Sage Reference Manual: Probability

Release 7.1

The Sage Development Team

March 20, 2016

CHAPTER ONE

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.gsl.probability_distribution.`**`GeneralDiscreteDistribution`**

Bases: `sage.gsl.probability_distribution.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()
1
```

Checking the distribution of samples:

```
sage: P = [0.3, 0.4, 0.3]
sage: counts = [0] * len(P)
sage: X = GeneralDiscreteDistribution(P)
sage: nr_samples = 10000
sage: for _ in range(nr_samples):
...       counts[X.get_random_element()] += 1
sage: [1.0*x/nr_samples for x in counts]
[0.304200000000000, 0.397300000000000, 0.298500000000000]
```

The distribution probabilities will automatically be normalised:

```
sage: P = [0.1, 0.3]
sage: X = GeneralDiscreteDistribution(P, seed = 0)
sage: counts = [0, 0]
sage: for _ in range(10000):
...       counts[X.get_random_element()] += 1
sage: float(counts[1]/counts[0])
3.042037186742118
```

TESTS:

Make sure that repeated initializations are randomly seeded (trac ticket #9770):

```
sage: P = [0.001] * 1000
sage: Xs = [GeneralDiscreteDistribution(P).get_random_element() for _ in range(1000)]
sage: len(set(Xs)) > 2^^32
True
```

The distribution probabilities must be non-negative:

```
sage: GeneralDiscreteDistribution([0.1, -0.1])
Traceback (most recent call last):
...
ValueError: The distribution probabilities must be non-negative
```

get_random_element()

Get a random sample from the probability distribution.

EXAMPLE:

```
sage: P = [0.3, 0.4, 0.3]
sage: X = GeneralDiscreteDistribution(P)
sage: [X.get_random_element() for _ in range(10)]
[1, 0, 1, 1, 0, 1, 1, 1, 1, 1]
```

```
sage: isinstance(X.get_random_element(), sage.rings.integer.Integer)
True
```

reset_distribution()

This method resets the distribution.

EXAMPLE:

```
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.

EXAMPLE:

```
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.

EXAMPLE:

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

class `sage.gsl.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.

EXAMPLE:

```
sage: from sage.gsl.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
[1.6299999999999999,
 0.0,
 0.0,
```

```
 0.0,
 0.0,
 1.9049999999999985,
 0.0,
 0.0,
 0.0,
 1.4650000000000003]
sage: b
[0.0, 0.20000000000000001, 0.40000000000000002, 0.60000000000000009, 0.80000000000000004, 1.
```

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.

EXAMPLE:

This saves the histogram plot to `my_general_distribution_plot.png` in the temporary directory `SAGE_TMP`:

```
sage: from sage.gsl.probability_distribution import GeneralDiscreteDistribution
sage: import os
sage: P = [0.3, 0.4, 0.3]
sage: X = GeneralDiscreteDistribution(P)
sage: file = os.path.join(SAGE_TMP, "my_general_distribution_plot")
sage: X.generate_histogram_plot(file)
```

get_random_element ()

To be implemented by a derived class:

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

class `sage.gsl.probability_distribution.`**RealDistribution**

Bases: `sage.gsl.probability_distribution.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: T.get_random_element()
0.8175557665526867
sage: T.distribution_function(0)
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: T.get_random_element()
-0.5860943109756299
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: T.get_random_element()
5.748307572643492
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: T.get_random_element()  # abs tol 1e-16
0.3876433713532701
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: T.get_random_element()
10.418714048916407
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: T.get_random_element()  # rel tol 1e-15
-8.404911172800615
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: F.get_random_element()  # rel tol 1e-14
1.239233786115256
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: T.get_random_element()
0.4603367753992381
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: T.get_random_element()
0.16442075306686463
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: T.get_random_element()  # rel tol 1e-14
0.7110581877139808
sage: T.distribution_function(0)
0.0
sage: T.cum_distribution_function(1)
1.0
```

The weibull distribution has two parameters a and b:

```
sage: a = 1
sage: b = 1
sage: T = RealDistribution('weibull', [a, b])
sage: T.get_random_element()
1.1867854542468694
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)
```

TESTS:

Make sure that repeated initializations are randomly seeded (trac ticket #9770):

```
sage: Xs = [RealDistribution('gaussian', 1).get_random_element() for _ in range(1000)]
sage: len(set(Xs)) > 2^^32
True
```

cum_distribution_function (*x*)

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

EXAMPLE:

```
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 x.

EXAMPLE:

```
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.

EXAMPLE:

```
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.plot` can be passed through `*args` and `**kwds`.

EXAMPLE:

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

reset_distribution()

This method resets the distribution.

EXAMPLE:

```
sage: T = RealDistribution('gaussian', 1, seed = 10)
sage: [T.get_random_element() for _ in range(10)]  # rel tol 4e-16
[-0.7460999595745819, -0.004644606626413462, -0.8720538317207641, 0.6916259921666037, 2.6766
sage: T.reset_distribution()
sage: [T.get_random_element() for _ in range(10)]  # rel tol 4e-16
[-0.7460999595745819, -0.004644606626413462, -0.8720538317207641, 0.6916259921666037, 2.6766
```

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`.

EXAMPLE:

```
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.

EXAMPLE:

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

class sage.gsl.probability_distribution.**SphericalDistribution**

Bases: sage.gsl.probability_distribution.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: T.get_random_element()  # rel tol 1e-14
(-0.2922296724828204, -0.9563459345927822, 0.0020668595602153454)
sage: T = SphericalDistribution(dimension = 4, rng = 'luxury')
sage: T.get_random_element()  # rel tol 1e-14
(-0.0363300434761631, 0.6459885817544098, 0.24825817345598158, 0.7209346430129753)
```

TESTS:

Make sure that repeated initializations are randomly seeded (trac ticket #9770):

```
sage: Xs = [tuple(SphericalDistribution(2).get_random_element()) for _ in range(1000)]
sage: len(set(Xs)) > 2^^32
True
```

get_random_element()

Get a random sample from the probability distribution.

EXAMPLE:

```
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.

EXAMPLE:

```
sage: T = SphericalDistribution(seed = 0)
sage: [T.get_random_element() for _ in range(4)]  # rel tol 4e-16
[(0.07961564104639995, -0.05237671627581255, 0.9954486572862178), (0.4123599490593727, 0.560
sage: T.reset_distribution()
sage: [T.get_random_element() for _ in range(4)]  # rel tol 4e-16
[(0.07961564104639995, -0.05237671627581255, 0.9954486572862178), (0.4123599490593727, 0.560
```

set_random_number_generator(*rng='default'*)

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

EXAMPLE:

```
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.

EXAMPLE:

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

CHAPTER

TWO

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.

class sage.probability.random_variable.**DiscreteProbabilitySpace**(*X*, *P*, *codomain=None*, *check=False*)

Bases: sage.probability.random_variable.ProbabilitySpace_generic, sage.probability.random_variable.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).

class sage.probability.random_variable.**DiscreteRandomVariable**(*X*, *f*, *codomain=None*, *check=False*)

Bases: sage.probability.random_variable.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:

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

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:

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

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:

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

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:

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

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

Bases: sage.probability.random_variable.RandomVariable_generic

A probability space.

domain()

class `sage.probability.random_variable.`**`RandomVariable_generic`**`(`*X*, *RR*`)`

Bases: `sage.structure.parent_base.ParentWithBase`

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

C

D

E

F

G

I

P

R

S

T

V