Statistics

*Release 10.3*

The Sage Development Team

Mar 20, 2024
This file contains basic descriptive functions. Included are the mean, median, mode, moving average, standard deviation, and the variance. When calling a function on data, there are checks for functions already defined for that data type.

The `mean()` function returns the arithmetic mean (the sum of all the members of a list, divided by the number of members). Further revisions may include the geometric and harmonic mean. The `median()` function returns the number separating the higher half of a sample from the lower half. The `mode()` returns the most common occurring member of a sample, plus the number of times it occurs. If entries occur equally common, the smallest of a list of the most common entries is returned. The `moving_average()` is a finite impulse response filter, creating a series of averages using a user-defined number of subsets of the full data set. The `std()` and the `variance()` return a measurement of how far data points tend to be from the arithmetic mean.

Functions are available in the namespace `stats`, i.e. you can use them by typing `stats.mean`, `stats.median`, etc.

REMARK: If all the data you are working with are floating point numbers, you may find `stats.TimeSeries` helpful, since it is extremely fast and offers many of the same descriptive statistics as in the module.

AUTHOR:

- Andrew Hou (11/06/2009)

`sage.stats.basic_stats.mean(v)`

Return the mean of the elements of `v`.

We define the mean of the empty list to be the (symbolic) NaN, following the convention of MATLAB, Scipy, and R.

This function is deprecated. Use `numpy.mean()` or `numpy.nanmean()` instead.

INPUT:

- `v` – a list of numbers

OUTPUT:

- a number

EXAMPLES:

```python
sage: mean([pi, e])  # ...
→needs sage.symbolic
doctest:warning...
DeprecationWarning: sage.stats.basic_stats.mean is deprecated; use numpy.mean or numpy.nanmean instead
See https://github.com/sagemath/sage/issues/29662 for details.
1/2*pi + 1/2*e
sage: mean([])      # ...
```
Return the median (middle value) of the elements of $v$.

If $v$ is empty, we define the median to be NaN, which is consistent with NumPy (note that R returns NULL). If $v$ is comprised of strings, TypeError occurs. For elements other than numbers, the median is a result of \texttt{sorted()}.

This function is deprecated. Use \texttt{numpy.median()} or \texttt{numpy.nanmedian()} instead.

**INPUT:**
- $v$ - a list

**OUTPUT:**
- median element of $v$

**EXAMPLES:**

```python
sage: median([1,2,3,4,5])
```

.. deprecated::

   Use \texttt{numpy.median()} or \texttt{numpy.nanmedian()} instead.

   See https://github.com/sagemath/sage/issues/29662 for details.

   3

```python
sage: median([e, pi])
```

.. needs sagemath

1/2*pi + 1/2*e

```python
sage: median(['sage', 'linux', 'python'])
```

'python'

```python
sage: median([])
```

.. needs sagemath

NaN

```python
class MyClass:
    ....:    def median(self):
    ....:        return 1
sage: stats.median(MyClass())
```

1

Return the mode of $v$.

```python
sage: stats.basic_stats.mode(v)
```

.. needs sagemath

1
The mode is the list of the most frequently occurring elements in \( v \). If \( n \) is the most times that any element occurs in \( v \), then the mode is the list of elements of \( v \) that occur \( n \) times. The list is sorted if possible.

This function is deprecated. Use `scipy.stats.mode()` or `statistics.mode()` instead.

**Note:** The elements of \( v \) must be hashable.

**INPUT:**
- \( v \) – a list

**OUTPUT:**
- a list (sorted if possible)

**EXAMPLES:**

```python
sage: v = [1, 2, 4, 1, 6, 2, 6, 7, 1]
sage: mode(v)
DeprecationWarning: sage.stats.basic_stats.mode is deprecated; use scipy.stats.mode or statistics.mode instead
See https://github.com/sagemath/sage/issues/29662 for details.
[1]
sage: v.count(1)
3
sage: mode([])
[]
```

```python
sage: mode([1, 2, 3, 4, 5])
[1, 2, 3, 4, 5]
sage: mode([3, 1, 2, 1, 2, 3])
[1, 2, 3]
sage: mode([0, 2, 7, 7, 13, 20, 2, 13])
[2, 7, 13]
sage: mode(['sage', 'four', 'I', 'three', 'sage', 'pi'])
['sage']
```

```python
sage: class MyClass:
.....    def mode(self):
.....        return [1]
sage: stats.mode(MyClass())
[1]
```

**sage.stats.basic_stats.moving_average** \((v, n)\)

Return the moving average of a list \( v \).

The moving average of a list is often used to smooth out noisy data.

If \( v \) is empty, we define the entries of the moving average to be NaN.

This method is deprecated. Use `pandas.Series.rolling()` instead.

**INPUT:**
- \( v \) – a list
- \( n \) – the number of values used in computing each average.

**OUTPUT:**
• a list of length \( \text{len}(v) - n + 1 \), since we do not fabric any values

EXMODES:

```sage
sage: moving_average([1..10], 1)
doctest:warning...
DeprecationWarning: sage.stats.basic_stats.moving_average is deprecated; use pandas.Series.rolling instead
See https://github.com/sagemath/sage/issues/29662 for details.
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
sage: moving_average([1..10], 4)
[5/2, 7/2, 9/2, 11/2, 13/2, 15/2, 17/2]
sage: moving_average([], 1)
[]
sage: moving_average([pi, e, I, sqrt(2), 3/5], 2)  #...

→ needs sage.symbolic
[1/2*pi + 1/2*e, 1/2*e + 1/2*I, 1/2*sqrt(2) + 1/2*I,
1/2*sqrt(2) + 3/10]
```

We check if the input is a time series, and if so use the optimized `simple_moving_average()` method, but with (slightly different) meaning as defined above (the point is that the `simple_moving_average()` on time series returns \( n \) values:

```sage
sage: a = stats.TimeSeries([1..10])  #...
→ needs numpy
sage: stats.moving_average(a, 3)  #...
→ needs numpy
[2.0000, 3.0000, 4.0000, 5.0000, 6.0000, 7.0000, 8.0000, 9.0000]
sage: stats.moving_average(list(a), 3)  #...
→ needs numpy
[2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0]
```

\[\text{sage.stats.basic_stats.std}(v, \text{bias}=\text{False})\]

Return the standard deviation of the elements of \( v \).

We define the standard deviation of the empty list to be NaN, following the convention of MATLAB, Scipy, and R.

This function is deprecated. Use `numpy.std()` or `numpy.nanstd()` instead.

INPUT:

• \( v \) – a list of numbers

• \text{bias} – bool (default: False); if False, divide by \( \text{len}(v) - 1 \) instead of \( \text{len}(v) \) to give a less biased estimator (sample) for the standard deviation.

OUTPUT:

• a number

EXMODES:

```sage
sage: # needs sage.symbolic
sage: std([1..6], bias=True)
doctest:warning...
DeprecationWarning: sage.stats.basic_stats.std is deprecated; use numpy.std or numpy.nanstd instead
See https://github.com/sagemath/sage/issues/29662 for details.
doctest:warning...
```
DeprecationWarning: sage.stats.basic_stats.variance is deprecated; use numpy.var or numpy.nanvar instead
See https://github.com/sagemath/sage/issues/29662 for details.
doctest:warning...
DeprecationWarning: sage.stats.basic_stats.mean is deprecated; use numpy.mean or numpy.nanmean instead
See https://github.com/sagemath/sage/issues/29662 for details.

1/2*sqrt(35/3)
sage: std([1..6], bias=False)
sqrt(7/2)
sage: std([e, pi])
sqrt(1/2)*abs(pi - e)
sage: std([])
NaN
sage: std([I, sqrt(2), 3/5])
1/15*sqrt(1/2)*sqrt((10*sqrt(2) - 5*I - 3)^2 + (5*sqrt(2) - 10*I + 3)^2 + (5*sqrt(2) + 5*I - 6)^2)
sage: std([RIF(1.0103, 1.0103), RIF(2)])
0.6998235813403261?
sage: # needs numpy
sage: import numpy
sage: x = numpy.array([1,2,3,4,5])
sage: std(x, bias=False)
1.5811388300841898
sage: x = stats.TimeSeries([1..100])
sage: std(x)
29.011491975882016

sage.stats.basic_stats.variance(v, bias=False)

Return the variance of the elements of v.

We define the variance of the empty list to be NaN, following the convention of MATLAB, Scipy, and R.

This function is deprecated. Use numpy.var() or numpy.nanvar() instead.

INPUT:

• v – a list of numbers

• bias – bool (default: False); if False, divide by len(v) – 1 instead of len(v) to give a less biased estimator (sample) for the standard deviation.

OUTPUT:

• a number

EXAMPLES:

sage: variance([1..6])
doctest:warning...
DeprecationWarning: sage.stats.basic_stats.variance is deprecated; use numpy.var or numpy.nanvar instead
See https://github.com/sagemath/sage/issues/29662 for details.
7/2
sage: variance([1..6], bias=True)
35/12
sage: variance([e, pi])
#...
←needs sage.symbolic
\[
\frac{1}{2}(\pi - e)^2
\]

**sage:** `variance([])`
NaN

**sage:** `variance([I, sqrt(2), 3/5])`  
\[
\frac{1}{450}(10\sqrt{2} - 5I - 3)^2 + \frac{1}{450}(5\sqrt{2} - 10I + 3)^2 \\
+ \frac{1}{450}(5\sqrt{2} + 5I - 6)^2
\]

**sage:** `variance([RIF(1.0103, 1.0103), RIF(2)])`
\approx 0.4897530450000000

**sage:** `import numpy`  
\[\text{needs numpy}\]

**sage:** `x = numpy.array([1, 2, 3, 4, 5])`  
\[\text{needs numpy}\]

**sage:** `variance(x, bias=False)`  
\[\text{needs numpy}\]

2.5

**sage:** `x = stats.TimeSeries([1..100])`

**sage:** `variance(x)`
841.6666666666666

**sage:** `variance(x, bias=True)`
833.25

**sage:** `class MyClass:`

....: def variance(self, bias=False):
....: return 1

**sage:** `stats.variance(MyClass())`
1

**sage:** `class SillyPythonList:`

....: def __init__(self):
....: self.__list = [2, 4]

....: def __len__(self):
....: return len(self.__list)

....: def __iter__(self):
....: return self.__list.__iter__()

....: def mean(self):
....: return 3

**sage:** `R = SillyPythonList()`

**sage:** `variance(R)`
2

**sage:** `variance(R, bias=True)`
1
This is a class for fast basic operations with lists of C ints. It is similar to the double precision TimeSeries class. It has all the standard C int semantics, of course, including overflow. It is also similar to the Python list class, except all elements are C ints, which makes some operations much, much faster. For example, concatenating two IntLists can be over 10 times faster than concatenating the corresponding Python lists of ints, and taking slices is also much faster.

AUTHOR:

- William Stein, 2010-03

```python
class sage.stats.intlist.IntList
    Bases: object

    A list of C int's.

    list()
        Return Python list version of self with Python ints as entries.

        EXAMPLES:
        sage: a = stats.IntList([1..15]); a
        [1, 2, 3, 4, 5 ... 11, 12, 13, 14, 15]
        sage: a.list()
        [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]
        sage: list(a) == a.list()
        True
        sage: type(a.list()[0])
        <... 'int'>

    max(index=False)
        Return the largest value in this time series. If this series has length 0 we raise a ValueError

        INPUT:
        - index – bool (default: False); if True, also return index of maximum entry.

        OUTPUT:
        - int – largest value
        - int – index of largest value; only returned if index=True

        EXAMPLES:
        sage: v = stats.IntList([1,-4,3,-2,-4,3])
        sage: v.max()
        3
```

(continues on next page)
\texttt{sage: v.max(index=\texttt{True})}
\texttt{(3, 2)}

\textbf{\texttt{min} (index=\texttt{False})}

Return the smallest value in this integer list. If this series has length 0 we raise a \texttt{ValueError}.

\textbf{INPUT:}

- \texttt{index} – \texttt{bool} (default: \texttt{False}); if \texttt{True}, also return index of minimal entry.

\textbf{OUTPUT:}

- \texttt{float} – smallest value
- \texttt{integer} – index of smallest value; only returned if \texttt{index=True}

\textbf{EXAMPLES:}

\begin{verbatim}
\texttt{sage: v = stats.IntList([1,-4,3,-2,-4])}
\texttt{sage: v.min()}
\texttt{-4}
\texttt{sage: v.min(index=\texttt{True})}
\texttt{(-4, 1)}
\end{verbatim}

\textbf{\texttt{plot} (*\texttt{args}, **\texttt{kwds})}

Return a plot of this \texttt{IntList}.

This just constructs the corresponding double-precision floating point \texttt{TimeSeries} object, passing on all arguments.

\textbf{EXAMPLES:}

\begin{verbatim}
\texttt{sage: stats.IntList([3,7,19,-2]).plot()}
\texttt{...: needs sage.plot}
\texttt{Graphics object consisting of 1 graphics primitive}
\texttt{sage: stats.IntList([3,7,19,-2]).plot(color='red',}
\texttt{......: needs sage.plot}
\texttt{......: points=\texttt{True})}
\texttt{Graphics object consisting of 1 graphics primitive}
\end{verbatim}

\textbf{\texttt{plot_histogram} (*\texttt{args}, **\texttt{kwds})}

Return a histogram plot of this \texttt{IntList}.

This just constructs the corresponding double-precision floating point \texttt{TimeSeries} object, and plots it, passing on all arguments.

\textbf{EXAMPLES:}

\begin{verbatim}
\texttt{sage: stats.IntList([1..15]).plot_histogram()}
\texttt{...: needs sage.plot}
\texttt{Graphics object consisting of 50 graphics primitives}
\end{verbatim}

\textbf{\texttt{prod}()}

Return the product of the entries of \texttt{self}.

\textbf{EXAMPLES:}
```python
sage: a = stats.IntList([1..10]); a
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```
```
sage: a.prod()
3628800
```
```
sage: factorial(10)
3628800
```

Note that there can be overflow:
```
```
sage: a = stats.IntList([2^30, 2]); a
[1073741824, 2]
```
```
sage: a.prod()
-2147483648
```

**sum()**

Return the sum of the entries of self.

**EXAMPLES:**
```
sage: stats.IntList([1..100]).sum()
5050
```
```
Note that there can be overflow, since the entries are C ints:
```
sage: a = stats.IntList([2^30,2^30]); a
[1073741824, 1073741824]
```
```
sage: a.sum()
-2147483648
```

**time_series()**

Return TimeSeries version of self, which involves changing each entry to a double.

**EXAMPLES:**
```
sage: T = stats.IntList([-2,3,5]).time_series(); T
[-2.0000, 3.0000, 5.0000]
```
```
sage: type(T)
<... sage.stats.time_series.TimeSeries>
```

`sage.stats.intlist.unpickle_intlist_v1(v, n)`

Version 1 unpickle method.

**INPUT:**

- v – a raw char buffer

**EXAMPLES:**
```
sage: v = stats.IntList([1,2,3])
sage: s = v.__reduce__()[1][0]
sage: type(s) == type(b'')
```
```
True
```
```
sage: sage.stats.intlist.unpickle_intlist_v1(s, 3)
[1, 2, 3]
```
```
sage: sage.stats.intlist.unpickle_intlist_v1(s+s,6)
[1, 2, 3, 1, 2, 3]
```
```
sage: sage.stats.intlist.unpickle_intlist_v1(b'',0)
[]
```
```
CHAPTER THREE

HIDDEN MARKOV MODELS

This is a complete pure-Cython optimized implementation of Hidden Markov Models. It fully supports Discrete, Gaussian, and Mixed Gaussian emissions.

The best references for the basic HMM algorithms implemented here are:

- Tapas Kanungo’s “Hidden Markov Models”
- Jackson’s HMM tutorial:
  http://personal.ee.surrey.ac.uk/Personal/P.Jackson/tutorial/

LICENSE: Some of the code in this file is based on reading Kanungo’s GPLv2+ implementation of discrete HMM’s, hence the present code must be licensed with a GPLv2+ compatible license.

AUTHOR:

- William Stein, 2010-03

class sage.stats.hmm.hmm.DiscreteHiddenMarkovModel
Bases: HiddenMarkovModel

A discrete Hidden Markov model implemented using double precision floating point arithmetic.

INPUT:

- A – a list of lists or a square $N \times N$ matrix, whose $(i,j)$ entry gives the probability of transitioning from state $i$ to state $j$.
- B – a list of $N$ lists or a matrix with $N$ rows, such that $B[i,k]$ gives the probability of emitting symbol $k$ while in state $i$.
- pi – the probabilities of starting in each initial state, i.e., $pi[i]$ is the probability of starting in state $i$.
- emission_symbols – None or list (default: None); if None, the emission_symbols are the ints [0..N-1], where $N$ is the number of states. Otherwise, they are the entries of the list emission_symbols, which must all be hashable.
- normalize – bool (default: True); if given, input is normalized to define valid probability distributions, e.g., the entries of $A$ are made nonnegative and the rows sum to 1, and the probabilities in $pi$ are normalized.

EXAMPLES:

```
sage: m = hmm.DiscreteHiddenMarkovModel([[0.4,0.6],[0.1,0.9]],
.....: [[0.1,0.9],[0.5,0.5]],
.....: [.5,.5]); m
Discrete Hidden Markov Model with 2 States and 2 Emissions
Transition matrix:
[0.4 0.6]
[0.1 0.9]
```

(continues on next page)
Emission matrix:
[[0.1, 0.9],
 [0.5, 0.5]]
Initial probabilities: [0.5000, 0.5000]
sage: m.log_likelihood([0, 1, 0, 1, 0, 1])
-4.66693474691329...
sage: m.viterbi([0, 1, 0, 1, 0, 1])
([1, 1, 1, 1, 1, 1], -5.378832842208748)
sage: m.baum_welch([0, 1, 0, 1, 0, 1])
(0.0, 22)
sage: m
# rel tol 1e-10
Discrete Hidden Markov Model with 2 States and 2 Emissions
Transition matrix:
[1.0134345614745788e-70, 1.0]
[1.0, 3.9974352713558623e-19]
Emission matrix:
[7.380221566254936e-54, 1.0]
[1.0, 3.9974352626002193e-19]
Initial probabilities: [0.0000, 1.0000]
sage: m.sample(10)
[0, 1, 0, 1, 0, 1, 0, 1, 0, 1]
sage: m.graph().plot()
# needs sage.plot
Graphics object consisting of 6 graphics primitives

A 3-state model that happens to always outputs 'b':
sage: m = hmm.DiscreteHiddenMarkovModel([[1/3]*3]*3, [[0, 1, 0]]*3, [1/3]*3, ['a', 'b', 'c'])
sage: m.sample(10)
['b', 'b', 'b', 'b', 'b', 'b', 'b', 'b', 'b', 'b']

baum_welch (obs, max_iter=100, log_likelihood_cutoff=0.0001, fix_emissions=False)
Given an observation sequence obs, improve this HMM using the Baum-Welch algorithm to increase the probability of observing obs.

INPUT:
• obs – list of emissions
• max_iter – integer (default: 100) maximum number of Baum-Welch steps to take
• log_likelihood_cutoff – positive float (default: 1e-4); the minimal improvement in likelihood with respect to the last iteration required to continue. Relative value to log likelihood.
• fix_emissions – bool (default: False); if True, do not change emissions when updating

OUTPUT:
changes the model in place, and returns the log likelihood and number of iterations.

EXAMPLES:
sage: m = hmm.DiscreteHiddenMarkovModel([[0.1, 0.9], [0.9, 0.1]],
....: [[.5, .5], [0.1]],
....: [.2, .8])
sage: m.baum_welch([1, 0]*20, log_likelihood_cutoff=0)
(0.0, 4)
sage: m
# rel tol 1e-14
Discrete Hidden Markov Model with 2 States and 2 Emissions

Transition matrix:
\[
\begin{bmatrix}
1.351269707707603e-51 & 1.0 \\
1.0 & 0.0
\end{bmatrix}
\]

Emission matrix:
\[
\begin{bmatrix}
1.0 & 6.462537138850569e-52 \\
0.0 & 1.0
\end{bmatrix}
\]

Initial probabilities: [0.0000, 1.0000]

The following illustrates how Baum-Welch is only a local optimizer, i.e., the above model is far more likely to produce the sequence [1,0]*20 than the one we get below:

```python
sage: m = hmm.DiscreteHiddenMarkovModel([[0.5,0.5],[0.5,0.5]],
.....: [[.5,.5],[.5,.5]],
.....: [.5,.5])
sage: m.baum_welch([1,0]*20, log_likelihood_cutoff=0)
(-27.725887222397784, 1)
sage: m
Discrete Hidden Markov Model with 2 States and 2 Emissions
Transition matrix:
[0.5 0.5]
[0.5 0.5]
Emission matrix:
[0.5 0.5]
[0.5 0.5]
Initial probabilities: [0.5000, 0.5000]
```

We illustrate fixing emissions:

```python
sage: m = hmm.DiscreteHiddenMarkovModel([[0.1,0.9],[0.9,0.1]],
.....: [[.5,.5],[.2,.8]],
.....: [.2,.8])
sage: set_random_seed(0); v = m.sample(100)
sage: m.baum_welch(v,fix_emissions=True)
(-66.98630856918774, 100)
sage: m.emission_matrix()
[0.5 0.5]
[0.2 0.8]
sage: m = hmm.DiscreteHiddenMarkovModel([[0.1,0.9],[0.9,0.1]],
.....: [[.5,.5],[.2,.8]],
.....: [.2,.8])
sage: m.baum_welch(v)
(-66.782360659293..., 100)
sage: m.emission_matrix()  # rel tol 1e-14
[ 0.5303085748626447 0.46969142513735535]
[ 0.2909775550173978 0.709022449826023]
```

**emission_matrix()**

Return the matrix whose \(i\)-th row specifies the emission probability distribution for the \(i\)-th state.

More precisely, the \(i, j\) entry of the matrix is the probability of the Markov model outputting the \(j\)-th symbol when it is in the \(i\)-th state.

**OUTPUT:** a Sage matrix with real double precision (RDF) entries.

**EXAMPLES:**
sage: m = hmm.DiscreteHiddenMarkovModel(
    ...:     [[0.4,0.6],[0.1,0.9]],
    ...:     [[0.1,0.9],[0.5,0.5]],
    ...:     [.5,.5])
sage: E = m.emission_matrix(); E
[0.1  0.9]
[0.5  0.5]

The returned matrix is mutable, but changing it does not change the transition matrix for the model:

sage: E[0,0] = 0; E[0,1] = 1
sage: m.emission_matrix()
[0.1  0.9]
[0.5  0.5]

**generate_sequence** *(length, starting_state=None)*

Return a sample of the given length from this HMM.

**INPUT:**

- length – positive integer
- starting_state – int (or None): if specified, generate a sequence using this model starting with the given state instead of the initial probabilities to determine the starting state.

**OUTPUT:**

- an *IntList* or list of emission symbols
- *IntList* of the actual states the model was in when emitting the corresponding symbols

**EXAMPLES:**

In this example, the emission symbols are not set:

sage: set_random_seed(0)
sage: a = hmm.DiscreteHiddenMarkovModel([[0.1,0.9],[0.1,0.9]],
    ...:     [[1,0],[0,1]],
    ...:     [0,1])
sage: a.generate_sequence(5)
([1, 0, 1, 1, 1], [1, 0, 1, 1, 1])
sage: list(a.generate_sequence(1000)[0]).count(0)
90

Here the emission symbols are set:

sage: set_random_seed(0)
sage: a = hmm.DiscreteHiddenMarkovModel([[0.5,0.5],[0.1,0.9]],
    ...:     [[1,0],[0,1]],
    ...:     [0,1], ['up', 'down'])
sage: a.generate_sequence(5)
(['down', 'up', 'down', 'down', 'down'], [1, 0, 1, 1, 1])

Specify the starting state:

sage: set_random_seed(0); a.generate_sequence(5, starting_state=0)
(['up', 'up', 'down', 'down', 'down'], [0, 0, 1, 1, 1])

**log_likelihood** *(obs, scale=True)*

Return the logarithm of the probability that this model produced the given observation sequence. Thus the output is a non-positive number.
INPUT:
- obs – sequence of observations
- scale – boolean (default: True); if True, use rescaling to avoid loss of precision due to the very limited dynamic range of floats. You should leave this as True unless the obs sequence is very small.

EXAMPLES:

```python
sage: m = hmm.DiscreteHiddenMarkovModel([[0.4,0.6],[0.1,0.9]],
....: [[0.1,0.9],[0.5,0.5]],
....: [.2,.8])
sage: m.log_likelihood([0, 1, 0, 1, 0, 1, 0, 0, 0])
-7.3301308009370825
sage: m.log_likelihood([0, 1, 0, 1, 0, 1, 0, 0, 0], scale=False)
-7.330130800937082
sage: m.log_likelihood([])
0.0
sage: m = hmm.DiscreteHiddenMarkovModel([[0.4,0.6],[0.1,0.9]],
....: [[0.1,0.9],[0.5,0.5]],
....: [.2,.8], ['happy','sad'])
sage: m.log_likelihood(['happy','happy'])
-1.6565295199679506
sage: m.log_likelihood(['happy','sad'])
-1.4731602941415523
```

Overflow from not using the scale option:

```python
sage: m = hmm.DiscreteHiddenMarkovModel([[0.4,0.6],[0.1,0.9]],
....: [[0.1,0.9],[0.5,0.5]],
....: [.2,.8])
```

```python
date
```

viterbi (obs, log_scale=True)

Determine “the” hidden sequence of states that is most likely to produce the given sequence seq of observations, along with the probability that this hidden sequence actually produced the observation.

INPUT:
- seq – sequence of emitted ints or symbols
- log_scale – bool (default: True) whether to scale the sequence in order to avoid numerical overflow.

OUTPUT:
- list – “the” most probable sequence of hidden states, i.e., the Viterbi path.
- float – log of probability that the observed sequence was produced by the Viterbi sequence of states.

EXAMPLES:

```python
sage: a = hmm.DiscreteHiddenMarkovModel([[0.1,0.9],[0.1,0.9]],
....: [[0.9,0.1],[0.1,0.9]],
....: [0.5,0.5])
sage: a.viterbi([1,0,0,1,0,0,1,1])
([1, 0, 0, 1, ..., 0, 1, 1], -11.06245322477221...)
```

We predict the state sequence when the emissions are 3/4 and ‘abc’.
Note that state 0 is common below, despite the model trying hard to switch to state 1:

```python
([0, 1, 1, 0, 0 ... 0, 0, 0, 0, 0], -25.299405845367794)
```

### class sage.stats.hmm.HiddenMarkovModel

Abstract base class for all Hidden Markov Models.

#### graph (eps=0.001)

Create a weighted directed graph from the transition matrix, not including any edge with a probability less than eps.

**INPUT:**

- **eps** – nonnegative real number

**OUTPUT:** a DiGraph

**EXAMPLES:**

```python
sage: m = hmm.DiscreteHiddenMarkovModel([[.3,0,.7], [0,0,1], [.5,.5,0]],
          [[.5,.5,.2]]*3,
          [.2,.8])
```

#### initial_probabilities()

Return the initial probabilities as a TimeSeries of length $N$, where $N$ is the number of states of the Markov model.

**EXAMPLES:**

```python
sage: m = hmm.DiscreteHiddenMarkovModel([[.4,0.6], [0.1,0.9]],
          [[.01,.9], [.5,.5]],
          [.2,.8])
```

The returned time series is a copy, so changing it does not change the model:

```python
sage: pi[0] = .1; pi[1] = .9
sage: m.initial_probabilities()  
[0.2000, 0.8000]
```
Some other models:

```
sage: m = hmm.GaussianHiddenMarkovModel([[.9], [.5, .5]],
            [(1,1), (-1,1)],
            [.1, .9])

sage: m.initial_probabilities()
[0.1000, 0.9000]

sage: m = hmm.GaussianMixtureHiddenMarkovModel(
            [[.9, .1], [.4, .6]],
            [[(1, 0, 1), (1, 0.1)], [(1, 0, 1)]],
            [.7, .3])

sage: m.initial_probabilities()
[0.7000, 0.3000]
```

**sample**(length, number=None, starting_state=None)

Return number samples from this HMM of given length.

**INPUT:**

- **length** – positive integer
- **number** – (default: None) if given, compute list of this many sample sequences
- **starting_state** – int (or None); if specified, generate a sequence using this model starting with the given state instead of the initial probabilities to determine the starting state.

**OUTPUT:**

- if `number` is not given, return a single `TimeSeries`
- if `number` is given, return a list of `TimeSeries`

**EXAMPLES:**

```
sage: set_random_seed(0)
sage: a = hmm.DiscreteHiddenMarkovModel([[0.1, 0.9], [0.1, 0.9]],
            [1, 0],
            [0, 1])

sage: print(a.sample(10, 3))
[[1, 1, 1, 1], [0, 1, 1, 1], [1, 1, 1, 1], [1, 1, 1, 1], [1, 1, 1, 1]]

sage: a.sample(15)
[1, 1, 1, 0 ... 1, 1, 1, 1, 1]

sage: a.sample(3, 1)
[[1, 1, 1]]

sage: list(a.sample(1000)).count(0)
88
```

If the emission symbols are set:

```
sage: set_random_seed(0)
sage: a = hmm.DiscreteHiddenMarkovModel([[0.5, 0.5], [0.1, 0.9]],
            [1, 0],
            ['up', 'down'])

sage: a.sample(10)
['down', 'up', 'down', 'down', 'down', 'down', 'up', 'up', 'up', 'up']
```

Force a starting state:
sage: set_random_seed(0); a.sample(10, starting_state=0)
['up', 'up', 'down', 'down', 'down', 'up', 'up', 'up', 'up', 'up']

**transition_matrix()**

Return the state transition matrix.

**OUTPUT:** a Sage matrix with real double precision (RDF) entries.

**EXAMPLES:**

```sage
sage: M = hmm.DiscreteHiddenMarkovModel([[0.7,0.3],[0.9,0.1]],
                                           ....:       [[0.5,.5],[.1,.9]],
                                           ....:       [0.3,0.7])
sage: T = M.transition_matrix(); T
[0.7 0.3]
[0.9 0.1]
```

The returned matrix is mutable, but changing it does not change the transition matrix for the model:

```sage
sage: T[0,0] = .1; T[0,1] = .9
sage: M.transition_matrix()
[0.7 0.3]
[0.9 0.1]
```

Transition matrices for other types of models:

```sage
sage: M = hmm.GaussianHiddenMarkovModel([[.1,.9],[.5,.5]],
                                             ....:       [(1,1), (-1,1)],
                                             ....:       [.5,.5])
sage: M.transition_matrix()
[0.1 0.9]
[0.5 0.5]
sage: M = hmm.GaussianMixtureHiddenMarkovModel(
                                             ....:       [[.9,.1],[.4,.6]],
                                             ....:       [[(.4,(0,1)), (.6,(1,0.1))],[1,(0,1)]]],
                                             ....:       [.7,.3])
sage: M.transition_matrix()
[0.9 0.1]
[0.4 0.6]
```

sage.stats.hmm.hmm.unpickle_discrete_hmm_v0(A, B, pi, emission_symbols, name)

sage.stats.hmm.hmm.unpickle_discrete_hmm_v1(A, B, pi, n_out, emission_symbols, emission_symbols_dict)

Return a **DiscreteHiddenMarkovModel**, restored from the arguments.

This function is used internally for unpickling.
CONTINUOUS EMISSION HIDDEN MARKOV MODELS

AUTHOR:
• William Stein, 2010-03

```python
class sage.stats.hmm.chmm.GaussianHiddenMarkovModel
    Bases: HiddenMarkovModel

    Gaussian emissions Hidden Markov Model.

    INPUT:
    • A – matrix; the \( N \times N \) transition matrix
    • B – list of pairs (mu, sigma) that define the distributions
    • pi – initial state probabilities
    • normalize – bool (default: True)

    EXAMPLES:
    We illustrate the primary functions with an example 2-state Gaussian HMM:

    sage: m = hmm.GaussianHiddenMarkovModel([[0.1, 0.9], [0.5, 0.5]], [(1,1), (-1,1)], [.5,.5]); m
    Gaussian Hidden Markov Model with 2 States
    Transition matrix:
    [0.1 0.9]
    [0.5 0.5]
    Emission parameters:
    [(1.0, 1.0), (-1.0, 1.0)]
    Initial probabilities: [0.5000, 0.5000]
```

We query the defining transition matrix, emission parameters, and initial state probabilities:

```python
sage: m.transition_matrix()
[0.1 0.9]
[0.5 0.5]
sage: m.emission_parameters()
[(1.0, 1.0), (-1.0, 1.0)]
sage: m.initial_probabilities()
[0.5000, 0.5000]
```

We obtain a sample sequence with 10 entries in it, and compute the logarithm of the probability of obtaining this sequence, given the model:
We compute the Viterbi path, and probability that the given path of states produced obs:

```
sage: m.viterbi(obs) # random
([[1, 0, 1, 0, 1], -8.714092684611794]
```

We use the Baum-Welch iterative algorithm to find another model for which our observation sequence is more likely:

```
sage: try:
....:     p, s = m.baum_welch(obs)
....:     assert p > log_likelihood
....:     assert (1 <= s <= 500)
....: except RuntimeError:
....:     pass
```

Notice that running Baum-Welch changed our model:

```
sage: m # random
Gaussian Hidden Markov Model with 2 States
Transition matrix:
0.4154981366185841 0.584501863381416
[ 0.9999993174253741 6.825746258991804e-07
Emission parameters:
[(0.4178882427119503, 0.5173109664360919),
(-1.5025208631331122, 0.5085512836055119)]
Initial probabilities: [0.0000, 1.0000]
```

**baum_welch**(obs, max_iter=500, log_likelihood_cutoff=0.0001, min_sd=0.01, fix_emissions=False, v=False)

Given an observation sequence obs, improve this HMM using the Baum-Welch algorithm to increase the probability of observing obs.

**INPUT:**

- obs – a time series of emissions
- max_iter – integer (default: 500) maximum number of Baum-Welch steps to take
- log_likelihood_cutoff – positive float (default: 1e-4); the minimal improvement in likelihood with respect to the last iteration required to continue. Relative value to log likelihood.
- min_sd – positive float (default: 0.01); when reestimating, the standard deviation of emissions is not allowed to be less than min_sd.
• **fix_emissions** – bool (default: False); if True, do not change emissions when updating

**OUTPUT:**
changes the model in place, and returns the log likelihood and number of iterations.

**EXAMPLES:**

```python
sage: m = hmm.GaussianHiddenMarkovModel([[.1,.9],[.5,.5]],
.....:     [(1,.5), (-1,3)],
.....:     [.1,.9])
sage: m.log_likelihood([-2,-1,.1,0.1])
-8.85822215986275
sage: m.baum_welch([-2,-1,.1,0.1])
(4.534646052182..., 7)
sage: m.log_likelihood([-2,-1,.1,0.1])
4.534646052182...
```

```
sage: m # rel tol 3e-14
Gaussian Hidden Markov Model with 2 States
Transition matrix:
[ 0.9999999999999999 7.56983934440382e-10]
[ 0.49998462791192644 0.5000153720880736]
Emission parameters:
[(0.09999999999999999, 0.01), (-1.4999508147591902, 0.5000710504895474)]
Initial probabilities: [0.0000, 1.0000]
```

We illustrate bounding the standard deviation below. Note that above we had different emission parameters when the min_sd was the default of 0.01:

```python
sage: m = hmm.GaussianHiddenMarkovModel([[.1,.9],[.9,.1]],
.....:     [(1,2), (-1,.5)],
.....:     [.3,.7])
sage: m.baum_welch([-2,-1,.1,0.1], min_sd=1)
(-4.07939572755..., 32)
sage: m.emission_parameters()
[(-0.2663018798..., 1.0), (-1.99850979..., 1.0)]
```

We watch the log likelihoods of the model converge, step by step:

```python
sage: m = hmm.GaussianHiddenMarkovModel([[.1,.9],[.5,.5]],
.....:     [(1,.5), (-1,3)],
.....:     [.1,.9])
sage: v = m.sample(10)
sage: l = stats.TimeSeries([m.baum_welch(v, max_iter=1)[0]
.....:     for _ in range(len(v))])
sage: all(l[i] <= l[i+1] + 0.0001 for i in range(9))
True
sage: l # random
```

We illustrate fixing emissions:

```python
sage: m = hmm.GaussianHiddenMarkovModel([[.1,.9],[.9,.1]],
.....:     [(1,2), (-1,.5)],
.....:     [.3,.7])
sage: set_random_seed(0); v = m.sample(100)
sage: m.baum_welch(v, fix_emissions=True)
(-164.72944548204..., 23)
```
sage: m.emission_parameters()
[(1.0, 2.0), (-1.0, 0.5)]

sage: m = hmm.GaussianHiddenMarkovModel([[.1,.9],[.9,.1]],
....: [(1,2),(-1,.5)],
....: [.3,.7])

sage: m.baum_welch(v)
(-162.854370397998..., 49)

sage: m.emission_parameters()  # rel tol 3e-14
[(1.2722419172602375, 2.371368751761901),
 (-0.9486174675179113, 0.5762360385123765)]

emission_parameters()
Return the parameters that define the normal distributions associated to all of the states.

OUTPUT:

a list \( B \) of pairs \( B[i] = (\mu, \sigma) \), such that the distribution associated to state \( i \) is normal with mean \( \mu \) and standard deviation \( \sigma \).

EXAMPLES:

sage: M = hmm.GaussianHiddenMarkovModel([[.1,.9],[.5,.5]],
....: [(1,.5), (-1,3)],
....: [.1,.9])

sage: M.emission_parameters()
[(1.0, 0.5), (-1.0, 3.0)]

generate_sequence(length, starting_state=None)
Return a sample of the given length from this HMM.

INPUT:

• length – positive integer

• starting_state – int (or None); if specified then generate a sequence using this model starting with the given state instead of the initial probabilities to determine the starting state.

OUTPUT:

• an \texttt{IntList} or list of emission symbols

• \texttt{TimeSeries} of emissions

EXAMPLES:

sage: m = hmm.GaussianHiddenMarkovModel([[.1,.9],[.5,.5]],
....: [(1,.5), (-1,3)],
....: [.1,.9])

sage: m.generate_sequence(5)  # random
([-3.0505, 0.5317, -4.5065, 0.6521, 1.0435], [1, 0, 1, 0, 1])

sage: m.generate_sequence(0)
([], [])

sage: m.generate_sequence(-1)
Traceback (most recent call last):
...
ValueError: length must be nonnegative

Verify numerically that the starting state is 0 with probability about 0.1:
```sage
counter = 0
n = 0
def add_samples(i):
    global counter, n
    for i in range(i):
        n += 1
        if m.generate_sequence(1)[1][0] == 0:
            counter += 1
add_samples(10^5)
while abs(counter*1.0 / n - 0.1) > 0.01: add_samples(10^5)
```

Example in which the starting state is 0 (see github issue #11452):

```sage
set_random_seed(23); m.generate_sequence(2)
([0.6501, -2.0151], [0, 1])
```

Force a starting state of 1 even though as we saw above it would be 0:

```sage
set_random_seed(23); m.generate_sequence(2, starting_state=1)
([-3.1491, -1.0244], [1, 1])
```

**log_likelihood**(obs)

Return the logarithm of a continuous analogue of the probability that this model produced the given observation sequence.

Note that the “continuous analogue of the probability” above can be bigger than 1, hence the logarithm can be positive.

**INPUT:**

- obs – sequence of observations

**OUTPUT:**

- float

**EXAMPLES:**

```sage
m = hmm.GaussianHiddenMarkovModel([[.1,.9],[.5,.5]],
                                        [(1,.5), (-1,3)],
                                        [.1,.9])
log_likelihood([1,1,1])
-4.297880766072486
s = m.sample(20)
log_likelihood(s) < -20
True
```

**viterbi**(obs)

Determine “the” hidden sequence of states that is most likely to produce the given sequence obs of observations, along with the probability that this hidden sequence actually produced the observation.

**INPUT:**

- obs – sequence of emitted ints or symbols

**OUTPUT:**

- list – “the” most probable sequence of hidden states, i.e., the Viterbi path.
- float – log of probability that the observed sequence was produced by the Viterbi sequence of states.
EXAMPLES:

We find the optimal state sequence for a given model:

\[
\text{sage: } m = \text{hmm.GaussianHiddenMarkovModel([[0.5,0.5],[0.5,0.5]],} \\
\text{.....: } [(0.1),(1.0,1)], \\
\text{.....: } [0.5,0.5]) \\
\text{sage: } m.\text{viterbi([0,1,10,10,1])} \\
\text{([0, 1, 1, 1, 0], -9.0604285868230...)}
\]

Another example in which the most likely states change based on the last observation:

\[
\text{sage: } m = \text{hmm.GaussianHiddenMarkovModel([[.1,.9],[.5,.5]],} \\
\text{.....: } [(1.5), (-1,3)], \\
\text{.....: } [.1,.9]) \\
\text{sage: } m.\text{viterbi([-2,-1,.1,0.1])} \\
\text{([1, 1, 0, 1], -9.61823698847639...)} \\
\text{sage: } m.\text{viterbi([-2,-1,.1,0.3])} \\
\text{([1, 1, 1, 0], -9.56602365378513)}
\]

class sage.stats.hmm.chmm.GaussianMixtureHiddenMarkovModel

Bases: GaussianHiddenMarkovModel

Gaussian mixture Hidden Markov Model.

INPUT:

- \(A\) – matrix; the \(N \times N\) transition matrix
- \(B\) – list of mixture definitions for each state. Each state may have a varying number of gaussians with selection probabilities that sum to 1 and encoded as \((p, (\mu,\sigma))\)
- \(\pi\) – initial state probabilities
- normalize – bool (default: \(\text{True}\)); if given, input is normalized to define valid probability distributions, e.g., the entries of \(A\) are made nonnegative and the rows sum to 1, and the probabilities in \(\pi\) are normalized.

EXAMPLES:

\[
\text{sage: } A = [[0.5,0.5],[0.5,0.5]] \\
\text{sage: } B = [[(0.9,(0.0,1.0)), (0.1,(1,10000))],[(1,(1,1)), (0,(0,0.1))]] \\
\text{sage: } \text{hmm.GaussianMixtureHiddenMarkovModel(A, B, [1,0])} \\
\] Gaussian Mixture Hidden Markov Model with 2 States

Transition matrix:

\[\begin{bmatrix}
0.5 & 0.5 \\
0.5 & 0.5
\end{bmatrix}\]

Emission parameters:

\[0.9*N(0.0,1.0) + 0.1*N(1.0,10000.0), 1.0*N(1.0,1.0) + 0.0*N(0.0,0.1)\]

Initial probabilities: \([1.0000, 0.0000]\)

baum_welch \((\text{obs}, \text{max_iter}=1000, \text{log_likelihood_cutoff}=1e-12, \text{min_sd}=0.01, \text{fix_emissions}=\text{False})\)

Given an observation sequence \(\text{obs}\), improve this HMM using the Baum-Welch algorithm to increase the probability of observing \(\text{obs}\).

INPUT:

- \(\text{obs}\) – a time series of emissions
- \(\text{max_iter}\) – integer (default: 1000) maximum number of Baum-Welch steps to take
- \(\text{log_likelihood_cutoff}\) – positive float (default: \(1e-12\)); the minimal improvement in likelihood with respect to the last iteration required to continue. Relative value to log likelihood.
• \texttt{min\_sd} – positive float (default: 0.01); when reestimating, the standard deviation of emissions is not allowed to be less than \texttt{min\_sd}.

• \texttt{fix\_emissions} – bool (default: False); if True, do not change emissions when updating

\textbf{OUTPUT:}

changes the model in place, and returns the log likelihood and number of iterations.

\textbf{EXAMPLES:}

\begin{verbatim}
sage: m = hmm.GaussianMixtureHiddenMarkovModel(
.....:  [[.9,.1],[.4,.6]],
.....:  [[.4,(0,1)), (.6,(1,0.1))], [(1,(0,1))]],
.....:  [.7,.3])
sage: set_random_seed(0); v = m.sample(10); v
[0.3576, -0.9365, 0.9449, -0.6957, 1.0217, 0.9644, 0.9987, -0.5950, -1.0219, 0.6477]
sage: m.log_likelihood(v)
-8.31408655939536...
sage: m.baum_welch(v)
(2.18905068682..., 15)
sage: m.log_likelihood(v)
2.18905068682...
sage: m  # rel tol 6e-12
Gaussian Mixture Hidden Markov Model with 2 States
Transition matrix:
[ 0.8746363339773399 0.12536366602266016]
[ 1.0 1.451685202290174e-40]
Emission parameters:
[0.500161629343*N(-0.812298726239,0.173329026744)
 + 0.499838370657*N(0.982433690378,0.029719932009),
  1.0*N(0.503260056832,0.145881515324)]
Initial probabilities: [0.0000, 1.0000]
\end{verbatim}

We illustrate bounding the standard deviation below. Note that above we had different emission parameters when the \texttt{min\_sd} was the default of 0.01:

\begin{verbatim}
sage: m = hmm.GaussianMixtureHiddenMarkovModel(
.....:  [[.9,.1],[.4,.6]],
.....:  [[.4,(0,1)), (.6,(1,0.1))], [(1,(0,1))]],
.....:  [.7,.3])
sage: m.baum_welch(v, min\_sd=1)
(-12.617885761692..., 1000)
sage: m.emission_parameters()  # rel tol 6e-12
[0.503545634447*N(0.200166509595,1.0) + 0.496454365553*N(0.200166509595,1.0),
  1.0*N(0.0543433426535,1.0)]
\end{verbatim}

We illustrate fixing all emissions:

\begin{verbatim}
sage: m = hmm.GaussianMixtureHiddenMarkovModel(
.....:  [[.9,.1],[.4,.6]],
.....:  [[.4,(0,1)), (.6,(1,0.1))], [(1,(0,1))]],
.....:  [.7,.3])
sage: set_random_seed(0); v = m.sample(10)
sage: m.baum_welch(v, fix\_emissions=True)
(-7.58656858997..., 36)
sage: m.emission_parameters()
[0.4*N(0.0,1.0) + 0.6*N(1.0,0.1),
  1.0*N(0.0,1.0)]
\end{verbatim}
emission_parameters()  
Returns a list of all the emission distributions.

OUTPUT:
list of Gaussian mixtures

EXAMPLES:

```python
sage: m = hmm.GaussianMixtureHiddenMarkovModel([[.9,.1],[.4,.6]],
                                           [[(.4,(0,1)), (.6,(1,0.1))],
                                           [(1,(0,1))],
                                           [.7,.3])
sage: m.emission_parameters()  
[0.4*N(0.0,1.0) + 0.6*N(1.0,0.1), 1.0*N(0.0,1.0)]
```
These distribution classes are designed specifically for HMM’s and not for general use in statistics. For example, they have fixed or non-fixed status, which only make sense relative to being used in a hidden Markov model.

AUTHOR:

• William Stein, 2010-03

```python
class sage.stats.hmm.distributions.DiscreteDistribution
    Bases: Distribution
class sage.stats.hmm.distributions.Distribution
    Bases: object
    A distribution.

plot(*args, **kwds)
    Return a plot of the probability density function.
    INPUT:
    • args and kwds, passed to the Sage plot() function
    OUTPUT:
    • a Graphics object

EXAMPLES:

    sage: P = hmm.GaussianMixtureDistribution([(0.2,-10,.5),(.6,1,1),(0.2,20,.5)])
    sage: P.plot(-10,30)  # needs sage.plot
    Graphics object consisting of 1 graphics primitive
```

```python
prob(x)
    The probability density function evaluated at x.
    INPUT:
    • x – object
    OUTPUT:
    • float

EXAMPLES:
    This method must be defined in a derived class:
```
```
sage: import sage.stats.hmm.distributions
sage: sage.stats.hmm.distributions.Distribution().prob(0)
Traceback (most recent call last):
...
NotImplementedError
```

**sample** (*n=None*)

Return either a single sample (the default) or *n* samples from this probability distribution.

**INPUT:**

- *n* – *None* or a positive integer

**OUTPUT:**

- a single sample if *n* is 1; otherwise many samples

**EXAMPLES:**

This method must be defined in a derived class:

```
sage: import sage.stats.hmm.distributions
sage: sage.stats.hmm.distributions.Distribution().sample()
Traceback (most recent call last):
...
NotImplementedError
```

class **sage.stats.hmm.distributions.GaussianDistribution**

**Bases:** *Distribution*

**class** **sage.stats.hmm.distributions.GaussianMixtureDistribution**

**Bases:** *Distribution*

A probability distribution defined by taking a weighted linear combination of Gaussian distributions.

**EXAMPLES:**

```
sage: P = hmm.GaussianMixtureDistribution([(0.3, 1, 2), (0.7, -1, 1)]); P
0.3*N(1.0, 2.0) + 0.7*N(-1.0, 1.0)
sage: P[0]
(0.3, 1.0, 2.0)
sage: P.is_fixed()
False
sage: P.fix(1)
sage: P.is_fixed(0)
False
sage: P.is_fixed(1)
True
sage: P.unfix(1)
sage: P.is_fixed(1)
False
```

**fix** (*i=None*)

Set that this **GaussianMixtureDistribution** (or its *i*-th component) is fixed when using
Baum-Welch to update the corresponding HMM.

**INPUT:**

- *i* – *None* (default) or integer; if given, only fix the *i*-th component

**EXAMPLES:**
.. code:: python

    sage: P = hmm.GaussianMixtureDistribution([(0.2,-10,.5),(.6,1,1),(.2,20,.5)])
    sage: P.is_fixed()  # Doctest: +NORMALIZE_WHITESPACE
    False
    sage: P.is_fixed(1)  # Doctest: +NORMALIZE_WHITESPACE
    True
    sage: P.fix(); P.is_fixed()  # Doctest: +NORMALIZE_WHITESPACE
    True

is_fixed\(i=None\)

Return whether or not this :class:`GaussianMixtureDistribution` is fixed when using Baum-Welch to update the corresponding HMM.

**INPUT:**

- \(i\) – None (default) or integer; if given, only return whether the \(i\)-th component is fixed

**EXAMPLES:**

.. code:: python

    sage: P = hmm.GaussianMixtureDistribution([(0.2,-10,.5),(.6,1,1),(.2,20,.5)])
    sage: P.is_fixed()  # Doctest: +NORMALIZE_WHITESPACE
    False
    sage: P.is_fixed(0)  # Doctest: +NORMALIZE_WHITESPACE
    False
    sage: P.fix(0); P.is_fixed()  # Doctest: +NORMALIZE_WHITESPACE
    False
    sage: P.is_fixed(0)  # Doctest: +NORMALIZE_WHITESPACE
    True
    sage: P.fix(); P.is_fixed()  # Doctest: +NORMALIZE_WHITESPACE
    True

prob\(x\)

Return the probability of \(x\).

Since this is a continuous distribution, this is defined to be the limit of the \(p\)'s such that the probability of \([x,x+h]\) is \(p\cdot h\).

**INPUT:**

- \(x\) – float

**OUTPUT:**

- float

**EXAMPLES:**

.. code:: python

    sage: P = hmm.GaussianMixtureDistribution([(0.2,-10,.5),(.6,1,1),(.2,20,.5)])
    sage: P.prob(.5)  # Doctest: +NORMALIZE_WHITESPACE
    0.21123919605857971
    sage: P.prob(-100)  # Doctest: +NORMALIZE_WHITESPACE
    0.0
    sage: P.prob(20)  # Doctest: +NORMALIZE_WHITESPACE
    0.1595769121605731

prob_m\((x, m)\)

Return the probability of \(x\) using just the \(m\)-th summand.

**INPUT:**

- \(x\) – float
Statistics, Release 10.3

- \( m \) – integer

**OUTPUT:**
- float

**EXAMPLES:**

```sage
sage: P = hmm.GaussianMixtureDistribution([(0.2,-10,.5),(.6,1,1),(.2,20,.5)])
sage: P.prob_m(.5, 0)
2.760817680508...e-97
sage: P.prob_m(.5, 1)
0.21123919605857971
sage: P.prob_m(.5, 2)
0.0
```

**sample** *(n=None)*

Return a single sample from this distribution (by default), or if \( n > 1 \), return a `TimeSeries` of samples.

**INPUT:**
- \( n \) – integer or `None` (default: `None`)

**OUTPUT:**
- float if \( n \) is `None` (default); otherwise a `TimeSeries`

**EXAMPLES:**

```sage
sage: P = hmm.GaussianMixtureDistribution([(0.2,-10,.5),(.6,1,1),(.2,20,.5)])
sage: type(P.sample())
<class 'float'>
sage: l = P.sample(1)
sage: len(l)
1
sage: type(l)
<class 'sage.stats.time_series.TimeSeries'>
sage: l = P.sample(5)
sage: len(l)
5
sage: type(l)
<class 'sage.stats.time_series.TimeSeries'>
sage: l = P.sample(0)
sage: len(l)
0
sage: type(l)
<class 'sage.stats.time_series.TimeSeries'>
sage: P.sample(-3)
Traceback (most recent call last):
  ... ValueError: n must be nonnegative
```

**unfix** *(i=None)*

Set that this `GaussianMixtureDistribution` (or its \( i \)-th component) is not fixed when using Baum-Welch to update the corresponding HMM.

**INPUT:**
- \( i \) – `None` (default) or integer; if given, only fix the \( i \)-th component

**EXAMPLES:**
sage: P = hmm.GaussianMixtureDistribution([(0.2, -10, 0.5), (0.6, 1, 1), (0.2, 20, 0.5)])
sage: P.fix(1); P.is_fixed(1)
True
sage: P.unfix(1); P.is_fixed(1)
False
sage: P.fix(); P.is_fixed()
True
sage: P.unfix(); P.is_fixed()
False

sage.stats.hmm.distributions.unpickle_gaussian_mixture_distribution_v1(c0, c1, param, fixed)

Used in unpickling GaussianMixtureDistribution objects.

EXAMPLES:

sage: P = hmm.GaussianMixtureDistribution([(0.2, -10, 0.5), (0.6, 1, 1), (0.2, 20, 0.5)])
sage: loads(dumps(P)) == P  # indirect doctest
True
HIDDEN MARKOV MODELS – UTILITY FUNCTIONS

AUTHOR:

• William Stein, 2010-03

class sage.stats.hmm.util.HMM_Util
Bases: object

A class used in order to share cdef’s methods between different files.

initial_probs_to_TimeSeries(pi, normalize)

This function is used internally by the __init__ methods of various Hidden Markov Models.

INPUT:

• pi – vector, list, or TimeSeries
• normalize – if True, replace negative entries by 0 and rescale to ensure that the sum of the entries in each row is equal to 1. If the sum of the entries in a row is 0, replace them all by $1/N$.

OUTPUT:

• a TimeSeries of length $N$

EXAMPLES:

```
sage: import sage.stats.hmm.util
sage: u = sage.stats.hmm.util.HMM_Util()
sage: u.initial_probs_to_TimeSeries([0.1,0.2,0.9], True)
[0.0833, 0.1667, 0.7500]
sage: u.initial_probs_to_TimeSeries([0.1,0.2,0.9], False)
[0.1000, 0.2000, 0.9000]
```

normalize_probability_TimeSeries(T, i, j)

This function is used internally by the Hidden Markov Models code.

Replace entries of $T[i:j]$ in place so that they are all nonnegative and sum to 1. Negative entries are replaced by 0 and $T[i:j]$ is then rescaled to ensure that the sum of the entries in each row is equal to 1. If all entries are 0, replace them by $1/(j-i)$.

INPUT:

• $T$ – a TimeSeries
• $i$ – nonnegative integer
• $j$ – nonnegative integer

OUTPUT:

• $T$ is modified
EXAMPLES:

```python
sage: import sage.stats.hmm.util
sage: T = stats.TimeSeries([.1, .3, .7, .5])
sage: u = sage.stats.hmm.util.HMM_Util()
sage: u.normalize_probability_TimeSeries(T,0,3)
sage: T
[0.0909, 0.2727, 0.6364, 0.5000]
sage: u.normalize_probability_TimeSeries(T,0,4)
sage: T
[0.0606, 0.1818, 0.4242, 0.3333]
sage: abs(T.sum()-1) < 1e-8  # might not exactly equal 1 due to rounding
True
```

`state_matrix_to_TimeSeries(A, N, normalize)`

This function is used internally by the `__init__` methods of Hidden Markov Models to make a transition matrix from A.

**INPUT:**

- A – matrix, list, list of lists, or TimeSeries
- N – number of states
- normalize – if True, replace negative entries by 0 and rescale to ensure that the sum of the entries in each row is equal to 1. If the sum of the entries in a row is 0, replace them all by 1/N.

**OUTPUT:**

- a TimeSeries

**EXAMPLES:**

```python
sage: import sage.stats.hmm.util
sage: u = sage.stats.hmm.util.HMM_Util()
sage: u.state_matrix_to_TimeSeries([[.1,.7],[3/7,4/7]], 2, True)
[0.1250, 0.8750, 0.4286, 0.5714]
sage: u.state_matrix_to_TimeSeries([[.1,.7],[3/7,4/7]], 2, False)
[0.1000, 0.7000, 0.4286, 0.5714]
```
This class realizes oracles which returns integers proportionally to \( \exp\left(-\frac{(x - c)^2}{(2^2)}\right) \). All oracles are implemented using rejection sampling. See \code{DiscreteGaussianDistributionIntegerSampler.__init__()\} for which algorithms are available.

**AUTHORS:**

**EXAMPLES:**

We construct a sampler for the distribution \( D_{3,c} \) with width \( c = 3 \) and center \( c = 0 \):

```python
sage: from sage.stats.distributions.discrete_gaussian_integer import...
   -DiscreteGaussianDistributionIntegerSampler
sage: sigma = 3.0
sage: D = DiscreteGaussianDistributionIntegerSampler(sigma=sigma)
```

We ask for 100000 samples:

```python
sage: from collections import defaultdict
sage: counter = defaultdict(Integer)
sage: n = 0
sage: def add_samples(i):
    ....:     global counter, n
    ....:     for _ in range(i):
    ....:         counter[D()] += 1
    ....:     n += 1
sage: add_samples(100000)
```

These are sampled with a probability proportional to \( \exp\left(-\frac{x^2}{18}\right) \). More precisely we have to normalise by dividing by the overall probability over all integers. We use the fact that hitting anything more than 6 standard deviations away is very unlikely and compute:

```python
sage: bound = (6*sigma).floor()
sage: norm_factor = sum([exp(-x^2/(2*sigma^2)) for x in range(-bound,bound+1)])
```

With this normalisation factor, we can now test if our samples follow the expected distribution:

```python
sage: expected = lambda x : ZZ(round(n*exp(-x^2/(2*sigma^2))/norm_factor))
sage: observed = lambda x : counter[x]
sage: add_samples(10000)
```
We construct an instance with a larger width:

```python
sage: from sage.stats.distributions.discrete_gaussian_integer import DiscreteGaussianDistributionIntegerSampler
sage: sigma = 127
sage: D = DiscreteGaussianDistributionIntegerSampler(sigma=sigma, algorithm='uniform+online')
```

ask for 100000 samples:

```python
sage: from collections import defaultdict
sage: counter = defaultdict(Integer)
sage: n = 0
sage: def add_samples(i):
    ....:     global counter, n
    ....:     for _ in range(i):
    ....:         counter[D()] += 1
    ....:     n += 1
sage: add_samples(100000)
```

and check if the proportions fit:

```python
sage: expected = lambda x, y: (exp(-x^2/(2*sigma^2))/exp(-y^2/(2*sigma^2)).n())
sage: observed = lambda x, y: float(counter[x])/counter[y]
sage: while not all(v in counter for v in (0, 1, -100)): add_samples(10000)
sage: while abs(expected(0, 1) - observed(0, 1)) > 2e-1: add_samples(10000)
sage: while abs(expected(0, -100) - observed(0, -100)) > 2e-1: add_samples(10000)
```

We construct a sampler with $c \% 1 = 0$:

```python
sage: from sage.stats.distributions.discrete_gaussian_integer import DiscreteGaussianDistributionIntegerSampler
sage: sigma = 3
sage: D = DiscreteGaussianDistributionIntegerSampler(sigma=sigma, c=1/2)
sage: s = 0
sage: n = 0
sage: def add_samples(i):
    ....:     global s, n
    ....:     for _ in range(i):
    ....:         s += D()
    ....:     n += 1
sage: add_samples(100000)
sage: while abs(float(s)/n - 0.5) > 5e-2: add_samples(10000)
```

REFERENCES:

- [DDLL2013]
class sage.stats.distributions.discrete_gaussian_integer.DiscreteGaussianDistributionIntegerSampler

Bases: SageObject

A Discrete Gaussian Sampler using rejection sampling.

__init__(sigma, c=0, tau=6, algorithm=None, precision='mp')

Construct a new sampler for a discrete Gaussian distribution.

INPUT:

• sigma - samples \(x\) are accepted with probability proportional to \(\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\)

• c - the mean of the distribution. The value of \(c\) does not have to be an integer. However, some algorithms only support integer-valued \(c\) (default: 0)

• tau - samples outside the range \((-\tau c, \ldots, \tau c\)) are considered to have probability zero. This bound applies to algorithms which sample from the uniform distribution (default: 6)

• algorithm - see list below (default: "uniform+table" for \(t\) bounded by \(\text{DiscreteGaussianDistributionIntegerSampler.table_cutoff}\) and "uniform+online" for bigger)

• precision - either "mp" for multi-precision where the actual precision used is taken from sigma or "dp" for double precision. In the latter case results are not reproducible. (default: "mp")

ALGORITHMS:

• "uniform+table" - classical rejection sampling, sampling from the uniform distribution and accepted with probability proportional to \(\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\) where \(\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\) is precomputed and stored in a table. Any real-valued \(c\) is supported.

• "uniform+logtable" - samples are drawn from a uniform distribution and accepted with probability proportional to \(\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\) where \(\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\) is computed using logarithmically many calls to Bernoulli distributions. See [DDLL2013] for details. Only integer-valued \(c\) are supported.

• "uniform+online" - samples are drawn from a uniform distribution and accepted with probability proportional to \(\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\) where \(\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\) is computed in each invocation. Typically this is very slow. See [DDLL2013] for details. Any real-valued \(c\) is accepted.

• "sigma2+logtable" - samples are drawn from an easily samplable distribution with \(\sigma = k_2\) with \(k_2 = \sqrt{\frac{1}{2\log 2}}\) and accepted with probability proportional to \(\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\) where \(\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\) is computed using logarithmically many calls to Bernoulli distributions (but no calls to exp). See [DDLL2013] for details. Note that this sampler adjusts \(\sigma\) to match \(k_2\) for some integer \(k\). Only integer-valued \(c\) are supported.

EXAMPLES:

sage: from sage.stats.distributions.discrete_gaussian_integer import...
 sage: DiscreteGaussianDistributionIntegerSampler(3.0, algorithm=
 "uniform+online")
 Discrete Gaussian sampler over the Integers with sigma = 3.000000 and c = 0.
 sage: DiscreteGaussianDistributionIntegerSampler(3.0, algorithm="uniform+table"
 )
 Discrete Gaussian sampler over the Integers with sigma = 3.000000 and c = 0.
 sage: DiscreteGaussianDistributionIntegerSampler(3.0, algorithm=
 "uniform+logtable")
 (continues on next page)
Discrete Gaussian sampler over the Integers with sigma = 3.000000 and c = 0.000000

Note that "sigma2+logtable" adjusts:

```
sage: DiscreteGaussianDistributionIntegerSampler(3.0, algorithm="sigma2+logtable")
```

Discrete Gaussian sampler over the Integers with sigma = 3.397287 and c = 0.000000

```
__call__()
```

Return a new sample.

EXAMPLES:

```
sage: from sage.stats.distributions.discrete_gaussian_integer import DiscreteGaussianDistributionIntegerSampler
sage: DiscreteGaussianDistributionIntegerSampler(3.0, algorithm="uniform+online")()
# random -3
sage: DiscreteGaussianDistributionIntegerSampler(3.0, algorithm="uniform+table")()
# random
```

```
algorithm
csigma
table_cutoff = 1000000tau
```

Chapter 7. Discrete Gaussian Samplers over the Integers
This class realizes oracles which returns polynomials in \( \mathbb{Z}[x] \) where each coefficient is sampled independently with a probability proportional to \( \exp\left(-\frac{(x - c)^2}{2^s}\right) \).

**AUTHORS:**
- Martin Albrecht, Robert Fitzpatrick, Daniel Cabracas, Florian Göpfert, Michael Schneider: initial version

**EXAMPLES:**

```python
sage: from sage.stats.distributions.discrete_gaussian_polynomial import DiscreteGaussianDistributionPolynomialSampler
sage: sigma = 3.0; n = 1000
sage: l = [DiscreteGaussianDistributionPolynomialSampler(ZZ['x'], 64, sigma)() for _ in range(n)]

sage: from numpy import mean

sage: mean(l), sqrt(64)*sigma  # abs tol 5e-1
(24.0, 24.0)
```

---

**class** `sage.stats.distributions.discrete_gaussian_polynomial.DiscreteGaussianDistributionPolynomialSampler`

**Bases:** `SageObject`

Discrete Gaussian sampler for polynomials.

**EXAMPLES:**

```python
sage: from sage.stats.distributions.discrete_gaussian_polynomial import DiscreteGaussianDistributionPolynomialSampler
sage: p = DiscreteGaussianDistributionPolynomialSampler(ZZ['x'], 8, 3.0)()

sage: p.parent()
Univariate Polynomial Ring in x over Integer Ring

sage: p.degree() < 8
True

sage: gs = DiscreteGaussianDistributionPolynomialSampler(ZZ['x'], 8, 3.0)

sage: [gs() for _ in range(3)]  # random
[4*x^7 + 4*x^6 - 4*x^5 + 2*x^4 + x^3 - 4*x + 7, -5*x^6 + 4*x^5 - 3*x^3 + 4*x^2 + x, 2*x^7 + 2*x^6 + 2*x^5 - x^4 - 2*x^2 + 3*x + 1]
```

**__init__** \((P, n, sigma)\)
Construct a sampler for univariate polynomials of degree \( n-1 \) where coefficients are drawn independently with standard deviation \( \sigma \).

**INPUT:**

- \( P \) - a univariate polynomial ring over the Integers
- \( n \) - number of coefficients to be sampled
- \( \sigma \) - coefficients \( x \) are accepted with probability proportional to \( \exp(-x^2/(2\sigma^2)) \). If an object of type `sage.stats.distributions.discrete_gaussian_integer.DiscreteGaussianDistributionIntegerSampler` is passed, then this sampler is used to sample coefficients.

**EXAMPLES:**

```python
sage: from sage.stats.distributions.discrete_gaussian_polynomial import DiscreteGaussianDistributionPolynomialSampler
sage: p = DiscreteGaussianDistributionPolynomialSampler(ZZ['x'], 8, 3.0)()
sage: p.parent()
Univariate Polynomial Ring in x over Integer Ring
sage: p.degree() < 8
True
sage: gs = DiscreteGaussianDistributionPolynomialSampler(ZZ['x'], 8, 3.0)
[sage: [gs() for _ in range(3)]  # random
[4*x^7 + 4*x^6 - 4*x^5 + 2*x^4 + x^3 - 4*x + 7, -5*x^6 + 4*x^5 - 3*x^3 + 4*x^2 + x, 2*x^7 + 2*x^6 + 2*x^5 - x^4 - 2*x^2 + 3*x + 1]
```

**__call__**

Return a new sample.

**EXAMPLES:**

```python
sage: from sage.stats.distributions.discrete_gaussian_polynomial import DiscreteGaussianDistributionPolynomialSampler
sage: sampler = DiscreteGaussianDistributionPolynomialSampler(ZZ['x'], 8, 12.0)
sage: sampler().parent()
Univariate Polynomial Ring in x over Integer Ring
sage: sampler().degree() <= 7
True
```
This file implements oracles which return samples from a lattice following a discrete Gaussian distribution. That is, if \( \sigma \) is big enough relative to the provided basis, then vectors are returned with a probability proportional to \( \exp(-|x - c|^2/(2^2)) \). More precisely, lattice vectors in \( x \in \mathbb{Z}^n \) are returned with probability:

\[
\exp(-|x - c|^2/(2^2))/\left(\sum_{x \in \mathbb{Z}^n} \exp(-|x|^2/(2^2))\right)
\]

AUTHORS:

EXAMPLES:

```python
sage: from sage.stats.distributions.discrete_gaussian_lattice import DiscreteGaussianDistributionLatticeSampler
sage: D = DiscreteGaussianDistributionLatticeSampler(ZZ^10, 3.0)

sage: D(), D(), D()
# random
((3, 0, -5, 0, -1, -3, 3, -7, 2), (4, 0, 1, -2, -4, -4, 4, 0, 1, -4), (-3, 0, 4, 5, 0, 1, 3, 2, 0, -1))

sage: a = D()
sage: a.parent()
Ambient free module of rank 10 over the principal ideal domain Integer Ring
```

GPV sampler for Discrete Gaussians over Lattices.

EXAMPLES:

```python
sage: from sage.stats.distributions.discrete_gaussian_lattice import DiscreteGaussianDistributionLatticeSampler
sage: D = DiscreteGaussianDistributionLatticeSampler(ZZ^10, 3.0); D
Discrete Gaussian sampler with \( \sigma = 3.000000 \), \( c = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0) \) over lattice with basis

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]
```
We plot a histogram:

```python
sage: from sage.stats.distributions.discrete_gaussian_lattice import DiscreteGaussianDistributionLatticeSampler
sage: import warnings
sage: warnings.simplefilter('ignore', UserWarning)
sage: D = DiscreteGaussianDistributionLatticeSampler(identity_matrix(2), 3.0)
sage: S = [D() for _ in range(2^12)]
```
....:    m += 1
sage: v = vector(ZZ, n, (-3, -3))
sage: v.set_immutable()
sage: while v not in counter:
....:      add_samples(1000)
sage: while abs(m*f(v)*1.0/c/counter[v] - 1.0) >= 0.1:
      # needs sage.symbolic
      ....:      add_samples(1000)

sage: v = vector(ZZ, n, (0, 0))
sage: v.set_immutable()
sage: while v not in counter:
....:      add_samples(1000)
sage: while abs(m*f(v)*1.0/c/counter[v] - 1.0) >= 0.1:
      # needs sage.symbolic
      ....:      add_samples(1000)

sage: from sage.stats.distributions.discrete_gaussian_lattice import...
   # needs sage.symbolic
sage: qf = QuadraticForm(matrix(3, [2, 1, 1, 1, 2, 1, 1, 1, 2]))
sage: D = DiscreteGaussianDistributionLatticeSampler(qf, 3.0); D
Discrete Gaussian sampler with \( \sigma = 3.000000 \), \( c = (0, 0, 0) \) over lattice with
basis

\[ \begin{bmatrix} 2 & 1 & 1 \\
1 & 2 & 1 \\
1 & 1 & 2 \end{bmatrix} \]
sage: D().parent() is D.c.parent()
# needs sage.symbolic
True

__call__()
Return a new sample.

EXAMPLES:

sage: from sage.stats.distributions.discrete_gaussian_lattice import...
   # needs sage.symbolic
sage: D = DiscreteGaussianDistributionLatticeSampler(ZZ^3, 3.0, c=(1,0,0))
sage: L = [D() for _ in range(2^12)]
sage: mean_L = sum(L) / len(L)
sage: norm(mean_L.n() - D.c) < 0.25
True

sage: D = DiscreteGaussianDistributionLatticeSampler(ZZ^3, 3.0, c=(1/2,0,0))
sage: L = [D() for _ in range(2^12)]  # long time
sage: mean_L = sum(L) / len(L)  # long time
sage: norm(mean_L.n() - D.c) < 0.25  # long time
True

property c
Center \( c \).

Examples from this sampler will be centered at \( c \).

EXAMPLES:
```python
sage: from sage.stats.distributions.discrete_gaussian_lattice import DiscreteGaussianDistributionLatticeSampler
sage: D = DiscreteGaussianDistributionLatticeSampler(ZZ^3, 3.0, c=(1,0,0)); D
Discrete Gaussian sampler with σ = 3.000000, c=(1, 0, 0) over lattice with
basis
[1 0 0]
[0 1 0]
[0 0 1]
```

```python
sage: D.c
(1, 0, 0)
```

**static compute_precision**(precision, sigma)

Compute precision to use.

**INPUT:**

- precision - an integer > 53 nor None.
- sigma - if precision is None then the precision of sigma is used.

**EXAMPLES:**

```python
sage: from sage.stats.distributions.discrete_gaussian_lattice import DiscreteGaussianDistributionLatticeSampler
sage: DiscreteGaussianDistributionLatticeSampler.compute_precision(100, RR(3))
100
sage: DiscreteGaussianDistributionLatticeSampler.compute_precision(100, RealField(200)(3))
100
sage: DiscreteGaussianDistributionLatticeSampler.compute_precision(100, 3)
100
sage: DiscreteGaussianDistributionLatticeSampler.compute_precision(None, RR(3))
53
sage: DiscreteGaussianDistributionLatticeSampler.compute_precision(None, RealField(200)(3))
200
sage: DiscreteGaussianDistributionLatticeSampler.compute_precision(None, 3)
53
```

**property sigma**

Gaussian parameter .

Samples from this sampler will have expected norm \( \sqrt{n} \) where \( n \) is the dimension of the lattice.

**EXAMPLES:**

```python
sage: from sage.stats.distributions.discrete_gaussian_lattice import DiscreteGaussianDistributionLatticeSampler
sage: D = DiscreteGaussianDistributionLatticeSampler(ZZ^3, 3.0, c=(1,0,0))
sage: D.sigma
3.00000000000000
```
CHAPTER
TEN

T-TEST USING R

\texttt{sage.stats.r.ttest(x, y, conf\_level=0.95, **kw)}

T-Test using R

INPUT:

- \(x, y\) – vectors of same length
- \(\text{conf\_level}\) – confidence level of the interval, \([0,1)\) in percent

OUTPUT:

Tuple: (p-value, R return object)

EXAMPLES:

\begin{verbatim}
sage: a, b = ttest([1,2,3,4,5],[1,2,3,3.5,5.121]); a
# abs tol 1e-12 # optional
\end{verbatim}

\begin{verbatim}
→ rpy2
0.9410263720274274
\end{verbatim}
INDICES AND TABLES

• Index
• Module Index
• Search Page
S
sage.stats.basic_stats, 1
sage.stats.distributions.discrete_gaussian_integer, 35
sage.stats.distributions.discrete_gaussian_lattice, 41
sage.stats.distributions.discrete_gaussian_polynomial, 39
sage.stats.hmm.chmm, 19
sage.stats.hmm.distributions, 27
sage.stats.hmm.hmm, 11
sage.stats.hmm.util, 33
sage.stats.intlist, 7
sage.stats.r, 45
Non-alphabetical

__call__() (sage.stats.distributions.discrete_gaussian_integer.DiscreteGaussianDistributionIntegerSampler method), 38
__call__() (sage.stats.distributions.discrete_gaussian_lattice.DiscreteGaussianDistributionLatticeSampler method), 43
__call__() (sage.stats.distributions.discrete_gaussian_polynomial.DiscreteGaussianDistributionPolynomialSampler method), 40
__init__() (sage.stats.distributions.discrete_gaussian_integer.DiscreteGaussianDistributionIntegerSampler method), 37
__init__() (sage.stats.distributions.discrete_gaussian_lattice.DiscreteGaussianDistributionLatticeSampler method), 42
__init__() (sage.stats.distributions.discrete_gaussian_polynomial.DiscreteGaussianDistributionPolynomialSampler method), 39

A
algorithm (sage.stats.distributions.discrete_gaussian_integer.DiscreteGaussianDistributionIntegerSampler attribute), 38

B
baum_welch() (sage.stats.hmm.chmm.GaussianHiddenMarkovModel method), 20
baum_welch() (sage.stats.hmm.chmm.GaussianMixtureHiddenMarkovModel method), 24
baum_welch() (sage.stats.hmm.hmm.DiscreteHiddenMarkovModel method), 12

C
c (sage.stats.distributions.discrete_gaussian_integer.DiscreteGaussianDistributionIntegerSampler attribute), 38
c (sage.stats.distributions.discrete_gaussian_lattice.DiscreteGaussianDistributionLatticeSampler property), 43

D
DiscreteDistribution (class in sage.stats.hmm.distributions), 27
DiscreteGaussianDistributionIntegerSampler (class in sage.stats.distributions.discrete_gaussian_integer), 36
DiscreteGaussianDistributionLatticeSampler (class in sage.stats.distributions.discrete_gaussian_lattice), 41
DiscreteGaussianDistributionPolynomialSampler (class in sage.stats.distributions.discrete_gaussian_polynomial), 39
DiscreteHiddenMarkovModel (class in sage.stats.hmm.hmm), 11
Distribution (class in sage.stats.hmm.distributions), 27

E
emission_matrix() (sage.stats.hmm.hmm.DiscreteHiddenMarkovModel method), 13
emission_parameters() (sage.stats.hmm.chmm.GaussianHiddenMarkovModel method), 22
emission_parameters() (sage.stats.hmm.chmm.GaussianMixtureHiddenMarkovModel method), 25

F
fix() (sage.stats.hmm.distributions.GaussianMixtureDistribution method), 28

G
GaussianDistribution (class in sage.stats.hmm.distributions), 28
GaussianHiddenMarkovModel (class in sage.stats.hmm.hmm), 19
GaussianMixtureDistribution (class in sage.stats.hmm.distributions), 28

compute_precision() (sage.stats.distributions.discrete_gaussian_lattice.DiscreteGaussianDistributionLatticeSampler static method), 44
<table>
<thead>
<tr>
<th>Class</th>
<th>Module</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaussianMixtureHiddenMarkovModel</td>
<td>sage.stats.hmm.chmm</td>
<td>24</td>
</tr>
<tr>
<td>generate_sequence()</td>
<td>sage.stats.hmm.chmm.GaussianHiddenMarkovModel method</td>
<td>22</td>
</tr>
<tr>
<td>generate_sequence()</td>
<td>sage.stats.hmm.hmm.DiscreteHiddenMarkovModel method</td>
<td>14</td>
</tr>
<tr>
<td>graph()</td>
<td>sage.stats.hmm.hmm.HiddenMarkovModel method</td>
<td>16</td>
</tr>
<tr>
<td>HiddenMarkovModel</td>
<td>sage.stats.hmm.hmm</td>
<td>16</td>
</tr>
<tr>
<td>HMM_Util</td>
<td>sage.stats.hmm.util</td>
<td>33</td>
</tr>
<tr>
<td>initial_probabilities()</td>
<td>sage.stats.hmm.hmm.HiddenMarkovModel method</td>
<td>16</td>
</tr>
<tr>
<td>initial_probs_to_TimeSeries()</td>
<td>sage.stats.hmm.util.HMM_Util method</td>
<td>33</td>
</tr>
<tr>
<td>IntList</td>
<td>sage.stats.intlist</td>
<td>7</td>
</tr>
<tr>
<td>is_fixed()</td>
<td>sage.stats.hmm.distributions.GaussianMixtureDistribution method</td>
<td>29</td>
</tr>
<tr>
<td>list()</td>
<td>sage.stats.intlist.IntList method</td>
<td>7</td>
</tr>
<tr>
<td>log_likelihood()</td>
<td>sage.stats.hmm.chmm.GaussianHiddenMarkovModel method</td>
<td>23</td>
</tr>
<tr>
<td>log_likelihood()</td>
<td>sage.stats.hmm.hmm.DiscreteHiddenMarkovModel method</td>
<td>14</td>
</tr>
<tr>
<td>max()</td>
<td>sage.stats.intlist.IntList method</td>
<td>7</td>
</tr>
<tr>
<td>mean()</td>
<td>in module sage.stats.basic_stats</td>
<td>1</td>
</tr>
<tr>
<td>median()</td>
<td>in module sage.stats.basic_stats</td>
<td>2</td>
</tr>
<tr>
<td>min()</td>
<td>sage.stats.intlist.IntList method</td>
<td>8</td>
</tr>
<tr>
<td>mode()</td>
<td>in module sage.stats.basic_stats</td>
<td>2</td>
</tr>
<tr>
<td>module</td>
<td>sage.stats.basic_stats</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>sage.stats.distributions.discrete_gaussian_integer</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>sage.stats.distributions.discrete_gaussian_lattice</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>sage.stats.distributions.discrete_gaussian_polynomial</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>sage.stats.hmm.chmm</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>sage.stats.hmm.distributions</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>sage.stats.hmm.hmm</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>sage.stats.hmm.util</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>sage.stats.intlist</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>sage.stats.r</td>
<td>45</td>
</tr>
<tr>
<td>moving_average()</td>
<td>in module sage.stats.basic_stats</td>
<td>3</td>
</tr>
<tr>
<td>normalize_probability_TimeSeries()</td>
<td>sage.stats.hmm.util.HMM_Util method</td>
<td>33</td>
</tr>
<tr>
<td>plot()</td>
<td>sage.stats.hmm.distributions.Distribution method</td>
<td>27</td>
</tr>
<tr>
<td>plot()</td>
<td>sage.stats.intlist.IntList method</td>
<td>8</td>
</tr>
<tr>
<td>plot_histogram()</td>
<td>sage.stats.intlist.IntList method</td>
<td>8</td>
</tr>
<tr>
<td>prob()</td>
<td>sage.stats.hmm.distributions.Distribution method</td>
<td>27</td>
</tr>
<tr>
<td>prob()</td>
<td>sage.stats.hmm.distributions.GaussianMixtureDistribution method</td>
<td>29</td>
</tr>
<tr>
<td>prob_m()</td>
<td>sage.stats.hmm.distributions.GaussianMixtureDistribution method</td>
<td>29</td>
</tr>
<tr>
<td>prod()</td>
<td>sage.stats.intlist.IntList method</td>
<td>8</td>
</tr>
<tr>
<td>list()</td>
<td>sage.stats.basic_stats</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>sage.stats.distributions.discrete_gaussian_integer</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>sage.stats.distributions.discrete_gaussian_lattice</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>sage.stats.distributions.discrete_gaussian_polynomial</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>sage.stats.hmm.chmm</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>sage.stats.hmm.distributions</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>sage.stats.hmm.hmm</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>sage.stats.hmm.util</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>sage.stats.intlist</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>sage.stats.r</td>
<td>45</td>
</tr>
<tr>
<td>sample()</td>
<td>sage.stats.hmm.distributions.Distribution method</td>
<td>28</td>
</tr>
<tr>
<td>sample()</td>
<td>sage.stats.hmm.distributions.GaussianMixtureDistribution method</td>
<td>30</td>
</tr>
<tr>
<td>sample()</td>
<td>sage.stats.hmm.hmm.HiddenMarkovModel method</td>
<td>17</td>
</tr>
<tr>
<td>sigma</td>
<td>sage.stats.distributions.discrete_gaussian_integer.DiscreteGaussianDistributionIntegerSampler attribute</td>
<td>38</td>
</tr>
<tr>
<td>sigma</td>
<td>sage.stats.distributions.discrete_gaussian_lattice.DiscreteGaussianDistributionLatticeSampler property</td>
<td>44</td>
</tr>
</tbody>
</table>

52 Index
state_matrix_to_TimeSeries()
   (sage.stats.hmm.util.HMM_Util method), 34
std() (in module sage.stats.basic_stats), 4
sum() (sage.stats.intlist.IntList method), 9

T
table_cutoff (sage.stats.distributions.discrete_gaussian_integer.DiscreteGaussianDistributionIntegerSampler attribute), 38
tau (sage.stats.distributions.discrete_gaussian_integer.DiscreteGaussianDistributionIntegerSampler attribute), 38
time_series() (sage.stats.intlist.IntList method), 9
transition_matrix() (sage.stats.hmm.HiddenMarkovModel method), 18
ttest() (in module sage.stats.r), 45

U
unfix() (sage.stats.hmm.distributions.GaussianMixtureDistribution method), 30
unpickle_discrete_hmm_v0() (in module sage.stats.hmm), 18
unpickle_discrete_hmm_v1() (in module sage.stats.hmm), 18
unpickle_gaussian_hmm_v0() (in module sage.stats.hmm.chmm), 26
unpickle_gaussian_hmm_v1() (in module sage.stats.hmm.chmm), 26
unpickle_gaussian_mixture_distribution_v1() (in module sage.stats.hmm.distributions), 31
unpickle_gaussian_mixture_hmm_v1() (in module sage.stats.hmm.chmm), 26
unpickle_intlist_v1() (in module sage.stats.intlist), 9

V
variance() (in module sage.stats.basic_stats), 5
viterbi() (sage.stats.hmm.chmm.GaussianHiddenMarkovModel method), 23
viterbi() (sage.stats.hmm.chmm.DiscreteHiddenMarkovModel method), 15