Overview over probability theory

05. May 2021.

Random Variables

- Random variable X
- A random variable is a function that assigns a number to each possible outcome of a probabilistic experiment.
- A random variable is a quantity taking different values on different occasions
- The set of possible values that a random variable X can take is called the state space $\mathbb{X}$ or the range of X
- $p(x) = p(X = x)$, $x \in \mathbb{X}$
- $0 \leq p(x) \leq 1$
- $\sum_{x \in X} p(x) = 1$

Types of random variables

Types of random variables

- discrete random variables obtained by counting values for which there are no in-between values (such as the integers 0, 1, 2, ...); e.g. the number of female students in the class, the number of applicants who have applied for a vacant position at a company.
- continuous random variables obtained from data that can take infinitely many values, the set of possible values is the set of real numbers R, one interval, or a disjoint union of intervals on the real line; e.g. the height of a randomly selected student, the length of time it takes to go by car from Las Vegas to Miami. In general, quantities such as volume, temperature, pressure, height, weight, density, and distance are examples of continuous random variables.

Probability Distributions

How to describe the behaviour of a random variable

- Suppose that a random variable X can only takes values $\{x_1, x_2, ..., x_n\}$, i.e. the state space of X is the set of n values $\mathbb{X} = \{x_1, x_2, ..., x_n\}$
- The Probability Function of a discrete random variable X is the function $p(x)$ satisfying $p(X = x_i)$, for all $x_i \in \mathbb{X}$
- $0 \leq p(x) \leq 1$
- $\sum_{x \in X} p(x) = 1$

Event	Probability
$X = x_1$	$\text{Pr}(X = x_1)$
$X = x_2$	$\text{Pr}(X = x_2)$
$\vdots$	$\vdots$
$X = x_n$	$\text{Pr}(X = x_n)$

The cumulative distribution function

The cumulative distribution function (cdf)

The cumulative distribution function of a random variable X is the function $F_X(x)$ of x given by $F_X(x) = P(X \leq x)$, for all values $x \in \mathbb{X}$.

- The cdf is sometimes given the alternative name of distribution function.
- For a discrete rv X we have $F_X(x) = \sum_{y \leq x} P(X = y)$

Example

- Assume that a rv X has range $\mathbb{X} = \{0, 1, 2, ...\}$.
- Then, we have $F_X(3) = P(X \leq 3) = p(0) + p(1) + p(2) + p(3)$
- $p(2) = F_X(2) - F_X(1)$

- Suppose that a coin is tossed two times.
- X represents the number of heads that can come up
- the state space $\mathbb{X} = \{HH, HT, TH, TT\}$
- in the case of HH H (i.e., 2 heads), $X = 2$ while n the case of TH (1 head), $X = 1$

Sample Point	HH	HT	TH	TT
X	2	1	1	0

The coin is fair

$$P(HH) = \frac{1}{4} \qquad P(HT) = \frac{1}{4} \qquad P(TH) = \frac{1}{4} \qquad P(TT) = \frac{1}{4}$$

$$P(X = 0) = P(TT) = \frac{1}{4}$$

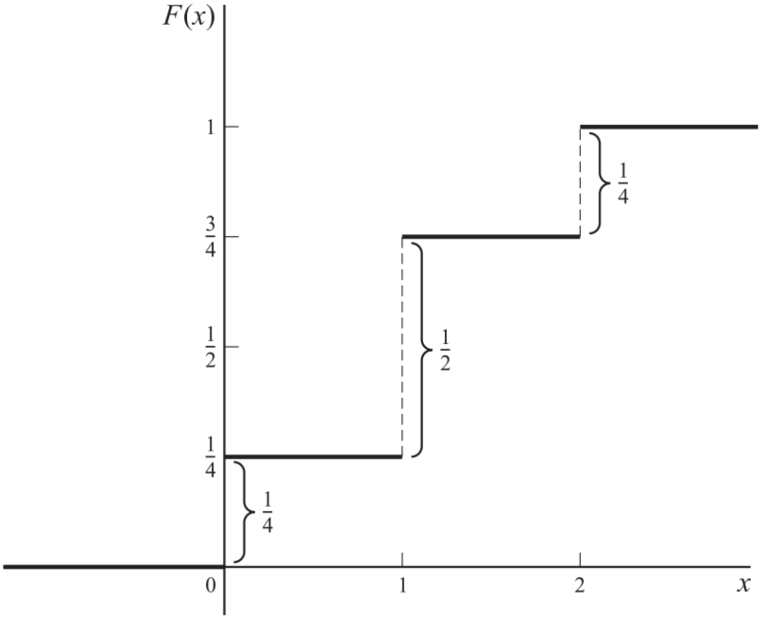
$$P(X = 1) = P(HT \cup TH) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(X = 2) = P(HH) = \frac{1}{4}$$

x	0	1	2
$f(x)$	1/4	1/2	1/4

The distribution function for the random variable X

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ \frac{1}{4} & 0 \leq x < 1 \\ \frac{3}{4} & 1 \leq x < 2 \\ 1 & 2 \leq x < \infty \end{cases}$$



Joint Distributions

Let X and Y be two discrete random variables with the state space $\mathbb{X} = \{x_1, x_2, ..., x_m\}$ and $\mathbb{Y} = \{y_1, y_2, ..., y_n\}$ respectively.

- Then the probability of the event that $X = x_j$ and $Y = y_k$ is given by $P(X = x_j, Y = y_k) = f(x_j, y_k)$
- The probability that $X = x_j$ is obtained by adding all entries in the row corresponding to x_j and is given by $P(X = x_j) = \sum_{k=1}^{n} f(x_j, y_k)$

$X \backslash Y$	y_1	y_2	$\dots$	y_n	Totals $\downarrow$
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\dots$	$f(x_1, y_n)$	$f_1(x_1)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\dots$	$f(x_2, y_n)$	$f_1(x_2)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	$\dots$	$f(x_m, y_n)$	$f_1(x_m)$
Totals $\rightarrow$	$f_2(y_1)$	$f_2(y_2)$	$\dots$	$f_2(y_n)$	1

$$\sum_{j=1}^{m} f_1(x_j) = 1 \quad \sum_{k=1}^{n} f_2(y_k) = 1$$

$$\sum_{j=1}^{m} \sum_{k=1}^{n} f(x_j, y_k) = 1$$

Fundamental rules

Given two events A, B

- $p(A \cup B) = p(A) + p(B) - p(A \cap B)$
- $p(A, B) = p(A \cap B) = p(A|B)p(B)$ the product rule

Given two independent events A, B

- $p(A|B) = p(A)$
- $p(B|A) = p(B)$
- $p(A, B) = p(A \cap B) = p(A)p(B)$

The two events A, B are mutually exclusive events if they cannot occur at the same time, i.e. $p(A, B) = 0$.

- Let A = the event of getting **at most one tail**. (At most one tail means zero or one tail.) Then A can be written as $\{HH, HT, TH\}$. The outcome HH shows zero tails. HT and TH each show one tail.
- Let B = the event of getting all tails. B can be written as $\{TT\}$. B is the **complement** of A, so $B = A'$. Also, $P(A) + P(B) = P(A) + P(A') = 1$.
- The probabilities for A and for B are $P(A) = \frac{3}{4}$ and $P(B) = \frac{1}{4}$.
- Let C = the event of getting all heads. $C = \{HH\}$. Since $B = \{TT\}$, $P(B \text{ AND } C) = 0$. B and C are mutually exclusive. (B and C have no members in common because you cannot have all tails and all heads at the same time.)
- Let D = event of getting **more than one** tail. $D = \{TT\}$. $P(D) = \frac{1}{4}$
- Let E = event of getting a head on the first roll. (This implies you can get either a head or tail on the second roll.) $E = \{HT, HH\}$. $P(E) = \frac{2}{4}$
- Find the probability of getting **at least one** (one or two) tail in two flips. Let F = event of getting at least one tail in two flips. $F = \{HT, TH, TT\}$. $P(F) = \frac{3}{4}$

Mean and Variance

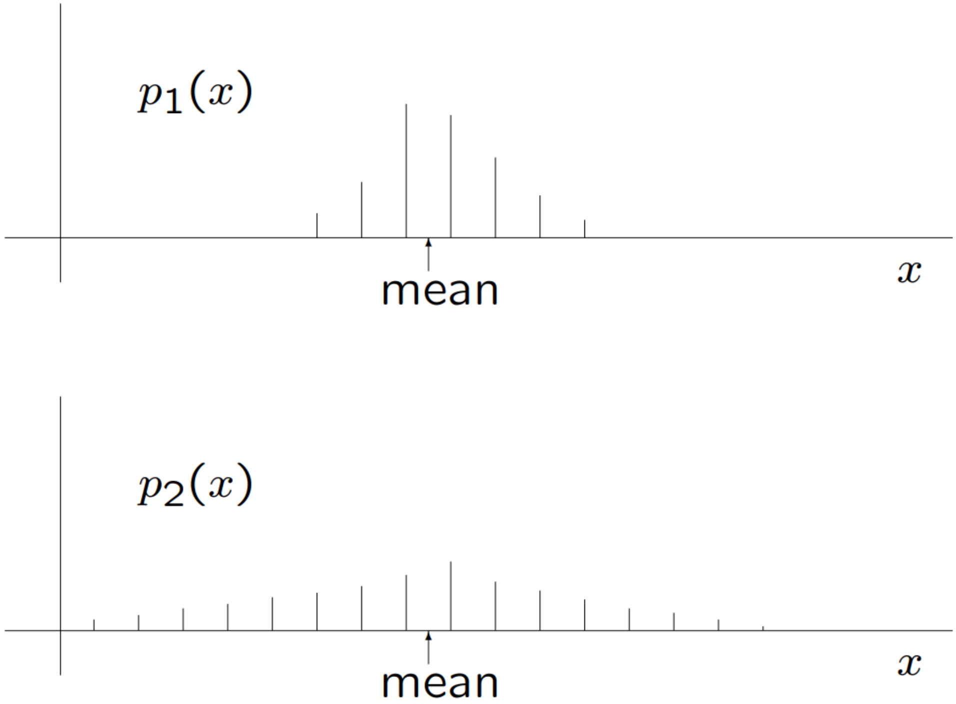
The mean or expectation of a discrete rv X, $E(X)$, is defined as

$$E(X) = \sum_{x \in \mathbb{X}} xp(x)$$

Properties

- If c is a constant, then $E(c) = 0$
- $E(X + Y) = E(X) + E(Y)$
- $E(aX + b) = aE(X) + b$, where a and b are constants
- Non-negativity: If $X \geq 0$ (a.s.), then $E(X) \geq 0$.
- Monotonicity: If $X \leq Y$ and both $E(X)$ and $E(Y)$ exist, then $E(X) \leq E(Y)$.
- If X and Y are independent, then $E(XY) = E(X)E(Y)$

Measures of dispersion



Variance

The variance, $V(X)$ or $Var(X)$, of a discrete random variable X is defined as

$$V(X) = Var(X) = E(X - E(X))^2.$$

Properties

- $Var(X) \geq 0$
- If a is constant $Var(a) = 0$ and $Var(aX) = a^2 Var(X)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- If X and Y are independent then $Var(X + Y) = Var(X) + Var(Y)$

Measures of dispersion

x	1	2	3	4	5
$p(x) = \Pr(X = x)$	0.1	0.1	0.2	0.4	0.2

$\mathrm{E}(X) = \sum_x x \Pr(X = x)$, so

$$\mathrm{E}(X) = (1 \times 0.1) + (2 \times 0.1) + (3 \times 0.2) \\ +(4 \times 0.4) + (5 \times 0.2) = 3.5.$$

$\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2$, and

$$\mathrm{E}(X^2) = (1^2 \times 0.1) + (2^2 \times 0.1) + (3^2 \times 0.2) \\ +(4^2 \times 0.4) + (5^2 \times 0.2) = 13.7\,,$$

so $\mathrm{Var}(X) = 13.7 - (3.5)^2 = 1.45$.

Standard deviation of X: $\sqrt{1.45}$, or 1.20.

The conditional probability

$p(A|B) = p(A, B)/p(B)$

Bayes rule

$$p(X = x|Y = y) = \frac{p(X = x, Y = y)}{p(Y = y)}$$

the chain rule of probability

$p(X_1, ...X_M) =$
$p(X_1)p(X_2|X_1)p(X_3|X_1, X_2) \cdots p(X_M|X_1, X_2, ..., X_{M-1})$

Fundamental rules

Given two events A, B

- $p(A \cup B) = p(A) + p(B) - p(A \cap B)$
- $p(A, B) = p(A \cap B) = p(A|B)p(B)$ the product rule

Given a joint distribution on two events $p(A, B)$, the marginal distribution is defined as

- $p(A) = \sum_b p(A, B = b) = \sum_b p(A|B = b)p(B = b)$
- the sum rule or the rule of total probability