Statistics for Astronomers: Lecture 3, 2020.09.28

Prof. Sundar Srinivasan

IRyA/UNAM

Review

Random variables.
Probability distributions: discrete (PMF) and continuous (PDF).
Populations vs. samples.
The cumulative distribution function and its inverse (the quantile function).
Expectation value: mean and variance.

Variance, contd.

X draws values from a distribution $p_X(x)$. A sample consisting of multiple draws for X will have points distributed around $\mathbb{E}[X]$.

Recall: $\mathrm{Var}[X] \equiv \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$; the variance is non-negative by definition.

Variance, contd.

X draws values from a distribution $p_X(x)$. A sample consisting of multiple draws for X will have points distributed around $\mathbb{E}[X]$.

Recall: $\mathrm{Var}[X] \equiv \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$; the variance is non-negative by definition.

$\mathbb{E}$ is a linear operator. What about Var?

For any constant α, $\mathbb{E}[\alpha] = \alpha$. What is $\mathrm{Var}[\alpha]$?

Variance, contd.

X draws values from a distribution $p_X(x)$. A sample consisting of multiple draws for X will have points distributed around $\mathbb{E}[X]$.

Recall: $\mathrm{Var}[X] \equiv \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$; the variance is non-negative by definition.

$\mathbb{E}$ is a linear operator. What about Var?

For any constant α, $\mathbb{E}[\alpha] = \alpha$. What is $\mathrm{Var}[\alpha]$?

Effect of shifting: $\mathbb{E}[X + \alpha] = \mathbb{E}[X] + \alpha$. What is $\mathrm{Var}[X + \alpha]$?

Variance, contd.

X draws values from a distribution $p_X(x)$. A sample consisting of multiple draws for X will have points distributed around $\mathbb{E}[X]$.

Recall: $\mathrm{Var}[X] \equiv \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$; the variance is non-negative by definition.

$\mathbb{E}$ is a linear operator. What about Var?

- For any constant α, $\mathbb{E}[\alpha] = \alpha$. What is $\mathrm{Var}[\alpha]$?
- Effect of shifting: $\mathbb{E}[X + \alpha] = \mathbb{E}[X] + \alpha$. What is $\mathrm{Var}[X + \alpha]$?
- Effect of scaling: $\mathbb{E}[\alpha X] = \alpha\ \mathbb{E}[X]$. What is $\mathrm{Var}[\alpha X]$?

Variance, contd.

X draws values from a distribution $p_X(x)$. A sample consisting of multiple draws for X will have points distributed around $\mathbb{E}[X]$.

Recall: $\mathrm{Var}[X] \equiv \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$; the variance is non-negative by definition.

$\mathbb{E}$ is a linear operator. What about Var?

For any constant α, $\mathbb{E}[\alpha] = \alpha$. What is $\mathrm{Var}[\alpha]$?

Effect of shifting: $\mathbb{E}[X + \alpha] = \mathbb{E}[X] + \alpha$. What is $\mathrm{Var}[X + \alpha]$?

Effect of scaling: $\mathbb{E}[\alpha X] = \alpha\, \mathbb{E}[X]$. What is $\mathrm{Var}[\alpha X]$?

Practice:

```
import numpy as np
x = np.random.uniform(size = 20) #sample of 20 uniform random numbers
x.mean()
x.var()
(x+3).mean() #should be 3 + x.mean()
(x+3).var() #should be x.var()
```

Variance, contd.

X draws values from a distribution $p_X(x)$. A sample consisting of multiple draws for X will have points distributed around $\mathbb{E}[X]$.

Recall: $\mathrm{Var}[X] \equiv \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$; the variance is non-negative by definition.

$\mathbb{E}$ is a linear operator. What about Var?

For any constant α, $\mathbb{E}[\alpha] = \alpha$. What is $\mathrm{Var}[\alpha]$?

Effect of shifting: $\mathbb{E}[X + \alpha] = \mathbb{E}[X] + \alpha$. What is $\mathrm{Var}[X + \alpha]$?

Effect of scaling: $\mathbb{E}[\alpha X] = \alpha\ \mathbb{E}[X]$. What is $\mathrm{Var}[\alpha X]$?

Practice:

```
import numpy as np
x = np.random.uniform(size = 20) #sample of 20 uniform random numbers
x.mean()
x.var()
(x+3).mean() #should be 3 + x.mean()
(x+3).var() #should be x.var()
```

Linear combination: $\mathbb{E}[\alpha X + \beta Y] = \alpha\ \mathbb{E}[X] + \beta\ \mathbb{E}[Y]$. What is $\mathrm{Var}[\alpha X + \beta Y]$?
(use the definition for variance above)

Covariance

$$\mathrm{Var}[\alpha X + \beta Y] = \alpha^2\ \mathrm{Var}[X] + \beta^2\ \mathrm{Var}[Y] + 2\alpha\beta\ \overbrace{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}^{\text{interpretation?}} \tag{1}$$

Covariance

$$\mathrm{Var}[\alpha X + \beta Y] = \alpha^2\ \mathrm{Var}[X] + \beta^2\ \mathrm{Var}[Y] + 2\alpha\beta\ \overbrace{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}^{\text{interpretation?}} \tag{1}$$

If $X = Y$, then this term becomes $\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}[X]$.
If $X \perp Y$, then the Y deviations and X deviations are independent, and this term vanishes.
$\implies$ the term quantifies a variance-like dependence between X and Y.

Covariance

$$\mathrm{Var}[\alpha X + \beta Y] = \alpha^2\ \mathrm{Var}[X] + \beta^2\ \mathrm{Var}[Y] + 2\alpha\beta\ \overbrace{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}^{\text{interpretation?}} \tag{1}$$

If $X = Y$, then this term becomes $\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}[X]$.
If $X \perp Y$, then the Y deviations and X deviations are independent, and this term vanishes.
$\implies$ the term quantifies a variance-like dependence between X and Y.
If X, Y not independent::
$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) > 0$ if both deviations are in the same direction,
$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) < 0$ if deviations are in opposite directions.

Covariance

$$\mathrm{Var}[\alpha X + \beta Y] = \alpha^2\ \mathrm{Var}[X] + \beta^2\ \mathrm{Var}[Y] + 2\alpha\beta\ \overbrace{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}^{\text{interpretation?}} \tag{1}$$

If $X = Y$, then this term becomes $\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}[X]$.
If $X \perp Y$, then the Y deviations and X deviations are independent, and this term vanishes.
$\implies$ the term quantifies a variance-like dependence between X and Y.
If X, Y not independent::
$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) > 0$ if both deviations are in the same direction,
$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) < 0$ if deviations are in opposite directions.

The third term in Eq (1) is the mean of the products of X and Y deviations.

Covariance

$$\mathrm{Var}[\alpha X + \beta Y] = \alpha^2\ \mathrm{Var}[X] + \beta^2\ \mathrm{Var}[Y] + 2\alpha\beta\ \overbrace{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}^{\text{interpretation?}} \quad (1)$$

If $X = Y$, then this term becomes $\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}[X]$.
If $X \perp Y$, then the Y deviations and X deviations are independent, and this term vanishes.
$\implies$ the term quantifies a variance-like dependence between X and Y.
If X, Y not independent::
$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) > 0$ if both deviations are in the same direction,
$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) < 0$ if deviations are in opposite directions.

The third term in Eq (1) is the mean of the products of X and Y deviations.

Definition (Covariance)

The covariance is a measure of joint variability of two random variables:
$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. By definition, $\mathrm{Cov}(X, X) = \mathrm{Var}[X]$.

Covariance

$$\mathrm{Var}[\alpha X + \beta Y] = \alpha^2\ \mathrm{Var}[X] + \beta^2\ \mathrm{Var}[Y] + 2\alpha\beta\ \overbrace{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}^{\text{interpretation?}} \qquad (1)$$

If $X = Y$, then this term becomes $\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}[X]$.
If $X \perp Y$, then the Y deviations and X deviations are independent, and this term vanishes.
$\implies$ the term quantifies a variance-like dependence between X and Y.
If X, Y not independent::
$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) > 0$ if both deviations are in the same direction,
$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) < 0$ if deviations are in opposite directions.

The third term in Eq (1) is the mean of the products of X and Y deviations.

Definition (Covariance)

The covariance is a measure of joint variability of two random variables:
$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. By definition, $\mathrm{Cov}(X, X) = \mathrm{Var}[X]$.

Therefore, $\mathrm{Var}[\alpha X + \beta Y] = \alpha^2\ \mathrm{Var}[X] + \beta^2\ \mathrm{Var}[Y] + 2\alpha\beta\ \mathrm{Cov}(X, Y)$.

If the two variables are uncorrelated, then the third term vanishes.

Covariance matrix

Pairwise comparison of N random variables $\longrightarrow$ $N \times N$ covariances $\longrightarrow$ covariance matrix.
If $\mathbf{X}$ is a $1 \times N$ random vector (each component is a random variable), then

$$\Sigma_{\mathbf{XX}} = \mathrm{Cov}(\mathbf{X}, \mathbf{X}), \text{ such that } \Sigma_{ij} = \mathrm{Cov}(X_i, X_j). \qquad \text{(diagonal elements: } \mathrm{Var}[X_i])$$

Covariance matrix

Pairwise comparison of N random variables $\longrightarrow$ $N \times N$ covariances $\longrightarrow$ covariance matrix.
If $\mathbf{X}$ is a $1 \times N$ random vector (each component is a random variable), then

$$\Sigma_{\mathbf{XX}} = \mathrm{Cov}(\mathbf{X}, \mathbf{X}), \text{ such that } \Sigma_{ij} = \mathrm{Cov}(X_i, X_j). \qquad \text{(diagonal elements: } \mathrm{Var}[X_i])$$

Practice:

```
import numpy as np
np.random.seed(0)                   #for reproducibility
x = np.random.uniform(size = 20)  #sample of 20 uniform random numbers
y = x**2; z = 1/x                  #positively- and anti-correlated with x
X = np.array([x, y, z])            #20 samples of a 4-element random vector
print(np.cov(X))
```

What are the values of $\mathrm{Cov}(x, y)$ and $\mathrm{Cov}(x, z)$?

Covariance matrix

Pairwise comparison of N random variables $\longrightarrow$ $N \times N$ covariances $\longrightarrow$ covariance matrix.
If $\mathbf{X}$ is a $1 \times N$ random vector (each component is a random variable), then

$$\Sigma_{\mathbf{XX}} = \mathrm{Cov}(\mathbf{X}, \mathbf{X}), \text{ such that } \Sigma_{ij} = \mathrm{Cov}(X_i, X_j). \qquad \text{(diagonal elements: } \mathrm{Var}[X_i])$$

Practice:

```
import numpy as np
np.random.seed(0)                  #for reproducibility
x = np.random.uniform(size = 20)  #sample of 20 uniform random numbers
y = x**2; z = 1/x                  #positively- and anti-correlated with x
X = np.array([x, y, z])           #20 samples of a 4-element random vector
print(np.cov(X))
```

What are the values of $\mathrm{Cov}(x, y)$ and $\mathrm{Cov}(x, z)$?

```
np.random.seed(0)                  #for reproducibility
x = np.random.uniform(size = 20)  #sample of 20 uniform random numbers
np.random.seed(1)                  #for reproducibility
w = np.random.uniform(size = 20)  #no correlation expected with x
print(np.cov(np.array([x, w])))
```

What is the value of $\mathrm{Cov}(x, w)$? How does it change if you increase size to 1000?

Correlation coefficient

While the sign of $\mathrm{Cov}(X, Y)$ is useful, the magnitude isn't.

Cov is **not** scale-invariant: $\mathrm{Cov}(\alpha X, \beta Y) = \alpha\beta\ \mathrm{Cov}(X, Y)$.

Correlation coefficient

While the sign of $\mathrm{Cov}(X, Y)$ is useful, the magnitude isn't.

Cov is **not** scale-invariant: $\mathrm{Cov}(\alpha X, \beta Y) = \alpha\beta\ \mathrm{Cov}(X, Y)$.

Definition ((Pearson's) Correlation coefficient)

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$\rho_{XX} = 1$ (by definition). "Perfect correlation".
$\rho_{XY} = -1 \implies$ "perfect anticorrelation".
ρ_{XY} probes the strength and direction of a linear relationship.

To explore more general monotonic relations, see rank correlation (*e.g.*, Spearman's r).

Correlation coefficient

While the sign of $\mathrm{Cov}(X, Y)$ is useful, the magnitude isn't.

Cov is **not** scale-invariant: $\mathrm{Cov}(\alpha X, \beta Y) = \alpha\beta\ \mathrm{Cov}(X, Y)$.

Definition ((Pearson's) Correlation coefficient)

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$\rho_{XX} = 1$ (by definition). "Perfect correlation".
$\rho_{XY} = -1 \implies$ "perfect anticorrelation".
ρ_{XY} probes the strength and direction of a linear relationship.

To explore more general monotonic relations, see rank correlation (*e.g.*, Spearman's r).

Correlation coefficient

While the sign of $\mathrm{Cov}(X, Y)$ is useful, the magnitude isn't.

Cov is **not** scale-invariant: $\mathrm{Cov}(\alpha X, \beta Y) = \alpha\beta\ \mathrm{Cov}(X, Y)$.

Definition ((Pearson's) Correlation coefficient)

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$\rho_{XX} = 1$ (by definition). "Perfect correlation".
$\rho_{XY} = -1 \Longrightarrow$ "perfect anticorrelation".
ρ_{XY} probes the strength and direction of a linear relationship.

To explore more general monotonic relations, see rank correlation (*e.g.*, Spearman's r).

Correlation coefficient

While the sign of $\mathrm{Cov}(X, Y)$ is useful, the magnitude isn't.

Cov is **not** scale-invariant: $\mathrm{Cov}(\alpha X, \beta Y) = \alpha\beta\ \mathrm{Cov}(X, Y)$.

Definition ((Pearson's) Correlation coefficient)

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{Var(X) Var(Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$\rho_{XX} = 1$ (by definition). "Perfect correlation".
$\rho_{XY} = -1 \implies$ "perfect anticorrelation".
ρ_{XY} probes the strength and direction of a linear relationship.

To explore more general monotonic relations, see rank correlation (*e.g.*, Spearman's r).

Practice:

```
import numpy as np
np.random.seed(0); x = np.random.uniform(size = 20); y = x**2; z = 1/x
np.random.seed(1); w = np.random.uniform(size = 20)
Sigma = np.cov(np.array([x, y, z, w]))
#The following should return the same value
print(Sigma[0, 1] / np.sqrt(Sigma[0, 0] * Sigma[1, 1]))
print(np.corrcoef(x, y = y)[0, 1])
```

Repeat for (x, z) and (x, w).

Uncorrelated random variables

Random variables X_i $(i = 1, 2, \cdots N)$ are uncorrelated if $\mathrm{Cov}(X_i, X_j) = \delta_{ij}\mathrm{Var}[X_i]$.

Uncorrelated random variables

Random variables X_i $(i = 1, 2, \cdots N)$ are uncorrelated if $\mathrm{Cov}(X_i, X_j) = \delta_{ij}\mathrm{Var}[X_i]$.

Bienaymé's Identity

The variance of the sum of N uncorrelated variables is the sum of their variances.
(corollary: standard deviations add in quadrature)

$$X_i \perp X_j \text{ for } (i \neq j) \implies$$
$$\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathrm{Var}[X_i]$$

Uncorrelated random variables

Random variables X_i $(i = 1, 2, \cdots N)$ are uncorrelated if $\mathrm{Cov}(X_i, X_j) = \delta_{ij}\mathrm{Var}[X_i]$.

Bienaymé's Identity

The variance of the sum of N uncorrelated variables is the sum of their variances.
(corollary: standard deviations add in quadrature)

$$X_i \perp X_j \text{ for } (i \neq j) \implies$$
$$\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathrm{Var}[X_i]$$

The mean of iid variables is drawn from the same distribution, but with $1/N^{\mathrm{th}}$ the variance.

$$X_i \sim p_X(\mu, \sigma^2) \implies \bar{X} \sim p_X(\mu, \sigma^2/N)$$

$$\mathrm{Var}[\bar{X}] = \frac{1}{N^2}\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}[X_i] = \frac{1}{N^2} N\ \mathrm{Var}[X] = \frac{\mathrm{Var}[X]}{N}$$

Uncorrelated random variables

Random variables X_i ($i = 1, 2, \cdots N$) are uncorrelated if $\mathrm{Cov}(X_i, X_j) = \delta_{ij}\mathrm{Var}[X_i]$.

Bienaymé's Identity

The variance of the sum of N uncorrelated variables is the sum of their variances.
(corollary: standard deviations add in quadrature)

$$X_i \perp X_j \text{ for } (i \neq j) \implies$$
$$\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathrm{Var}[X_i]$$

The mean of iid variables is drawn from the same distribution, but with $1/N^{\mathrm{th}}$ the variance.

$$X_i \sim p_X(\mu, \sigma^2) \implies \bar{X} \sim p_X(\mu, \sigma^2/N)$$

$$\mathrm{Var}[\bar{X}] = \frac{1}{N^2}\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}[X_i] = \frac{1}{N^2} N\ \mathrm{Var}[X] = \frac{\mathrm{Var}[X]}{N}$$

Uncorrelated random variables

Random variables X_i ($i = 1, 2, \cdots N$) are uncorrelated if $\mathrm{Cov}(X_i, X_j) = \delta_{ij}\mathrm{Var}[X_i]$.

Bienaymé's Identity

The variance of the sum of N uncorrelated variables is the sum of their variances.
(corollary: standard deviations add in quadrature)

$$X_i \perp X_j \text{ for } (i \neq j) \implies$$
$$\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathrm{Var}[X_i]$$

The mean of iid variables is drawn from the same distribution, but with $1/N^{\mathrm{th}}$ the variance.

$$X_i \sim p_X(\mu, \sigma^2) \implies \bar{X} \sim p_X(\mu, \sigma^2/N)$$

$$\mathrm{Var}[\bar{X}] = \frac{1}{N^2}\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}[X_i] = \frac{1}{N^2} N\ \mathrm{Var}[X] = \frac{\mathrm{Var}[X]}{N}$$

Verify:

```
import numpy as np
x=np.random.uniform(size=10)
print(x.mean());
print(x.var())
```

`x.var()`: variance of the 10 samples (observations) of `x`.

Variance of the mean of the 10 observations: `x.var()`/10.

Uncorrelated random variables

Random variables X_i ($i = 1, 2, \cdots N$) are uncorrelated if $\mathrm{Cov}(X_i, X_j) = \delta_{ij}\mathrm{Var}[X_i]$.

Bienaymé's Identity

The variance of the sum of N uncorrelated variables is the sum of their variances.
(corollary: standard deviations add in quadrature)

$$X_i \perp X_j \text{ for } (i \neq j) \Longrightarrow$$
$$\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathrm{Var}[X_i]$$

The mean of iid variables is drawn from the same distribution, but with $1/N^{\mathrm{th}}$ the variance.

$$X_i \sim p_X(\mu, \sigma^2) \Longrightarrow \bar{X} \sim p_X(\mu, \sigma^2/N)$$

$$\mathrm{Var}[\bar{X}] = \frac{1}{N^2}\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}[X_i] = \frac{1}{N^2} N\ \mathrm{Var}[X] = \frac{\mathrm{Var}[X]}{N}$$

Verify:

```
import numpy as np
x=np.random.uniform(size=10)
print(x.mean());
print(x.var())
```

`x.var()`: variance of the 10 samples (observations) of `x`.

Variance of the mean of the 10 observations: `x.var()`/10.

$\mathrm{Var}[\bar{X}] \downarrow$ as $N \uparrow \Longrightarrow \bar{X} \to \mu$ as $N \to \infty$ (Law of Large Numbers).

Some common probability distributions

Attributes of probability distributions via `Python/Scipy`

(See documentation for each distribution in `scipy.stats`)

- `rvs` - random variates (sample from the distribution)
- `pmf/pdf` - PMF or PDF
- `logpmf/logpdf` - log of the PMF or PDF
- `cdf` - CDF
- `logcdf` - log of the CDF
- `ppf` - percent point function (inverse of `cdf`; percentiles)
- `stats` - Mean('m'), variance('v'), skew('s'), kurtosis('k')
  (also see `mean`, `median`, `var`, `std`)
- `expect` - Compute expectation value of a function of this random variable
- `interval` - Confidence interval

Bernoulli (discrete; `scipy.stats.bernoulli`)

A Bernoulli random variable is the result of an experiment that asks a single yes-no question. Example: outcome of tossing a single (not necessarily fair) coin.

State space: $\{0, 1\}$. Probability distribution: $(1 - p, p)$, where p = probability of success.

Bernoulli (discrete; `scipy.stats.bernoulli`)

A Bernoulli random variable is the result of an experiment that asks a single yes-no question.
Example: outcome of tossing a single (not necessarily fair) coin.

State space: $\{0, 1\}$. Probability distribution: $(1 - p, p)$, where p = probability of success.

Definition (Bernoulli Distribution)

$$\mathrm{Bernoulli}(p) = p^x (1-p)^{1-x}\ \mathbb{I}_{x \in \{0,1\}}(x),$$

with $\mathbb{I}(x)$ the Indicator (or Heaviside) function: $\mathbb{I}_{x \in \{0,1\}}(x) = \begin{cases} 1, & \text{if } x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$

Bernoulli (discrete; `scipy.stats.bernoulli`)

A Bernoulli random variable is the result of an experiment that asks a single yes-no question.
Example: outcome of tossing a single (not necessarily fair) coin.

State space: $\{0, 1\}$. Probability distribution: $(1 - p, p)$, where p = probability of success.

Definition (Bernoulli Distribution)

$$\mathrm{Bernoulli}(p) = p^x (1-p)^{1-x} \; \mathbb{I}_{x \in \{0,1\}}(x),$$

with $\mathbb{I}(x)$ the Indicator (or Heaviside) function: $\mathbb{I}_{x \in \{0,1\}}(x) = \begin{cases} 1, & \text{if } x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$

Bernoulli (discrete; `scipy.stats.bernoulli`)

A Bernoulli random variable is the result of an experiment that asks a single yes-no question. Example: outcome of tossing a single (not necessarily fair) coin.

State space: $\{0, 1\}$. Probability distribution: $(1 - p, p)$, where p = probability of success.

Definition (Bernoulli Distribution)

$$\mathrm{Bernoulli}(p) = p^x (1-p)^{1-x}\ \mathbb{I}_{x\in\{0,1\}}(x),$$

with $\mathbb{I}(x)$ the Indicator (or Heaviside) function: $\mathbb{I}_{x\in\{0,1\}}(x) = \begin{cases} 1, & \text{if } x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$

Mean: $\mathbb{E}[X] = 1 \times P(X=1) + 0 \times P(X=0) = 1 \times p + 0 \times (1-p) = p$

Bernoulli (discrete; `scipy.stats.bernoulli`)

A Bernoulli random variable is the result of an experiment that asks a single yes-no question. Example: outcome of tossing a single (not necessarily fair) coin.

State space: $\{0, 1\}$. Probability distribution: $(1-p, p)$, where p = probability of success.

Definition (Bernoulli Distribution)

$$\text{Bernoulli}(p) = p^x (1-p)^{1-x}\ \mathbb{I}_{x\in\{0,1\}}(x),$$

with $\mathbb{I}(x)$ the Indicator (or Heaviside) function: $\mathbb{I}_{x\in\{0,1\}}(x) = \begin{cases} 1, & \text{if } x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$

Mean: $\mathbb{E}[X] = 1 \times P(X=1) + 0 \times P(X=0) = 1 \times p + 0 \times (1-p) = p$

Variance: First, $E[X^2] = 1^2 \times P(X=1) + 0^2 \times P(X=0) = 1^2 \times p + 0^2 \times (1-p) = p$

$\Rightarrow Var[X] = E[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$

Bernoulli (discrete; `scipy.stats.bernoulli`)

A Bernoulli random variable is the result of an experiment that asks a single yes-no question. Example: outcome of tossing a single (not necessarily fair) coin.

State space: $\{0, 1\}$. Probability distribution: $(1-p, p)$, where p = probability of success.

Definition (Bernoulli Distribution)

$$\text{Bernoulli}(p) = p^x (1-p)^{1-x}\ \mathbb{I}_{x\in\{0,1\}}(x),$$

with $\mathbb{I}(x)$ the Indicator (or Heaviside) function: $\mathbb{I}_{x\in\{0,1\}}(x) = \begin{cases} 1, & \text{if } x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$

Mean: $\mathbb{E}[X] = 1 \times P(X=1) + 0 \times P(X=0) = 1 \times p + 0 \times (1-p) = p$

Variance: First, $E[X^2] = 1^2 \times P(X=1) + 0^2 \times P(X=0) = 1^2 \times p + 0^2 \times (1-p) = p$

$\Rightarrow Var[X] = E[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$

Bernoulli (discrete; `scipy.stats.bernoulli`)

A Bernoulli random variable is the result of an experiment that asks a single yes-no question. Example: outcome of tossing a single (not necessarily fair) coin.

State space: $\{0, 1\}$. Probability distribution: $(1 - p, p)$, where p = probability of success.

Definition (Bernoulli Distribution)

$$\mathrm{Bernoulli}(p) = p^x (1-p)^{1-x}\ \mathbb{I}_{x\in\{0,1\}}(x),$$

with $\mathbb{I}(x)$ the Indicator (or Heaviside) function: $\mathbb{I}_{x\in\{0,1\}}(x) = \begin{cases} 1, & \text{if } x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$

Mean: $\mathbb{E}[X] = 1 \times P(X=1) + 0 \times P(X=0) = 1 \times p + 0 \times (1-p) = p$

Variance: First, $E[X^2] = 1^2 \times P(X=1) + 0^2 \times P(X=0) = 1^2 \times p + 0^2 \times (1-p) = p$

$\Rightarrow Var[X] = E[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$

```
from scipy.stats import bernoulli
p = 0.25; mean_th = p; var_th = p * (1-p)      #prob(success), theor. mean/var
x = bernoulli.rvs(p, size = 10)               #10 random deviates
print(x.mean()/mean_th); print(x.var()/var_th) #compare sample mean/var to theory
```

Bernoulli (discrete; `scipy.stats.bernoulli`)

A Bernoulli random variable is the result of an experiment that asks a single yes-no question. Example: outcome of tossing a single (not necessarily fair) coin.

State space: $\{0, 1\}$. Probability distribution: $(1-p, p)$, where p = probability of success.

Definition (Bernoulli Distribution)

$$\mathrm{Bernoulli}(p) = p^x (1-p)^{1-x}\ \mathbb{I}_{x \in \{0,1\}}(x),$$

with $\mathbb{I}(x)$ the Indicator (or Heaviside) function: $\mathbb{I}_{x \in \{0,1\}}(x) = \begin{cases} 1, & \text{if } x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$

Mean: $\mathbb{E}[X] = 1 \times P(X=1) + 0 \times P(X=0) = 1 \times p + 0 \times (1-p) = p$

Variance: First, $E[X^2] = 1^2 \times P(X=1) + 0^2 \times P(X=0) = 1^2 \times p + 0^2 \times (1-p) = p$

$\Rightarrow Var[X] = E[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$

```
from scipy.stats import bernoulli
p = 0.25; mean_th = p; var_th = p * (1-p)      #prob(success), theor. mean/var
x = bernoulli.rvs(p, size = 10)                #10 random deviates
print(x.mean()/mean_th); print(x.var()/var_th) #compare sample mean/var to theory
```

Law of Large Numbers: sample mean $\rightarrow \mathbb{E}[X]$ as sample size $\rightarrow \infty$.

Binomial (discrete, `scipy.stats.binom`)

Distribution of # successes in n independent experiments (n Bernoulli trials).

Binomial (discrete, `scipy.stats.binom`)

Distribution of # successes in n independent experiments (n Bernoulli trials).

Examples:

# of heads obtained in n tosses of a fair coin $= \mathrm{Binomial}(n, p = \frac{1}{2})$.

of "point" masses in a volume fraction V_1/V of space with N points in volume V

$= \mathrm{Binomial}(N, p = \frac{V_1}{V})$ (Meszaros, A. 1997 A&A 328, 1).

Binomial (discrete, scipy.stats.binom)

Distribution of # successes in n independent experiments (n Bernoulli trials).

Examples:

# of heads obtained in n tosses of a fair coin $= \mathrm{Binomial}(n, p = \frac{1}{2})$.

of "point" masses in a volume fraction V_1/V of space with N points in volume V $= \mathrm{Binomial}(N, p = \frac{V_1}{V})$ (Meszaros, A. 1997 A&A 328, 1).

Distribution = probability of k successes (and $n-k$ failures) in n trials:

$$\mathrm{Binomial}(n,k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$ (Binomial distribution)

Binomial (discrete, `scipy.stats.binom`)

Distribution of # successes in n independent experiments (n Bernoulli trials).

Examples:

# of heads obtained in n tosses of a fair coin $= \mathrm{Binomial}(n, p = \frac{1}{2})$.

of "point" masses in a volume fraction V_1/V of space with N points in volume V $= \mathrm{Binomial}(N, p = \frac{V_1}{V})$ (Meszaros, A. 1997 A&A 328, 1).

Distribution = probability of k successes (and $n - k$ failures) in n trials:

$$\mathrm{Binomial}(n, k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$ (Binomial distribution)

Mean: $\mathbb{E}[X] = np$ Variance: $Var[X] = np(1-p)$

Binomial (discrete, `scipy.stats.binom`)

Distribution of # successes in n independent experiments (n Bernoulli trials).

Examples:

# of heads obtained in n tosses of a fair coin $= \mathrm{Binomial}(n, p = \frac{1}{2})$.

of "point" masses in a volume fraction V_1/V of space with N points in volume V $= \mathrm{Binomial}(N, p = \frac{V_1}{V})$ (Meszaros, A. 1997 A&A 328, 1).

Distribution = probability of k successes (and $n - k$ failures) in n trials:

$$\mathrm{Binomial}(n, k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$ (Binomial distribution)

Mean: $\mathbb{E}[X] = np$ Variance: $Var[X] = np(1-p)$

Binomial (discrete, `scipy.stats.binom`)

Distribution of # successes in n independent experiments (n Bernoulli trials).

Examples:

# of heads obtained in n tosses of a fair coin $= \mathrm{Binomial}(n, p = \frac{1}{2})$.

of "point" masses in a volume fraction V_1/V of space with N points in volume V $= \mathrm{Binomial}(N, p = \frac{V_1}{V})$ (Meszaros, A. 1997 A&A 328, 1).

Distribution = probability of k successes (and $n-k$ failures) in n trials:

$$\mathrm{Binomial}(n, k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$ (Binomial distribution)

Mean: $\mathbb{E}[X] = np$ Variance: $Var[X] = np(1-p)$

Both are n times the values for $\mathrm{Bernoulli}(p)$ as expected (n independent Bernoulli trials).

Binomial (discrete, `scipy.stats.binom`)

Distribution of # successes in n independent experiments (n Bernoulli trials).

Examples:

# of heads obtained in n tosses of a fair coin $= \mathrm{Binomial}(n, p = \frac{1}{2})$.

of "point" masses in a volume fraction V_1/V of space with N points in volume V $= \mathrm{Binomial}(N, p = \frac{V_1}{V})$ (Meszaros, A. 1997 A&A 328, 1).

Distribution = probability of k successes (and $n - k$ failures) in n trials:

$$\mathrm{Binomial}(n, k) = \binom{n}{k} p^k (1 - p)^{(n-k)} \text{ (Binomial distribution)}$$

Mean: $\mathbb{E}[X] = np$ Variance: $Var[X] = np(1 - p)$

Both are n times the values for $\mathrm{Bernoulli}(p)$ as expected (n independent Bernoulli trials).

Practice:

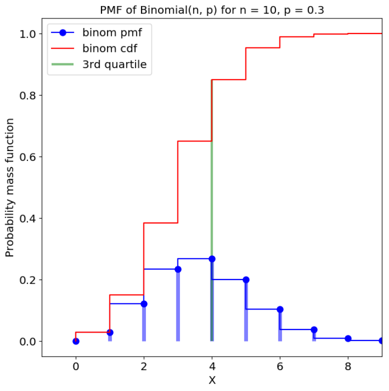
If prob. of single success $p = 0.25$, compute prob. of $k = 2$ successes in $n = 10$ trials

```
from scipy.stats import binom
n, k, p = 10, 2, 0.25          #total trials, # successes, prob of 1 success
print(binom.pmf(k, n, p))      #prob of k successes in n trials
```

Binomial distribution, contd.

Use the python module scipy.stats.binom to answer the following:

1) For $n = 10, p = 0.3$, what is the probability of $k = 6$ successes?
2) For $n = 10, p = 0.3$, what is the probability of having $k > 2$ successes?
3) For $n = 10, p = 0.3$, for what k is the probability of $\leq k$ successes at least 0.75?



Code for plot available here