Probability

Release 10.0

The Sage Development Team

CONTENTS

1	Probability Distributions	1
2	Random variables and probability spaces	13
3	Indices and Tables	17
Python Module Index		
In	dex	2.1

PROBABILITY DISTRIBUTIONS

This module provides three types of probability distributions:

- RealDistribution: various real-valued probability distributions.
- SphericalDistribution: uniformly distributed points on the surface of an n-1 sphere in n dimensional euclidean space.
- GeneralDiscreteDistribution: user-defined discrete distributions.

AUTHORS:

- Josh Kantor (2007-02): first version
- William Stein (2007-02): rewrite of docs, conventions, etc.
- Carlo Hamalainen (2008-08): full doctest coverage, more documentation, GeneralDiscreteDistribution, misc fixes.
- Kwankyu Lee (2010-05-29): F-distribution support.

REFERENCES:

GNU gsl library, General discrete distributions http://www.gnu.org/software/gsl/manual/html_node/General-Discrete-Distributions.html

GNU gsl library, Random number distributions http://www.gnu.org/software/gsl/manual/html_node/Random-Number-Distributions.html

class sage.probability.probability_distribution.GeneralDiscreteDistribution

Bases: ProbabilityDistribution

Create a discrete probability distribution.

INPUT:

- P list of probabilities. The list will automatically be normalised if sum(P) is not equal to 1.
- rng (optional) random number generator to use. May be one of 'default', 'luxury', or 'taus'.
- seed (optional) seed to use with the random number generator.

OUTPUT:

• a probability distribution where the probability of selecting x is P[x].

EXAMPLES:

Constructs a GeneralDiscreteDistribution with the probability distribution P where P(0) = 0.3, P(1) = 0.4, P(2) = 0.3:

```
sage: P = [0.3, 0.4, 0.3]
sage: X = GeneralDiscreteDistribution(P)
sage: X.get_random_element() in (0, 1, 2)
True
```

Checking the distribution of samples:

The distribution probabilities will automatically be normalised:

get_random_element()

Get a random sample from the probability distribution.

EXAMPLES:

```
sage: P = [0.3, 0.4, 0.3]
sage: X = GeneralDiscreteDistribution(P)
sage: all(X.get_random_element() in (0,1,2) for _ in range(10))
True
sage: isinstance(X.get_random_element(), sage.rings.integer.Integer)
True
```

reset_distribution()

This method resets the distribution.

EXAMPLES:

```
sage: T = GeneralDiscreteDistribution([0.1, 0.3, 0.6])
sage: T.set_seed(0)
sage: [T.get_random_element() for _ in range(10)]
[2, 2, 2, 2, 2, 1, 2, 2, 1, 2]
sage: T.reset_distribution()
sage: [T.get_random_element() for _ in range(10)]
[2, 2, 2, 2, 2, 1, 2, 2, 1, 2]
```

set_random_number_generator(rng='default')

Set the random number generator to be used by gsl.

EXAMPLES:

```
sage: X = GeneralDiscreteDistribution([0.3, 0.4, 0.3])
sage: X.set_random_number_generator('taus')
```

set_seed(seed)

Set the seed to be used by the random number generator.

EXAMPLES:

```
sage: X = GeneralDiscreteDistribution([0.3, 0.4, 0.3])
sage: X.set_seed(1)
sage: X.get_random_element()
1
```

class sage.probability.probability_distribution.ProbabilityDistribution

Bases: object

Concrete probability distributions should be derived from this abstract class.

```
generate_histogram_data(num_samples=1000, bins=50)
```

Compute a histogram of the probability distribution.

INPUT:

- num_samples (optional) number of times to sample from the probability distribution
- bins (optional) number of bins to divide the samples into.

OUTPUT:

• a tuple. The first element of the tuple is a list of length bins, consisting of the normalised histogram of the random samples. The second list is the bins.

EXAMPLES:

```
sage: set_random_seed(0)
sage: from sage.probability.probability_distribution import_
→ GeneralDiscreteDistribution
sage: P = [0.3, 0.4, 0.3]
sage: X = GeneralDiscreteDistribution(P)
sage: h, b = X.generate_histogram_data(bins = 10)
sage: h # rel tol 1e-08
[1.6299999999999999.
0.0,
0.0,
 0.0,
0.0,
 1.904999999999985,
0.0,
0.0.
0.0,
1.4650000000000003]
sage: b
[0.0,
0.2,
0.4,
0.60000000000000001,
 0.8,
```

generate_histogram_plot(name, num_samples=1000, bins=50)

Save the histogram from generate_histogram_data() to a file.

INPUT:

- name file to save the histogram plot (as a PNG).
- num_samples (optional) number of times to sample from the probability distribution
- bins (optional) number of bins to divide the samples into.

EXAMPLES:

This saves the histogram plot to a temporary file:

```
sage: from sage.probability.probability_distribution import_
GeneralDiscreteDistribution
sage: import tempfile
sage: P = [0.3, 0.4, 0.3]
sage: X = GeneralDiscreteDistribution(P)
sage: with tempfile.NamedTemporaryFile() as f:
....: X.generate_histogram_plot(f.name)
```

get_random_element()

To be implemented by a derived class:

```
sage: P = sage.probability.probability_distribution.ProbabilityDistribution()
sage: P.get_random_element()
Traceback (most recent call last):
...
NotImplementedError: implement in derived class
```

class sage.probability.probability_distribution.RealDistribution

Bases: ProbabilityDistribution

The RealDistribution class provides a number of routines for sampling from and analyzing and visualizing probability distributions. For precise definitions of the distributions and their parameters see the gsl reference manuals chapter on random number generators and probability distributions.

EXAMPLES:

Uniform distribution on the interval [a, b]:

```
sage: a = 0
sage: b = 2
sage: T = RealDistribution('uniform', [a, b])
sage: a <= T.get_random_element() <= b
True
sage: T.distribution_function(0)</pre>
```

```
0.5
sage: T.cum_distribution_function(1)
0.5
sage: T.cum_distribution_function_inv(.5)
1.0
```

The gaussian distribution takes 1 parameter sigma. The standard gaussian distribution has sigma = 1:

```
sage: sigma = 1
sage: T = RealDistribution('gaussian', sigma)
sage: s = T.get_random_element()
sage: s.parent()
Real Double Field
sage: T.distribution_function(0)
0.3989422804014327
sage: T.cum_distribution_function(1)
0.8413447460685429
sage: T.cum_distribution_function_inv(.5)
0.0
```

The rayleigh distribution has 1 parameter sigma:

```
sage: sigma = 3
sage: T = RealDistribution('rayleigh', sigma)
sage: s = T.get_random_element()
sage: s >= 0
True
sage: s.parent()
Real Double Field
sage: T.distribution_function(0)
0.0
sage: T.cum_distribution_function(1)
0.054040531093234534
sage: T.cum_distribution_function_inv(.5)
3.532230067546424...
```

The lognormal distribution has two parameters sigma and zeta:

```
sage: zeta = 0
sage: sigma = 1
sage: T = RealDistribution('lognormal', [zeta, sigma])
sage: s = T.get_random_element()
sage: s >= 0
True
sage: s.parent()
Real Double Field
sage: T.distribution_function(0)
0.0
sage: T.cum_distribution_function(1)
0.5
sage: T.cum_distribution_function_inv(.5)
1.0
```

The pareto distribution has two parameters a, and b:

```
sage: a = 1
sage: b = 1
sage: T = RealDistribution('pareto', [a, b])
sage: s = T.get_random_element()
sage: s >= b
True
sage: s.parent()
Real Double Field
sage: T.distribution_function(0)
0.0
sage: T.cum_distribution_function(1)
0.0
sage: T.cum_distribution_function_inv(.5)
2.0
```

The t-distribution has one parameter nu:

```
sage: nu = 1
sage: T = RealDistribution('t', nu)
sage: s = T.get_random_element()
sage: s.parent()
Real Double Field
sage: T.distribution_function(0)  # rel tol 1e-15
0.3183098861837906
sage: T.cum_distribution_function(1)  # rel tol 1e-15
0.75
sage: T.cum_distribution_function_inv(.5)
0.0
```

The F-distribution has two parameters nu1 and nu2:

```
sage: nu1 = 9; nu2 = 17
sage: F = RealDistribution('F', [nu1,nu2])
sage: s = F.get_random_element()
sage: s >= 0
True
sage: s.parent()
Real Double Field
sage: F.distribution_function(1) # rel tol 1e-14
0.6695025505192798
sage: F.cum_distribution_function(3.68) # rel tol 1e-14
0.9899717772300652
sage: F.cum_distribution_function_inv(0.99) # rel tol 1e-14
3.682241524045864
```

The chi-squared distribution has one parameter nu:

```
sage: nu = 1
sage: T = RealDistribution('chisquared', nu)
sage: s = T.get_random_element()
sage: s >= 0
True
```

```
sage: s.parent()
Real Double Field
sage: T.distribution_function(0)
+infinity
sage: T.cum_distribution_function(1) # rel tol 1e-14
0.6826894921370856
sage: T.cum_distribution_function_inv(.5) # rel tol 1e-14
0.45493642311957305
```

The exponential power distribution has two parameters a and b:

```
sage: a = 1
sage: b = 2.5
sage: T = RealDistribution('exppow', [a, b])
sage: s = T.get_random_element()
sage: s.parent()
Real Double Field
sage: T.distribution_function(0) # rel tol 1e-14
0.5635302489930136
sage: T.cum_distribution_function(1) # rel tol 1e-14
0.940263052542855
```

The beta distribution has two parameters a and b:

```
sage: a = 2
sage: b = 2
sage: T = RealDistribution('beta', [a, b])
sage: s = T.get_random_element()
sage: 0 <= s <= 1
True
sage: s.parent()
Real Double Field
sage: T.distribution_function(0)
0.0
sage: T.cum_distribution_function(1)
1.0</pre>
```

The weibull distribution has two parameters a and b:

```
sage: a = 1
sage: b = 1
sage: T = RealDistribution('weibull', [a, b])
sage: s = T.get_random_element()
sage: s >= 0
True
sage: s.parent()
Real Double Field
sage: T.distribution_function(0)
1.0
sage: T.cum_distribution_function(1)
0.6321205588285577
sage: T.cum_distribution_function_inv(.5)
0.6931471805599453
```

It is possible to select which random number generator drives the sampling as well as the seed. The default is the Mersenne twister. Also available are the RANDLXS algorithm and the Tausworthe generator (see the gsl reference manual for more details). These are all supposed to be simulation quality generators. For RANDLXS use rng = 'luxury' and for tausworth use rng = 'taus':

```
sage: T = RealDistribution('gaussian', 1, rng = 'luxury', seed = 10)
```

To change the seed at a later time use set_seed:

```
sage: T.set_seed(100)
```

cum_distribution_function(x)

Evaluate the cumulative distribution function of the probability distribution at x.

EXAMPLES:

```
sage: T = RealDistribution('uniform', [0, 2])
sage: T.cum_distribution_function(1)
0.5
```

cum_distribution_function_inv(x)

Evaluate the inverse of the cumulative distribution distribution function of the probability distribution at \mathbf{x} .

EXAMPLES:

```
sage: T = RealDistribution('uniform', [0, 2])
sage: T.cum_distribution_function_inv(.5)
1.0
```

distribution_function(x)

Evaluate the distribution function of the probability distribution at x.

EXAMPLES:

```
sage: T = RealDistribution('uniform', [0, 2])
sage: T.distribution_function(0)
0.5
sage: T.distribution_function(1)
0.5
sage: T.distribution_function(1.5)
0.5
sage: T.distribution_function(2)
0.0
```

get_random_element()

Get a random sample from the probability distribution.

EXAMPLES:

```
sage: T = RealDistribution('gaussian', 1, seed = 0)
sage: T.get_random_element() # rel tol 4e-16
0.13391860811867587
```

```
plot(*args, **kwds)
```

Plot the distribution function for the probability distribution. Parameters to sage.plot.plot.plot() can be passed through *args and **kwds.

EXAMPLES:

```
sage: T = RealDistribution('uniform', [0, 2])
sage: P = T.plot()
```

reset_distribution()

This method resets the distribution.

EXAMPLES:

set_distribution(name='uniform', parameters=[])

This method can be called to change the current probability distribution.

EXAMPLES:

```
sage: T = RealDistribution('gaussian', 1)
sage: T.set_distribution('gaussian', 1)
sage: T.set_distribution('pareto', [0, 1])
```

set_random_number_generator(rng='default')

Set the gsl random number generator to be one of default, luxury, or taus.

EXAMPLES:

```
sage: T = SphericalDistribution()
sage: T.set_random_number_generator('default')
sage: T.set_seed(0)
sage: T.get_random_element() # rel tol 4e-16
(0.07961564104639995, -0.05237671627581255, 0.9954486572862178)
sage: T.set_random_number_generator('luxury')
sage: T.set_seed(0)
sage: T.get_random_element() # rel tol 4e-16
(0.07961564104639995, -0.05237671627581255, 0.9954486572862178)
```

set_seed(seed)

Set the seed for the underlying random number generator.

EXAMPLES:

```
sage: T = RealDistribution('gaussian', 1, rng = 'luxury', seed = 10)
sage: T.set_seed(100)
```

class sage.probability.probability_distribution.SphericalDistribution

Bases: ProbabilityDistribution

This class is capable of producing random points uniformly distributed on the surface of an n-1 sphere in n dimensional euclidean space. The dimension, n is selected via the keyword dimension. The random number generator which drives it can be selected using the keyword rng. Valid choices are default which uses the Mersenne-Twister, luxury which uses RANDLXS, and taus which uses the tausworth generator. The default dimension is 3.

EXAMPLES:

```
sage: T = SphericalDistribution()
sage: s = T.get_random_element()
sage: s.norm() # rel tol 1e-14
1.0
sage: len(s)
3
sage: T = SphericalDistribution(dimension=4, rng='luxury')
sage: s = T.get_random_element()
sage: s.norm() # rel tol 1e-14
1.0
sage: len(s)
4
```

get_random_element()

Get a random sample from the probability distribution.

EXAMPLES:

```
sage: T = SphericalDistribution(seed = 0)
sage: T.get_random_element() # rel tol 4e-16
(0.07961564104639995, -0.05237671627581255, 0.9954486572862178)
```

reset_distribution()

This method resets the distribution.

EXAMPLES:

set_random_number_generator(rng='default')

Set the gsl random number generator to be one of default, luxury, or taus.

EXAMPLES:

```
sage: T = SphericalDistribution()
sage: T.set_random_number_generator('default')
```

```
sage: T.set_seed(0)
sage: T.get_random_element() # rel tol 4e-16
(0.07961564104639995, -0.05237671627581255, 0.9954486572862178)
sage: T.set_random_number_generator('luxury')
sage: T.set_seed(0)
sage: T.get_random_element() # rel tol 4e-16
(0.07961564104639995, -0.05237671627581255, 0.9954486572862178)
```

set_seed(seed)

Set the seed for the underlying random number generator.

EXAMPLES:

```
sage: T = SphericalDistribution(seed = 0)
sage: T.set_seed(100)
```

RANDOM VARIABLES AND PROBABILITY SPACES

This introduces a class of random variables, with the focus on discrete random variables (i.e. on a discrete probability space). This avoids the problem of defining a measure space and measurable functions.

Bases: ProbabilitySpace_generic, DiscreteRandomVariable

The discrete probability space

entropy()

The entropy of the probability space.

set()

The set of values of the probability space taking possibly nonzero probability (a subset of the domain).

Bases: RandomVariable_generic

A random variable on a discrete probability space.

correlation(other)

The correlation of the probability space X = self with Y = other.

covariance(other)

The covariance of the discrete random variable X = self with Y = other.

Let S be the probability space of X = self, with probability function p, and E(X) be the expectation of X. Then the variance of X is:

$$\mathrm{cov}(X,Y) = E((X-E(X)\cdot(Y-E(Y)) = \sum_{x\in S} p(x)(X(x)-E(X))(Y(x)-E(Y))$$

expectation()

The expectation of the discrete random variable, namely $\sum_{x \in S} p(x)X[x]$, where X = self and S is the probability space of X.

function()

The function defining the random variable.

standard_deviation()

The standard deviation of the discrete random variable.

Let S be the probability space of X = self, with probability function p, and E(X) be the expectation of X. Then the standard deviation of X is defined to be

$$\sigma(X) = \sqrt{\sum_{x \in S} p(x)(X(x) - E(x))^2}$$

translation_correlation(other, map)

The correlation of the probability space X = self with image of Y = other under map.

translation_covariance(other, map)

The covariance of the probability space X = self with image of Y = other under the given map of the probability space.

Let S be the probability space of X = self, with probability function p, and E(X) be the expectation of X. Then the variance of X is:

$$\mathrm{cov}(X,Y) = E((X-E(X)\cdot(Y-E(Y)) = \sum_{x\in S} p(x)(X(x)-E(X))(Y(x)-E(Y))$$

translation_expectation(map)

The expectation of the discrete random variable, namely $\sum_{x \in S} p(x) X[e(x)]$, where X = self, S is the probability space of X, and e = map.

translation_standard_deviation(map)

The standard deviation of the translated discrete random variable $X \circ e$, where X = self and e = map.

Let S be the probability space of X = self, with probability function p, and E(X) be the expectation of X. Then the standard deviation of X is defined to be

$$\sigma(X) = \sqrt{\sum_{x \in S} p(x)(X(x) - E(x))^2}$$

translation_variance(map)

The variance of the discrete random variable $X \circ e$, where X = self, and e = map.

Let S be the probability space of X = self, with probability function p, and E(X) be the expectation of X. Then the variance of X is:

$$var(X) = E((X - E(x))^{2}) = \sum_{x \in S} p(x)(X(x) - E(x))^{2}$$

variance()

The variance of the discrete random variable.

Let S be the probability space of X = self, with probability function p, and E(X) be the expectation of X. Then the variance of X is:

$$var(X) = E((X - E(x))^{2}) = \sum_{x \in S} p(x)(X(x) - E(x))^{2}$$

class sage.probability.random_variable.ProbabilitySpace_generic(domain, RR)

Bases: RandomVariable_generic

A probability space.

```
domain()
class sage.probability.random_variable.RandomVariable_generic(X, RR)
    Bases: Parent
    A random variable.
    codomain()
    domain()
    field()
    probability_space()
sage.probability.random_variable.is_DiscreteProbabilitySpace(S)
sage.probability.random_variable.is_DiscreteRandomVariable(X)
sage.probability.random_variable.is_ProbabilitySpace(S)
sage.probability.random_variable.is_RandomVariable(X)
```

CHAPTER

THREE

INDICES AND TABLES

- Index
- Module Index
- Search Page

PYTHON MODULE INDEX

p

sage.probability.probability_distribution, 1
sage.probability.random_variable, 13

20 Python Module Index

INDEX

U	sage.probability.probability_distribution),
<pre>codomain() (sage.probability.random_variable.RandomV</pre>	generate_histogram_data()
correlation() (sage.probability.random_variable.Discre	eteRandom (sqsqangobability.probability_distribution.ProbabilityDistribution method), 3
covariance() (sage.probability.random_variable.Discret method), 13	(sage.probability.probability_alstribution.FrobabilityDistribution
cum_distribution_function()	method), 4
(sage.probability.probability_distribution.RealDimethod), 8	(sage.probability.probability_aistribution.GeneralDiscreteDistrib
cum_distribution_function_inv()	method), 2
(sage.probability.probability_distribution.RealDimethod), 8	is get_{urion}dom_element() (sage.probability.probability_distribution.ProbabilityDistribution method), 4
D	get_random_element()
DiscreteProbabilitySpace (class in	(sage.probability.probability_distribution.RealDistribution method), 8
sage.probability.random_variable), 13 DiscreteRandomVariable (class in sage.probability.random_variable), 13	<pre>get_random_element()</pre>
distribution_function() (sage.probability.probability_distribution.RealDimethod) 8	
domain() (sage.probability.random_variable.ProbabilityS	sage.probability.ranaom_variable), 13
domain() (sage.probability.random_variable.RandomVari	iable_DiscreteRandomVariable() (in module
method), 15	sage.probability.random_variable), 13
E	is_ProbabilitySpace() (in module sage.probability.random_variable), 15
<pre>entropy() (sage.probability.random_variable.DiscretePro</pre>	is_RandomVariable() (in module obabilitySpace sage.probability.random_variable), 15
<pre>expectation() (sage.probability.random_variable.Discre method), 13</pre>	
_	module
F	sage.probability.probability_distribution,
<pre>field() (sage.probability.random_variable.RandomVaria</pre>	uble_generic sage.probability.random_variable,13
function() (sage.probability.random_variable.DiscreteR	a m lomVariable
method), 13	plot() (sage.probability.probability_distribution.RealDistribution
G	method), 8 probability_space()
GeneralDiscreteDistribution (class in	(sage probability random variable Random Variable generic

```
T
        method), 15
ProbabilityDistribution
                                    (class
                                                  in
                                                      translation_correlation()
        sage.probability.probability distribution),
                                                                (sage.probability.random variable.DiscreteRandomVariable
                                                                method), 14
ProbabilitySpace_generic
                                     (class
                                                  in
                                                      translation_covariance()
        sage.probability.random variable), 14
                                                                (sage.probability.random_variable.DiscreteRandomVariable
                                                                method), 14
R
                                                      translation_expectation()
RandomVariable_generic
                                   (class
                                                  in
                                                                (sage.probability.random_variable.DiscreteRandomVariable
         sage.probability.random variable), 15
                                                                method), 14
RealDistribution
                                (class
                                                  in
                                                      translation_standard_deviation()
        sage.probability.probability_distribution),
                                                                (sage.probability.random_variable.DiscreteRandomVariable
                                                                method), 14
reset_distribution()
                                                      translation_variance()
         (sage.probability.probability_distribution.GeneralDiscreteD(sangbyptiobability.random_variable.DiscreteRandomVariable
        method), 2
                                                               method), 14
reset_distribution()
         (sage.probability.probability_distribution.RealDistribution
         method), 9
                                                      variance() (sage.probability.random variable.DiscreteRandomVariable
reset_distribution()
                                                                method), 14
         (sage.probability_probability_distribution.SphericalDistribution
         method), 10
S
sage.probability.probability_distribution
    module, 1
sage.probability.random_variable
    module, 13
set() (sage.probability.random_variable.DiscreteProbabilitySpace
         method), 13
set_distribution() (sage.probability.probability distribution.RealDistribution
        method), 9
set_random_number_generator()
         (sage.probability_probability_distribution.GeneralDiscreteDistribution
         method), 2
set_random_number_generator()
         (sage.probability_probability_distribution.RealDistribution
        method), 9
set_random_number_generator()
         (sage.probability\_distribution. Spherical Distribution
\mathtt{set\_seed}() (sage.probability.probability_distribution.GeneralDiscreteDistribution
         method), 3
set_seed() (sage.probability.probability_distribution.RealDistribution
set\_seed() (sage.probability_distribution.SphericalDistribution
        method), 11
SphericalDistribution
                                   (class
                                                  in
        sage.probability.probability distribution),
standard_deviation()
         (sage.probability.random\_variable.DiscreteRandomVariable
        method), 13
```

22 Index