
BayesPy Documentation

Release 0.4.0

Jaakko Luttinen

November 02, 2015

CONTENTS

| | | |
|----------|--|------------|
| 1 | Introduction | 1 |
| 1.1 | Project information | 1 |
| 1.2 | Similar projects | 1 |
| 1.3 | Contributors | 2 |
| 1.4 | Version history | 2 |
| 2 | User guide | 7 |
| 2.1 | Installation | 7 |
| 2.2 | Quick start guide | 9 |
| 2.3 | Constructing the model | 11 |
| 2.4 | Performing inference | 19 |
| 2.5 | Examining the results | 23 |
| 2.6 | Advanced topics | 29 |
| 3 | Examples | 33 |
| 3.1 | Linear regression | 33 |
| 3.2 | Gaussian mixture model | 38 |
| 3.3 | Bernoulli mixture model | 42 |
| 3.4 | Hidden Markov model | 46 |
| 3.5 | Principal component analysis | 53 |
| 3.6 | Linear state-space model | 56 |
| 3.7 | Latent Dirichlet allocation | 64 |
| 4 | Developer guide | 69 |
| 4.1 | Workflow | 69 |
| 4.2 | Variational message passing | 70 |
| 4.3 | Implementing inference engines | 73 |
| 4.4 | Implementing nodes | 73 |
| 5 | User API | 79 |
| 5.1 | bayespy.nodes | 79 |
| 5.2 | bayespy.inference | 80 |
| 5.3 | bayespy.plot | 87 |
| 6 | Developer API | 93 |
| 6.1 | Developer nodes | 93 |
| 6.2 | Moments | 106 |
| 6.3 | Distributions | 116 |
| 6.4 | Utility functions | 145 |
| | Bibliography | 167 |

| | |
|----------------------------|------------|
| Python Module Index | 169 |
| Index | 171 |

INTRODUCTION

BayesPy provides tools for Bayesian inference with Python. The user constructs a model as a Bayesian network, observes data and runs posterior inference. The goal is to provide a tool which is efficient, flexible and extendable enough for expert use but also accessible for more casual users.

Currently, only variational Bayesian inference for conjugate-exponential family (variational message passing) has been implemented. Future work includes variational approximations for other types of distributions and possibly other approximate inference methods such as expectation propagation, Laplace approximations, Markov chain Monte Carlo (MCMC) and other methods. Contributions are welcome.

1.1 Project information

Copyright (C) 2011-2015 Jaakko Luttinen and other contributors (see below)

BayesPy including the documentation is licensed under the MIT License. See LICENSE file for a text of the license or visit <http://opensource.org/licenses/MIT>.

- Documentation:
 - <http://bayespy.org>
 - PDF file
 - RST format in doc directory
- Repository: <https://github.com/bayespy/bayespy.git>
- Bug reports: <https://github.com/bayespy/bayespy/issues>
- Mailing list: bayespy@googlegroups.com
- IRC: #bayespy @ freenode
- Author: Jaakko Luttinen jaakko.luttinen@iki.fi
- Latest release:
- Build status:
- Unit test coverage:

1.2 Similar projects

VIBES (<http://vibes.sourceforge.net/>) allows variational inference to be performed automatically on a Bayesian network. It is implemented in Java and released under revised BSD license.

Bayes Blocks (<http://research.ics.aalto.fi/bayes/software/>) is a C++/Python implementation of the variational building block framework. The framework allows easy learning of a wide variety of models using variational Bayesian learning. It is available as free software under the GNU General Public License.

Infer.NET (<http://research.microsoft.com/infernet/>) is a .NET framework for machine learning. It provides message-passing algorithms and statistical routines for performing Bayesian inference. It is partly closed source and licensed for non-commercial use only.

PyMC (<https://github.com/pymc-devs/pymc>) provides MCMC methods in Python. It is released under the Academic Free License.

OpenBUGS (<http://www.openbugs.info>) is a software package for performing Bayesian inference using Gibbs sampling. It is released under the GNU General Public License.

Dimple (<http://dimple.problog.org/>) provides Gibbs sampling, belief propagation and a few other inference algorithms for Matlab and Java. It is released under the Apache License.

Stan (<http://mc-stan.org/>) provides inference using MCMC with an interface for R and Python. It is released under the New BSD License.

PBNT - Python Bayesian Network Toolbox (<http://pbnt.berlios.de/>) is Bayesian network library in Python supporting static networks with discrete variables. There was no information about the license.

1.3 Contributors

The list of contributors:

- Jaakko Luttinen
- Hannu Hartikainen

Each file or the git log can be used for more detailed information.

1.4 Version history

1.4.1 Version 0.4.0 (2015-11-02)

- Implement Add node for Gaussian nodes
- Raise error if attempting to install on Python 2
- Return both relative and absolute errors from numerical gradient checking
- Add nose plugin to filter unit test warnings appropriately

1.4.2 Version 0.3.9 (2015-10-16)

- Fix Gaussian ARD node sampling

1.4.3 Version 0.3.8 (2015-10-16)

- Fix Gaussian node sampling

1.4.4 Version 0.3.7 (2015-09-23)

- Enable keyword arguments when plotting via the inference engine
- Add initial support for logging

1.4.5 Version 0.3.6 (2015-08-12)

- Add maximum likelihood node for the shape parameter of Gamma
- Fix Hinton diagrams for 1-D and 0-D Gaussians
- Fix autosave interval counter
- Fix bugs in constant nodes

1.4.6 Version 0.3.5 (2015-06-09)

- Fix indexing bug in VB optimization (not VB-EM)
- Fix demos

1.4.7 Version 0.3.4 (2015-06-09)

- Fix computation of probability density of Dirichlet nodes
- Use unit tests for all code snippets in docstrings and documentation

1.4.8 Version 0.3.3 (2015-06-05)

- Change license to the MIT license
- Improve SumMultiply efficiency
- Hinton diagrams for gamma variables
- Possible to load only nodes from HDF5 results

1.4.9 Version 0.3.2 (2015-03-16)

- Concatenate node added
- Unit tests for plotting fixed

1.4.10 Version 0.3.1 (2015-03-12)

- Gaussian mixture 2D plotting improvements
- Covariance matrix sampling improvements
- Minor documentation fixes

1.4.11 Version 0.3 (2015-03-05)

- Add gradient-based optimization methods (Riemannian/natural gradient or normal)
- Add collapsed inference
- Add the pattern search method
- Add deterministic annealing
- Add stochastic variational inference
- Add optional input signals to Gaussian Markov chains
- Add unit tests for plotting functions (by Hannu Hartikainen)
- Add printing support to nodes
- Drop Python 3.2 support

1.4.12 Version 0.2.3 (2014-12-03)

- Fix matplotlib compatibility broken by recent changes in matplotlib
- Add random sampling for Binomial and Bernoulli nodes
- Fix minor bugs, for instance, in plot module

1.4.13 Version 0.2.2 (2014-11-01)

- Fix normalization of categorical Markov chain probabilities (fixes HMM demo)
- Fix initialization from parameter values

1.4.14 Version 0.2.1 (2014-09-30)

- Add workaround for matplotlib 1.4.0 bug related to interactive mode which affected monitoring
- Fix bugs in Hinton diagrams for Gaussian variables

1.4.15 Version 0.2 (2014-08-06)

- Added all remaining common distributions: Bernoulli, binomial, multinomial, Poisson, beta, exponential.
- Added Gaussian arrays (not just scalars or vectors).
- Added Gaussian Markov chains with time-varying or switching dynamics.
- Added discrete Markov chains (enabling hidden Markov models).
- Added joint Gaussian-Wishart and Gaussian-gamma nodes.
- Added deterministic gating node.
- Added deterministic general sum-product node.
- Added parameter expansion for Gaussian arrays and time-varying/switching Gaussian Markov chains.
- Added new plotting functions: pdf, Hinton diagram.
- Added monitoring of posterior distributions during iteration.

- Finished documentation and added API.

1.4.16 Version 0.1 (2013-07-25)

- Added variational message passing inference engine.
- Added the following common distributions: Gaussian vector, gamma, Wishart, Dirichlet, categorical.
- Added Gaussian Markov chain.
- Added parameter expansion for Gaussian vectors and Gaussian Markov chain.
- Added stochastic mixture node.
- Added deterministic dot product node.
- Created preliminary version of the documentation.

2.1 Installation

BayesPy is a Python 3 package and it can be installed from PyPI or the latest development version from GitHub. The instructions below explain how to set up the system by installing required packages, how to install BayesPy and how to compile this documentation yourself. However, if these instructions contain errors or some relevant details are missing, please file a bug report at <https://github.com/bayespy/bayespy/issues>.

2.1.1 Installing BayesPy

BayesPy can be installed easily by using Pip if the system has been properly set up. If you have problems with the following methods, see the following section for some help on installing the requirements. For instance, a bug in recent versions of h5py and pip may require you to install some of the requirements manually.

For users

First, you may want to set up a virtual environment. Using virtual environment is optional but recommended. To create and activate a new virtual environment, run (in the folder in which you want to create the environment):

```
virtualenv -p python3 --system-site-packages ENV
source ENV/bin/activate
```

The latest release of BayesPy can be installed from PyPI simply as

```
pip install bayespy
```

If you want to install the latest development version of BayesPy, use GitHub instead:

```
pip install https://github.com/bayespy/bayespy/archive/master.zip
```

For developers

If you want to install the development version of BayesPy in such a way that you can easily edit the package, follow these instructions. Get the git repository:

```
git clone https://github.com/bayespy/bayespy.git
cd bayespy
```

Create and activate a new virtual environment (optional but recommended):

```
virtualenv -p python3 --system-site-packages ENV
source ENV/bin/activate
```

Install BayesPy in editable mode:

```
pip install -e .
```

Checking installation

If you have problems installing BayesPy, read the next section for more details. It is recommended to run the unit tests in order to check that BayesPy is working properly. Thus, install Nose and run the unit tests:

```
pip install nose
nosetests bayespy
```

2.1.2 Installing requirements

BayesPy requires Python 3.3 (or later) and the following packages:

- NumPy ($\geq 1.8.0$),
- SciPy ($\geq 0.13.0$)
- matplotlib (≥ 1.2)
- h5py

Ideally, Pip should install the necessary requirements and a manual installation of these dependencies is not required. However, there are several reasons why the installation of these dependencies needs to be done manually in some cases. Thus, this section tries to give some details on how to set up your system. A proper installation of the dependencies for Python 3 can be a bit tricky and you may refer to <http://www.scipy.org/install.html> for more detailed instructions about the SciPy stack. Detailed instructions on installing recent SciPy stack for various platforms is out of the scope of these instructions, but we provide some general guidance here. There are basically three ways to install the dependencies:

1. Install a Python distribution which includes the packages. For Windows, Mac and Linux, there are several Python distributions which include all the necessary packages: <http://www.scipy.org/install.html#scientific-python-distributions>. For instance, you may try *Anaconda* or *Enthought*.
2. Install the packages using the system package manager. On Linux, the packages might be called something like `python-scipy` or `scipy`. However, it is possible that these system packages are not recent enough for BayesPy.
3. Install the packages using Pip:

```
pip install "distribute>=0.6.28"
pip install "numpy>=1.8.0" "scipy>=0.13.0" "matplotlib>=1.2" h5py
```

This also makes sure you have recent enough version of Distribute (required by Matplotlib). However, this installation method may require that the system has some libraries needed for compiling (e.g., C compiler, Python development files, BLAS/LAPACK). For instance, on Ubuntu (≥ 12.10), you may install the required system libraries for each package as:

```
sudo apt-get build-dep python3-numpy
sudo apt-get build-dep python3-scipy
sudo apt-get build-dep python3-matplotlib
sudo apt-get build-dep python-h5py
```

Then installation using Pip should work.

2.1.3 Compiling documentation

This documentation can be found at <http://bayespy.org/> in HTML and PDF formats. The documentation source files are also readable as such in reStructuredText format in `doc/source/` directory. It is possible to compile the documentation into HTML or PDF yourself. In order to compile the documentation, Sphinx is required and a few extensions for it. Those can be installed as:

```
pip install "sphinx>=1.2.3" sphinxcontrib-tikz sphinxcontrib-bayesnet sphinxcontrib-bibtex "numpydoc"
```

In order to visualize graphical models in HTML, you need to have ImageMagick or Netpbm installed. The documentation can be compiled to HTML and PDF by running the following commands in the `doc` directory:

```
make html
make latexpdf
```

You can also run doctest to test code snippets in the documentation:

```
make doctest
```

or in the docstrings:

```
nosetests --with-doctest --doctest-options="+ELLIPSIS" bayespy
```

2.2 Quick start guide

This short guide shows the key steps in using BayesPy for variational Bayesian inference by applying BayesPy to a simple problem. The key steps in using BayesPy are the following:

- Construct the model
- Observe some of the variables by providing the data in a proper format
- Run variational Bayesian inference
- Examine the resulting posterior approximation

To demonstrate BayesPy, we'll consider a very simple problem: we have a set of observations from a Gaussian distribution with unknown mean and variance, and we want to learn these parameters. In this case, we do not use any real-world data but generate some artificial data. The dataset consists of ten samples from a Gaussian distribution with mean 5 and standard deviation 10. This dataset can be generated with NumPy as follows:

```
>>> import numpy as np
>>> data = np.random.normal(5, 10, size=(10,))
```

2.2.1 Constructing the model

Now, given this data we would like to estimate the mean and the standard deviation as if we didn't know their values. The model can be defined as follows:

$$p(\mathbf{y}|\mu, \tau) = \prod_{n=0}^9 \mathcal{N}(y_n|\mu, \tau)$$

$$p(\mu) = \mathcal{N}(\mu|0, 10^{-6})$$

$$p(\tau) = \mathcal{G}(\tau|10^{-6}, 10^{-6})$$

where \mathcal{N} is the Gaussian distribution parameterized by its mean and precision (i.e., inverse variance), and \mathcal{G} is the gamma distribution parameterized by its shape and rate parameters. Note that we have given quite uninformative

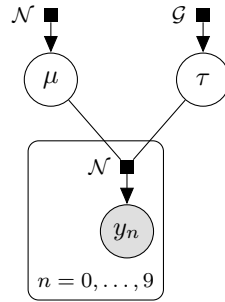


Fig. 2.1: Directed factor graph of the example model.

priors for the variables μ and τ . This simple model can also be shown as a directed factor graph: This model can be constructed in BayesPy as follows:

```
>>> from bayespy.nodes import GaussianARD, Gamma
>>> mu = GaussianARD(0, 1e-6)
>>> tau = Gamma(1e-6, 1e-6)
>>> y = GaussianARD(mu, tau, plates=(10,))
```

This is quite self-explanatory given the model definitions above. We have used two types of nodes `GaussianARD` and `Gamma` to represent Gaussian and gamma distributions, respectively. There are much more distributions in `bayespy.nodes` so you can construct quite complex conjugate exponential family models. The node `y` uses keyword argument `plates` to define the plates $n = 0, \dots, 9$.

2.2.2 Performing inference

Now that we have created the model, we can provide our data by setting `y` as observed:

```
>>> y.observe(data)
```

Next we want to estimate the posterior distribution. In principle, we could use different inference engines (e.g., MCMC or EP) but currently only variational Bayesian (VB) engine is implemented. The engine is initialized by giving all the nodes of the model:

```
>>> from bayespy.inference import VB
>>> Q = VB(mu, tau, y)
```

The inference algorithm can be run as long as wanted (max. 20 iterations in this case):

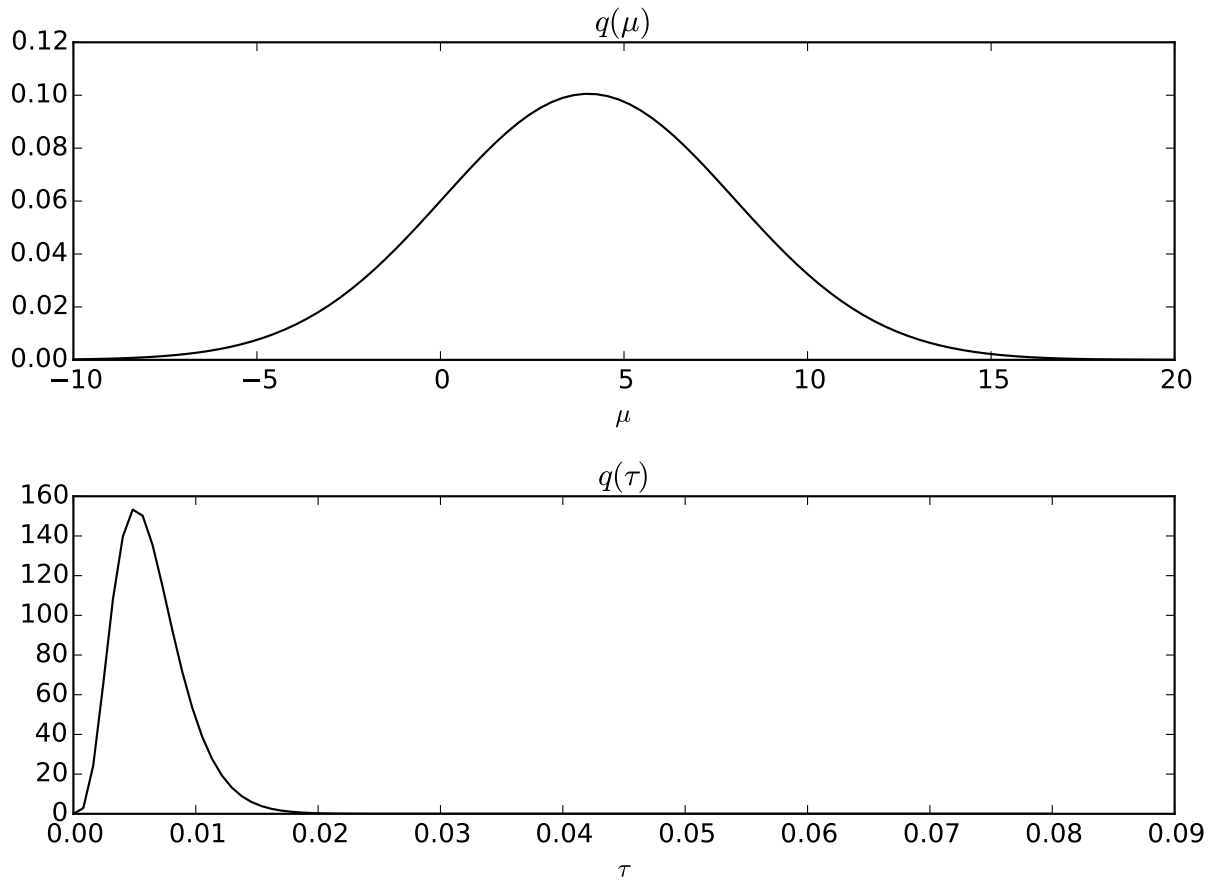
```
>>> Q.update(repeat=20)
Iteration 1: loglike=-6.020956e+01 (... seconds)
Iteration 2: loglike=-5.820527e+01 (... seconds)
Iteration 3: loglike=-5.820290e+01 (... seconds)
Iteration 4: loglike=-5.820288e+01 (... seconds)
Converged at iteration 4.
```

Now the algorithm converged after four iterations, before the requested 20 iterations. VB approximates the true posterior $p(\mu, \tau | y)$ with a distribution which factorizes with respect to the nodes: $q(\mu)q(\tau)$.

2.2.3 Examining posterior approximation

The resulting approximate posterior distributions $q(\mu)$ and $q(\tau)$ can be examined, for instance, by plotting the marginal probability density functions:

```
>>> import bayespy.plot as bpplt
>>> bpplt.pyplot.subplot(2, 1, 1)
<matplotlib.axes...AxesSubplot object at 0x...>
>>> bpplt.pdf(mu, np.linspace(-10, 20, num=100), color='k', name=r'\mu')
[<matplotlib.lines.Line2D object at 0x...>]
>>> bpplt.pyplot.subplot(2, 1, 2)
<matplotlib.axes...AxesSubplot object at 0x...>
>>> bpplt.pdf(tau, np.linspace(1e-6, 0.08, num=100), color='k', name=r'\tau')
[<matplotlib.lines.Line2D object at 0x...>]
>>> bpplt.pyplot.tight_layout()
>>> bpplt.pyplot.show()
```



This example was a very simple introduction to using BayesPy. The model can be much more complex and each phase contains more options to give the user more control over the inference. The following sections give more details about the phases.

2.3 Constructing the model

In BayesPy, the model is constructed by creating nodes which form a directed network. There are two types of nodes: stochastic and deterministic. A stochastic node corresponds to a random variable (or a set of random variables) from a specific probability distribution. A deterministic node corresponds to a deterministic function of its parents. For a list of built-in nodes, see the [User API](#).

2.3.1 Creating nodes

Creating a node is basically like writing the conditional prior distribution of the variable in Python. The node is constructed by giving the parent nodes, that is, the conditioning variables as arguments. The number of parents and their meaning depend on the node. For instance, a `Gaussian` node is created by giving the mean vector and the precision matrix. These parents can be constant numerical arrays if they are known:

```
>>> from bayespy.nodes import Gaussian
>>> X = Gaussian([2, 5], [[1.0, 0.3], [0.3, 1.0]])
```

or other nodes if they are unknown and given prior distributions:

```
>>> from bayespy.nodes import Gaussian, Wishart
>>> mu = Gaussian([0, 0], [[1e-6, 0], [0, 1e-6]])
>>> Lambda = Wishart(2, [[1, 0], [0, 1]])
>>> X = Gaussian(mu, Lambda)
```

Nodes can also be named by providing `name` keyword argument:

```
>>> X = Gaussian(mu, Lambda, name='x')
```

The name may be useful when referring to the node using an inference engine.

For the parent nodes, there are two main restrictions: non-constant parent nodes must be conjugate and the parent nodes must be mutually independent in the posterior approximation.

Conjugacy of the parents

In Bayesian framework in general, one can give quite arbitrary probability distributions for variables. However, one often uses distributions that are easy to handle in practice. Quite often this means that the parents are given conjugate priors. This is also one of the limitations in BayesPy: only conjugate family prior distributions are accepted currently. Thus, although in principle one could give, for instance, gamma prior for the mean parameter `mu`, only Gaussian-family distributions are accepted because of the conjugacy. If the parent is not of a proper type, an error is raised. This conjugacy is checked automatically by BayesPy and `NoConverterError` is raised if a parent cannot be interpreted as being from a conjugate distribution.

Independence of the parents

Another a bit rarely encountered limitation is that the parents must be mutually independent (in the posterior factorization). Thus, a node cannot have the same stochastic node as several parents without intermediate stochastic nodes. For instance, the following leads to an error:

```
>>> from bayespy.nodes import Dot
>>> Y = Dot(X, X)
Traceback (most recent call last):
...
ValueError: Parent nodes are not independent
```

The error is raised because `X` is given as two parents for `Y`, and obviously `X` is not independent of `X` in the posterior approximation. Even if `X` is not given several times directly but there are some intermediate deterministic nodes, an error is raised because the deterministic nodes depend on their parents and thus the parents of `Y` would not be independent. However, it is valid that a node is a parent of another node via several paths if all the paths or all except one path has intermediate stochastic nodes. This is valid because the intermediate stochastic nodes have independent posterior approximations. Thus, for instance, the following construction does not raise errors:


```
>>> from bayespy.nodes import Dot
>>> Z = Gaussian(X, [[1,0], [0,1]])
>>> Y = Dot(X, Z)
```

This works because there is now an intermediate stochastic node Z on the other path from X node to Y node.

2.3.2 Effects of the nodes on inference

When constructing the network with nodes, the stochastic nodes actually define three important aspects:

1. the prior probability distribution for the variables,
2. the factorization of the posterior approximation,
3. the functional form of the posterior approximation for the variables.

Prior probability distribution

First, the most intuitive feature of the nodes is that they define the prior distribution. In the previous example, `mu` was a stochastic `GaussianARD` node corresponding to μ from the normal distribution, `tau` was a stochastic `Gamma` node corresponding to τ from the gamma distribution, and `y` was a stochastic `GaussianARD` node corresponding to y from the normal distribution with mean μ and precision τ . If we denote the set of all stochastic nodes by Ω , and by π_X the set of parents of a node X , the model is defined as

$$p(\Omega) = \prod_{X \in \Omega} p(X|\pi_X),$$

where nodes correspond to the terms $p(X|\pi_X)$.

Posterior factorization

Second, the nodes define the structure of the posterior approximation. The variational Bayesian approximation factorizes with respect to nodes, that is, each node corresponds to an independent probability distribution in the posterior approximation. In the previous example, `mu` and `tau` were separate nodes, thus the posterior approximation factorizes with respect to them: $q(\mu)q(\tau)$. Thus, the posterior approximation can be written as:

$$p(\tilde{\Omega}|\hat{\Omega}) \approx \prod_{X \in \tilde{\Omega}} q(X),$$

where $\tilde{\Omega}$ is the set of latent stochastic nodes and $\hat{\Omega}$ is the set of observed stochastic nodes. Sometimes one may want to avoid the factorization between some variables. For this purpose, there are some nodes which model several variables jointly without factorization. For instance, `GaussianGammaISO` is a joint node for μ and τ variables from the normal-gamma distribution and the posterior approximation does not factorize between μ and τ , that is, the posterior approximation is $q(\mu, \tau)$.

Functional form of the posterior

Last, the nodes define the functional form of the posterior approximation. Usually, the posterior approximation has the same or similar functional form as the prior. For instance, `Gamma` uses gamma distribution to also approximate the posterior distribution. Similarly, `GaussianARD` uses Gaussian distribution for the posterior. However, the posterior approximation of `GaussianARD` uses a full covariance matrix although the prior assumes a diagonal covariance matrix. Thus, there can be slight differences in the exact functional form of the posterior approximation but the rule of thumb is that the functional form of the posterior approximation is the same as or more general than the functional form of the prior.

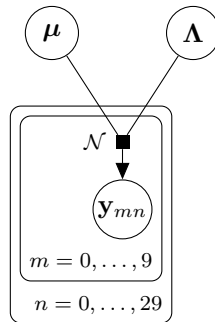
2.3.3 Using plate notation

Defining plates

Stochastic nodes take the optional parameter `plates`, which can be used to define plates of the variable. A plate defines the number of repetitions of a set of variables. For instance, a set of random variables y_{mn} could be defined as

$$y_{mn} \sim \mathcal{N}(\mu, \Lambda), \quad m = 0, \dots, 9, \quad n = 0, \dots, 29.$$

This can also be visualized as a graphical model:



The variable has two plates: one for the index m and one for the index n . In BayesPy, this random variable can be constructed as:

```
>>> y = Gaussian(mu, Lambda, plates=(10, 30))
```

Note: The plates are always given as a tuple of positive integers.

Plates also define indexing for the nodes, thus you can use simple NumPy-style slice indexing to obtain a subset of the plates:

```
>>> y_0 = y[0]
>>> y_0.plates
(30,)
>>> y_even = y[:, ::2]
>>> y_even.plates
(10, 15)
>>> y_complex = y[:, 5, 10:20:5]
>>> y_complex.plates
(5, 2)
```

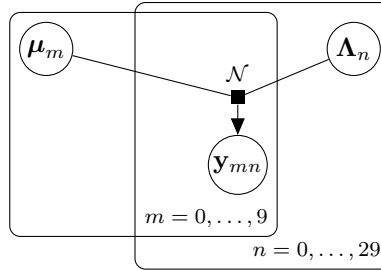
Note that this indexing is for the plates only, not for the random variable dimensions.

Sharing and broadcasting plates

Instead of having a common mean and precision matrix for all y_{mn} , it is also possible to share plates with parents. For instance, the mean could be different for each index m and the precision for each index n :

$$y_{mn} \sim \mathcal{N}(\mu_m, \Lambda_n), \quad m = 0, \dots, 9, \quad n = 0, \dots, 29.$$

which has the following graphical representation:



This can be constructed in BayesPy, for instance, as:

```
>>> from bayespy.nodes import Gaussian, Wishart
>>> mu = Gaussian([0, 0], [[1e-6, 0], [0, 1e-6]], plates=(10,1))
>>> Lambda = Wishart(2, [[1, 0], [0, 1]], plates=(1,30))
>>> X = Gaussian(mu, Lambda)
```

There are a few things to notice here. First, the plates are defined similarly as shapes in NumPy, that is, they use similar broadcasting rules. For instance, the plates (10, 1) and (1, 30) broadcast to (10, 30). In fact, one could use plates (10, 1) and (30,) to get the broadcasted plates (10, 30) because broadcasting compares the plates from right to left starting from the last axis. Second, X is not given plates keyword argument because the default plates are the plates broadcasted from the parents and that was what we wanted so it was not necessary to provide the keyword argument. If we wanted, for instance, plates (20, 10, 30) for X, then we would have needed to provide plates=(20, 10, 30).

The validity of the plates between a child and its parents is checked as follows. The plates are compared plate-wise starting from the last axis and working the way forward. A plate of the child is compatible with a plate of the parent if either of the following conditions is met:

1. The two plates have equal size
2. The parent has size 1 (or no plate)

Table below shows an example of compatible plates for a child node and its two parent nodes:

| node | plates | | | | | |
|---------|--------|---|---|---|---|---|
| parent1 | | 3 | 1 | 1 | 1 | 8 |
| parent2 | | | 1 | 1 | 5 | 1 |
| child | 5 | 3 | 1 | 7 | 5 | 8 |

Plates in deterministic nodes

Note that plates can be defined explicitly only for stochastic nodes. For deterministic nodes, the plates are defined implicitly by the plate broadcasting rules from the parents. Deterministic nodes do not need more plates than this because there is no randomness. The deterministic node would just have the same value over the extra plates, but it is not necessary to do this explicitly because the child nodes of the deterministic node can utilize broadcasting anyway. Thus, there is no point in having extra plates in deterministic nodes, and for this reason, deterministic nodes do not use plates keyword argument.

Plates in constants

It is useful to understand how the plates and the shape of a random variable are connected. The shape of an array which contains all the plates of a random variable is the concatenation of the plates and the shape of the variable. For instance, consider a 2-dimensional Gaussian variable with plates (3,). If you want the value of the constant mean vector and constant precision matrix to vary between plates, they are given as (3, 2)-shape and (3, 2, 2)-shape arrays, respectively:

```
>>> import numpy as np
>>> mu = [ [0,0], [1,1], [2,2] ]
>>> Lambda = [ [[1.0, 0.0],
...             [0.0, 1.0]],
...            [[1.0, 0.9],
...             [0.9, 1.0]],
...            [[1.0, -0.3],
...             [-0.3, 1.0]] ]
>>> X = Gaussian(mu, Lambda)
>>> np.shape(mu)
(3, 2)
>>> np.shape(Lambda)
(3, 2, 2)
>>> X.plates
(3,)
```

Thus, the leading axes of an array are the plate axes and the trailing axes are the random variable axes. In the example above, the mean vector has plates $(3,)$ and shape $(2,)$, and the precision matrix has plates $(3,)$ and shape $(2, 2)$.

Factorization of plates

It is important to understand the independency structure the plates induce for the model. First, the repetitions defined by a plate are independent a priori given the parents. Second, the repetitions are independent in the posterior approximation, that is, the posterior approximation factorizes with respect to plates. Thus, the plates also have an effect on the independence structure of the posterior approximation, not only prior. If dependencies between a set of variables need to be handled, that set must be handled as a some kind of multi-dimensional variable.

Irregular plates

The handling of plates is not always as simple as described above. There are cases in which the plates of the parents do not map directly to the plates of the child node. The user API should mention such irregularities.

For instance, the parents of a mixture distribution have a plate which contains the different parameters for each cluster, but the variable from the mixture distribution does not have that plate:

```
>>> from bayespy.nodes import Gaussian, Wishart, Categorical, Mixture
>>> mu = Gaussian([[0], [0], [0]], [ [[1]], [[1]], [[1]] ])
>>> Lambda = Wishart(1, [ [[1]], [[1]], [[1]] ])
>>> Z = Categorical([1/3, 1/3, 1/3], plates=(100,))
>>> X = Mixture(Z, Gaussian, mu, Lambda)
>>> mu.plates
(3,)
>>> Lambda.plates
(3,)
>>> Z.plates
(100,)
>>> X.plates
(100,)
```

The plates $(3,)$ and $(100,)$ should not broadcast according to the rules mentioned above. However, when validating the plates, `Mixture` removes the plate which corresponds to the clusters in `mu` and `Lambda`. Thus, `X` has plates which are the result of broadcasting plates $()$ and $(100,)$ which equals $(100,)$.

Also, sometimes the plates of the parents may be mapped to the variable axes. For instance, an automatic relevance determination (ARD) prior for a Gaussian variable is constructed by giving the diagonal elements of the precision

matrix (or tensor). The Gaussian variable itself can be a scalar, a vector, a matrix or a tensor. A set of five 4×3 -dimensional Gaussian matrices with ARD prior is constructed as:

```
>>> from bayespy.nodes import GaussianARD, Gamma
>>> tau = Gamma(1, 1, plates=(5,4,3))
>>> X = GaussianARD(0, tau, shape=(4,3))
>>> tau.plates
(5, 4, 3)
>>> X.plates
(5,)
```

Note how the last two plate axes of `tau` are mapped to the variable axes of `X` with shape $(4, 3)$ and the plates of `X` are obtained by taking the remaining leading plate axes of `tau`.

2.3.4 Example model: Principal component analysis

Now, we'll construct a bit more complex model which will be used in the following sections. The model is a probabilistic version of principal component analysis (PCA):

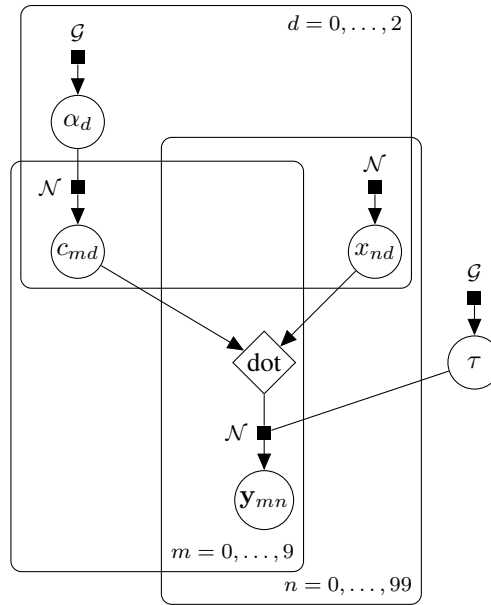
$$\mathbf{Y} = \mathbf{C}\mathbf{X}^T + \text{noise}$$

where \mathbf{Y} is $M \times N$ data matrix, \mathbf{C} is $M \times D$ loading matrix, \mathbf{X} is $N \times D$ state matrix, and noise is isotropic Gaussian. The dimensionality D is usually assumed to be much smaller than M and N .

A probabilistic formulation can be written as:

$$\begin{aligned} p(\mathbf{Y}) &= \prod_{m=0}^{M-1} \prod_{n=0}^{N-1} \mathcal{N}(y_{mn} | \mathbf{c}_m^T \mathbf{x}_n, \tau) \\ p(\mathbf{X}) &= \prod_{n=0}^{N-1} \prod_{d=0}^{D-1} \mathcal{N}(x_{nd} | 0, 1) \\ p(\mathbf{C}) &= \prod_{m=0}^{M-1} \prod_{d=0}^{D-1} \mathcal{N}(c_{md} | 0, \alpha_d) \\ p(\boldsymbol{\alpha}) &= \prod_{d=0}^{D-1} \mathcal{G}(\alpha_d | 10^{-3}, 10^{-3}) \\ p(\tau) &= \mathcal{G}(\tau | 10^{-3}, 10^{-3}) \end{aligned}$$

where we have given automatic relevance determination (ARD) prior for \mathbf{C} . This can be visualized as a graphical model:



Now, let us construct this model in BayesPy. First, we'll define the dimensionality of the latent space in our model:

```
>>> D = 3
```

Then the prior for the latent states **X**:

```
>>> X = GaussianARD(0, 1,
...                 shape=(D, ),
...                 plates=(1, 100),
...                 name='X')
```

Note that the shape of **X** is $(D,)$, although the latent dimensions are marked with a plate in the graphical model and they are conditionally independent in the prior. However, we want to (and need to) model the posterior dependency of the latent dimensions, thus we cannot factorize them, which would happen if we used `plates=(1, 100, D)` and `shape=()`. The first plate axis with size 1 is given just for clarity.

The prior for the ARD parameters α of the loading matrix:

```
>>> alpha = Gamma(1e-3, 1e-3,
...               plates=(D, ),
...               name='alpha')
```

The prior for the loading matrix **C**:

```
>>> C = GaussianARD(0, alpha,
...                 shape=(D, ),
...                 plates=(10, 1),
...                 name='C')
```

Again, note that the shape is the same as for **X** for the same reason. Also, the plates of **alpha**, $(D,)$, are mapped to the full shape of the node **C**, $(10, 1, D)$, using standard broadcasting rules.

The dot product is just a deterministic node:

```
>>> F = Dot(C, X)
```

However, note that `Dot` requires that the input Gaussian nodes have the same shape and that this shape has exactly one axis, that is, the variables are vectors. This the reason why we used `shape=(D,)` for **X** and **C** but from a bit different perspective. The node computes the inner product of D -dimensional vectors resulting in plates $(10, 100)$ broadcasted from the plates $(1, 100)$ and $(10, 1)$:

```
>>> F.plates
(10, 100)
```

The prior for the observation noise τ :

```
>>> tau = Gamma(1e-3, 1e-3, name='tau')
```

Finally, the observations are conditionally independent Gaussian scalars:

```
>>> Y = GaussianARD(F, tau, name='Y')
```

Now we have defined our model and the next step is to observe some data and to perform inference.

2.4 Performing inference

Approximation of the posterior distribution can be divided into several steps:

- Observe some nodes
- Choose the inference engine
- Initialize the posterior approximation
- Run the inference algorithm

In order to illustrate these steps, we'll be using the PCA model constructed in the previous section.

2.4.1 Observing nodes

First, let us generate some toy data:

```
>>> c = np.random.randn(10, 2)
>>> x = np.random.randn(2, 100)
>>> data = np.dot(c, x) + 0.1*np.random.randn(10, 100)
```

The data is provided by simply calling `observe` method of a stochastic node:

```
>>> Y.observe(data)
```

It is important that the shape of the `data` array matches the plates and shape of the node `Y`. For instance, if `Y` was `Wishart` node for 3×3 matrices with plates $(5, 1, 10)$, the full shape of `Y` would be $(5, 1, 10, 3, 3)$. The data array should have this shape exactly, that is, no broadcasting rules are applied.

Missing values

It is possible to mark missing values by providing a mask which is a boolean array:

```
>>> Y.observe(data, mask=[[True], [False], [False], [True], [True],
...                        [False], [True], [True], [True], [False]])
```

`True` means that the value is observed and `False` means that the value is missing. The shape of the above mask is $(10, 1)$, which broadcasts to the plates of `Y`, $(10, 100)$. Thus, the above mask means that the second, third, sixth and tenth rows of the 10×100 data matrix are missing.

The mask is applied to the *plates*, not to the data array directly. This means that it is not possible to observe a random variable partially, each repetition defined by the plates is either fully observed or fully missing. Thus, the mask is

applied to the plates. It is often possible to circumvent this seemingly tight restriction by adding an observable child node which factorizes more.

The shape of the mask is broadcasted to plates using standard NumPy broadcasting rules. So, if the variable has plates $(5, 1, 10)$, the mask could have a shape $(, (1,), (1, 1), (1, 1, 1), (10,), (1, 10), (1, 1, 10), (5, 1, 1))$ or $(5, 1, 10)$. In order to speed up the inference, missing values are automatically integrated out if they are not needed as latent variables to child nodes. This leads to faster convergence and more accurate approximations.

2.4.2 Choosing the inference method

Inference methods can be found in `bayespy.inference` package. Currently, only variational Bayesian approximation is implemented (`bayespy.inference.VB`). The inference engine is constructed by giving the stochastic nodes of the model.

```
>>> from bayespy.inference import VB
>>> Q = VB(Y, C, X, alpha, tau)
```

There is no need to give any deterministic nodes. Currently, the inference engine does not automatically search for stochastic parents and children, thus it is important that all stochastic nodes of the model are given. This should be made more robust in future versions.

A node of the model can be obtained by using the name of the node as a key:

```
>>> Q['X']
<bayespy.inference.vmp.nodes.gaussian.GaussianARD object at 0x...>
```

Note that the returned object is the same as the node object itself:

```
>>> Q['X'] is X
True
```

Thus, one may use the object `X` when it is available. However, if the model and the inference engine are constructed in another function or module, the node object may not be available directly and this feature becomes useful.

2.4.3 Initializing the posterior approximation

The inference engines give some initialization to the stochastic nodes by default. However, the inference algorithms can be sensitive to the initialization, thus it is sometimes necessary to have better control over the initialization. For VB, the following initialization methods are available:

- `initialize_from_prior`: Use the current states of the parent nodes to update the node. This is the default initialization.
- `initialize_from_parameters`: Use the given parameter values for the distribution.
- `initialize_from_value`: Use the given value for the variable.
- `initialize_from_random`: Draw a random value for the variable. The random sample is drawn from the current state of the node's distribution.

Note that `initialize_from_value` and `initialize_from_random` initialize the distribution with a value of the variable instead of parameters of the distribution. Thus, the distribution is actually a delta distribution with a peak on the value after the initialization. This state of the distribution does not have proper natural parameter values nor normalization, thus the VB lower bound terms are `np.nan` for this initial state.

These initialization methods can be used to perform even a bit more complex initializations. For instance, a Gaussian distribution could be initialized with a random mean and variance 0.1. In our PCA model, this can be obtained by


```
>>> X.initialize_from_parameters(np.random.randn(1, 100, D), 10)
```

Note that the shape of the random mean is the sum of the plates (1, 100) and the variable shape (D,). In addition, instead of variance, GaussianARD uses precision as the second parameter, thus we initialized the variance to $\frac{1}{10}$. This random initialization is important in our PCA model because the default initialization gives C and X zero mean. If the mean of the other variable was zero when the other is updated, the other variable gets zero mean too. This would lead to an update algorithm where both means remain zeros and effectively no latent space is found. Thus, it is important to give non-zero random initialization for X if C is updated before X the first time. It is typical that at least some nodes need be initialized with some randomness.

By default, nodes are initialized with the method `initialize_from_prior`. The method is not very time consuming but if for any reason you want to avoid that default initialization computation, you can provide `initialize=False` when creating the stochastic node. However, the node does not have a proper state in that case, which leads to errors in VB learning unless the distribution is initialized using the above methods.

2.4.4 Running the inference algorithm

The approximation methods are based on iterative algorithms, which can be run using `update` method. By default, it takes one iteration step updating all nodes once:

```
>>> Q.update()
Iteration 1: loglike=-9.305259e+02 (... seconds)
```

The `loglike` tells the VB lower bound. The order in which the nodes are updated is the same as the order in which the nodes were given when creating Q. If you want to change the order or update only some of the nodes, you can give as arguments the nodes you want to update and they are updated in the given order:

```
>>> Q.update(C, X)
Iteration 2: loglike=-8.818976e+02 (... seconds)
```

It is also possible to give the same node several times:

```
>>> Q.update(C, X, C, tau)
Iteration 3: loglike=-8.071222e+02 (... seconds)
```

Note that each call to `update` is counted as one iteration step although not variables are necessarily updated. Instead of doing one iteration step, `repeat` keyword argument can be used to perform several iteration steps:

```
>>> Q.update(repeat=10)
Iteration 4: loglike=-7.167588e+02 (... seconds)
Iteration 5: loglike=-6.827873e+02 (... seconds)
Iteration 6: loglike=-6.259477e+02 (... seconds)
Iteration 7: loglike=-4.725400e+02 (... seconds)
Iteration 8: loglike=-3.270816e+02 (... seconds)
Iteration 9: loglike=-2.208865e+02 (... seconds)
Iteration 10: loglike=-1.658761e+02 (... seconds)
Iteration 11: loglike=-1.469468e+02 (... seconds)
Iteration 12: loglike=-1.420311e+02 (... seconds)
Iteration 13: loglike=-1.405139e+02 (... seconds)
```

The VB algorithm stops automatically if it converges, that is, the relative change in the lower bound is below some threshold:

```
>>> Q.update(repeat=1000)
Iteration 14: loglike=-1.396481e+02 (... seconds)
...
Iteration 488: loglike=-1.224106e+02 (... seconds)
Converged at iteration 488.
```

Now the algorithm stopped before taking 1000 iteration steps because it converged. The relative tolerance can be adjusted by providing `tol` keyword argument to the `update` method:

```
>>> Q.update(repeat=10000, tol=1e-6)
Iteration 489: loglike=-1.224094e+02 (... seconds)
...
Iteration 847: loglike=-1.222506e+02 (... seconds)
Converged at iteration 847.
```

Making the tolerance smaller, may improve the result but it may also significantly increase the iteration steps until convergence.

Instead of using `update` method of the inference engine VB, it is possible to use the `update` methods of the nodes directly as

```
>>> C.update()
```

or

```
>>> Q['C'].update()
```

However, this is not recommended, because the `update` method of the inference engine VB is a wrapper which, in addition to calling the nodes' `update` methods, checks for convergence and does a few other useful minor things. But if for any reason these direct update methods are needed, they can be used.

Parameter expansion

Sometimes the VB algorithm converges very slowly. This may happen when the variables are strongly coupled in the true posterior but factorized in the approximate posterior. This coupling leads to zigzagging of the variational parameters which progresses slowly. One solution to this problem is to use parameter expansion. The idea is to add an auxiliary variable which parameterizes the posterior approximation of several variables. Then optimizing this auxiliary variable actually optimizes several posterior approximations jointly leading to faster convergence.

The parameter expansion is model specific. Currently in BayesPy, only state-space models have built-in parameter expansions available. These state-space models contain a variable which is a dot product of two variables (plus some noise):

$$y = \mathbf{c}^T \mathbf{x} + \text{noise}$$

The parameter expansion can be motivated by noticing that we can add an auxiliary variable which rotates the variables \mathbf{c} and \mathbf{x} so that the dot product is unaffected:

$$y = \mathbf{c}^T \mathbf{x} + \text{noise} = \mathbf{c}^T \mathbf{R} \mathbf{R}^{-1} \mathbf{x} + \text{noise} = (\mathbf{R}^T \mathbf{c})^T (\mathbf{R}^{-1} \mathbf{x}) + \text{noise}$$

Now, applying this rotation to the posterior approximations $q(\mathbf{c})$ and $q(\mathbf{x})$, and optimizing the VB lower bound with respect to the rotation leads to parameterized joint optimization of \mathbf{c} and \mathbf{x} .

The available parameter expansion methods are in module `transformations`:

```
>>> from bayespy.inference.vmp import transformations
```

First, you create the rotation transformations for the two variables:

```
>>> rotX = transformations.RotateGaussianARD(X)
>>> rotC = transformations.RotateGaussianARD(C, alpha)
```

Here, the rotation for \mathbf{C} provides the ARD parameters `alpha` so they are updated simultaneously. In addition to `RotateGaussianARD`, there are a few other built-in rotations defined, for instance, `RotateGaussian` and `RotateGaussianMarkovChain`. It is extremely important that the model satisfies the assumptions made by

the rotation class and the user is mostly responsible for this. The optimizer for the rotations is constructed by giving the two rotations and the dimensionality of the rotated space:

```
>>> R = transformations.RotationOptimizer(rotC, rotX, D)
```

Now, calling `rotate` method will find optimal rotation and update the relevant nodes (X, C and alpha) accordingly:

```
>>> R.rotate()
```

Let us see how our iteration would have gone if we had used this parameter expansion. First, let us re-initialize our nodes and VB algorithm:

```
>>> alpha.initialize_from_prior()
>>> C.initialize_from_prior()
>>> X.initialize_from_parameters(np.random.randn(1, 100, D), 10)
>>> tau.initialize_from_prior()
>>> Q = VB(Y, C, X, alpha, tau)
```

Then, the rotation is set to run after each iteration step:

```
>>> Q.callback = R.rotate
```

Now the iteration converges to the relative tolerance 10^{-6} much faster:

```
>>> Q.update(repeat=1000, tol=1e-6)
Iteration 1: loglike=-9.363500e+02 (... seconds)
...
Iteration 18: loglike=-1.221354e+02 (... seconds)
Converged at iteration 18.
```

The convergence took 18 iterations with rotations and 488 or 847 iterations without the parameter expansion. In addition, the lower bound is improved slightly. One can compare the number of iteration steps in this case because the cost per iteration step with or without parameter expansion is approximately the same. Sometimes the parameter expansion can have the drawback that it converges to a bad local optimum. Usually, this can be solved by updating the nodes near the observations a few times before starting to update the hyperparameters and to use parameter expansion. In any case, the parameter expansion is practically necessary when using state-space models in order to converge to a proper solution in a reasonable time.

2.5 Examining the results

After the results have been obtained, it is important to be able to examine the results easily. The results can be examined either numerically by inspecting numerical arrays or visually by plotting distributions of the nodes. In addition, the posterior distributions can be visualized during the learning algorithm and the results can be saved into a file.

2.5.1 Plotting the results

The module `plot` offers some plotting basic functionality:

```
>>> import bayespy.plot as bpplt
```

The module contains `matplotlib.pyplot` module if the user needs that. For instance, interactive plotting can be enabled as:

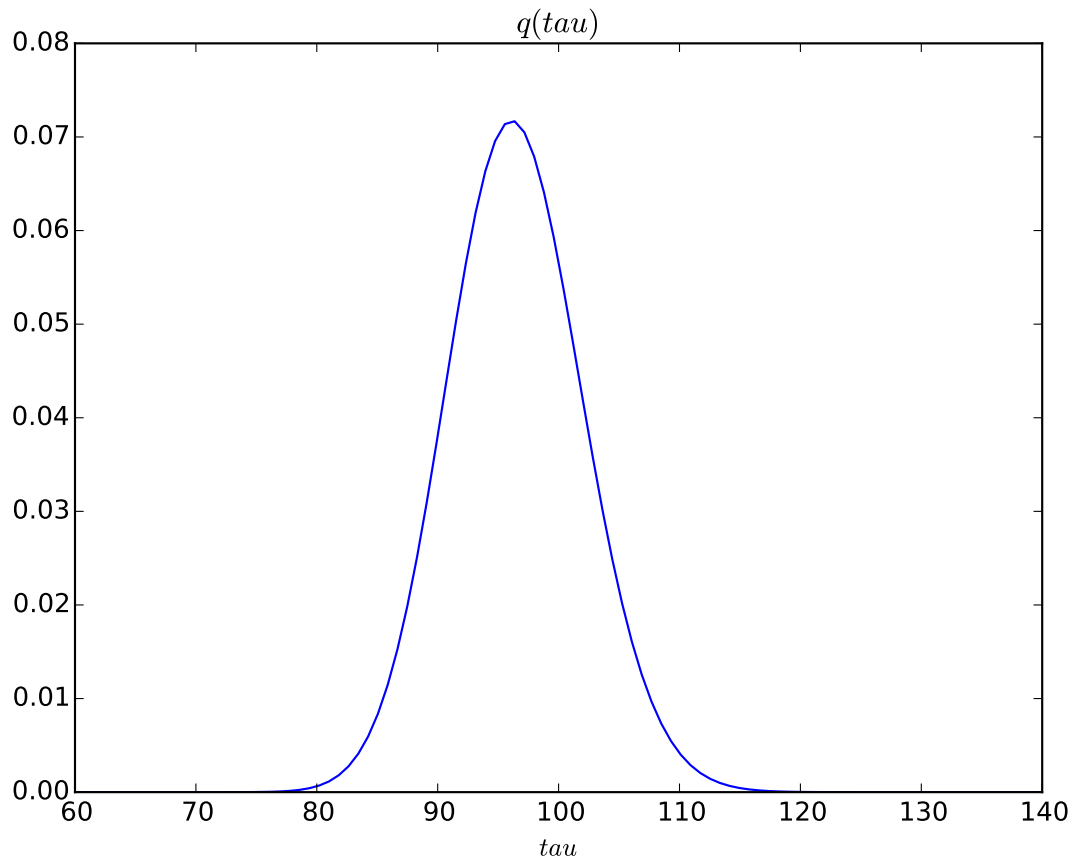
```
>>> bpplt.pyplot.ion()
```

The `plot` module contains some functions but it is not a very comprehensive collection, thus the user may need to write some problem- or model-specific plotting functions. The current collection is:

- `pdf()`: show probability density function of a scalar
- `contour()`: show probability density function of two-element vector
- `hinton()`: show the Hinton diagram
- `plot()`: show value as a function

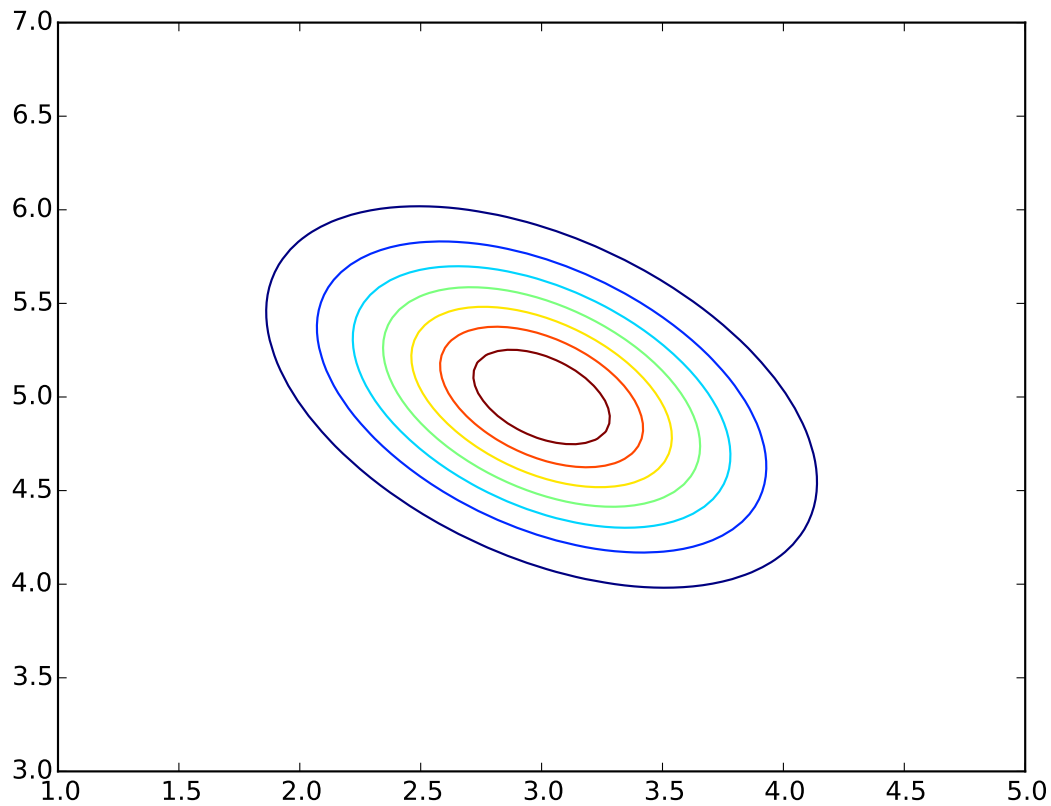
The probability density function of a scalar random variable can be plotted using the function `pdf()`:

```
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.pdf(Q['tau'], np.linspace(60, 140, num=100))
[<matplotlib.lines.Line2D object at 0x...>]
```



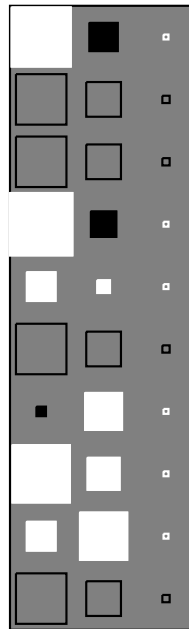
The variable `tau` models the inverse variance of the noise, for which the true value is $0.1^{-2} = 100$. Thus, the posterior captures the true value quite accurately. Similarly, the function `contour()` can be used to plot the probability density function of a 2-dimensional variable, for instance:

```
>>> V = Gaussian([3, 5], [[4, 2], [2, 5]])
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.contour(V, np.linspace(1, 5, num=100), np.linspace(3, 7, num=100))
<matplotlib.contour.QuadContourSet object at 0x...>
```



Both `pdf()` and `contour()` require that the user provides the grid on which the probability density function is computed. They also support several keyword arguments for modifying the output, similarly as `plot` and `contour` in `matplotlib.pyplot`. These functions can be used only for stochastic nodes. A few other plot types are also available as built-in functions. A Hinton diagram can be plotted as:

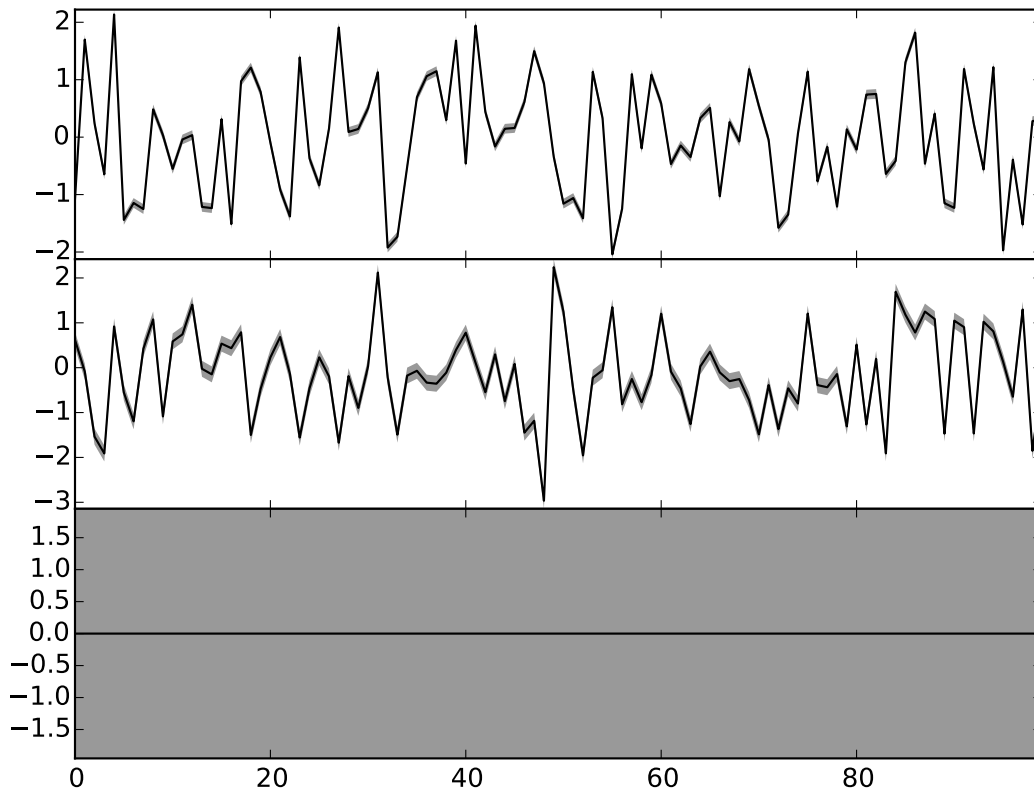
```
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.hinton(C)
```



The diagram shows the elements of the matrix C . The size of the filled rectangle corresponds to the absolute value of the element mean, and white and black correspond to positive and negative values, respectively. The non-filled rectangle shows standard deviation. From this diagram it is clear that the third column of C has been pruned out and the rows that were missing in the data have zero mean and column-specific variance. The function `hinton()` is a simple wrapper for node-specific Hinton diagram plotters, such as `gaussian_hinton()` and `dirichlet_hinton()`. Thus, the keyword arguments depend on the node which is plotted.

Another plotting function is `plot()`, which just plots the values of the node over one axis as a function:

```
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.plot(X, axis=-2)
```



Now, the `axis` is the second last axis which corresponds to $n = 0, \dots, N-1$. As $D = 3$, there are three subplots. For Gaussian variables, the function shows the mean and two standard deviations. The plot shows that the third component has been pruned out, thus the method has been able to recover the true dimensionality of the latent space. It also has similar keyword arguments to `plot` function in `matplotlib.pyplot`. Again, `plot()` is a simple wrapper over node-specific plotting functions, thus it supports only some node classes.

2.5.2 Monitoring during the inference algorithm

It is possible to plot the distribution of the nodes during the learning algorithm. This is useful when the user is interested to see how the distributions evolve during learning and what is happening to the distributions. In order to utilize monitoring, the user must set plotters for the nodes that he or she wishes to monitor. This can be done either when creating the node or later at any time.

The plotters are set by creating a plotter object and providing this object to the node. The plotter is a wrapper of one of the plotting functions mentioned above: `PDFPlotter`, `ContourPlotter`, `HintonPlotter` or `FunctionPlotter`. Thus, our example model could use the following plotters:

```
>>> tau.set_plotter(bpplt.PDFPlotter(np.linspace(60, 140, num=100)))
>>> C.set_plotter(bpplt.HintonPlotter())
>>> X.set_plotter(bpplt.FunctionPlotter(axis=-2))
```

These could have been given at node creation as a keyword argument `plotter`:

```
>>> V = Gaussian([3, 5], [[4, 2], [2, 5]],
...               plotter=bpplt.ContourPlotter(np.linspace(1, 5, num=100),
...               np.linspace(3, 7, num=100)))
```

When the plotter is set, one can use the `plot` method of the node to perform plotting:

```
>>> V.plot()
<matplotlib.contour.QuadContourSet object at 0x...>
```

Nodes can also be plotted using the `plot` method of the inference engine:

```
>>> Q.plot('C')
```

This method remembers the figure in which a node has been plotted and uses that every time it plots the same node. In order to monitor the nodes during learning, it is possible to use the keyword argument `plot`:

```
>>> Q.update(repeat=5, plot=True, tol=np.nan)
Iteration 19: loglike=-1.221354e+02 (... seconds)
Iteration 20: loglike=-1.221354e+02 (... seconds)
Iteration 21: loglike=-1.221354e+02 (... seconds)
Iteration 22: loglike=-1.221354e+02 (... seconds)
Iteration 23: loglike=-1.221354e+02 (... seconds)
```

Each node which has a plotter set will be plotted after it is updated. Note that this may slow down the inference significantly if the plotting operation is time consuming.

2.5.3 Posterior parameters and moments

If the built-in plotting functions are not sufficient, it is possible to use `matplotlib.pyplot` for custom plotting. Each node has `get_moments` method which returns the moments and they can be used for plotting. Stochastic exponential family nodes have natural parameter vectors which can also be used. In addition to plotting, it is also possible to just print the moments or parameters in the console.

2.5.4 Saving and loading results

The results of the inference engine can be easily saved and loaded using `VB.save()` and `VB.load()` methods:

```
>>> Q.save(filename='tmp.hdf5')
>>> Q.load(filename='tmp.hdf5')
```

The results are stored in a HDF5 file. The user may set an autosave file in which the results are automatically saved regularly. Autosave filename can be set at creation time by `autosave_filename` keyword argument or later using `VB.set_autosave()` method. If autosave file has been set, the `VB.save()` and `VB.load()` methods use that file by default. In order for the saving to work, all stochastic nodes must have been given (unique) names.

However, note that these methods do *not* save nor load the node definitions. It means that the user must create the nodes and the inference engine and then use `VB.load()` to set the state of the nodes and the inference engine. If there are any differences in the model that was saved and the one which is tried to update using loading, then loading does not work. Thus, the user should keep the model construction unmodified in a Python file in order to be able to load the results later. Or if the user wishes to share the results, he or she must share the model construction Python file with the HDF5 results file.

2.6 Advanced topics

This section contains brief information on how to implement some advanced methods in BayesPy. These methods include Riemannian conjugate gradient methods, pattern search, simulated annealing, collapsed variational inference and stochastic variational inference. In order to use these methods properly, the user should understand them to some extent. They are also considered experimental, thus you may encounter bugs or unimplemented features. In any case, these methods may provide huge performance improvements easily compared to the standard VB-EM algorithm.

2.6.1 Gradient-based optimization

Variational Bayesian learning basically means that the parameters of the approximate posterior distributions are optimized to maximize the lower bound of the marginal log likelihood [3]. This optimization can be done by using gradient-based optimization methods. In order to improve the gradient-based methods, it is recommended to take into account the information geometry by using the Riemannian (a.k.a. natural) gradient. In fact, the standard VB-EM algorithm is equivalent to a gradient ascent method which uses the Riemannian gradient and step length 1. Thus, it is natural to try to improve this method by using non-linear conjugate gradient methods instead of gradient ascent. These optimization methods are especially useful when the VB-EM update equations are not available but one has to use fixed form approximation. But it is possible that the Riemannian conjugate gradient method improve performance even when the VB-EM update equations are available.

The optimization algorithm in `VB.optimize()` has a simple interface. Instead of using the default Riemannian geometry, one can use the Euclidean geometry by giving `riemannian=False`. It is also possible to choose the optimization method from gradient ascent (`method='gradient'`) or conjugate gradient methods (only `method='fletcher-reeves'` implemented at the moment). For instance, we could optimize nodes C and X jointly using Euclidean gradient ascent as:

```
>>> Q = VB(Y, C, X, alpha, tau)
>>> Q.optimize(C, X, riemannian=False, method='gradient', maxiter=5)
Iteration ...
```

Note that this is very inefficient way of updating those nodes (bad geometry and not using conjugate gradients). Thus, one should understand the idea of these optimization methods, otherwise one may do something extremely inefficient. Most likely this method can be found useful in combination with the advanced tricks in the following sections.

Note: The Euclidean gradient has not been implemented for all nodes yet. The Euclidean gradient is required by the Euclidean geometry based optimization but also by the conjugate gradient methods in the Riemannian geometry. Thus, the Riemannian conjugate gradient may not yet work for all models.

It is possible to construct custom optimization algorithms with the tools provided by VB. For instance, `VB.get_parameters()` and `VB.set_parameters()` can be used to handle the parameters of nodes. `VB.get_gradients()` is used for computing the gradients of nodes. The parameter and gradient objects are not numerical arrays but more complex nested lists not meant to be accessed by the user. Thus, for simple arithmetics with the parameter and gradient objects, use functions `VB.add()` and `VB.dot()`. Finally, `VB.compute_lowerbound()` and `VB.has_converged()` can be used to monitor the lower bound.

2.6.2 Collapsed inference

The optimization method can be used efficiently in such a way that some of the variables are collapsed, that is, marginalized out [1]. The collapsed variables must be conditionally independent given the observations and all other variables. Probably, one also wants that the size of the marginalized variables is large and the size of the optimized variables is small. For instance, in our PCA example, we could optimize as follows:

```
>>> Q.optimize(C, tau, maxiter=10, collapsed=[X, alpha])
Iteration ...
```

The collapsed variables are given as a list. This optimization does basically the following: It first computes the gradients for `C` and `tau` and takes an update step using the desired optimization method. Then, it updates the collapsed variables by using the standard VB-EM update equations. These two steps are taken in turns. Effectively, this corresponds to collapsing the variables `X` and `alpha` in a particular way. The point of this method is that the number of parameters in the optimization reduces significantly and the collapsed variables are updated optimally. For more details, see [1].

It is possible to use this method in such a way, that the collapsed variables are not conditionally independent given the observations and all other variables. However, in that case, the method does not anymore correspond to collapsing the variables but just using VB-EM updates after gradient-based updates. The method does not check for conditional independence, so the user is free to do this.

Note: Although the Riemannian conjugate gradient method has not yet been implemented for all nodes, it may be possible to collapse those nodes and optimize the other nodes for which the Euclidean gradient is already implemented.

2.6.3 Pattern search

The pattern search method estimates the direction in which the approximate posterior distributions are updating and performs a line search in that direction [4]. The search direction is based on the difference in the VB parameters on successive updates (or several updates). The idea is that the VB-EM algorithm may be slow because it just zigzags and this can be fixed by moving to the direction in which the VB-EM is slowly moving.

BayesPy offers a simple built-in pattern search method `VB.pattern_search()`. The method updates the nodes twice, measures the difference in the parameters and performs a line search with a small number of function evaluations:

```
>>> Q.pattern_search(C, X)
Iteration ...
```

Similarly to the collapsed optimization, it is possible to collapse some of the variables in the pattern search. The same rules of conditional independence apply as above. The collapsed variables are given as list:

```
>>> Q.pattern_search(C, tau, collapsed=[X, alpha])
Iteration ...
```

Also, a maximum number of iterations can be set by using `maxiter` keyword argument. It is not always obvious whether a pattern search will improve the rate of convergence or not but if it seems that the convergence is slow because of zigzagging, it may be worth a try. Note that the computational cost of the pattern search is quite high, thus it is not recommended to perform it after every VB-EM update but every now and then, for instance, after every 10 iterations. In addition, it is possible to write a more customized VB learning algorithm which uses pattern searches by using the different methods of VB discussed above.

2.6.4 Deterministic annealing

The standard VB-EM algorithm converges to a local optimum which can often be inferior to the global optimum and many other local optima. Deterministic annealing aims at finding a better local optimum, hopefully even the global optimum [5]. It does this by increasing the weight on the entropy of the posterior approximation in the VB lower bound. Effectively, the annealed lower bound becomes closer to a uniform function instead of the original multimodal lower bound. The weight on the entropy is recovered slowly and the optimization is much more robust to initialization.

In BayesPy, the annealing can be set by using `VB.set_annealing()`. The given annealing should be in range $(0, 1]$ but this is not validated in case the user wants to do something experimental. If annealing is set to 1, the original

VB lower bound is recovered. Annealing with 0 would lead to an improper uniform distribution, thus it will lead to errors. The entropy term is weighted by the inverse of this annealing term. An alternative view is that the model probability density functions are raised to the power of the annealing term.

Typically, the annealing is used in such a way that the annealing is small at the beginning and increased after every convergence of the VB algorithm until value 1 is reached. After the annealing value is increased, the algorithm continues from where it had just converged. The annealing can be used for instance as:

```
>>> beta = 0.1
>>> while beta < 1.0:
...     beta = min(beta*1.5, 1.0)
...     Q.set_annealing(beta)
...     Q.update(repeat=100, tol=1e-4)
Iteration ...
```

Here, the `tol` keyword argument is used to adjust the threshold for convergence. In this case, it is a bit larger than by default so the algorithm does not need to converge perfectly but a rougher convergence is sufficient for the next iteration with a new annealing value.

2.6.5 Stochastic variational inference

In stochastic variational inference [2], the idea is to use mini-batches of large datasets to compute noisy gradients and learn the VB distributions by using stochastic gradient ascent. In order for it to be useful, the model must be such that it can be divided into “intermediate” and “global” variables. The number of intermediate variables increases with the data but the number of global variables remains fixed. The global variables are learnt in the stochastic optimization.

By denoting the data as $Y = [Y_1, \dots, Y_N]$, the intermediate variables as $Z = [Z_1, \dots, Z_N]$ and the global variables as θ , the model needs to have the following structure:

$$p(Y, Z, \theta) = p(\theta) \prod_{n=1}^N p(Y_n | Z_n, \theta) p(Z_n | \theta)$$

The algorithm consists of three steps which are iterated: 1) a random mini-batch of the data is selected, 2) the corresponding intermediate variables are updated by using normal VB update equations, and 3) the global variables are updated with (stochastic) gradient ascent as if there was as many replications of the mini-batch as needed to recover the original dataset size.

The learning rate for the gradient ascent must satisfy:

$$\sum_{i=1}^{\infty} \alpha_i = \infty \quad \text{and} \quad \sum_{i=1}^{\infty} \alpha_i^2 < \infty,$$

where i is the iteration number. An example of a valid learning parameter is $\alpha_i = (\delta + i)^{-\gamma}$, where $\delta \geq 0$ is a delay and $\gamma \in (0.5, 1]$ is a forgetting rate.

Stochastic variational inference is relatively easy to use in BayesPy. The idea is that the user creates a model for the size of a mini-batch and specifies a multiplier for those plate axes that are replicated. For the PCA example, the mini-batch model can be constructed as follows. We decide to use `X` as an intermediate variable and the other variables are global. The global variables `alpha`, `C` and `tau` are constructed identically as before. The intermediate variable `X` is constructed as:

```
>>> X = GaussianARD(0, 1,
...                 shape=(D, ),
...                 plates=(1, 5),
...                 plates_multiplier=(1, 20),
...                 name='X')
```

Note that the plates are $(1, 5)$ whereas they are $(1, 100)$ in the full model. Thus, we need to provide a plates multiplier $(1, 20)$ to define how the plates are replicated to get the full dataset. These multipliers do not need to be integers, in this case the latter plate axis is multiplied by $100/5 = 20$. The remaining variables are defined as before:

```
>>> F = Dot(C, X)
>>> Y = GaussianARD(F, tau, name='Y')
```

Note that the plates of Y and F also correspond to the size of the mini-batch and they also deduce the plate multipliers from their parents, thus we do not need to specify the multiplier here explicitly (although it is ok to do so).

Let us construct the inference engine for the new mini-batch model:

```
>>> Q = VB(Y, C, X, alpha, tau)
```

Use random initialization for C to break the symmetry in C and X :

```
>>> C.initialize_from_random()
```

Then, stochastic variational inference algorithm could look as follows:

```
>>> Q.ignore_bound_checks = True
>>> for n in range(200):
...     subset = np.random.choice(100, 5)
...     Y.observe(data[:, subset])
...     Q.update(X)
...     learning_rate = (n + 2.0) ** (-0.7)
...     Q.gradient_step(C, alpha, tau, scale=learning_rate)
Iteration ...
```

First, we ignore the bound checks because they are noisy. Then, the loop consists of three parts: 1) Draw a random mini-batch of the data (5 samples from 100). 2) Update the intermediate variable X . 3) Update global variables with gradient ascent using a proper learning rate.

2.6.6 Black-box variational inference

NOT YET IMPLEMENTED.

EXAMPLES

3.1 Linear regression

3.1.1 Data

The true parameters of the linear regression:

```
>>> k = 2 # slope
>>> c = 5 # bias
>>> s = 2 # noise standard deviation
```

Generate data:

```
>>> import numpy as np
>>> x = np.arange(10)
>>> y = k*x + c + s*np.random.randn(10)
```

3.1.2 Model

The regressors, that is, the input data:

```
>>> X = np.vstack([x, np.ones(len(x))]).T
```

Note that we added a column of ones to the regressor matrix for the bias term. We model the slope and the bias term in the same node so we do not factorize between them:

```
>>> from bayespy.nodes import GaussianARD
>>> B = GaussianARD(0, 1e-6, shape=(2,))
```

The first element is the slope which multiplies x and the second element is the bias term which multiplies the constant ones. Now we compute the dot product of X and B :

```
>>> from bayespy.nodes import SumMultiply
>>> F = SumMultiply('i,i', B, X)
```

The noise parameter:

```
>>> from bayespy.nodes import Gamma
>>> tau = Gamma(1e-3, 1e-3)
```

The noisy observations:

```
>>> Y = GaussianARD(F, tau)
```

3.1.3 Inference

Observe the data:

```
>>> Y.observe(y)
```

Construct the variational Bayesian (VB) inference engine by giving all stochastic nodes:

```
>>> from bayespy.inference import VB
>>> Q = VB(Y, B, tau)
```

Iterate until convergence:

```
>>> Q.update(repeat=1000)
Iteration 1: loglike=-4.595948e+01 (... seconds)
...
Iteration 5: loglike=-4.495017e+01 (... seconds)
Converged at iteration 5.
```

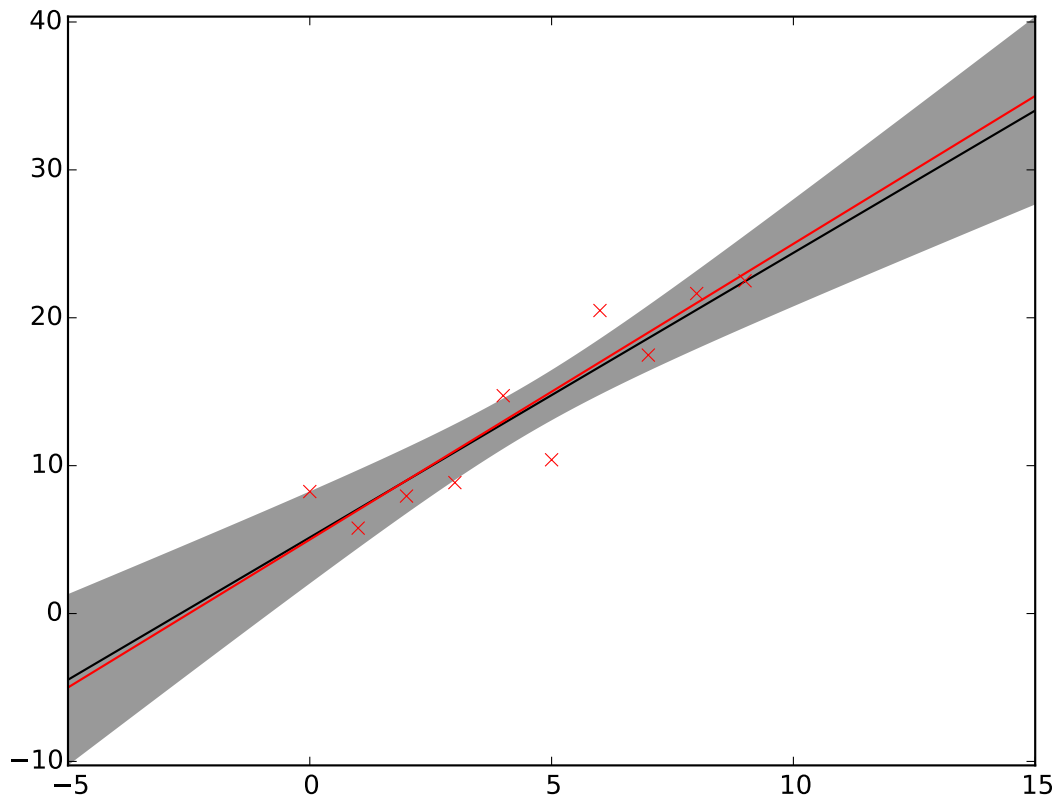
3.1.4 Results

Create a simple predictive model for new inputs:

```
>>> xh = np.linspace(-5, 15, 100)
>>> Xh = np.vstack([xh, np.ones(len(xh))]).T
>>> Fh = SumMultiply('i,i', B, Xh)
```

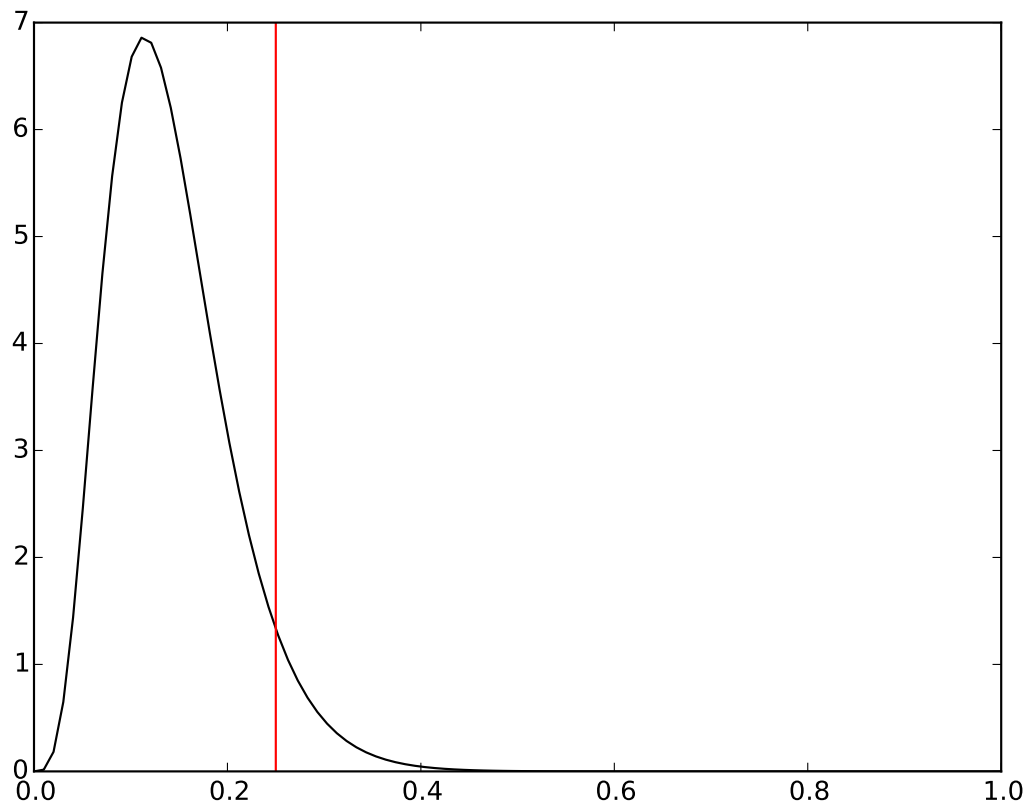
Note that we use the learned node B but create a new regressor array for predictions. Plot the predictive distribution of noiseless function values:

```
>>> import bayespy.plot as bpplt
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.plot(Fh, x=xh, scale=2)
>>> bpplt.plot(y, x=x, color='r', marker='x', linestyle='None')
>>> bpplt.plot(k*xh+c, x=xh, color='r');
```



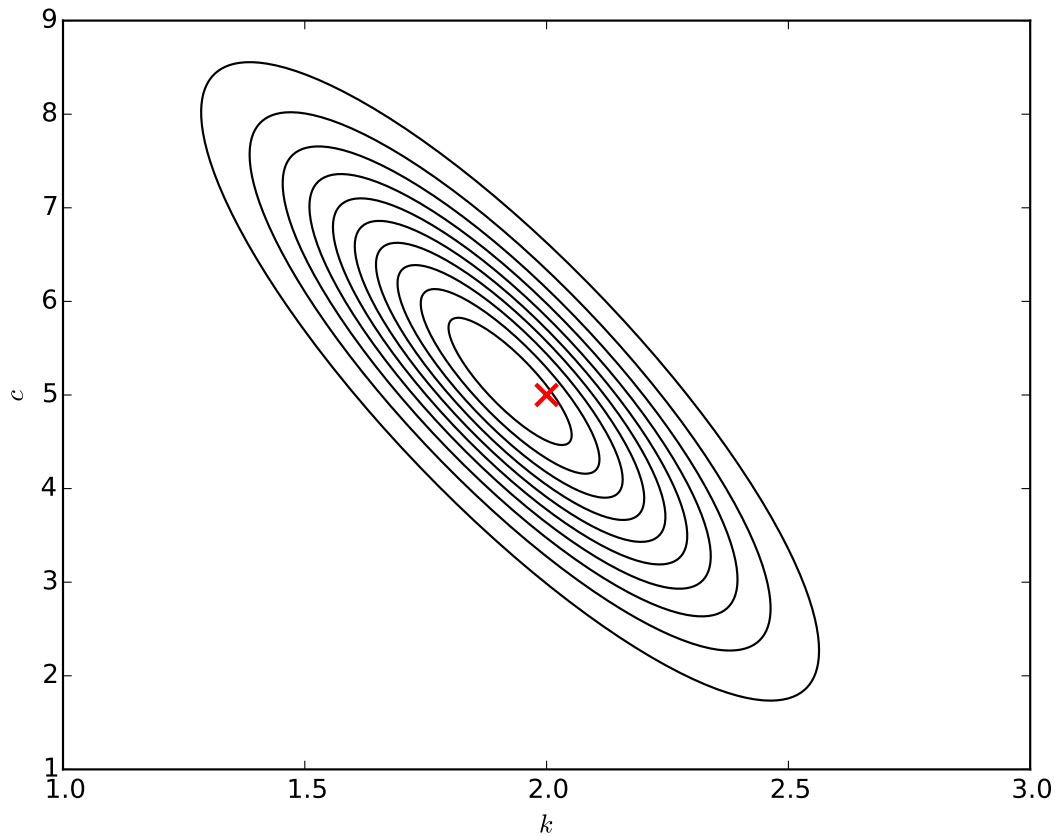
Note that the above plot shows two standard deviation of the posterior of the noiseless function, thus the data points may lie well outside this range. The red line shows the true linear function. Next, plot the distribution of the noise parameter and the true value, $2^{-2} = 0.25$:

```
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.pdf(tau, np.linspace(1e-6,1,100), color='k')
[<matplotlib.lines.Line2D object at 0x...>]
>>> bpplt.pyplot.axvline(s**(-2), color='r')
<matplotlib.lines.Line2D object at 0x...>
```



The noise level is captured quite well, although the posterior has more mass on larger noise levels (smaller precision parameter values). Finally, plot the distribution of the regression parameters and mark the true value:

```
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.contour(B, np.linspace(1,3,1000), np.linspace(1,9,1000),
...               n=10, colors='k')
<matplotlib.contour.QuadContourSet object at 0x...>
>>> bpplt.plot(c, x=k, color='r', marker='x', linestyle='None',
...           markersize=10, markeredgewidth=2)
>>> bpplt.pyplot.xlabel(r'$k$')
<matplotlib.text.Text object at 0x...>
>>> bpplt.pyplot.ylabel(r'$c$');
<matplotlib.text.Text object at 0x...>
```

In this case, the true parameters are captured well by the posterior distribution.

3.1.5 Improving accuracy

The model can be improved by not factorizing between B and τ but learning their joint posterior distribution. This requires a slight modification to the model by using `GaussianGammaISO` node:

```
>>> from bayespy.nodes import GaussianGammaISO
>>> B_tau = GaussianGammaISO(np.zeros(2), 1e-6*np.identity(2), 1e-3, 1e-3)
```

This node contains both the regression parameter vector and the noise parameter. We compute the dot product similarly as before:

```
>>> F_tau = SumMultiply('i,i', B_tau, X)
```

However, Y is constructed as follows:

```
>>> Y = GaussianARD(F_tau, 1)
```

Because the noise parameter is already in F_{τ} we can give a constant one as the second argument. The total noise parameter for Y is the product of the noise parameter in F_{τ} and one. Now, inference is run similarly as before:

```
>>> Y.observe(y)
>>> Q = VB(Y, B_tau)
>>> Q.update(repeat=1000)
```

```
Iteration 1: loglike=-4.678478e+01 (... seconds)
Iteration 2: loglike=-4.678478e+01 (... seconds)
Converged at iteration 2.
```

Note that the method converges immediately. This happens because there is only one unobserved stochastic node so there is no need for iteration and the result is actually the exact true posterior distribution, not an approximation. Currently, the main drawback of using this approach is that BayesPy does not yet contain any plotting utilities for nodes that contain both Gaussian and gamma variables jointly.

3.1.6 Further extensions

The approach discussed in this example can easily be extended to non-linear regression and multivariate regression. For non-linear regression, the inputs are first transformed by some known non-linear functions and then linear regression is applied to this transformed data. For multivariate regression, X and B are concatenated appropriately: If there are more regressors, add more columns to both X and B . If there are more output dimensions, add plates to B .

3.2 Gaussian mixture model

This example demonstrates the use of Gaussian mixture model for flexible density estimation, clustering or classification.

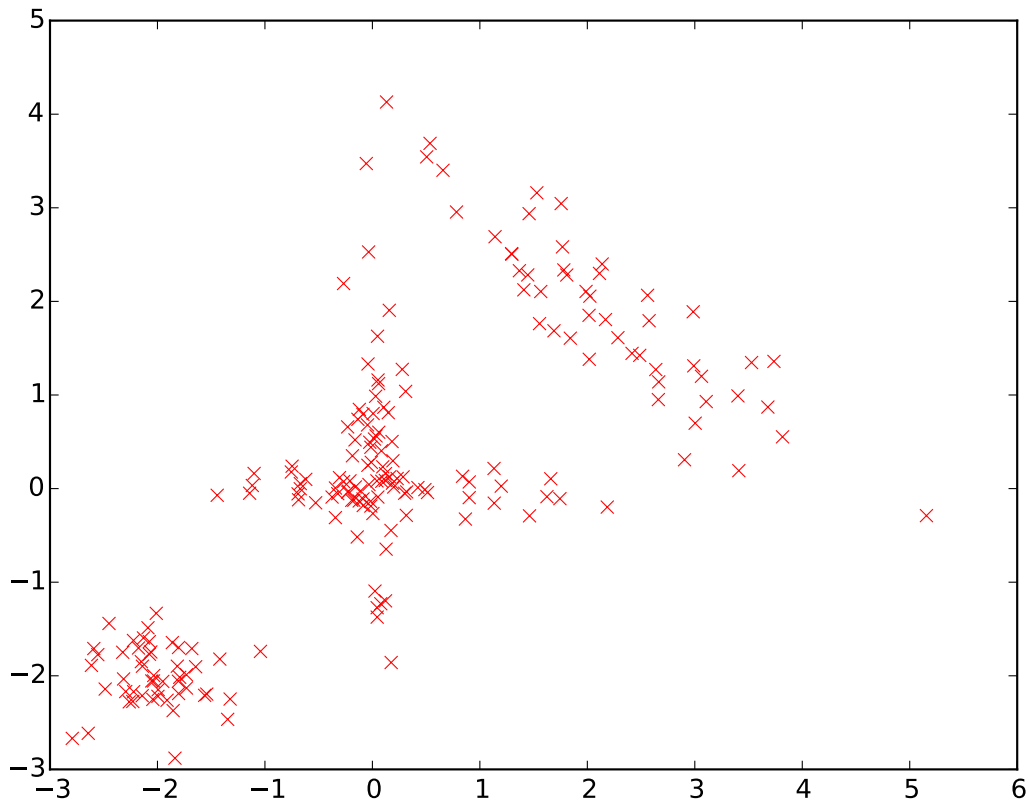
3.2.1 Data

First, let us generate some artificial data for the analysis. The data are two-dimensional vectors from one of the four different Gaussian distributions:

```
>>> import numpy as np
>>> y0 = np.random.multivariate_normal([0, 0], [[2, 0], [0, 0.1]], size=50)
>>> y1 = np.random.multivariate_normal([0, 0], [[0.1, 0], [0, 2]], size=50)
>>> y2 = np.random.multivariate_normal([2, 2], [[2, -1.5], [-1.5, 2]], size=50)
>>> y3 = np.random.multivariate_normal([-2, -2], [[0.5, 0], [0, 0.5]], size=50)
>>> y = np.vstack([y0, y1, y2, y3])
```

Thus, there are 200 data vectors in total. The data looks as follows:

```
>>> import bayespy.plot as bpplt
>>> bpplt.pyplot.plot(y[:,0], y[:,1], 'rx')
[<matplotlib.lines.Line2D object at 0x...>]
```



3.2.2 Model

For clarity, let us denote the number of the data vectors with N

```
>>> N = 200
```

and the dimensionality of the data vectors with D :

```
>>> D = 2
```

We will use a “large enough” number of Gaussian clusters in our model:

```
>>> K = 10
```

Cluster assignments Z and the prior for the cluster assignment probabilities α :

```
>>> from bayespy.nodes import Dirichlet, Categorical
>>> alpha = Dirichlet(1e-5*np.ones(K),
...                  name='alpha')
>>> Z = Categorical(alpha,
...                 plates=(N,),
...                 name='z')
```

The mean vectors and the precision matrices of the clusters:

```
>>> from bayespy.nodes import Gaussian, Wishart
>>> mu = Gaussian(np.zeros(D), 1e-5*np.identity(D),
...               plates=(K,),
...               name='mu')
>>> Lambda = Wishart(D, 1e-5*np.identity(D),
...                  plates=(K,),
...                  name='Lambda')
```

If either the mean or precision should be shared between clusters, then that node should not have plates, that is, `plates=()`. The data vectors are from a Gaussian mixture with cluster assignments `Z` and Gaussian component parameters `mu` and `Lambda`:

```
>>> from bayespy.nodes import Mixture
>>> Y = Mixture(Z, Gaussian, mu, Lambda,
...             name='Y')
```

```
>>> Z.initialize_from_random()
```

```
>>> from bayespy.inference import VB
>>> Q = VB(Y, mu, Lambda, Z, alpha)
```

3.2.3 Inference

Before running the inference algorithm, we provide the data:

```
>>> Y.observe(y)
```

Then, run VB iteration until convergence:

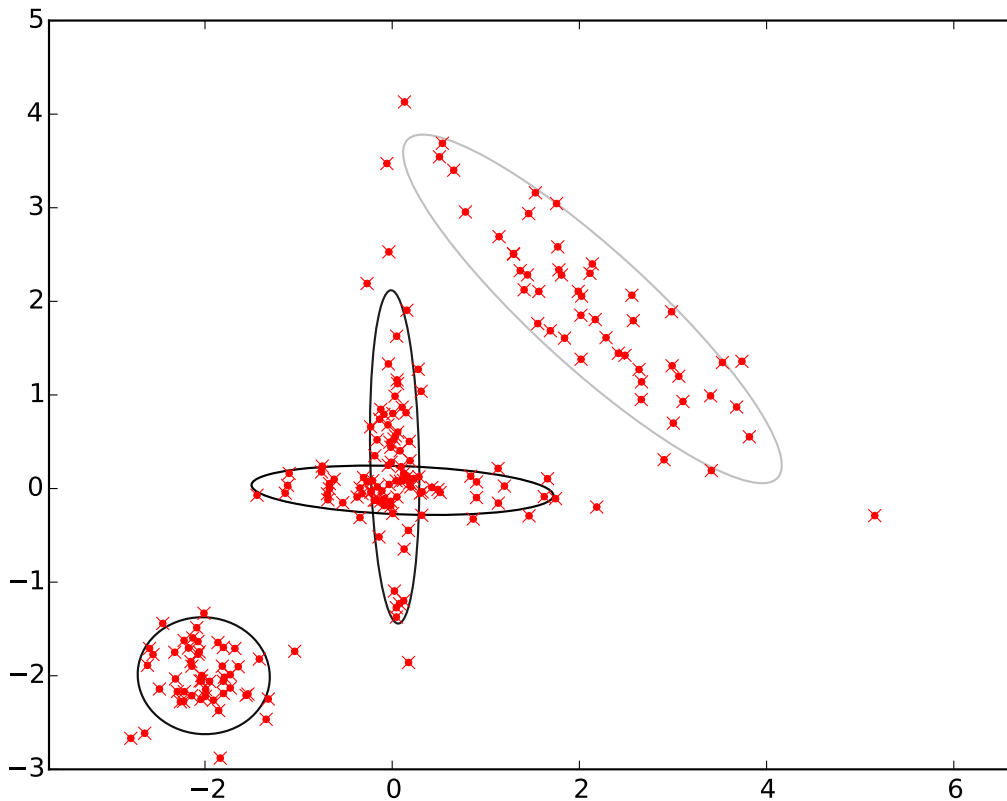
```
>>> Q.update(repeat=1000)
Iteration 1: loglike=-1.402345e+03 (... seconds)
...
Iteration 61: loglike=-8.888464e+02 (... seconds)
Converged at iteration 61.
```

The algorithm converges very quickly. Note that the default update order of the nodes was such that `mu` and `Lambda` were updated before `Z`, which is what we wanted because `Z` was initialized randomly.

3.2.4 Results

For two-dimensional Gaussian mixtures, the mixture components can be plotted using `gaussian_mixture_2d()`:

```
>>> bpplt.gaussian_mixture_2d(Y, alpha=alpha, scale=2)
```



The function is called with `scale=2` which means that each ellipse shows two standard deviations. From the ten cluster components, the model uses effectively the correct number of clusters (4). These clusters capture the true density accurately.

In addition to clustering and density estimation, this model could also be used for classification by setting the known class assignments as observed.

3.2.5 Advanced next steps

Joint node for mean and precision

The next step for improving the results could be to use `GaussianWishart` node for modelling the mean vectors `mu` and precision matrices `Lambda` jointly without factorization. This should improve the accuracy of the posterior approximation and the speed of the VB estimation. However, the implementation is a bit more complex.

Fast collapsed inference

3.3 Bernoulli mixture model

This example considers data generated from a Bernoulli mixture model. One simple example process could be a questionnaire for election candidates. We observe a set of binary vectors, where each vector represents a candidate in the election and each element in these vectors correspond to a candidate's answer to a yes-or-no question. The goal is to find groups of similar candidates and analyze the answer patterns of these groups.

3.3.1 Data

First, we generate artificial data to analyze. Let us assume that the questionnaire contains ten yes-or-no questions. We assume that there are three groups with similar opinions. These groups could represent parties. These groups have the following answering patterns, which are represented by vectors with probabilities of a candidate answering yes to the questions:

```
>>> p0 = [0.1, 0.9, 0.1, 0.9, 0.1, 0.9, 0.1, 0.9, 0.1, 0.9]
>>> p1 = [0.1, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.9, 0.9]
>>> p2 = [0.9, 0.9, 0.9, 0.9, 0.9, 0.1, 0.1, 0.1, 0.1, 0.1]
```

Thus, the candidates in the first group are likely to answer no to questions 1, 3, 5, 7 and 9, and yes to questions 2, 4, 6, 8, 10. The candidates in the second group are likely to answer yes to the last five questions, whereas the candidates in the third group are likely to answer yes to the first five questions. For convenience, we form a NumPy array of these vectors:

```
>>> import numpy as np
>>> p = np.array([p0, p1, p2])
```

Next, we generate a hundred candidates. First, we randomly select the group for each candidate:

```
>>> from bayespy.utils import random
>>> z = random.categorical([1/3, 1/3, 1/3], size=100)
```

Using the group patterns, we generate yes-or-no answers for the candidates:

```
>>> x = random.bernoulli(p[z])
```

This is our simulated data to be analyzed.

3.3.2 Model

Now, we construct a model for learning the structure in the data. We have a dataset of hundred 10-dimensional binary vectors:

```
>>> N = 100
>>> D = 10
```

We will create a Bernoulli mixture model. We assume that the true number of groups is unknown to us, so we use a large enough number of clusters:

```
>>> K = 10
```

We use the categorical distribution for the group assignments and give the group assignment probabilities an uninformative Dirichlet prior:

```
>>> from bayespy.nodes import Categorical, Dirichlet
>>> R = Dirichlet(K*[1e-5],
...              name='R')
>>> Z = Categorical(R,
...                 plates=(N,1),
...                 name='Z')
```

Each group has a probability of a yes answer for each question. These probabilities are given beta priors:

```
>>> from bayespy.nodes import Beta
>>> P = Beta([0.5, 0.5],
...          plates=(D,K),
...          name='P')
```

The answers of the candidates are modelled with the Bernoulli distribution:

```
>>> from bayespy.nodes import Mixture, Bernoulli
>>> X = Mixture(Z, Bernoulli, P)
```

Here, Z defines the group assignments and P the answering probability patterns for each group. Note how the plates of the nodes are matched: Z has plates (N, 1) and P has plates (D, K), but in the mixture node the last plate axis of P is discarded and thus the node broadcasts plates (N, 1) and (D,) resulting in plates (N, D) for X.

3.3.3 Inference

In order to infer the variables in our model, we construct a variational Bayesian inference engine:

```
>>> from bayespy.inference import VB
>>> Q = VB(Z, R, X, P)
```

This also gives the default update order of the nodes. In order to find different groups, they must be initialized differently, thus we use random initialization for the group probability patterns:

```
>>> P.initialize_from_random()
```

We provide our simulated data:

```
>>> X.observe(x)
```

Now, we can run inference:

```
>>> Q.update(repeat=1000)
Iteration 1: loglike=-6.872145e+02 (... seconds)
...
Iteration 17: loglike=-5.236921e+02 (... seconds)
Converged at iteration 17.
```

The algorithm converges in 17 iterations.

3.3.4 Results

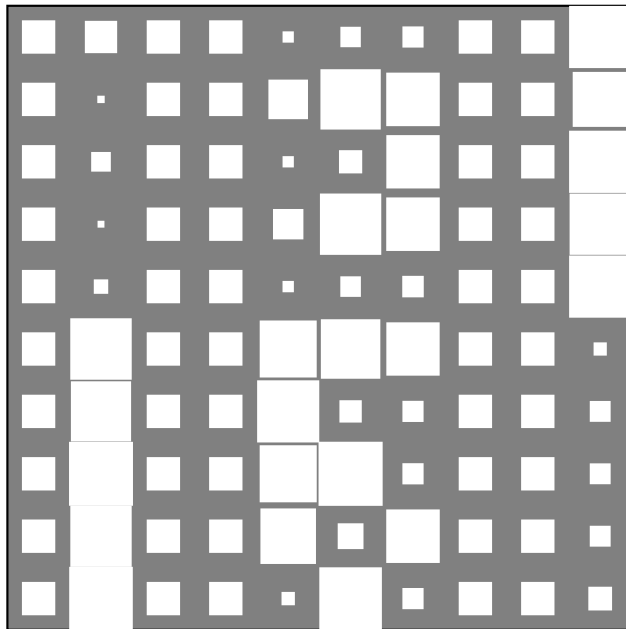
Now we can examine the approximate posterior distribution. First, let us plot the group assignment probabilities:

```
>>> import bayespy.plot as bpplt
>>> bpplt.hinton(R)
```



This plot shows that there are three dominant groups, which is equal to the true number of groups used to generate the data. However, there are still two smaller groups as the data does not give enough evidence to prune them out. The yes-or-no answer probability patterns for the groups can be plotted as:

```
>>> bpplt.hinton(P)
```

The three dominant groups have found the true patterns accurately. The patterns of the two minor groups some kind of mixtures of the three groups and they exist because the generated data happened to contain a few samples giving evidence for these groups. Finally, we can plot the group assignment probabilities for the candidates:

```
>>> bpplt.hinton(Z)
```



This plot shows the clustering of the candidates. It is possible to use `HintonPlotter` to enable monitoring during the VB iteration by providing `plotter=HintonPlotter()` for Z, P and R when creating the nodes.

3.4 Hidden Markov model

In this example, we will demonstrate the use of hidden Markov model in the case of known and unknown parameters. We will also use two different emission distributions to demonstrate the flexibility of the model construction.

3.4.1 Known parameters

This example follows the one presented in [Wikipedia](#).

Model

Each day, the state of the weather is either ‘rainy’ or ‘sunny’. The weather follows a first-order discrete Markov process. It has the following initial state probabilities

```
>>> a0 = [0.6, 0.4] # p(rainy)=0.6, p(sunny)=0.4
```

and state transition probabilities:

```
>>> A = [[0.7, 0.3], # p(rainy->rainy)=0.7, p(rainy->sunny)=0.3
...      [0.4, 0.6]] # p(sunny->rainy)=0.4, p(sunny->sunny)=0.6
```

We will be observing one hundred samples:

```
>>> N = 100
```

The discrete first-order Markov chain is constructed as:

```
>>> from bayespy.nodes import CategoricalMarkovChain
>>> Z = CategoricalMarkovChain(a0, A, states=N)
```

However, instead of observing this process directly, we observe whether Bob is ‘walking’, ‘shopping’ or ‘cleaning’. The probability of each activity depends on the current weather as follows:

```
>>> P = [[0.1, 0.4, 0.5],
...      [0.6, 0.3, 0.1]]
```

where the first row contains activity probabilities on a rainy weather and the second row contains activity probabilities on a sunny weather. Using these emission probabilities, the observed process is constructed as:

```
>>> from bayespy.nodes import Categorical, Mixture
>>> Y = Mixture(Z, Categorical, P)
```

Data

In order to test our method, we’ll generate artificial data from the model itself. First, draw realization of the weather process:

```
>>> weather = Z.random()
```

Then, using this weather, draw realizations of the activities:

```
>>> activity = Mixture(weather, Categorical, P).random()
```

Inference

Now, using this data, we set our variable Y to be observed:

```
>>> Y.observe(activity)
```

In order to run inference, we construct variational Bayesian inference engine:

```
>>> from bayespy.inference import VB
>>> Q = VB(Y, Z)
```

Note that we need to give all random variables to VB. In this case, the only random variables were Y and Z . Next we run the inference, that is, compute our posterior distribution:

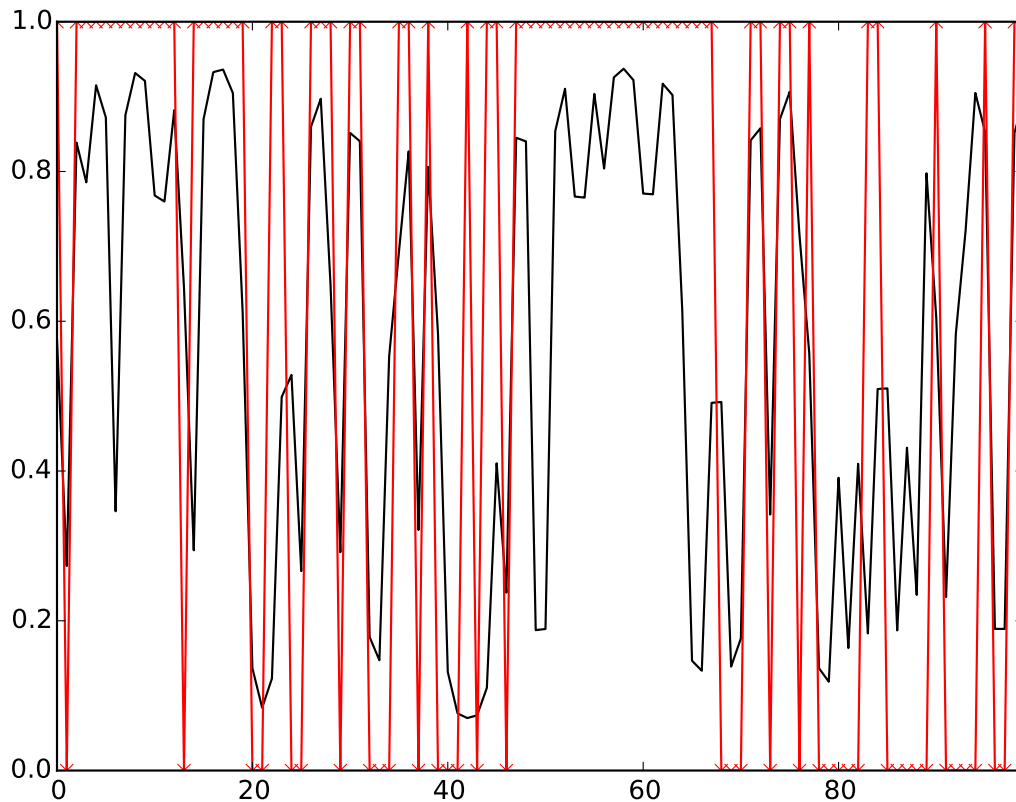
```
>>> Q.update()
Iteration 1: loglike=-1.095883e+02 (... seconds)
```

In this case, because there is only one unobserved random variable, we recover the exact posterior distribution and there is no need to iterate more than one step.

Results

One way to plot a 2-class categorical timeseries is to use the basic `plot()` function:

```
>>> import bayespy.plot as bpplt
>>> bpplt.plot(Z)
>>> bpplt.plot(1-weather, color='r', marker='x')
```



The black line shows the posterior probability of rain and the red line and crosses show the true state. Clearly, the method is not able to infer the weather very accurately in this case because the activities do not give that much information about the weather.

3.4.2 Unknown parameters

In this example, we consider unknown parameters for the Markov process and different emission distribution.

Data

We generate data from three 2-dimensional Gaussian distributions with different mean vectors and common standard deviation:

```
>>> import numpy as np
>>> mu = np.array([ [0,0], [3,4], [6,0] ])
>>> std = 2.0
```

Thus, the number of clusters is three:

```
>>> K = 3
```

And the number of samples is 200:

```
>>> N = 200
```

Each initial state is equally probable:

```
>>> p0 = np.ones(K) / K
```

State transition matrix is such that with probability 0.9 the process stays in the same state. The probability to move one of the other two states is 0.05 for both of those states.

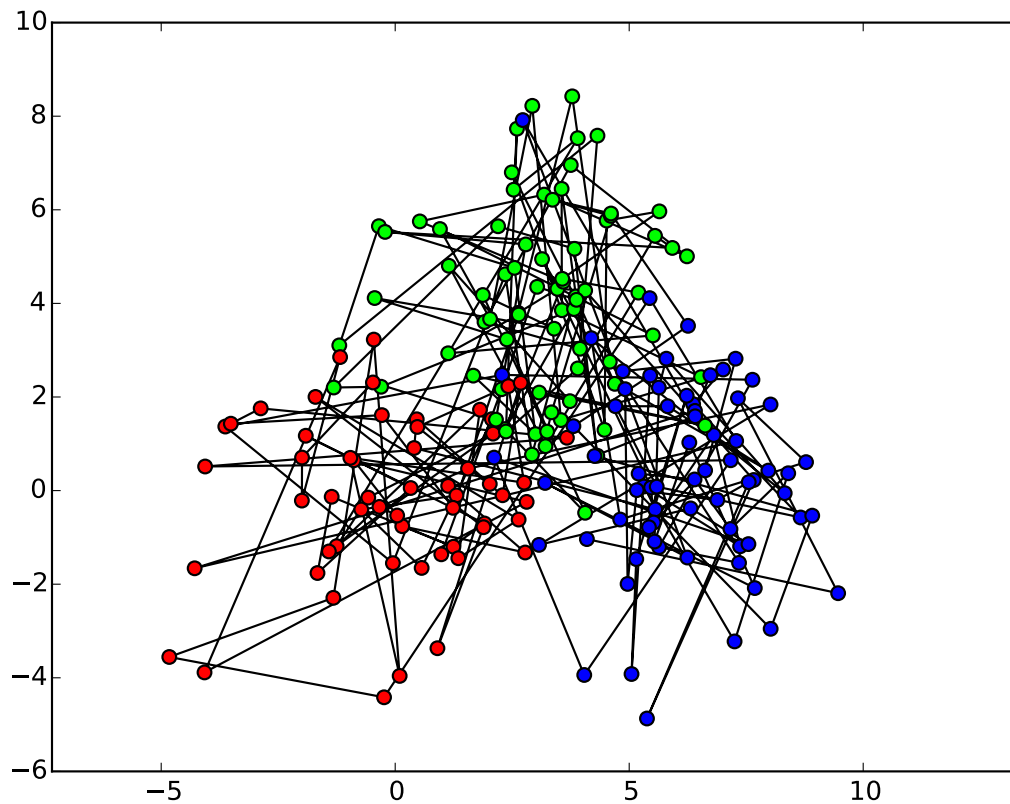
```
>>> q = 0.9
>>> r = (1-q) / (K-1)
>>> P = q*np.identity(K) + r*(np.ones((3,3))-np.identity(3))
```

Simulate the data:

```
>>> y = np.zeros((N,2))
>>> z = np.zeros(N)
>>> state = np.random.choice(K, p=p0)
>>> for n in range(N):
...     z[n] = state
...     y[n,:] = std*np.random.randn(2) + mu[state]
...     state = np.random.choice(K, p=P[state])
```

Then, let us visualize the data:

```
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.pyplot.axis('equal')
(...)
>>> colors = [ [1,0,0], [0,1,0], [0,0,1]][int(state)] for state in z ]
>>> bpplt.pyplot.plot(y[:,0], y[:,1], 'k-', zorder=-10)
<matplotlib.lines.Line2D object at 0x...>
>>> bpplt.pyplot.scatter(y[:,0], y[:,1], c=colors, s=40)
<matplotlib.collections.PathCollection object at 0x...>
```



Consecutive states are connected by a solid black line and the dot color shows the true class.

Model

Now, assume that we do not know the parameters of the process (initial state probability and state transition probabilities). We give these parameters quite non-informative priors, but it is possible to provide more informative priors if such information is available:

```
>>> from bayespy.nodes import Dirichlet
>>> a0 = Dirichlet(1e-3*np.ones(K))
>>> A = Dirichlet(1e-3*np.ones((K,K)))
```

The discrete Markov chain is constructed as:

```
>>> Z = CategoricalMarkovChain(a0, A, states=N)
```

Now, instead of using categorical emission distribution as before, we'll use Gaussian distribution. For simplicity, we use the true parameters of the Gaussian distributions instead of giving priors and estimating them. The known standard deviation can be converted to a precision matrix as:

```
>>> Lambda = std**(-2) * np.identity(2)
```

Thus, the observed process is a Gaussian mixture with cluster assignments from the hidden Markov process Z :

```
>>> from bayespy.nodes import Gaussian
>>> Y = Mixture(Z, Gaussian, mu, Lambda)
```

Note that `Lambda` does not have cluster plate axis because it is shared between the clusters.

Inference

Let us use the simulated data:

```
>>> Y.observe(y)
```

Because VB takes all the random variables, we need to provide `A` and `a0` also:

```
>>> Q = VB(Y, Z, A, a0)
```

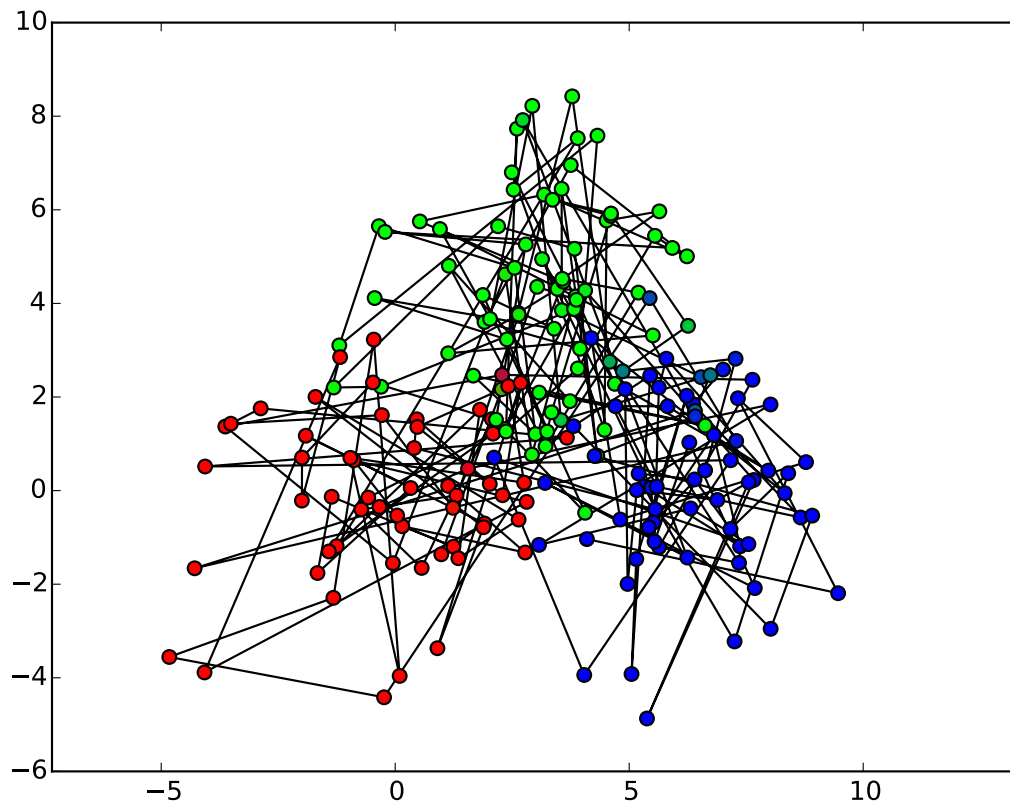
Then, run VB iteration until convergence:

```
>>> Q.update(repeat=1000)
Iteration 1: loglike=-9.963054e+02 (... seconds)
...
Iteration 8: loglike=-9.235053e+02 (... seconds)
Converged at iteration 8.
```

Results

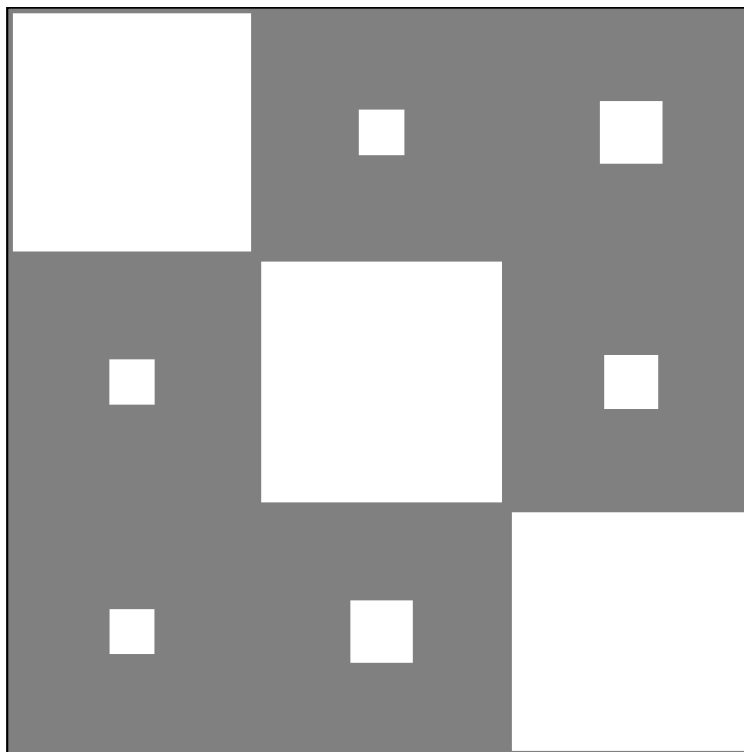
Plot the classification of the data similarly as the data:

```
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.pyplot.axis('equal')
(...)
>>> colors = Y.parents[0].get_moments()[0]
>>> bpplt.pyplot.plot(y[:,0], y[:,1], 'k-', zorder=-10)
[<matplotlib.lines.Line2D object at 0x...>]
>>> bpplt.pyplot.scatter(y[:,0], y[:,1], c=colors, s=40)
<matplotlib.collections.PathCollection object at 0x...>
```



The data has been classified quite correctly. Even samples that are more in the region of another cluster are classified correctly if the previous and next sample provide enough evidence for the correct class. We can also plot the state transition matrix:

```
>>> bpplt.hinton(A)
```

Clearly, the learned state transition matrix is close to the true matrix. The models described above could also be used for classification by providing the known class assignments as observed data to Z and the unknown class assignments as missing data.

3.5 Principal component analysis

This example uses a simple principal component analysis to find a two-dimensional latent subspace in a higher dimensional dataset.

3.5.1 Data

Let us create a Gaussian dataset with latent space dimensionality two and some observation noise:

```
>>> M = 20
>>> N = 100
```

```
>>> import numpy as np
>>> x = np.random.randn(N, 2)
>>> w = np.random.randn(M, 2)
>>> f = np.einsum('ik,jk->ij', w, x)
>>> y = f + 0.1*np.random.randn(M, N)
```

3.5.2 Model

We will use 10-dimensional latent space in our model and let it learn the true dimensionality:

```
>>> D = 10
```

Import relevant nodes:

```
>>> from bayespy.nodes import GaussianARD, Gamma, SumMultiply
```

The latent states:

```
>>> X = GaussianARD(0, 1, plates=(1,N), shape=(D,))
```

The loading matrix with automatic relevance determination (ARD) prior:

```
>>> alpha = Gamma(1e-5, 1e-5, plates=(D,))
>>> C = GaussianARD(0, alpha, plates=(M,1), shape=(D,))
```

Compute the dot product of the latent states and the loading matrix:

```
>>> F = SumMultiply('d,d->', X, C)
```

The observation noise:

```
>>> tau = Gamma(1e-5, 1e-5)
```

The observed variable:

```
>>> Y = GaussianARD(F, tau)
```

3.5.3 Inference

Observe the data:

```
>>> Y.observe(y)
```

We do not have missing data now, but they could be easily handled with `mask` keyword argument. Construct variational Bayesian (VB) inference engine:

```
>>> from bayespy.inference import VB
>>> Q = VB(Y, X, C, alpha, tau)
```

Initialize the latent subspace randomly, otherwise both `X` and `C` would converge to zero:

```
>>> C.initialize_from_random()
```

Now we could use `VB.update()` to run the inference. However, let us first create a parameter expansion to speed up the inference. The expansion is based on rotating the latent subspace optimally. This is optional but will usually improve the speed of the inference significantly, especially in high-dimensional problems:

```
>>> from bayespy.inference.vmp.transformations import RotateGaussianARD
>>> rot_X = RotateGaussianARD(X)
>>> rot_C = RotateGaussianARD(C, alpha)
```

By giving `alpha` for `rot_C`, the rotation will also optimize `alpha` jointly with `C`. Now that we have defined the rotations for our variables, we need to construct an optimizer:

```
>>> from bayespy.inference.vmp.transformations import RotationOptimizer
>>> R = RotationOptimizer(rot_X, rot_C, D)
```

In order to use the rotations automatically, we need to set it as a callback function:

```
>>> Q.set_callback(R.rotate)
```

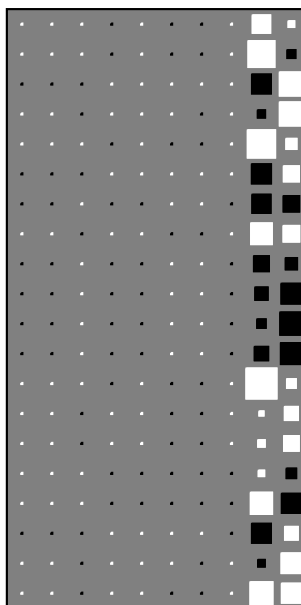
For more information about the rotation parameter expansion, see [\[7\]](#) and [\[6\]](#). Now we can run the actual inference until convergence:

```
>>> Q.update(repeat=1000)
Iteration 1: loglike=-2.339710e+03 (... seconds)
...
Iteration 22: loglike=6.500...e+02 (... seconds)
Converged at iteration 2...
```

3.5.4 Results

The results can be visualized, for instance, by plotting the Hinton diagram of the loading matrix:

```
>>> import bayespy.plot as bpplt
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.hinton(C)
```

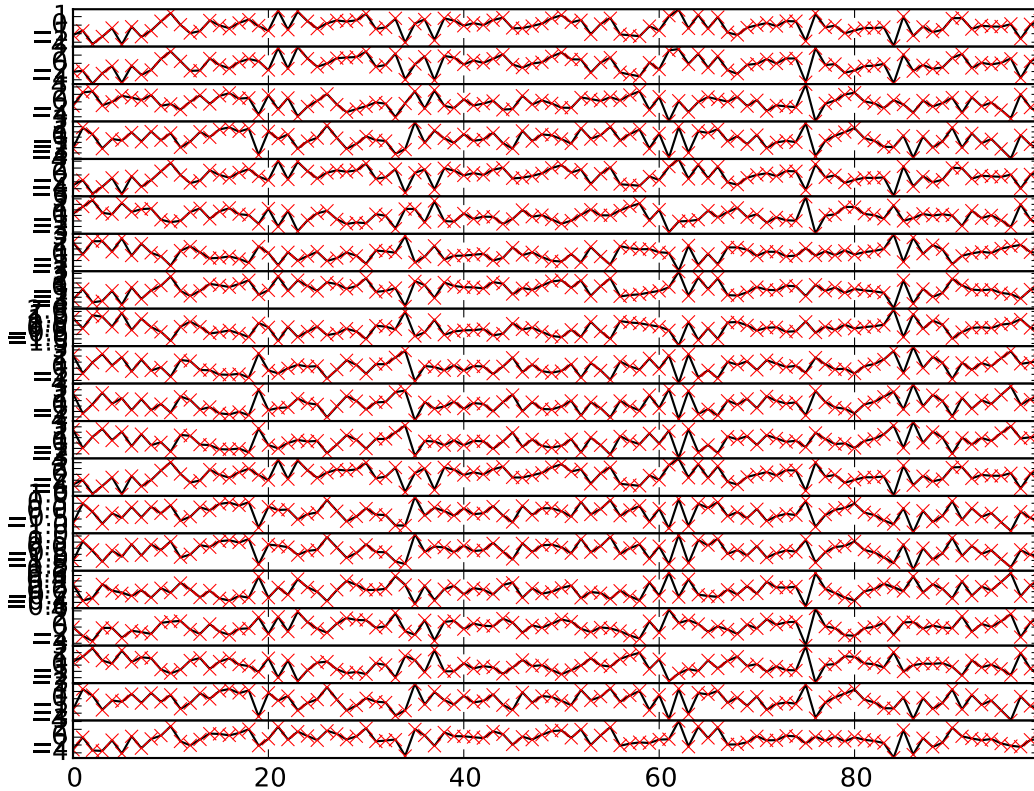


The method has been able to prune out unnecessary latent dimensions and keep two components, which is the true number of components.

```

>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.plot(F)
>>> bpplt.plot(f, color='r', marker='x', linestyle='None')

```



The reconstruction of the noiseless function values are practically perfect in this simple example. Larger noise variance, more latent space dimensions and missing values would make this problem more difficult. The model construction could also be improved by having, for instance, C and τ in the same node without factorizing between them in the posterior approximation. This can be achieved by using `GaussianGammaISO` node.

3.6 Linear state-space model

3.6.1 Model

In linear state-space models a sequence of M -dimensional observations $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ is assumed to be generated from latent D -dimensional states $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ which follow a first-order Markov process:

$$\begin{aligned}\mathbf{x}_n &= \mathbf{A}\mathbf{x}_{n-1} + \text{noise}, \\ \mathbf{y}_n &= \mathbf{C}\mathbf{x}_n + \text{noise},\end{aligned}$$

where the noise is Gaussian, \mathbf{A} is the $D \times D$ state dynamics matrix and \mathbf{C} is the $M \times D$ loading matrix. Usually, the latent space dimensionality D is assumed to be much smaller than the observation space dimensionality M in order to model the dependencies of high-dimensional observations efficiently.

In order to construct the model in BayesPy, first import relevant nodes:

```
>>> from bayespy.nodes import GaussianARD, GaussianMarkovChain, Gamma, Dot
```

The data vectors will be 30-dimensional:

```
>>> M = 30
```

There will be 400 data vectors:

```
>>> N = 400
```

Let us use 10-dimensional latent space:

```
>>> D = 10
```

The state dynamics matrix **A** has ARD prior:

```
>>> alpha = Gamma(1e-5,
...               1e-5,
...               plates=(D, ),
...               name='alpha')
>>> A = GaussianARD(0,
...                 alpha,
...                 shape=(D, ),
...                 plates=(D, ),
...                 name='A')
```

Note that **A** is a $D \times D$ -dimensional matrix. However, in BayesPy it is modelled as a collection (`plates=(D,)`) of D -dimensional vectors (`shape=(D,)`) because this is how the variables factorize in the posterior approximation of the state dynamics matrix in `GaussianMarkovChain`. The latent states are constructed as

```
>>> X = GaussianMarkovChain(np.zeros(D),
...                          1e-3*np.identity(D),
...                          A,
...                          np.ones(D),
...                          n=N,
...                          name='X')
```

where the first two arguments are the mean and precision matrix of the initial state, the third argument is the state dynamics matrix and the fourth argument is the diagonal elements of the precision matrix of the innovation noise. The node also needs the length of the chain given as the keyword argument `n=N`. Thus, the shape of this node is (N, D) .

The linear mapping from the latent space to the observation space is modelled with the loading matrix which has ARD prior:

```
>>> gamma = Gamma(1e-5,
...               1e-5,
...               plates=(D, ),
...               name='gamma')
>>> C = GaussianARD(0,
...                 gamma,
...                 shape=(D, ),
...                 plates=(M, 1),
...                 name='C')
```

Note that the plates for **C** are $(M, 1)$, thus the full shape of the node is $(M, 1, D)$. The unit plate axis is added so that **C** broadcasts with **X** when computing the dot product:

```
>>> F = Dot(C,
...          X,
...          name='F')
```

This dot product is computed over the D -dimensional latent space, thus the result is a $M \times N$ -dimensional matrix which is now represented with plates (M, N) in BayesPy:

```
>>> F.plates
(30, 400)
```

We also need to use random initialization either for C or X in order to find non-zero latent space because by default both C and X are initialized to zero because of their prior distributions. We use random initialization for C and then we must update X the first time before updating C :

```
>>> C.initialize_from_random()
```

The precision of the observation noise is given gamma prior:

```
>>> tau = Gamma(1e-5,
...             1e-5,
...             name='tau')
```

The observations are noisy versions of the dot products:

```
>>> Y = GaussianARD(F,
...                 tau,
...                 name='Y')
```

The variational Bayesian inference engine is then construed as:

```
>>> from bayespy.inference import VB
>>> Q = VB(X, C, gamma, A, alpha, tau, Y)
```

Note that X is given before C , thus X is updated before C by default.

3.6.2 Data

Now, let us generate some toy data for our model. Our true latent space is four dimensional with two noisy oscillator components, one random walk component and one white noise component.

```
>>> w = 0.3
>>> a = np.array([[np.cos(w), -np.sin(w), 0, 0],
...               [np.sin(w), np.cos(w), 0, 0],
...               [0, 0, 1, 0],
...               [0, 0, 0, 0]])
```

The true linear mapping is just random:

```
>>> c = np.random.randn(M, 4)
```

Then, generate the latent states and the observations using the model equations:

```
>>> x = np.empty((N, 4))
>>> f = np.empty((M, N))
>>> y = np.empty((M, N))
>>> x[0] = 10*np.random.randn(4)
>>> f[:, 0] = np.dot(c, x[0])
>>> y[:, 0] = f[:, 0] + 3*np.random.randn(M)
>>> for n in range(N-1):
...     x[n+1] = np.dot(a, x[n]) + [1, 1, 10, 10]*np.random.randn(4)
...     f[:, n+1] = np.dot(c, x[n+1])
...     y[:, n+1] = f[:, n+1] + 3*np.random.randn(M)
```

We want to simulate missing values, thus we create a mask which randomly removes 80% of the data:

```
>>> from bayespy.utils import random
>>> mask = random.mask(M, N, p=0.2)
>>> Y.observe(y, mask=mask)
```

3.6.3 Inference

As we did not define plotters for our nodes when creating the model, it is done now for some of the nodes:

```
>>> import bayespy.plot as bpplt
>>> X.set_plotter(bpplt.FunctionPlotter(center=True, axis=-2))
>>> A.set_plotter(bpplt.HintonPlotter())
>>> C.set_plotter(bpplt.HintonPlotter())
>>> tau.set_plotter(bpplt.PDFPlotter(np.linspace(0.02, 0.5, num=1000)))
```

This enables plotting of the approximate posterior distributions during VB learning. The inference engine can be run using `VB.update()` method:

```
>>> Q.update(repeat=10)
Iteration 1: loglike=-1.439704e+05 (... seconds)
...
Iteration 10: loglike=-1.051441e+04 (... seconds)
```

The iteration progresses a bit slowly, thus we'll consider parameter expansion to speed it up.

Parameter expansion

Section *Parameter expansion* discusses parameter expansion for state-space models to speed up inference. It is based on a rotating the latent space such that the posterior in the observation space is not affected:

$$\mathbf{y}_n = \mathbf{C}\mathbf{x}_n = (\mathbf{C}\mathbf{R}^{-1})(\mathbf{R}\mathbf{x}_n).$$

Thus, the transformation is $\mathbf{C} \rightarrow \mathbf{C}\mathbf{R}^{-1}$ and $\mathbf{X} \rightarrow \mathbf{R}\mathbf{X}$. In order to keep the dynamics of the latent states unaffected by the transformation, the state dynamics matrix \mathbf{A} must be transformed accordingly:

$$\mathbf{R}\mathbf{x}_n = \mathbf{R}\mathbf{A}\mathbf{R}^{-1}\mathbf{R}\mathbf{x}_{n-1},$$

resulting in a transformation $\mathbf{A} \rightarrow \mathbf{R}\mathbf{A}\mathbf{R}^{-1}$. For more details, refer to [6] and [7]. In BayesPy, the transformations are available in `bayespy.inference.vmp.transformations`:

```
>>> from bayespy.inference.vmp import transformations
```

The rotation of the loading matrix along with the ARD parameters is defined as:

```
>>> rotC = transformations.RotateGaussianARD(C, gamma)
```

For rotating \mathbf{X} , we first need to define the rotation of the state dynamics matrix:

```
>>> rotA = transformations.RotateGaussianARD(A, alpha)
```

Now we can define the rotation of the latent states:

```
>>> rotX = transformations.RotateGaussianMarkovChain(X, rotA)
```

The optimal rotation for all these variables is found using rotation optimizer:

```
>>> R = transformations.RotationOptimizer(rotX, rotC, D)
```

Set the parameter expansion to be applied after each iteration:

```
>>> Q.callback = R.rotate
```

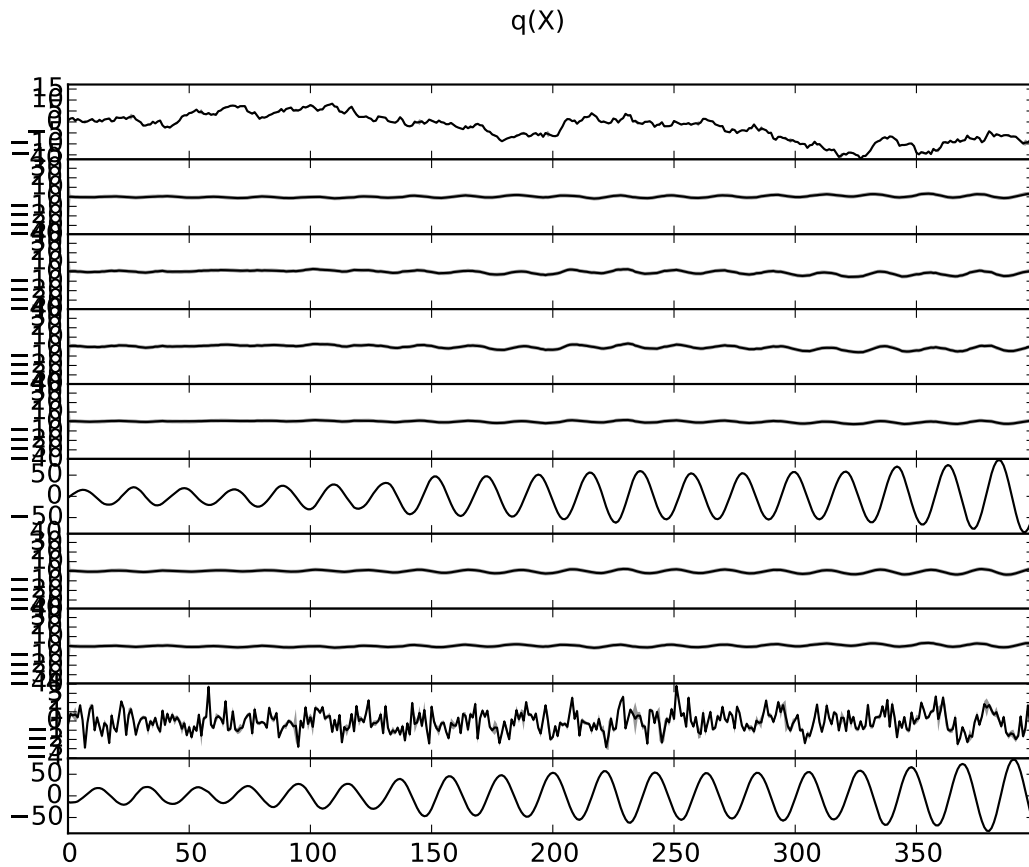
Now, run iterations until convergence:

```
>>> Q.update(repeat=1000)
Iteration 11: loglike=-1.010806e+04 (... seconds)
...
Iteration 58: loglike=-8.906...e+03 (... seconds)
Converged at iteration ...
```

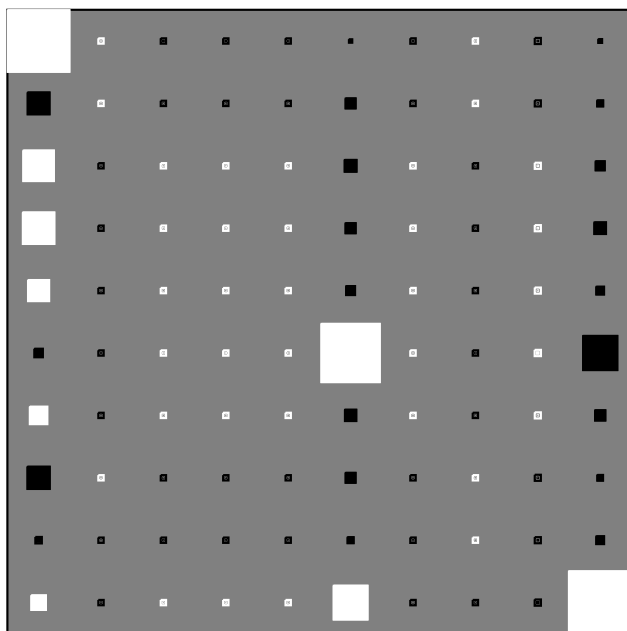
3.6.4 Results

Because we have set the plotters, we can plot those nodes as:

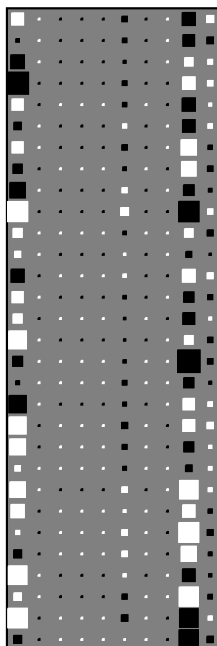
```
>>> Q.plot(X, A, C, tau)
```

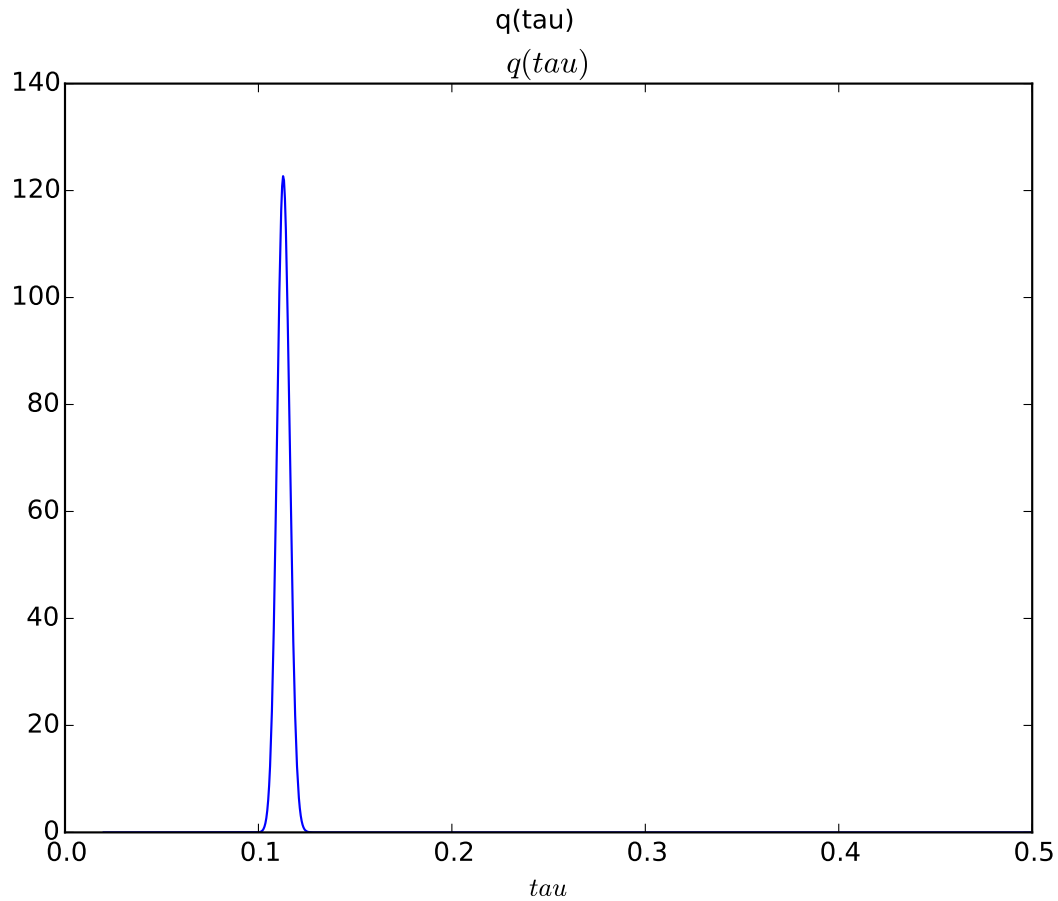


$q(A)$



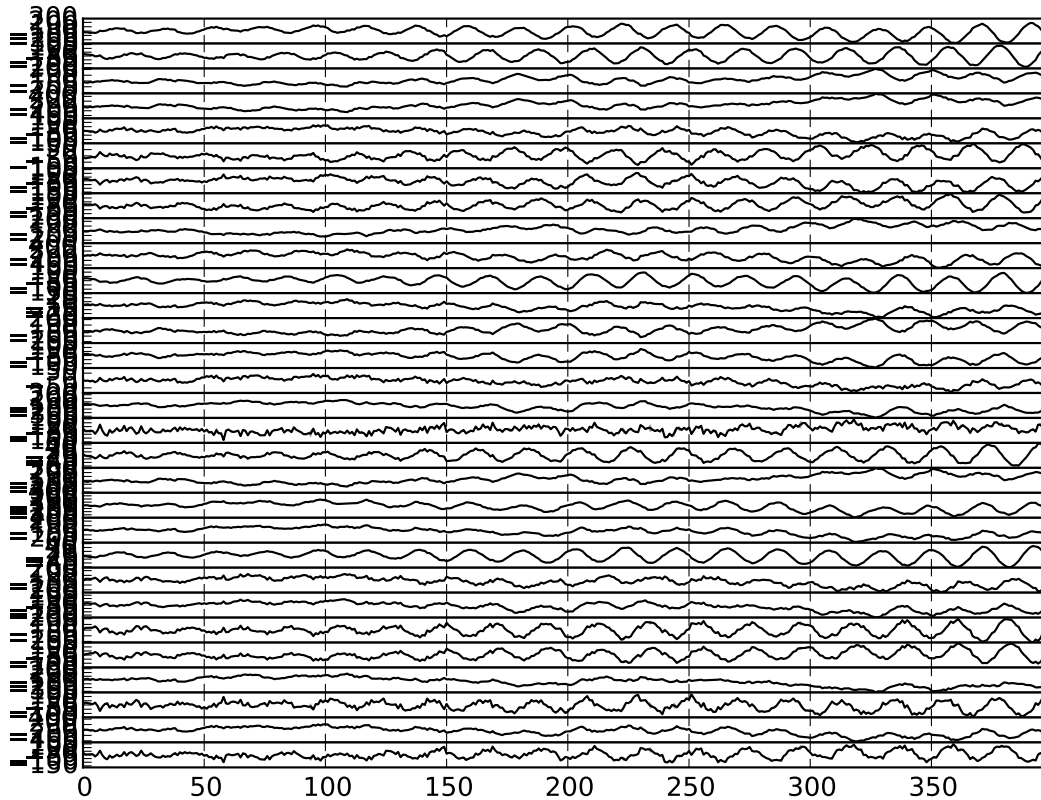
$q(C)$





There are clearly four effective components in X : random walk (component number 1), random oscillation (7 and 10), and white noise (9). These dynamics are also visible in the state dynamics matrix Hinton diagram. Note that the white noise component does not have any dynamics. Also C shows only four effective components. The posterior of τ captures the true value $3^{-2} \approx 0.111$ accurately. We can also plot predictions in the observation space:

```
>>> bpplt.plot(F, center=True)
```



We can also measure the performance numerically by computing root-mean-square error (RMSE) of the missing values:

```
>>> from bayespy.utils import misc
>>> misc.rmse(y[~mask], F.get_moments()[0][~mask])
5.18...
```

This is relatively close to the standard deviation of the noise (3), so the predictions are quite good considering that only 20% of the data was used.

3.7 Latent Dirichlet allocation

Latent Dirichlet allocation is a widely used topic model. The data is a collection of documents which contain words. The goal of the analysis is to find topics (distribution of words in topics) and document topics (distribution of topics in documents).

3.7.1 Data

The data consists of two vectors of equal length. The elements in these vectors correspond to the words in all documents combined. If there were M documents and each document had K words, the vectors contain $M \cdot K$ elements. Let M be the number of documents in total. The first vector gives each word a document index $i \in \{0, \dots, M - 1\}$

defining to which document the word belongs. Let N be the size of the whole available vocabulary. The second vector gives each word a vocabulary index $j \in \{0, \dots, N - 1\}$ defining which word it is from the vocabulary.

For this demo, we will just generate an artificial dataset for simplicity. We use the LDA model itself to generate the dataset. First, import relevant packages:

```
>>> import numpy as np
>>> from bayespy import nodes
```

Let us decide the number of documents and the number of words in those documents:

```
>>> n_documents = 10
>>> n_words = 10000
```

Randomly choose into which document each word belongs to:

```
>>> word_documents = nodes.Categorical(np.ones(n_documents)/n_documents,
...                                   plates=(n_words,)).random()
```

Let us also decide the size of our vocabulary:

```
>>> n_vocabulary = 100
```

Also, let us decide the true number of topics:

```
>>> n_topics = 5
```

Generate some random distributions for the topics in each document:

```
>>> p_topic = nodes.Dirichlet(1e-1*np.ones(n_topics),
...                            plates=(n_documents,)).random()
```

Generate some random distributions for the words in each topic:

```
>>> p_word = nodes.Dirichlet(1e-1*np.ones(n_vocabulary),
...                           plates=(n_topics,)).random()
```

Sample topic assignments for each word in each document:

```
>>> topic = nodes.Categorical(p_topic[word_documents],
...                           plates=(n_words,)).random()
```

And finally, draw vocabulary indices for each word in all the documents:

```
>>> corpus = nodes.Categorical(p_word[topic],
...                             plates=(n_words,)).random()
```

Now, our dataset consists of `word_documents` and `corpus`, which define the document and vocabulary indices for each word in our dataset.

Todo

Use some large real-world dataset, for instance, Wikipedia.

3.7.2 Model

Variable for learning the topic distribution for each document:

```
>>> p_topic = nodes.Dirichlet(np.ones(n_topics),
...                           plates=(n_documents,),
...                           name='p_topic')
```

Variable for learning the word distribution for each topic:

```
>>> p_word = nodes.Dirichlet(np.ones(n_vocabulary),
...                           plates=(n_topics,),
...                           name='p_word')
```

The document indices for each word in the corpus:

```
>>> from bayespy.inference.vmp.nodes.categorical import CategoricalMoments
>>> document_indices = nodes.Constant(CategoricalMoments(n_documents), word_documents,
...                                   name='document_indices')
```

Variable for learning the topic assignments of each word in the corpus:

```
>>> topics = nodes.Categorical(nodes.Gate(document_indices, p_topic),
...                             plates=(len(corpus),),
...                             name='topics')
```

The vocabulary indices for each word in the corpus:

```
>>> words = nodes.Categorical(nodes.Gate(topics, p_word),
...                             name='words')
```

3.7.3 Inference

Observe the corpus:

```
>>> words.observe(corpus)
```

Break symmetry by random initialization:

```
>>> p_topic.initialize_from_random()
>>> p_word.initialize_from_random()
```

Construct inference engine:

```
>>> from bayespy.inference import VB
>>> Q = VB(words, topics, p_word, p_topic, document_indices)
```

Run the VB learning algorithm:

```
>>> Q.update(repeat=1000)
Iteration ...
```

3.7.4 Results

Use `bayespy.plot` to plot the results:

```
>>> import bayespy.plot as bpplt
```

Plot the topic distributions for each document:

```
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.hinton(Q['p_topic'])
>>> bpplt.pyplot.title("Posterior topic distribution for each document")
<matplotlib.text.Text object at 0x...>
>>> bpplt.pyplot.xlabel("Topics")
<matplotlib.text.Text object at 0x...>
>>> bpplt.pyplot.ylabel("Documents")
<matplotlib.text.Text object at 0x...>
```

Plot the word distributions for each topic:

```
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.hinton(Q['p_word'])
>>> bpplt.pyplot.title("Posterior word distributions for each topic")
<matplotlib.text.Text object at 0x...>
>>> bpplt.pyplot.xlabel("Words")
<matplotlib.text.Text object at 0x...>
>>> bpplt.pyplot.ylabel("Topics")
<matplotlib.text.Text object at 0x...>
```

Todo

Create more illustrative plots.

3.7.5 Stochastic variational inference

LDA is a popular example for stochastic variational inference (SVI). Using SVI for LDA is quite simple in BayesPy. In SVI, only a subset of the dataset is used at each iteration step but this subset is “repeated” to get the same size as the original dataset. Let us define a size for the subset:

```
>>> subset_size = 1000
```

Thus, our subset will be repeat this many times:

```
>>> plates_multiplier = n_words / subset_size
```

Note that this multiplier doesn’t need to be an integer.

Now, let us repeat the model construction with only one minor addition. The following variables are identical to previous:

```
>>> p_topic = nodes.Dirichlet(np.ones(n_topics),
...                           plates=(n_documents,),
...                           name='p_topic')
>>> p_word = nodes.Dirichlet(np.ones(n_vocabulary),
...                          plates=(n_topics,),
...                          name='p_word')
```

The document indices vector is now a bit shorter, using only a subset:

```
>>> document_indices = nodes.Constant(CategoricalMoments(n_documents),
...                                  word_documents[:subset_size],
...                                  name='document_indices')
```

Note that at this point, it doesn't matter which elements we chose for the subset. For the topic assignments of each word in the corpus we need to use `plates_multiplier` because these topic assignments for the subset are "repeated" to recover the full dataset:

```
>>> topics = nodes.Categorical(nodes.Gate(document_indices, p_topic),
...                             plates=(subset_size,),
...                             plates_multiplier=(plates_multiplier,),
...                             name='topics')
```

Finally, the vocabulary indices for each word in the corpus are constructed as before:

```
>>> words = nodes.Categorical(nodes.Gate(topics, p_word),
...                             name='words')
```

This node inherits the plates and multipliers from its parent `topics`, so there is no need to define them here. Again, break symmetry by random initialization:

```
>>> p_topic.initialize_from_random()
>>> p_word.initialize_from_random()
```

Construct inference engine:

```
>>> from bayespy.inference import VB
>>> Q = VB(words, topics, p_word, p_topic, document_indices)
```

In order to use SVI, we need to disable some lower bound checks, because the lower bound doesn't anymore necessarily increase at each iteration step:

```
>>> Q.ignore_bound_checks = True
```

For the stochastic gradient ascent, we'll define some learning parameters:

```
>>> delay = 1
>>> forgetting_rate = 0.7
```

Run the inference:

```
>>> for n in range(1000):
...     # Observe a random mini-batch
...     subset = np.random.choice(n_words, subset_size)
...     Q['words'].observe(corpus[subset])
...     Q['document_indices'].set_value(word_documents[subset])
...     # Learn intermediate variables
...     Q.update('topics')
...     # Set step length
...     step = (n + delay) ** (-forgetting_rate)
...     # Stochastic gradient for the global variables
...     Q.gradient_step('p_topic', 'p_word', scale=step)
Iteration 1: ...
```

If one is interested, the lower bound values during the SVI algorithm can be plotted as:

```
>>> bpplt.pyplot.figure()
<matplotlib.figure.Figure object at 0x...>
>>> bpplt.pyplot.plot(Q.L)
[<matplotlib.lines.Line2D object at 0x...>]
```

The other results can be plotted as before.

DEVELOPER GUIDE

This chapter provides basic information for developers about contributing, the theoretical background and the core structure. It is assumed that the reader has read and is familiar with *User guide*.

4.1 Workflow

The main forum for BayesPy development is [GitHub](https://github.com/bayespy/bayespy). Bugs and other issues can be reported at <https://github.com/bayespy/bayespy/issues>. Contributions to the code and documentation are welcome and should be given as pull requests at <https://github.com/bayespy/bayespy/pulls>. In order to create pull requests, it is recommended to fork the git repository, make local changes and submit these changes as a pull request. The style guide for writing docstrings follows the style guide of NumPy, available at https://github.com/numpy/numpy/blob/master/doc/HOWTO_DOCUMENT.rst.txt. Detailed instructions on development workflow can be read from NumPy guide, available at http://docs.scipy.org/doc/numpy/dev/gitwash/development_workflow.html. BayesPy uses the following acronyms to start the commit message:

- API: an (incompatible) API change
- BLD: change related to building numpy
- BUG: bug fix
- DEMO: modification in demo code
- DEP: deprecate something, or remove a deprecated object
- DEV: development tool or utility
- DOC: documentation
- ENH: enhancement
- MAINT: maintenance commit (refactoring, typos, etc.)
- REV: revert an earlier commit
- STY: style fix (whitespace, PEP8)
- TST: addition or modification of tests
- REL: related to releasing

Since version 0.3.7, we have started following [Vincent Driessen's branching model](#) in how git is used.

4.1.1 Making releases

- Commit any current changes to git.
- Start a release branch: `git flow release start x.y.z`
- Edit version number in `setup.py` and commit.
- Add changes to `CHANGELOG.rst` and commit.
- Publish the release branch: `git flow release publish x.y.z`
- Finish the release: `git flow release finish x.y.z`. Write the following commit message: `REL: Version x.y.z.`
- Push to GitHub: `git push && git push --tags`
- Publish in PyPI: `python setup.py release_pypi`
- Update the documentation web page: `cd doc && make gh-pages`
- Publish in mloss.org.
- Announcements to bayespy@googlegroups.com, scipy-user@scipy.org and numpy-discussion@scipy.org.

4.2 Variational message passing

This section briefly describes the variational message passing (VMP) framework, which is currently the only implemented inference engine in BayesPy. The variational Bayesian (VB) inference engine in BayesPy assumes that the posterior approximation factorizes with respect to nodes and plates. VMP is based on updating one node at a time (the plates in one node can be updated simultaneously) and iteratively updating all nodes in turns until convergence.

4.2.1 Standard update equation

The general update equation for the factorized approximation of node θ is the following:

$$\log q(\theta) = \langle \log p(\theta | \text{pa}(\theta)) \rangle + \sum_{\mathbf{x} \in \text{ch}(\theta)} \langle \log p(\mathbf{x} | \text{pa}(\mathbf{x})) \rangle + \text{const}, \quad (4.1)$$

where $\text{pa}(\theta)$ and $\text{ch}(\theta)$ are the set of parents and children of θ , respectively. Thus, the posterior approximation of a node is updated by taking a sum of the expectations of all log densities in which the node variable appears. The expectations are over the approximate distribution of all other variables than θ . Actually, not all the variables are needed, because the non-constant part depends only on the Markov blanket of θ . This leads to a local optimization scheme, which uses messages from neighbouring nodes.

The messages are simple for conjugate exponential family models. An exponential family distribution has the following log probability density function:

$$\log p(\mathbf{x} | \Theta) = \mathbf{u}_{\mathbf{x}}(\mathbf{x})^T \phi_{\mathbf{x}}(\Theta) + g_{\mathbf{x}}(\Theta) + f_{\mathbf{x}}(\mathbf{x}), \quad (4.2)$$

where $\Theta = \{\theta_j\}$ is the set of parents, \mathbf{u} is the sufficient statistic vector, ϕ is the natural parameter vector, g is the negative log normalizer, and f is the log base function. Note that the log density is linear with respect to the terms that are functions of \mathbf{x} : \mathbf{u} and f . If a parent has a conjugate prior, (4.2) is also linear with respect to the parent's sufficient statistic vector. Thus, (4.2) can be re-organized with respect to a parent θ_j as

$$\log p(\mathbf{x} | \Theta) = \mathbf{u}_{\theta_j}(\theta_j)^T \phi_{\mathbf{x} \rightarrow \theta_j}(\mathbf{x}, \{\theta_k\}_{k \neq j}) + \text{const},$$

where \mathbf{u}_{θ_j} is the sufficient statistic vector of θ_j and the constant part is constant with respect to θ_j . Thus, the update equation (4.1) for θ_j can be written as

$$\begin{aligned}\log q(\theta_j) &= \mathbf{u}_{\theta_j}(\theta_j)^T \langle \phi_{\theta_j} \rangle + f_{\theta_j}(\theta_j) + \mathbf{u}_{\theta_j}(\theta_j)^T \sum_{\mathbf{x} \in \text{ch}(\theta_j)} \langle \phi_{\mathbf{x} \rightarrow \theta_j} \rangle + \text{const}, \\ &= \mathbf{u}_{\theta_j}(\theta_j)^T \left(\langle \phi_{\theta_j} \rangle + \sum_{\mathbf{x} \in \text{ch}(\theta_j)} \langle \phi_{\mathbf{x} \rightarrow \theta_j} \rangle \right) + f_{\theta_j}(\theta_j) + \text{const},\end{aligned}$$

where the summation is over all the child nodes of θ_j . Because of the conjugacy, $\langle \phi_{\theta_j} \rangle$ depends (multi)linearly on the parents' sufficient statistic vector. Similarly, $\langle \phi_{\mathbf{x} \rightarrow \theta_j} \rangle$ depends (multi)linearly on the expectations of the children's and co-parents' sufficient statistics. This gives the following update equation for the natural parameter vector of the posterior approximation $q(\phi_j)$:

$$\tilde{\phi}_j = \langle \phi_{\theta_j} \rangle + \sum_{\mathbf{x} \in \text{ch}(\theta_j)} \langle \phi_{\mathbf{x} \rightarrow \theta_j} \rangle. \quad (4.3)$$

4.2.2 Variational messages

The update equation (4.3) leads to a message passing scheme: the term $\langle \phi_{\theta_j} \rangle$ is a function of the parents' sufficient statistic vector and the term $\langle \phi_{\mathbf{x} \rightarrow \theta_j} \rangle$ can be interpreted as a message from the child node \mathbf{x} . Thus, the message from the child node \mathbf{x} to the parent node θ is

$$\mathbf{m}_{\mathbf{x} \rightarrow \theta} \equiv \langle \phi_{\mathbf{x} \rightarrow \theta} \rangle,$$

which can be computed as a function of the sufficient statistic vector of the co-parent nodes of θ and the sufficient statistic vector of the child node \mathbf{x} . The message from the parent node θ to the child node \mathbf{x} is simply the expectation of the sufficient statistic vector:

$$\mathbf{m}_{\theta \rightarrow \mathbf{x}} \equiv \langle \mathbf{u}_{\theta} \rangle.$$

In order to compute the expectation of the sufficient statistic vector we need to write $q(\theta)$ as

$$\log q(\theta) = \mathbf{u}(\theta)^T \tilde{\phi} + \tilde{g}(\tilde{\phi}) + f(\theta),$$

where $\tilde{\phi}$ is the natural parameter vector of $q(\theta)$. Now, the expectation of the sufficient statistic vector is defined as

$$\langle \mathbf{u}_{\theta} \rangle = -\frac{\partial \tilde{g}}{\partial \tilde{\phi}_{\theta}}(\tilde{\phi}_{\theta}). \quad (4.4)$$

We call this expectation of the sufficient statistic vector as the moments vector.

4.2.3 Lower bound

Computing the VB lower bound is not necessary in order to find the posterior approximation, although it is extremely useful in monitoring convergence and possible bugs. The VB lower bound can be written as

$$\mathcal{L} = \langle \log p(\mathbf{Y}, \mathbf{X}) \rangle - \langle \log q(\mathbf{X}) \rangle,$$

where \mathbf{Y} is the set of all observed variables and \mathbf{X} is the set of all latent variables. It can also be written as

$$\mathcal{L} = \sum_{\mathbf{y} \in \mathbf{Y}} \langle \log p(\mathbf{y} | \text{pa}(\mathbf{y})) \rangle + \sum_{\mathbf{x} \in \mathbf{X}} [\langle \log p(\mathbf{x} | \text{pa}(\mathbf{x})) \rangle - \langle \log q(\mathbf{x}) \rangle],$$

which shows that observed and latent variables contribute differently to the lower bound. These contributions have simple forms for exponential family nodes. Observed exponential family nodes contribute to the lower bound as follows:

$$\langle \log p(\mathbf{y}|\text{pa}(\mathbf{y})) \rangle = \mathbf{u}(\mathbf{y})^T \langle \phi \rangle + \langle g \rangle + f(\mathbf{x}),$$

where \mathbf{y} is the observed data. On the other hand, latent exponential family nodes contribute to the lower bound as follows:

$$\langle \log p(\mathbf{x}|\theta) \rangle - \langle \log q(\mathbf{x}) \rangle = \langle \mathbf{u} \rangle^T (\langle \phi \rangle - \tilde{\phi}) + \langle g \rangle - \tilde{g}.$$

If a node is partially observed and partially unobserved, these formulas are applied plate-wise appropriately.

4.2.4 Terms

To summarize, implementing VMP requires one to write for each stochastic exponential family node:

$\langle \phi \rangle$: the expectation of the prior natural parameter vector

Computed as a function of the messages from parents.

$\tilde{\phi}$: natural parameter vector of the posterior approximation

Computed as a sum of $\langle \phi \rangle$ and the messages from children.

$\langle \mathbf{u} \rangle$: the posterior moments vector

Computed as a function of $\tilde{\phi}$ as defined in (4.4).

$\mathbf{u}(\mathbf{x})$: the moments vector for given data

Computed as a function of the observed data \mathbf{x} .

$\langle g \rangle$: the expectation of the negative log normalizer of the prior

Computed as a function of parent moments.

\tilde{g} : the negative log normalizer of the posterior approximation

Computed as a function of $\tilde{\phi}$.

$f(\mathbf{x})$: the log base measure for given data

Computed as a function of the observed data \mathbf{x} .

$\langle \phi_{\mathbf{x} \rightarrow \theta} \rangle$: the message to parent θ

Computed as a function of the moments of this node and the other parents.

Deterministic nodes require only the following terms:

$\langle \mathbf{u} \rangle$: the posterior moments vector

Computed as a function of the messages from the parents.

\mathbf{m} : the message to a parent

Computed as a function of the messages from the other parents and all children.

4.3 Implementing inference engines

Currently, only variational Bayesian inference engine is implemented. This implementation is not very modular, that is, the inference engine is not well separated from the model construction. Thus, it is not straightforward to implement other inference engines at the moment. Improving the modularity of the inference engine and model construction is future work with high priority. In any case, BayesPy aims to be an efficient, simple and modular Bayesian package for variational inference at least.

4.4 Implementing nodes

The main goal of BayesPy is to provide a package which enables easy and flexible construction of simple and complex models with efficient inference. However, users may sometimes be unable to construct their models because the built-in nodes do not implement some specific features. Thus, one may need to implement new nodes in order to construct the model. BayesPy aims to make the implementation of new nodes both simple and fast. Probably, a large complex model can be constructed almost completely with the built-in nodes and the user needs to implement only a few nodes.

4.4.1 Moments

In order to implement nodes, it is important to understand the messaging framework of the nodes. A node is a unit of calculation which communicates to its parent and child nodes using messages. These messages have types that need to match between nodes, that is, the child node needs to understand the messages its parents are sending and vice versa. Thus, a node defines which message type it requires from each of its parents, and only nodes that have that type of output message (i.e., the message to a child node) are valid parent nodes for that node.

The message type is defined by the moments of the parent node. The moments are a collection of expectations: $\{\langle f_1(X) \rangle, \dots, \langle f_N(X) \rangle\}$. The functions f_1, \dots, f_N (and the number of the functions) define the message type and they are the sufficient statistic as discussed in the previous section. Different message types are represented by *Moments* class hierarchy. For instance, *GaussianMoments* represents a message type with parent moments $\{\langle \mathbf{x} \rangle, \langle \mathbf{x}\mathbf{x}^T \rangle\}$ and *WishartMoments* a message type with parent moments $\{\langle \mathbf{\Lambda} \rangle, \langle \log |\mathbf{\Lambda}| \rangle\}$.

Let us give an example: *Gaussian* node outputs *GaussianMoments* messages and *Wishart* node outputs *WishartMoments* messages. *Gaussian* node requires that it receives *GaussianMoments* messages from the mean parent node and *WishartMoments* messages from the precision parent node. Thus, *Gaussian* and *Wishart* are valid node classes as the mean and precision parent nodes of *Gaussian* node.

Note that several nodes may have the same output message type and some message types can be transformed to other message types using deterministic converter nodes. For instance, *Gaussian* and *GaussianARD* nodes both output *GaussianMoments* messages, deterministic *SumMultiply* also outputs *GaussianMoments* messages, and deterministic converter *MarkovChainToGaussian* converts *GaussianMarkovChainMoments* to *GaussianMoments*.

Each node specifies the message type requirements of its parents by *Node.parent_moments* attribute which is a list of *Moments* sub-class instances. These moments objects have a few purpose when creating the node: 1) check that parents are sending proper messages; 2) if parents use different message type, try to add a converter which converts the messages to the correct type if possible; 3) if given parents are not nodes but numeric arrays, convert them to constant nodes with correct output message type.

When implementing a new node, it is not always necessary to implement a new moments class. If another node has the same sufficient statistic vector, thus the same moments, that moments class can be used. Otherwise, one must implement a simple moments class which has the following methods:

- *Moments.compute_fixed_moments()*

Computes the moments for a known value. This is used to compute the moments of constant numeric arrays and wrap them into constant nodes.

- `Moments.compute_dims_from_values()`

Given a known value of the variable, return the shape of the variable dimensions in the moments. This is used to solve the shape of the moments array for constant nodes.

4.4.2 Distributions

In order to implement a stochastic exponential family node, one must first write down the log probability density function of the node and derive the terms discussed in section [Terms](#). These terms are implemented and collected as a class which is a subclass of `Distribution`. The main reason to implement these methods in another class instead of the node class itself is that these methods can be used without creating a node, for instance, in `Mixture` class.

For exponential family distributions, the distribution class is a subclass of `ExponentialFamilyDistribution`, and the relation between the terms in section [Terms](#) and the methods is as follows:

- `ExponentialFamilyDistribution.compute_phi_from_parents()`
Computes the expectation of the natural parameters $\langle \phi \rangle$ in the prior distribution given the moments of the parents.
- `ExponentialFamilyDistribution.compute_cgf_from_parents()`
Computes the expectation of the negative log normalizer $\langle g \rangle$ of the prior distribution given the moments of the parents.
- `ExponentialFamilyDistribution.compute_moments_and_cgf()`
Computes the moments $\langle \mathbf{u} \rangle$ and the negative log normalizer \tilde{g} of the posterior distribution given the natural parameters $\tilde{\phi}$.
- `ExponentialFamilyDistribution.compute_message_to_parent()`
Computes the message $\langle \phi_{\mathbf{x} \rightarrow \theta} \rangle$ from the node \mathbf{x} to its parent node θ given the moments of the node and the other parents.
- `ExponentialFamilyDistribution.compute_fixed_moments_and_f()`
Computes $\mathbf{u}(\mathbf{x})$ and $f(\mathbf{x})$ for given observed value \mathbf{x} . Without this method, variables from this distribution cannot be observed.

For each stochastic exponential family node, one must write a distribution class which implements these methods. After that, the node class is basically a simple wrapper and it also stores the moments and the natural parameters of the current posterior approximation. Note that the distribution classes do not store node-specific information, they are more like static collections of methods. However, sometimes the implementations depend on some information, such as the dimensionality of the variable, and this information must be provided, if needed, when constructing the distribution object.

In addition to the methods listed above, it is necessary to implement a few more methods in some cases. This happens when the plates of the parent do not map to the plates directly as discussed in section [Irregular plates](#). Then, one must write methods that implement this plate mapping and apply the same mapping to the mask array:

- `ExponentialFamilyDistribution.plates_from_parent()`
Given the plates of the parent, return the resulting plates of the child.
- `ExponentialFamilyDistribution.plates_to_parent()`
Given the plates of the child, return the plates of the parent that would have resulted them.
- `ExponentialFamilyDistribution.compute_mask_to_parent()`

Given the mask array of the child, apply the plate mapping.

It is important to understand when one must implement these methods, because the default implementations in the base class will lead to errors or weird results.

4.4.3 Stochastic exponential family nodes

After implementing the distribution class, the next task is to implement the node class. First, we need to explain a few important attributes before we can explain how to implement a node class.

Stochastic exponential family nodes have two attributes that store the state of the posterior distribution:

- `phi`
The natural parameter vector $\tilde{\phi}$ of the posterior approximation.
- `u`
The moments $\langle \mathbf{u} \rangle$ of the posterior approximation.

Instead of storing these two variables as vectors (as in the mathematical formulas), they are stored as lists of arrays with convenient shapes. For instance, `Gaussian` node stores the moments as a list consisting of a vector $\langle \mathbf{x} \rangle$ and a matrix $\langle \mathbf{x}\mathbf{x}^T \rangle$ instead of reshaping and concatenating these into a single vector. The same applies for the natural parameters `phi` because it has the same shape as `u`.

The shapes of the arrays in the lists `u` and `phi` consist of the shape caused by the plates and the shape caused by the variable itself. For instance, the moments of `Gaussian` node have shape $(D,)$ and (D, D) , where D is the dimensionality of the Gaussian vector. In addition, if the node has plates, they are added to these shapes. Thus, for instance, if the `Gaussian` node has plates $(3, 7)$ and D is 5, the shape of `u[0]` and `phi[0]` would be $(3, 7, 5)$ and the shape of `u[1]` and `phi[1]` would be $(3, 7, 5, 5)$. This shape information is stored in the following attributes:

- `plates` : a tuple
The plates of the node. In our example, $(3, 7)$.
- `dims` : a list of tuples
The shape of each of the moments arrays (or natural parameter arrays) without plates. In our example, $[(5,), (5, 5)]$.

Finally, three attributes define VMP for the node:

- `_moments` : `Moments` sub-class instance
An object defining the moments of the node.
- `_parent_moments` : list of `Moments` sub-class instances
A list defining the moments requirements for each parent.
- `_distribution` : `Distribution` sub-class instance
An object implementing the VMP formulas.

Basically, a node class is a collection of the above attributes. When a node is created, these attributes are defined. The base class for exponential family nodes, `ExponentialFamily`, provides a simple default constructor which does not need to be overwritten if `dims`, `_moments`, `_parent_moments` and `_distribution` can be provided as static class attributes. For instance, `Gamma` node defines these attributes statically. However, usually at least one of these attributes cannot be defined statically in the class. In that case, one must implement a class method which overloads `ExponentialFamily._constructor()`. The purpose of this method is to define all the attributes given the parent nodes. These are defined using a class method instead of `__init__` method in order to be able to use

the class constructors statically, for instance, in `Mixture` class. This construction allows users to create mixtures of any exponential family distribution with simple syntax.

The parents of a node must be converted so that they have a correct message type, because the user may have provided numeric arrays or nodes with incorrect message type. Numeric arrays should be converted to constant nodes with correct message type. Incorrect message type nodes should be converted to correct message type nodes if possible. Thus, the constructor should use `Node._ensure_moments` method to make sure the parent is a node with correct message type. Instead of calling this method for each parent node in the constructor, one can use `ensure_parents` decorator to do this automatically. However, the decorator requires that `_parent_moments` attribute has already been defined statically. If this is not possible, the parent nodes must be converted manually in the constructor, because one should never assume that the parent nodes given to the constructor are nodes with correct message type or even nodes at all.

4.4.4 Deterministic nodes

Deterministic nodes are nodes that do not correspond to any probability distribution but rather a deterministic function. It does not have any moments or natural parameters to store. A deterministic node is implemented as a subclass of `Deterministic` base class. The new node class must implement the following methods:

- `Deterministic._compute_moments()`

Computes the moments given the moments of the parents.

- `Deterministic._compute_message_to_parent()`

Computes the message to a parent node given the message from the children and the moments of the other parents. In some cases, one may want to implement `Deterministic._compute_message_and_mask_to_parent()` or `Deterministic._message_to_parent()` instead in order to gain more control over efficient computation.

Similarly as in `Distribution` class, if the node handles plates irregularly, it is important to implement the following methods:

- `Deterministic._plates_from_parent()`

Given the plates of the parent, return the resulting plates of the child.

- `Deterministic._plates_to_parent()`

Given the plates of the child, return the plates of the parent that would have resulted them.

- `Deterministic._compute_mask_to_parent()`

Given the mask array, convert it to a plate mask of the parent.

Converter nodes

Sometimes a node has incorrect message type but the message can be converted into a correct type. For instance, `GaussianMarkovChain` has `GaussianMarkovChainMoments` message type, which means moments $\{\langle \mathbf{x}_n \rangle, \langle \mathbf{x}_n \mathbf{x}_n^T \rangle, \langle \mathbf{x}_n \mathbf{x}_{n-1}^T \rangle\}_{n=1}^N$. These moments can be converted to `GaussianMoments` by ignoring the third element and considering the time axis as a plate axis. Thus, if a node requires `GaussianMoments` message from its parent, `GaussianMarkovChain` is a valid parent if its messages are modified properly. This conversion is implemented in `_MarkovChainToGaussian` converter class. Converter nodes are simple deterministic nodes that have one parent node and they convert the messages to another message type.

For the user, it is not convenient if the exact message type has to be known and an explicit converter node needs to be created. Thus, the conversions are done automatically and the user will be unaware of them. In order to enable this automatization, when writing a converter node, one should register the converter to the moments class using

`Moments.add_converter()`. For instance, a class `X` which converts moments `A` to moments `B` is registered as `A.add_converter(B, X)`. After that, `Node.ensure_moments()` and `Node._convert()` methods are used to perform the conversions automatically. The conversion can consist of several consecutive converter nodes, and the least number of conversions is used.

USER API

| | |
|--------------------------|--|
| <i>bayespy.nodes</i> | Package for nodes used to construct the model. |
| <i>bayespy.inference</i> | Package for Bayesian inference engines |
| <i>bayespy.plot</i> | Functions for plotting nodes. |

5.1 bayespy.nodes

Package for nodes used to construct the model.

5.1.1 Stochastic nodes

Nodes for Gaussian variables:

Nodes for precision and scale variables:

Nodes for modelling Gaussian and precision variables jointly (useful as prior for Gaussian nodes):

Nodes for discrete count variables:

Nodes for probabilities:

Nodes for dynamic variables:

Other stochastic nodes:

5.1.2 Deterministic nodes

5.2 bayespy.inference

Package for Bayesian inference engines

5.2.1 Inference engines

5.2.2 Parameter expansions

| | |
|---|--|
| <code>vmp.transformations.RotationOptimizer(...)</code> | Optimizer for rotation parameter expansion in state-space models |
| <code>vmp.transformations.RotateGaussian(X)</code> | Rotation parameter expansion for <code>bayespy.nodes.VaryingGaussian</code> |
| <code>vmp.transformations.RotateGaussianARD(X, *alpha)</code> | Rotation parameter expansion for <code>bayespy.nodes.VaryingGaussianARD</code> |
| <code>vmp.transformations.RotateGaussianMarkovChain(X, ...)</code> | Rotation parameter expansion for <code>bayespy.nodes.VaryingGaussianMarkovChain</code> |
| <code>vmp.transformations.RotateSwitchingMarkovChain(X, ...)</code> | Rotation for <code>bayespy.nodes.SwitchingGaussianMarkovChain</code> |
| <code>vmp.transformations.RotateVaryingMarkovChain(X, ...)</code> | Rotation for <code>bayespy.nodes.VaryingGaussianMarkovChain</code> |
| <code>vmp.transformations.RotateMultiple(*rotators)</code> | Identical parameter expansion for several nodes simultaneously |

bayespy.inference.vmp.transformations.RotationOptimizer

class `bayespy.inference.vmp.transformations.RotationOptimizer` (*block1*, *block2*, *D*)

Optimizer for rotation parameter expansion in state-space models

Rotates one model block with \mathbf{R} and one model block with \mathbf{R}^{-1} .

Parameters **block1** : rotator object

The first rotation parameter expansion object

block2 : rotator object

The second rotation parameter expansion object

D : int

Dimensionality of the latent space

References

[7], [6]

`__init__` (*block1*, *block2*, *D*)

Methods

`__init__` (*block1*, *block2*, *D*)

`rotate` ([*maxiter*, *check_gradient*, *verbose*, ...]) Optimize the rotation of two separate model blocks jointly.

bayespy.inference.vmp.transformations.RotationOptimizer.__init__

RotationOptimizer.__init__(block1, block2, D)

bayespy.inference.vmp.transformations.RotationOptimizer.rotate

RotationOptimizer.rotate(maxiter=10, check_gradient=False, verbose=False, check_bound=False)

Optimize the rotation of two separate model blocks jointly.

If some variable is the dot product of two Gaussians, rotating the two Gaussians optimally can make the inference algorithm orders of magnitude faster.

First block is rotated with \mathbf{R} and the second with \mathbf{R}^{-T} .

Blocks must have methods: *bound*(U,s,V) and *rotate*(R).

bayespy.inference.vmp.transformations.RotateGaussian

class bayespy.inference.vmp.transformations.RotateGaussian(X)

Rotation parameter expansion for bayespy.nodes.Gaussian

__init__(X)

Methods

| | |
|--|--|
| <code>__init__(X)</code> | |
| <code>bound(R[, logdet, inv])</code> | |
| <code>get_bound_terms(R[, logdet, inv])</code> | |
| <code>nodes()</code> | |
| <code>rotate(R[, inv, logdet])</code> | |
| <code>setup()</code> | This method should be called just before optimization. |

bayespy.inference.vmp.transformations.RotateGaussian.__init__

RotateGaussian.__init__(X)

bayespy.inference.vmp.transformations.RotateGaussian.bound

RotateGaussian.bound(R, logdet=None, inv=None)

bayespy.inference.vmp.transformations.RotateGaussian.get_bound_terms

RotateGaussian.get_bound_terms(R, logdet=None, inv=None)

bayespy.inference.vmp.transformations.RotateGaussian.nodes

RotateGaussian.nodes()

bayespy.inference.vmp.transformations.RotateGaussian.rotate

`RotateGaussian.rotate(R, inv=None, logdet=None)`

bayespy.inference.vmp.transformations.RotateGaussian.setup

`RotateGaussian.setup()`

This method should be called just before optimization.

bayespy.inference.vmp.transformations.RotateGaussianARD

`class bayespy.inference.vmp.transformations.RotateGaussianARD(X, *alpha, axis=-1, precompute=False, subset=None)`

Rotation parameter expansion for `bayespy.nodes.GaussianARD`

The model:

$\alpha \sim N(a, b)$ $X \sim N(\mu, \alpha)$

X can be an array (e.g., `GaussianARD`).

Transform $q(X)$ and $q(\alpha)$ by rotating X.

Requirements: * X and alpha do not contain any observed values

`__init__(X, *alpha, axis=-1, precompute=False, subset=None)`

Precompute tells whether to compute some moments once in the setup function instead of every time in the bound function. However, they are computed a bit differently in the bound function so it can be useful too. Precomputation is probably beneficial only when there are large axes that are not rotated (by R nor Q) and they are not contained in the plates of alpha, and the dimensions for R and Q are quite small.

Methods

| | |
|--|--|
| <code>__init__(X, *alpha[, axis, precompute, subset])</code> | Precompute tells whether to compute some moments once in the setup function instead of every time in the bound function. However, they are computed a bit differently in the bound function so it can be useful too. Precomputation is probably beneficial only when there are large axes that are not rotated (by R nor Q) and they are not contained in the plates of alpha, and the dimensions for R and Q are quite small. |
| <code>bound(R[, logdet, inv, Q])</code> | |
| <code>get_bound_terms(R[, logdet, inv, Q])</code> | |
| <code>nodes()</code> | |
| <code>rotate(R[, inv, logdet, Q])</code> | |
| <code>setup([plate_axis])</code> | This method should be called just before optimization. |

bayespy.inference.vmp.transformations.RotateGaussianARD.__init__

`RotateGaussianARD.__init__(X, *alpha, axis=-1, precompute=False, subset=None)`

Precompute tells whether to compute some moments once in the setup function instead of every time in the bound function. However, they are computed a bit differently in the bound function so it can be useful too. Precomputation is probably beneficial only when there are large axes that are not rotated (by R nor Q) and they are not contained in the plates of alpha, and the dimensions for R and Q are quite small.

bayespy.inference.vmp.transformations.RotateGaussianARD.bound

`RotateGaussianARD.bound(R, logdet=None, inv=None, Q=None)`

bayespy.inference.vmp.transformations.RotateGaussianARD.get_bound_terms

`RotateGaussianARD.get_bound_terms (R, logdet=None, inv=None, Q=None)`

bayespy.inference.vmp.transformations.RotateGaussianARD.nodes

`RotateGaussianARD.nodes ()`

bayespy.inference.vmp.transformations.RotateGaussianARD.rotate

`RotateGaussianARD.rotate (R, inv=None, logdet=None, Q=None)`

bayespy.inference.vmp.transformations.RotateGaussianARD.setup

`RotateGaussianARD.setup (plate_axis=None)`

This method should be called just before optimization.

For efficiency, sum over axes that are not in mu, alpha nor rotation.

If using Q, set rotate_plates to True.

bayespy.inference.vmp.transformations.RotateGaussianMarkovChain

class `bayespy.inference.vmp.transformations.RotateGaussianMarkovChain (X, *args)`

Rotation parameter expansion for `bayespy.nodes.GaussianMarkovChain`

Assume the following model.

Constant, unit isotropic innovation noise. Unit variance only?

Maybe: Assume innovation noise with unit variance? Would it help make this function more general with respect to A.

TODO: Allow constant A or not rotating A.

A may vary in time.

Shape of A: (N,D,D) Shape of AA: (N,D,D,D)

No plates for X.

`__init__ (X, *args)`

Methods

| | |
|--|--|
| <code>__init__(X, *args)</code> | |
| <code>bound(R[, logdet, inv])</code> | |
| <code>get_bound_terms(R[, logdet, inv])</code> | |
| <code>nodes()</code> | |
| <code>rotate(R[, inv, logdet])</code> | |
| <code>setup()</code> | This method should be called just before optimization. |

bayespy.inference.vmp.transformations.RotateGaussianMarkovChain.__init__

`RotateGaussianMarkovChain.__init__(X, *args)`

bayespy.inference.vmp.transformations.RotateGaussianMarkovChain.bound

`RotateGaussianMarkovChain.bound(R, logdet=None, inv=None)`

bayespy.inference.vmp.transformations.RotateGaussianMarkovChain.get_bound_terms

`RotateGaussianMarkovChain.get_bound_terms(R, logdet=None, inv=None)`

bayespy.inference.vmp.transformations.RotateGaussianMarkovChain.nodes

`RotateGaussianMarkovChain.nodes()`

bayespy.inference.vmp.transformations.RotateGaussianMarkovChain.rotate

`RotateGaussianMarkovChain.rotate(R, inv=None, logdet=None)`

bayespy.inference.vmp.transformations.RotateGaussianMarkovChain.setup

`RotateGaussianMarkovChain.setup()`
This method should be called just before optimization.

bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain

class bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain (*X, B, Z, B_rotator*)

Rotation for bayespy.nodes.VaryingGaussianMarkovChain

Assume the following model.

Constant, unit isotropic innovation noise.

$$A_n = B_{z_n}$$

Gaussian B: (... , K, D) x (D) Categorical Z: (... , N-1) x (K) GaussianMarkovChain X: (...) x (N,D)

No plates for X.

`__init__(X, B, Z, B_rotator)`

Methods

`__init__(X, B, Z, B_rotator)`

`bound(R[, logdet, inv])`

`get_bound_terms(R[, logdet, inv])`

`nodes()`

Continued on next page

Table 5.16 – continued from previous page

| | |
|---------------------------------------|--|
| <code>rotate(R[, inv, logdet])</code> | |
| <code>setup()</code> | This method should be called just before optimization. |

bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain.__init__

`RotateSwitchingMarkovChain.__init__(X, B, Z, B_rotator)`

bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain.bound

`RotateSwitchingMarkovChain.bound(R, logdet=None, inv=None)`

bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain.get_bound_terms

`RotateSwitchingMarkovChain.get_bound_terms(R, logdet=None, inv=None)`

bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain.nodes

`RotateSwitchingMarkovChain.nodes()`

bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain.rotate

`RotateSwitchingMarkovChain.rotate(R, inv=None, logdet=None)`

bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain.setup

`RotateSwitchingMarkovChain.setup()`
This method should be called just before optimization.

bayespy.inference.vmp.transformations.RotateVaryingMarkovChain

class `bayespy.inference.vmp.transformations.RotateVaryingMarkovChain(X, B, S, B_rotator)`

Rotation for `bayespy.nodes.SwitchingGaussianMarkovChain`

Assume the following model.

Constant, unit isotropic innovation noise.

$$A_n = \sum_k B_k s_{kn}$$

Gaussian B: (1,D) x (D,K) Gaussian S: (N,1) x (K) MC X: () x (N+1,D)

No plates for X.

`__init__(X, B, S, B_rotator)`

Methods

| | |
|--|--|
| <code>__init__(X, B, S, B_rotator)</code> | |
| <code>bound(R[, logdet, inv])</code> | |
| <code>get_bound_terms(R[, logdet, inv])</code> | |
| <code>nodes()</code> | |
| <code>rotate(R[, inv, logdet])</code> | |
| <code>setup()</code> | This method should be called just before optimization. |

bayespy.inference.vmp.transformations.RotateVaryingMarkovChain.__init__

`RotateVaryingMarkovChain.__init__(X, B, S, B_rotator)`

bayespy.inference.vmp.transformations.RotateVaryingMarkovChain.bound

`RotateVaryingMarkovChain.bound(R, logdet=None, inv=None)`

bayespy.inference.vmp.transformations.RotateVaryingMarkovChain.get_bound_terms

`RotateVaryingMarkovChain.get_bound_terms(R, logdet=None, inv=None)`

bayespy.inference.vmp.transformations.RotateVaryingMarkovChain.nodes

`RotateVaryingMarkovChain.nodes()`

bayespy.inference.vmp.transformations.RotateVaryingMarkovChain.rotate

`RotateVaryingMarkovChain.rotate(R, inv=None, logdet=None)`

bayespy.inference.vmp.transformations.RotateVaryingMarkovChain.setup

`RotateVaryingMarkovChain.setup()`
This method should be called just before optimization.

bayespy.inference.vmp.transformations.RotateMultiple

class `bayespy.inference.vmp.transformations.RotateMultiple(*rotators)`

Identical parameter expansion for several nodes simultaneously

Performs the same rotation for multiple nodes and combines the cost effect.

`__init__(*rotators)`

Methods

| |
|--------------------------------------|
| <code>__init__(*rotators)</code> |
| <code>bound(R[, logdet, inv])</code> |
| Continued on next page |

Table 5.18 – continued from previous page

| |
|--|
| <code>get_bound_terms(R[, logdet, inv])</code> |
| <code>nodes()</code> |
| <code>rotate(R[, inv, logdet])</code> |
| <code>setup()</code> |

bayespy.inference.vmp.transformations.RotateMultiple.__init__

`RotateMultiple.__init__(*rotators)`

bayespy.inference.vmp.transformations.RotateMultiple.bound

`RotateMultiple.bound(R, logdet=None, inv=None)`

bayespy.inference.vmp.transformations.RotateMultiple.get_bound_terms

`RotateMultiple.get_bound_terms(R, logdet=None, inv=None)`

bayespy.inference.vmp.transformations.RotateMultiple.nodes

`RotateMultiple.nodes()`

bayespy.inference.vmp.transformations.RotateMultiple.rotate

`RotateMultiple.rotate(R, inv=None, logdet=None)`

bayespy.inference.vmp.transformations.RotateMultiple.setup

`RotateMultiple.setup()`

5.3 bayespy.plot

Functions for plotting nodes.

5.3.1 Functions

| | |
|--|--|
| <code>pdf(Z, x, *args[, name, axes, fig])</code> | Plot probability density function of a scalar variable. |
| <code>contour(Z, x, y[, n, axes, fig])</code> | Plot 2-D probability density function of a 2-D variable. |
| <code>plot(Y[, axis, scale, center])</code> | Plot a variable or an array as 1-D function with errorbars |
| <code>hinton(X, **kwargs)</code> | Plot the Hinton diagram of a node |
| <code>gaussian.mixture_2d(X[, alpha, scale, fill, ...])</code> | Plot Gaussian mixture as ellipses in 2-D |

bayespy.plot.pdf

`bayespy.plot.pdf` (*Z*, *x*, **args*, *name=None*, *axes=None*, *fig=None*, ***kwargs*)
 Plot probability density function of a scalar variable.

Parameters *Z* : node or function

Stochastic node or log pdf function

x : array

Grid points

bayespy.plot.contour

`bayespy.plot.contour` (*Z*, *x*, *y*, *n=None*, *axes=None*, *fig=None*, ***kwargs*)
 Plot 2-D probability density function of a 2-D variable.

Parameters *Z* : node or function

Stochastic node or log pdf function

x : array

Grid points on x axis

y : array

Grid points on y axis

bayespy.plot.plot

`bayespy.plot.plot` (*Y*, *axis=-1*, *scale=2*, *center=False*, ***kwargs*)
 Plot a variable or an array as 1-D function with errorbars

bayespy.plot.hinton

`bayespy.plot.hinton` (*X*, ***kwargs*)
 Plot the Hinton diagram of a node

The keyword arguments depend on the node type. For some node types, the diagram also shows uncertainty with non-filled rectangles. Currently, beta-like, Gaussian-like and Dirichlet-like nodes are supported.

Parameters *X* : node

bayespy.plot.gaussian_mixture_2d

`bayespy.plot.gaussian_mixture_2d` (*X*, *alpha=None*, *scale=2*, *fill=False*, *axes=None*, ***kwargs*)
 Plot Gaussian mixture as ellipses in 2-D

Parameters *X* : Mixture node

alpha : Dirichlet-like node (optional)

Probabilities for the clusters

scale : float (optional)

Scale for the covariance ellipses (by default, 2)

5.3.2 Plotters

| | |
|---|---|
| <code>Plotter(plotter, *args, **kwargs)</code> | Wrapper for plotting functions and base class for node plotters |
| <code>PDFPlotter(x_grid, **kwargs)</code> | Plotter of probability density function of a scalar node |
| <code>ContourPlotter(x1_grid, x2_grid, **kwargs)</code> | Plotter of probability density function of a two-dimensional node |
| <code>HintonPlotter(**kwargs)</code> | Plotter of the Hinton diagram of a node |
| <code>FunctionPlotter(**kwargs)</code> | Plotter of a node as a 1-dimensional function |
| <code>GaussianTimeseriesPlotter(**kwargs)</code> | Plotter of a Gaussian node as a timeseries |
| <code>CategoricalMarkovChainPlotter(**kwargs)</code> | Plotter of a Categorical timeseries |

bayespy.plot.Plotter

class bayespy.plot.**Plotter** (*plotter, *args, **kwargs*)

Wrapper for plotting functions and base class for node plotters

The purpose of this class is to collect all the parameters needed by a plotting function and provide a callable interface which needs only the node as the input.

Plotter instances are callable objects that plot a given node using a specified plotting function.

Parameters **plotter** : function

Plotting function to use

args : defined by the plotting function

Additional inputs needed by the plotting function

kwargs : defined by the plotting function

Additional keyword arguments supported by the plotting function

Examples

First, create a gamma variable:

```
>>> import numpy as np
>>> from bayespy.nodes import Gamma
>>> x = Gamma(4, 5)
```

The probability density function can be plotted as:

```
>>> import bayespy.plot as bpplt
>>> bpplt.pdf(x, np.linspace(0.1, 10, num=100))
[<matplotlib.lines.Line2D object at 0x...>]
```

However, this can be problematic when one needs to provide a plotting function for the inference engine as the inference engine gives only the node as input. Thus, we need to create a simple plotter wrapper:

```
>>> p = bpplt.Plotter(bpplt.pdf, np.linspace(0.1, 10, num=100))
```

Now, this callable object `p` needs only the node as the input:

```
>>> p(x)
[<matplotlib.lines.Line2D object at 0x...>]
```

Thus, it can be given to the inference engine to use as a plotting function:

```
>>> x = Gamma(4, 5, plotter=p)
>>> x.plot()
[<matplotlib.lines.Line2D object at 0x...>]
```

`__init__` (*plotter*, *args, **kwargs)

Methods

`__init__` (*plotter*, *args, **kwargs)

bayespy.plot.Plotter.__init__

Plotter.`__init__` (*plotter*, *args, **kwargs)

bayespy.plot.PDFPlotter

class bayespy.plot.**PDFPlotter** (*x_grid*, **kwargs)

Plotter of probability density function of a scalar node

Parameters **x_grid** : array

Numerical grid on which the density function is computed and plotted

See also:

[*pdf*](#)

`__init__` (*x_grid*, **kwargs)

Methods

`__init__` (*x_grid*, **kwargs)

bayespy.plot.PDFPlotter.__init__

PDFPlotter.`__init__` (*x_grid*, **kwargs)

bayespy.plot.ContourPlotter

class bayespy.plot.**ContourPlotter** (*x1_grid*, *x2_grid*, **kwargs)

Plotter of probability density function of a two-dimensional node

Parameters **x1_grid** : array

Grid for the first dimension

x2_grid : array

Grid for the second dimension

See also:

[*contour*](#)

```
__init__(x1_grid, x2_grid, **kwargs)
```

Methods

```
__init__(x1_grid, x2_grid, **kwargs)
```

bayespy.plot.ContourPlotter.__init__

```
ContourPlotter.__init__(x1_grid, x2_grid, **kwargs)
```

bayespy.plot.HintonPlotter

class bayespy.plot.HintonPlotter (**kwargs)
 Plotter of the Hinton diagram of a node

See also:

hinton

```
__init__(**kwargs)
```

Methods

```
__init__(**kwargs)
```

bayespy.plot.HintonPlotter.__init__

```
HintonPlotter.__init__(**kwargs)
```

bayespy.plot.FunctionPlotter

class bayespy.plot.FunctionPlotter (**kwargs)
 Plotter of a node as a 1-dimensional function

See also:

plot

```
__init__(**kwargs)
```

Methods

```
__init__(**kwargs)
```

bayespy.plot.FunctionPlotter.__init__

```
FunctionPlotter.__init__(**kwargs)
```

bayespy.plot.GaussianTimeseriesPlotter

class bayespy.plot.**GaussianTimeseriesPlotter** (**kwargs)
 Plotter of a Gaussian node as a timeseries

__init__ (**kwargs)

Methods

__init__ (**kwargs)

bayespy.plot.GaussianTimeseriesPlotter.__init__

GaussianTimeseriesPlotter.**__init__** (**kwargs)

bayespy.plot.CategoricalMarkovChainPlotter

class bayespy.plot.**CategoricalMarkovChainPlotter** (**kwargs)
 Plotter of a Categorical timeseries

__init__ (**kwargs)

Methods

__init__ (**kwargs)

bayespy.plot.CategoricalMarkovChainPlotter.__init__

CategoricalMarkovChainPlotter.**__init__** (**kwargs)

DEVELOPER API

This chapter contains API specifications which are relevant to BayesPy developers and contributors.

6.1 Developer nodes

The following base classes are useful if writing new nodes:

| | |
|---|--|
| <code>node.Node(*parents, **kwargs)</code> | Base class for all nodes. |
| <code>stochastic.Stochastic(*args[, initialize, dims])</code> | Base class for nodes that are stochastic. |
| <code>expfamily.ExponentialFamily(*args, **kwargs)</code> | A base class for nodes using natural parameterization <i>phi</i> . |
| <code>deterministic.Deterministic(*args, **kwargs)</code> | Base class for deterministic nodes. |

6.1.1 bayespy.inference.vmp.nodes.node.Node

class bayespy.inference.vmp.nodes.node.**Node** (**parents, **kwargs*)

Base class for all nodes.

mask dims plates parents children name

Sub-classes must implement: 1. For computing the message to children:

`get_moments(self):`

2. For computing the message to parents: `_get_message_and_mask_to_parent(self, index)`

Sub-classes may need to re-implement: 1. If they manipulate plates:

`_compute_mask_to_parent(index, mask)` `_plates_to_parent(self, index)` `_plates_from_parent(self, index)`

`__init__` (**parents, **kwargs*)

Methods

`__init__` (**parents, **kwargs*)

`add_plate_axis` (*to_plate*)

`broadcasting_multiplier` (*plates, *args*)

`delete` ()

Delete this node and the children

`get_mask` ()

Continued on next page

Table 6.2 – continued from previous page

| | |
|--|--|
| <code>get_moments()</code> | |
| <code>get_shape(ind)</code> | |
| <code>has_plotter()</code> | Return True if the node has a plotter |
| <code>move_plates(from_plate, to_plate)</code> | |
| <code>plot([fig])</code> | Plot the node distribution using the plotter of the node |
| <code>set_plotter(plotter)</code> | |

bayespy.inference.vmp.nodes.node.Node.__init__

`Node.__init__ (*parents, **kwargs)`

bayespy.inference.vmp.nodes.node.Node.add_plate_axis

`Node.add_plate_axis (to_plate)`

bayespy.inference.vmp.nodes.node.Node.broadcasting_multiplier

static `Node.broadcasting_multiplier (plates, *args)`

bayespy.inference.vmp.nodes.node.Node.delete

`Node.delete ()`
Delete this node and the children

bayespy.inference.vmp.nodes.node.Node.get_mask

`Node.get_mask ()`

bayespy.inference.vmp.nodes.node.Node.get_moments

`Node.get_moments ()`

bayespy.inference.vmp.nodes.node.Node.get_shape

`Node.get_shape (ind)`

bayespy.inference.vmp.nodes.node.Node.has_plotter

`Node.has_plotter ()`
Return True if the node has a plotter

bayespy.inference.vmp.nodes.node.Node.move_plates

`Node.move_plates (from_plate, to_plate)`

bayespy.inference.vmp.nodes.node.Node.plot

Node.**plot** (*fig=None, **kwargs*)

Plot the node distribution using the plotter of the node

Because the distributions are in general very difficult to plot, the user must specify some functions which performs the plotting as wanted. See, for instance, bayespy.plot.plotting for available plotters, that is, functions that perform plotting for a node.

bayespy.inference.vmp.nodes.node.Node.set_plotter

Node.**set_plotter** (*plotter*)

Attributes

| | |
|--------------------------|--|
| <i>plates</i> | |
| <i>plates_multiplier</i> | Plate multiplier is applied to messages to parents |

bayespy.inference.vmp.nodes.node.Node.plates

Node.**plates** = None

bayespy.inference.vmp.nodes.node.Node.plates_multiplier

Node.**plates_multiplier**

Plate multiplier is applied to messages to parents

6.1.2 bayespy.inference.vmp.nodes.stochastic.Stochastic

class bayespy.inference.vmp.nodes.stochastic.**Stochastic** (**args, initialize=True, dims=None, **kwargs*)

Base class for nodes that are stochastic.

u observed

Sub-classes must implement: `_compute_message_to_parent(parent, index, u_self, *u_parents)`
`_update_distribution_and_lowerbound(self, m, *u)` `lowerbound(self)` `_compute_dims` `initialize_from_prior()`

If you want to be able to observe the variable: `_compute_fixed_moments_and_f`

Sub-classes may need to re-implement: 1. If they manipulate plates:

`_compute_mask_to_parent(index, mask)` `_compute_plates_to_parent(self, index, plates)`
`_compute_plates_from_parent(self, index, plates)`

`__init__` (**args, initialize=True, dims=None, **kwargs*)

Methods

`__init__` (**args[, initialize, dims]*)

Continued on next page

Table 6.4 – continued from previous page

| | |
|----------------------------------|---|
| <code>get_moments()</code> | |
| <code>load(filename)</code> | |
| <code>lowerbound()</code> | |
| <code>observe(x[, mask])</code> | Fix moments, compute f and propagate mask. |
| <code>random()</code> | Draw a random sample from the distribution. |
| <code>save(filename)</code> | |
| <code>show()</code> | Print the distribution using standard parameterization. |
| <code>unobserve()</code> | |
| <code>update([annealing])</code> | |

bayespy.inference.vmp.nodes.stochastic.Stochastic.__init__

`Stochastic.__init__(*args, initialize=True, dims=None, **kwargs)`

bayespy.inference.vmp.nodes.stochastic.Stochastic.get_moments

`Stochastic.get_moments()`

bayespy.inference.vmp.nodes.stochastic.Stochastic.load

`Stochastic.load(filename)`

bayespy.inference.vmp.nodes.stochastic.Stochastic.lowerbound

`Stochastic.lowerbound()`

bayespy.inference.vmp.nodes.stochastic.Stochastic.observe

`Stochastic.observe(x, mask=True)`
Fix moments, compute f and propagate mask.

bayespy.inference.vmp.nodes.stochastic.Stochastic.random

`Stochastic.random()`
Draw a random sample from the distribution.

bayespy.inference.vmp.nodes.stochastic.Stochastic.save

`Stochastic.save(filename)`

bayespy.inference.vmp.nodes.stochastic.Stochastic.show

`Stochastic.show()`
Print the distribution using standard parameterization.

bayespy.inference.vmp.nodes.stochastic.Stochastic.unobserve

`Stochastic.unobserve()`

bayespy.inference.vmp.nodes.stochastic.Stochastic.update

`Stochastic.update (annealing=1.0)`

Attributes

| | |
|--------------------------------|--|
| <code>plates</code> | |
| <code>plates_multiplier</code> | Plate multiplier is applied to messages to parents |

bayespy.inference.vmp.nodes.stochastic.Stochastic.plates

`Stochastic.plates = None`

bayespy.inference.vmp.nodes.stochastic.Stochastic.plates_multiplier

`Stochastic.plates_multiplier`
Plate multiplier is applied to messages to parents

6.1.3 bayespy.inference.vmp.nodes.expfamily.ExponentialFamily

class `bayespy.inference.vmp.nodes.expfamily.ExponentialFamily (*args, **kwargs)`

A base class for nodes using natural parameterization *phi*.

phi

Sub-classes must implement the following static methods: `_compute_message_to_parent(index, u_self, *u_parents)` `_compute_phi_from_parents(*u_parents, mask)` `_compute_moments_and_cgf(phi, mask)` `_compute_fixed_moments_and_f(x, mask=True)`

Sub-classes may need to re-implement: 1. If they manipulate plates:

`_compute_mask_to_parent(index, mask)` `_compute_plates_to_parent(self, index, plates)`
`_compute_plates_from_parent(self, index, plates)`

`__init__ (*args, **kwargs)`

Methods

| | |
|--|--|
| <code>__init__(*args, **kwargs)</code> | |
| <code>get_gradient(rg)</code> | Computes gradient with respect to the natural parameters. |
| <code>get_parameters()</code> | Return parameters of the VB distribution. |
| <code>get_riemannian_gradient()</code> | Computes the Riemannian/natural gradient. |
| <code>initialize_from_parameters(*args)</code> | |
| <code>initialize_from_prior()</code> | |
| <code>initialize_from_random()</code> | Set the variable to a random sample from the current distribution. |
| Continued on next page | |

Table 6.6 – continued from previous page

| | |
|---|---|
| <i>initialize_from_value</i> (x, *args) | |
| <i>logpdf</i> (X[, mask]) | Compute the log probability density function $Q(X)$ of this node. |
| <i>lower_bound_contribution</i> ([gradient, ...]) | Compute $E[\log p(X \text{parents}) - \log q(X)]$ |
| <i>observe</i> (x, *args[, mask]) | Fix moments, compute f and propagate mask. |
| <i>pdf</i> (X[, mask]) | Compute the probability density function of this node. |
| <i>random</i> () | Draw a random sample from the distribution. |
| <i>set_parameters</i> (x) | Set the parameters of the VB distribution. |

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.__init__

`ExponentialFamily.__init__(*args, **kwargs)`

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.get_gradient

`ExponentialFamily.get_gradient(rg)`

Computes gradient with respect to the natural parameters.

The function takes the Riemannian gradient as an input. This is for three reasons: 1) You probably want to use the Riemannian gradient anyway so this helps avoiding accidental use of this function. 2) The gradient is computed by using the Riemannian gradient and chain rules. 3) Probably you need both Riemannian and normal gradients anyway so you can provide it to this function to avoid re-computing it.

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.get_parameters

`ExponentialFamily.get_parameters()`

Return parameters of the VB distribution.

The parameters should be such that they can be used for optimization, that is, use log transformation for positive parameters.

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.get_riemannian_gradient

`ExponentialFamily.get_riemannian_gradient()`

Computes the Riemannian/natural gradient.

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.initialize_from_parameters

`ExponentialFamily.initialize_from_parameters(*args)`

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.initialize_from_prior

`ExponentialFamily.initialize_from_prior()`

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.initialize_from_random

`ExponentialFamily.initialize_from_random()`

Set the variable to a random sample from the current distribution.

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.initialize_from_value

`ExponentialFamily.initialize_from_value(x, *args)`

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.logpdf

`ExponentialFamily.logpdf(X, mask=True)`
 Compute the log probability density function $Q(X)$ of this node.

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.lower_bound_contribution

`ExponentialFamily.lower_bound_contribution(gradient=False, ignore_masked=True)`
 Compute $E[\log p(X|\text{parents}) - \log q(X)]$
 If deterministic annealing is used, the term $E[-\log q(X)]$ is divided by the annealing coefficient. That is, ϕ and cgf of q are multiplied by the temperature (inverse annealing coefficient).

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.observe

`ExponentialFamily.observe(x, *args, mask=True)`
 Fix moments, compute f and propagate mask.

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.pdf

`ExponentialFamily.pdf(X, mask=True)`
 Compute the probability density function of this node.

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.random

`ExponentialFamily.random()`
 Draw a random sample from the distribution.

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.set_parameters

`ExponentialFamily.set_parameters(x)`
 Set the parameters of the VB distribution.
 The parameters should be such that they can be used for optimization, that is, use log transformation for positive parameters.

Attributes

| | |
|--------------------------------|--|
| <code>dims</code> | |
| <code>plates</code> | |
| <code>plates_multiplier</code> | Plate multiplier is applied to messages to parents |

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.dims

ExponentialFamily.dims = None

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.plates

ExponentialFamily.plates = None

bayespy.inference.vmp.nodes.expfamily.ExponentialFamily.plates_multiplier

ExponentialFamily.plates_multiplier
Plate multiplier is applied to messages to parents

6.1.4 bayespy.inference.vmp.nodes.deterministic.Deterministic

class bayespy.inference.vmp.nodes.deterministic.Deterministic(*args, **kwargs)
Base class for deterministic nodes.

Sub-classes must implement: 1. For implementing the deterministic function:

 _compute_moments(self, *u)

2. One of the following options: a) Simple methods:

 _compute_message_to_parent(self, index, m, *u) not? _compute_mask_to_parent(self, index, mask)

(a) More control with: _compute_message_and_mask_to_parent(self, index, m, *u)

Sub-classes may need to re-implement: 1. If they manipulate plates:

 _compute_mask_to_parent(index, mask) _compute_plates_to_parent(self, index, plates)
 _compute_plates_from_parent(self, index, plates)

__init__(*args, **kwargs)

Methods

| |
|--------------------------------------|
| __init__(*args, **kwargs) |
| get_moments() |
| lower_bound_contribution([gradient]) |

bayespy.inference.vmp.nodes.deterministic.Deterministic.__init__

Deterministic.__init__(*args, **kwargs)

bayespy.inference.vmp.nodes.deterministic.Deterministic.get_moments

Deterministic.get_moments()

bayespy.inference.vmp.nodes.deterministic.Deterministic.lower_bound_contribution

Deterministic.lower_bound_contribution (*gradient=False, **kwargs*)

Attributes

| | |
|--------------------------------|--|
| <code>plates</code> | |
| <code>plates_multiplier</code> | Plate multiplier is applied to messages to parents |

bayespy.inference.vmp.nodes.deterministic.Deterministic.plates

Deterministic.plates = None

bayespy.inference.vmp.nodes.deterministic.Deterministic.plates_multiplier

Deterministic.plates_multiplier

Plate multiplier is applied to messages to parents

The following nodes are examples of special nodes that remain hidden for the user although they are often implicitly used:

| | |
|--|--|
| <code>constant.Constant(moments, x, **kwargs)</code> | Node for presenting constant values. |
| <code>gaussian.GaussianToGaussianGammaISO(X, **kwargs)</code> | Converter for Gaussian moments to Gaussian-gamma isotropic |
| <code>gaussian.GaussianGammaISOToGaussianGammaARD(X, ...)</code> | Converter for Gaussian-gamma ISO moments to Gaussian-g |
| <code>gaussian.GaussianGammaARDToGaussianWishart(...)</code> | |
| <code>gaussian.WrapToGaussianGammaISO(*parents, ...)</code> | |
| <code>gaussian.WrapToGaussianGammaARD(mu_alpha, ...)</code> | |
| <code>gaussian.WrapToGaussianWishart(X, Lambda, ...)</code> | Wraps Gaussian and Wishart nodes into a Gaussian-Wishart |

6.1.5 bayespy.inference.vmp.nodes.constant.Constant

class bayespy.inference.vmp.nodes.constant.Constant (*moments, x, **kwargs*)

Node for presenting constant values.

The node wraps arrays into proper node type.

`__init__` (*moments, x, **kwargs*)

Methods

| |
|---|
| <code>__init__(moments, x, **kwargs)</code> |
| <code>get_moments()</code> |
| <code>lower_bound_contribution([gradient])</code> |
| <code>set_value(x)</code> |

bayespy.inference.vmp.nodes.constant.Constant.__init__

Constant.__init__ (*moments, x, **kwargs*)

`bayespy.inference.vmp.nodes.constant.Constant.get_moments`

`Constant.get_moments()`

`bayespy.inference.vmp.nodes.constant.Constant.lower_bound_contribution`

`Constant.lower_bound_contribution (gradient=False, **kwargs)`

`bayespy.inference.vmp.nodes.constant.Constant.set_value`

`Constant.set_value(x)`

Attributes

| | |
|--------------------------------|--|
| <code>plates</code> | |
| <code>plates_multiplier</code> | Plate multiplier is applied to messages to parents |

`bayespy.inference.vmp.nodes.constant.Constant.plates`

`Constant.plates = None`

`bayespy.inference.vmp.nodes.constant.Constant.plates_multiplier`

`Constant.plates_multiplier`
Plate multiplier is applied to messages to parents

6.1.6 `bayespy.inference.vmp.nodes.gaussian.GaussianToGaussianGammaISO`

`class bayespy.inference.vmp.nodes.gaussian.GaussianToGaussianGammaISO (X, **kwargs)`

Converter for Gaussian moments to Gaussian-gamma isotropic moments

Combines the Gaussian moments with gamma moments for a fixed value 1.

`__init__ (X, **kwargs)`

Methods

`__init__(X, **kwargs)`

`bayespy.inference.vmp.nodes.gaussian.GaussianToGaussianGammaISO.__init__`

`GaussianToGaussianGammaISO.__init__ (X, **kwargs)`

Attributes

| | |
|--------------------------|--|
| <i>plates</i> | |
| <i>plates_multiplier</i> | Plate multiplier is applied to messages to parents |

bayespy.inference.vmp.nodes.gaussian.GaussianToGaussianGammaISO.plates

GaussianToGaussianGammaISO.**plates** = None

bayespy.inference.vmp.nodes.gaussian.GaussianToGaussianGammaISO.plates_multiplier

GaussianToGaussianGammaISO.**plates_multiplier**
Plate multiplier is applied to messages to parents

6.1.7 bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOToGaussianGammaARD

class bayespy.inference.vmp.nodes.gaussian.**GaussianGammaISOToGaussianGammaARD** (*X*,
***kwargs*)

Converter for Gaussian-gamma ISO moments to Gaussian-gamma ARD moments

__init__ (*X*, ***kwargs*)

Methods

__init__ (*X*, ***kwargs*)

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOToGaussianGammaARD.__init__

GaussianGammaISOToGaussianGammaARD.**__init__** (*X*, ***kwargs*)

Attributes

| | |
|--------------------------|--|
| <i>plates</i> | |
| <i>plates_multiplier</i> | Plate multiplier is applied to messages to parents |

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOToGaussianGammaARD.plates

GaussianGammaISOToGaussianGammaARD.**plates** = None

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOToGaussianGammaARD.plates_multiplier

GaussianGammaISOToGaussianGammaARD.**plates_multiplier**
Plate multiplier is applied to messages to parents

6.1.8 bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDToGaussianWishart

class bayespy.inference.vmp.nodes.gaussian.**GaussianGammaARDToGaussianWishart** (*X_alpha*,
***kwargs*)

__init__ (*X_alpha*, ***kwargs*)

Methods

__init__ (*X_alpha*, ***kwargs*)

bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDToGaussianWishart.__init__

GaussianGammaARDToGaussianWishart.**__init__** (*X_alpha*, ***kwargs*)

Attributes

plates
plates_multiplier Plate multiplier is applied to messages to parents

bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDToGaussianWishart.plates

GaussianGammaARDToGaussianWishart.**plates** = None

bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDToGaussianWishart.plates_multiplier

GaussianGammaARDToGaussianWishart.**plates_multiplier**
Plate multiplier is applied to messages to parents

6.1.9 bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaISO

class bayespy.inference.vmp.nodes.gaussian.**WrapToGaussianGammaISO** (**parents*,
***kwargs*)

__init__ (**parents*, ***kwargs*)

Methods

Attributes

plates
plates_multiplier Plate multiplier is applied to messages to parents

bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaISO.plates

WrapToGaussianGammaISO.**plates** = None

bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaISO.plates_multiplier

WrapToGaussianGammaISO.**plates_multiplier**
Plate multiplier is applied to messages to parents

6.1.10 bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaARD

class bayespy.inference.vmp.nodes.gaussian.**WrapToGaussianGammaARD** (*mu_alpha*, *tau*,
***kwargs*)

__init__ (*mu_alpha*, *tau*, ***kwargs*)

Methods

__init__ (*mu_alpha*, *tau*, ***kwargs*)

bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaARD.__init__

WrapToGaussianGammaARD.**__init__** (*mu_alpha*, *tau*, ***kwargs*)

Attributes

| | |
|--------------------------|--|
| <i>plates</i> | |
| <i>plates_multiplier</i> | Plate multiplier is applied to messages to parents |

bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaARD.plates

WrapToGaussianGammaARD.**plates** = None

bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaARD.plates_multiplier

WrapToGaussianGammaARD.**plates_multiplier**
Plate multiplier is applied to messages to parents

6.1.11 bayespy.inference.vmp.nodes.gaussian.WrapToGaussianWishart

class bayespy.inference.vmp.nodes.gaussian.**WrapToGaussianWishart** (*X*, *Lambda*,
***kwargs*)

Wraps Gaussian and Wishart nodes into a Gaussian-Wishart node.

The following node combinations can be wrapped:

- Gaussian and Wishart

- Gaussian-gamma and Wishart
- Gaussian-Wishart and gamma

`__init__(X, Lambda, **kwargs)`

Methods

`__init__(X, Lambda, **kwargs)`

`bayespy.inference.vmp.nodes.gaussian.WrapToGaussianWishart.__init__`

`WrapToGaussianWishart.__init__(X, Lambda, **kwargs)`

Attributes

| | |
|--------------------------------|--|
| <code>plates</code> | |
| <code>plates_multiplier</code> | Plate multiplier is applied to messages to parents |

`bayespy.inference.vmp.nodes.gaussian.WrapToGaussianWishart.plates`

`WrapToGaussianWishart.plates = None`

`bayespy.inference.vmp.nodes.gaussian.WrapToGaussianWishart.plates_multiplier`

`WrapToGaussianWishart.plates_multiplier`
Plate multiplier is applied to messages to parents

6.2 Moments

| | |
|--|---|
| <code>node.Moments</code> | Base class for defining the expectation of the su |
| <code>gaussian.GaussianMoments(ndim)</code> | Class for the moments of Gaussian variables. |
| <code>gaussian_markov_chain.GaussianMarkovChainMoments</code> | |
| <code>gaussian.GaussianGammaISOMoments(ndim)</code> | Class for the moments of Gaussian-gamma-ISO |
| <code>gaussian.GaussianGammaARDMoments(ndim)</code> | Class for the moments of Gaussian-gamma-AR |
| <code>gaussian.GaussianWishartMoments</code> | Class for the moments of Gaussian-Wishart vari |
| <code>gamma.GammaMoments</code> | Class for the moments of gamma variables. |
| <code>wishart.WishartMoments</code> | |
| <code>beta.BetaMoments</code> | Class for the moments of beta variables. |
| <code>dirichlet.DirichletMoments</code> | Class for the moments of Dirichlet variables. |
| <code>bernoulli.BernoulliMoments()</code> | Class for the moments of Bernoulli variables. |
| <code>binomial.BinomialMoments(N)</code> | Class for the moments of binomial variables |
| <code>categorical.CategoricalMoments(categories)</code> | Class for the moments of categorical variables. |
| <code>categorical_markov_chain.CategoricalMarkovChainMoments(...)</code> | Class for the moments of categorical Markov ch |
| <code>multinomial.MultinomialMoments</code> | Class for the moments of multinomial variables |
| <code>poisson.PoissonMoments</code> | Class for the moments of Poisson variables |

6.2.1 bayespy.inference.vmp.nodes.node.Moments

class bayespy.inference.vmp.nodes.node.Moments

Base class for defining the expectation of the sufficient statistics.

The benefits:

- Write statistic-specific features in one place only. For instance, covariance from Gaussian message.
- Different nodes may have identically defined statistic so you need to implement related features only once. For instance, Gaussian and GaussianARD differ on the prior but the moments are the same.
- General processing nodes which do not change the type of the moments may “inherit” the features from the parent node. For instance, slicing operator.
- Conversions can be done easily in both of the above cases if the message conversion is defined in the moments class. For instance, GaussianMarkovChain to Gaussian and VaryingGaussianMarkovChain to Gaussian.

__init__ ()

Initialize self. See help(type(self)) for accurate signature.

Methods

| | |
|--|---|
| <code>add_converter</code> (moments_to, converter) | |
| <code>compute_dims_from_values</code> (x) | |
| <code>compute_fixed_moments</code> (x) | |
| <code>get_converter</code> (moments_to) | Finds conversion to another moments type if possible. |

bayespy.inference.vmp.nodes.node.Moments.add_converter

classmethod Moments.add_converter (moments_to, converter)

bayespy.inference.vmp.nodes.node.Moments.compute_dims_from_values

Moments.compute_dims_from_values (x)

bayespy.inference.vmp.nodes.node.Moments.compute_fixed_moments

Moments.compute_fixed_moments (x)

bayespy.inference.vmp.nodes.node.Moments.get_converter

Moments.get_converter (moments_to)

Finds conversion to another moments type if possible.

Note that a conversion from moments A to moments B may require intermediate conversions. For instance: A->C->D->B. This method finds the path which uses the least amount of conversions and returns that path as a single conversion. If no conversion path is available, an error is raised.

The search algorithm starts from the original moments class and applies all possible converters to get a new list of moments classes. This list is extended by adding recursively all parent classes because their converters are applicable. Then, all possible converters are applied to this list to get a new list of current moments classes. This is iterated until the algorithm hits the target moments class or its subclass.

6.2.2 bayespy.inference.vmp.nodes.gaussian.GaussianMoments

class bayespy.inference.vmp.nodes.gaussian.**GaussianMoments** (*ndim*)
 Class for the moments of Gaussian variables.

__init__ (*ndim*)

Methods

| | |
|--|--|
| <i>__init__</i> (<i>ndim</i>) | |
| <i>compute_dims_from_values</i> (<i>x</i>) | Return the shape of the moments for a fixed value. |
| <i>compute_fixed_moments</i> (<i>x</i>) | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.gaussian.GaussianMoments.__init__

GaussianMoments.**__init__** (*ndim*)

bayespy.inference.vmp.nodes.gaussian.GaussianMoments.compute_dims_from_values

GaussianMoments.**compute_dims_from_values** (*x*)
 Return the shape of the moments for a fixed value.

bayespy.inference.vmp.nodes.gaussian.GaussianMoments.compute_fixed_moments

GaussianMoments.**compute_fixed_moments** (*x*)
 Compute the moments for a fixed value

6.2.3 bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainMoments

class bayespy.inference.vmp.nodes.gaussian_markov_chain.**GaussianMarkovChainMoments**

__init__ ()
 Initialize self. See help(type(self)) for accurate signature.

Methods

| |
|---|
| <i>compute_fixed_moments</i> (<i>x</i>) |
|---|

bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainMoments.compute_fixed_moments

GaussianMarkovChainMoments.**compute_fixed_moments** (*x*)

6.2.4 bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOMoments

class bayespy.inference.vmp.nodes.gaussian.**GaussianGammaISOMoments** (*ndim*)
 Class for the moments of Gaussian-gamma-ISO variables.


```
__init__(ndim)
    Create moments object for Gaussian-gamma isotropic variables
    ndim=0: scalar ndim=1: vector ndim=2: matrix ...
```

Methods

| | |
|---|--|
| <code>__init__(ndim)</code> | Create moments object for Gaussian-gamma isotropic variables |
| <code>compute_dims_from_values(x, alpha)</code> | Return the shape of the moments for a fixed value. |
| <code>compute_fixed_moments(x, alpha)</code> | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOMoments.__init__

```
GaussianGammaISOMoments.__init__(ndim)
    Create moments object for Gaussian-gamma isotropic variables
    ndim=0: scalar ndim=1: vector ndim=2: matrix ...
```

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOMoments.compute_dims_from_values

```
GaussianGammaISOMoments.compute_dims_from_values(x, alpha)
    Return the shape of the moments for a fixed value.
```

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOMoments.compute_fixed_moments

```
GaussianGammaISOMoments.compute_fixed_moments(x, alpha)
    Compute the moments for a fixed value
    x is a mean vector. alpha is a precision scale
```

6.2.5 bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDMoments

```
class bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDMoments(ndim)
    Class for the moments of Gaussian-gamma-ARD variables.
```

```
__init__(ndim)
    Create moments object for Gaussian-gamma isotropic variables
    ndim=0: scalar ndim=1: vector ndim=2: matrix ...
```

Methods

| | |
|---|--|
| <code>__init__(ndim)</code> | Create moments object for Gaussian-gamma isotropic variables |
| <code>compute_dims_from_values(x, alpha)</code> | Return the shape of the moments for a fixed value. |
| <code>compute_fixed_moments(x, alpha)</code> | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDMoments.__init__

```
GaussianGammaARDMoments.__init__(ndim)
    Create moments object for Gaussian-gamma isotropic variables
```

ndim=0: scalar ndim=1: vector ndim=2: matrix ...

bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDMoments.compute_dims_from_values

`GaussianGammaARDMoments.compute_dims_from_values` (*x*, *alpha*)
Return the shape of the moments for a fixed value.

bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDMoments.compute_fixed_moments

`GaussianGammaARDMoments.compute_fixed_moments` (*x*, *alpha*)
Compute the moments for a fixed value
x is a mean vector. *alpha* is a precision scale

6.2.6 bayespy.inference.vmp.nodes.gaussian.GaussianWishartMoments

class `bayespy.inference.vmp.nodes.gaussian.GaussianWishartMoments`
Class for the moments of Gaussian-Wishart variables.

`__init__` ()
Initialize self. See `help(type(self))` for accurate signature.

Methods

| | |
|--|--|
| <code>compute_dims_from_values</code> (<i>x</i> , <i>Lambda</i>) | Return the shape of the moments for a fixed value. |
| <code>compute_fixed_moments</code> (<i>x</i> , <i>Lambda</i>) | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.gaussian.GaussianWishartMoments.compute_dims_from_values

`GaussianWishartMoments.compute_dims_from_values` (*x*, *Lambda*)
Return the shape of the moments for a fixed value.

bayespy.inference.vmp.nodes.gaussian.GaussianWishartMoments.compute_fixed_moments

`GaussianWishartMoments.compute_fixed_moments` (*x*, *Lambda*)
Compute the moments for a fixed value
x is a vector. *Lambda* is a precision matrix

6.2.7 bayespy.inference.vmp.nodes.gamma.GammaMoments

class `bayespy.inference.vmp.nodes.gamma.GammaMoments`
Class for the moments of gamma variables.

`__init__` ()
Initialize self. See `help(type(self))` for accurate signature.

Methods

| | |
|--|--|
| <code>compute_dims_from_values(x)</code> | Return the shape of the moments for a fixed value. |
| <code>compute_fixed_moments(x)</code> | Compute the moments for a fixed value |

`bayespy.inference.vmp.nodes.gamma.GammaMoments.compute_dims_from_values`

`GammaMoments.compute_dims_from_values(x)`
Return the shape of the moments for a fixed value.

`bayespy.inference.vmp.nodes.gamma.GammaMoments.compute_fixed_moments`

`GammaMoments.compute_fixed_moments(x)`
Compute the moments for a fixed value

6.2.8 `bayespy.inference.vmp.nodes.wishart.WishartMoments`

`class bayespy.inference.vmp.nodes.wishart.WishartMoments`

`__init__()`
Initialize self. See `help(type(self))` for accurate signature.

Methods

| | |
|--|--------------------------------------|
| <code>compute_dims_from_values(x)</code> | Compute the dimensions of phi and u. |
| <code>compute_fixed_moments(Lambda)</code> | Compute moments for fixed x. |

`bayespy.inference.vmp.nodes.wishart.WishartMoments.compute_dims_from_values`

`WishartMoments.compute_dims_from_values(x)`
Compute the dimensions of phi and u.

`bayespy.inference.vmp.nodes.wishart.WishartMoments.compute_fixed_moments`

`WishartMoments.compute_fixed_moments(Lambda)`
Compute moments for fixed x.

6.2.9 `bayespy.inference.vmp.nodes.beta.BetaMoments`

`class bayespy.inference.vmp.nodes.beta.BetaMoments`
Class for the moments of beta variables.

`__init__()`
Initialize self. See `help(type(self))` for accurate signature.

Methods

| | |
|--|--|
| <code>compute_dims_from_values(p)</code> | Return the shape of the moments for a fixed value. |
| <code>compute_fixed_moments(p)</code> | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.beta.BetaMoments.compute_dims_from_values

`BetaMoments.compute_dims_from_values(p)`
Return the shape of the moments for a fixed value.

bayespy.inference.vmp.nodes.beta.BetaMoments.compute_fixed_moments

`BetaMoments.compute_fixed_moments(p)`
Compute the moments for a fixed value

6.2.10 bayespy.inference.vmp.nodes.dirichlet.DirichletMoments

class `bayespy.inference.vmp.nodes.dirichlet.DirichletMoments`
Class for the moments of Dirichlet variables.

`__init__()`
Initialize self. See `help(type(self))` for accurate signature.

Methods

| | |
|--|--|
| <code>compute_dims_from_values(x)</code> | Return the shape of the moments for a fixed value. |
| <code>compute_fixed_moments(p)</code> | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.dirichlet.DirichletMoments.compute_dims_from_values

`DirichletMoments.compute_dims_from_values(x)`
Return the shape of the moments for a fixed value.

bayespy.inference.vmp.nodes.dirichlet.DirichletMoments.compute_fixed_moments

`DirichletMoments.compute_fixed_moments(p)`
Compute the moments for a fixed value

6.2.11 bayespy.inference.vmp.nodes.bernoulli.BernoulliMoments

class `bayespy.inference.vmp.nodes.bernoulli.BernoulliMoments`
Class for the moments of Bernoulli variables.

`__init__()`

Methods

`__init__()`

bayespy.inference.vmp.nodes.bernoulli.BernoulliMoments.__init__

BernoulliMoments.__init__()

6.2.12 bayespy.inference.vmp.nodes.binomial.BinomialMoments

class bayespy.inference.vmp.nodes.binomial.**BinomialMoments** (*N*)

Class for the moments of binomial variables

__init__(*N*)

Methods

| | |
|--------------------------------------|--|
| __init__(<i>N</i>) | |
| compute_dims_from_values(<i>x</i>) | Return the shape of the moments for a fixed value. |
| compute_fixed_moments(<i>x</i>) | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.binomial.BinomialMoments.__init__

BinomialMoments.__init__(*N*)

bayespy.inference.vmp.nodes.binomial.BinomialMoments.compute_dims_from_values

BinomialMoments.compute_dims_from_values(*x*)

Return the shape of the moments for a fixed value.

The realizations are scalars, thus the shape of the moment is ().

bayespy.inference.vmp.nodes.binomial.BinomialMoments.compute_fixed_moments

BinomialMoments.compute_fixed_moments(*x*)

Compute the moments for a fixed value

6.2.13 bayespy.inference.vmp.nodes.categorical.CategoricalMoments

class bayespy.inference.vmp.nodes.categorical.**CategoricalMoments** (*categories*)

Class for the moments of categorical variables.

__init__(*categories*)

Create moments object for categorical variables

Methods

| | |
|--------------------------------------|--|
| __init__(<i>categories</i>) | Create moments object for categorical variables |
| compute_dims_from_values(<i>x</i>) | Return the shape of the moments for a fixed value. |
| compute_fixed_moments(<i>x</i>) | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.categorical.CategoricalMoments.__init__

`CategoricalMoments.__init__(categories)`
 Create moments object for categorical variables

bayespy.inference.vmp.nodes.categorical.CategoricalMoments.compute_dims_from_values

`CategoricalMoments.compute_dims_from_values(x)`
 Return the shape of the moments for a fixed value.
 The observations are scalar.

bayespy.inference.vmp.nodes.categorical.CategoricalMoments.compute_fixed_moments

`CategoricalMoments.compute_fixed_moments(x)`
 Compute the moments for a fixed value

6.2.14 bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainMoments

class `bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainMoments(categories)`
 Class for the moments of categorical Markov chain variables.

`__init__(categories)`
 Create moments object for categorical Markov chain variables.

Methods

| | |
|--|---|
| <code>__init__(categories)</code> | Create moments object for categorical Markov chain variables. |
| <code>compute_dims_from_values(x)</code> | Return the shape of the moments for a fixed value. |
| <code>compute_fixed_moments(x)</code> | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainMoments.__init__

`CategoricalMarkovChainMoments.__init__(categories)`
 Create moments object for categorical Markov chain variables.

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainMoments.compute_dims_from_values

`CategoricalMarkovChainMoments.compute_dims_from_values(x)`
 Return the shape of the moments for a fixed value.

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainMoments.compute_fixed_moments

`CategoricalMarkovChainMoments.compute_fixed_moments(x)`
 Compute the moments for a fixed value

6.2.15 bayespy.inference.vmp.nodes.multinomial.MultinomialMoments

class bayespy.inference.vmp.nodes.multinomial.**MultinomialMoments**

Class for the moments of multinomial variables.

__init__()

Initialize self. See help(type(self)) for accurate signature.

Methods

| | |
|-------------------------------------|--|
| <i>compute_dims_from_values</i> (x) | Return the shape of the moments for a fixed value. |
| <i>compute_fixed_moments</i> (x) | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.multinomial.MultinomialMoments.compute_dims_from_values

MultinomialMoments.compute_dims_from_values(x)

Return the shape of the moments for a fixed value.

bayespy.inference.vmp.nodes.multinomial.MultinomialMoments.compute_fixed_moments

MultinomialMoments.compute_fixed_moments(x)

Compute the moments for a fixed value

x must be a vector of counts.

6.2.16 bayespy.inference.vmp.nodes.poisson.PoissonMoments

class bayespy.inference.vmp.nodes.poisson.**PoissonMoments**

Class for the moments of Poisson variables

__init__()

Initialize self. See help(type(self)) for accurate signature.

Methods

| | |
|-------------------------------------|--|
| <i>compute_dims_from_values</i> (x) | Return the shape of the moments for a fixed value. |
| <i>compute_fixed_moments</i> (x) | Compute the moments for a fixed value |

bayespy.inference.vmp.nodes.poisson.PoissonMoments.compute_dims_from_values

PoissonMoments.compute_dims_from_values(x)

Return the shape of the moments for a fixed value.

The realizations are scalars, thus the shape of the moment is ().

bayespy.inference.vmp.nodes.poisson.PoissonMoments.compute_fixed_moments

PoissonMoments.compute_fixed_moments(x)

Compute the moments for a fixed value

6.3 Distributions

| | |
|--|---|
| <code>stochastic.Distribution</code> | A base class for the VMP formulas of variables. |
| <code>expfamily.ExponentialFamilyDistribution</code> | Sub-classes implement distribution specific computations. |
| <code>gaussian.GaussianDistribution</code> | Class for the VMP formulas of variables. |
| <code>gaussian.GaussianARDDistribution(shape, ndim_mu)</code> | ... |
| <code>gaussian.GaussianGammaISODistribution</code> | Class for the VMP formulas of variables. |
| <code>gaussian.GaussianGammaARDDistribution()</code> | ... |
| <code>gaussian.GaussianWishartDistribution</code> | Class for the VMP formulas of variables. |
| <code>gaussian_markov_chain.GaussianMarkovChainDistribution(N, D)</code> | Sub-classes implement distribution specific computations. |
| <code>gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution(N, D, K)</code> | Sub-classes implement distribution specific computations. |
| <code>gaussian_markov_chain.VaryingGaussianMarkovChainDistribution(N, D)</code> | Sub-classes implement distribution specific computations. |
| <code>gamma.GammaDistribution</code> | Class for the VMP formulas of variables. |
| <code>wishart.WishartDistribution</code> | Sub-classes implement distribution specific computations. |
| <code>beta.BetaDistribution</code> | Class for the VMP formulas of variables. |
| <code>dirichlet.DirichletDistribution</code> | Class for the VMP formulas of variables. |
| <code>bernoulli.BernoulliDistribution()</code> | Class for the VMP formulas of variables. |
| <code>binomial.BinomialDistribution(N)</code> | Class for the VMP formulas of variables. |
| <code>categorical.CategoricalDistribution(categories)</code> | Class for the VMP formulas of variables. |
| <code>categorical_markov_chain.CategoricalMarkovChainDistribution(...)</code> | Class for the VMP formulas of variables. |
| <code>multinomial.MultinomialDistribution(trials)</code> | Class for the VMP formulas of variables. |
| <code>poisson.PoissonDistribution</code> | Class for the VMP formulas of variables. |

6.3.1 bayespy.inference.vmp.nodes.stochastic.Distribution

class `bayespy.inference.vmp.nodes.stochastic.Distribution`

A base class for the VMP formulas of variables.

Sub-classes implement distribution specific computations.

If a sub-class maps the plates differently, it needs to overload the following methods:

- `compute_mask_to_parent`
- `plates_to_parent`
- `plates_from_parent`

`__init__()`

Initialize self. See `help(type(self))` for accurate signature.

Methods

| | |
|--|---|
| <code>compute_mask_to_parent(index, mask)</code> | Maps the mask to the plates of a parent. |
| <code>compute_message_to_parent(parent, index, ...)</code> | Compute the message to a parent node. |
| <code>plates_from_parent(index, plates)</code> | Resolve the plate mapping from a parent. |
| <code>plates_to_parent(index, plates)</code> | Resolves the plate mapping to a parent. |
| <code>random(*params[, plates])</code> | Draw a random sample from the distribution. |

bayespy.inference.vmp.nodes.stochastic.Distribution.compute_mask_to_parent

`Distribution.compute_mask_to_parent` (*index, mask*)
Maps the mask to the plates of a parent.

bayespy.inference.vmp.nodes.stochastic.Distribution.compute_message_to_parent

`Distribution.compute_message_to_parent` (*parent, index, u_self, *u_parents*)
Compute the message to a parent node.

bayespy.inference.vmp.nodes.stochastic.Distribution.plates_from_parent

`Distribution.plates_from_parent` (*index, plates*)
Resolve the plate mapping from a parent.

Given the plates of a parent's moments, this method returns the plates that the moments has for this distribution.

bayespy.inference.vmp.nodes.stochastic.Distribution.plates_to_parent

`Distribution.plates_to_parent` (*index, plates*)
Resolves the plate mapping to a parent.

Given the plates of the node's moments, this method returns the plates that the message to a parent has for the parent's distribution.

bayespy.inference.vmp.nodes.stochastic.Distribution.random

`Distribution.random` (**params, plates=None*)
Draw a random sample from the distribution.

6.3.2 bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution

`class bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution`
Sub-classes implement distribution specific computations.

`__init__` ()
Initialize self. See help(type(self)) for accurate signature.

Methods

| | |
|--|---|
| <code>compute_cgf_from_parents(*u_parents)</code> | |
| <code>compute_fixed_moments_and_f(x[, mask])</code> | |
| <code>compute_gradient(g, u, phi)</code> | Compute the standard gradient with respect to the natural parameters. |
| <code>compute_logpdf(u, phi, g, f, ndims)</code> | Compute $E[\log p(X)]$ given $E[u]$, $E[\phi]$, $E[g]$ and $E[f]$. |
| <code>compute_message_to_parent(parent, index, ...)</code> | |
| <code>compute_moments_and_cgf(phi[, mask])</code> | |
| <code>compute_phi_from_parents(*u_parents[, mask])</code> | |

bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution.compute_cgf_from_parents

`ExponentialFamilyDistribution.compute_cgf_from_parents(*u_parents)`

bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution.compute_fixed_moments_and_f

`ExponentialFamilyDistribution.compute_fixed_moments_and_f(x, mask=True)`

bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution.compute_gradient

`ExponentialFamilyDistribution.compute_gradient(g, u, phi)`
 Compute the standard gradient with respect to the natural parameters.

bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution.compute_logpdf

`ExponentialFamilyDistribution.compute_logpdf(u, phi, g, f, ndims)`
 Compute $E[\log p(X)]$ given $E[u]$, $E[\phi]$, $E[g]$ and $E[f]$. Does not sum over plates.

bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution.compute_message_to_parent

`ExponentialFamilyDistribution.compute_message_to_parent(parent, index, u_self, *u_parents)`

bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution.compute_moments_and_cgf

`ExponentialFamilyDistribution.compute_moments_and_cgf(phi, mask=True)`

bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution.compute_phi_from_parents

`ExponentialFamilyDistribution.compute_phi_from_parents(*u_parents, mask=True)`

6.3.3 bayespy.inference.vmp.nodes.gaussian.GaussianDistribution

class bayespy.inference.vmp.nodes.gaussian.GaussianDistribution

Class for the VMP formulas of Gaussian variables.

Currently, supports only vector variables.

Notes

Message passing equations:

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}),$$

$$\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^D, \quad \boldsymbol{\Lambda} \in \mathbb{R}^{D \times D}, \quad \boldsymbol{\Lambda} \text{ symmetric positive definite}$$

$$\log \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Lambda}) = -\frac{1}{2} \mathbf{x}^T \boldsymbol{\Lambda} \mathbf{x} + \mathbf{x}^T \boldsymbol{\Lambda} \boldsymbol{\mu} - \frac{1}{2} \boldsymbol{\mu}^T \boldsymbol{\Lambda} \boldsymbol{\mu} + \frac{1}{2} \log |\boldsymbol{\Lambda}| - \frac{D}{2} \log(2\pi)$$

`--init--()`

Initialize self. See help(type(self)) for accurate signature.

Methods

| | |
|---|---|
| <code>compute_cgf_from_parents(u_mu_Lambda)</code> | Compute $E_{q(p)}[g(p)]$ |
| <code>compute_fixed_moments_and_f(x[, mask])</code> | Compute the moments and $f(x)$ for a fixed value. |
| <code>compute_gradient(g, u, phi)</code> | Compute the standard gradient with respect to the natural parameters. |
| <code>compute_message_to_parent(parent, index, u, ...)</code> | Compute the message to a parent node. |
| <code>compute_moments_and_cgf(phi[, mask])</code> | Compute the moments and $g(\phi)$. |
| <code>compute_phi_from_parents(u_mu_Lambda[, mask])</code> | Compute the natural parameter vector given parent moments. |
| <code>random(*phi[, plates])</code> | Draw a random sample from the distribution. |

`bayespy.inference.vmp.nodes.gaussian.GaussianDistribution.compute_cgf_from_parents`

`GaussianDistribution.compute_cgf_from_parents(u_mu_Lambda)`

Compute $E_{q(p)}[g(p)]$

$$g(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = -\frac{1}{2} \text{tr}(\boldsymbol{\mu} \boldsymbol{\mu}^T \boldsymbol{\Lambda}) + \frac{1}{2} \log |\boldsymbol{\Lambda}|$$

`bayespy.inference.vmp.nodes.gaussian.GaussianDistribution.compute_fixed_moments_and_f`

`GaussianDistribution.compute_fixed_moments_and_f(x, mask=True)`

Compute the moments and $f(x)$ for a fixed value.

$$\mathbf{u}(\mathbf{x}) = \begin{bmatrix} \mathbf{x} \\ \mathbf{x} \mathbf{x}^T \end{bmatrix}$$

$$f(\mathbf{x}) = -\frac{D}{2} \log(2\pi)$$

`bayespy.inference.vmp.nodes.gaussian.GaussianDistribution.compute_gradient`

`GaussianDistribution.compute_gradient(g, u, phi)`

Compute the standard gradient with respect to the natural parameters.

Gradient of the moments:

$$\begin{aligned} d\bar{\mathbf{u}} &= \begin{bmatrix} -\frac{1}{4}\phi_2^{-1}d\phi_2\phi_2^{-1}\phi_1\phi_1^T\phi_2^{-1} - \frac{1}{4}\phi_2^{-1}\phi_1\phi_1^T\phi_2^{-1}d\phi_2\phi_2^{-1} + \frac{1}{2}\phi_2^{-1}d\phi_2\phi_2^{-1} + \frac{1}{4}\phi_2^{-1}d\phi_1\phi_1^T\phi_2^{-1} + \frac{1}{4}\phi_2^{-1}\phi_1d\phi_1^T\phi_2^{-1} \end{bmatrix} \\ &= \begin{bmatrix} 2(\bar{u}_2 - \bar{u}_1\bar{u}_1^T)d\phi_2\bar{u}_1 + (\bar{u}_2 - \bar{u}_1\bar{u}_1^T)d\phi_1 \\ u_2d\phi_2u_2 - 2u_1u_1^Td\phi_2u_1u_1^T + 2(u_2 - u_1u_1^T)d\phi_1u_1^T \end{bmatrix} \end{aligned}$$

Standard gradient given the gradient with respect to the moments, that is, given the Riemannian gradient $\tilde{\nabla}$:

$$\nabla = \begin{bmatrix} (\bar{u}_2 - \bar{u}_1\bar{u}_1^T)\tilde{\nabla}_1 + 2(u_2 - u_1u_1^T)\tilde{\nabla}_2u_1 \\ (u_2 - u_1u_1^T)\tilde{\nabla}_1u_1^T + u_1\tilde{\nabla}_1^T(u_2 - u_1u_1^T) + 2u_2\tilde{\nabla}_2u_2 - 2u_1u_1^T\tilde{\nabla}_2u_1u_1^T \end{bmatrix}$$

bayespy.inference.vmp.nodes.gaussian.GaussianDistribution.compute_message_to_parent

`GaussianDistribution.compute_message_to_parent` (*parent, index, u, u_mu_Lambda*)
 Compute the message to a parent node.

$$\phi_{\mu}(\mathbf{x}, \Lambda) = \begin{bmatrix} \Lambda \mathbf{x} \\ -\frac{1}{2} \Lambda \end{bmatrix}$$

$$\phi_{\Lambda}(\mathbf{x}, \mu) = \begin{bmatrix} -\frac{1}{2} \mathbf{x} \mathbf{x}^T + \frac{1}{2} \mathbf{x} \mu^T + \frac{1}{2} \mu \mathbf{x}^T - \frac{1}{2} \mu \mu^T \\ \frac{1}{2} \end{bmatrix}$$

bayespy.inference.vmp.nodes.gaussian.GaussianDistribution.compute_moments_and_cgf

`GaussianDistribution.compute_moments_and_cgf` (*phi, mask=True*)
 Compute the moments and $g(\phi)$.

$$\bar{\mathbf{u}}(\phi) = \begin{bmatrix} -\frac{1}{2} \phi_2^{-1} \phi_1 \\ \frac{1}{4} \phi_2^{-1} \phi_1 \phi_1^T \phi_2^{-1} - \frac{1}{2} \phi_2^{-1} \end{bmatrix}$$

$$g_{\phi}(\phi) = \frac{1}{4} \phi_1^T \phi_2^{-1} \phi_1 + \frac{1}{2} \log | -2 \phi_2 |$$

bayespy.inference.vmp.nodes.gaussian.GaussianDistribution.compute_phi_from_parents

`GaussianDistribution.compute_phi_from_parents` (*u_mu_Lambda, mask=True*)
 Compute the natural parameter vector given parent moments.

$$\phi(\mu, \Lambda) = \begin{bmatrix} \Lambda \mu \\ -\frac{1}{2} \Lambda \end{bmatrix}$$

bayespy.inference.vmp.nodes.gaussian.GaussianDistribution.random

`GaussianDistribution.random` (**phi, plates=None*)
 Draw a random sample from the distribution.

6.3.4 bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution

class `bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution` (*shape, ndim_mu*)
 ...

Log probability density function:

$$\log p(x|\mu, \alpha) = -\frac{1}{2} x^T \text{diag}(\alpha) x + x^T \text{diag}(\alpha) \mu - \frac{1}{2} \mu^T \text{diag}(\alpha) \mu + \frac{1}{2} \sum_i \log \alpha_i - \frac{D}{2} \log(2\pi)$$

Parent has moments:

$$\begin{bmatrix} \alpha \circ \mu \\ \alpha \circ \mu \circ \mu \\ \alpha \\ \log(\alpha) \end{bmatrix}$$

`__init__` (*shape, ndim_mu*)

Methods

| | |
|---|---|
| <code>__init__(shape, ndim_mu)</code> | |
| <code>compute_cgf_from_parents(u_mu_alpha)</code> | Compute the value of the cumulant generating function. |
| <code>compute_fixed_moments_and_f(x[, mask])</code> | Compute $u(x)$ and $f(x)$ for given x . |
| <code>compute_gradient(g, u, phi)</code> | Compute the standard gradient with respect to the natural parameters. |
| <code>compute_mask_to_parent(index, mask)</code> | Maps the mask to the plates of a parent. |
| <code>compute_message_to_parent(parent, index, u, ...)</code> | ... |
| <code>compute_moments_and_cgf(phi[, mask])</code> | |
| <code>compute_phi_from_parents(u_mu_alpha[, mask])</code> | |
| <code>plates_from_parent(index, plates)</code> | Resolve the plate mapping from a parent. |
| <code>plates_to_parent(index, plates)</code> | Resolves the plate mapping to a parent. |
| <code>random(*phi[, plates])</code> | Draw a random sample from the Gaussian distribution. |

`bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.__init__`

`GaussianARDDistribution.__init__(shape, ndim_mu)`

`bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.compute_cgf_from_parents`

`GaussianARDDistribution.compute_cgf_from_parents(u_mu_alpha)`
 Compute the value of the cumulant generating function.

`bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.compute_fixed_moments_and_f`

`GaussianARDDistribution.compute_fixed_moments_and_f(x, mask=True)`
 Compute $u(x)$ and $f(x)$ for given x .

`bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.compute_gradient`

`GaussianARDDistribution.compute_gradient(g, u, phi)`
 Compute the standard gradient with respect to the natural parameters.

Gradient of the moments:

$$\begin{aligned} d\bar{u} &= \left[-\frac{1}{4}\phi_2^{-1}d\phi_2\phi_2^{-1}\phi_1\phi_1^T\phi_2^{-1} - \frac{1}{4}\phi_2^{-1}\phi_1\phi_1^T\phi_2^{-1}d\phi_2\phi_2^{-1} + \frac{1}{2}\phi_2^{-1}d\phi_2\phi_2^{-1} + \frac{1}{4}\phi_2^{-1}d\phi_1\phi_1^T\phi_2^{-1} + \frac{1}{4}\phi_2^{-1}\phi_1d\phi_1^T\phi_2^{-1} \right] \\ &= \left[\begin{array}{c} 2(\bar{u}_2 - \bar{u}_1\bar{u}_1^T)d\phi_2\bar{u}_1 + (\bar{u}_2 - \bar{u}_1\bar{u}_1^T)d\phi_1 \\ u_2d\phi_2u_2 - 2u_1u_1^Td\phi_2u_1u_1^T + 2(u_2 - u_1u_1^T)d\phi_1u_1^T \end{array} \right] \end{aligned}$$

Standard gradient given the gradient with respect to the moments, that is, given the Riemannian gradient $\tilde{\nabla}$:

$$\nabla = \left[\begin{array}{c} (\bar{u}_2 - \bar{u}_1\bar{u}_1^T)\tilde{\nabla}_1 + 2(u_2 - u_1u_1^T)\tilde{\nabla}_2u_1 \\ (u_2 - u_1u_1^T)\tilde{\nabla}_1u_1^T + u_1\tilde{\nabla}_1^T(u_2 - u_1u_1^T) + 2u_2\tilde{\nabla}_2u_2 - 2u_1u_1^T\tilde{\nabla}_2u_1u_1^T \end{array} \right]$$

`bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.compute_mask_to_parent`

`GaussianARDDistribution.compute_mask_to_parent(index, mask)`
 Maps the mask to the plates of a parent.

bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.compute_message_to_parent

`GaussianARDDistribution.compute_message_to_parent` (*parent, index, u, u_mu_alpha*)
...

$$m = \begin{bmatrix} x \\ [-\frac{1}{2}, \dots, -\frac{1}{2}] \\ -\frac{1}{2} \text{diag}(xx^T) \\ [\frac{1}{2}, \dots, \frac{1}{2}] \end{bmatrix}$$

bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.compute_moments_and_cgf

`GaussianARDDistribution.compute_moments_and_cgf` (*phi, mask=True*)

bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.compute_phi_from_parents

`GaussianARDDistribution.compute_phi_from_parents` (*u_mu_alpha, mask=True*)

bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.plates_from_parent

`GaussianARDDistribution.plates_from_parent` (*index, plates*)

Resolve the plate mapping from a parent.

Given the plates of a parent's moments, this method returns the plates that the moments has for this distribution.

bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.plates_to_parent

`GaussianARDDistribution.plates_to_parent` (*index, plates*)

Resolves the plate mapping to a parent.

Given the plates of the node's moments, this method returns the plates that the message to a parent has for the parent's distribution.

bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution.random

`GaussianARDDistribution.random` (**phi, plates=None*)

Draw a random sample from the Gaussian distribution.

6.3.5 bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution

class `bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution`

Class for the VMP formulas of Gaussian-Gamma-ISO variables.

Currently, supports only vector variables.

`__init__` ()

Initialize self. See `help(type(self))` for accurate signature.

Methods

| | |
|---|--|
| <code>compute_cgf_from_parents(u_mu_Lambda, u_a, u_b)</code> | Compute $E_{q(p)}[g(p)]$ |
| <code>compute_fixed_moments_and_f(x, alpha[, mask])</code> | Compute the moments and $f(x)$ for a fixed value. |
| <code>compute_message_to_parent(parent, index, u, ...)</code> | Compute the message to a parent node. |
| <code>compute_moments_and_cgf(phi[, mask])</code> | Compute the moments and $g(\phi)$. |
| <code>compute_phi_from_parents(u_mu_Lambda, u_a, u_b)</code> | Compute the natural parameter vector given parent moments. |
| <code>random(*params[, plates])</code> | Draw a random sample from the distribution. |

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution.compute_cgf_from_parents

`GaussianGammaISODistribution.compute_cgf_from_parents(u_mu_Lambda, u_a, u_b)`
 Compute $E_{q(p)}[g(p)]$

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution.compute_fixed_moments_and_f

`GaussianGammaISODistribution.compute_fixed_moments_and_f(x, alpha, mask=True)`
 Compute the moments and $f(x)$ for a fixed value.

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution.compute_message_to_parent

`GaussianGammaISODistribution.compute_message_to_parent(parent, index, u, u_mu_Lambda, u_a, u_b)`
 Compute the message to a parent node.

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution.compute_moments_and_cgf

`GaussianGammaISODistribution.compute_moments_and_cgf(phi, mask=True)`
 Compute the moments and $g(\phi)$.

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution.compute_phi_from_parents

`GaussianGammaISODistribution.compute_phi_from_parents(u_mu_Lambda, u_a, u_b, mask=True)`
 Compute the natural parameter vector given parent moments.

bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution.random

`GaussianGammaISODistribution.random(*params, plates=None)`
 Draw a random sample from the distribution.

6.3.6 bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDDistribution

class `bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDDistribution`

`__init__()`

Methods

```
__init__()
```

`bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDDistribution.__init__`

`GaussianGammaARDDistribution.__init__()`

6.3.7 `bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution`

class `bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution`

Class for the VMP formulas of Gaussian-Wishart variables.

Currently, supports only vector variables.

__init__()

Initialize self. See `help(type(self))` for accurate signature.

Methods

| | |
|---|--|
| <code>compute_cgf_from_parents(u_mu_alpha, u_V, u_n)</code> | Compute $E_{q(p)}[g(p)]$ |
| <code>compute_fixed_moments_and_f(x, Lambda[, mask])</code> | Compute the moments and $f(x)$ for a fixed value. |
| <code>compute_message_to_parent(parent, index, u, ...)</code> | Compute the message to a parent node. |
| <code>compute_moments_and_cgf(phi[, mask])</code> | Compute the moments and $g(\phi)$. |
| <code>compute_phi_from_parents(u_mu_alpha, u_V, u_n)</code> | Compute the natural parameter vector given parent moments. |
| <code>random(*params[, plates])</code> | Draw a random sample from the distribution. |

`bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution.compute_cgf_from_parents`

`GaussianWishartDistribution.compute_cgf_from_parents(u_mu_alpha, u_V, u_n)`

Compute $E_{q(p)}[g(p)]$

`bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution.compute_fixed_moments_and_f`

`GaussianWishartDistribution.compute_fixed_moments_and_f(x, Lambda, mask=True)`

Compute the moments and $f(x)$ for a fixed value.

`bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution.compute_message_to_parent`

`GaussianWishartDistribution.compute_message_to_parent(parent, index, u, u_mu_alpha, u_V, u_n)`

Compute the message to a parent node.

`bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution.compute_moments_and_cgf`

`GaussianWishartDistribution.compute_moments_and_cgf(phi, mask=True)`

Compute the moments and $g(\phi)$.

bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution.compute_phi_from_parents

GaussianWishartDistribution.**compute_phi_from_parents**(*u_mu_alpha*, *u_V*, *u_n*,
mask=True)

Compute the natural parameter vector given parent moments.

bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution.random

GaussianWishartDistribution.**random**(**params*, *plates=None*)

Draw a random sample from the distribution.

6.3.8 bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution

class bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution(*N*,
D)

Sub-classes implement distribution specific computations.

__init__(*N*, *D*)

Methods

| | |
|---|--|
| __init__ (<i>N</i> , <i>D</i>) | |
| compute_cgf_from_parents (<i>u_mu</i> , <i>u_Lambda</i> , ...) | Compute CGF using the moments of the parents. |
| compute_fixed_moments_and_f (<i>x</i> [, <i>mask</i>]) | Compute <i>u</i> (<i>x</i>) and <i>f</i> (<i>x</i>) for given <i>x</i> . |
| compute_mask_to_parent (<i>index</i> , <i>mask</i>) | |
| compute_message_to_parent (<i>parent</i> , <i>index</i> , <i>u</i> , ...) | Compute a message to a parent. |
| compute_moments_and_cgf (<i>phi</i> [, <i>mask</i>]) | Compute the moments and the cumulant-generating function. |
| compute_phi_from_parents (<i>u_mu</i> , <i>u_Lambda</i> , ...) | Compute the natural parameters using parents' moments. |
| plates_from_parent (<i>index</i> , <i>plates</i>) | Compute the plates using information of a parent node. |
| plates_to_parent (<i>index</i> , <i>plates</i>) | Computes the plates of this node with respect to a parent. |

bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution.__init__

GaussianMarkovChainDistribution.**__init__**(*N*, *D*)

bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution.compute_cgf_from_parents

GaussianMarkovChainDistribution.**compute_cgf_from_parents**(*u_mu*, *u_Lambda*,
u_A, *u_v*, **u_inputs*)

Compute CGF using the moments of the parents.

bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution.compute_fixed_moments_and_f

GaussianMarkovChainDistribution.**compute_fixed_moments_and_f**(*x*, *mask=True*)

Compute *u*(*x*) and *f*(*x*) for given *x*.

bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution.compute_mask_to_parent

GaussianMarkovChainDistribution.**compute_mask_to_parent**(*index*, *mask*)

bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution.compute_message_to_parent

`GaussianMarkovChainDistribution.compute_message_to_parent` (*parent*, *index*, *u*,
u_mu, *u_Lambda*,
u_A, *u_v*, **u_inputs*)

Compute a message to a parent.

Parameters *index* : int

Index of the parent requesting the message.

u : list of ndarrays

Moments of this node.

u_mu : list of ndarrays

Moments of parent *mu*.

u_Lambda : list of ndarrays

Moments of parent *Lambda*.

u_A : list of ndarrays

Moments of parent *A*.

u_v : list of ndarrays

Moments of parent *v*.

u_inputs : list of ndarrays

Moments of input signals.

bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution.compute_moments_and_cgf

`GaussianMarkovChainDistribution.compute_moments_and_cgf` (*phi*, *mask=True*)

Compute the moments and the cumulant-generating function.

This basically performs the filtering and smoothing for the variable.

Parameters *phi*

Returns *u*

g

bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution.compute_phi_from_parents

`GaussianMarkovChainDistribution.compute_phi_from_parents` (*u_mu*, *u_Lambda*,
u_A, *u_v*, **u_inputs*,
mask=True)

Compute the natural parameters using parents' moments.

Parameters *u_parents* : list of list of arrays

List of parents' lists of moments.

Returns *phi* : list of arrays

Natural parameters.

dims : tuple

Shape of the variable part of phi.

bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution.plates_from_parent

`GaussianMarkovChainDistribution.plates_from_parent` (*index, plates*)

Compute the plates using information of a parent node.

If the plates of the parents are: `mu: (...) Lambda: (...) A: (...N-1,D) v: (...N-1,D) N: ()`

the resulting plates of this node are (...)

Parameters `index` : int

Index of the parent to use.

bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution.plates_to_parent

`GaussianMarkovChainDistribution.plates_to_parent` (*index, plates*)

Computes the plates of this node with respect to a parent.

If this node has plates (...), the latent dimensionality is D and the number of time instances is N, the plates with respect to the parents are:

`mu: (...) Lambda: (...) A: (...N-1,D) v: (...N-1,D)`

Parameters `index` : int

The index of the parent node to use.

6.3.9 bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution

class `bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution`

Sub-classes implement distribution specific computations.

`__init__` (*N, D, K*)

Methods

| | |
|---|--|
| <code>__init__(N, D, K)</code> | |
| <code>compute_cgf_from_parents(u_mu, u_Lambda, ...)</code> | Compute CGF using the moments of the parents. |
| <code>compute_fixed_moments_and_f(x[, mask])</code> | Compute $u(x)$ and $f(x)$ for given x . |
| <code>compute_mask_to_parent(index, mask)</code> | |
| <code>compute_message_to_parent(parent, index, u, ...)</code> | Compute a message to a parent. |
| <code>compute_moments_and_cgf(phi[, mask])</code> | Compute the moments and the cumulant-generating function. |
| <code>compute_phi_from_parents(u_mu, u_Lambda, ...)</code> | Compute the natural parameters using parents' moments. |
| <code>plates_from_parent(index, plates)</code> | Compute the plates using information of a parent node. |
| <code>plates_to_parent(index, plates)</code> | Computes the plates of this node with respect to a parent. |

bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution.__init__

`SwitchingGaussianMarkovChainDistribution.__init__` (*N, D, K*)

bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution.compute_cg

```
SwitchingGaussianMarkovChainDistribution.compute_cgf_from_parents (u_mu,
                                                                    u_Lambda,
                                                                    u_B,
                                                                    u_Z,
                                                                    u_v)
```

Compute CGF using the moments of the parents.

bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution.compute_f

```
SwitchingGaussianMarkovChainDistribution.compute_fixed_moments_and_f (x,
                                                                        mask=True)
```

Compute $u(x)$ and $f(x)$ for given x .

bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution.compute_mask

```
SwitchingGaussianMarkovChainDistribution.compute_mask_to_parent (index,
                                                                    mask)
```

bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution.compute_message

```
SwitchingGaussianMarkovChainDistribution.compute_message_to_parent (parent,
                                                                      index,
                                                                      u,
                                                                      u_mu,
                                                                      u_Lambda,
                                                                      u_B,
                                                                      u_Z,
                                                                      u_v)
```

Compute a message to a parent.

Parameters **index** : int

Index of the parent requesting the message.

u : list of ndarrays

Moments of this node.

u_mu : list of ndarrays

Moments of parent μ .

u_Lambda : list of ndarrays

Moments of parent Λ .

u_B : list of ndarrays

Moments of parent B .

u_Z : list of ndarrays

Moments of parent Z .

u_v : list of ndarrays

Moments of parent v .

bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution.compute_moments_and_cgf

`SwitchingGaussianMarkovChainDistribution.compute_moments_and_cgf` (*phi*,
mask=True)

Compute the moments and the cumulant-generating function.

This basically performs the filtering and smoothing for the variable.

Parameters *phi*

Returns *u*

g

bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution.compute_phi_from_parents

`SwitchingGaussianMarkovChainDistribution.compute_phi_from_parents` (*u_mu*,
u_Lambda,
u_B,
u_Z,
u_v,
mask=True)

Compute the natural parameters using parents' moments.

Parameters *u_parents* : list of list of arrays

List of parents' lists of moments.

Returns *phi* : list of arrays

Natural parameters.

dims : tuple

Shape of the variable part of phi.

bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution.plates_from_parent

`SwitchingGaussianMarkovChainDistribution.plates_from_parent` (*index*, *plates*)

Compute the plates using information of a parent node.

If the plates of the parents are: *mu*: (...) *Lambda*: (...) *B*: (...,D) *S*: (...,N-1) *v*: (...,N-1,D) *N*: ()

the resulting plates of this node are (...)

Parameters *index* : int

Index of the parent to use.

bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution.plates_to_parent

`SwitchingGaussianMarkovChainDistribution.plates_to_parent` (*index*, *plates*)

Computes the plates of this node with respect to a parent.

If this node has plates (...), the latent dimensionality is D and the number of time instances is N, the plates with respect to the parents are:

mu: (...) *Lambda*: (...) *A*: (...,N-1,D) *v*: (...,N-1,D)

Parameters *index* : int

The index of the parent node to use.

6.3.10 bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution

class bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution

Sub-classes implement distribution specific computations.

__init__ (*N*, *D*)

Methods

| | |
|---|--|
| <i>__init__</i> (<i>N</i> , <i>D</i>) | |
| <i>compute_cgf_from_parents</i> (<i>u_mu</i> , <i>u_Lambda</i> , ...) | Compute CGF using the moments of the parents. |
| <i>compute_fixed_moments_and_f</i> (<i>x</i> [, <i>mask</i>]) | Compute <i>u</i> (<i>x</i>) and <i>f</i> (<i>x</i>) for given <i>x</i> . |
| <i>compute_mask_to_parent</i> (<i>index</i> , <i>mask</i>) | |
| <i>compute_message_to_parent</i> (<i>parent</i> , <i>index</i> , <i>u</i> , ...) | Compute a message to a parent. |
| <i>compute_moments_and_cgf</i> (<i>phi</i> [, <i>mask</i>]) | Compute the moments and the cumulant-generating function. |
| <i>compute_phi_from_parents</i> (<i>u_mu</i> , <i>u_Lambda</i> , ...) | Compute the natural parameters using parents' moments. |
| <i>plates_from_parent</i> (<i>index</i> , <i>plates</i>) | Compute the plates using information of a parent node. |
| <i>plates_to_parent</i> (<i>index</i> , <i>plates</i>) | Computes the plates of this node with respect to a parent. |

bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution.__init__

VaryingGaussianMarkovChainDistribution.**__init__** (*N*, *D*)

bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution.compute_cgf_from_parents

VaryingGaussianMarkovChainDistribution.**compute_cgf_from_parents** (*u_mu*,
u_Lambda,
u_B, *u_S*,
u_v)

Compute CGF using the moments of the parents.

bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution.compute_fixed_moments_and_f

VaryingGaussianMarkovChainDistribution.**compute_fixed_moments_and_f** (*x*,
mask=True)

Compute *u*(*x*) and *f*(*x*) for given *x*.

bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution.compute_mask_to_parent

VaryingGaussianMarkovChainDistribution.**compute_mask_to_parent** (*index*, *mask*)

bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution.compute_message

```
VaryingGaussianMarkovChainDistribution.compute_message_to_parent (parent,
                                                                    index,
                                                                    u, u_mu,
                                                                    u_Lambda,
                                                                    u_B, u_S,
                                                                    u_v)
```

Compute a message to a parent.

Parameters `index` : int

Index of the parent requesting the message.

u : list of ndarrays

Moments of this node.

u_mu : list of ndarrays

Moments of parent *mu*.

u_Lambda : list of ndarrays

Moments of parent *Lambda*.

u_B : list of ndarrays

Moments of parent *B*.

u_S : list of ndarrays

Moments of parent *S*.

u_v : list of ndarrays

Moments of parent *v*.

bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution.compute_moments_and_cgf

```
VaryingGaussianMarkovChainDistribution.compute_moments_and_cgf (phi,
                                                                    mask=True)
```

Compute the moments and the cumulant-generating function.

This basically performs the filtering and smoothing for the variable.

Parameters `phi`

Returns `u`

`g`

bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution.compute_phi

```
VaryingGaussianMarkovChainDistribution.compute_phi_from_parents (u_mu,
                                                                    u_Lambda,
                                                                    u_B,
                                                                    u_S, u_v,
                                                                    mask=True)
```

Compute the natural parameters using parents' moments.

Parameters `u_parents` : list of list of arrays

List of parents' lists of moments.

Returns **phi** : list of arrays

Natural parameters.

dims : tuple

Shape of the variable part of phi.

bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution.plates_from

`VaryingGaussianMarkovChainDistribution.plates_from_parent (index, plates)`

Compute the plates using information of a parent node.

If the plates of the parents are: mu: (...) Lambda: (...) B: (...,D) S: (...,N-1) v: (...,N-1,D) N: ()

the resulting plates of this node are (...)

Parameters **index** : int

Index of the parent to use.

bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution.plates_to_pa

`VaryingGaussianMarkovChainDistribution.plates_to_parent (index, plates)`

Computes the plates of this node with respect to a parent.

If this node has plates (...), the latent dimensionality is D and the number of time instances is N, the plates with respect to the parents are:

mu: (...) Lambda: (...) A: (...,N-1,D) v: (...,N-1,D)

Parameters **index** : int

The index of the parent node to use.

6.3.11 bayespy.inference.vmp.nodes.gamma.GammaDistribution

class bayespy.inference.vmp.nodes.gamma.GammaDistribution

Class for the VMP formulas of gamma variables.

__init__ ()

Initialize self. See help(type(self)) for accurate signature.

Methods

| | |
|--|--|
| <code>compute_cgf_from_parents(*u_parents)</code> | Compute $E_{q(p)}[g(p)]$ |
| <code>compute_fixed_moments_and_f(x[, mask])</code> | Compute the moments and $f(x)$ for a fixed value. |
| <code>compute_gradient(g, u, phi)</code> | Compute the moments and $g(\phi)$. |
| <code>compute_message_to_parent(parent, index, ...)</code> | Compute the message to a parent node. |
| <code>compute_moments_and_cgf(phi[, mask])</code> | Compute the moments and $g(\phi)$. |
| <code>compute_phi_from_parents(*u_parents[, mask])</code> | Compute the natural parameter vector given parent moments. |
| <code>random(*phi[, plates])</code> | Draw a random sample from the distribution. |

bayespy.inference.vmp.nodes.gamma.GammaDistribution.compute_cgf_from_parents

`GammaDistribution.compute_cgf_from_parents (*u_parents)`
 Compute $E_{q(p)}[g(p)]$

bayespy.inference.vmp.nodes.gamma.GammaDistribution.compute_fixed_moments_and_f

`GammaDistribution.compute_fixed_moments_and_f (x, mask=True)`
 Compute the moments and $f(x)$ for a fixed value.

bayespy.inference.vmp.nodes.gamma.GammaDistribution.compute_gradient

`GammaDistribution.compute_gradient (g, u, phi)`
 Compute the moments and $g(\phi)$.

$$d\bar{u} = \begin{bmatrix} -\frac{d\phi_2}{\phi_1^2} + \frac{\phi_2}{\phi_1^2} d\phi_1 \\ \psi^{(1)}(\phi_2) d\phi_2 - \frac{1}{\phi_1} d\phi_1 \end{bmatrix}$$

Standard gradient given the gradient with respect to the moments, that is, given the Riemannian gradient $\tilde{\nabla}$:

$$\nabla = \begin{bmatrix} \nabla_1 \frac{\phi_2}{\phi_1^2} - \nabla_2 \frac{1}{\phi_1} \\ \nabla_2 \psi^{(1)}(\phi_2) - \nabla_1 \frac{1}{\phi_1} \end{bmatrix}$$

bayespy.inference.vmp.nodes.gamma.GammaDistribution.compute_message_to_parent

`GammaDistribution.compute_message_to_parent (parent, index, u_self, u_a, u_b)`
 Compute the message to a parent node.

bayespy.inference.vmp.nodes.gamma.GammaDistribution.compute_moments_and_cgf

`GammaDistribution.compute_moments_and_cgf (phi, mask=True)`
 Compute the moments and $g(\phi)$.

$$\bar{u}(\phi) = \begin{bmatrix} -\frac{\phi_2}{\phi_1} \\ \psi(\phi_2) - \log(-\phi_1) \end{bmatrix}$$

$$g_\phi(\phi) = \text{TODO}$$

bayespy.inference.vmp.nodes.gamma.GammaDistribution.compute_phi_from_parents

`GammaDistribution.compute_phi_from_parents (*u_parents, mask=True)`
 Compute the natural parameter vector given parent moments.

bayespy.inference.vmp.nodes.gamma.GammaDistribution.random

`GammaDistribution.random (*phi, plates=None)`
 Draw a random sample from the distribution.

6.3.12 bayespy.inference.vmp.nodes.wishart.WishartDistribution

class bayespy.inference.vmp.nodes.wishart.WishartDistribution

Sub-classes implement distribution specific computations.

Distribution for *kimesk* symmetric positive definite matrix.

$$\Lambda \sim \mathcal{W}(n, V)$$

Note: V is inverse scale matrix.

$$p(\Lambda|n, V) = ..$$

__init__()

Initialize self. See help(type(self)) for accurate signature.

Methods

| | |
|--|---|
| <code>compute_cgf_from_parents(u_n, u_V)</code> | CGF from parents |
| <code>compute_fixed_moments_and_f(Lambda[, mask])</code> | Compute u(x) and f(x) for given x. |
| <code>compute_message_to_parent(parent, index, ...)</code> | |
| <code>compute_moments_and_cgf(phi[, mask])</code> | Return moments and cgf for given natural parameters |
| <code>compute_phi_from_parents(u_n, u_V[, mask])</code> | Compute natural parameters |

bayespy.inference.vmp.nodes.wishart.WishartDistribution.compute_cgf_from_parents

WishartDistribution.**compute_cgf_from_parents** (u_n, u_V)

CGF from parents

$$g(n, V) = \frac{n}{2} \log |V| - \frac{nk}{2} \log 2 - \log \Gamma_k\left(\frac{n}{2}\right)$$

bayespy.inference.vmp.nodes.wishart.WishartDistribution.compute_fixed_moments_and_f

WishartDistribution.**compute_fixed_moments_and_f** (Λ , $mask=True$)

Compute u(x) and f(x) for given x.

bayespy.inference.vmp.nodes.wishart.WishartDistribution.compute_message_to_parent

WishartDistribution.**compute_message_to_parent** ($parent, index, u_{self}, u_n, u_V$)

bayespy.inference.vmp.nodes.wishart.WishartDistribution.compute_moments_and_cgf

WishartDistribution.**compute_moments_and_cgf** (ϕ , $mask=True$)

Return moments and cgf for given natural parameters

$$\langle u \rangle = \begin{bmatrix} \phi_2(-\phi_1)^{-1} \\ -\log |-\phi_1| + \psi_k(\phi_2) \end{bmatrix}$$

$$g(\phi) = \phi_2 \log |-\phi_1| - \log \Gamma_k(\phi_2)$$

bayespy.inference.vmp.nodes.wishart.WishartDistribution.compute_phi_from_parents

WishartDistribution.**compute_phi_from_parents** (*u_n, u_V, mask=True*)
 Compute natural parameters

$$\phi(n, V) = \begin{bmatrix} -\frac{1}{2}V \\ \frac{1}{2}n \end{bmatrix}$$

6.3.13 bayespy.inference.vmp.nodes.beta.BetaDistribution

class bayespy.inference.vmp.nodes.beta.**BetaDistribution**
 Class for the VMP formulas of beta variables.

Although the realizations are scalars (probability p), the moments is a two-dimensional vector: $[\log(p), \log(1-p)]$.

__init__ ()
 Initialize self. See help(type(self)) for accurate signature.

Methods

| | |
|--|--|
| <i>compute_cgf_from_parents</i> (<i>u_alpha</i>) | Compute $E_{q(p)}[g(p)]$ |
| <i>compute_fixed_moments_and_f</i> (<i>p[, mask]</i>) | Compute the moments and $f(x)$ for a fixed value. |
| <i>compute_message_to_parent</i> (<i>parent, index, ...</i>) | Compute the message to a parent node. |
| <i>compute_moments_and_cgf</i> (<i>phi[, mask]</i>) | Compute the moments and $g(\phi)$. |
| <i>compute_phi_from_parents</i> (<i>u_alpha[, mask]</i>) | Compute the natural parameter vector given parent moments. |
| <i>random</i> (* <i>phi[, plates]</i>) | Draw a random sample from the distribution. |

bayespy.inference.vmp.nodes.beta.BetaDistribution.compute_cgf_from_parents

BetaDistribution.**compute_cgf_from_parents** (*u_alpha*)
 Compute $E_{q(p)}[g(p)]$

bayespy.inference.vmp.nodes.beta.BetaDistribution.compute_fixed_moments_and_f

BetaDistribution.**compute_fixed_moments_and_f** (*p, mask=True*)
 Compute the moments and $f(x)$ for a fixed value.

bayespy.inference.vmp.nodes.beta.BetaDistribution.compute_message_to_parent

BetaDistribution.**compute_message_to_parent** (*parent, index, u_self, u_alpha*)
 Compute the message to a parent node.

bayespy.inference.vmp.nodes.beta.BetaDistribution.compute_moments_and_cgf

BetaDistribution.**compute_moments_and_cgf** (*phi, mask=True*)
 Compute the moments and $g(\phi)$.

bayespy.inference.vmp.nodes.beta.BetaDistribution.compute_phi_from_parents

`BetaDistribution.compute_phi_from_parents(u_alpha, mask=True)`
 Compute the natural parameter vector given parent moments.

bayespy.inference.vmp.nodes.beta.BetaDistribution.random

`BetaDistribution.random(*phi, plates=None)`
 Draw a random sample from the distribution.

6.3.14 bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution

class `bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution`
 Class for the VMP formulas of Dirichlet variables.

`__init__()`
 Initialize self. See `help(type(self))` for accurate signature.

Methods

| | |
|--|--|
| <code>compute_cgf_from_parents(u_alpha)</code> | Compute $E_{q(p)}[g(p)]$ |
| <code>compute_fixed_moments_and_f(p[, mask])</code> | Compute the moments and $f(x)$ for a fixed value. |
| <code>compute_gradient(g, u, phi)</code> | Compute the moments and $g(\phi)$. |
| <code>compute_message_to_parent(parent, index, ...)</code> | Compute the message to a parent node. |
| <code>compute_moments_and_cgf(phi[, mask])</code> | Compute the moments and $g(\phi)$. |
| <code>compute_phi_from_parents(u_alpha[, mask])</code> | Compute the natural parameter vector given parent moments. |
| <code>random(*phi[, plates])</code> | Draw a random sample from the distribution. |

bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution.compute_cgf_from_parents

`DirichletDistribution.compute_cgf_from_parents(u_alpha)`
 Compute $E_{q(p)}[g(p)]$

bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution.compute_fixed_moments_and_f

`DirichletDistribution.compute_fixed_moments_and_f(p, mask=True)`
 Compute the moments and $f(x)$ for a fixed value.

$$u(p) = \begin{bmatrix} \log(p_1) \\ \vdots \\ \log(p_D) \end{bmatrix}$$

$$f(p) = - \sum_d \log(p_d)$$

bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution.compute_gradient

`DirichletDistribution.compute_gradient(g, u, phi)`
 Compute the moments and $g(\phi)$.

$$\psi(\phi_{1,1}) - \psi(\sum_d \phi_{1,d})$$

Standard gradient given the gradient with respect to the moments, that is, given the Riemannian gradient $\tilde{\nabla}$:

$$\nabla = [(\psi^{(1)}(\phi_1) - \psi^{(1)}(\sum_d \phi_{1,d})\nabla_1]$$

bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution.compute_message_to_parent

`DirichletDistribution.compute_message_to_parent` (*parent, index, u_self, u_alpha*)

Compute the message to a parent node.

bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution.compute_moments_and_cgf

`DirichletDistribution.compute_moments_and_cgf` (*phi, mask=True*)

Compute the moments and $g(\phi)$.

$$\begin{aligned}\bar{\mathbf{u}}(\phi) &= [\psi(\phi_1) - \psi(\sum_d \phi_{1,d})] \\ g_\phi(\phi) &= TODO\end{aligned}$$

bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution.compute_phi_from_parents

`DirichletDistribution.compute_phi_from_parents` (*u_alpha, mask=True*)

Compute the natural parameter vector given parent moments.

bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution.random

`DirichletDistribution.random` (**phi, plates=None*)

Draw a random sample from the distribution.

6.3.15 bayespy.inference.vmp.nodes.bernoulli.BernoulliDistribution

class `bayespy.inference.vmp.nodes.bernoulli.BernoulliDistribution`

Class for the VMP formulas of Bernoulli variables.

`__init__()`

Methods

`__init__()`

bayespy.inference.vmp.nodes.bernoulli.BernoulliDistribution.__init__

`BernoulliDistribution.__init__()`

6.3.16 bayespy.inference.vmp.nodes.binomial.BinomialDistribution

class bayespy.inference.vmp.nodes.binomial.**BinomialDistribution** (*N*)
 Class for the VMP formulas of binomial variables.

__init__ (*N*)

Methods

| | |
|--|--|
| <i>__init__</i> (<i>N</i>) | |
| <i>compute_cgf_from_parents</i> (<i>u_p</i>) | Compute $E_{q(p)}[g(p)]$ |
| <i>compute_fixed_moments_and_f</i> (<i>x</i> [, <i>mask</i>]) | Compute the moments and $f(x)$ for a fixed value. |
| <i>compute_message_to_parent</i> (<i>parent</i> , <i>index</i> , ...) | Compute the message to a parent node. |
| <i>compute_moments_and_cgf</i> (<i>phi</i> [, <i>mask</i>]) | Compute the moments and $g(\phi)$. |
| <i>compute_phi_from_parents</i> (<i>u_p</i> [, <i>mask</i>]) | Compute the natural parameter vector given parent moments. |
| <i>random</i> (* <i>phi</i> [, <i>plates</i>]) | Draw a random sample from the distribution. |

bayespy.inference.vmp.nodes.binomial.BinomialDistribution.__init__

BinomialDistribution.**__init__** (*N*)

bayespy.inference.vmp.nodes.binomial.BinomialDistribution.compute_cgf_from_parents

BinomialDistribution.**compute_cgf_from_parents** (*u_p*)
 Compute $E_{q(p)}[g(p)]$

bayespy.inference.vmp.nodes.binomial.BinomialDistribution.compute_fixed_moments_and_f

BinomialDistribution.**compute_fixed_moments_and_f** (*x*, *mask*=*True*)
 Compute the moments and $f(x)$ for a fixed value.

bayespy.inference.vmp.nodes.binomial.BinomialDistribution.compute_message_to_parent

BinomialDistribution.**compute_message_to_parent** (*parent*, *index*, *u_self*, *u_p*)
 Compute the message to a parent node.

bayespy.inference.vmp.nodes.binomial.BinomialDistribution.compute_moments_and_cgf

BinomialDistribution.**compute_moments_and_cgf** (*phi*, *mask*=*True*)
 Compute the moments and $g(\phi)$.

bayespy.inference.vmp.nodes.binomial.BinomialDistribution.compute_phi_from_parents

BinomialDistribution.**compute_phi_from_parents** (*u_p*, *mask*=*True*)
 Compute the natural parameter vector given parent moments.

bayespy.inference.vmp.nodes.binomial.BinomialDistribution.random

`BinomialDistribution.random(*phi, plates=None)`
Draw a random sample from the distribution.

6.3.17 bayespy.inference.vmp.nodes.categorical.CategoricalDistribution

class `bayespy.inference.vmp.nodes.categorical.CategoricalDistribution(categories)`
Class for the VMP formulas of categorical variables.

__init__(*categories*)
Create VMP formula node for a categorical variable
categories is the total number of categories.

Methods

| | |
|---|--|
| <code>__init__(categories)</code> | Create VMP formula node for a categorical variable |
| <code>compute_cgf_from_parents(u_p)</code> | Compute $E_{q(p)}[g(p)]$ |
| <code>compute_fixed_moments_and_f(x[, mask])</code> | Compute the moments and $f(x)$ for a fixed value. |
| <code>compute_message_to_parent(parent, index, u, u_p)</code> | Compute the message to a parent node. |
| <code>compute_moments_and_cgf(phi[, mask])</code> | Compute the moments and $g(\phi)$. |
| <code>compute_phi_from_parents(u_p[, mask])</code> | Compute the natural parameter vector given parent moments. |
| <code>random(*phi[, plates])</code> | Draw a random sample from the distribution. |

bayespy.inference.vmp.nodes.categorical.CategoricalDistribution.__init__

`CategoricalDistribution.__init__(categories)`
Create VMP formula node for a categorical variable
categories is the total number of categories.

bayespy.inference.vmp.nodes.categorical.CategoricalDistribution.compute_cgf_from_parents

`CategoricalDistribution.compute_cgf_from_parents(u_p)`
Compute $E_{q(p)}[g(p)]$

bayespy.inference.vmp.nodes.categorical.CategoricalDistribution.compute_fixed_moments_and_f

`CategoricalDistribution.compute_fixed_moments_and_f(x, mask=True)`
Compute the moments and $f(x)$ for a fixed value.

bayespy.inference.vmp.nodes.categorical.CategoricalDistribution.compute_message_to_parent

`CategoricalDistribution.compute_message_to_parent(parent, index, u, u_p)`
Compute the message to a parent node.

bayespy.inference.vmp.nodes.categorical.CategoricalDistribution.compute_moments_and_cgf

`CategoricalDistribution.compute_moments_and_cgf(phi, mask=True)`
 Compute the moments and $g(\phi)$.

bayespy.inference.vmp.nodes.categorical.CategoricalDistribution.compute_phi_from_parents

`CategoricalDistribution.compute_phi_from_parents(u_p, mask=True)`
 Compute the natural parameter vector given parent moments.

bayespy.inference.vmp.nodes.categorical.CategoricalDistribution.random

`CategoricalDistribution.random(*phi, plates=None)`
 Draw a random sample from the distribution.

6.3.18 bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution

class bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution

Class for the VMP formulas of categorical Markov chain variables.

__init__(categories, states)
 Create VMP formula node for a categorical variable
categories is the total number of categories. *states* is the length of the chain.

Methods

| | |
|---|--|
| <code>__init__(categories, states)</code> | Create VMP formula node for a categorical variable |
| <code>compute_cgf_from_parents(u_p0, u_P)</code> | Compute $E_{q(p)}[g(p)]$ |
| <code>compute_fixed_moments_and_f(x[, mask])</code> | Compute the moments and $f(x)$ for a fixed value. |
| <code>compute_mask_to_parent(index, mask)</code> | Maps the mask to the plates of a parent. |
| <code>compute_message_to_parent(parent, index, u, ...)</code> | Compute the message to a parent node. |
| <code>compute_moments_and_cgf(phi[, mask])</code> | Compute the moments and $g(\phi)$. |
| <code>compute_phi_from_parents(u_p0, u_P[, mask])</code> | Compute the natural parameter vector given parent moments. |
| <code>plates_from_parent(index, plates)</code> | Resolve the plate mapping from a parent. |
| <code>plates_to_parent(index, plates)</code> | Resolves the plate mapping to a parent. |
| <code>random(*phi[, plates])</code> | Draw a random sample from the distribution. |

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution.__init__

`CategoricalMarkovChainDistribution.__init__(categories, states)`
 Create VMP formula node for a categorical variable
categories is the total number of categories. *states* is the length of the chain.

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution.compute_cgf_from_parents

`CategoricalMarkovChainDistribution.compute_cgf_from_parents(u_p0, u_P)`
 Compute $E_{q(p)}[g(p)]$

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution.compute_fixed_moments_and_f

`CategoricalMarkovChainDistribution.compute_fixed_moments_and_f(x, mask=True)`
 Compute the moments and $f(x)$ for a fixed value.

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution.compute_mask_to_parent

`CategoricalMarkovChainDistribution.compute_mask_to_parent(index, mask)`
 Maps the mask to the plates of a parent.

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution.compute_message_to_parent

`CategoricalMarkovChainDistribution.compute_message_to_parent(parent, index, u, u_p0, u_P)`
 Compute the message to a parent node.

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution.compute_moments_and_cg

`CategoricalMarkovChainDistribution.compute_moments_and_cg(phi, mask=True)`
 Compute the moments and $g(\phi)$.

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution.compute_phi_from_parents

`CategoricalMarkovChainDistribution.compute_phi_from_parents(u_p0, u_P, mask=True)`
 Compute the natural parameter vector given parent moments.

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution.plates_from_parents

`CategoricalMarkovChainDistribution.plates_from_parent(index, plates)`
 Resolve the plate mapping from a parent.
 Given the plates of a parent's moments, this method returns the plates that the moments has for this distribution.

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution.plates_to_parents

`CategoricalMarkovChainDistribution.plates_to_parent(index, plates)`
 Resolves the plate mapping to a parent.
 Given the plates of the node's moments, this method returns the plates that the message to a parent has for the parent's distribution.

bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution.random

`CategoricalMarkovChainDistribution.random(*phi, plates=None)`
 Draw a random sample from the distribution.

6.3.19 bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution

class bayespy.inference.vmp.nodes.multinomial.**MultinomialDistribution** (*trials*)

Class for the VMP formulas of multinomial variables.

__init__ (*trials*)

Create VMP formula node for a multinomial variable

trials is the total number of trials.

Methods

| | |
|---|--|
| __init__ (<i>trials</i>) | Create VMP formula node for a multinomial variable |
| compute_cgf_from_parents (<i>u_p</i>) | Compute $E_{q(p)}[g(p)]$ |
| compute_fixed_moments_and_f (<i>x</i> [, <i>mask</i>]) | Compute the moments and $f(x)$ for a fixed value. |
| compute_message_to_parent (<i>parent</i> , <i>index</i> , <i>u</i> , <i>u_p</i>) | Compute the message to a parent node. |
| compute_moments_and_cgf (<i>phi</i> [, <i>mask</i>]) | Compute the moments and $g(\phi)$. |
| compute_phi_from_parents (<i>u_p</i> [, <i>mask</i>]) | Compute the natural parameter vector given parent moments. |
| random (* <i>phi</i>) | Draw a random sample from the distribution. |

bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution.__init__

MultinomialDistribution.__init__ (*trials*)

Create VMP formula node for a multinomial variable

trials is the total number of trials.

bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution.compute_cgf_from_parents

MultinomialDistribution.compute_cgf_from_parents (*u_p*)

Compute $E_{q(p)}[g(p)]$

bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution.compute_fixed_moments_and_f

MultinomialDistribution.compute_fixed_moments_and_f (*x*, *mask*=*True*)

Compute the moments and $f(x)$ for a fixed value.

bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution.compute_message_to_parent

MultinomialDistribution.compute_message_to_parent (*parent*, *index*, *u*, *u_p*)

Compute the message to a parent node.

bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution.compute_moments_and_cgf

MultinomialDistribution.compute_moments_and_cgf (*phi*, *mask*=*True*)

Compute the moments and $g(\phi)$.

bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution.compute_phi_from_parents

`MultinomialDistribution.compute_phi_from_parents (u_p, mask=True)`
 Compute the natural parameter vector given parent moments.

bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution.random

`MultinomialDistribution.random (*phi)`
 Draw a random sample from the distribution.

6.3.20 bayespy.inference.vmp.nodes.poisson.PoissonDistribution

class bayespy.inference.vmp.nodes.poisson.PoissonDistribution
 Class for the VMP formulas of Poisson variables.

__init__()
 Initialize self. See help(type(self)) for accurate signature.

Methods

| | |
|---|--|
| <code>compute_cgf_from_parents(u_lambda)</code> | Compute $E_{q(p)}[g(p)]$ |
| <code>compute_fixed_moments_and_f(x[, mask])</code> | Compute the moments and $f(x)$ for a fixed value. |
| <code>compute_message_to_parent(parent, index, u, ...)</code> | Compute the message to a parent node. |
| <code>compute_moments_and_cgf(phi[, mask])</code> | Compute the moments and $g(\phi)$. |
| <code>compute_phi_from_parents(u_lambda[, mask])</code> | Compute the natural parameter vector given parent moments. |
| <code>random(*phi)</code> | Draw a random sample from the distribution. |

bayespy.inference.vmp.nodes.poisson.PoissonDistribution.compute_cgf_from_parents

`PoissonDistribution.compute_cgf_from_parents (u_lambda)`
 Compute $E_{q(p)}[g(p)]$

bayespy.inference.vmp.nodes.poisson.PoissonDistribution.compute_fixed_moments_and_f

`PoissonDistribution.compute_fixed_moments_and_f (x, mask=True)`
 Compute the moments and $f(x)$ for a fixed value.

bayespy.inference.vmp.nodes.poisson.PoissonDistribution.compute_message_to_parent

`PoissonDistribution.compute_message_to_parent (parent, index, u, u_lambda)`
 Compute the message to a parent node.

bayespy.inference.vmp.nodes.poisson.PoissonDistribution.compute_moments_and_cgf

`PoissonDistribution.compute_moments_and_cgf (phi, mask=True)`
 Compute the moments and $g(\phi)$.

bayespy.inference.vmp.nodes.poisson.PoissonDistribution.compute_phi_from_parents

`PoissonDistribution.compute_phi_from_parents(u_lambda, mask=True)`
Compute the natural parameter vector given parent moments.

bayespy.inference.vmp.nodes.poisson.PoissonDistribution.random

`PoissonDistribution.random(*phi)`
Draw a random sample from the distribution.

6.4 Utility functions

| | |
|-----------------|--|
| <i>linalg</i> | General numerical functions and methods. |
| <i>random</i> | General functions random sampling and distributions. |
| <i>optimize</i> | |
| <i>misc</i> | General numerical functions and methods. |

6.4.1 bayespy.utils.linalg

General numerical functions and methods.

Functions

| | |
|--|--|
| <i>block_banded_solve</i> (A, B, y) | Invert symmetric, banded, positive-definite matrix. |
| <i>chol</i> (C) | |
| <i>chol_inv</i> (U) | |
| <i>chol_logdet</i> (U) | |
| <i>chol_solve</i> (U, b[, out, matrix]) | |
| <i>dot</i> (*arrays) | Compute matrix-matrix product. |
| <i>inner</i> (*args[, ndim]) | Compute inner product. |
| <i>inv</i> (A[, ndim]) | General array inversion. |
| <i>logdet_chol</i> (U) | |
| <i>logdet_cov</i> (C) | |
| <i>logdet_tri</i> (R) | Logarithm of the absolute value of the determinant of a triangular matrix. |
| <i>m_dot</i> (A, b) | |
| <i>mmdot</i> (A, B[, ndim]) | Compute matrix-matrix product. |
| <i>mvdot</i> (A, b[, ndim]) | Compute matrix-vector product. |
| <i>outer</i> (A, B[, ndim]) | Computes outer product over the last axes of A and B. |
| <i>solve_triangular</i> (U, B, **kwargs) | |
| <i>tracedot</i> (A, B) | Computes trace(A*B). |
| <i>transpose</i> (X[, ndim]) | Transpose the matrix. |

bayespy.utils.linalg.block_banded_solve

`bayespy.utils.linalg.block_banded_solve(A, B, y)`
Invert symmetric, banded, positive-definite matrix.
A contains the diagonal blocks.

B contains the superdiagonal blocks (their transposes are the subdiagonal blocks).

Shapes: A: (... , N, D, D) B: (... , N-1, D, D) y: (... , N, D)

The algorithm is basically LU decomposition.

Computes only the diagonal and super-diagonal blocks of the inverse. The true inverse is dense, in general.

Assume each block has the same size.

Return: * inverse blocks * solution to the system * log-determinant

bayespy.utils.linalg.chol

`bayespy.utils.linalg.chol(C)`

bayespy.utils.linalg.chol_inv

`bayespy.utils.linalg.chol_inv(U)`

bayespy.utils.linalg.chol_logdet

`bayespy.utils.linalg.chol_logdet(U)`

bayespy.utils.linalg.chol_solve

`bayespy.utils.linalg.chol_solve(U, b, out=None, matrix=False)`

bayespy.utils.linalg.dot

`bayespy.utils.linalg.dot(*arrays)`

Compute matrix-matrix product.

You can give multiple arrays, the dot product is computed from left to right: $A_1 * A_2 * A_3 * \dots * A_N$. The dot product is computed over the last two axes of each arrays. All other axes must be broadcastable.

bayespy.utils.linalg.inner

`bayespy.utils.linalg.inner(*args, ndim=1)`

Compute inner product.

The number of arrays is arbitrary. The number of dimensions is arbitrary.

bayespy.utils.linalg.inv

`bayespy.utils.linalg.inv(A, ndim=1)`

General array inversion.

Supports broadcasting and inversion of multidimensional arrays. For instance, an array with shape (4,3,2,3,2) could mean that there are four (3*2) x (3*2) matrices to be inverted. This can be done by `inv(A, ndim=2)`. For inverting scalars, `ndim=0`. For inverting matrices, `ndim=1`.

bayespy.utils.linalg.logdet_chol

`bayespy.utils.linalg.logdet_chol(U)`

bayespy.utils.linalg.logdet_cov

`bayespy.utils.linalg.logdet_cov(C)`

bayespy.utils.linalg.logdet_tri

`bayespy.utils.linalg.logdet_tri(R)`

Logarithm of the absolute value of the determinant of a triangular matrix.

bayespy.utils.linalg.m_dot

`bayespy.utils.linalg.m_dot(A, b)`

bayespy.utils.linalg.mmdot

`bayespy.utils.linalg.mmdot(A, B, ndim=1)`

Compute matrix-matrix product.

Applies broadcasting.

bayespy.utils.linalg.mvdot

`bayespy.utils.linalg.mvdot(A, b, ndim=1)`

Compute matrix-vector product.

Applies broadcasting.

bayespy.utils.linalg.outer

`bayespy.utils.linalg.outer(A, B, ndim=1)`

Computes outer product over the last axes of A and B.

The other axes are broadcasted. Thus, if A has shape $(..., N)$ and B has shape $(..., M)$, then the result has shape $(..., N, M)$.

Using the argument *ndim* it is possible to change that how many axes trailing axes are used for the outer product. For instance, if *ndim*=3, A and B have shapes $(..., N1, N2, N3)$ and $(..., M1, M2, M3)$, the result has shape $(..., N1, M1, N2, M2, N3, M3)$.

bayespy.utils.linalg.solve_triangular

`bayespy.utils.linalg.solve_triangular(U, B, **kwargs)`

bayespy.utils.linalg.tracedot

`bayespy.utils.linalg.tracedot(A, B)`
 Computes $\text{trace}(A*B)$.

bayespy.utils.linalg.transpose

`bayespy.utils.linalg.transpose(X, ndim=1)`
 Transpose the matrix.

6.4.2 bayespy.utils.random

General functions random sampling and distributions.

Functions

| | |
|--|--|
| <code>alpha_beta_recursion(logp0, logP)</code> | Compute alpha-beta recursion for Markov chain |
| <code>bernoulli(p[, size])</code> | Draw random samples from the Bernoulli distribution. |
| <code>categorical(p[, size])</code> | Draw random samples from a categorical distribution. |
| <code>correlation(D)</code> | Draw a random correlation matrix. |
| <code>covariance(D[, size, nu])</code> | Draw a random covariance matrix. |
| <code>dirichlet(alpha[, size])</code> | Draw random samples from the Dirichlet distribution. |
| <code>gamma_entropy(a, log_b, gammaln_a, psi_a, ...)</code> | Entropy of $\mathcal{G}(a, b)$. |
| <code>gamma_logpdf(bx, logx, a_logx, a_logb, gammaln_a)</code> | Log-density of $\mathcal{G}(x a, b)$. |
| <code>gaussian_entropy(logdet_V, D)</code> | Compute the entropy of a Gaussian distribution. |
| <code>gaussian_gamma_to_t(mu, Cov, a, b[, ndim])</code> | Integrates gamma distribution to obtain parameters of t distribution |
| <code>gaussian_logpdf(yVy, yVmu, muVmu, logdet_V, D)</code> | Log-density of a Gaussian distribution. |
| <code>intervals(N, length[, amount, gap])</code> | Return random non-overlapping parts of a sequence. |
| <code>invwishart_rand(nu, V)</code> | |
| <code>logodds_to_probability(x)</code> | Solves p from $\log(p/(1-p))$ |
| <code>mask(*shape[, p])</code> | Return a boolean array of the given shape. |
| <code>orth(D)</code> | Draw random orthogonal matrix. |
| <code>sphere([N])</code> | Draw random points uniformly on a unit sphere. |
| <code>svd(s)</code> | Draw a random matrix given its singular values. |
| <code>t_logpdf(z2, logdet_cov, nu, D)</code> | |
| <code>wishart(nu, V)</code> | Draw a random sample from the Wishart distribution. |
| <code>wishart_rand(nu, V)</code> | Draw a random sample from the Wishart distribution. |

bayespy.utils.random.alpha_beta_recursion

`bayespy.utils.random.alpha_beta_recursion(logp0, logP)`
 Compute alpha-beta recursion for Markov chain

Initial state log-probabilities are in $p0$ and state transition log-probabilities are in P . The probabilities do not need to be scaled to sum to one, but they are interpreted as below:

$$\log p0 = \log P(z_{-0}) + \log P(y_{-0}|z_{-0}) \quad \log P[...n,:,:) = \log P(z_{-n+1}|z_{-n}) + \log P(y_{-n+1}|z_{-n+1})$$

bayespy.utils.random.bernoulli

`bayespy.utils.random.bernoulli` (*p*, *size=None*)
 Draw random samples from the Bernoulli distribution.

bayespy.utils.random.categorical

`bayespy.utils.random.categorical` (*p*, *size=None*)
 Draw random samples from a categorical distribution.

bayespy.utils.random.correlation

`bayespy.utils.random.correlation` (*D*)
 Draw a random correlation matrix.

bayespy.utils.random.covariance

`bayespy.utils.random.covariance` (*D*, *size=()*, *nu=None*)
 Draw a random covariance matrix.

Draws from inverse-Wishart distribution. The distribution of each element is independent of the dimensionality of the matrix.

$C \sim \text{Inv-W}(I, D)$

Parameters *D* : int

Dimensionality of the covariance matrix.

bayespy.utils.random.dirichlet

`bayespy.utils.random.dirichlet` (*alpha*, *size=None*)
 Draw random samples from the Dirichlet distribution.

bayespy.utils.random.gamma_entropy

`bayespy.utils.random.gamma_entropy` (*a*, *log_b*, *gamma_ln_a*, *psi_a*, *a_psi_a*)
 Entropy of $\mathcal{G}(a, b)$.

If you want to get the gradient, just let each parameter be a gradient of that term.

Parameters *a* : ndarray

a

log_b : ndarray

$\log(b)$

gamma_ln_a : ndarray

$\log \Gamma(a)$

psi_a : ndarray

$\psi(a)$

a_psi_a : ndarray

$$a\psi(a)$$

bayespy.utils.random.gamma_logpdf

bayespy.utils.random.**gamma_logpdf** (*bx, logx, a_logx, a_logb, gammaln_a*)

Log-density of $\mathcal{G}(x|a, b)$.

If you want to get the gradient, just let each parameter be a gradient of that term.

Parameters **bx** : ndarray

$$bx$$

logx : ndarray

$$\log(x)$$

a_logx : ndarray

$$a \log(x)$$

a_logb : ndarray

$$a \log(b)$$

gammaln_a : ndarray

$$\log \Gamma(a)$$

bayespy.utils.random.gaussian_entropy

bayespy.utils.random.**gaussian_entropy** (*logdet_V, D*)

Compute the entropy of a Gaussian distribution.

If you want to get the gradient, just let each parameter be a gradient of that term.

Parameters **logdet_V** : ndarray or double

The log-determinant of the precision matrix.

D : int

The dimensionality of the distribution.

bayespy.utils.random.gaussian_gamma_to_t

bayespy.utils.random.**gaussian_gamma_to_t** (*mu, Cov, a, b, ndim=1*)

Integrates gamma distribution to obtain parameters of t distribution

bayespy.utils.random.gaussian_logpdf

bayespy.utils.random.**gaussian_logpdf** (*yVy, yVmu, muVmu, logdet_V, D*)

Log-density of a Gaussian distribution.

$$\mathcal{G}(y|\mu, V^{-1})$$

Parameters **yVy** : ndarray or double

$$y^T V y$$

yVmu : ndarray or double
 $y^T V \mu$
muVmu : ndarray or double
 $\mu^T V \mu$
logdet_V : ndarray or double
 Log-determinant of the precision matrix, $\log |V|$.
D : int
 Dimensionality of the distribution.

bayespy.utils.random.intervals

`bayespy.utils.random.intervals` (*N*, *length*, *amount*=1, *gap*=0)

Return random non-overlapping parts of a sequence.

For instance, N=16, length=2 and amount=4: [0, **1, 2**], 3, 4, 5, **6, 7**, 8, 9, **10, 11**, **12, 13**, 14, 15] that is, [1,2,6,7,10,11,12,13]

However, the function returns only the indices of the beginning of the sequences, that is, in the example: [1,6,10,12]

bayespy.utils.random.invwishart_rand

`bayespy.utils.random.invwishart_rand` (*nu*, *V*)

bayespy.utils.random.logodds_to_probability

`bayespy.utils.random.logodds_to_probability` (*x*)

Solves p from $\log(p/(1-p))$

bayespy.utils.random.mask

`bayespy.utils.random.mask` (**shape*, *p*=0.5)

Return a boolean array of the given shape.

Parameters **d0, d1, ..., dn** : int

Shape of the output.

p : value in range [0,1]

A probability that the elements are *True*.

bayespy.utils.random.orth

`bayespy.utils.random.orth` (*D*)

Draw random orthogonal matrix.

bayespy.utils.random.sphere

`bayespy.utils.random.sphere` ($N=1$)
 Draw random points uniformly on a unit sphere.
 Returns (latitude,longitude) in degrees.

bayespy.utils.random.svd

`bayespy.utils.random.svd` (s)
 Draw a random matrix given its singular values.

bayespy.utils.random.t_logpdf

`bayespy.utils.random.t_logpdf` ($z2$, \logdet_cov , nu , D)

bayespy.utils.random.wishart

`bayespy.utils.random.wishart` (nu , V)
 Draw a random sample from the Wishart distribution.
Parameters nu : int

bayespy.utils.random.wishart_rand

`bayespy.utils.random.wishart_rand` (nu , V)
 Draw a random sample from the Wishart distribution.
Parameters nu : int

6.4.3 bayespy.utils.optimize

Functions

| | |
|---|--|
| <code>check_gradient</code> (f , $x0$ [, $verbose$, $epsilon$, ...]) | Simple wrapper for SciPy's gradient checker. |
| <code>minimize</code> (f , $x0$ [, $maxiter$, $verbose$]) | Simple wrapper for SciPy's optimize. |

bayespy.utils.optimize.check_gradient

`bayespy.utils.optimize.check_gradient` (f , $x0$, $verbose=True$, $epsilon=1.4901161193847656e-08$, $return_abserr=False$)
 Simple wrapper for SciPy's gradient checker.
 The given function must return a tuple: (value, gradient).
 Returns absolute and relative errors

bayespy.utils.optimize.minimize

`bayespy.utils.optimize.minimize` (f , $x0$, $maxiter=None$, $verbose=False$)
 Simple wrapper for SciPy's optimize.

The given function must return a tuple: (value, gradient).

6.4.4 bayespy.utils.misc

General numerical functions and methods.

Functions

| | |
|--|---|
| <i>T(X)</i> | Transpose the matrix. |
| <i>add_axes(X[, num, axis])</i> | |
| <i>add_leading_axes(x, n)</i> | |
| <i>add_trailing_axes(x, n)</i> | |
| <i>array_to_scalar(x)</i> | |
| <i>atleast_nd(X, d)</i> | |
| <i>axes_to_collapse(shape_x, shape_to)</i> | |
| <i>block_banded(D, B)</i> | Construct a symmetric block-banded matrix. |
| <i>broadcasted_shape(*shapes)</i> | Computes the resulting broadcasted shape for a given set of shapes. |
| <i>broadcasted_shape_from_arrays(*arrays)</i> | Computes the resulting broadcasted shape for a given set of arrays. |
| <i>broadcasting_multiplier(plates, *args)</i> | Compute the plate multiplier for given shapes. |
| <i>ceildiv(a, b)</i> | Compute a divided by b and rounded up. |
| <i>chol(C)</i> | |
| <i>chol_inv(U)</i> | |
| <i>chol_logdet(U)</i> | |
| <i>chol_solve(U, b)</i> | |
| <i>cholesky(K)</i> | |
| <i>composite_function(function_list)</i> | Construct a function composition from a list of functions. |
| <i>diag(X[, ndim])</i> | Create a diagonal array given the diagonal elements. |
| <i>diagonal(A)</i> | |
| <i>dist_haversine(c1, c2[, radius])</i> | |
| <i>find_set_index(index, set_lengths)</i> | Given set sizes and an index, returns the index of the set |
| <i>first(L)</i> | |
| <i>gaussian_logpdf(y_invcov_y, y_invcov_mu, ...)</i> | |
| <i>get_diag(X[, ndim])</i> | Get the diagonal of an array. |
| <i>grid(x1, x2)</i> | Returns meshgrid as a (M*N,2)-shape array. |
| <i>identity(*shape)</i> | |
| <i>invgamma(x)</i> | Inverse gamma function. |
| <i>invpsi(x)</i> | Inverse digamma (psi) function. |
| <i>is_callable(f)</i> | |
| <i>is_numeric(a)</i> | |
| <i>is_shape_subset(sub_shape, full_shape)</i> | |
| <i>is_string(s)</i> | |
| <i>isinteger(x)</i> | |
| <i>kalman_filter(y, U, A, V, mu0, Cov0[, out])</i> | Perform Kalman filtering to obtain filtered mean and covariance. |
| <i>logdet_chol(U)</i> | |
| <i>logsumexp(X[, axis, keepdims])</i> | Compute log(sum(exp(X)) in a numerically stable way |
| <i>m_digamma(a, d)</i> | Returns the derivative of the log of multivariate gamma. |
| <i>m_dot(A, b)</i> | |
| <i>m_outer(A, B)</i> | |
| <i>m_solve_triangular(U, B, **kwargs)</i> | |

Continued on next page

Table 6.67 – continued from previous page

| | |
|--|---|
| <code>make_equal_length(*shapes)</code> | Make tuples equal length. |
| <code>make_equal_ndim(*arrays)</code> | Add trailing unit axes so that arrays have equal ndim |
| <code>mean(X[, axis, keepdims])</code> | Compute the mean, ignoring NaNs. |
| <code>moveaxis(A, axis_from, axis_to)</code> | Move the axis <i>axis_from</i> to position <i>axis_to</i> . |
| <code>multidigamma(a, d)</code> | Returns the derivative of the log of multivariate gamma. |
| <code>multiply_shapes(*shapes)</code> | Compute element-wise product of lists/tuples. |
| <code>nans([size])</code> | |
| <code>nested_iterator(max_inds)</code> | |
| <code>parse_command_line_arguments(mandatory_args, ...)</code> | Parse command line arguments of style “-parameter=value”. |
| <code>remove_whitespace(s)</code> | |
| <code>repeat_to_shape(A, s)</code> | |
| <code>rmse(y1, y2[, axis])</code> | |
| <code>rts_smoother(mu, Cov, A, V[, removethis])</code> | Perform Rauch-Tung-Striebel smoothing to obtain the posterior. |
| <code>safe_indices(inds, shape)</code> | Makes sure that indices are valid for given shape. |
| <code>squeeze(X)</code> | Remove leading axes that have unit length. |
| <code>squeeze_to_dim(X, dim)</code> | |
| <code>sum_multiply(*args[, axis, sumaxis, keepdims])</code> | |
| <code>sum_multiply_to_plates(*arrays[, to_plates, ...])</code> | Compute the product of the arguments and sum to the target shape. |
| <code>sum_product(*args[, axes_to_keep, ...])</code> | |
| <code>sum_to_dim(A, dim)</code> | Sum leading axes of A such that A has dim dimensions. |
| <code>sum_to_shape(X, s)</code> | Sum axes of the array such that the resulting shape is as given. |
| <code>symm(X)</code> | Make X symmetric. |
| <code>tempfile([prefix, suffix])</code> | |
| <code>trues(shape)</code> | |
| <code>unique(l)</code> | Remove duplicate items from a list while preserving order. |
| <code>write_to_hdf5(group, data, name)</code> | Writes the given array into the HDF5 file. |
| <code>zipper_merge(*lists)</code> | Combines lists by alternating elements from them. |

bayespy.utils.misc.T

`bayespy.utils.misc.T(X)`
Transpose the matrix.

bayespy.utils.misc.add_axes

`bayespy.utils.misc.add_axes(X, num=1, axis=0)`

bayespy.utils.misc.add_leading_axes

`bayespy.utils.misc.add_leading_axes(x, n)`

bayespy.utils.misc.add_trailing_axes

`bayespy.utils.misc.add_trailing_axes(x, n)`

bayespy.utils.misc.array_to_scalar

`bayespy.utils.misc.array_to_scalar(x)`

bayespy.utils.misc.atleast_nd

`bayespy.utils.misc.atleast_nd(X, d)`

bayespy.utils.misc.axes_to_collapse

`bayespy.utils.misc.axes_to_collapse(shape_x, shape_to)`

bayespy.utils.misc.block_banded

`bayespy.utils.misc.block_banded(D, B)`

Construct a symmetric block-banded matrix.

D contains square diagonal blocks. *B* contains super-diagonal blocks.

The resulting matrix is:

$D[0], B[0], 0, 0, \dots, 0, 0, 0, B[0].T, D[1], B[1], 0, \dots, 0, 0, 0, B[1].T, D[2], B[2], \dots, \dots, \dots, \dots, \dots, \dots, B[N-2].T, D[N-1], B[N-1], 0, 0, 0, 0, \dots, 0, B[N-1].T, D[N]$

bayespy.utils.misc.broadcasted_shape

`bayespy.utils.misc.broadcasted_shape(*shapes)`

Computes the resulting broadcasted shape for a given set of shapes.

Uses the broadcasting rules of NumPy. Raises an exception if the shapes do not broadcast.

bayespy.utils.misc.broadcasted_shape_from_arrays

`bayespy.utils.misc.broadcasted_shape_from_arrays(*arrays)`

Computes the resulting broadcasted shape for a given set of arrays.

Raises an exception if the shapes do not broadcast.

bayespy.utils.misc.broadcasting_multiplier

`bayespy.utils.misc.broadcasting_multiplier(plates, *args)`

Compute the plate multiplier for given shapes.

The first shape is compared to all other shapes (using NumPy broadcasting rules). All the elements which are non-unit in the first shape but 1 in all other shapes are multiplied together.

This method is used, for instance, for computing a correction factor for messages to parents: If this node has non-unit plates that are unit plates in the parent, those plates are summed. However, if the message has unit axis for that plate, it should be first broadcasted to the plates of this node and then summed to the plates of the parent. In order to avoid this broadcasting and summing, it is more efficient to just multiply by the correct factor. This method computes that factor. The first argument is the full plate shape of this node (with respect to the parent). The other arguments are the shape of the message array and the plates of the parent (with respect to this node).

bayespy.utils.misc.ceildiv

`bayespy.utils.misc.ceildiv(a, b)`

Compute a divided by b and rounded up.

bayespy.utils.misc.chol

`bayespy.utils.misc.chol(C)`

bayespy.utils.misc.chol_inv

`bayespy.utils.misc.chol_inv(U)`

bayespy.utils.misc.chol_logdet

`bayespy.utils.misc.chol_logdet(U)`

bayespy.utils.misc.chol_solve

`bayespy.utils.misc.chol_solve(U, b)`

bayespy.utils.misc.cholesky

`bayespy.utils.misc.cholesky(K)`

bayespy.utils.misc.composite_function

`bayespy.utils.misc.composite_function(function_list)`

Construct a function composition from a list of functions.

Given a list of functions $[f, g, h]$, constructs a function $h \circ g \circ f$. That is, returns a function z , for which $z(x) = h(g(f(x)))$.

bayespy.utils.misc.diag

`bayespy.utils.misc.diag(X, ndim=1)`

Create a diagonal array given the diagonal elements.

The diagonal array can be multi-dimensional. By default, the last axis is transformed to two axes (diagonal matrix) but this can be changed using `ndim` keyword. For instance, an array with shape (K, L, M, N) can be transformed to a set of diagonal 4-D tensors with shape (K, L, M, N, M, N) by giving `ndim=2`. If `ndim=3`, the result has shape (K, L, M, N, L, M, N) , and so on.

Diagonality means that for the resulting array Y holds: $Y[..., i_1, i_2, ..., i_{ndim}, j_1, j_2, ..., j_{ndim}]$ is zero if $i_n \neq j_n$ for any n .

bayespy.utils.misc.diagonal

`bayespy.utils.misc.diagonal(A)`

bayespy.utils.misc.dist_haversine

`bayespy.utils.misc.dist_haversine(c1, c2, radius=6372795)`

bayespy.utils.misc.find_set_index

`bayespy.utils.misc.find_set_index(index, set_lengths)`

Given set sizes and an index, returns the index of the set

The given index is for the concatenated list of the sets.

bayespy.utils.misc.first

`bayespy.utils.misc.first(L)`

bayespy.utils.misc.gaussian_logpdf

`bayespy.utils.misc.gaussian_logpdf(y_invcov_y, y_invcov_mu, mu_invcov_mu, logdetcov, D)`

bayespy.utils.misc.get_diag

`bayespy.utils.misc.get_diag(X, ndim=1)`

Get the diagonal of an array.

If `ndim > 1`, take the diagonal of the last `2*ndim` axes.

bayespy.utils.misc.grid

`bayespy.utils.misc.grid(x1, x2)`

Returns meshgrid as a $(M*N, 2)$ -shape array.

bayespy.utils.misc.identity

`bayespy.utils.misc.identity(*shape)`

bayespy.utils.misc.invgamma

`bayespy.utils.misc.invgamma(x)`

Inverse gamma function.

See: <http://mathoverflow.net/a/28977>

bayespy.utils.misc.invpsi

`bayespy.utils.misc.invpsi(x)`

Inverse digamma (psi) function.

The digamma function is the derivative of the log gamma function. This calculates the value $Y > 0$ for a value X such that $\text{digamma}(Y) = X$.

See: <http://www4.ncsu.edu/~pfackler/>

bayespy.utils.misc.is_callable

`bayespy.utils.misc.is_callable(f)`

bayespy.utils.misc.is_numeric

`bayespy.utils.misc.is_numeric(a)`

bayespy.utils.misc.is_shape_subset

`bayespy.utils.misc.is_shape_subset(sub_shape, full_shape)`

bayespy.utils.misc.is_string

`bayespy.utils.misc.is_string(s)`

bayespy.utils.misc.isinteger

`bayespy.utils.misc.isinteger(x)`

bayespy.utils.misc.kalman_filter

`bayespy.utils.misc.kalman_filter(y, U, A, V, mu0, Cov0, out=None)`

Perform Kalman filtering to obtain filtered mean and covariance.

The parameters of the process may vary in time, thus they are given as iterators instead of fixed values.

Parameters **y** : (N,D) array

“Normalized” noisy observations of the states, that is, the observations multiplied by the precision matrix U (and possibly other transformation matrices).

U : (N,D,D) array or N-list of (D,D) arrays

Precision matrix (i.e., inverse covariance matrix) of the observation noise for each time instance.

A : (N-1,D,D) array or (N-1)-list of (D,D) arrays

Dynamic matrix for each time instance.

V : (N-1,D,D) array or (N-1)-list of (D,D) arrays

Covariance matrix of the innovation noise for each time instance.

Returns **mu** : array

Filtered mean of the states.

Cov : array

Filtered covariance of the states.

See also:

rts_smoother

bayespy.utils.misc.logdet_chol

`bayespy.utils.misc.logdet_chol(U)`

bayespy.utils.misc.logsumexp

`bayespy.utils.misc.logsumexp` (*X*, *axis=None*, *keepdims=False*)
 Compute $\log(\sum(\exp(X)))$ in a numerically stable way

bayespy.utils.misc.m_digamma

`bayespy.utils.misc.m_digamma` (*a*, *d*)
 Returns the derivative of the log of multivariate gamma.

bayespy.utils.misc.m_dot

`bayespy.utils.misc.m_dot` (*A*, *b*)

bayespy.utils.misc.m_outer

`bayespy.utils.misc.m_outer` (*A*, *B*)

bayespy.utils.misc.m_solve_triangular

`bayespy.utils.misc.m_solve_triangular` (*U*, *B*, ***kwargs*)

bayespy.utils.misc.make_equal_length

`bayespy.utils.misc.make_equal_length` (**shapes*)
 Make tuples equal length.
 Add leading 1s to shorter tuples.

bayespy.utils.misc.make_equal_ndim

`bayespy.utils.misc.make_equal_ndim` (**arrays*)
 Add trailing unit axes so that arrays have equal ndim

bayespy.utils.misc.mean

`bayespy.utils.misc.mean` (*X*, *axis=None*, *keepdims=False*)
 Compute the mean, ignoring NaNs.

bayespy.utils.misc.moveaxis

`bayespy.utils.misc.moveaxis` (*A*, *axis_from*, *axis_to*)
 Move the axis *axis_from* to position *axis_to*.

bayespy.utils.misc.multidigamma

`bayespy.utils.misc.multidigamma` (*a*, *d*)
 Returns the derivative of the log of multivariate gamma.

bayespy.utils.misc.multiply_shapes

`bayespy.utils.misc.multiply_shapes(*shapes)`

Compute element-wise product of lists/tuples.

Shorter lists are concatenated with leading 1s in order to get lists with the same length.

bayespy.utils.misc.nans

`bayespy.utils.misc.nans(size=())`

bayespy.utils.misc.nested_iterator

`bayespy.utils.misc.nested_iterator(max_inds)`

bayespy.utils.misc.parse_command_line_arguments

`bayespy.utils.misc.parse_command_line_arguments(mandatory_args, *optional_args_list, argv=None)`

Parse command line arguments of style “--parameter=value”.

Parameter specification is tuple: (name, converter, description).

Some special handling:

- If converter is None, the command line does not accept any value for it, but instead use either “--option” to enable or “--no-option” to disable.
- If argument name contains hyphens, those are converted to underscores in the keys of the returned dictionaries.

Parameters `mandatory_args` : list of tuples

Specs for mandatory arguments

`optional_args_list` : list of lists of tuples

Specs for each optional arguments set

`argv` : list of strings (optional)

The command line arguments. By default, read `sys.argv`.

Returns `args` : dictionary

The parsed mandatory arguments

`kwargs` : dictionary

The parsed optional arguments

Examples

```
>>> from pprint import pprint as print
>>> (args, kwargs) = parse_command_line_arguments(
...     # Mandatory arguments
...     [
...         ('name',      str,  "Full name"),
```

```

...     ('age', int, "Age (years)"),
...     ('employed', None, "Working"),
... ],
... # Optional arguments
... [
...     ('phone', str, "Phone number"),
...     ('favorite-color', str, "Favorite color")
... ],
... argv=['--name=John Doe',
...       '--age=42',
...       '--no-employed',
...       '--favorite-color=pink']
... )
>>> print(args)
{'age': 42, 'employed': False, 'name': 'John Doe'}
>>> print(kwargs)
{'favorite_color': 'pink'}

```

It is possible to have several optional argument sets:

```

>>> (args, kw_info, kw_fav) = parse_command_line_arguments(
...     # Mandatory arguments
...     [
...         ('name', str, "Full name"),
...     ],
...     # Optional arguments (contact information)
...     [
...         ('phone', str, "Phone number"),
...         ('email', str, "E-mail address")
...     ],
...     # Optional arguments (preferences)
...     [
...         ('favorite-color', str, "Favorite color"),
...         ('favorite-food', str, "Favorite food")
...     ],
...     argv=['--name=John Doe',
...           '--favorite-color=pink',
...           '--email=john.doe@email.com',
...           '--favorite-food=spaghetti']
... )
>>> print(args)
{'name': 'John Doe'}
>>> print(kw_info)
{'email': 'john.doe@email.com'}
>>> print(kw_fav)
{'favorite_color': 'pink', 'favorite_food': 'spaghetti'}

```

bayespy.utils.misc.remove_whitespace

bayespy.utils.misc.remove_whitespace(s)

bayespy.utils.misc.repeat_to_shape

bayespy.utils.misc.repeat_to_shape(A, s)

bayespy.utils.misc.rmse

`bayespy.utils.misc.rmse(y1, y2, axis=None)`

bayespy.utils.misc.rts_smoother

`bayespy.utils.misc.rts_smoother(mu, Cov, A, V, removethis=None)`

Perform Rauch-Tung-Striebel smoothing to obtain the posterior.

The function returns the posterior mean and covariance of each state. The parameters of the process may vary in time, thus they are given as iterators instead of fixed values.

Parameters **mu** : (N,D) array

Mean of the states from Kalman filter.

Cov : (N,D,D) array

Covariance of the states from Kalman filter.

A : (N-1,D,D) array or (N-1)-list of (D,D) arrays

Dynamic matrix for each time instance.

V : (N-1,D,D) array or (N-1)-list of (D,D) arrays

Covariance matrix of the innovation noise for each time instance.

Returns **mu** : array

Posterior mean of the states.

Cov : array

Posterior covariance of the states.

See also:

`kalman_filter`

bayespy.utils.misc.safe_indices

`bayespy.utils.misc.safe_indices(inds, shape)`

Makes sure that indices are valid for given shape.

The shorter shape determines the length.

For instance,

```
>>> safe_indices( (3, 4, 5), (1, 6) )
(0, 5)
```

bayespy.utils.misc.squeeze

`bayespy.utils.misc.squeeze(X)`

Remove leading axes that have unit length.

For instance, a shape (1,1,4,1,3) will be reshaped to (4,1,3).

bayespy.utils.misc.squeeze_to_dim

`bayespy.utils.misc.squeeze_to_dim(X, dim)`

bayespy.utils.misc.sum_multiply

`bayespy.utils.misc.sum_multiply(*args, axis=None, sumaxis=True, keepdims=False)`

bayespy.utils.misc.sum_multiply_to_plates

`bayespy.utils.misc.sum_multiply_to_plates(*arrays, to_plates=(), from_plates=None, ndim=0)`
 Compute the product of the arguments and sum to the target shape.

bayespy.utils.misc.sum_product

`bayespy.utils.misc.sum_product(*args, axes_to_keep=None, axes_to_sum=None, keepdims=False)`

bayespy.utils.misc.sum_to_dim

`bayespy.utils.misc.sum_to_dim(A, dim)`
 Sum leading axes of A such that A has dim dimensions.

bayespy.utils.misc.sum_to_shape

`bayespy.utils.misc.sum_to_shape(X, s)`
 Sum axes of the array such that the resulting shape is as given.
 Thus, the shape of the result will be s or an error is raised.

bayespy.utils.misc.symm

`bayespy.utils.misc.symm(X)`
 Make X symmetric.

bayespy.utils.misc.tempfile

`bayespy.utils.misc.tempfile(prefix='', suffix='')`

bayespy.utils.misc.trues

`bayespy.utils.misc.trues(shape)`

bayespy.utils.misc.unique

`bayespy.utils.misc.unique(l)`
 Remove duplicate items from a list while preserving order.

bayespy.utils.misc.write_to_hdf5

`bayespy.utils.misc.write_to_hdf5 (group, data, name)`
 Writes the given array into the HDF5 file.

bayespy.utils.misc.zipper_merge

`bayespy.utils.misc.zipper_merge (*lists)`
 Combines lists by alternating elements from them.
 Combining lists [1,2,3], ['a','b','c'] and [42,666,99] results in [1,'a',42,2,'b',666,3,'c',99]
 The lists should have equal length or they are assumed to have the length of the shortest list.
 This is known as alternating merge or zipper merge.

Classes

| | |
|-------------------------------------|-------------------------------------|
| <code>CholeskyDense(K)</code> | |
| <code>CholeskySparse(K)</code> | |
| <code>TestCase([methodName])</code> | Simple base class for unit testing. |

bayespy.utils.misc.CholeskyDense

`class bayespy.utils.misc.CholeskyDense (K)`

`__init__ (K)`

Methods

| |
|---------------------------------------|
| <code>__init__(K)</code> |
| <code>logdet()</code> |
| <code>solve(b)</code> |
| <code>trace_solve_gradient(dK)</code> |

bayespy.utils.misc.CholeskyDense.__init__

`CholeskyDense.__init__ (K)`

bayespy.utils.misc.CholeskyDense.logdet

`CholeskyDense.logdet ()`

bayespy.utils.misc.CholeskyDense.solve

`CholeskyDense.solve (b)`

bayespy.utils.misc.CholeskyDense.trace_solve_gradient

`CholeskyDense.trace_solve_gradient(dK)`

bayespy.utils.misc.CholeskySparse

class bayespy.utils.misc.CholeskySparse(K)

`__init__(K)`

Methods

| |
|---------------------------------------|
| <code>__init__(K)</code> |
| <code>logdet()</code> |
| <code>solve(b)</code> |
| <code>trace_solve_gradient(dK)</code> |

bayespy.utils.misc.CholeskySparse.__init__

`CholeskySparse.__init__(K)`

bayespy.utils.misc.CholeskySparse.logdet

`CholeskySparse.logdet()`

bayespy.utils.misc.CholeskySparse.solve

`CholeskySparse.solve(b)`

bayespy.utils.misc.CholeskySparse.trace_solve_gradient

`CholeskySparse.trace_solve_gradient(dK)`

bayespy.utils.misc.TestCase

class bayespy.utils.misc.TestCase(methodName='runTest')

Simple base class for unit testing.

Adds NumPy's features to Python's unittest.

`__init__(methodName='runTest')`

Create an instance of the class that will use the named test method when executed. Raises a ValueError if the instance does not have a method with the specified name.

Methods

| |
|--|
| <code>assertAllClose(A, B[, msg, rtol, atol])</code> |
| <code>assertArrayEqual(A, B[, msg])</code> |
| <code>assertMessage(M1, M2)</code> |
| <code>assertMessageToChild(X, u)</code> |

bayespy.utils.misc.TestCase.assertAllClose

`TestCase.assertAllClose(A, B, msg='Arrays not almost equal', rtol=0.0001, atol=0)`

bayespy.utils.misc.TestCase.assertArrayEqual

`TestCase.assertArrayEqual(A, B, msg='Arrays not equal')`

bayespy.utils.misc.TestCase.assertMessage

`TestCase.assertMessage(M1, M2)`

bayespy.utils.misc.TestCase.assertMessageToChild

`TestCase.assertMessageToChild(X, u)`

Attributes

| |
|--------------------------|
| <code>longMessage</code> |
| <code>maxDiff</code> |

bayespy.utils.misc.TestCase.longMessage

`TestCase.longMessage = True`

bayespy.utils.misc.TestCase.maxDiff

`TestCase.maxDiff = 640`

Bibliography

- Bibliography
- genindex
- modindex
- search

BIBLIOGRAPHY

- [1] James Hensman, Magnus Rattray, and Neil D. Lawrence. Fast variational inference in the conjugate exponential family. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, 2888–2896. Lake Tahoe, Nevada, USA, 2012.
- [2] Matthew D. Hoffman, David M. Blei, Chong Wang, and John Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14:1303–47, 2013.
- [3] Antti Honkela, Tapani Raiko, Mikael Kuusela, Matti Törnio, and Juha Karhunen. Approximate Riemannian conjugate gradient learning for fixed-form variational Bayes. *Journal of Machine Learning Research*, 11:3235–3268, 2010.
- [4] Antti Honkela, Harri Valpola, and Juha Karhunen. Accelerating cyclic update algorithms for parameter estimation by pattern searches. *Neural Processing Letters*, 17(2):191–203, 2003. doi:[10.1023/A:1023655202546](https://doi.org/10.1023/A:1023655202546).
- [5] K. Katahira, K. Watanabe, and M. Okada. Deterministic annealing variant of variational Bayes method. *Journal of Physics: Conference Series*, 2008.
- [6] Jaakko Luttinen. Fast variational Bayesian linear state-space model. In Hendrik Blockeel, Kristian Kersting, Siegfried Nijssen, and Filip Železný, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 8188 of Lecture Notes in Computer Science, pages 305–320. Springer, 2013. doi:[10.1007/978-3-642-40988-2_20](https://doi.org/10.1007/978-3-642-40988-2_20).
- [7] Jaakko Luttinen and Alexander Ilin. Transformations in variational Bayesian factor analysis to speed up learning. *Neurocomputing*, 73:1093–1102, 2010. doi:[10.1016/j.neucom.2009.11.018](https://doi.org/10.1016/j.neucom.2009.11.018).
- [8] Jaakko Luttinen, Tapani Raiko, and Alexander Ilin. Linear state-space model with time-varying dynamics. In Toon Calders, Floriana Esposito, Eyke Hüllermeier, and Rosa Meo, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 8725 of Lecture Notes in Computer Science, pages 338–353. Springer, 2014. doi:[10.1007/978-3-662-44851-9_22](https://doi.org/10.1007/978-3-662-44851-9_22).

b

`bayespy.inference`, 80
`bayespy.nodes`, 79
`bayespy.plot`, 87
`bayespy.utils.linalg`, 145
`bayespy.utils.misc`, 153
`bayespy.utils.optimize`, 152
`bayespy.utils.random`, 148

Symbols

- `__init__()` (bayespy.inference.vmp.nodes.bernoulli.BernoulliDistribution method), 138
- `__init__()` (bayespy.inference.vmp.nodes.bernoulli.BernoulliMoments method), 112, 113
- `__init__()` (bayespy.inference.vmp.nodes.beta.BetaDistribution method), 136
- `__init__()` (bayespy.inference.vmp.nodes.beta.BetaMoments method), 111
- `__init__()` (bayespy.inference.vmp.nodes.binomial.BinomialDistribution method), 139
- `__init__()` (bayespy.inference.vmp.nodes.binomial.BinomialMoments method), 113
- `__init__()` (bayespy.inference.vmp.nodes.categorical.CategoricalDistribution method), 140
- `__init__()` (bayespy.inference.vmp.nodes.categorical.CategoricalMoments method), 113, 114
- `__init__()` (bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution method), 141
- `__init__()` (bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainMoments method), 114
- `__init__()` (bayespy.inference.vmp.nodes.constant.Constant method), 101
- `__init__()` (bayespy.inference.vmp.nodes.deterministic.Deterministic method), 100
- `__init__()` (bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution method), 137
- `__init__()` (bayespy.inference.vmp.nodes.dirichlet.DirichletMoments method), 112
- `__init__()` (bayespy.inference.vmp.nodes.expfamily.ExponentialFamily method), 97, 98
- `__init__()` (bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution method), 117
- `__init__()` (bayespy.inference.vmp.nodes.gamma.GammaDistribution method), 133
- `__init__()` (bayespy.inference.vmp.nodes.gamma.GammaMoments method), 110
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution method), 120, 121
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianDistribution method), 118
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDDistribution method), 123, 125
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDMoments method), 109
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDToGaussianMoments method), 104
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution method), 122
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOMoments method), 108, 109
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOTOGaussianMoments method), 103
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianMoments method), 108
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianToGaussianGammaMoments method), 102
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution method), 125
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.GaussianWishartMoments method), 110
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaARDMoments method), 105
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaISOMoments method), 104
- `__init__()` (bayespy.inference.vmp.nodes.gaussian.WrapToGaussianWishartMoments method), 106
- `__init__()` (bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution method), 126
- `__init__()` (bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainMoments method), 108
- `__init__()` (bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMoments method), 128
- `__init__()` (bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMoments method), 131
- `__init__()` (bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution method), 143
- `__init__()` (bayespy.inference.vmp.nodes.multinomial.MultinomialMoments method), 115
- `__init__()` (bayespy.inference.vmp.nodes.node.Moments method), 107
- `__init__()` (bayespy.inference.vmp.nodes.node.Node method), 93, 94
- `__init__()` (bayespy.inference.vmp.nodes.poisson.PoissonDistribution method), 123, 125

method), 144
__init__() (bayespy.inference.vmp.nodes.poisson.PoissonMoments method), 115
__init__() (bayespy.inference.vmp.nodes.stochastic.Distribution method), 116
__init__() (bayespy.inference.vmp.nodes.stochastic.Stochastic method), 95, 96
__init__() (bayespy.inference.vmp.nodes.wishart.WishartDistribution method), 135
__init__() (bayespy.inference.vmp.nodes.wishart.WishartMoments method), 111
__init__() (bayespy.inference.vmp.transformations.RotateGaussian method), 81
__init__() (bayespy.inference.vmp.transformations.RotateGaussianARD method), 82
__init__() (bayespy.inference.vmp.transformations.RotateGaussianMarkovChain method), 83, 84
__init__() (bayespy.inference.vmp.transformations.RotateMultipleBernoulli method), 86, 87
__init__() (bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain method), 84, 85
__init__() (bayespy.inference.vmp.transformations.RotateVaryingMarkovChain method), 85, 86
__init__() (bayespy.inference.vmp.transformations.RotationOptimizer method), 80, 81
__init__() (bayespy.plot.CategoricalMarkovChainPlotter method), 92
__init__() (bayespy.plot.ContourPlotter method), 90, 91
__init__() (bayespy.plot.FunctionPlotter method), 91
__init__() (bayespy.plot.GaussianTimeseriesPlotter method), 92
__init__() (bayespy.plot.HintonPlotter method), 91
__init__() (bayespy.plot.PDFPlotter method), 90
__init__() (bayespy.plot.Plotter method), 90
__init__() (bayespy.utils.misc.CholeskyDense method), 164
__init__() (bayespy.utils.misc.CholeskySparse method), 165
__init__() (bayespy.utils.misc.TestCase method), 165

A

add_axes() (in module bayespy.utils.misc), 154
add_converter() (bayespy.inference.vmp.nodes.node.Moments class method), 107
add_leading_axes() (in module bayespy.utils.misc), 154
add_plate_axis() (bayespy.inference.vmp.nodes.node.Node method), 94
add_trailing_axes() (in module bayespy.utils.misc), 154
alpha_beta_recursion() (in module bayespy.utils.random), 148
array_to_scalar() (in module bayespy.utils.misc), 154
assertAllClose() (bayespy.utils.misc.TestCase method), 166

assertArrayEqual() (bayespy.utils.misc.TestCase method), 166
assertMessage() (bayespy.utils.misc.TestCase method), 166
assertMessageToChild() (bayespy.utils.misc.TestCase method), 166
atleast_nd() (in module bayespy.utils.misc), 155
collapse() (in module bayespy.utils.misc), 155

B

bayespy.inference (module), 80
bayespy.nodes (module), 79
bayespy.plot (module), 87
bayespy.utils.linalg (module), 145
bayespy.utils.misc (module), 153
bayespy.utils.optimize (module), 152
bayespy.utils.random (module), 148
bernoulli() (in module bayespy.utils.random), 149
BernoulliDistribution (class in bayespy.inference.vmp.nodes.bernoulli), 138
BetaDistribution (class in bayespy.inference.vmp.nodes.bernoulli), 112
BetaMoments (class in bayespy.inference.vmp.nodes.beta), 136
BetaMoments (class in bayespy.inference.vmp.nodes.beta), 111
BinomialDistribution (class in bayespy.inference.vmp.nodes.binomial), 139
BinomialMoments (class in bayespy.inference.vmp.nodes.binomial), 113
block_banded() (in module bayespy.utils.misc), 155
block_banded_solve() (in module bayespy.utils.linalg), 145
bound() (bayespy.inference.vmp.transformations.RotateGaussian method), 81
bound() (bayespy.inference.vmp.transformations.RotateGaussianARD method), 82
bound() (bayespy.inference.vmp.transformations.RotateGaussianMarkovChain method), 84
bound() (bayespy.inference.vmp.transformations.RotateMultiple method), 87
bound() (bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain method), 85
bound() (bayespy.inference.vmp.transformations.RotateVaryingMarkovChain method), 86
broadcasted_shape() (in module bayespy.utils.misc), 155
broadcasted_shape_from_arrays() (in module bayespy.utils.misc), 155
broadcasting_multiplier() (bayespy.inference.vmp.nodes.node.Node

- static method), 94
- broadcasting_multiplier() (in module bayespy.utils.misc), 155
- ## C
- categorical() (in module bayespy.utils.random), 149
- CategoricalDistribution (class in bayespy.inference.vmp.nodes.categorical), 140
- CategoricalMarkovChainDistribution (class in bayespy.inference.vmp.nodes.categorical_markov_chain), 141
- CategoricalMarkovChainMoments (class in bayespy.inference.vmp.nodes.categorical_markov_chain), 114
- CategoricalMarkovChainPlotter (class in bayespy.plot), 92
- CategoricalMoments (class in bayespy.inference.vmp.nodes.categorical), 113
- ceildiv() (in module bayespy.utils.misc), 155
- check_gradient() (in module bayespy.utils.optimize), 152
- chol() (in module bayespy.utils.linalg), 146
- chol() (in module bayespy.utils.misc), 156
- chol_inv() (in module bayespy.utils.linalg), 146
- chol_inv() (in module bayespy.utils.misc), 156
- chol_logdet() (in module bayespy.utils.linalg), 146
- chol_logdet() (in module bayespy.utils.misc), 156
- chol_solve() (in module bayespy.utils.linalg), 146
- chol_solve() (in module bayespy.utils.misc), 156
- cholesky() (in module bayespy.utils.misc), 156
- CholeskyDense (class in bayespy.utils.misc), 164
- CholeskySparse (class in bayespy.utils.misc), 165
- composite_function() (in module bayespy.utils.misc), 156
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.beta.BetaDistribution method), 136
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.binomial.BinomialDistribution method), 139
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.categorical.CategoricalDistribution method), 140
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution method), 141
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution method), 137
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution method), 118
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.gamma.GammaDistribution method), 134
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution method), 121
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.gaussian.GaussianDistribution method), 119
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution method), 123
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution method), 125
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution method), 126
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution method), 129
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution method), 131
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution method), 143
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.poisson.PoissonDistribution method), 144
- compute_cgf_from_parents()
(bayespy.inference.vmp.nodes.wishart.WishartDistribution method), 135
- compute_dims_from_values()
(bayespy.inference.vmp.nodes.beta.BetaMoments method), 112
- compute_dims_from_values()
(bayespy.inference.vmp.nodes.binomial.BinomialMoments method), 113
- compute_dims_from_values()
(bayespy.inference.vmp.nodes.categorical.CategoricalMoments method), 114
- compute_dims_from_values()
(bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainMoments method), 114
- compute_dims_from_values()
(bayespy.inference.vmp.nodes.dirichlet.DirichletMoments method), 112
- compute_dims_from_values()
(bayespy.inference.vmp.nodes.gamma.GammaMoments method), 111
- compute_dims_from_values()
(bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDMoments method), 110
- compute_dims_from_values()
(bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOMoments method), 110

| | |
|--|---|
| method), 109 | method), 115 |
| compute_dims_from_values() (bayespy.inference.vmp.nodes.gaussian.GaussianMoments method), 108 | compute_fixed_moments() (bayespy.inference.vmp.nodes.node.Moments method), 107 |
| compute_dims_from_values() (bayespy.inference.vmp.nodes.gaussian.GaussianWishartMoments method), 110 | compute_fixed_moments() (bayespy.inference.vmp.nodes.poisson.PoissonMoments method), 115 |
| compute_dims_from_values() (bayespy.inference.vmp.nodes.multinomial.MultinomialMoments method), 115 | compute_fixed_moments() (bayespy.inference.vmp.nodes.wishart.WishartMoments method), 111 |
| compute_dims_from_values() (bayespy.inference.vmp.nodes.node.Moments method), 107 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.beta.BetaDistribution method), 136 |
| compute_dims_from_values() (bayespy.inference.vmp.nodes.poisson.PoissonMoments method), 115 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.binomial.BinomialDistribution method), 139 |
| compute_dims_from_values() (bayespy.inference.vmp.nodes.wishart.WishartMoments method), 111 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.categorical.CategoricalDistribution method), 140 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.beta.BetaMoments method), 112 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainMoments method), 142 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.binomial.BinomialMoments method), 113 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution method), 137 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.categorical.CategoricalMoments method), 114 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution method), 118 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainMoments method), 114 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.gamma.GammaDistribution method), 134 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.dirichlet.DirichletMoments method), 112 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution method), 121 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.gamma.GammaMoments method), 111 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.gaussian.GaussianDistribution method), 119 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDDistribution method), 110 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution method), 123 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution method), 109 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution method), 125 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.gaussian.GaussianMoments method), 108 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainMoments method), 126 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.gaussian.GaussianWishartMoments method), 110 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainMoments method), 129 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainMoments method), 108 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainMoments method), 131 |
| compute_fixed_moments() (bayespy.inference.vmp.nodes.multinomial.MultinomialMoments method), 108 | compute_fixed_moments_and_f() (bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution method), 108 |

method), 143

compute.fixed_moments_and_cf()
(bayespy.inference.vmp.nodes.poisson.PoissonDistribution
method), 144

compute.fixed_moments_and_cf()
(bayespy.inference.vmp.nodes.wishart.WishartDistribution
method), 135

compute.gradient()(bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution
method), 137

compute.gradient()(bayespy.inference.vmp.nodes.exponential.ExponentialDistribution
method), 118

compute.gradient()(bayespy.inference.vmp.nodes.gamma.GammaDistribution
method), 134

compute.gradient()(bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution
method), 121

compute.gradient()(bayespy.inference.vmp.nodes.gaussian.GaussianDistribution
method), 119

compute.logpdf()(bayespy.inference.vmp.nodes.exponential.ExponentialDistribution
method), 118

compute.mask_to_parent()
(bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution
method), 142

compute.mask_to_parent()
(bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution
method), 121

compute.mask_to_parent()
(bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution
method), 126

compute.mask_to_parent()
(bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution
method), 129

compute.mask_to_parent()
(bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution
method), 131

compute.mask_to_parent()
(bayespy.inference.vmp.nodes.stochastic.Distribution
method), 117

compute.message_to_parent()
(bayespy.inference.vmp.nodes.beta.BetaDistribution
method), 136

compute.message_to_parent()
(bayespy.inference.vmp.nodes.binomial.BinomialDistribution
method), 139

compute.message_to_parent()
(bayespy.inference.vmp.nodes.categorical.CategoricalDistribution
method), 140

compute.message_to_parent()
(bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution
method), 142

compute.message_to_parent()
(bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution
method), 138

compute.message_to_parent()
(bayespy.inference.vmp.nodes.exponential.ExponentialDistribution
method), 118

compute.message_to_parent()
(bayespy.inference.vmp.nodes.gamma.GammaDistribution
method), 134

compute.message_to_parent()
(bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution
method), 122

compute.message_to_parent()
(bayespy.inference.vmp.nodes.gaussian.GaussianDistribution
method), 125

compute.message_to_parent()
(bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution
method), 123

compute.message_to_parent()
(bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution
method), 125

compute.message_to_parent()
(bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution
method), 127

compute.message_to_parent()
(bayespy.inference.vmp.nodes.gaussian_markov_chain.SwitchingGaussianMarkovChainDistribution
method), 129

compute.message_to_parent()
(bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution
method), 132

compute.message_to_parent()
(bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution
method), 143

compute.message_to_parent()
(bayespy.inference.vmp.nodes.poisson.PoissonDistribution
method), 144

compute.message_to_parent()
(bayespy.inference.vmp.nodes.stochastic.Distribution
method), 117

compute.message_to_parent()
(bayespy.inference.vmp.nodes.wishart.WishartDistribution
method), 135

compute.moments_and_cgfs()
(bayespy.inference.vmp.nodes.beta.BetaDistribution
method), 136

compute.moments_and_cgfs()
(bayespy.inference.vmp.nodes.binomial.BinomialDistribution
method), 139

compute.moments_and_cgfs()
(bayespy.inference.vmp.nodes.categorical.CategoricalDistribution
method), 141

compute.moments_and_cgfs()
(bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution
method), 142

compute.moments_and_cgfs()
(bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution
method), 138

compute.moments_and_cgfs()
(bayespy.inference.vmp.nodes.exponential.ExponentialDistribution
method), 118

method), 118
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.gamma.GammaDistribution method), 134
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution method), 122
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.gaussian.GaussianDistribution method), 120
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution method), 123
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution method), 126
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.gaussian.markov_chain.GaussianMarkovChainDistribution method), 127
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.gaussian.markov_chain.SwitchingGaussianMarkovChainDistribution method), 130
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.gaussian.markov_chain.VaryingGaussianMarkovChainDistribution method), 132
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution method), 143
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.poisson.PoissonDistribution method), 144
compute_moments_and_cgf()
(bayespy.inference.vmp.nodes.wishart.WishartDistribution method), 136
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.beta.BetaDistribution method), 137
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.binomial.BinomialDistribution method), 139
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.categorical.CategoricalDistribution method), 141
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.categorical.markov_chain.CategoricalMarkovChainDistribution method), 142
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.dirichlet.DirichletDistribution method), 138
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.expfamily.ExponentialFamilyDistribution method), 118
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.gamma.GammaDistribution method), 134
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution method), 122
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.gaussian.GaussianDistribution method), 120
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.gaussian.GaussianGammaISODistribution method), 123
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.gaussian.GaussianWishartDistribution method), 126
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.gaussian.markov_chain.GaussianMarkovChainDistribution method), 127
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.gaussian.markov_chain.SwitchingGaussianMarkovChainDistribution method), 130
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.gaussian.markov_chain.VaryingGaussianMarkovChainDistribution method), 132
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.multinomial.MultinomialDistribution method), 143
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.poisson.PoissonDistribution method), 145
compute_phi_from_parents()
(bayespy.inference.vmp.nodes.wishart.WishartDistribution method), 136
Constant (class in bayespy.inference.vmp.nodes.constant), 101
contour() (in module bayespy.plot), 88
ContourPlotter (class in bayespy.plot), 90
correlation() (in module bayespy.utils.random), 149
covariance() (in module bayespy.utils.random), 149
D
delete() (bayespy.inference.vmp.nodes.node.Node method), 94
Deterministic (class in bayespy.inference.vmp.nodes.deterministic), 100
diag() (in module bayespy.utils.misc), 156
diagonal() (in module bayespy.utils.misc), 156
dims (bayespy.inference.vmp.nodes.expfamily.ExponentialFamily attribute), 100
dirichlet() (in module bayespy.utils.random), 149
DirichletDistribution (class in bayespy.inference.vmp.nodes.dirichlet), 137
DirichletMoments (class in bayespy.inference.vmp.nodes.dirichlet), 112
dist_inversine() (in module bayespy.utils.misc), 156

Distribution (class in bayespy.inference.vmp.nodes.stochastic), 116

dot() (in module bayespy.utils.linalg), 146

E

ExponentialFamily (class in bayespy.inference.vmp.nodes.expfamily), 97

ExponentialFamilyDistribution (class in bayespy.inference.vmp.nodes.expfamily), 117

F

find_set_index() (in module bayespy.utils.misc), 157

first() (in module bayespy.utils.misc), 157

FunctionPlotter (class in bayespy.plot), 91

G

gamma_entropy() (in module bayespy.utils.random), 149

gamma_logpdf() (in module bayespy.utils.random), 150

GammaDistribution (class in bayespy.inference.vmp.nodes.gamma), 133

GammaMoments (class in bayespy.inference.vmp.nodes.gamma), 110

gaussian_entropy() (in module bayespy.utils.random), 150

gaussian_gamma_to_t() (in module bayespy.utils.random), 150

gaussian_logpdf() (in module bayespy.utils.misc), 157

gaussian_logpdf() (in module bayespy.utils.random), 150

gaussian_mixture_2d() (in module bayespy.plot), 88

GaussianARDDistribution (class in bayespy.inference.vmp.nodes.gaussian), 120

GaussianDistribution (class in bayespy.inference.vmp.nodes.gaussian), 118

GaussianGammaARDDistribution (class in bayespy.inference.vmp.nodes.gaussian), 123

GaussianGammaARDMoments (class in bayespy.inference.vmp.nodes.gaussian), 109

GaussianGammaARDToGaussianWishart (class in bayespy.inference.vmp.nodes.gaussian), 104

GaussianGammaISODistribution (class in bayespy.inference.vmp.nodes.gaussian), 122

GaussianGammaISOMoments (class in bayespy.inference.vmp.nodes.gaussian), 108

GaussianGammaISOToGaussianGammaARD (class in bayespy.inference.vmp.nodes.gaussian), 103

GaussianMarkovChainDistribution (class in bayespy.inference.vmp.nodes.gaussian_markov_chain), 126

GaussianMarkovChainMoments (class in bayespy.inference.vmp.nodes.gaussian_markov_chain), 108

GaussianMoments (class in bayespy.inference.vmp.nodes.gaussian), 108

GaussianTimeseriesPlotter (class in bayespy.plot), 92

GaussianToGaussianGammaISO (class in bayespy.inference.vmp.nodes.gaussian), 102

GaussianWishartDistribution (class in bayespy.inference.vmp.nodes.gaussian), 125

GaussianWishartMoments (class in bayespy.inference.vmp.nodes.gaussian), 110

get_bound_terms() (bayespy.inference.vmp.transformations.RotateGaussian method), 81

get_bound_terms() (bayespy.inference.vmp.transformations.RotateGaussian method), 83

get_bound_terms() (bayespy.inference.vmp.transformations.RotateGaussian method), 84

get_bound_terms() (bayespy.inference.vmp.transformations.RotateMultiple method), 87

get_bound_terms() (bayespy.inference.vmp.transformations.RotateSwitching method), 85

get_bound_terms() (bayespy.inference.vmp.transformations.RotateVaryingM method), 86

get_converter() (bayespy.inference.vmp.nodes.node.Moments method), 107

get_diag() (in module bayespy.utils.misc), 157

get_gradient() (bayespy.inference.vmp.nodes.expfamily.ExponentialFamily method), 98

get_mask() (bayespy.inference.vmp.nodes.node.Node method), 94

get_moments() (bayespy.inference.vmp.nodes.constant.Constant method), 102

get_moments() (bayespy.inference.vmp.nodes.deterministic.Deterministic method), 100

get_moments() (bayespy.inference.vmp.nodes.node.Node method), 94

get_moments() (bayespy.inference.vmp.nodes.stochastic.Stochastic method), 96

get_parameters() (bayespy.inference.vmp.nodes.expfamily.ExponentialFamily method), 98

get_riemannian_gradient() (bayespy.inference.vmp.nodes.expfamily.ExponentialFamily method), 98

get_shape() (bayespy.inference.vmp.nodes.node.Node method), 94

grid() (in module bayespy.utils.misc), 157

H

has_plotter() (bayespy.inference.vmp.nodes.node.Node method), 94

hinton() (in module bayespy.plot), 88

HintonPlotter (class in bayespy.plot), 91

I

identity() (in module bayespy.utils.misc), 157

initialize_from_parameters()

(bayespy.inference.vmp.nodes.expfamily.ExponentialFamily method), 98

initialize_from_prior() (bayespy.inference.vmp.nodes.exponential.ExponentialFamily method), 98
 initialize_from_random() (bayespy.inference.vmp.nodes.exponential.ExponentialFamily method), 98
 initialize_from_value() (bayespy.inference.vmp.nodes.exponential.ExponentialFamily method), 99
 inner() (in module bayespy.utils.linalg), 146
 intervals() (in module bayespy.utils.random), 151
 inv() (in module bayespy.utils.linalg), 146
 invgamma() (in module bayespy.utils.misc), 157
 invpsi() (in module bayespy.utils.misc), 157
 invwishart_rand() (in module bayespy.utils.random), 151
 is_callable() (in module bayespy.utils.misc), 157
 is_numeric() (in module bayespy.utils.misc), 158
 is_shape_subset() (in module bayespy.utils.misc), 158
 is_string() (in module bayespy.utils.misc), 158
 isinteger() (in module bayespy.utils.misc), 158

K

kalman_filter() (in module bayespy.utils.misc), 158

L

load() (bayespy.inference.vmp.nodes.stochastic.Stochastic method), 96
 logdet() (bayespy.utils.misc.CholeskyDense method), 164
 logdet() (bayespy.utils.misc.CholeskySparse method), 165
 logdet_chol() (in module bayespy.utils.linalg), 147
 logdet_chol() (in module bayespy.utils.misc), 158
 logdet_cov() (in module bayespy.utils.linalg), 147
 logdet_tri() (in module bayespy.utils.linalg), 147
 logodds_to_probability() (in module bayespy.utils.random), 151
 logpdf() (bayespy.inference.vmp.nodes.exponential.ExponentialFamily method), 99
 logsumexp() (in module bayespy.utils.misc), 159
 longMessage (bayespy.utils.misc.TestCase attribute), 166
 lower_bound_contribution() (bayespy.inference.vmp.nodes.constant.Constant method), 102
 lower_bound_contribution() (bayespy.inference.vmp.nodes.deterministic.Deterministic method), 101
 lower_bound_contribution() (bayespy.inference.vmp.nodes.exponential.ExponentialFamily method), 99
 lowerbound() (bayespy.inference.vmp.nodes.stochastic.Stochastic method), 96

M

m_digamma() (in module bayespy.utils.misc), 159
 m_dot() (in module bayespy.utils.linalg), 147
 m_dot() (in module bayespy.utils.misc), 159

m_solve_triangular() (in module bayespy.utils.misc), 159
 make_equal_ndim() (in module bayespy.utils.misc), 159
 make_equal_ndim() (in module bayespy.utils.random), 151
 maxDiff (bayespy.utils.misc.TestCase attribute), 166
 mean() (in module bayespy.utils.misc), 159
 minimize() (in module bayespy.utils.optimize), 152
 mmdot() (in module bayespy.utils.linalg), 147
 Moments (class in bayespy.inference.vmp.nodes.node), 107
 move_plates() (bayespy.inference.vmp.nodes.node.Node method), 94
 moveaxis() (in module bayespy.utils.misc), 159
 multidigamma() (in module bayespy.utils.misc), 159
 MultinomialDistribution (class in bayespy.inference.vmp.nodes.multinomial), 143
 MultinomialMoments (class in bayespy.inference.vmp.nodes.multinomial), 115
 multiply_shapes() (in module bayespy.utils.misc), 160
 mvdot() (in module bayespy.utils.linalg), 147

N

nans() (in module bayespy.utils.misc), 160
 nested_iterator() (in module bayespy.utils.misc), 160
 Node (class in bayespy.inference.vmp.nodes.node), 93
 nodes() (bayespy.inference.vmp.transformations.RotateGaussian method), 81
 nodes() (bayespy.inference.vmp.transformations.RotateGaussianARD method), 83
 nodes() (bayespy.inference.vmp.transformations.RotateGaussianMarkovChain method), 84
 nodes() (bayespy.inference.vmp.transformations.RotateMultiple method), 87
 nodes() (bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain method), 85
 nodes() (bayespy.inference.vmp.transformations.RotateVaryingMarkovChain method), 86

O

observe() (bayespy.inference.vmp.nodes.exponential.ExponentialFamily method), 99
 observe() (bayespy.inference.vmp.nodes.stochastic.Stochastic method), 96
 orth() (in module bayespy.utils.random), 151
 orth() (in module bayespy.utils.linalg), 147

P

parse_command_line_arguments() (in module bayespy.utils.misc), 160
 pdf() (bayespy.inference.vmp.nodes.exponential.ExponentialFamily method), 99

pdf() (in module bayespy.plot), 88
 PDFPlotter (class in bayespy.plot), 90
 plates (bayespy.inference.vmp.nodes.constant.Constant attribute), 102
 plates (bayespy.inference.vmp.nodes.deterministic.Deterministic attribute), 101
 plates (bayespy.inference.vmp.nodes.expfamily.ExponentialFamily attribute), 100
 plates (bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDToGaussianWishart attribute), 104
 plates (bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOToGaussianGammaARD attribute), 103
 plates (bayespy.inference.vmp.nodes.gaussian.GaussianToGaussianGammaISOD attribute), 103
 plates (bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaARD attribute), 105
 plates (bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaISOD attribute), 105
 plates (bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaISOD attribute), 106
 plates (bayespy.inference.vmp.nodes.node.Node attribute), 95
 plates (bayespy.inference.vmp.nodes.stochastic.Stochastic attribute), 97
 plates_from_parent() (bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution method), 142
 plates_from_parent() (bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution method), 122
 plates_from_parent() (bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution method), 128
 plates_from_parent() (bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianSwitchingGaussianMarkovChainDistribution method), 130
 plates_from_parent() (bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution method), 133
 plates_from_parent() (bayespy.inference.vmp.nodes.stochastic.Distribution method), 117
 plates_multiplier (bayespy.inference.vmp.nodes.constant.Constant attribute), 102
 plates_multiplier (bayespy.inference.vmp.nodes.deterministic.Deterministic attribute), 101
 plates_multiplier (bayespy.inference.vmp.nodes.expfamily.ExponentialFamily attribute), 100
 plates_multiplier (bayespy.inference.vmp.nodes.gaussian.GaussianGammaARDToGaussianWishart attribute), 104
 plates_multiplier (bayespy.inference.vmp.nodes.gaussian.GaussianGammaISOToGaussianGammaARD attribute), 103
 plates_multiplier (bayespy.inference.vmp.nodes.gaussian.GaussianToGaussianGammaISOD attribute), 103
 plates_multiplier (bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaARD attribute), 105
 plates_multiplier (bayespy.inference.vmp.nodes.gaussian.WrapToGaussianGammaISOD attribute), 105
 plates_multiplier (bayespy.inference.vmp.nodes.gaussian.WrapToGaussianWishart attribute), 106
 plates_multiplier (bayespy.inference.vmp.nodes.node.Node attribute), 95
 plates_multiplier (bayespy.inference.vmp.nodes.stochastic.Stochastic attribute), 97
 plates_to_parent() (bayespy.inference.vmp.nodes.categorical_markov_chain.CategoricalMarkovChainDistribution method), 142
 plates_to_parent() (bayespy.inference.vmp.nodes.gaussian.GaussianARDDistribution method), 122
 plates_to_parent() (bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianMarkovChainDistribution method), 128
 plates_to_parent() (bayespy.inference.vmp.nodes.gaussian_markov_chain.GaussianSwitchingGaussianMarkovChainDistribution method), 130
 plates_to_parent() (bayespy.inference.vmp.nodes.gaussian_markov_chain.VaryingGaussianMarkovChainDistribution method), 133
 plates_to_parent() (bayespy.inference.vmp.nodes.stochastic.Distribution method), 117
 plates_to_parent() (bayespy.inference.vmp.nodes.stochastic.Distribution method), 117
 plot (in module bayespy.plot), 88
 Plotter (class in bayespy.plot), 89
 PoissonDistribution (class in bayespy.inference.vmp.nodes.poisson), 144
 PoissonMoments (class in bayespy.inference.vmp.nodes.poisson), 115
 R

[random\(\)](#) (bayespy.inference.vmp.nodes.stochastic.Stochastic method), 96
[remove_whitespace\(\)](#) (in module bayespy.utils.misc), 161
[repeat_to_shape\(\)](#) (in module bayespy.utils.misc), 161
[rmse\(\)](#) (in module bayespy.utils.misc), 162
[rotate\(\)](#) (bayespy.inference.vmp.transformations.RotateGaussian method), 82
[rotate\(\)](#) (bayespy.inference.vmp.transformations.RotateGaussianARD method), 85
[rotate\(\)](#) (bayespy.inference.vmp.transformations.RotateGaussianMarkovChain method), 86
[rotate\(\)](#) (bayespy.inference.vmp.transformations.RotateMultiple method), 87
[rotate\(\)](#) (bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain method), 85
[rotate\(\)](#) (bayespy.inference.vmp.transformations.RotateVaryingMarkovChain method), 86
[rotate\(\)](#) (bayespy.inference.vmp.transformations.RotationOptimizer method), 81
[RotateGaussian](#) (class in bayespy.inference.vmp.transformations), 81
[RotateGaussianARD](#) (class in bayespy.inference.vmp.transformations), 82
[RotateGaussianMarkovChain](#) (class in bayespy.inference.vmp.transformations), 83
[RotateMultiple](#) (class in bayespy.inference.vmp.transformations), 86
[RotateSwitchingMarkovChain](#) (class in bayespy.inference.vmp.transformations), 84
[RotateVaryingMarkovChain](#) (class in bayespy.inference.vmp.transformations), 85
[RotationOptimizer](#) (class in bayespy.inference.vmp.transformations), 80
[rts_smoother\(\)](#) (in module bayespy.utils.misc), 162

S

[safe_indices\(\)](#) (in module bayespy.utils.misc), 162
[save\(\)](#) (bayespy.inference.vmp.nodes.stochastic.Stochastic method), 96
[set_parameters\(\)](#) (bayespy.inference.vmp.nodes.expfamily.ExponentialFamily method), 99
[set_plotter\(\)](#) (bayespy.inference.vmp.nodes.node.Node method), 95
[set_value\(\)](#) (bayespy.inference.vmp.nodes.constant.Constant method), 102
[setup\(\)](#) (bayespy.inference.vmp.transformations.RotateGaussian method), 82

[setup\(\)](#) (bayespy.inference.vmp.transformations.RotateGaussianARD method), 83
[setup\(\)](#) (bayespy.inference.vmp.transformations.RotateGaussianMarkovChain method), 84
[setup\(\)](#) (bayespy.inference.vmp.transformations.RotateMultiple method), 87
[setup\(\)](#) (bayespy.inference.vmp.transformations.RotateSwitchingMarkovChain method), 85
[setup\(\)](#) (bayespy.inference.vmp.transformations.RotateVaryingMarkovChain method), 86
[show\(\)](#) (bayespy.inference.vmp.nodes.stochastic.Stochastic method), 96
[solve\(\)](#) (bayespy.utils.misc.CholeskyDense method), 164
[solve\(\)](#) (bayespy.utils.misc.CholeskySparse method), 165
[solve_triangular\(\)](#) (in module bayespy.utils.linalg), 147
[split\(\)](#) (in module bayespy.utils.random), 152
[squeeze\(\)](#) (in module bayespy.utils.misc), 162
[squeeze_to_dim\(\)](#) (in module bayespy.utils.misc), 163
[Stochastic](#) (class in bayespy.inference.vmp.nodes.stochastic), 95
[sum_multiply\(\)](#) (in module bayespy.utils.misc), 163
[sum_multiply_to_plates\(\)](#) (in module bayespy.utils.misc), 163
[sum_product\(\)](#) (in module bayespy.utils.misc), 163
[sum_to_dim\(\)](#) (in module bayespy.utils.misc), 163
[sum_to_shape\(\)](#) (in module bayespy.utils.misc), 163
[svd\(\)](#) (in module bayespy.utils.random), 152
[SwitchingGaussianMarkovChainDistribution](#) (class in bayespy.inference.vmp.nodes.gaussian_markov_chain), 128
[symm\(\)](#) (in module bayespy.utils.misc), 163

T

[T\(\)](#) (in module bayespy.utils.misc), 154
[t_logpdf\(\)](#) (in module bayespy.utils.random), 152
[tempfile\(\)](#) (in module bayespy.utils.misc), 163
[TestCase](#) (class in bayespy.utils.misc), 165
[trace_solve_gradient\(\)](#) (bayespy.utils.misc.CholeskyDense method), 165
[trace_solve_gradient\(\)](#) (bayespy.utils.misc.CholeskySparse method), 165
[tracedot\(\)](#) (in module bayespy.utils.linalg), 148
[transpose\(\)](#) (in module bayespy.utils.linalg), 148
[trues\(\)](#) (in module bayespy.utils.misc), 163

U

[unobserve\(\)](#) (bayespy.inference.vmp.nodes.stochastic.Stochastic method), 97
[update\(\)](#) (bayespy.inference.vmp.nodes.stochastic.Stochastic method), 97

V

[VaryingGaussianMarkovChainDistribution](#) (class in

bayespy.inference.vmp.nodes.gaussian_markov_chain),
131

W

wishart() (in module bayespy.utils.random), 152
wishart_rand() (in module bayespy.utils.random), 152
WishartDistribution (class in
bayespy.inference.vmp.nodes.wishart), 135
WishartMoments (class in
bayespy.inference.vmp.nodes.wishart), 111
WrapToGaussianGammaARD (class in
bayespy.inference.vmp.nodes.gaussian), 105
WrapToGaussianGammaISO (class in
bayespy.inference.vmp.nodes.gaussian), 104
WrapToGaussianWishart (class in
bayespy.inference.vmp.nodes.gaussian), 105
write_to_hdf5() (in module bayespy.utils.misc), 164

Z

zipper_merge() (in module bayespy.utils.misc), 164