

---

# **PyStan Documentation**

*Release 2.17.1.0*

**Allen B. Riddell**

**Jan 16, 2018**



---

## Contents

---

<b>1</b>	<b>Documentation</b>	<b>3</b>
<b>2</b>	<b>Stan documentation</b>	<b>33</b>
<b>3</b>	<b>Important links</b>	<b>35</b>
<b>4</b>	<b>Similar projects</b>	<b>37</b>
<b>5</b>	<b>Indices and tables</b>	<b>39</b>
	<b>Python Module Index</b>	<b>41</b>



PyStan provides an interface to [Stan](#), a package for Bayesian inference using the No-U-Turn sampler, a variant of Hamiltonian Monte Carlo.

License: Open source, GPL3



## 1.1 What's New

### 1.1.1 v2.17.1.0 (16. Jan 2018)

- Update Stan source to v2.17.1 ([release notes](#)) (bugfix release)
- Remove deprecation warnings (Thanks to Alexander Rudiuk)
- Drop testing of Python 3.4.

### 1.1.2 v2.17.0.0 (6. Oct 2017)

- Update Stan source to v2.17.0 ([release notes](#)),
- Added `pystan.stansummary` function. Patch by @ahartikainen.
- Marked `pystan.stan` as deprecated. It will be removed in version 3.0. Please compile and use a Stan program as two separate steps.
- Reminder: Consider using Python 3.5 or higher. In a future release, Python 3.4 wheels will no longer be built.
- Reminder: Stan v2.16.0 is the final release which will not require a C++11 compatible compiler. Future releases will require a C++11 compatible compiler. (This does not affect most users.)

### 1.1.3 v2.16.0.0 (22. June 2017)

- Update Stan source to v2.16.0 ([release notes](#)),
- Ari Hartikainen (Aalto University) @ahartikainen joins the Stan development team.
- Added `pystan.lookup` (contributed by Marco Inacio, @randommm)
- NOTE: Stan v2.16.0 is the final release which will not require a C++11 compatible compiler. Future releases will require a C++11 compatible compiler. The vast majority of users have a compatible compiler.

### 1.1.4 v2.15.0.1 (2. May 2017)

- Python 2.7 compatibility fix (#332). Thanks to @monga for the report.

### 1.1.5 v2.15.0.0 (21. Apr 2017)

- Update Stan source to v2.15.0 ([release notes](#))
- Allow UTF-8 comments in Stan Program code. Thanks to @ahartikainen
- Expose constrain\_pars method, thanks to Lars Mescheder.

### 1.1.6 V2.14.0.0 (1. Jan 2017)

- Update Stan source to v2.14.0 ([release notes](#)), includes important fix to the default sampling algorithm (NUTS). All users are encouraged to upgrade.
- Several documentation and minor bug fixes (thanks @ahartikainen, @jrings, @nesanders)
- New OpenPGP signing key for use with PyPI. Key fingerprint is C3542448245BEC68F43070E4CCB669D9761F0CAC.

### 1.1.7 V2.12.0.0 (15. Sept 2016)

- Update Stan source to v2.12.0 ([release notes](#))
- #239 Fix bug in array indexing (thanks @stephen-hoover)
- #254 Fix off-by-one error in estimated sample size calculation

### 1.1.8 V2.11.0.0 (28. July 2016)

- Update Stan source to v2.11.0 ([release notes](#))

### 1.1.9 V2.10.0.0 (18. July 2016)

- Update Stan source to v2.10.0 ([release notes](#))
- Sampling in Fixed\_param mode now works. Thanks to @luac for the fix and @axch for the original report.
- Detailed installation instructions from @chendaniely added to the documentation.

### 1.1.10 v2.9.0.0 (7. Jan 2016)

- Update Stan source to v2.9.0 ([release notes](#))
- Bugs fixed in \_chains.pyx and model.py (thanks to @stephen-hoover, Paul Kernfeld)

### 1.1.11 v2.8.0.2 (6. Nov 2015)

- Cython 0.22 or higher requirement included on PyPI



### 1.1.12 v2.8.0.1 (5. Nov 2015)

- Python 3.5 support added
- Cython 0.22 or higher now required
- Compiler optimization (`-O2`) turned on for model compilation. This should increase sampling speed.
- Significant bug fixes (pickling, `pars` keyword)

### 1.1.13 v2.8.0.0 (1. Oct 2015)

- Update Stan source to v2.8.0 ([release notes](#))

### 1.1.14 v2.7.0.1 (22. August 2015)

- Minor Cython 0.23.1 compatibility fixes
- Bug preventing `mean_pars` from being recorded

### 1.1.15 v2.7.0.0 (21. July 2015)

- Update Stan source to v2.7.0 ([release notes](#))

### 1.1.16 v2.6.3.0 (21. Mar 2015)

- Update Stan source to v2.6.3 ([release notes](#)).

### 1.1.17 v2.6.0.0 (9. Feb 2015)

- Update Stan source to v2.6.0 ([release notes](#)).

### 1.1.18 v2.5.0.2 (30. Jan 2015)

- Fix bug in `rdump` (for `>1` dimensional arrays)

### 1.1.19 v2.5.0.1 (14. Nov 2014)

- Support for pickling fit objects (experimental)
- Fix bug that occurs when printing fit summary

### 1.1.20 v2.5.0.0 (21. Oct 2014)

- Update Stan source to v2.5.0
- Fix several significant bugs in the `extract` method

### 1.1.21 v2.4.0.3 (9. Sept 2014)

- Performance improvements for the printed summary of a fit.

### 1.1.22 v2.4.0.2 (6. Sept 2014)

- Performance improvements for the `extract` method (5-10 times faster)
- Performance improvements for the printed summary of a fit. Printing a summary of a model with more than a hundred parameters is not recommended. Consider using `extract` and calculating summary statistics for the parameters of interest.

### 1.1.23 v2.4.0.1 (31. July 2014)

- Sets LBFGS as default optimizer.
- Adds preliminary support for Python binary wheels on OS X and Windows.
- Fixes bug in edge case in new summary code.

### 1.1.24 v2.4.0.0 (26. July 2014)

- Stan 2.4 (LBFGS optimizer added, Nesterov removed)
- Improve display of fit summaries

### 1.1.25 v2.3.0.0 (26. June 2014)

- Stan 2.3 (includes user-defined functions, among other improvements).
- Optimizing returns a vector (array) by default instead of a dictionary.

### 1.1.26 v2.2.0.1 (30. April 2014)

- Add support for reading Stan's R dump files.
- Add support for specifying parameters of interest in `stan`.
- Add Windows installation instructions. Thanks to @patricksnape.
- Lighten source distribution.

### 1.1.27 v2.2.0.0 (16. February 2014)

- Updates Stan to v2.2.0.

### 1.1.28 v2.1.0.1 (27. January 2014)

- Implement model name obfuscation. Thanks to @karnold
- Improve documentation of StanFit objects

### 1.1.29 v2.1.0.0 (26. December 2013)

- Updates Stan code to v2.1.0.

### 1.1.30 v2.0.1.3 (18. December 2013)

- Sampling is parallel by default.
- `grad_log_prob` method of fit objects is available.

### 1.1.31 v2.0.1.2 (1. December 2013)

- Improves `setuptools` support.
- Allows sampling chains in parallel using multiprocessing. See the `n_jobs` parameter for `stan()` and the `sampling` method.
- Allows users to specify initial values for chains.

### 1.1.32 v2.0.1.1 (18. November 2013)

- Clean up `random_seed` handling (Stephan Hoyer).
- Add fit methods `get_seed`, `get_inits`, and `get_stancode`.

### 1.1.33 v2.0.1.0 (24. October 2013)

- Updated to Stan 2.0.1.
- Specifying `sample_file` now works as expected.

### 1.1.34 v2.0.0.1 (23. October 2013)

- Stan `array` parameters are now handled correctly.
- Ancillary methods added to fit instances.
- Fixed bug that caused parameters in `control` dict to be ignored.

### 1.1.35 v2.0.0.0 (21. October 2013)

- Stan source updated to to 2.0.0.
- PyStan version now mirrors Stan version.
- Rudimentary plot and traceplot methods have been added to fit instances.
- Warmup and sampling progress now visible.

### 1.1.36 v.0.2.2 (28. September 2013)

- `log_prob` method available from `StanFit` instances.
- Estimated sample size and `Rhat` included in summary.

### 1.1.37 v.0.2.1 (17. September 2013)

- StanModel instances can now be pickled.
- Adds basic support for saving output to `sample_file`.

### 1.1.38 v.0.2.0 (25. August 2013)

- `optimizing` method working for scalar, vector, and matrix parameters
- stanfit objects now have `summary` and `__str__` methods à la RStan
- stan source updated to commit `cc82d51d492d26f754fd56efe22a99191c80217b` (July 26, 2013)
- IPython-relevant bug fixes

### 1.1.39 v.0.1.1 (19. July 2013)

- Support for Python 2.7 and Python 3.3
- `stan` and `stanc` working with common arguments

## 1.2 Getting started

PyStan is the [Python](#) interface for [Stan](#).

### 1.2.1 Prerequisites

PyStan has the following dependencies:

- Python: 2.7, >=3.3
- Cython: >=0.22
- NumPy: >=1.7

PyStan also requires that a C++ compiler be available to Python during installation and at runtime. On Debian-based systems this is accomplished by issuing the command `apt-get install build-essential`.

### 1.2.2 Installation

---

**Note:** Installing PyStan involves compiling Stan. This may take a considerable amount of time.

---

#### Unix-based systems including Mac OS X

PyStan and the required packages may be installed from the [Python Package Index](#) using `pip`.

```
pip install pystan
```

To install PyStan from source, first make sure you have installed the dependencies, then issue the commands:

```
wget https://pypi.python.org/packages/source/p/pystan/pystan-2.17.1.0.tar.gz
# alternatively, use curl, or a web browser
tar zxvf pystan-2.17.1.0.tar.gz
cd pystan-2.17.1.0
python setup.py install
cd .. # change out of the source directory before importing pystan
```

Mac OS X users encountering installation problems may wish to consult the [PyStan Wiki](#) for possible solutions.

## Windows

PyStan on Windows requires Python 3.5 or higher and a working C++ compiler. If you have already installed Python 3.5 (or higher) and the Microsoft Visual C++ 14.0 (or higher) compiler, running `pip install pystan` will install PyStan. Note that you must specify `n_jobs=1` when drawing samples using Stan because PyStan on Windows is not currently able to use multiple processors simultaneously.

If you need to install a C++ compiler, you will find detailed installation instructions in [PyStan on Windows](#).

### 1.2.3 Using PyStan

The module's name is `pystan` so we load the module as follows:

```
import pystan
```

#### Example 1: Eight Schools

The “eight schools” example appears in Section 5.5 of Gelman et al. (2003), which studied coaching effects from eight schools.

```
schools_code = """
data {
  int<lower=0> J; // number of schools
  real y[J]; // estimated treatment effects
  real<lower=0> sigma[J]; // s.e. of effect estimates
}
parameters {
  real mu;
  real<lower=0> tau;
  real eta[J];
}
transformed parameters {
  real theta[J];
  for (j in 1:J)
    theta[j] = mu + tau * eta[j];
}
model {
  eta ~ normal(0, 1);
  y ~ normal(theta, sigma);
}
"""

schools_dat = {'J': 8,
               'y': [28, 8, -3, 7, -1, 1, 18, 12],
               'sigma': [15, 10, 16, 11, 9, 11, 10, 18]}
```

```
sm = pystan.StanModel(model_code=schools_code)
fit = sm.sampling(data=schools_dat, iter=1000, chains=4)
```

In this model, we let `theta` be transformed parameters of `mu` and `eta` instead of directly declaring `theta` as parameters. By parameterizing this way, the sampler will run more efficiently.

In PyStan, we can also specify the Stan model using a file. For example, we can download the file `8schools.stan` into our working directory and use the following call to `stan` instead:

```
sm = pystan.StanModel(file='8schools.stan')
fit = sm.sampling(data=schools_dat, iter=1000, chains=4)
```

Once a model is compiled, we can use the `StanModel` object multiple times. This saves us time compiling the C++ code for the model. For example, if we want to sample more iterations, we proceed as follows:

```
fit2 = sm.sampling(data=schools_dat, iter=10000, chains=4)
```

The object `fit`, returned from function `stan` stores samples from the posterior distribution. The `fit` object has a number of methods, including `plot` and `extract`. We can also print the `fit` object and receive a summary of the posterior samples as well as the log-posterior (which has the name `lp__`).

The method `extract` extracts samples into a dictionary of arrays for parameters of interest, or just an array.

```
la = fit.extract(permuted=True) # return a dictionary of arrays
mu = la['mu']

## return an array of three dimensions: iterations, chains, parameters
a = fit.extract(permuted=False)
```

```
print(fit)
```

If `matplotlib` and `scipy` are installed, a visual summary may also be displayed using the `plot()` method.

```
fit.plot()
```

## 1.3 Detailed Installation Instructions

The following is addressed to an audience who is just getting started with Python and would benefit from additional guidance on how to install PyStan.

Installing PyStan requires installing:

- Python
- Python dependencies
- PyStan

## 1.4 Prerequisite knowledge

It is highly recommended to know what `bash` is and the basics of navigating the terminal. You can review or learn it from the [Software Carpentry bash lesson](http://swcarpentry.github.io/shell-novice/) here: <http://swcarpentry.github.io/shell-novice/>.

Lessons 1 - 3 are probably the most important.

### 1.4.1 Installing Python

The easiest way to install Python is to use the Anaconda distribution of python. It can be downloaded here: <http://continuum.io/downloads>.

This is because PyStan (and many python tools) require packages (aka modules) that have C dependencies. These types of dependencies are unable to be installed (at least easily) using `pip`, which is a common way to install python packages. Anaconda ships with it's own package manager (that also plays nicely with `pip`) called `conda`, and comes with many of the data analytics packages and dependencies pre-installed.

Don't worry about Anaconda ruining your current Python installation, it can be easily uninstalled (described below).

#### Anaconda is not a requirement

Anaconda is not an absolute requirement to get `pystan` to work. As long as you can get the necessary python dependencies installed, `pystan` will work. If you want to install Anaconda, follow the Windows, Macs, and Linux instructions below.

#### Windows

The Anaconda installer should be able to be double-clicked and installed. Use all of the defaults for installation except make sure to check Make Anaconda the default Python.

#### Macs

After downloading the installer, double click the `.pkg` file and follow the instructions on the screen. Use all of the defaults for installation.

#### Linux

After downloading the installer execute the associated shell script. For example, if the file downloaded were named `Anaconda3-4.1.1-Linux-x86_64.sh` you would enter `bash Anaconda3-4.1.1-Linux-x86_64.sh` in the directory where you downloaded the file.

#### Uninstalling Anaconda

The default location for anaconda can be found in your home directory. Typically this means it in in the `~/anaconda` or `~/anaconda3` directory when you open a terminal.

### 1.4.2 Python dependencies

If you used the Anaconda installer, `numpy` and `cython` should already be installed, so additional dependencies should not be needed. However, should you need to install additional dependencies, we can use `conda` to install them as such:

- open a terminal
- type `conda install numpy` to install `numpy` or replace `numpy` with the package you need to install

### 1.4.3 Installing PyStan

Since we have the `numpy` and `cython` dependencies we need, we can install the latest version of PyStan using `pip`. To do so:

- Open a terminal
- type `pip install pystan`

## 1.5 Optimization in Stan

PyStan provides an interface to Stan's optimization methods. These methods obtain a point estimate by maximizing the posterior function defined for a model. The following example estimates the mean from samples assumed to be drawn from normal distribution with known standard deviation:

Specifying an improper prior for  $\mu$  of  $p(\mu) \propto 1$ , the posterior obtains a maximum at the sample mean. The following Python code illustrates how to use Stan's optimizer methods via a call to `optimizing`:

```
import pystan
import numpy as np

ocode = """
data {
  int<lower=1> N;
  real y[N];
}
parameters {
  real mu;
}
model {
  y ~ normal(mu, 1);
}
"""

sm = pystan.StanModel(model_code=ocode)
y2 = np.random.normal(size=20)
np.mean(y2)

op = sm.optimizing(data=dict(y=y2, N=len(y2)))

op
```

## 1.6 Avoiding recompilation of Stan models

Compiling models takes time. It is in our interest to avoid recompiling models whenever possible. If the same model is going to be used repeatedly, we would like to compile it just once. The following demonstrates how to reuse a model in different scripts and between interactive Python sessions.

**Within sessions** you can avoid recompiling a model by reusing the *StanModel* instance:

```
from pystan import stan

# bernoulli model
model_code = """
data {
  int<lower=0> N;
```



```

    int<lower=0,upper=1> y[N];
  }
  parameters {
    real<lower=0,upper=1> theta;
  }
  model {
    theta ~ beta(0.5, 0.5); // Jeffreys' prior
    for (n in 1:N)
      y[n] ~ bernoulli(theta);
  }
"""

```

```

from pystan import StanModel

data = dict(N=10, y=[0, 1, 0, 1, 0, 1, 0, 1, 1, 1])
sm = StanModel(model_code=model_code)
fit = sm.sampling(data=data)
print(fit)

# reuse model with new data
new_data = dict(N=6, y=[0, 0, 0, 0, 0, 1])
fit2 = sm.sampling(data=new_data)
print(fit2)

```

It is also possible to share models **between sessions** (or between different Python scripts). We do this by saving compiled models (`StanModel` instances) in a file and then reloading it when we need it later. (In short, `StanModel` instances are `picklable`.)

The following two code blocks illustrate how a model may be compiled in one session and reloaded in a subsequent one using `pickle` (part of the Python standard library).

```

import pickle
from pystan import StanModel

# using the same model as before
data = dict(N=10, y=[0, 1, 0, 1, 0, 1, 0, 1, 1, 1])
sm = StanModel(model_code=model_code)
fit = sm.sampling(data=data)
print(fit)

# save it to the file 'model.pkl' for later use
with open('model.pkl', 'wb') as f:
    pickle.dump(sm, f)

```

The following block of code might appear in a different script (in the same directory).

```

import pickle

sm = pickle.load(open('model.pkl', 'rb'))

new_data = dict(N=6, y=[0, 0, 0, 0, 0, 1])
fit2 = sm.sampling(data=new_data)
print(fit2)

```

## 1.6.1 Automatically reusing models

For those who miss using variables across sessions in R, it is not difficult to write a function that automatically saves a copy of every model that gets compiled and opportunistically loads a copy of a model if one is available.

```
import pystan
import pickle
from hashlib import md5

def StanModel_cache(model_code, model_name=None, **kwargs):
    """Use just as you would `stan`"""
    code_hash = md5(model_code.encode('ascii')).hexdigest()
    if model_name is None:
        cache_fn = 'cached-model-{}.pkl'.format(code_hash)
    else:
        cache_fn = 'cached-{}-{}.pkl'.format(model_name, code_hash)
    try:
        sm = pickle.load(open(cache_fn, 'rb'))
    except:
        sm = pystan.StanModel(model_code=model_code)
        with open(cache_fn, 'wb') as f:
            pickle.dump(sm, f)
    else:
        print("Using cached StanModel")
    return sm

# with same model_code as before
data = dict(N=10, y=[0, 1, 0, 0, 0, 0, 0, 0, 0, 1])
sm = StanModel_cache(model_code=model_code)
fit = sm.sampling(data=data)
print(fit)

new_data = dict(N=6, y=[0, 0, 0, 0, 0, 1])
# the cached copy of the model will be used
sm = StanModel_cache(model_code=model_code)
fit2 = sm.sampling(data=new_data)
print(fit2)
```

## 1.7 Differences between PyStan and RStan

While PyStan attempts to maintain API compatibility with RStan, there are certain unavoidable differences between Python and R.

### 1.7.1 Methods and attributes

Methods are invoked in different ways: `fit.summary()` and `fit.extract()` (Python) vs. `summary(fit)` and `extract(fit)` (R).

Attributes are accessed in a different manner as well: `fit.sim` (Python) vs. `fit@sim` (R).

### 1.7.2 Dictionaries instead of Lists

Where RStan uses lists, PyStan uses (ordered) dictionaries.

Python:

```
fit.extract()['theta']
```

R:

```
extract(fit)$theta
```

### 1.7.3 Reusing models and saving objects

PyStan uses `pickle` to save objects for future use.

Python:

```
import pickle
import pystan

# bernoulli model
model_code = """
    data {
      int<lower=0> N;
      int<lower=0,upper=1> y[N];
    }
    parameters {
      real<lower=0,upper=1> theta;
    }
    model {
      for (n in 1:N)
        y[n] ~ bernoulli(theta);
    }
  """
data = dict(N=10, y=[0, 1, 0, 0, 0, 0, 0, 0, 0, 1])
model = pystan.StanModel(model_code=model_code)
fit = model.sampling(data=data)

with open('model.pkl', 'wb') as f:
    pickle.dump(model, f, protocol=pickle.HIGHEST_PROTOCOL)

# load it at some future point
with open('model.pkl', 'rb') as f:
    model = pickle.load(f)

# run with different data
fit = model.sampling(data=dict(N=5, y=[1, 1, 0, 1, 0]))
```

R:

```
library(rstan)

model = stan_model(model_code=model_code)
save(model, file='model.rdata')
```

See also *Avoiding recompilation of Stan models*.

If you are saving a large amount of data with `pickle.dump`, be sure to use the highest protocol version available. Earlier versions are limited in the amount of data they can save in a single file.

## 1.8 API

---

<code>stan</code> ([file, model_name, model_code, fit, ...])	Fit a model using Stan.
<code>stan</code> ([file, charset, model_code, ...])	Translate Stan model specification into C++ code.
<code>StanModel</code> ([file, charset, model_name, ...])	Model described in Stan's modeling language compiled from C++ code.

---

`StanFit4model` instances are also documented on this page.

```
pystan.stan(file=None, model_name='anon_model', model_code=None, fit=None, data=None,
            pars=None, chains=4, iter=2000, warmup=None, thin=1, init='random', seed=None,
            algorithm=None, control=None, sample_file=None, diagnostic_file=None, verbose=False,
            boost_lib=None, eigen_lib=None, n_jobs=-1, **kwargs)
```

Fit a model using Stan.

The `pystan.stan` function was deprecated in version 2.17 and will be removed in version 3.0. Compiling and using a Stan Program (e.g., for drawing samples) should be done in separate steps.

**Parameters** `file` : string { 'filename', file-like object }

Model code must found via one of the following parameters: `file` or `model_code`.

If `file` is a filename, the string passed as an argument is expected to be a filename containing the Stan model specification.

If `file` is a file object, the object passed must have a 'read' method (file-like object) that is called to fetch the Stan model specification.

**charset** : string, optional

If bytes or files are provided, this charset is used to decode. 'utf-8' by default.

**model\_code** : string

A string containing the Stan model specification. Alternatively, the model may be provided with the parameter `file`.

**model\_name**: string, optional :

A string naming the model. If none is provided 'anon\_model' is the default. However, if `file` is a filename, then the filename will be used to provide a name. 'anon\_model' by default.

**fit** : StanFit instance

An instance of StanFit derived from a previous fit, None by default. If `fit` is not None, the compiled model associated with a previous fit is reused and recompilation is avoided.

**data** : dict

A Python dictionary providing the data for the model. Variables for Stan are stored in the dictionary as expected. Variable names are the keys and the values are their associated values. Stan only accepts certain kinds of values; see Notes.

**pars** : list of string, optional

A list of strings indicating parameters of interest. By default all parameters specified in the model will be stored.

**chains** : int, optional

Positive integer specifying number of chains. 4 by default.

**iter** : int, 2000 by default

Positive integer specifying how many iterations for each chain including warmup.

**warmup** : int, iter//2 by default

Positive integer specifying number of warmup (aka burn) iterations. As *warmup* also specifies the number of iterations used for stepsize adaption, warmup samples should not be used for inference.

**thin** : int, optional

Positive integer specifying the period for saving samples. Default is 1.

**init** : {0, '0', 'random', function returning dict, list of dict}, optional

Specifies how initial parameter values are chosen: - 0 or '0' initializes all to be zero on the unconstrained support. - 'random' generates random initial values. An optional parameter

*init\_r* controls the range of randomly generated initial values for parameters in terms of their unconstrained support;

- **list of size equal to the number of chains (*chains*), where the** list contains a dict with initial parameter values;
- **function returning a dict with initial parameter values.** The function may take an optional argument *chain\_id*.

**seed** : int or np.random.RandomState, optional

The seed, a positive integer for random number generation. Only one seed is needed when multiple chains are used, as the other chain's seeds are generated from the first chain's to prevent dependency among random number streams. By default, seed is `random.randint(0, MAX_UINT)`.

**algorithm** : {"NUTS", "HMC", "Fixed\_param"}, optional

One of the algorithms that are implemented in Stan such as the No-U-Turn sampler (NUTS, Hoffman and Gelman 2011) and static HMC.

**sample\_file** : string, optional

File name specifying where samples for *all* parameters and other saved quantities will be written. If not provided, no samples will be written. If the folder given is not writable, a temporary directory will be used. When there are multiple chains, an underscore and chain number are appended to the file name. By default do not write samples to file.

**diagnostic\_file** : string, optional

File name specifying where diagnostic information should be written. By default no diagnostic information is recorded.

**boost\_lib** : string, optional

The path to a version of the Boost C++ library to use instead of the one supplied with PyStan.

**eigen\_lib** : string, optional

The path to a version of the Eigen C++ library to use instead of the one in the supplied with PyStan.

**verbose** : boolean, optional

Indicates whether intermediate output should be piped to the console. This output may be useful for debugging. False by default.

**control** : dict, optional

A dictionary of parameters to control the sampler's behavior. Default values are used if control is not specified. The following are adaptation parameters for sampling algorithms.

These are parameters used in Stan with similar names:

- *adapt\_engaged* : bool
- *adapt\_gamma* : float, positive, default 0.05
- *adapt\_delta* : float, between 0 and 1, default 0.8
- *adapt\_kappa* : float, between default 0.75
- *adapt\_t0* : float, positive, default 10
- *adapt\_init\_buffer* : int, positive, defaults to 75
- *adapt\_term\_buffer* : int, positive, defaults to 50
- *adapt\_window* : int, positive, defaults to 25

In addition, the algorithm HMC (called 'static HMC' in Stan) and NUTS share the following parameters:

- *stepsize*: float, positive
- *stepsize\_jitter*: float, between 0 and 1
- *metric* : str, {"unit\_e", "diag\_e", "dense\_e"}

In addition, depending on which algorithm is used, different parameters can be set as in Stan for sampling. For the algorithm HMC we can set

- *int\_time*: float, positive

For algorithm NUTS, we can set

- *max\_treedepth* : int, positive

**n\_jobs** : int, optional

Sample in parallel. If -1 all CPUs are used. If 1, no parallel computing code is used at all, which is useful for debugging.

**Returns** *fit* : StanFit instance

**Other Parameters** *chain\_id* : int, optional

*chain\_id* can be a vector to specify the *chain\_id* for all chains or an integer. For the former case, they should be unique. For the latter, the sequence of integers starting from the given *chain\_id* are used for all chains.

**init\_r** : float, optional

*init\_r* is only valid if *init* == "random". In this case, the initial values are simulated from  $[-init_r, init_r]$  rather than using the default interval (see the manual of (Cmd)Stan).

**test\_grad**: bool, optional :

If *test\_grad* is True, Stan will not do any sampling. Instead, the gradient calculation is tested and printed out and the fitted StanFit4Model object is in test gradient mode. By default, it is False.

**append\_samples': bool, optional :**

**refresh': int, optional :**

Argument *refresh* can be used to control how to indicate the progress during sampling (i.e. show the progress every code{refresh} iterations). By default, *refresh* is  $\max(\text{iter}/10, 1)$ .

**obfuscate\_model\_name** : boolean, optional

*obfuscate\_model\_name* is only valid if *fit* is None. True by default. If False the model name in the generated C++ code will not be made unique by the insertion of randomly generated characters. Generally it is recommended that this parameter be left as True.

## Examples

```
>>> from pystan import stan
>>> import numpy as np
>>> model_code = '''
... parameters {
...   real y[2];
... }
... model {
...   y[1] ~ normal(0, 1);
...   y[2] ~ double_exponential(0, 2);
... }'''
>>> fit1 = stan(model_code=model_code, iter=10)
>>> print(fit1)
>>> excode = '''
... transformed data {
...   real y[20];
...   y[1] <- 0.5796; y[2] <- 0.2276; y[3] <- -0.2959;
...   y[4] <- -0.3742; y[5] <- 0.3885; y[6] <- -2.1585;
...   y[7] <- 0.7111; y[8] <- 1.4424; y[9] <- 2.5430;
...   y[10] <- 0.3746; y[11] <- 0.4773; y[12] <- 0.1803;
...   y[13] <- 0.5215; y[14] <- -1.6044; y[15] <- -0.6703;
...   y[16] <- 0.9459; y[17] <- -0.382; y[18] <- 0.7619;
...   y[19] <- 0.1006; y[20] <- -1.7461;
... }
... parameters {
...   real mu;
...   real<lower=0, upper=10> sigma;
...   vector[2] z[3];
...   real<lower=0> alpha;
... }
... model {
...   y ~ normal(mu, sigma);
...   for (i in 1:3)
...     z[i] ~ normal(0, 1);
...   alpha ~ exponential(2);
... }'''
>>>
>>> def initfun1():
...     return dict(mu=1, sigma=4, z=np.random.normal(size=(3, 2)), alpha=1)
>>> exfit0 = stan(model_code=excode, init=initfun1)
>>> def initfun2(chain_id=1):
...     return dict(mu=1, sigma=4, z=np.random.normal(size=(3, 2)), alpha=1 +
↳chain_id)
```

```
>>> exfit1 = stan(model_code=excode, init=initfun2)
```

`pystan.stanc` (*file=None, charset='utf-8', model\_code=None, model\_name='anon\_model, verbose=False, obfuscate\_model\_name=True*)  
Translate Stan model specification into C++ code.

**Parameters** `file` : {string, file}, optional

If filename, the string passed as an argument is expected to be a filename containing the Stan model specification.

If file, the object passed must have a 'read' method (file-like object) that is called to fetch the Stan model specification.

**charset** : string, 'utf-8' by default

If bytes or files are provided, this charset is used to decode.

**model\_code** : string, optional

A string containing the Stan model specification. Alternatively, the model may be provided with the parameter *file*.

**model\_name**: string, 'anon\_model' by default :

A string naming the model. If none is provided 'anon\_model' is the default. However, if *file* is a filename, then the filename will be used to provide a name.

**verbose** : boolean, False by default

Indicates whether intermediate output should be piped to the console. This output may be useful for debugging.

**obfuscate\_model\_name** : boolean, True by default

If False the model name in the generated C++ code will not be made unique by the insertion of randomly generated characters. Generally it is recommended that this parameter be left as True.

**Returns** `stanc_ret` : dict

A dictionary with the following keys: `model_name`, `model_code`, `cpp_code`, and `status`. Status indicates the success of the translation from Stan code into C++ code (success = 0, error = -1).

**See also:**

`StanModel` Class representing a compiled Stan model

`stan` Fit a model using Stan

## Notes

C++ reserved words and Stan reserved words may not be used for variable names; see the Stan User's Guide for a complete list.

## References

The Stan Development Team (2013) *Stan Modeling Language User's Guide and Reference Manual*. <<http://mc-stan.org/>>.



## Examples

```

>>> stanmodelcode = '''
... data {
...   int<lower=0> N;
...   real y[N];
... }
...
... parameters {
...   real mu;
... }
...
... model {
...   mu ~ normal(0, 10);
...   y ~ normal(mu, 1);
... }
... '''
>>> r = stanc(model_code=stanmodelcode, model_name = "normal1")
>>> sorted(r.keys())
['cppcode', 'model_code', 'model_cppname', 'model_name', 'status']
>>> r['model_name']
'normal1'

```

**class** `pystan.StanModel` (*file=None*, *charset='utf-8'*, *model\_name='anon\_model'*,  
*model\_code=None*, *stanc\_ret=None*, *boost\_lib=None*, *eigen\_lib=None*,  
*verbose=False*, *obfuscate\_model\_name=True*, *extra\_compile\_args=None*)

Model described in Stan's modeling language compiled from C++ code.

Instances of `StanModel` are typically created indirectly by the functions `stan` and `stanc`.

**Parameters** `file` : string {'filename', 'file'}

If `filename`, the string passed as an argument is expected to be a filename containing the Stan model specification.

If `file`, the object passed must have a 'read' method (file-like object) that is called to fetch the Stan model specification.

**charset** : string, 'utf-8' by default

If bytes or files are provided, this charset is used to decode.

**model\_name**: string, 'anon\_model' by default :

A string naming the model. If none is provided 'anon\_model' is the default. However, if `file` is a filename, then the filename will be used to provide a name.

**model\_code** : string

A string containing the Stan model specification. Alternatively, the model may be provided with the parameter `file`.

**stanc\_ret** : dict

A dict returned from a previous call to `stanc` which can be used to specify the model instead of using the parameter `file` or `model_code`.

**boost\_lib** : string

The path to a version of the Boost C++ library to use instead of the one supplied with PyStan.

**eigen\_lib** : string

The path to a version of the Eigen C++ library to use instead of the one in the supplied with PyStan.

**verbose** : boolean, False by default

Indicates whether intermediate output should be piped to the console. This output may be useful for debugging.

**kwargs** : keyword arguments

Additional arguments passed to *stanc*.

**See also:**

*stanc* Compile a Stan model specification

*stan* Fit a model using Stan

## Notes

More details of Stan, including the full user's guide and reference manual can be found at <URL: <http://mc-stan.org/>>.

There are three ways to specify the model's code for *stan\_model*.

1. parameter *model\_code*, containing a string to whose value is the Stan model specification,
2. parameter *file*, indicating a file (or a connection) from which to read the Stan model specification, or
3. **parameter *stanc\_ret*, indicating the re-use of a model** generated in a previous call to *stanc*.

## References

The Stan Development Team (2013) *Stan Modeling Language User's Guide and Reference Manual*. <URL: <http://mc-stan.org/>>.

## Examples

```
>>> model_code = 'parameters {real y;} model {y ~ normal(0,1);}'
>>> model_code; m = StanModel(model_code=model_code)
...
'parameters ...
>>> m.model_name
'anon_model'
```

## Attributes

model_name	(string)
model_code	(string) Stan code for the model.
model_cpp	(string) C++ code for the model.
module	(builtins.module) Python module created by compiling the C++ code for the model.

## Methods

<code>sampling</code> ([data, pars, chains, iter, warmup, ...])	Draw samples from the model.
<code>optimizing</code> ([data, seed, init, sample_file, ...])	Obtain a point estimate by maximizing the joint posterior.

<code>show</code>	Print the Stan model specification.
<code>get_cppcode</code>	Return the C++ code for the module.
<code>get_cxxflags</code>	Return the 'CXXFLAGS' used for compiling the model.

**optimizing** (*data=None, seed=None, init='random', sample\_file=None, algorithm=None, verbose=False, as\_vector=True, \*\*kwargs*)  
Obtain a point estimate by maximizing the joint posterior.

**Parameters** *data* : dict

A Python dictionary providing the data for the model. Variables for Stan are stored in the dictionary as expected. Variable names are the keys and the values are their associated values. Stan only accepts certain kinds of values; see Notes.

**seed** : int or `np.random.RandomState`, optional

The seed, a positive integer for random number generation. Only one seed is needed when multiple chains are used, as the other chain's seeds are generated from the first chain's to prevent dependency among random number streams. By default, seed is `random.randint(0, MAX_UINT)`.

**init** : {0, '0', 'random', function returning dict, list of dict}, optional

Specifies how initial parameter values are chosen: - 0 or '0' initializes all to be zero on the unconstrained support. - 'random' generates random initial values. An optional parameter

*init\_r* controls the range of randomly generated initial values for parameters in terms of their unconstrained support;

- list of size equal to the number of chains (*chains*), where the list contains a dict with initial parameter values;
- function returning a dict with initial parameter values. The function may take an optional argument *chain\_id*.

**sample\_file** : string, optional

File name specifying where samples for *all* parameters and other saved quantities will be written. If not provided, no samples will be written. If the folder given is not writable, a temporary directory will be used. When there are multiple chains, an underscore and chain number are appended to the file name. By default do not write samples to file.

**algorithm** : {"LBFSGS", "BFGS", "Newton"}, optional

Name of optimization algorithm to be used. Default is LBFSGS.

**verbose** : boolean, optional

Indicates whether intermediate output should be piped to the console. This output may be useful for debugging. False by default.

**as\_vector** : boolean, optional

Indicates an OrderedDict will be returned rather than a nested dictionary with keys 'par' and 'value'.

**Returns optim** : OrderedDict

Depending on *as\_vector*, returns either an OrderedDict having parameters as keys and point estimates as values or an OrderedDict with components 'par' and 'value'. `optim['par']` is a dictionary of point estimates, indexed by the parameter name. `optim['value']` stores the value of the log-posterior (up to an additive constant, the `lp__` in Stan) corresponding to the point identified by `optim['par']`.

**Other Parameters iter** : int, optional

The maximum number of iterations.

**save\_iterations** : bool, optional

**refresh** : int, optional

**init\_alpha** : float, optional

For BFGS and LBFGS, default is 0.001

**tol\_obj** : float, optional

For BFGS and LBFGS, default is 1e-12.

**tol\_grad** : float, optional

For BFGS and LBFGS, default is 1e-8.

**tol\_param** : float, optional

For BFGS and LBFGS, default is 1e-8.

**tol\_rel\_grad** : float, optional

For BFGS and LBFGS, default is 1e7.

**history\_size** : int, optional

For LBFGS, default is 5.

**Refer to the manuals for both CmdStan and Stan for more details. :**

## Examples

```
>>> from pystan import StanModel
>>> m = StanModel(model_code='parameters {real y;} model {y ~ normal(0,1);}')
>>> f = m.optimizing()
```

**sampling** (*data=None, pars=None, chains=4, iter=2000, warmup=None, thin=1, seed=None, init='random', sample\_file=None, diagnostic\_file=None, verbose=False, algorithm=None, control=None, n\_jobs=-1, \*\*kwargs*)

Draw samples from the model.

**Parameters data** : dict

A Python dictionary providing the data for the model. Variables for Stan are stored in the dictionary as expected. Variable names are the keys and the values are their associated values. Stan only accepts certain kinds of values; see Notes.

**pars** : list of string, optional

A list of strings indicating parameters of interest. By default all parameters specified in the model will be stored.

**chains** : int, optional

Positive integer specifying number of chains. 4 by default.

**iter** : int, 2000 by default

Positive integer specifying how many iterations for each chain including warmup.

**warmup** : int, iter//2 by default

Positive integer specifying number of warmup (aka burn-in) iterations. As *warmup* also specifies the number of iterations used for step-size adaption, warmup samples should not be used for inference.

**thin** : int, 1 by default

Positive integer specifying the period for saving samples.

**seed** : int or np.random.RandomState, optional

The seed, a positive integer for random number generation. Only one seed is needed when multiple chains are used, as the other chain's seeds are generated from the first chain's to prevent dependency among random number streams. By default, seed is `random.randint(0, MAX_UINT)`.

**algorithm** : {"NUTS", "HMC", "Fixed\_param"}, optional

One of algorithms that are implemented in Stan such as the No-U-Turn sampler (NUTS, Hoffman and Gelman 2011), static HMC, or `Fixed_param`. Default is NUTS.

**init** : {0, '0', 'random', function returning dict, list of dict}, optional

Specifies how initial parameter values are chosen: 0 or '0' initializes all to be zero on the unconstrained support; 'random' generates random initial values; list of size equal to the number of chains (*chains*), where the list contains a dict with initial parameter values; function returning a dict with initial parameter values. The function may take an optional argument *chain\_id*.

**sample\_file** : string, optional

File name specifying where samples for *all* parameters and other saved quantities will be written. If not provided, no samples will be written. If the folder given is not writable, a temporary directory will be used. When there are multiple chains, an underscore and chain number are appended to the file name. By default do not write samples to file.

**verbose** : boolean, False by default

Indicates whether intermediate output should be piped to the console. This output may be useful for debugging.

**control** : dict, optional

A dictionary of parameters to control the sampler's behavior. Default values are used if control is not specified. The following are adaptation parameters for sampling algorithms.

These are parameters used in Stan with similar names:

- *adapt\_engaged* : bool, default True
- *adapt\_gamma* : float, positive, default 0.05
- *adapt\_delta* : float, between 0 and 1, default 0.8

- *adapt\_kappa* : float, between default 0.75
- *adapt\_t0* : float, positive, default 10

In addition, the algorithm HMC (called ‘static HMC’ in Stan) and NUTS share the following parameters:

- *stepsize*: float, positive
- *stepsize\_jitter*: float, between 0 and 1
- *metric* : str, {“unit\_e”, “diag\_e”, “dense\_e”}

In addition, depending on which algorithm is used, different parameters can be set as in Stan for sampling. For the algorithm HMC we can set

- *int\_time*: float, positive

For algorithm NUTS, we can set

- *max\_treedepth* : int, positive

**n\_jobs** : int, optional

Sample in parallel. If -1 all CPUs are used. If 1, no parallel computing code is used at all, which is useful for debugging.

**Returns** *fit* : StanFit4Model

Instance containing the fitted results.

**Other Parameters** *chain\_id* : int or iterable of int, optional

*chain\_id* can be a vector to specify the *chain\_id* for all chains or an integer. For the former case, they should be unique. For the latter, the sequence of integers starting from the given *chain\_id* are used for all chains.

**init\_r** : float, optional

*init\_r* is only valid if *init* == “random”. In this case, the initial values are simulated from  $[-init_r, init_r]$  rather than using the default interval (see the manual of Stan).

**test\_grad**: bool, optional :

If *test\_grad* is `True`, Stan will not do any sampling. Instead, the gradient calculation is tested and printed out and the fitted StanFit4Model object is in test gradient mode. By default, it is `False`.

**append\_samples**: bool, optional :

**refresh**: int, optional :

Argument *refresh* can be used to control how to indicate the progress during sampling (i.e. show the progress every `code{refresh}` iterations). By default, *refresh* is `max(iter/10, 1)`.

## Examples

```
>>> from pystan import StanModel
>>> m = StanModel(model_code='parameters {real y;} model {y ~ normal(0,1);}')
>>> m.sampling(iter=100)
```

**vb** (*data=None, pars=None, iter=10000, seed=None, init='random', sample\_file=None, diagnostic\_file=None, verbose=False, algorithm=None, \*\*kwargs*)  
 Call Stan's variational Bayes methods.

**Parameters data** : dict

A Python dictionary providing the data for the model. Variables for Stan are stored in the dictionary as expected. Variable names are the keys and the values are their associated values. Stan only accepts certain kinds of values; see Notes.

**pars** : list of string, optional

A list of strings indicating parameters of interest. By default all parameters specified in the model will be stored.

**seed** : int or np.random.RandomState, optional

The seed, a positive integer for random number generation. Only one seed is needed when multiple chains are used, as the other chain's seeds are generated from the first chain's to prevent dependency among random number streams. By default, seed is `random.randint(0, MAX_UINT)`.

**sample\_file** : string, optional

File name specifying where samples for *all* parameters and other saved quantities will be written. If not provided, samples will be written to a temporary file and read back in. If the folder given is not writable, a temporary directory will be used. When there are multiple chains, an underscore and chain number are appended to the file name. By default do not write samples to file.

**diagnostic\_file** : string, optional

File name specifying where diagnostics for the variational fit will be written.

**iter** : int, 10000 by default

Positive integer specifying how many iterations for each chain including warmup.

**algorithm** : {'meanfield', 'fullrank'}

algorithm}{One of "meanfield" and "fullrank" indicating which variational inference algorithm is used. meanfield: mean-field approximation; fullrank: full-rank covariance. The default is 'meanfield'.

**verbose** : boolean, False by default

Indicates whether intermediate output should be piped to the console. This output may be useful for debugging.

**Other optional parameters, refer to the manuals for both CmdStan :**

**and Stan. :**

- 'iter': the maximum number of iterations, defaults to 10000 :

- 'grad\_samples' the number of samples for Monte Carlo enumerate of :  
 gradients, defaults to 1.

- 'elbo\_samples' the number of samples for Monte Carlo estimate of ELBO :  
 (objective function), defaults to 100. (ELBO stands for "the evidence lower bound".)

- 'eta' positive stepsize weighting parameters for variational :

inference but is ignored if adaptation is engaged, which is the case by default.

- ‘**adapt\_engaged**’ flag indicating whether to automatically adapt the :  
stepsize and defaults to True.
- ‘**tol\_rel\_obj**’ convergence tolerance on the relative norm of the :  
objective, defaults to 0.01.
- ‘**eval\_elbo**’, evaluate ELBO every Nth iteration, defaults to 100 :
- ‘**output\_samples**’ number of posterior samples to draw and save, :  
defaults to 1000.
- ‘**adapt\_iter**’ number of iterations to adapt the stepsize if :  
*adapt\_engaged* is True and ignored otherwise.

**Returns results** : dict

Dictionary containing information related to results.

### Examples

```
>>> from pystan import StanModel
>>> m = StanModel(model_code='parameters {real y;} model {y ~ normal(0,1);}')
>>> results = m.vb()
>>> # results saved on disk in format inspired by CSV
>>> print(results['args']['sample_file'])
```

## 1.8.1 StanFit4model

Each StanFit instance is model-specific, so the name of the class will be something like: StanFit4anon\_model. The StanFit4model instances expose a number of methods.

**class** `pystan.StanFit4model`

**plot** (*pars=None*)

Visualize samples from posterior distributions

Parameters

**pars** [sequence of str] names of parameters

This is currently an alias for the *traceplot* method.

**extract** (*pars=None, permuted=True, inc\_warmup=False*)

Extract samples in different forms for different parameters.

Parameters

**pars** [sequence of str] names of parameters (including other quantities)

**permuted** [bool] If True, returned samples are permuted. All chains are merged and warmup samples are discarded.

**inc\_warmup** [bool] If True, warmup samples are kept; otherwise they are discarded. If *permuted* is True, *inc\_warmup* is ignored.



Returns

`samples` : dict or array If *permuted* is True, return dictionary with samples for each parameter (or other quantity) named in *pars*.

If *permuted* is False, an array is returned. The first dimension of the array is for the iterations; the second for the number of chains; the third for the parameters. Vectors and arrays are expanded to one parameter (a scalar) per cell, with names indicating the third dimension. Parameters are listed in the same order as *model\_pars* and *flatnames*.

**log\_prob** (*upar*, *adjust\_transform=True*, *gradient=False*)

Expose the `log_prob` of the model to `stan_fit` so user can call this function.

Parameters

**upar** : The real parameters on the unconstrained space.

**adjust\_transform** [bool] Whether we add the term due to the transform from constrained space to unconstrained space implicitly done in Stan.

Note

In Stan, the parameters need be defined with their supports. For example, for a variance parameter, we must define it on the positive real line. But inside Stan's sampler, all parameters defined on the constrained space are transformed to unconstrained space, so the log density function need be adjusted (i.e., adding the log of the absolute value of the Jacobian determinant). With the transformation, Stan's samplers work on the unconstrained space and once a new iteration is drawn, Stan transforms the parameters back to their supports. All the transformation are done inside Stan without interference from the users. However, when using the log density function for a model exposed to Python, we need to be careful. For example, if we are interested in finding the mode of parameters on the constrained space, we then do not need the adjustment. For this reason, there is an argument named *adjust\_transform* for functions *log\_prob* and *grad\_log\_prob*.

**grad\_log\_prob** (*upars*, *adjust\_transform=True*)

Expose the `grad_log_prob` of the model to `stan_fit` so user can call this function.

Parameters

**upar** [array] The real parameters on the unconstrained space.

**adjust\_transform** [bool] Whether we add the term due to the transform from constrained space to unconstrained space implicitly done in Stan.

**get\_adaptation\_info** ()

Obtain adaptation information for sampler, which now only NUTS2 has.

The results are returned as a list, each element of which is a character string for a chain.

**get\_logposterior** (*inc\_warmup=True*)

Get the log-posterior (up to an additive constant) for all chains.

Each element of the returned array is the log-posterior for a chain. Optional parameter *inc\_warmup* indicates whether to include the warmup period.

**get\_sampler\_params** (*inc\_warmup=True*)

Obtain the parameters used for the sampler such as *stepsize* and *treedepth*. The results are returned as a list, each element of which is an OrderedDict a chain. The dictionary has number of elements corresponding to the number of parameters used in the sampler. Optional parameter *inc\_warmup* indicates whether to include the warmup period.

**get\_posterior\_mean** ()

Get the posterior mean for all parameters

Returns

**means** [array of shape (num\_parameters, num\_chains)] Order of parameters is given by self.model\_pars or self.flatnames if parameters of interest include non-scalar parameters. An additional column for mean **lp\_\_** is also included.

**unconstrain\_pars** (*par*)

Transform parameters from defined support to unconstrained space

**get\_seed** ()

**get\_inits** ()

**get\_stancode** ()

## 1.9 Conversion utilities for Stan's R Dump format

---

<code>stan_rdump</code> (data, filename)	Dump a dictionary with model data into a file using the R dump format that Stan supports.
<code>read_rdump</code> (filename)	Read data formatted using the R dump format

---

`pystan.misc.stan_rdump` (*data*, *filename*)

Dump a dictionary with model data into a file using the R dump format that Stan supports.

**Parameters data** : dict

**filename** : str

`pystan.misc.read_rdump` (*filename*)

Read data formatted using the R dump format

**Parameters filename: str** :

**Returns data** : OrderedDict

## 1.10 PyStan on Windows

PyStan is partially supported under Windows with the following caveats:

- Python 3.5 or higher must be used.
- When drawing samples `n_jobs=1` must be used. (PyStan on Windows cannot use multiple processors in parallel.)
- Vectorized functions are not supported with MSVC (see GCC section)

PyStan requires a working C++ compiler. Configuring such a compiler is typically the most challenging step in getting PyStan running.

Due to problems with MSVC template deduction, functions with Eigen library are failing. Until this bug is fixed no full support is provided for Windows + MSVC. Currently, no fix is known for this problem, other than to change the compiler to GCC.

### 1.10.1 Installing Python

There several ways of installing PyStan on Windows. The following instructions assume you have installed Python 3.5 (or higher) as packaged in the [Anaconda Python distribution](#). (Make sure you install the Python 3 variant of Anaconda.)

The Anaconda distribution is well-maintained and includes packages such as Numpy which PyStan requires. The following instructions assume that you are using Windows 7. (Windows 10 disregards user choice and user privacy.)

## 1.10.2 Installing a C++ Compiler

This section describes how to install the Microsoft Visual C++ 14.0 compiler. It is likely that PyStan will work with more recent Microsoft Visual C++ compilers as well.

Navigate to the [Visual C++ Build Tools](#) page and click on “Download Visual C++ Build Tools 2015”.

If you encounter problems you may find the [Windows Compilers](#) page on the Python Wiki useful. Note that on Windows 7 you may need to update the installed version of the Microsoft .NET Framework before installing the Visual C++ Build Tools.

## 1.10.3 Installing PyStan

Once you have the compiler installed, installing PyStan is easy. Open the application called “Command Prompt” (`cmd.exe`) and enter the following command:

```
pip install pystan
```

You can verify that everything was installed successfully by opening up the Python terminal (run `python` from a command prompt) and drawing samples from a very simple model:

```
>>> import pystan
>>> model_code = 'parameters {real y;} model {y ~ normal(0,1);}'
>>> model = pystan.StanModel(model_code=model_code)
>>> y = model.sampling(n_jobs=1).extract()['y']
>>> y.mean() # with luck the result will be near 0
```

Again, remember that using `n_jobs=1` when calling `sampling` is required as PyStan on Windows does not support sampling using multiple processors in parallel.

## 1.10.4 GCC on Windows

For full support of vectorized functions, the user must use GCC compiler.

Supported GCC compilers

- Python 2.7 and <=3.4 : mingw32
- Python 3.5=< : MSYS2 mingw-w64

For mingw32 installation see <http://docs.cython.org/en/latest/src/tutorial/appendix.html>

or with conda use:

```
conda install mingw32 libpython
```

## 1.10.5 Setting up mingw-w64 on Windows

Install Python with either Anaconda or Miniconda

- Anaconda <https://www.anaconda.com/download/>
- Miniconda <https://conda.io/miniconda.html>

Create an virtual conda-environment [OPTIONAL]:

```
conda create -n stan_env python=3.6
```

Activate conda-env [OPTIONAL]:

```
activate stan_env
```

Install *libpython* and *m2w64-toolchain* packages:

```
conda install libpython
conda install -c msys2 m2w64-toolchain
```

In `PYTHONPATH\Lib\distutils` create `distutils.cfg` with text editor (e.g. *notepad*, *notepad++*) and add the following lines:

```
[build]
mingw32
```

To find the correct *distutils* path, run *python*:

```
>>> import distutils
>>> print(distutils.__file__)
```

Finally, install other packages:

```
conda install numpy cython matplotlib
pip install pystan
```

## CHAPTER 2

---

### Stan documentation

---

- Stan: <http://mc-stan.org/>
- Stan User's Guide and Reference Manual, available at <http://mc-stan.org>
- BUGS Examples



## CHAPTER 3

---

### Important links

---

- Source code repo: <https://github.com/stan-dev/pystan>
- HTML documentation: <http://pystan.readthedocs.org>
- Issue tracker: <https://github.com/stan-dev/pystan/issues>





## CHAPTER 4

---

### Similar projects

---

- PyMC: <http://pymc-devs.github.io/pymc/>
- emcee: <http://dan.iel.fm/emcee/current/>



## CHAPTER 5

---

### Indices and tables

---

- `genindex`
- `modindex`
- `search`



**p**

`pystan`, 16



## E

extract() (pystan.StanFit4model method), 28

## G

get\_adaptation\_info() (pystan.StanFit4model method), 29

get\_inits() (pystan.StanFit4model method), 30

get\_logposterior() (pystan.StanFit4model method), 29

get\_posterior\_mean() (pystan.StanFit4model method), 29

get\_sampler\_params() (pystan.StanFit4model method),  
29

get\_seed() (pystan.StanFit4model method), 30

get\_stancode() (pystan.StanFit4model method), 30

grad\_log\_prob() (pystan.StanFit4model method), 29

## L

log\_prob() (pystan.StanFit4model method), 29

## O

optimizing() (pystan.StanModel method), 23

## P

plot() (pystan.StanFit4model method), 28

pystan (module), 16

## R

read\_rdump() (in module pystan.misc), 30

## S

sampling() (pystan.StanModel method), 24

stan() (in module pystan), 16

stan\_rdump() (in module pystan.misc), 30

stanc() (in module pystan), 20

StanFit4model (class in pystan), 28

StanModel (class in pystan), 21

## U

unconstrain\_pars() (pystan.StanFit4model method), 30

## V

vb() (pystan.StanModel method), 26