Statistics for Astronomers: Lecture 2, 2020.09.23

Prof. Sundar Srinivasan

IRyA/UNAM

BY NC SA

Review

Probability: Classical vs. Frequentist vs. Bayesian.
Kolmogorov Axioms.
Conditional and marginal probability.
Independence and exclusivity.
Law of Total Probability from union of pairwise disjoint, collectively exhaustive sets.
Bayes' Theorem.

Random variables and probability distributions

Random variables

Random: Uncertain, no "pattern" can be detected.

Random variables

Random: Uncertain, no "pattern" can be detected.

Randomness: A measure of uncertainty of the outcome of an experiment.
Some sources of "true" randomness – initial conditions of the experiment (*e.g.*, throwing dice, chaos) and environmental effects (*e.g.*, Brownian Motion, dark current).

Random variables

Random: Uncertain, no "pattern" can be detected.

Randomness: A measure of uncertainty of the outcome of an experiment.
Some sources of "true" randomness – initial conditions of the experiment (*e.g.*, throwing dice, chaos) and environmental effects (*e.g.*, Brownian Motion, dark current).

Random variable: A function that assigns a numerical value to each distinct outcome.
The set of assigned numerical values is the state space S.
A random variable is a mapping from the sample space to the state space; $X : \Omega \longrightarrow S$.

Random variables

Random: Uncertain, no "pattern" can be detected.

Randomness: A measure of uncertainty of the outcome of an experiment.
Some sources of "true" randomness – initial conditions of the experiment (*e.g.*, throwing dice, chaos) and environmental effects (*e.g.*, Brownian Motion, dark current).

Random variable: A function that assigns a numerical value to each distinct outcome.
The set of assigned numerical values is the state space S.
A random variable is a mapping from the sample space to the state space; $X : \Omega \longrightarrow S$.

Random process: A sequence of random variables whose outcomes don't follow a pattern.
Their evolution can, however, be described probabilistically.
Example: Markov Chains.

Random variables

Random: Uncertain, no "pattern" can be detected.

Randomness: A measure of uncertainty of the outcome of an experiment.
Some sources of "true" randomness – initial conditions of the experiment (*e.g.*, throwing dice, chaos) and environmental effects (*e.g.*, Brownian Motion, dark current).

Random variable: A function that assigns a numerical value to each distinct outcome.
The set of assigned numerical values is the state space S.
A random variable is a mapping from the sample space to the state space; $X : \Omega \longrightarrow S$.

Random process: A sequence of random variables whose outcomes don't follow a pattern.
Their evolution can, however, be described probabilistically.
Example: Markov Chains.

Probability distribution: A function that maps a random variable to a real number; $p : X \rightarrow \mathbb{R}$.

Random variables

Random: Uncertain, no "pattern" can be detected.

Randomness: A measure of uncertainty of the outcome of an experiment.
Some sources of "true" randomness – initial conditions of the experiment (*e.g.*, throwing dice, chaos) and environmental effects (*e.g.*, Brownian Motion, dark current).

Random variable: A function that assigns a numerical value to each distinct outcome.
The set of assigned numerical values is the state space S.
A random variable is a mapping from the sample space to the state space; $X : \Omega \longrightarrow S$.

Random process: A sequence of random variables whose outcomes don't follow a pattern.
Their evolution can, however, be described probabilistically.
Example: Markov Chains.

Probability distribution: A function that maps a random variable to a real number; $p : X \to \mathbb{R}$.

Convention:
X: random variable, x: value assigned to random variable.
"Probability that X has value x": $P(X = x)$, $P_X(x)$ (x discrete) or $p_X(x)$ (x continuous).
Specifying $P_X(x)$ or $p_X(x)$ for all $x \in S$, the state space results in a probability distribution.
Discrete: mass function (PMF). Continuous: density function (PDF).
Note: I won't abbreviate "probability distribution function", so that "PDF" is unambiguous.

Illustration

3 balls are drawn (w/o replacement) from a box containing 2 red balls and 8 yellow balls. Sketch the probability distribution for the colour of the third ball.

Illustration

3 balls are drawn (w/o replacement) from a box containing 2 red balls and 8 yellow balls. Sketch the probability distribution for the colour of the third ball.

Sample space: $\Omega =$ {~~🔴🔴🔴~~, 🔴🔴🟡, 🔴🟡🔴, 🔴🟡🟡, 🟡🔴🔴, 🟡🔴🟡, 🟡🟡🔴, 🟡🟡🟡}. State space: $S = \{0, 1\}$.
Random variables: X_1, X_2, X_3 (one for the colour of each ball), each draws values from S.

Illustration

3 balls are drawn (w/o replacement) from a box containing 2 red balls and 8 yellow balls. Sketch the probability distribution for the colour of the third ball.

Sample space: $\Omega = \{$~~🔴🔴🔴~~, 🔴🔴🟡, 🔴🟡🔴, 🔴🟡🟡, 🟡🔴🔴, 🟡🔴🟡, 🟡🟡🔴, 🟡🟡🟡$\}$. State space: $S = \{0, 1\}$.
Random variables: X_1, X_2, X_3 (one for the colour of each ball), each draws values from S.

The PMF for X_3 is obtained by marginalising over X_1 and X_2:

$$P_{x_3}(x_3) = \sum_{x_2 \in S} \sum_{x_1 \in S} P(X_3 = x_3 | X_2 = x_2, X_1 = x_1)\ P(X_2 = x_2 | X_1 = x_1)\ P(X_1 = x_1)$$

Illustration

3 balls are drawn (w/o replacement) from a box containing 2 red balls and 8 yellow balls. Sketch the probability distribution for the colour of the third ball.

Sample space: $\Omega = \{$~~🔴🔴🔴~~, 🔴🔴🟡, 🔴🟡🔴, 🔴🟡🟡, 🟡🔴🔴, 🟡🔴🟡, 🟡🟡🔴, 🟡🟡🟡$\}$. State space: $S = \{0, 1\}$.
Random variables: X_1, X_2, X_3 (one for the colour of each ball), each draws values from S.

The PMF for X_3 is obtained by marginalising over X_1 and X_2:

$$P_{x_3}(x_3) = \sum_{x_2 \in S} \sum_{x_1 \in S} P(X_3 = x_3 | X_2 = x_2, X_1 = x_1)\ P(X_2 = x_2 | X_1 = x_1)\ P(X_1 = x_1)$$

The state $X_3 = 0$ ("third ball is red") can be achieved in three ways (see Ω).

Illustration

3 balls are drawn (w/o replacement) from a box containing 2 red balls and 8 yellow balls. Sketch the probability distribution for the colour of the third ball.

Sample space: $\Omega = \{$~~🔴🔴🔴~~, 🔴🔴🟡, 🔴🟡🔴, 🔴🟡🟡, 🟡🔴🔴, 🟡🔴🟡, 🟡🟡🔴, 🟡🟡🟡$\}$. State space: $S = \{0, 1\}$.
Random variables: X_1, X_2, X_3 (one for the colour of each ball), each draws values from S.

The PMF for X_3 is obtained by marginalising over X_1 and X_2:

$$P_{x_3}(x_3) = \sum_{x_2 \in S} \sum_{x_1 \in S} P(X_3 = x_3 | X_2 = x_2, X_1 = x_1)\ P(X_2 = x_2 | X_1 = x_1)\ P(X_1 = x_1)$$

The state $X_3 = 0$ ("third ball is red") can be achieved in three ways (see Ω).

$$\begin{aligned} P(X_3 = 0) = & \ P(X_3 = 0 | X_2 = 1, X1 = 0)\ P(X_2 = 1 | X_1 = 0)\ P(X_1 = 0) \\ & + P(X_3 = 0 | X_2 = 0, X1 = 1)\ P(X_2 = 0 | X_1 = 1)\ P(X_1 = 1) \\ & + P(X_3 = 0 | X_2 = 1, X1 = 1)\ P(X_2 = 1 | X_1 = 1)\ P(X_1 = 1) \\ & = \frac{2}{10} \cdot \frac{8}{9} \cdot \frac{1}{8} + \frac{8}{10} \cdot \frac{2}{9} \cdot \frac{1}{8} + \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{2}{8} = \frac{1}{5} \end{aligned}$$

Illustration

3 balls are drawn (w/o replacement) from a box containing 2 red balls and 8 yellow balls. Sketch the probability distribution for the colour of the third ball.

Sample space: $\Omega = \{$~~🔴🔴🔴~~, 🔴🔴🟡, 🔴🟡🔴, 🔴🟡🟡, 🟡🔴🔴, 🟡🔴🟡, 🟡🟡🔴, 🟡🟡🟡$\}$. State space: $S = \{0, 1\}$.
Random variables: X_1, X_2, X_3 (one for the colour of each ball), each draws values from S.

The PMF for X_3 is obtained by marginalising over X_1 and X_2:

$$P_{x_3}(x_3) = \sum_{x_2 \in S} \sum_{x_1 \in S} P(X_3 = x_3 | X_2 = x_2, X_1 = x_1)\ P(X_2 = x_2 | X_1 = x_1)\ P(X_1 = x_1)$$

The state $X_3 = 0$ ("third ball is red") can be achieved in three ways (see Ω).

$$\begin{aligned} P(X_3 = 0) &= P(X_3 = 0 | X_2 = 1, X1 = 0)\ P(X_2 = 1 | X_1 = 0)\ P(X_1 = 0) \\ &+ P(X_3 = 0 | X_2 = 0, X1 = 1)\ P(X_2 = 0 | X_1 = 1)\ P(X_1 = 1) \\ &+ P(X_3 = 0 | X_2 = 1, X1 = 1)\ P(X_2 = 1 | X_1 = 1)\ P(X_1 = 1) \\ &= \frac{2}{10} \cdot \frac{8}{9} \cdot \frac{1}{8} + \frac{8}{10} \cdot \frac{2}{9} \cdot \frac{1}{8} + \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{2}{8} = \frac{1}{5} \end{aligned}$$

$$\Longrightarrow P(X_3 = 1) = 1 - P(X_3 = 0) = \frac{4}{5}.$$

Illustration

3 balls are drawn (w/o replacement) from a box containing 2 red balls and 8 yellow balls. Sketch the probability distribution for the colour of the third ball.

Sample space: $\Omega = \{$~~🔴🔴🔴~~, 🔴🔴🟡, 🔴🟡🔴, 🔴🟡🟡, 🟡🔴🔴, 🟡🔴🟡, 🟡🟡🔴, 🟡🟡🟡$\}$. State space: $S = \{0, 1\}$.
Random variables: X_1, X_2, X_3 (one for the colour of each ball), each draws values from S.

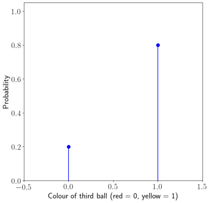
The PMF for X_3 is obtained by marginalising over X_1 and X_2:

$$P_{x_3}(x_3) = \sum_{x_2 \in S} \sum_{x_1 \in S} P(X_3 = x_3 | X_2 = x_2, X_1 = x_1)\ P(X_2 = x_2 | X_1 = x_1)\ P(X_1 = x_1)$$

The state $X_3 = 0$ ("third ball is red") can be achieved in three ways (see Ω).

$$\begin{aligned} P(X_3 = 0) &= P(X_3 = 0 | X_2 = 1, X1 = 0)\ P(X_2 = 1 | X_1 = 0)\ P(X_1 = 0) \\ &+ P(X_3 = 0 | X_2 = 0, X1 = 1)\ P(X_2 = 0 | X_1 = 1)\ P(X_1 = 1) \\ &+ P(X_3 = 0 | X_2 = 1, X1 = 1)\ P(X_2 = 1 | X_1 = 1)\ P(X_1 = 1) \\ &= \frac{2}{10} \cdot \frac{8}{9} \cdot \frac{1}{8} + \frac{8}{10} \cdot \frac{2}{9} \cdot \frac{1}{8} + \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{2}{8} = \frac{1}{5} \end{aligned}$$

$$\Longrightarrow P(X_3 = 1) = 1 - P(X_3 = 0) = \frac{4}{5}.$$



Code for plot available ▶ here.

Discrete and continuous probability distributions

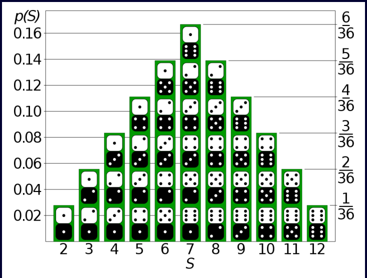
Sum of numbers displayed on two dice after one throw (discrete distribution):

State space $= \{2, 3, ..., 11, 12\}$

Discrete and continuous probability distributions

Sum of numbers displayed on two dice after one throw (discrete distribution):

State space = $\{2, 3, ..., 11, 12\}$

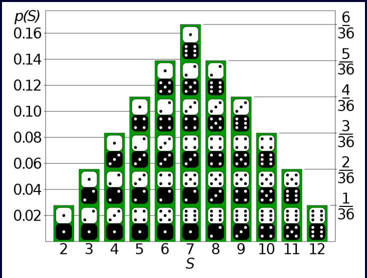


(Tim Stellmach/Public Domain)

Discrete and continuous probability distributions

Sum of numbers displayed on two dice after one throw (discrete distribution):

State space $= \{2, 3, ..., 11, 12\}$



(Tim Stellmach/Public Domain)

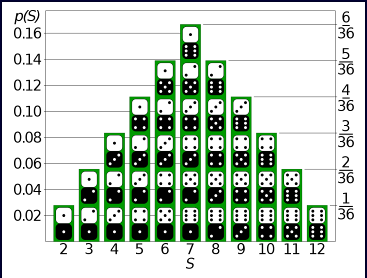
$P(X = 4)$ makes sense, is finite.

$$P(4 \leq X \leq 7) = \sum_{i=4}^{7} P(X = i) = \frac{18}{36} = \frac{1}{2}.$$

Discrete and continuous probability distributions

Sum of numbers displayed on two dice after one throw (discrete distribution):

State space $= \{2, 3, ..., 11, 12\}$



(Tim Stellmach/Public Domain)

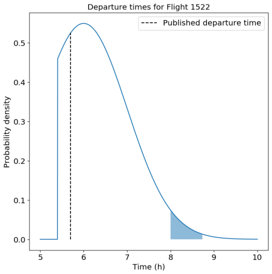
$P(X = 4)$ makes sense, is finite.

$$P(4 \le X \le 7) = \sum_{i=4}^{7} P(X = i) = \frac{18}{36} = \frac{1}{2}.$$

Actual departure time of Flight 1522 (continuous distribution):

State space $= (5.4, \infty)$

Code for plot available here.



Shaded region $= P(8{:}00\ \mathrm{AM} < T < 8{:}44\ \mathrm{AM})$.

$P(X = 8{:}15\ \mathrm{AM})$ is meaningless, zero.

$$P(8{:}00\ \mathrm{AM} < T < 8{:}44\ \mathrm{AM}) = \int_{8:00\ \mathrm{AM}}^{8:44\ \mathrm{AM}} p_X(x)\, dx \text{ is finite.}$$

Populations and samples

If a random variable X has probability distribution is $P_X(x)$ (discrete) or $p_X(x)$ (continuous), we say that X is drawn from the PMF/PDF: $X \sim P_X(x)$ or $X \sim p_X(x)$.

Populations and samples

If a random variable X has probability distribution is $P_X(x)$ (discrete) or $p_X(x)$ (continuous), we say that X is drawn from the PMF/PDF: $X \sim P_X(x)$ or $X \sim p_X(x)$.

Population: the underlying probability distribution.

Populations and samples

If a random variable X has probability distribution is $P_X(x)$ (discrete) or $p_X(x)$ (continuous), we say that X is drawn from the PMF/PDF: $X \sim P_X(x)$ or $X \sim p_X(x)$.

Population: the underlying probability distribution.
Sample: the results of a finite number of experiments/draws from the population (a subset).

Populations and samples

If a random variable X has probability distribution is $P_X(x)$ (discrete) or $p_X(x)$ (continuous), we say that X is drawn from the PMF/PDF: $X \sim P_X(x)$ or $X \sim p_X(x)$.

Population: the underlying probability distribution.
Sample: the results of a finite number of experiments/draws from the population (a subset).

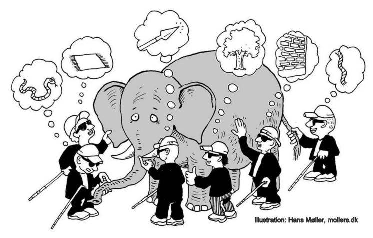
Experiment performed a finite number of times; sample unable to faithfully reproduce the population – statistics (quantities derived from the sample) are only guesses at (estimates of) the corresponding parameters (values that describe the population).
Convention: Greek symbols for parameters (*e.g.*, μ, σ), Latin symbols for statistics (*e.g.*, $\bar{x}$, s).

Populations and samples

If a random variable X has probability distribution is $P_X(x)$ (discrete) or $p_X(x)$ (continuous), we say that X is drawn from the PMF/PDF: $X \sim P_X(x)$ or $X \sim p_X(x)$.

Population: the underlying probability distribution.
Sample: the results of a finite number of experiments/draws from the population (a subset).

Experiment performed a finite number of times; sample unable to faithfully reproduce the population – statistics (quantities derived from the sample) are only guesses at (estimates of) the corresponding parameters (values that describe the population).
Convention: Greek symbols for parameters (*e.g.*, μ, σ), Latin symbols for statistics (*e.g.*, $\bar{x}$, s).



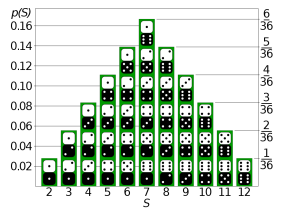
"More data is required."

Populations and samples

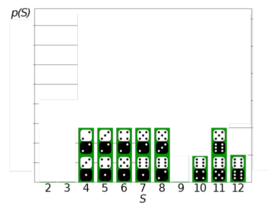
If a random variable X has probability distribution is $P_X(x)$ (discrete) or $p_X(x)$ (continuous), we say that X is drawn from the PMF/PDF: $X \sim P_X(x)$ or $X \sim p_X(x)$.

Population: the underlying probability distribution.
Sample: the results of a finite number of experiments/draws from the population (a subset).

Experiment performed a finite number of times; sample unable to faithfully reproduce the population – statistics (quantities derived from the sample) are only guesses at (estimates of) the corresponding parameters (values that describe the population).
Convention: Greek symbols for parameters (*e.g.*, μ, σ), Latin symbols for statistics (*e.g.*, $\bar{x}$, s).



(Tim Stellmach/Public Domain)
Population: $\mu = 7.00$, $Var(X) = 5.83$.



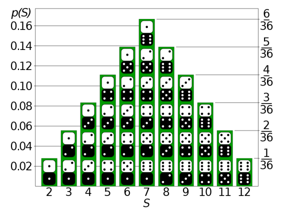
Sample of outcomes obtained from rolling two dice 14 times.
$\bar{x} = 7.43$, sample variance (discussed later) $= 6.67$

Populations and samples

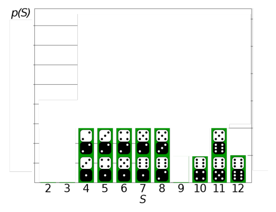
If a random variable X has probability distribution is $P_X(x)$ (discrete) or $p_X(x)$ (continuous), we say that X is drawn from the PMF/PDF: $X \sim P_X(x)$ or $X \sim p_X(x)$.

Population: the underlying probability distribution.
Sample: the results of a finite number of experiments/draws from the population (a subset).

Experiment performed a finite number of times; sample unable to faithfully reproduce the population – statistics (quantities derived from the sample) are only guesses at (estimates of) the corresponding parameters (values that describe the population).
Convention: Greek symbols for parameters (*e.g.*, μ, σ), Latin symbols for statistics (*e.g.*, $\bar{x}$, s).



(Tim Stellmach/Public Domain)
Population: $\mu = 7.00$, $Var(X) = 5.83$.



Sample of outcomes obtained from rolling two dice 14 times.
$\bar{x} = 7.43$, sample variance (discussed later) $= 6.67$

Sample distribution seems uniform, results in a larger sample mean/variance than the true (population) values.

Cumulative distribution function (CDF)

Definition (Cumulative distribution function)

A function $F_X(x)$ of a random variable X such that $F_X(x)$ is the probability that $X \leq x$.

For a discrete PMF:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i).$$

For a continuous PDF:

$$F_X(x) = P(X \leq x) = \int_{t=-\infty}^{t=x} p_X(t)\, dt.$$

Cumulative distribution function (CDF)

Definition (Cumulative distribution function)

A function $F_X(x)$ of a random variable X such that $F_X(x)$ is the probability that $X \leq x$.

For a discrete PMF:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i).$$

For a continuous PDF:

$$F_X(x) = P(X \leq x) = \int_{t=-\infty}^{t=x} p_X(t)\ dt.$$

From this definition, the probability of the variable ranging between two values a and b is

$P(a < X \leq b) = F_X(x = b) - F_X(x = a)$. For a PDF, this is also equal to $\int_{t=a}^{t=b} p_X(t)\ dt$.

Cumulative distribution function (CDF)

Definition (Cumulative distribution function)

A function $F_X(x)$ of a random variable X such that $F_X(x)$ is the probability that $X \leq x$.

For a discrete PMF:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i).$$

For a continuous PDF:

$$F_X(x) = P(X \leq x) = \int_{t=-\infty}^{t=x} p_X(t)\, dt.$$

From this definition, the probability of the variable ranging between two values a and b is

$P(a < X \leq b) = F_X(x = b) - F_X(x = a)$. For a PDF, this is also equal to $\int_{t=a}^{t=b} p_X(t)\, dt$.

The CDF is a monotonically increasing function.

For a discrete random variable, it is constant in between values.

For the continuous case, the PDF is the derivative of the CDF w.r.t. x.

Cumulative distribution function (contd.)

Definition (Quantile function)

The inverse of the CDF, a function $Q(p)$ that returns the value of x such that $F_X(X \leq x) = p$.

Cumulative distribution function (contd.)

Definition (Quantile function)

The inverse of the CDF, a function $Q(p)$ that returns the value of x such that $F_x(X \leq x) = p$.

e.g., $Q(p = 0.5)$ is the median (equal "mass" on either side of $x = Q(0.5)$).
$Q(p = 0.25)$ and $Q(p = 0.75)$ are the first and third quartiles.

Cumulative distribution function (contd.)

Definition (Quantile function)

The inverse of the CDF, a function $Q(p)$ that returns the value of x such that $F_X(X \leq x) = p$.

e.g., $Q(p = 0.5)$ is the median (equal "mass" on either side of $x = Q(0.5)$).
$Q(p = 0.25)$ and $Q(p = 0.75)$ are the first and third quartiles.

Definition (Independent and identically distributed variables)

Two random variables X and Y are said to be iid if and only if they are mutually independent and drawn from the same distribution:

$$F_{X,Y}(x, y) = F_X(x) \times F_Y(y)$$
$$F_X(x) = F_Y(x)$$

Cumulative distribution function (contd.)

Definition (Quantile function)

The inverse of the CDF, a function $Q(p)$ that returns the value of x such that $F_X(X \leq x) = p$.

e.g., $Q(p = 0.5)$ is the median (equal "mass" on either side of $x = Q(0.5)$).
$Q(p = 0.25)$ and $Q(p = 0.75)$ are the first and third quartiles.

Definition (Independent and identically distributed variables)

Two random variables X and Y are said to be iid if and only if they are mutually independent and drawn from the same distribution:

$$F_{X,Y}(x, y) = F_X(x) \times F_Y(y)$$
$$F_X(x) = F_Y(x)$$

Heteroskedastic: originally iid observations + measurement errors that aren't identical.

Typical case in astronomy.

Expectation value

Definition (Expectation value)

The expectation value $E[g(X)]$ of a function $g(X)$ of a random variable X, is the weighted average of $g(X)$, with the weights being the associated probabilities:

Discrete: $E[g(X)] = \sum_{i=1}^{N} g(x_i)\ P(X = x_i)$.

Continuous: $E[g(X)] = \int_{t=-\infty}^{t=\infty} g(x)\ p_X(x)\ dx$.

Expectation value

Definition (Expectation value)

The expectation value $E[g(X)]$ of a function $g(X)$ of a random variable X, is the weighted average of $g(X)$, with the weights being the associated probabilities:

Discrete: $E[g(X)] = \sum_{i=1}^{N} g(x_i)\ P(X = x_i)$. Continuous: $E[g(X)] = \int_{t=-\infty}^{t=\infty} g(x)\ p_X(x)\ dx$.

Linearity: $E[\alpha g(X) + \beta h(X)] = \alpha E[g(X)] + \beta E[h(X)]$. ($\sum$ and $\int$ are linear operators!)

Independence: $X \perp Y \implies E[XY] = E[X]E[Y]$.

Expectation value

Definition (Expectation value)

The expectation value $E[g(X)]$ of a function $g(X)$ of a random variable X, is the weighted average of $g(X)$, with the weights being the associated probabilities:

Discrete: $E[g(X)] = \sum_{i=1}^{N} g(x_i)\ P(X = x_i)$.

Continuous: $E[g(X)] = \int_{t=-\infty}^{t=\infty} g(x)\ p_X(x)\ dx$.

Linearity: $E[\alpha g(X) + \beta h(X)] = \alpha E[g(X)] + \beta E[h(X)]$. ($\sum$ and $\int$ are linear operators!)

Independence: $X \perp Y \implies E[XY] = E[X]E[Y]$.

Mean $\equiv E[X]$. An estimate of the distribution's location or central tendency.

Variance, $Var[X] \equiv E[(X - E[X])^2] = (\text{standard deviation})^2$. An estimate of spread.

Expectation value

Definition (Expectation value)

The expectation value $E[g(X)]$ of a function $g(X)$ of a random variable X, is the weighted average of $g(X)$, with the weights being the associated probabilities:

Discrete: $E[g(X)] = \sum_{i=1}^{N} g(x_i)\ P(X = x_i)$. Continuous: $E[g(X)] = \int_{t=-\infty}^{t=\infty} g(x)\ p_X(x)\ dx$.

Linearity: $E[\alpha g(X) + \beta h(X)] = \alpha E[g(X)] + \beta E[h(X)]$. ($\sum$ and $\int$ are linear operators!)

Independence: $X \perp Y \implies E[XY] = E[X]E[Y]$.

Mean $\equiv E[X]$. An estimate of the distribution's location or central tendency.

Variance, $Var[X] \equiv E[(X - E[X])^2] = (\text{standard deviation})^2$. An estimate of spread.

Why the square? What is $E[X - E(X)]$?

Expectation value

Definition (Expectation value)

The expectation value $E[g(X)]$ of a function $g(X)$ of a random variable X, is the weighted average of $g(X)$, with the weights being the associated probabilities:

Discrete: $E[g(X)] = \sum_{i=1}^{N} g(x_i)\ P(X = x_i)$. Continuous: $E[g(X)] = \int_{t=-\infty}^{t=\infty} g(x)\ p_X(x)\ dx$.

Linearity: $E[\alpha g(X) + \beta h(X)] = \alpha E[g(X)] + \beta E[h(X)]$. ($\sum$ and $\int$ are linear operators!)

Independence: $X \perp Y \implies E[XY] = E[X]E[Y]$.

Mean $\equiv E[X]$. An estimate of the distribution's location or central tendency.

Variance, $Var[X] \equiv E[(X - E[X])^2] = (\text{standard deviation})^2$. An estimate of spread.

Why the square? What is $E[X - E(X)]$?

The expression for the variance can be simplified: $Var[X] = E[X^2] - (E[X])^2$.

Note: When no other function is specified, "expectation value" refers to the mean, $E(X)$.

Some properties of the variance of a random variable

1. By definition, non-negative.

Some properties of the variance of a random variable

1. By definition, non-negative.

2. For any constant α:

   – $Var(\alpha) = 0$, because $E[\alpha] = \alpha$.

   – $Var(X + \alpha) = Var(X)$ – *i.e.*, invariant w.r.t. a location parameter.

   – $Var(\alpha X) = \alpha^2 Var(X)$.

Some properties of the variance of a random variable

1. By definition, non-negative.

2. For any constant α:

   – $Var(\alpha) = 0$, because $E[\alpha] = \alpha$.

   – $Var(X + \alpha) = Var(X)$ – *i.e.*, invariant w.r.t. a location parameter.

   – $Var(\alpha X) = \alpha^2 Var(X)$.

3. For constants α, β and random variables X, Y,

   $Var(\alpha X + \beta Y) = ??$

   Evaluate this expression using the definition of variance in terms of expectation values.

Covariance

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

Covariance

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

If $X \perp Y$, then $(Y - E[Y])$ independent of $(X - E[X])$ for any (X, Y) pair.
$\implies$ the term quantifies a dependence between X and Y.

Covariance

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

If $X \perp Y$, then $(Y - E[Y])$ independent of $(X - E[X])$ for any (X, Y) pair.
$\implies$ the term quantifies a dependence between X and Y.

If X and Y are not independent, then $(X - E[X])(Y - E[Y]) > 0$ if both deviations are in the same direction, and < 0 if the variables deviate from their means in opposite directions.

Covariance

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

If $X \perp Y$, then $(Y - E[Y])$ independent of $(X - E[X])$ for any (X, Y) pair.
$\Longrightarrow$ the term quantifies a dependence between X and Y.

If X and Y are not independent, then $(X - E[X])(Y - E[Y]) > 0$ if both deviations are in the same direction, and < 0 if the variables deviate from their means in opposite directions.

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

Covariance

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

If $X \perp Y$, then $(Y - E[Y])$ independent of $(X - E[X])$ for any (X, Y) pair.
$\implies$ the term quantifies a dependence between X and Y.

If X and Y are not independent, then $(X - E[X])(Y - E[Y]) > 0$ if both deviations are in the same direction, and < 0 if the variables deviate from their means in opposite directions.

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

Definition (Covariance)

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

Covariance

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

If $X \perp Y$, then $(Y - E[Y])$ independent of $(X - E[X])$ for any (X, Y) pair.
$\implies$ the term quantifies a dependence between X and Y.

If X and Y are not independent, then $(X - E[X])(Y - E[Y]) > 0$ if both deviations are in the same direction, and < 0 if the variables deviate from their means in opposite directions.

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

Definition (Covariance)

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

Therefore, $Var(\alpha X + \beta Y) = \alpha^2 Var(X) + \beta^2 Var(Y) + 2\alpha\beta Cov(X, Y)$.

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

Correlation coefficient

The sign of $Cov(X, Y)$ probes a linear relationship between the two variables X and Y.

We can define a scale-invariant of $Cov(X, Y)$ instead:

Correlation coefficient

The sign of $Cov(X, Y)$ probes a linear relationship between the two variables X and Y.

However, the magnitude isn't as useful, as the covariance is not scale-invariant:
$Cov(\alpha X, \beta Y) = \alpha\beta Cov(X, Y)$.

We can define a scale-invariant of $Cov(X, Y)$ instead:

Definition ((Pearson's) Correlation coefficient)

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

Correlation coefficient

The sign of $Cov(X, Y)$ probes a linear relationship between the two variables X and Y.

However, the magnitude isn't as useful, as the covariance is not scale-invariant:
$Cov(\alpha X, \beta Y) = \alpha\beta Cov(X, Y)$.

We can define a scale-invariant of $Cov(X, Y)$ instead:

Definition ((Pearson's) Correlation coefficient)

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

By definition, $\rho_{XX} = 1$. "Perfect correlation".

$\rho_{XX} = -1$: "perfect anticorrelation".

Uncorrelated random variables

If random variables X and Y are such that $Cov(X, Y) = 0$, then they are uncorrelated.

Uncorrelated random variables

If random variables X and Y are such that $Cov(X, Y) = 0$, then they are uncorrelated.

The variance of the sum of N uncorrelated variables is therefore the sum of their variances:

$$Var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} Var(X_i)$$ (Bienaymé formula).

Uncorrelated random variables

If random variables X and Y are such that $Cov(X, Y) = 0$, then they are uncorrelated.

The variance of the sum of N uncorrelated variables is therefore the sum of their variances:
$Var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} Var(X_i)$ (Bienaymé formula).

For N iid variables, $Var(\bar{X}) = \frac{1}{N^2} Var\left(\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2} \sum_{i=1}^{N} Var(X_i) = \frac{1}{N^2} N\ Var(X) = \frac{Var(X)}{N}$

Uncorrelated random variables

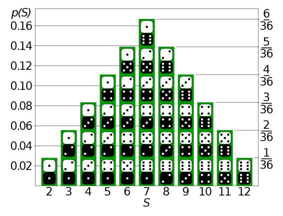
If random variables X and Y are such that $Cov(X,Y) = 0$, then they are uncorrelated.

The variance of the sum of N uncorrelated variables is therefore the sum of their variances:

$Var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} Var(X_i)$ (Bienaymé formula).

For N iid variables, $Var(\bar{X}) = \frac{1}{N^2} Var\left(\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2}\sum_{i=1}^{N} Var(X_i) = \frac{1}{N^2} N\ Var(X) = \frac{Var(X)}{N}$

Two-dice example



(Tim Stellmach/Public Domain)

$\mu = 7.00$, $Var(X) = 5.83$.

X_1, X_2: random variables for the values on the 1st and 2nd die after each throw, we record $X_1 + X_2$ each time.

Uncorrelated random variables

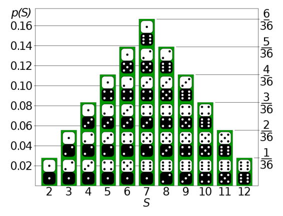
If random variables X and Y are such that $Cov(X, Y) = 0$, then they are uncorrelated.

The variance of the sum of N uncorrelated variables is therefore the sum of their variances:

$$Var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} Var(X_i) \text{ (Bienaymé formula).}$$

For N iid variables, $Var(\bar{X}) = \frac{1}{N^2} Var\left(\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2}\sum_{i=1}^{N} Var(X_i) = \frac{1}{N^2} N\ Var(X) = \frac{Var(X)}{N}$

Two-dice example



(Tim Stellmach/Public Domain)
$\mu = 7.00$, $Var(X) = 5.83$.

X_1, X_2: random variables for the values on the 1st and 2nd die after each throw, we record $X_1 + X_2$ each time.

Variance on a single measurement of this sum = 5.83.

Uncorrelated random variables

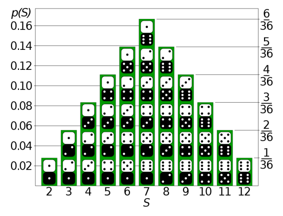
If random variables X and Y are such that $Cov(X, Y) = 0$, then they are uncorrelated.

The variance of the sum of N uncorrelated variables is therefore the sum of their variances:

$$Var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} Var(X_i) \text{ (Bienaymé formula).}$$

For N iid variables, $Var(\bar{X}) = \frac{1}{N^2} Var\left(\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2} \sum_{i=1}^{N} Var(X_i) = \frac{1}{N^2} N\ Var(X) = \frac{Var(X)}{N}$

Two-dice example



(Tim Stellmach/Public Domain)
$\mu = 7.00$, $Var(X) = 5.83$.

X_1, X_2: random variables for the values on the 1st and 2nd die after each throw, we record $X_1 + X_2$ each time.

Variance on a single measurement of this sum $= 5.83$.

Variance on average of 10 measurements: $\frac{5.83}{10} = 0.583$

Uncorrelated random variables

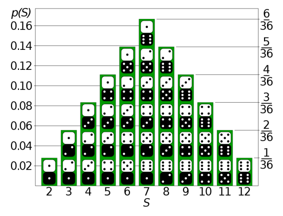
If random variables X and Y are such that $Cov(X, Y) = 0$, then they are uncorrelated.

The variance of the sum of N uncorrelated variables is therefore the sum of their variances:

$$Var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} Var(X_i) \text{ (Bienaymé formula).}$$

For N iid variables, $Var(\bar{X}) = \frac{1}{N^2} Var\left(\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2} \sum_{i=1}^{N} Var(X_i) = \frac{1}{N^2} N\ Var(X) = \frac{Var(X)}{N}$

Two-dice example



(Tim Stellmach/Public Domain)
$\mu = 7.00$, $Var(X) = 5.83$.

X_1, X_2: random variables for the values on the 1st and 2nd die after each throw, we record $X_1 + X_2$ each time.

Variance on a single measurement of this sum $= 5.83$.

Variance on average of 10 measurements: $\frac{5.83}{10} = 0.583$

As # measurements $N \uparrow$, variance on the mean of N measurements $\downarrow$.
Sample mean approaches population mean.

See 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.

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.

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

Definition (Bernoulli Distribution)

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

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.

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

Definition (Bernoulli Distribution)

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

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.

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

Definition (Bernoulli Distribution)

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

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

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

A Bernoulli random variable is the result of an experiment that asks a single yes-no question.

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

Definition (Bernoulli Distribution)

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

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

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

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

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

A Bernoulli random variable is the result of an experiment that asks a single yes-no question.

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

Definition (Bernoulli Distribution)

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

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

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

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

Example of a Bernoulli random variable: outcome of tossing a single (not necessarily fair) coin.

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

A Bernoulli random variable is the result of an experiment that asks a single yes-no question.

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

Definition (Bernoulli Distribution)

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

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

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

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

Example of a Bernoulli random variable: outcome of tossing a single (not necessarily fair) coin.

Generate 10 samples from `scipy.stats.bernoulli`:

```
from scipy.stats import bernoulli
p = 0.25 #probability of success
print(bernoulli.rvs(p, size = 10)) #10 random deviates
[0 0 0 1 1 0 0 0 0 0] #possible output
```

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).

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

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

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

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

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

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

Examples:

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

The number of "point" masses in a volume fraction V_1/V of space with N points in volume V = 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).

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

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

Examples:

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

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

Mean: $E[X] = np$ (demonstrated on following slide)

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

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

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

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

Examples:

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

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

Mean: $E[X] = np$ (demonstrated on following slide) Variance: $Var[X] = np(1-p)$

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

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

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

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

Examples:

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

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

Mean: $E[X] = np$ (demonstrated on following slide) Variance: $Var[X] = np(1-p)$

Both are n times the values for the Bernoulli distribution as expected (n independent Bernoulli trials).

Binomial (discrete, scipy.stats.binom)

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

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

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

Examples:

The number of heads obtained in n tosses of a fair coin = Binomial($n, p=\frac{1}{2}$).
The number of "point" masses in a volume fraction V_1/V of space with N points in volume V = Binomial($N, p=\frac{V_1}{V}$) (Meszaros, A. 1997 A&A 328, 1).

Mean: $E[X]=np$ (demonstrated on following slide) Variance: $Var[X]=np(1-p)$
Both are n times the values for the Bernoulli distribution as expected (n independent Bernoulli trials).

Compute probability of 2 successes in 10 trials

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

0.28156757354736334 #output
```

Expectation value of a Binomial Distribution

Recall:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \quad (2)$$

$$k\binom{n}{k} = n\binom{n-1}{k-1} \quad (3)$$

Expectation value of a Binomial Distribution

Recall:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \qquad (2)$$

$$k\binom{n}{k} = n\binom{n-1}{k-1} \qquad (3)$$

Expectation value for the binomial distribution:

$$E(X) = \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \text{ } (k = 0 \text{ term vanishes})$$

Expectation value of a Binomial Distribution

Recall:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \qquad (2)$$

$$k\binom{n}{k} = n\binom{n-1}{k-1} \qquad (3)$$

Expectation value for the binomial distribution:

$$E(X) = \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \text{ (}k=0\text{ term vanishes)}$$

$$= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k} \qquad \text{using Eq. (3)}$$

Expectation value of a Binomial Distribution

Recall:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \qquad (2)$$

$$k\binom{n}{k} = n\binom{n-1}{k-1} \qquad (3)$$

Expectation value for the binomial distribution:

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \text{ } (k=0 \text{ term vanishes}) \\ &= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k} \qquad \text{using Eq. (3)} \\ &= \sum_{s=0}^{n-1} n\binom{n-1}{s} p^{s+1} (1-p)^{n-s-1} \qquad (\text{setting } s = k-1) \end{aligned}$$

Expectation value of a Binomial Distribution

Recall:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \qquad (2)$$

$$k\binom{n}{k} = n\binom{n-1}{k-1} \qquad (3)$$

Expectation value for the binomial distribution:

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \text{ ($k=0$ term vanishes)} \\
&= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k} \qquad \text{using Eq. (3)} \\
&= \sum_{s=0}^{n-1} n\binom{n-1}{s} p^{s+1} (1-p)^{n-s-1} \qquad \text{(setting $s=k-1$)} \\
&= np \sum_{s=0}^{n-1} \binom{n-1}{s} p^s (1-p)^{n-1-s}
\end{aligned}$$

Expectation value of a Binomial Distribution

Recall:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \qquad (2)$$

$$k\binom{n}{k} = n\binom{n-1}{k-1} \qquad (3)$$

Expectation value for the binomial distribution:

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \ (k=0 \text{ term vanishes}) \\
&= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k} \qquad \text{using Eq. (3)} \\
&= \sum_{s=0}^{n-1} n\binom{n-1}{s} p^{s+1} (1-p)^{n-s-1} \qquad (\text{setting } s=k-1) \\
&= np\sum_{s=0}^{n-1} \binom{n-1}{s} p^{s} (1-p)^{n-1-s} = np \qquad \text{using Eq. (2)}
\end{aligned}$$

Expectation value of a Binomial Distribution

Recall:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \qquad (2)$$

$$k\binom{n}{k} = n\binom{n-1}{k-1} \qquad (3)$$

Expectation value for the binomial distribution:

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \text{ } (k=0 \text{ term vanishes}) \\
&= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k} && \text{using Eq. (3)} \\
&= \sum_{s=0}^{n-1} n\binom{n-1}{s} p^{s+1} (1-p)^{n-s-1} && (\text{setting } s = k-1) \\
&= np\sum_{s=0}^{n-1} \binom{n-1}{s} p^{s} (1-p)^{n-1-s} = np && \text{using Eq. (2)}
\end{aligned}$$

Similarly, we can compute $Var[X]$ using $k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}$