`expand()` (*Stable method*), 67  
`expand()` (*TorchDistribution method*), 51  
`expand()` (*TorchDistributionMixin method*), 49  
`expand()` (*Unit method*), 68  
`expand()` (*VonMises method*), 68  
`expand()` (*VonMises3D method*), 69  
`expand()` (*ZeroInflatedDistribution method*), 70  
`expand_by()` (*TorchDistributionMixin method*), 49  
`Exponent` (*class in pyro.contrib.gp.kernels*), 177  
`Exponential` (*class in pyro.contrib.gp.kernels*), 177  
`Exponential` (*class in pyro.distributions*), 44  
`ExponentialFamily` (*class in pyro.distributions*), 44  
`ExponentialLR()` (*in module pyro.optim.pytorch\_optimizers*), 102  
`ExponentialNet` (*class in pyro.contrib.cevae*), 155

## F

`factor()` (*in module pyro*), 6  
`filter()` (*DiscreteHMM method*), 55  
`filter()` (*GaussianHMM method*), 58  
`filter_states()` (*EKFDistribution method*), 208  
`FisherSnedecor` (*class in pyro.distributions*), 44  
`fit()` (*CEVAE method*), 152  
`fit_generalized_pareto()` (*in module pyro.ops.stats*), 138  
`FoldedDistribution` (*class in pyro.distributions*), 56  
`forecast()` (*DependentMaternGP method*), 202  
`forecast()` (*GenericLGSSM method*), 202  
`forecast()` (*GenericLGSSMWithGPNoiseModel method*), 203  
`forecast()` (*IndependentMaternGP method*), 200  
`forecast()` (*LinearlyCoupledMaternGP method*), 201  
`forecast()` (*TimeSeriesModel method*), 199  
`format_shapes()` (*Trace method*), 114  
`forward()` (*AutoCallable method*), 37  
`forward()` (*AutoContinuous method*), 38  
`forward()` (*AutoDelta method*), 37  
`forward()` (*AutoDiscreteParallel method*), 41

`forward()` (*AutoGuideList method*), 36  
`forward()` (*AutoRegressiveNN method*), 97  
`forward()` (*BernoulliNet method*), 155  
`forward()` (*Binary method*), 182  
`forward()` (*Brownian method*), 176  
`forward()` (*ConditionalAutoRegressiveNN method*), 98  
`forward()` (*Constant method*), 176  
`forward()` (*Coregionalize method*), 177  
`forward()` (*Cosine method*), 177  
`forward()` (*DiagNormalNet method*), 156  
`forward()` (*DynamicModel method*), 208  
`forward()` (*EasyGuide method*), 157  
`forward()` (*Exponent method*), 177  
`forward()` (*Exponential method*), 177  
`forward()` (*ExponentialNet method*), 155  
`forward()` (*Gaussian method*), 182  
`forward()` (*GPLVM method*), 175  
`forward()` (*GPModel method*), 166  
`forward()` (*GPRegression method*), 168  
`forward()` (*Guide method*), 154  
`forward()` (*Kernel method*), 175  
`forward()` (*LaplaceNet method*), 155  
`forward()` (*Likelihood method*), 181  
`forward()` (*Linear method*), 178  
`forward()` (*Matern32 method*), 178  
`forward()` (*Matern52 method*), 178  
`forward()` (*Model method*), 154  
`forward()` (*MultiClass method*), 183  
`forward()` (*Ncp method*), 210  
`forward()` (*Ncv method*), 210  
`forward()` (*NormalNet method*), 156  
`forward()` (*Periodic method*), 179  
`forward()` (*Poisson method*), 184  
`forward()` (*Polynomial method*), 179  
`forward()` (*Predictive method*), 27  
`forward()` (*Product method*), 179  
`forward()` (*RationalQuadratic method*), 180  
`forward()` (*RBF method*), 180  
`forward()` (*SparseGPRegression method*), 170  
`forward()` (*StudentTNet method*), 156  
`forward()` (*Sum method*), 180  
`forward()` (*VariationalGP method*), 172  
`forward()` (*VariationalSparseGP method*), 173  
`forward()` (*VerticalScaling method*), 180  
`forward()` (*Warping method*), 181  
`forward()` (*WhiteNoise method*), 181  
`frame_num` (*EKFState attribute*), 213  
`frame_num` (*Measurement attribute*), 216  
`FullyConnected` (*class in pyro.contrib.cevae*), 154

## G

`Gamma` (*class in pyro.distributions*), 44  
`GammaPoisson` (*class in pyro.distributions*), 56

- Gaussian (class in `pyro.contrib.gp.likelihoods`), 182
- Gaussian (class in `pyro.ops.gaussian`), 134
- `gaussian_tensordot()` (in module `pyro.ops.gaussian`), 136
- GaussianHMM (class in `pyro.distributions`), 57
- GaussianMRF (class in `pyro.distributions`), 58
- GaussianScaleMixture (class in `pyro.distributions`), 59
- `gelman_rubin()` (in module `pyro.ops.stats`), 136
- `generator()` (*MarkovMessenger* method), 122
- GenericLGSSM (class in `pyro.contrib.timeseries.lgssm`), 202
- GenericLGSSMWithGPNoiseModel (class in `pyro.contrib.timeseries.lgssm`), 203
- `geodesic_difference()` (*DynamicModel* method), 209
- `geodesic_difference()` (*Measurement* method), 216
- Geometric (class in `pyro.distributions`), 44
- `get_all_param_names()` (*ParamStoreDict* method), 90
- `get_class()` (*DistributionNet* static method), 155
- `get_covariance()` (*WelfordCovariance* method), 129
- `get_empirical()` (*SMCFilter* method), 22
- `get_ESS()` (*Importance* method), 19
- `get_log_normalizer()` (*Importance* method), 19
- `get_named_particles()` (*SVGD* method), 24
- `get_normalized_weights()` (*Importance* method), 19
- `get_param()` (*ParamStoreDict* method), 90
- `get_param_store()` (in module `pyro`), 8
- `get_param_store()` (in module `pyro.contrib.minipyro`), 188
- `get_permutation()` (*ConditionalAutoRegressiveNN* method), 98
- `get_posterior()` (*AutoContinuous* method), 38
- `get_posterior()` (*AutoDiagonalNormal* method), 40
- `get_posterior()` (*AutoIAFNormal* method), 41
- `get_posterior()` (*AutoLaplaceApproximation* method), 41
- `get_posterior()` (*AutoLowRankMultivariateNormal* method), 40
- `get_posterior()` (*AutoMultivariateNormal* method), 39
- `get_samples()` (*MCMC* method), 30
- `get_samples()` (*Predictive* method), 27
- `get_state()` (*DualAveraging* method), 128
- `get_state()` (*ParamStoreDict* method), 91
- `get_state()` (*PyroOptim* method), 99
- `get_step()` (*MixedMultiOptimizer* method), 103
- `get_step()` (*MultiOptimizer* method), 103
- `get_step()` (*Newton* method), 104
- `get_trace()` (*trace* method), 188
- `get_trace()` (*TraceHandler* method), 125
- `get_trace()` (*TraceMessenger* method), 125
- `get_vectorized_trace()` (*Predictive* method), 27
- GPLVM (class in `pyro.contrib.gp.models.gplvm`), 174
- GPMModel (class in `pyro.contrib.gp.models.model`), 164
- GPRegression (class in `pyro.contrib.gp.models.gpr`), 167
- Group (class in `pyro.contrib.easyguide.easyguide`), 158
- `group()` (*EasyGuide* method), 158
- Guide (class in `pyro.contrib.cevae`), 154
- `guide` (*Group* attribute), 159
- `guide()` (*EasyGuide* method), 157
- `guide()` (*GPLVM* method), 174
- `guide()` (*GPMModel* method), 166
- `guide()` (*GPRegression* method), 168
- `guide()` (*SparseGPRegression* method), 170
- `guide()` (*VariationalGP* method), 172
- `guide()` (*VariationalSparseGP* method), 173
- Gumbel (class in `pyro.distributions`), 44
- ## H
- HalfCauchy (class in `pyro.distributions`), 45
- HalfNormal (class in `pyro.distributions`), 45
- `has_enumerate_support` (*BetaBinomial* attribute), 52
- `has_enumerate_support` (*Distribution* attribute), 48
- `has_enumerate_support` (*Empirical* attribute), 56
- `has_enumerate_support` (*SpanningTree* attribute), 66
- `has_rsample` (*Delta* attribute), 53
- `has_rsample` (*Distribution* attribute), 48
- `has_rsample` (*GaussianScaleMixture* attribute), 60
- `has_rsample` (*InverseGamma* attribute), 60
- `has_rsample` (*LKJCorrCholesky* attribute), 61
- `has_rsample` (*MaskedMixture* attribute), 62
- `has_rsample` (*MixtureOfDiagNormals* attribute), 62
- `has_rsample` (*MixtureOfDiagNormalsSharedCovariance* attribute), 63
- `has_rsample` (*MultivariateStudentT* attribute), 63
- `has_rsample` (*Rejector* attribute), 65
- `has_rsample` (*Stable* attribute), 67
- `has_rsample` (*VonMises* attribute), 68
- HiddenLayer (class in `pyro.contrib.bnn.hidden_layer`), 149
- HMC (class in `pyro.infer.mcmc`), 31
- Householder (class in `pyro.distributions.transforms`), 78
- `householder()` (in module `pyro.distributions.transforms`), 85
- `hpdi()` (in module `pyro.ops.stats`), 138

## I

[identify\\_dense\\_edges\(\)](#) (in module [pyro.poutine.trace\\_messenger](#)), 126  
[Importance](#) (class in [pyro.infer.importance](#)), 19  
[IMQSteinKernel](#) (class in [pyro.infer.svgd](#)), 22  
[Independent](#) (class in [pyro.distributions](#)), 45  
[independent\(\)](#) ([TorchDistributionMixin](#) method), 50  
[IndependentMaternGP](#) (class in [pyro.contrib.timeseries.gp](#)), 200  
[IndepMessenger](#) (class in [pyro.poutine.indep\\_messenger](#)), 121  
[indices](#) ([IndepMessenger](#) attribute), 121  
[infer\\_config\(\)](#) (in module [pyro.poutine.handlers](#)), 108  
[infer\\_discrete\(\)](#) (in module [pyro.infer.discrete](#)), 25  
[InferConfigMessenger](#) (class in [pyro.poutine.infer\\_config\\_messenger](#)), 121  
[information\\_criterion\(\)](#) ([TracePosterior](#) method), 28  
[init\(\)](#) ([EasyGuide](#) method), 157  
[init\(\)](#) ([SMCFilter](#) method), 22  
[init\\_to\\_feasible\(\)](#) (in module [pyro.infer.autoguide.initialization](#)), 42  
[init\\_to\\_mean\(\)](#) (in module [pyro.infer.autoguide.initialization](#)), 42  
[init\\_to\\_median\(\)](#) (in module [pyro.infer.autoguide.initialization](#)), 42  
[init\\_to\\_sample\(\)](#) (in module [pyro.infer.autoguide.initialization](#)), 42  
[init\\_value](#) ([PyroParam](#) attribute), 93  
[initial\\_params](#) ([HMC](#) attribute), 32  
[initialize\\_model\(\)](#) (in module [pyro.infer.mcmc.util](#)), 34  
[InitMessenger](#) (class in [pyro.infer.autoguide.initialization](#)), 42  
[innovation\(\)](#) ([EKFState](#) method), 213  
[inv\\_permutation](#) ([Permute](#) attribute), 72  
[inverse\\_mass\\_matrix](#) ([HMC](#) attribute), 32  
[InverseGamma](#) (class in [pyro.distributions](#)), 60  
[is\\_validation\\_enabled\(\)](#) (in module [pyro.poutine.util](#)), 116  
[Isotropy](#) (class in [pyro.contrib.gp.kernels](#)), 178  
[ite\(\)](#) ([CEVAE](#) method), 153  
[items\(\)](#) ([ParamStoreDict](#) method), 89  
[iter\\_sample\(\)](#) ([GPRegression](#) method), 168  
[iter\\_stochastic\\_nodes\(\)](#) ([Trace](#) method), 114

## J

[jacobian\(\)](#) ([DifferentiableDynamicModel](#) method), 209  
[jacobian\(\)](#) ([DifferentiableMeasurement](#) method), 216  
[jacobian\(\)](#) ([Ncp](#) method), 210  
[jacobian\(\)](#) ([Ncv](#) method), 211

[jacobian\(\)](#) ([PositionMeasurement](#) method), 216  
[JitTrace\\_ELBO](#) (class in [pyro.contrib.minipyro](#)), 187  
[JitTrace\\_ELBO](#) (class in [pyro.infer.trace\\_elbo](#)), 13  
[JitTraceEnum\\_ELBO](#) (class in [pyro.infer.traceenum\\_elbo](#)), 15  
[JitTraceGraph\\_ELBO](#) (class in [pyro.infer.tracegraph\\_elbo](#)), 14  
[JitTraceMeanField\\_ELBO](#) (class in [pyro.infer.trace\\_mean\\_field\\_elbo](#)), 16

## K

[Kernel](#) (class in [pyro.contrib.gp.kernels](#)), 175  
[keys\(\)](#) ([ParamStoreDict](#) method), 90

## L

[LambdaLR\(\)](#) (in module [pyro.optim.pytorch\\_optimizers](#)), 102  
[Laplace](#) (class in [pyro.distributions](#)), 45  
[laplace\\_approximation\(\)](#) ([AutoLaplaceApproximation](#) method), 41  
[laplace\\_eig\(\)](#) (in module [pyro.contrib.oed.eig](#)), 190  
[LaplaceNet](#) (class in [pyro.contrib.cevae](#)), 155  
[leaky\\_relu\(\)](#) (in module [pyro.distributions.transforms](#)), 86  
[LeakyReLUTransform](#) (class in [pyro.distributions.transforms](#)), 71  
[lfire\\_eig\(\)](#) (in module [pyro.contrib.oed.eig](#)), 195  
[lift\(\)](#) (in module [pyro.poutine.handlers](#)), 109  
[LiftMessenger](#) (class in [pyro.poutine.lift\\_messenger](#)), 122  
[Likelihood](#) (class in [pyro.contrib.gp.likelihoods](#)), 181  
[Linear](#) (class in [pyro.contrib.gp.kernels](#)), 178  
[LinearlyCoupledMaternGP](#) (class in [pyro.contrib.timeseries.gp](#)), 200  
[List](#) (class in [pyro.contrib.autoname.named](#)), 144  
[lkj\\_constant\(\)](#) ([LKJCorrCholesky](#) method), 61  
[LKJCorrCholesky](#) (class in [pyro.distributions](#)), 60  
[load\(\)](#) ([ParamStoreDict](#) method), 91  
[load\(\)](#) ([PyroOptim](#) method), 100  
[log\\_abs\\_det\\_jacobian\(\)](#) ([AffineAutoregressive](#) method), 74  
[log\\_abs\\_det\\_jacobian\(\)](#) ([AffineCoupling](#) method), 75  
[log\\_abs\\_det\\_jacobian\(\)](#) ([BatchNorm](#) method), 76  
[log\\_abs\\_det\\_jacobian\(\)](#) ([BlockAutoregressive](#) method), 77  
[log\\_abs\\_det\\_jacobian\(\)](#) ([CorrLCholeskyTransform](#) method), 70  
[log\\_abs\\_det\\_jacobian\(\)](#) ([ELUTransform](#) method), 70  
[log\\_abs\\_det\\_jacobian\(\)](#) ([Householder](#) method), 79

- [log\\_abs\\_det\\_jacobian\(\) \(LeakyReLUTransform method\), 71](#)  
[log\\_abs\\_det\\_jacobian\(\) \(LowerCholeskyAffine method\), 71](#)  
[log\\_abs\\_det\\_jacobian\(\) \(NeuralAutoregressive method\), 80](#)  
[log\\_abs\\_det\\_jacobian\(\) \(Permute method\), 72](#)  
[log\\_abs\\_det\\_jacobian\(\) \(Polynomial method\), 81](#)  
[log\\_abs\\_det\\_jacobian\(\) \(Radial method\), 82](#)  
[log\\_abs\\_det\\_jacobian\(\) \(Sylvester method\), 83](#)  
[log\\_abs\\_det\\_jacobian\(\) \(TanhTransform method\), 72](#)  
[log\\_density\(\) \(Gaussian method\), 134](#)  
[log\\_kernel\\_and\\_grad\(\) \(IMQSteinKernel method\), 23](#)  
[log\\_kernel\\_and\\_grad\(\) \(RBFSteinKernel method\), 23](#)  
[log\\_kernel\\_and\\_grad\(\) \(SteinKernel method\), 24](#)  
[log\\_likelihood\\_of\\_update\(\) \(EKFSState method\), 214](#)  
[log\\_partition\\_function \(SpanningTree attribute\), 66](#)  
[log\\_prob\(\) \(BetaBinomial method\), 52](#)  
[log\\_prob\(\) \(Delta method\), 53](#)  
[log\\_prob\(\) \(DependentMaternGP method\), 202](#)  
[log\\_prob\(\) \(DirichletMultinomial method\), 54](#)  
[log\\_prob\(\) \(DiscreteHMM method\), 55](#)  
[log\\_prob\(\) \(Distribution method\), 48](#)  
[log\\_prob\(\) \(EKFDistribution method\), 208](#)  
[log\\_prob\(\) \(Empirical method\), 56](#)  
[log\\_prob\(\) \(FoldedDistribution method\), 56](#)  
[log\\_prob\(\) \(GammaPoisson method\), 57](#)  
[log\\_prob\(\) \(GaussianHMM method\), 58](#)  
[log\\_prob\(\) \(GaussianMRF method\), 59](#)  
[log\\_prob\(\) \(GaussianScaleMixture method\), 60](#)  
[log\\_prob\(\) \(GenericLGSSM method\), 202](#)  
[log\\_prob\(\) \(GenericLGSSMWithGPNoiseModel method\), 203](#)  
[log\\_prob\(\) \(IndependentMaternGP method\), 200](#)  
[log\\_prob\(\) \(LinearlyCoupledMaternGP method\), 201](#)  
[log\\_prob\(\) \(LKJCorrCholesky method\), 61](#)  
[log\\_prob\(\) \(MaskedMixture method\), 62](#)  
[log\\_prob\(\) \(MixtureOfDiagNormals method\), 62](#)  
[log\\_prob\(\) \(MixtureOfDiagNormalsSharedCovariance method\), 63](#)  
[log\\_prob\(\) \(MultivariateStudentT method\), 63](#)  
[log\\_prob\(\) \(Rejector method\), 65](#)  
[log\\_prob\(\) \(RelaxedBernoulliStraightThrough method\), 64](#)  
[log\\_prob\(\) \(RelaxedOneHotCategoricalStraightThrough method\), 65](#)  
[log\\_prob\(\) \(SpanningTree method\), 66](#)  
[log\\_prob\(\) \(Stable method\), 67](#)  
[log\\_prob\(\) \(TimeSeriesModel method\), 199](#)  
[log\\_prob\(\) \(Unit method\), 68](#)  
[log\\_prob\(\) \(VonMises method\), 68](#)  
[log\\_prob\(\) \(VonMises3D method\), 69](#)  
[log\\_prob\(\) \(ZeroInflatedDistribution method\), 70](#)  
[log\\_prob\\_sum\(\) \(Trace method\), 114](#)  
[log\\_weights \(Empirical attribute\), 56](#)  
[logging\(\) \(HMC method\), 32](#)  
[LogisticNormal \(class in `pyro.distributions`\), 45](#)  
[LogNormal \(class in `pyro.distributions`\), 45](#)  
[loss\(\) \(EnergyDistance method\), 25](#)  
[loss\(\) \(RenyiELBO method\), 18](#)  
[loss\(\) \(ReweightedWakeSleep method\), 21](#)  
[loss\(\) \(Trace\\_ELBO method\), 13](#)  
[loss\(\) \(TraceCausalEffect\\_ELBO method\), 154](#)  
[loss\(\) \(TraceEnum\\_ELBO method\), 15](#)  
[loss\(\) \(TraceGraph\\_ELBO method\), 14](#)  
[loss\(\) \(TraceMeanField\\_ELBO method\), 16](#)  
[loss\(\) \(TraceTailAdaptive\\_ELBO method\), 17](#)  
[loss\(\) \(TraceTMC\\_ELBO method\), 18](#)  
[loss\\_and\\_grads\(\) \(JitTrace\\_ELBO method\), 14](#)  
[loss\\_and\\_grads\(\) \(JitTraceEnum\\_ELBO method\), 16](#)  
[loss\\_and\\_grads\(\) \(JitTraceGraph\\_ELBO method\), 14](#)  
[loss\\_and\\_grads\(\) \(JitTraceMeanField\\_ELBO method\), 17](#)  
[loss\\_and\\_grads\(\) \(RenyiELBO method\), 18](#)  
[loss\\_and\\_grads\(\) \(ReweightedWakeSleep method\), 21](#)  
[loss\\_and\\_grads\(\) \(Trace\\_ELBO method\), 13](#)  
[loss\\_and\\_grads\(\) \(TraceEnum\\_ELBO method\), 15](#)  
[loss\\_and\\_grads\(\) \(TraceGraph\\_ELBO method\), 14](#)  
[loss\\_and\\_grads\(\) \(TraceTMC\\_ELBO method\), 18](#)  
[loss\\_and\\_surrogate\\_loss\(\) \(JitTrace\\_ELBO method\), 13](#)  
[LowerCholeskyAffine \(class in `pyro.distributions.transforms`\), 71](#)  
[LowRankMultivariateNormal \(class in `pyro.distributions`\), 45](#)  
[LSH \(class in `pyro.contrib.tracking.hashing`\), 214](#)
- ## M
- [make\\_dist\(\) \(BernoulliNet static method\), 155](#)  
[make\\_dist\(\) \(ExponentialNet static method\), 155](#)  
[make\\_dist\(\) \(LaplaceNet static method\), 155](#)  
[make\\_dist\(\) \(NormalNet static method\), 156](#)  
[make\\_dist\(\) \(StudentTNet static method\), 156](#)  
[map\\_estimate\(\) \(EasyGuide method\), 158](#)  
[map\\_estimate\(\) \(Group method\), 159](#)  
[marginal\(\) \(TracePosterior method\), 29](#)  
[marginal\(\) \(TracePredictive method\), 29](#)



`marginal_eig()` (in module `pyro.contrib.oed.eig`), 194  
`MarginalAssignment` (class in `pyro.contrib.tracking.assignment`), 205  
`MarginalAssignmentPersistent` (class in `pyro.contrib.tracking.assignment`), 206  
`MarginalAssignmentSparse` (class in `pyro.contrib.tracking.assignment`), 205  
`marginalize()` (Gaussian method), 135  
`Marginals` (class in `pyro.infer.abstract_infer`), 28  
`markov()` (in module `pyro.poutine.handlers`), 109  
`MarkovMessenger` (class in `pyro.poutine.markov_messenger`), 122  
`mask()` (in module `pyro.poutine.handlers`), 110  
`mask()` (`TorchDistributionMixin` method), 50  
`MaskedMixture` (class in `pyro.distributions`), 61  
`MaskMessenger` (class in `pyro.poutine.mask_messenger`), 123  
`match()` (`ParamStoreDict` method), 90  
`Matern32` (class in `pyro.contrib.gp.kernels`), 178  
`Matern52` (class in `pyro.contrib.gp.kernels`), 178  
`MaternKernel` (class in `pyro.ops.ssm_gp`), 139  
`matrix_and_mvn_to_gaussian()` (in module `pyro.ops.gaussian`), 135  
`mc_extend()` (in module `pyro.poutine.util`), 116  
`MCMC` (class in `pyro.infer.mcmc.api`), 29  
`mean` (`BetaBinomial` attribute), 52  
`mean` (`Delta` attribute), 53  
`mean` (`DirichletMultinomial` attribute), 54  
`mean` (`EKFState` attribute), 213  
`mean` (`Empirical` attribute), 56  
`mean` (`GammaPoisson` attribute), 57  
`mean` (`MaskedMixture` attribute), 62  
`mean` (`Measurement` attribute), 216  
`mean` (`MultivariateStudentT` attribute), 63  
`mean` (`VonMises` attribute), 68  
`mean` (`ZeroInflatedDistribution` attribute), 70  
`mean2pv()` (`DynamicModel` method), 209  
`mean2pv()` (`Ncp` method), 210  
`mean2pv()` (`Ncv` method), 211  
`mean_pv` (`EKFState` attribute), 213  
`Measurement` (class in `pyro.contrib.tracking.measurements`), 216  
`median()` (`AutoContinuous` method), 38  
`median()` (`AutoDelta` method), 38  
`median()` (`AutoGuide` method), 35  
`median()` (`AutoGuideList` method), 36  
`merge_points()` (in module `pyro.contrib.tracking.hashing`), 215  
`Messenger` (class in `pyro.contrib.minipyro`), 187  
`Messenger` (class in `pyro.poutine.messenger`), 117  
`MixedMultiOptimizer` (class in `pyro.optim.multi`), 103  
`MixtureOfDiagNormals` (class in `pyro.distributions`), 62  
`MixtureOfDiagNormalsSharedCovariance` (class in `pyro.distributions`), 62  
`mode` (`Parameterized` attribute), 164, 185  
`model` (`AutoGuide` attribute), 35  
`Model` (class in `pyro.contrib.cevae`), 153  
`model` (`EasyGuide` attribute), 157  
`model()` (`GPLVM` method), 174  
`model()` (`GPModel` method), 166  
`model()` (`GPRegression` method), 168  
`model()` (`SparseGPRegression` method), 170  
`model()` (`VariationalGP` method), 172  
`model()` (`VariationalSparseGP` method), 173  
`module()` (in module `pyro`), 6  
`module_from_param_with_module_name()` (in module `pyro.params.param_store`), 91  
`MultiClass` (class in `pyro.contrib.gp.likelihoods`), 183  
`Multinomial` (class in `pyro.distributions`), 45  
`MultiOptimizer` (class in `pyro.optim.multi`), 102  
`MultiStepLR()` (in module `pyro.optim.pytorch_optimizers`), 102  
`MultivariateNormal` (class in `pyro.distributions`), 45  
`MultivariateStudentT` (class in `pyro.distributions`), 63  
`mvn_to_gaussian()` (in module `pyro.ops.gaussian`), 135

## N

`name_count()` (in module `pyro.contrib.autoname`), 142  
`name_count()` (in module `pyro.contrib.autoname.scoping`), 146  
`NameCountMessenger` (class in `pyro.contrib.autoname.scoping`), 145  
`named_parameters()` (`ParamStoreDict` method), 90  
`named_pyro_params()` (`PyroModule` method), 95  
`Ncp` (class in `pyro.contrib.tracking.dynamic_models`), 209  
`NcpContinuous` (class in `pyro.contrib.tracking.dynamic_models`), 211  
`NcpDiscrete` (class in `pyro.contrib.tracking.dynamic_models`), 212  
`Ncv` (class in `pyro.contrib.tracking.dynamic_models`), 210  
`NcvContinuous` (class in `pyro.contrib.tracking.dynamic_models`), 211  
`NcvDiscrete` (class in `pyro.contrib.tracking.dynamic_models`), 212  
`nearby()` (`LSH` method), 215  
`NegativeBinomial` (class in `pyro.distributions`), 46  
`neural_autoregressive()` (in module `pyro.distributions.transforms`), 86

- NeuralAutoregressive (class in `pyro.distributions.transforms`), 79
- Newton (class in `pyro.optim.multi`), 103
- `newton_step()` (in module `pyro.ops.newton`), 129
- `newton_step_1d()` (in module `pyro.ops.newton`), 129
- `newton_step_2d()` (in module `pyro.ops.newton`), 130
- `newton_step_3d()` (in module `pyro.ops.newton`), 130
- `next_context()` (*IndepMessenger* method), 121
- `nmc_eig()` (in module `pyro.contrib.oed.eig`), 191
- NonlocalExit, 115
- `nonreparam_stochastic_nodes` (*Trace attribute*), 114
- Normal (class in `pyro.distributions`), 46
- NormalNet (class in `pyro.contrib.cevae`), 155
- `num_process_noise_parameters` (*DynamicModel attribute*), 208
- `num_steps` (*HMC attribute*), 32
- NUTS (class in `pyro.infer.mcmc`), 32
- ## O
- Object (class in `pyro.contrib.autoname.named`), 143
- `observation_nodes` (*Trace attribute*), 114
- OMTMultivariateNormal (class in `pyro.distributions`), 64
- `OneCycleLR()` (in module `pyro.optim.pytorch_optimizers`), 102
- `OneHotCategorical` (class in `pyro.distributions`), 46
- ## P
- `pack_tensors()` (*Trace method*), 114
- `param()` (in module `pyro`), 5
- `param()` (in module `pyro.contrib.minipyro`), 188
- `param_()` (*Object method*), 144
- `param_name()` (*ParamStoreDict method*), 91
- `param_nodes` (*Trace attribute*), 114
- `param_with_module_name()` (in module `pyro.params.param_store`), 91
- `Parameterized` (class in `pyro.contrib.gp`), 163
- `Parameterized` (class in `pyro.contrib.gp.parameterized`), 184
- `ParamStoreDict` (class in `pyro.params.param_store`), 89
- Pareto (class in `pyro.distributions`), 46
- Periodic (class in `pyro.contrib.gp.kernels`), 179
- Permute (class in `pyro.distributions.transforms`), 71
- `permute()` (in module `pyro.distributions.transforms`), 86
- `pi()` (in module `pyro.ops.stats`), 137
- Planar (class in `pyro.distributions.transforms`), 80
- `planar()` (in module `pyro.distributions.transforms`), 86
- `plate` (class in `pyro`), 7
- `plate()` (*EasyGuide method*), 158
- `plate()` (in module `pyro.contrib.minipyro`), 188
- `PlateMessenger` (class in `pyro.contrib.minipyro`), 187
- `PlateMessenger` (class in `pyro.poutine.plate_messenger`), 123
- Poisson (class in `pyro.contrib.gp.likelihoods`), 184
- Poisson (class in `pyro.distributions`), 46
- Polynomial (class in `pyro.contrib.gp.kernels`), 179
- Polynomial (class in `pyro.distributions.transforms`), 81
- `polynomial()` (in module `pyro.distributions.transforms`), 86
- `PositionMeasurement` (class in `pyro.contrib.tracking.measurements`), 216
- `posterior_eig()` (in module `pyro.contrib.oed.eig`), 193
- `postprocess_message()` (*Messenger method*), 187
- `postprocess_message()` (*trace method*), 188
- `potential_grad()` (in module `pyro.ops.integrator`), 128
- `precision_matrix` (*MultivariateStudentT attribute*), 63
- `predecessors()` (*Trace method*), 114
- `predict()` (*EKFState method*), 213
- `Predictive` (class in `pyro.infer.predictive`), 26
- `prior` (*PyroSample attribute*), 93
- `process_covariance()` (*MaternKernel method*), 139
- `process_message()` (*block method*), 188
- `process_message()` (*Messenger method*), 187
- `process_message()` (*PlateMessenger method*), 187
- `process_message()` (*replay method*), 188
- `process_noise_cov()` (*DynamicModel method*), 209
- `process_noise_cov()` (*Ncp method*), 210
- `process_noise_cov()` (*NcpContinuous method*), 211
- `process_noise_cov()` (*NcpDiscrete method*), 212
- `process_noise_cov()` (*Ncv method*), 211
- `process_noise_cov()` (*NcvContinuous method*), 212
- `process_noise_cov()` (*NcvDiscrete method*), 212
- `process_noise_dist()` (*DynamicModel method*), 209
- `Product` (class in `pyro.contrib.gp.kernels`), 179
- `prune_subsample_sites()` (in module `pyro.poutine.util`), 117
- `psis_diagnostic()` (in module `pyro.infer.importance`), 19
- `pyro.contrib.autoname` (module), 141
- `pyro.contrib.autoname.named` (module), 143
- `pyro.contrib.autoname.scoping` (module), 145

pyro.contrib.bnn (*module*), 149  
 pyro.contrib.bnn.hidden\_layer (*module*), 149  
 pyro.contrib.cevae (*module*), 151  
 pyro.contrib.easyguide (*module*), 157  
 pyro.contrib.gp (*module*), 163  
 pyro.contrib.gp.kernels (*module*), 175  
 pyro.contrib.gp.likelihoods (*module*), 181  
 pyro.contrib.gp.models.gplvm (*module*), 174  
 pyro.contrib.gp.models.gpr (*module*), 167  
 pyro.contrib.gp.models.model (*module*), 164  
 pyro.contrib.gp.models.sgpr (*module*), 169  
 pyro.contrib.gp.models.vgp (*module*), 171  
 pyro.contrib.gp.models.vsgp (*module*), 172  
 pyro.contrib.gp.parameterized (*module*), 184  
 pyro.contrib.gp.util (*module*), 185  
 pyro.contrib.minipyro (*module*), 186  
 pyro.contrib.oed (*module*), 189  
 pyro.contrib.oed.eig (*module*), 190  
 pyro.contrib.oed.glmm (*module*), 196  
 pyro.contrib.timeseries.base (*module*), 199  
 pyro.contrib.timeseries.gp (*module*), 200  
 pyro.contrib.timeseries.lgssm (*module*), 202  
 pyro.contrib.timeseries.lgssmgp (*module*), 203  
 pyro.contrib.tracking (*module*), 205  
 pyro.contrib.tracking.assignment (*module*), 205  
 pyro.contrib.tracking.distributions (*module*), 208  
 pyro.contrib.tracking.dynamic\_models (*module*), 208  
 pyro.contrib.tracking.extended\_kalman\_filter (*module*), 212  
 pyro.contrib.tracking.hashing (*module*), 214  
 pyro.contrib.tracking.measurements (*module*), 216  
 pyro.distributions.torch (*module*), 43  
 pyro.infer.abstract\_infer (*module*), 27  
 pyro.infer.autoguide (*module*), 35  
 pyro.infer.autoguide.initialization (*module*), 42  
 pyro.infer.discrete (*module*), 25  
 pyro.infer.elbo (*module*), 12  
 pyro.infer.energy\_distance (*module*), 24  
 pyro.infer.importance (*module*), 19  
 pyro.infer.predictive (*module*), 26  
 pyro.infer.renyi\_elbo (*module*), 17  
 pyro.infer.rws (*module*), 20  
 pyro.infer.smcfilter (*module*), 22  
 pyro.infer.svgd (*module*), 22  
 pyro.infer.svi (*module*), 11  
 pyro.infer.trace\_elbo (*module*), 13  
 pyro.infer.trace\_mean\_field\_elbo (*module*), 16  
 pyro.infer.trace\_tail\_adaptive\_elbo (*module*), 17  
 pyro.infer.traceenum\_elbo (*module*), 14  
 pyro.infer.tracegraph\_elbo (*module*), 14  
 pyro.infer.tracetmc\_elbo (*module*), 18  
 pyro.nn.module (*module*), 93  
 pyro.ops.dual\_averaging (*module*), 127  
 pyro.ops.einsum (*module*), 132  
 pyro.ops.gaussian (*module*), 134  
 pyro.ops.indexing (*module*), 131  
 pyro.ops.integrator (*module*), 128  
 pyro.ops.newton (*module*), 129  
 pyro.ops.ssm\_gp (*module*), 139  
 pyro.ops.stats (*module*), 136  
 pyro.ops.welford (*module*), 128  
 pyro.optim.adagrad\_rmsprop (*module*), 100  
 pyro.optim.clipped\_adam (*module*), 101  
 pyro.optim.lr\_scheduler (*module*), 100  
 pyro.optim.multi (*module*), 102  
 pyro.optim.optim (*module*), 99  
 pyro.optim.pytorch\_optimizers (*module*), 101  
 pyro.params.param\_store (*module*), 89  
 pyro.poutine.block\_messenger (*module*), 118  
 pyro.poutine.broadcast\_messenger (*module*), 119  
 pyro.poutine.condition\_messenger (*module*), 119  
 pyro.poutine.do\_messenger (*module*), 120  
 pyro.poutine.enum\_messenger (*module*), 120  
 pyro.poutine.escape\_messenger (*module*), 121  
 pyro.poutine.handlers (*module*), 105  
 pyro.poutine.indep\_messenger (*module*), 121  
 pyro.poutine.infer\_config\_messenger (*module*), 121  
 pyro.poutine.lift\_messenger (*module*), 122  
 pyro.poutine.markov\_messenger (*module*), 122  
 pyro.poutine.mask\_messenger (*module*), 123  
 pyro.poutine.messenger (*module*), 117  
 pyro.poutine.plate\_messenger (*module*), 123  
 pyro.poutine.reentrant\_messenger (*module*), 123  
 pyro.poutine.replay\_messenger (*module*), 123  
 pyro.poutine.runtime (*module*), 115  
 pyro.poutine.scale\_messenger (*module*), 124  
 pyro.poutine.seed\_messenger (*module*), 124

- [pyro.poutine.subsample\\_messenger \(module\)](#), 125  
[pyro.poutine.trace\\_messenger \(module\)](#), 125  
[pyro.poutine.uncondition\\_messenger \(module\)](#), 126  
[pyro.poutine.util \(module\)](#), 116  
[pyro\\_method\(\)](#) (in module [pyro.nn.module](#)), 96  
[PyroLRScheduler \(class in \[pyro.optim.lr\\\_scheduler\]\(#\)\)](#), 100  
[PyroModule \(class in \[pyro.nn.module\]\(#\)\)](#), 93  
[PyroMultiOptimizer \(class in \[pyro.optim.multi\]\(#\)\)](#), 103  
[PyroOptim \(class in \[pyro.optim.optim\]\(#\)\)](#), 99  
[PyroParam \(class in \[pyro.nn.module\]\(#\)\)](#), 93  
[PyroSample \(class in \[pyro.nn.module\]\(#\)\)](#), 93
- ## Q
- [Q\(\)](#) ([Sylvester method](#)), 83  
[quantile\(\)](#) (in module [pyro.ops.stats](#)), 137  
[quantiles\(\)](#) ([AutoContinuous method](#)), 38  
[queue\(\)](#) (in module [pyro.poutine.handlers](#)), 110
- ## R
- [R\(\)](#) ([Sylvester method](#)), 83  
[Radial \(class in \[pyro.distributions.transforms\]\(#\)\)](#), 82  
[radial\(\)](#) (in module [pyro.distributions.transforms](#)), 87  
[random\\_module\(\)](#) (in module [pyro](#)), 6  
[rate \(\[GammaPoisson\]\(#\) attribute\)](#), 57  
[rate \(\[InverseGamma\]\(#\) attribute\)](#), 60  
[RationalQuadratic \(class in \[pyro.contrib.gp.kernels\]\(#\)\)](#), 180  
[RBF \(class in \[pyro.contrib.gp.kernels\]\(#\)\)](#), 179  
[RBFSteinKernel \(class in \[pyro.infer.svgd\]\(#\)\)](#), 23  
[ReduceLROnPlateau\(\)](#) (in module [pyro.optim.pytorch\\_optimizers](#)), 102  
[ReentrantMessenger \(class in \[pyro.poutine.reentrant\\\_messenger\]\(#\)\)](#), 123  
[register\(\)](#) ([pyro.poutine.messenger.Messenger](#) class method), 117  
[Rejector \(class in \[pyro.distributions\]\(#\)\)](#), 65  
[RelaxedBernoulli \(class in \[pyro.distributions\]\(#\)\)](#), 46  
[RelaxedBernoulliStraightThrough \(class in \[pyro.distributions\]\(#\)\)](#), 64  
[RelaxedOneHotCategorical \(class in \[pyro.distributions\]\(#\)\)](#), 46  
[RelaxedOneHotCategoricalStraightThrough \(class in \[pyro.distributions\]\(#\)\)](#), 64  
[remove\(\)](#) ([LSH method](#)), 215  
[remove\\_node\(\)](#) ([Trace method](#)), 115  
[RenyiELBO \(class in \[pyro.infer.renyi\\\_elbo\]\(#\)\)](#), 17  
[reparameterized\\_nodes \(Trace attribute\)](#), 115  
[replace\\_param\(\)](#) ([ParamStoreDict method](#)), 90  
[replay \(class in \[pyro.contrib.minipyro\]\(#\)\)](#), 188  
[replay\(\)](#) (in module [pyro.poutine.handlers](#)), 110  
[ReplayMessenger \(class in \[pyro.poutine.replay\\\_messenger\]\(#\)\)](#), 123  
[resample\(\)](#) (in module [pyro.ops.stats](#)), 137  
[reset\(\)](#) ([DualAveraging method](#)), 128  
[reset\(\)](#) ([WelfordCovariance method](#)), 128  
[reset\\_parameters\(\)](#) ([Householder method](#)), 79  
[reset\\_parameters\(\)](#) ([Planar method](#)), 80  
[reset\\_parameters\(\)](#) ([Polynomial method](#)), 82  
[reset\\_parameters\(\)](#) ([Radial method](#)), 82  
[reset\\_parameters2\(\)](#) ([Sylvester method](#)), 83  
[reset\\_stack\(\)](#) ([NonlocalExit method](#)), 115  
[reshape\(\)](#) ([Gaussian method](#)), 134  
[reshape\(\)](#) ([TorchDistributionMixin method](#)), 49  
[ReweightedWakeSleep \(class in \[pyro.infer.rws\]\(#\)\)](#), 20  
[RMSprop\(\)](#) (in module [pyro.optim.pytorch\\_optimizers](#)), 102  
[Rprop\(\)](#) (in module [pyro.optim.pytorch\\_optimizers](#)), 102  
[rsample\(\)](#) ([AVFMMultivariateNormal method](#)), 52  
[rsample\(\)](#) ([Delta method](#)), 53  
[rsample\(\)](#) ([GaussianScaleMixture method](#)), 60  
[rsample\(\)](#) ([MaskedMixture method](#)), 62  
[rsample\(\)](#) ([MixtureOfDiagNormals method](#)), 62  
[rsample\(\)](#) ([MixtureOfDiagNormalsSharedCovariance method](#)), 63  
[rsample\(\)](#) ([MultivariateStudentT method](#)), 63  
[rsample\(\)](#) ([OMTMultivariateNormal method](#)), 64  
[rsample\(\)](#) ([Rejector method](#)), 65  
[rsample\(\)](#) ([RelaxedBernoulliStraightThrough method](#)), 64  
[rsample\(\)](#) ([RelaxedOneHotCategoricalStraightThrough method](#)), 65  
[rsample\(\)](#) ([Stable method](#)), 67  
[run \(MCMC attribute\)](#), 30  
[run\(\)](#) ([SVI method](#)), 12  
[run\(\)](#) ([TracePosterior method](#)), 29
- ## S
- [S\(\)](#) ([Sylvester method](#)), 83  
[sample\(\)](#) ([BetaBinomial method](#)), 52  
[sample\(\)](#) ([DirichletMultinomial method](#)), 54  
[sample\(\)](#) ([Distribution method](#)), 48  
[sample\(\)](#) ([Empirical method](#)), 56  
[sample\(\)](#) ([GammaPoisson method](#)), 57  
[sample\(\)](#) ([Group method](#)), 159  
[sample\(\)](#) ([HMC method](#)), 32  
[sample\(\)](#) (in module [pyro](#)), 5  
[sample\(\)](#) (in module [pyro.contrib.minipyro](#)), 188  
[sample\(\)](#) ([LKJCorrCholesky method](#)), 61  
[sample\(\)](#) ([MaskedMixture method](#)), 62  
[sample\(\)](#) ([NUTS method](#)), 34  
[sample\(\)](#) ([SpanningTree method](#)), 66  
[sample\(\)](#) ([Unit method](#)), 68  
[sample\(\)](#) ([VonMises method](#)), 68



`sample()` (*ZeroInflatedDistribution method*), 70  
`sample_()` (*Object method*), 144  
`sample_latent()` (*AutoContinuous method*), 39  
`sample_latent()` (*AutoGuide method*), 36  
`sample_posterior()` (*TraceEnum\_ELBO method*), 15  
`sample_saved()` (*TraceEnumSample\_ELBO method*), 26  
`sample_size` (*Empirical attribute*), 56  
`save()` (*ParamStoreDict method*), 91  
`save()` (*PyroOptim method*), 99  
`scale()` (*in module pyro.poutine.handlers*), 111  
`ScaleMessenger` (*class pyro.poutine.scale\_messenger*), 124  
`scope()` (*in module pyro.contrib.autoname*), 141  
`scope()` (*in module pyro.contrib.autoname.scoping*), 145  
`ScopeMessenger` (*class pyro.contrib.autoname.scoping*), 145  
`score_parts()` (*Distribution method*), 48  
`score_parts()` (*Rejector method*), 65  
`seed` (*class in pyro.contrib.minipyro*), 188  
`seed()` (*in module pyro.poutine.handlers*), 111  
`SeedMessenger` (*class pyro.poutine.seed\_messenger*), 124  
`set_data()` (*GPMModel method*), 166  
`set_mode()` (*Parameterized method*), 164, 185  
`set_prior()` (*Parameterized method*), 163, 185  
`set_rng_seed()` (*in module pyro*), 9  
`set_state()` (*ParamStoreDict method*), 91  
`set_state()` (*PyroOptim method*), 99  
`setdefault()` (*ParamStoreDict method*), 90  
`setup()` (*HMC method*), 32  
`SGD()` (*in module pyro.optim.pytorch\_optimizers*), 102  
`shape()` (*TorchDistributionMixin method*), 49  
`share_memory()` (*AdagradRMSProp method*), 101  
`sign` (*CorrLCholeskyTransform attribute*), 70  
`sign` (*ELUTransform attribute*), 71  
`sign` (*LeakyReLUTransform attribute*), 71  
`sign` (*TanhTransform attribute*), 72  
`site_is_subsample()` (*in module pyro.poutine.util*), 117  
`SMCFilter` (*class in pyro.infer.smcfilter*), 22  
`SMCState` (*class in pyro.infer.smcfilter*), 22  
`SpanningTree` (*class in pyro.distributions*), 65  
`SparseAdam()` (*in module pyro.optim.pytorch\_optimizers*), 101  
`SparseGPRegression` (*class in pyro.contrib.gp.models.sgpr*), 169  
`split_gelman_rubin()` (*in module pyro.ops.stats*), 136  
`Stable` (*class in pyro.distributions*), 67  
`stationary_covariance()` (*MaternKernel method*), 139  
`SteinKernel` (*class in pyro.infer.svgd*), 24  
`step()` (*AdagradRMSProp method*), 101  
`step()` (*ClippedAdam method*), 101  
`step()` (*DualAveraging method*), 128  
`step()` (*MixedMultiOptimizer method*), 103  
`step()` (*MultiOptimizer method*), 103  
`step()` (*PyroLRScheduler method*), 100  
`step()` (*PyroMultiOptimizer method*), 103  
`step()` (*SMCFilter method*), 22  
`step()` (*SVGD method*), 24  
`step()` (*SVI method*), 12, 187  
`step_size` (*HMC attribute*), 32  
`in StepLR()` (*in module pyro.optim.pytorch\_optimizers*), 102  
`stochastic_nodes` (*Trace attribute*), 115  
`StudentT` (*class in pyro.distributions*), 46  
`StudentTNet` (*class in pyro.contrib.cevae*), 156  
`in SubsampleMessenger` (*class in pyro.poutine.subsample\_messenger*), 125  
`successors()` (*Trace method*), 115  
`Sum` (*class in pyro.contrib.gp.kernels*), 180  
`summary()` (*MCMC method*), 30  
`support` (*BetaBinomial attribute*), 52  
`in support` (*Delta attribute*), 53  
`support` (*DirichletMultinomial attribute*), 54  
`support` (*Empirical attribute*), 56  
`support` (*FoldedDistribution attribute*), 56  
`support` (*GammaPoisson attribute*), 57  
`support` (*InverseGamma attribute*), 60  
`support` (*LKJCorrCholesky attribute*), 61  
`support` (*MaskedMixture attribute*), 62  
`support` (*MultivariateStudentT attribute*), 63  
`support` (*SpanningTree attribute*), 66  
`support` (*Stable attribute*), 67  
`support` (*Unit attribute*), 68  
`support` (*VonMises attribute*), 68  
`support` (*VonMises3D attribute*), 69  
`support` (*ZeroInflatedNegativeBinomial attribute*), 69  
`support` (*ZeroInflatedPoisson attribute*), 69  
`support()` (*Marginals method*), 28  
`SVGD` (*class in pyro.infer.svgd*), 23  
`SVI` (*class in pyro.contrib.minipyro*), 187  
`SVI` (*class in pyro.infer.svi*), 11  
`Sylvester` (*class in pyro.distributions.transforms*), 82  
`sylvester()` (*in module pyro.distributions.transforms*), 87  
`symbolize_dims()` (*Trace method*), 115  

## T

`t_dist()` (*Guide method*), 154  
`t_dist()` (*Model method*), 154  
`tanh()` (*in module pyro.distributions.transforms*), 87  
`TanhTransform` (*class in pyro.distributions.transforms*), 72

time (*EKFState attribute*), 213  
 time (*Measurement attribute*), 216  
 TimeSeriesModel (class in *pyro.contrib.timeseries.base*), 199  
 to\_event() (*TorchDistributionMixin method*), 49  
 to\_pyro\_module() (in module *pyro.nn.module*), 96  
 to\_script\_module() (*CEVAE method*), 153  
 topological\_sort() (*Trace method*), 115  
 TorchDistribution (class in *pyro.distributions*), 50  
 TorchDistributionMixin (class in *pyro.distributions.torch\_distribution*), 48  
 TorchMultiOptimizer (class in *pyro.optim.multi*), 103  
 trace (class in *pyro.contrib.minipyro*), 188  
 Trace (class in *pyro.poutine*), 113  
 trace (*TraceHandler attribute*), 125  
 trace() (in module *pyro.ops.jit*), 9  
 trace() (in module *pyro.poutine.handlers*), 111  
 Trace\_ELBO (class in *pyro.infer.trace\_elbo*), 13  
 Trace\_ELBO() (in module *pyro.contrib.minipyro*), 187  
 TraceCausalEffect\_ELBO (class in *pyro.contrib.cevae*), 154  
 TraceEnum\_ELBO (class in *pyro.infer.traceenum\_elbo*), 15  
 TraceEnumSample\_ELBO (class in *pyro.infer.discrete*), 26  
 TraceGraph\_ELBO (class in *pyro.infer.tracegraph\_elbo*), 14  
 TraceHandler (class in *pyro.poutine.trace\_messenger*), 125  
 TraceMeanField\_ELBO (class in *pyro.infer.trace\_mean\_field\_elbo*), 16  
 TraceMessenger (class in *pyro.poutine.trace\_messenger*), 125  
 TracePosterior (class in *pyro.infer.abstract\_infer*), 28  
 TracePredictive (class in *pyro.infer.abstract\_infer*), 29  
 TraceTailAdaptive\_ELBO (class in *pyro.infer.trace\_tail\_adaptive\_elbo*), 17  
 TraceTMC\_ELBO (class in *pyro.infer.tracetmc\_elbo*), 18  
 train() (in module *pyro.contrib.gp.util*), 186  
 TransformedDistribution (class in *pyro.distributions*), 47  
 Transforming (class in *pyro.contrib.gp.kernels*), 180  
 TransformModule (class in *pyro.distributions*), 83  
 transition\_matrix() (*MaternKernel method*), 139  
 transition\_matrix\_and\_covariance() (*MaternKernel method*), 139  
 try\_add() (*ApproxSet method*), 215

**U**  
 u() (*Householder method*), 79  
 ubersum() (in module *pyro.ops.contract*), 134  
 uncondition() (in module *pyro.poutine.handlers*), 112  
 UnconditionMessenger (class in *pyro.poutine.uncondition\_messenger*), 126  
 Uniform (class in *pyro.distributions*), 47  
 Unit (class in *pyro.distributions*), 67  
 unregister() (*pyro.poutine.messenger.Messenger class method*), 117  
 update() (*EKFState method*), 214  
 update() (*WelfordCovariance method*), 129  
 user\_param\_name() (in module *pyro.params.param\_store*), 91

**V**  
 validate\_edges() (*SpanningTree method*), 66  
 validation\_enabled() (in module *pyro*), 8  
 values() (*ParamStoreDict method*), 90  
 variance (*BetaBinomial attribute*), 52  
 variance (*Delta attribute*), 53  
 variance (*DirichletMultinomial attribute*), 54  
 variance (*Empirical attribute*), 56  
 variance (*GammaPoisson attribute*), 57  
 variance (*MaskedMixture attribute*), 62  
 variance (*MultivariateStudentT attribute*), 64  
 variance (*VonMises attribute*), 68  
 variance (*ZeroInflatedDistribution attribute*), 70  
 VariationalGP (class in *pyro.contrib.gp.models.vgp*), 171  
 VariationalSparseGP (class in *pyro.contrib.gp.models.vsgp*), 172  
 vectorize() (in module *pyro.infer.svgd*), 24  
 vectorized (*CondIndepStackFrame attribute*), 121  
 vectorized\_importance\_weights() (in module *pyro.infer.importance*), 19  
 velocity\_verlet() (in module *pyro.ops.integrator*), 128  
 VerticalScaling (class in *pyro.contrib.gp.kernels*), 180  
 vi\_eig() (in module *pyro.contrib.oed.eig*), 191  
 Vindex (class in *pyro.ops.indexing*), 132  
 vindex() (in module *pyro.ops.indexing*), 131  
 vnmcmc\_eig() (in module *pyro.contrib.oed.eig*), 195  
 volume\_preserving (*Householder attribute*), 79  
 volume\_preserving (*LowerCholeskyAffine attribute*), 71  
 volume\_preserving (*Permute attribute*), 72  
 VonMises (class in *pyro.distributions*), 68  
 VonMises3D (class in *pyro.distributions*), 68

**W**  
 waic() (in module *pyro.ops.stats*), 138

Warping (*class in* `pyro.contrib.gp.kernels`), [181](#)  
Weibull (*class in* `pyro.distributions`), [47](#)  
WelfordCovariance (*class in* `pyro.ops.welford`), [128](#)  
WhiteNoise (*class in* `pyro.contrib.gp.kernels`), [181](#)

## X

`x_dist()` (*Model method*), [154](#)

## Y

`y_dist()` (*Guide method*), [154](#)  
`y_dist()` (*Model method*), [154](#)  
`y_mean()` (*Model method*), [154](#)

## Z

`z_dist()` (*Guide method*), [154](#)  
`z_dist()` (*Model method*), [154](#)  
`ZeroInflatedDistribution` (*class in* `pyro.distributions`), [69](#)  
`ZeroInflatedNegativeBinomial` (*class in* `pyro.distributions`), [69](#)  
`ZeroInflatedPoisson` (*class in* `pyro.distributions`), [69](#)