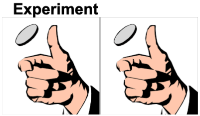
Random variables and Distributions

Phùng Khánh Lâm, MD, PhD

Deparment of Epidemiology, Faculty of Public Health, University of Medicine and Pharmacy at Ho Chi Minh City

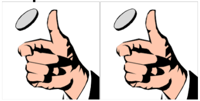
22-07-2020

Recap



Recap

Experiment



Outcome

HH

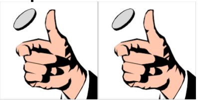
TT

HT

TH

Recap

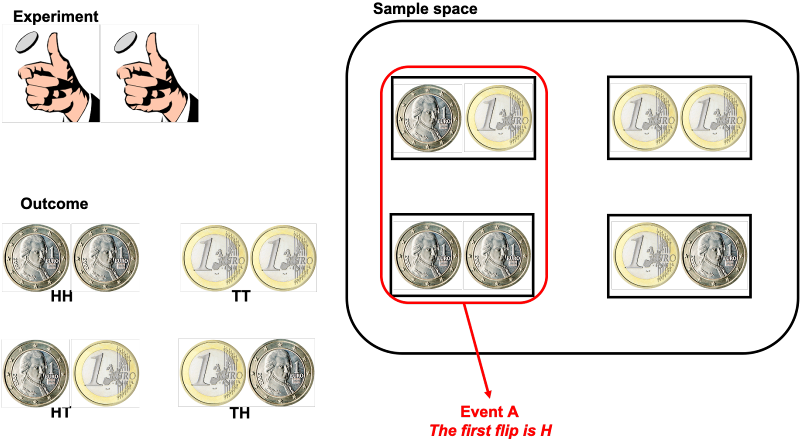
Experiment



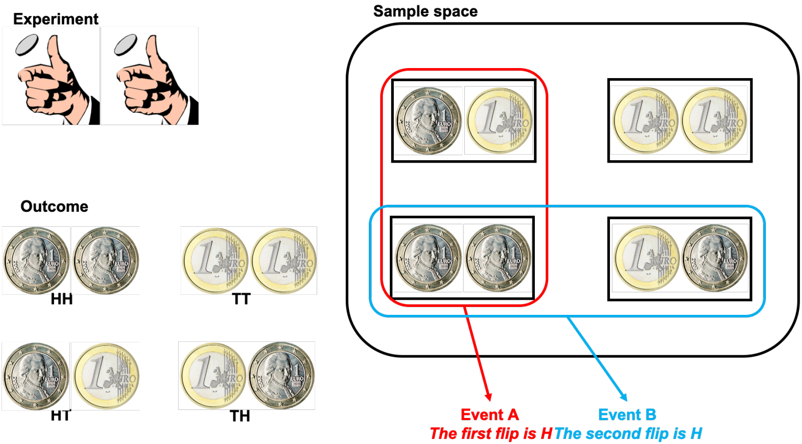
Outcome

Sample space

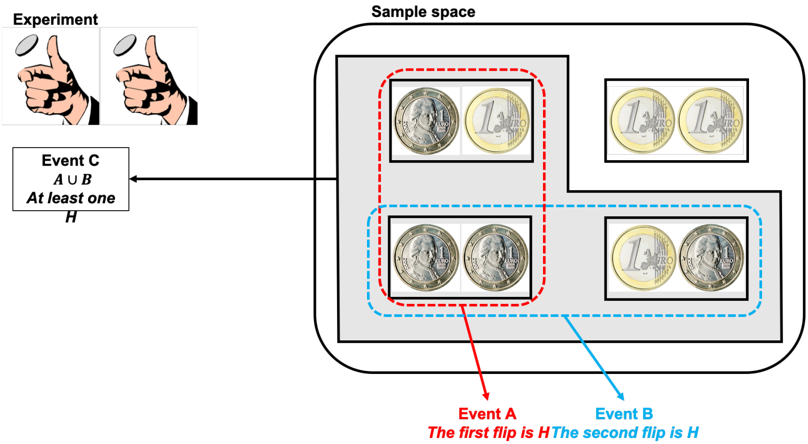
Recap


Experiment
Outcome
HH
TT
HT
TH
Sample space
Event A
The first flip is H

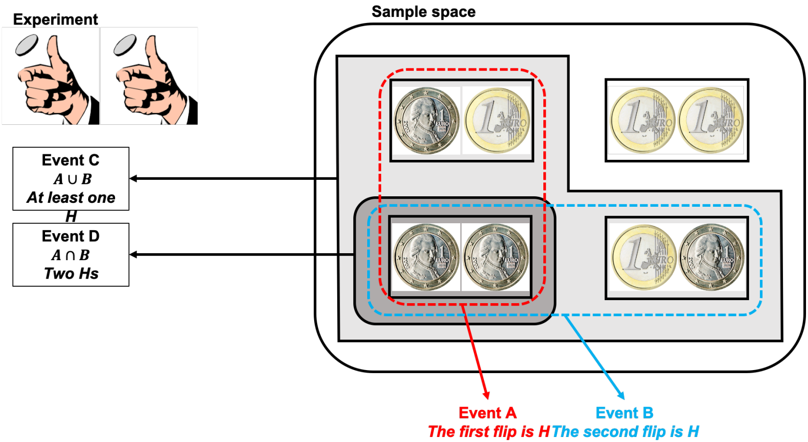
Recap


Experiment
Outcome
HH
TT
HT
TH
Sample space
Event A
The first flip is H
Event B
The second flip is H

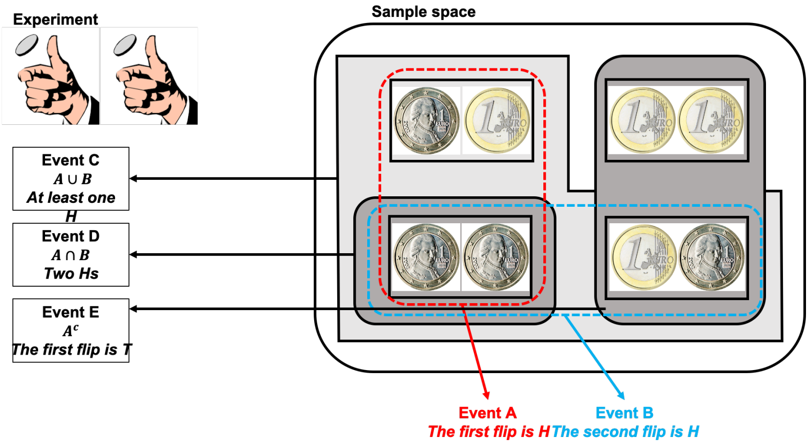
Recap


Experiment
Sample space
Event C
$A \cup B$
At least one H
Event A
The first flip is H
Event B
The second flip is H

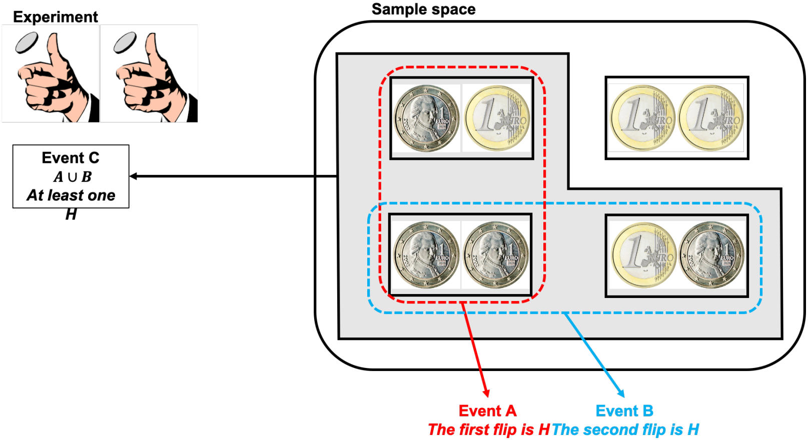
Recap


Experiment
Sample space
Event C
$A \cup B$
At least one H
Event D
$A \cap B$
Two Hs
Event A
The first flip is H
Event B
The second flip is H

Recap


Experiment
Sample space
Event C
$A \cup B$
At least one H
Event D
$A \cap B$
Two Hs
Event E
A^c
The first flip is T
Event A
The first flip is H
Event B
The second flip is H

Recap


Experiment
Sample space
Event C
$A \cup B$
At least one H
Event A
The first flip is H
Event B
The second flip is H

Recap

Random variables

Given an experiment with sample space S, a random variable (r.v.) is a function from the sample space S to the real numbers R.
r.v. X assigns a value $X(s)$ for each outcome s of experiment.

Random variables

Flip 2 coins.
Sample space: $S = \{SS, NN, SN, NS\}$
Example of r.v.:

- X is the number of heads. r.v with 3 possible values: 0, 1, 2:
  $X(NN) = 2$, $X(NS) = X(SN) = 1 X(SS) = 0$

Random variables

Flip 2 coins.
Sample space: $S = \{SS, NN, SN, NS\}$
Example of r.v.:

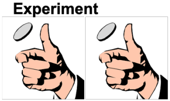
- X is the number of heads. r.v with 3 possible values: 0, 1, 2: $X(NN) = 2$, $X(NS) = X(SN) = 1X(SS) = 0$
- Y is the number of tails. $Y(s) = 2 - X(s)$

Random variables

Flip 2 coins.
Sample space: $S = \{SS, NN, SN, NS\}$
Example of r.v.:

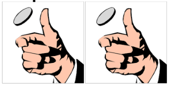
- X is the number of heads. r.v with 3 possible values: 0, 1, 2: $X(NN) = 2$, $X(NS) = X(SN) = 1X(SS) = 0$
- Y is the number of tails. $Y(s) = 2 - X(s)$
- $I = 1$ if heads first and $I = 0$ otherwise. $I(NN) = I(NS) = 1$, $I(SN) = I(SS) = 0$

Random variables



Random variables

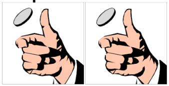
Experiment



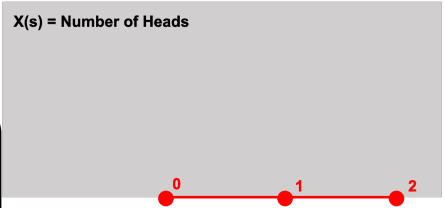
Sample space

Random variables

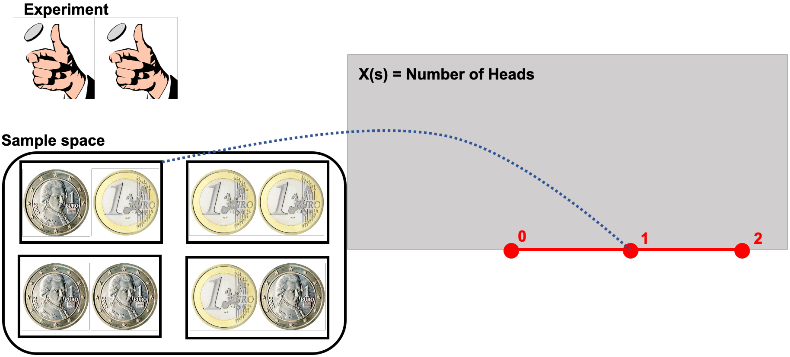
Experiment



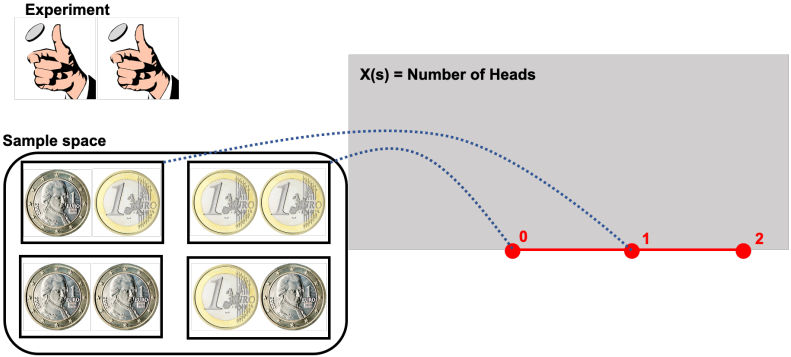
Sample space



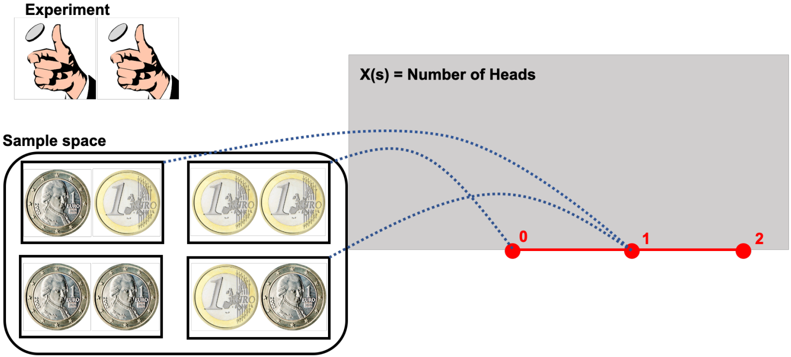
Random variables



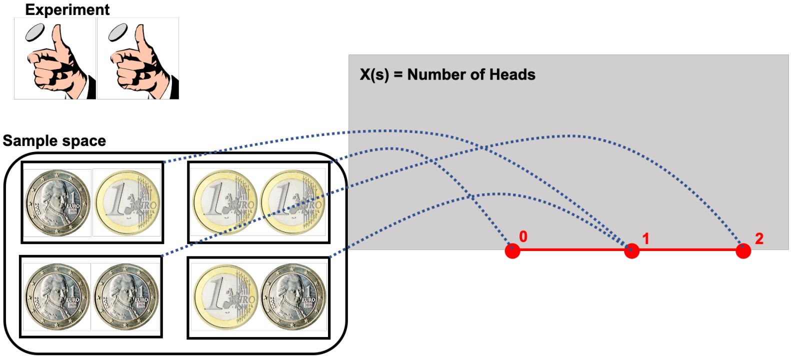
Random variables



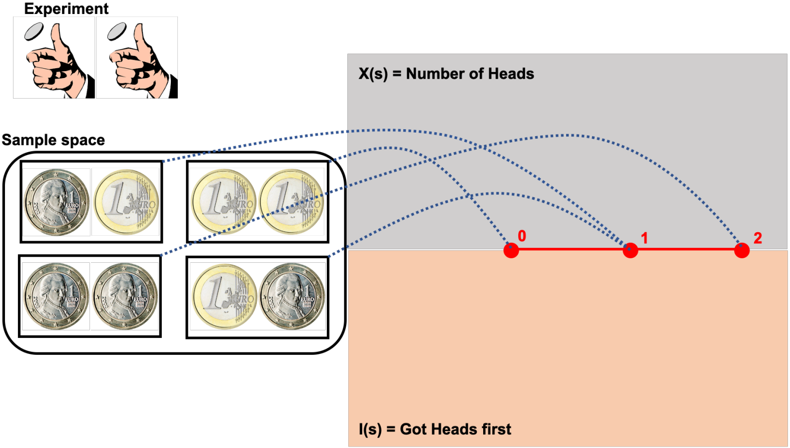
Random variables



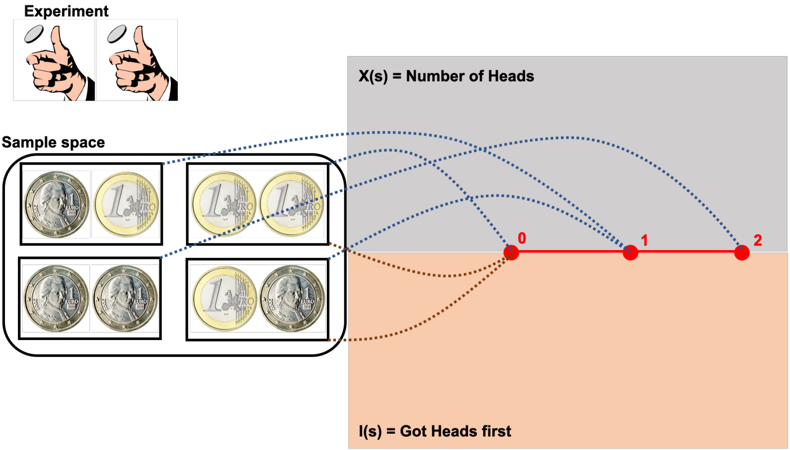
Random variables



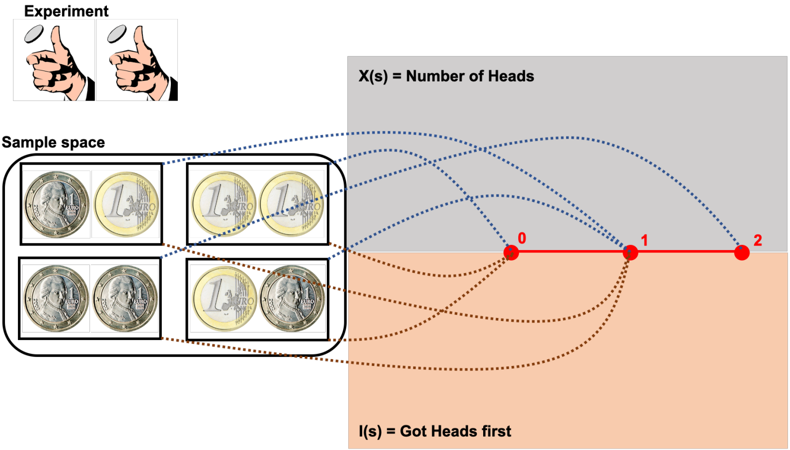
Random variables



Random variables



Random variables



Types of random variables

- Discrete random variables: possible values of X is finite or countable infinite
- Continuous random variables: can take on any real value in an interval

Probability distribution

Distribution of a r.v. specifies the probabilities of all events associated with that r.v.

Probability mass function (PMF)

PMF of a discrete r.v. X is the function p_X given by $p_X(x) = P(X = x)$. This function is positive if x is in the support of X, and 0 otherwise.

Properties of PMF p_X:

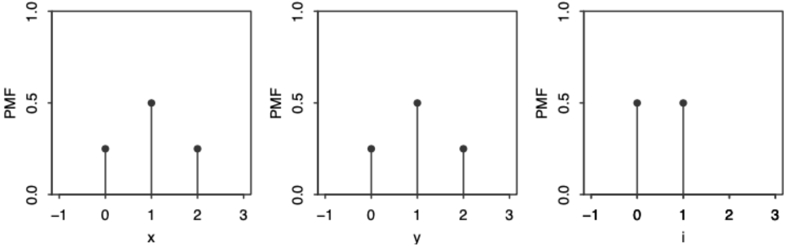
- Non-negative: $p_X(x) > 0$ for some x and $p_X(x) = 0$ otherwise
- Sum to 1: $\Sigma_{j=1}^{\infty} p_X(x_j) = 1$

Example 1

Find PMF of the following r.v.:

- X is the number of heads
- Y is the number of tails
- $I = 1$ if heads first and $I = 0$ otherwise

Example 1

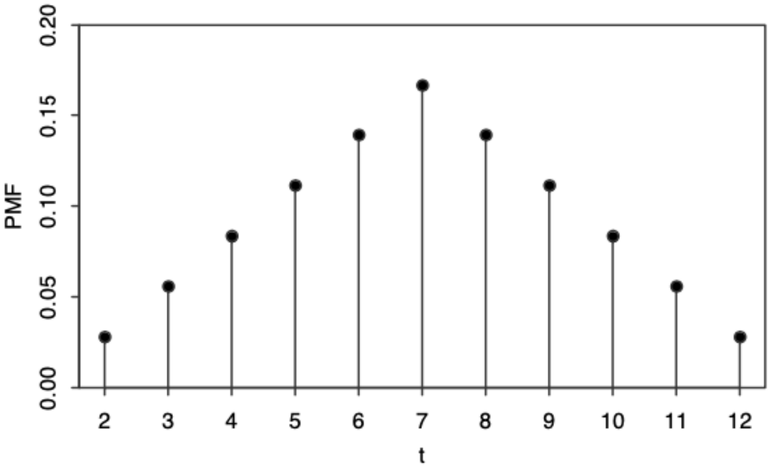


Example 2

Roll 2 fair dices A and B. Calculate PMF of the following r.v.:

- X is the score of dice A.
- Y is the score of dice B
- $Z = 7 - X$
- $T = X + Y$

Example 2



Example 2

Calculate probability that the total score of 2 dices (T) has value from 1 to 4

Example 2

Calculate probability that the total score of 2 dices (T) has value from 1 to 4
$P(1 \leq T \leq 4) = P(T = 2) + P(T = 3) + P(T = 4) = 6/36$

Bernoulli distribution

A r.v. X has Bernoulli distribution with parameter p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$, with $0 < p < 1$. Denote as $X \sim Bern(p)$

Bernoulli distribution

A r.v. X has Bernoulli distribution with parameter p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$, with $0 < p < 1$. Denote as $X \sim Bern(p)$

Any r.v. whose possible values are 0 and 1 has a $Bern(p)$, with p is the probability of the r.v. equalling 1.

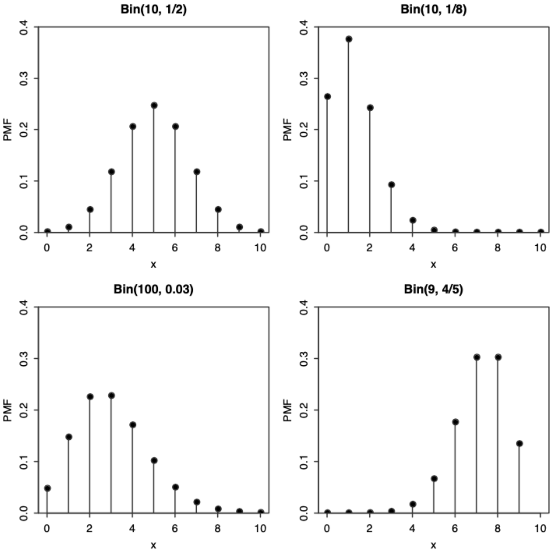
An experiment that can result in either a "success" or a "failure" (but not both) is called a **Bernoulli trial**. A Bernoulli random variable can be thought of as the indicator of success in a Bernoulli trial.

Binomial distribution

Perform n independent Bernoulli trials, each with the same success probability p. Let X be the number of successes. The distribution of X is the Binomial distribution with parameters n and p. We write $X \sim Bin(n, p)$
If $X \sim Bin(n, p)$, PMF of X is

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

Binomial distribution



Cumulative distribution functions (CDF)

CDF of a r.v. X is a function F_X with $F_X(x) = P(X \leq x)$

Cumulative distribution functions (CDF)

CDF of a r.v. X is a function F_X with $F_X(x) = P(X \leq x)$

- Increasing: if $x_1 \leq x_2$ then $F(x_1) \leq F(x_2)$

Cumulative distribution functions (CDF)

CDF of a r.v. X is a function F_X with $F_X(x) = P(X \leq x)$

- Increasing: if $x_1 \leq x_2$ then $F(x_1) \leq F(x_2)$
- Convergence to 0 and 1 at the limits: $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$

Expectation

Expectation helps to summarize distribution of random variables.

Expectation

Expectation helps to summarize distribution of random variables.

- Expected value: describe the center of the distribution

Expectation

Expectation helps to summarize distribution of random variables.

- Expected value: describe the center of the distribution
- Variance: describe the spread of the distribution

Expected value

Expected value of a discrete r.v. X with possible values $x_1, x_2, \ldots$ is defined as:

$$E(X) = \Sigma_{j=1}^{\infty} x_j P(X = x_j)$$

Expected value

Expected value of a discrete r.v. X with possible values $x_1, x_2, \ldots$ is defined as:

$$E(X) = \Sigma_{j=1}^{\infty} x_j P(X = x_j)$$

For all r.v. X, Y and constant c: $E(X + Y) = E(X) + E(Y)$ and $E(cX) = cE(X)$

Expected value

Expected value of a discrete r.v. X with possible values $x_1, x_2, \ldots$ is defined as:

$$E(X) = \Sigma_{j=1}^{\infty} x_j P(X = x_j)$$

For all r.v. X, Y and constant c: $E(X + Y) = E(X) + E(Y)$ and $E(cX) = cE(X)$
Expected value of Bernoulli distribution: if $X \sim Bern(p)$ then $E(X) = p$

Expected value

Expected value of a discrete r.v. X with possible values $x_1, x_2, \ldots$ is defined as:

$$E(X) = \Sigma_{j=1}^{\infty} x_j P(X = x_j)$$

For all r.v. X, Y and constant c: $E(X + Y) = E(X) + E(Y)$ and $E(cX) = cE(X)$

Expected value of Bernoulli distribution: if $X \sim Bern(p)$ then $E(X) = p$

Expected value of Binomial distribution: if $X \sim Bin(n, p)$ then $E(X) = np$

Variance

Variance and standard deviation of a r.v. X are defined as:

$$Var(X) = E(X - EX)^2 = E(X^2) - (EX)^2$$

$$SD(X) = \sqrt{Var(X)}$$

Variance

Variance and standard deviation of a r.v. X are defined as:

$$Var(X) = E(X - EX)^2 = E(X^2) - (EX)^2$$
$$SD(X) = \sqrt{Var(X)}$$

Properties of variance:

- $Var(X + c) = Var(X)$
- $Var(cX) = c^2 Var(X)$
- If X and Y are independence,
  $Var(X + Y) = Var(X) + Var(Y)$

Variance

Variance and standard deviation of a r.v. X are defined as:

$$Var(X) = E(X - EX)^2 = E(X^2) - (EX)^2$$
$$SD(X) = \sqrt{Var(X)}$$

Properties of variance:

- $Var(X + c) = Var(X)$
- $Var(cX) = c^2 Var(X)$
- If X and Y are independence, $Var(X + Y) = Var(X) + Var(Y)$

Variance of Bernoulli distribution: if $X \sim Bern(p)$ then $Var(X) = p(1 - p)$

Variance

Variance and standard deviation of a r.v. X are defined as:

$$Var(X) = E(X - EX)^2 = E(X^2) - (EX)^2$$
$$SD(X) = \sqrt{Var(X)}$$

Properties of variance:

- $Var(X + c) = Var(X)$
- $Var(cX) = c^2 Var(X)$
- If X and Y are independence,
  $Var(X + Y) = Var(X) + Var(Y)$

Variance of Bernoulli distribution: if $X \sim Bern(p)$ then
$Var(X) = p(1 - p)$
Variance of Binomial distribution: if $X \sim Bin(n, p)$ then
$Var(X) = np(1 - p)$

Probability density function (PDF)

For a continuous r.v. X with CDF F, the probability density function (PDF) of X is the derivative f of the CDF, given by $f(x) = F'(x)$.

Probability density function (PDF)

For a continuous r.v. X with CDF F, the probability density function (PDF) of X is the derivative f of the CDF, given by $f(x) = F'(x)$.
Of note, $f(x)$ is not a probability ($f(x)$ can > 1)

Probability density function (PDF)

For a continuous r.v. X with CDF F, the probability density function (PDF) of X is the derivative f of the CDF, given by $f(x) = F'(x)$.
Of note, $f(x)$ is not a probability ($f(x)$ can > 1)
X is a r.v. with PDF f. Then CDF của X is:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

và probability that X is from a to b is:

$$P(a < X \leq b) = \int_{a}^{b} f(x)dx$$

Probability density function (PDF)

Properties of PDF

- Non-negative: $f(x) \geq 0$
- Integrates to 1: $\int_{-\infty}^{\infty} f(x)dx = 1$

Probability density function (PDF)

Properties of PDF

- Non-negative: $f(x) \geq 0$
- Integrates to 1: $\int_{-\infty}^{\infty} f(x)dx = 1$

Expected value of r.v. X with PDF f is

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Standard normal distribution

A continuous r.v. Z is said to have the standard Normal distribution if its PDF φ is given by

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp^{-z^2/2}, -\infty < z < \infty$$

Z has mean 0 and variance 1. Denote as $Z \sim N(0, 1)$

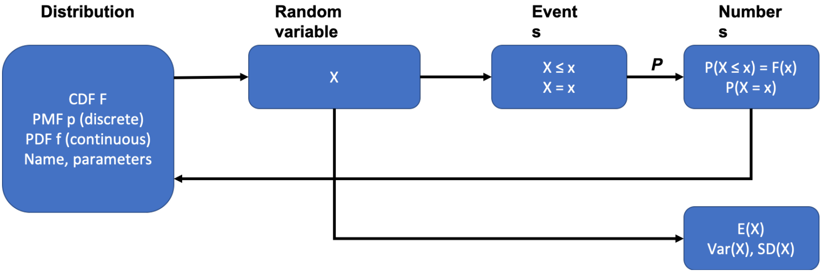
Normal distribution

If $Z \sim N(0,1)$ then

$$X = \mu + \sigma Z$$

has the Normal distribution with mean μ và variance σ^2. Denote as $X \sim N(\mu, \sigma^2)$

Recap


Distribution
Random variable
Event s
Number s
CDF F
PMF p (discrete)
PDF f (continuous)
Name, parameters
X
X ≤ x
X = x
P
P(X ≤ x) = F(x)
P(X = x)
E(X)
Var(X), SD(X)