Lecture note 4: Random variables

Foundation of Probability Theory/STA 203

Zhuosong ZHANG

Department of Statistics and Data Science, SUSTech

Fall, 2023

Learning Objectives: Random Variables

- Define a random variable and explain how it relates to a probability distribution.
- Distinguish between discrete and continuous random variables and give examples of each.
- Define the probability mass function (PMF) and probability density function (PDF) of a random variable, and use them to compute probabilities.
- Define the cumulative distribution function (CDF) of a random variable, and use it to compute probabilities and quantiles.
- Define the expected value (or mean) and variance of a random variable, and compute them for both discrete and continuous random variables.

Learning Objectives: Random Variables

- Explain the properties of expected value and variance, including linearity and additivity, and use them to compute expected values and variances of linear combinations of random variables.
- Define covariance and correlation between two random variables, and explain how they relate to independence.

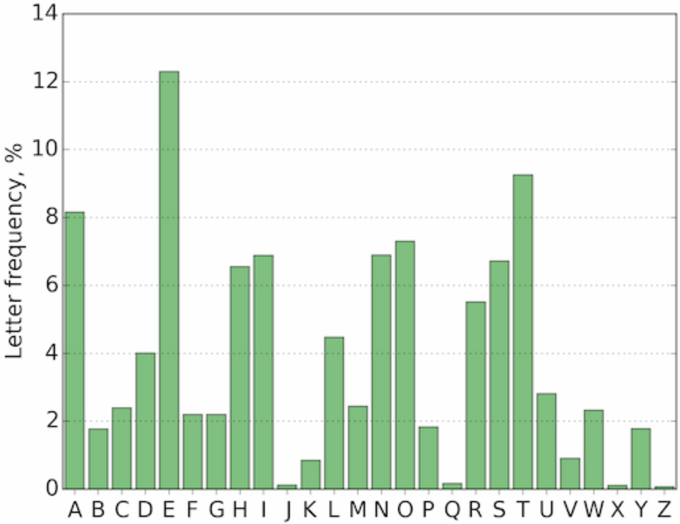
Random variables

Introduction

- How long do products last?
- Should you expect your computer to die just after the warranty runs out?
- How can you reduce your risk for developing hepatitis?
- Businesses, medical researchers, and other scientists all use probability to determine risk factors to help answer questions like these.

To do that, they model the probability of outcomes using a special kind of variable — a random variable. Using random variables can help us talk about and predict random behavior.

What is the average?



Why random variables

- Random variables allow us to quantify and analyze the outcomes of random events or experiments in a mathematical way.
- They provide a way to convert the outcomes of a random experiment into numbers, which can be manipulated using mathematical operations such as addition and multiplication.
- Random variables allow us to calculate probabilities of events associated with the outcomes of a random experiment, such as the probability of getting a certain result or the probability of a certain range of outcomes.
- They allow us to compute summary statistics such as means, variances, and correlations, which provide a more comprehensive understanding of the distribution of outcomes than just looking at individual outcomes.

Why random variables

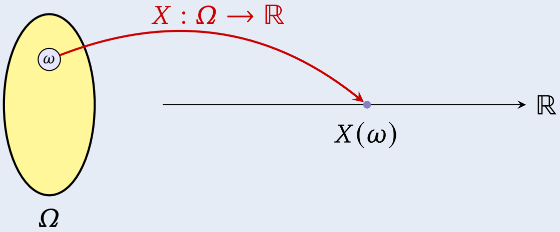
- Random variables provide a framework for modeling and analyzing complex systems and processes that involve random events or inputs.

Definition of random variables

Definition 1 (Random variable)

A random variable is a real-valued function defined on the sample space.

A random variable is a mapping from the outcomes in the sample space to numbers on the real line. We can think of a random variable X as a translator that translates a statement to a number.



We usually use capital letters X, Y, Z, W etc. to represent random variables.

Example

Example 2 (Coin tossing)

Suppose that our experiment consists of tossing 3 fair coins. If we let Y denote the number of tails that appear. Figure out the sample space and the definition of Y.

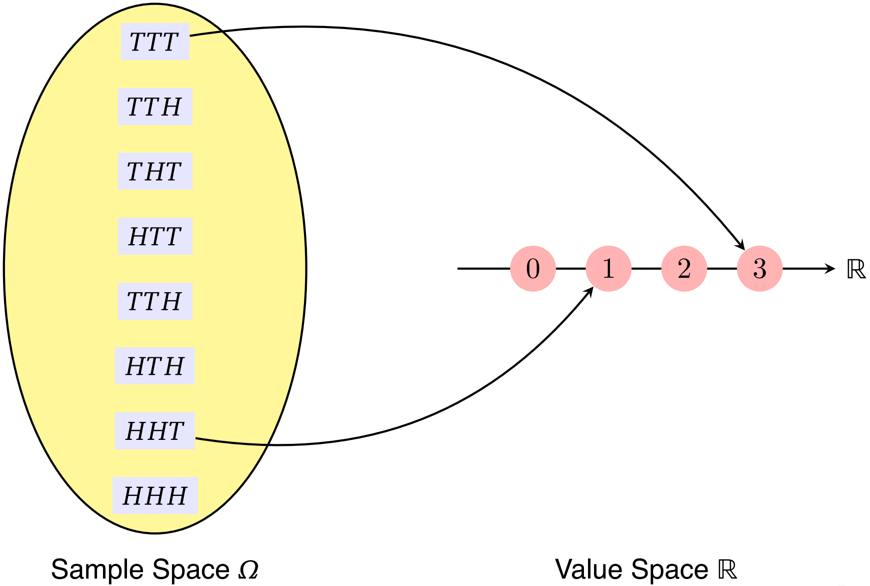
Example

Solution.

The sample space can be written as $\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$, and the random variable $Y : \Omega \to \mathbb{R}$ can be defined as

$$Y(\omega) = \begin{cases} 3 & \text{if } \omega = TTT, \\ 2 & \text{if } \omega \in \{TTH, THT, HTT\}, \\ 1 & \text{if } \omega \in \{THH, HTH, HHT\}, \\ 0 & \text{if } \omega = HHH. \end{cases}$$

Example


TTT
TTH
THT
HTT
TTH
HTH
HHT
HHH
0
1
2
3
ℝ
Sample Space Ω
Value Space ℝ

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 3 (Tossing a Coin)

Suppose we perform the simple experiment of tossing a fair coin.

- The sample space of this experiment is $\Omega = \{H, T\}$.
- The σ-field is the power set of Ω, which is $\mathscr{F} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$.
- We can define a random variable X that takes the value 1 if the outcome is heads and 0 otherwise. So, $X(H) = 1$ and $X(T) = 0$.

Random Variables - Real World Examples

Example 4 (Rolling a Die)

Suppose we perform the experiment of rolling a six-sided die.

- The sample space of this experiment is $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- The σ-field is the power set of Ω.
- We can define a random variable X that takes the value 1 if the outcome is an even number and 0 otherwise. So, $X(2) = X(4) = X(6) = 1$ and $X(1) = X(3) = X(5) = 0$.

Random Variables - Real World Examples

Example 5 (Weather Forecast)

Suppose we are interested in the weather condition (sunny, cloudy, rainy) of a certain day in a city.

- The sample space of this experiment is $\Omega = \{$sunny, cloudy, rainy$\}$.
- The σ-field is the power set of Ω.
- We can define a random variable X that takes the value 1 if the weather is sunny and 0 otherwise. So, $X(\text{sunny}) = 1$, $X(\text{cloudy}) = X(\text{rainy}) = 0$.

Because the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable.

Example 6 (Coin tossing)

Let $\mathbb{P}$ be the classical probability defined on Ω. We have

$$\mathbb{P}\{Y = 3\} = \mathbb{P}\{TTT\} = \frac{1}{8}$$

$$\mathbb{P}\{Y = 2\} = \mathbb{P}\{TTH, THT, HTT\} = \frac{3}{8},$$

$$\mathbb{P}\{Y = 1\} = \mathbb{P}\{THH, HTH, HHT\} = \frac{3}{8},$$

$$\mathbb{P}\{Y = 0\} = \mathbb{P}\{HHH\} = \frac{1}{8}.$$

Example

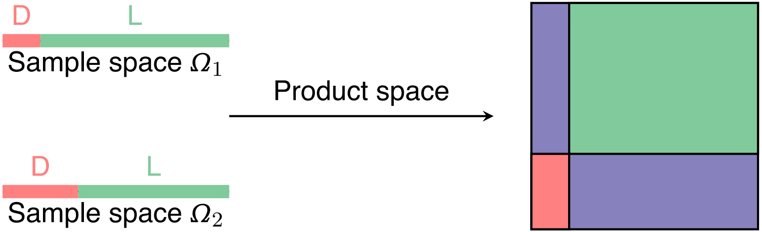
南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 7

A life insurance agent has 2 elderly clients, each of whom has a life insurance policy that pays $100,000 upon death. Let A be the event that the younger one dies in the following year, and let B be the event that the older one dies in the following year. Assume that A and B are independent, with respective probabilities $\mathbb{P}(A) = .05$ and $\mathbb{P}(B) = .10$. If X denotes the total amount of money (in units of $100, 000) that will be paid out this year to any of these clients' beneficiaries, then X is a random variable that takes on one of the possible values $0, 1, 2$ with respective probabilities

$$\mathbb{P}(X = 0) = \mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c)\,\mathbb{P}(B^c) = (0.95)(0.9) = 0.855,$$
$$\mathbb{P}(X = 1) = \mathbb{P}(A \cap B^c) + \mathbb{P}(A^c \cap B) = (0.05)(0.9) + (0.95)(0.1) = 0.140,$$
$$\mathbb{P}(X = 2) = \mathbb{P}(A \cap B) = (0.05)(0.1) = 0.005.$$

Example


D
L
Sample space Ω_1
Product space
D
L
Sample space Ω_2

Discrete random variables

Discrete and continuous random variables

Definition 8 (Discrete random varibles)

A random variable is called discrete if its **support** (the set of values that it can take, usually denoted by $\mathbb{S}$) is either finite or countably infinite.

Continuous random variables, an informal definition

A random variable is typically a continuous random variable if there are no "gaps" in its support.

Example 9

(a) Lifetime of a bulb.

(b) Number of customers visiting a bank during 9am to 10 am.

(c) The height of a random selected person...

Example 10 (Ball selection)

Three balls are to be randomly selected without replacement from an urn containing 20 balls numbered 1 through 20. If we bet that at least one of the balls that are drawn has a number as large as or larger than 17, what is the probability that we win the bet?

Solution.

Let X denote the largest number selected. Then, X is a random variable taking on one of the values $3, 4, \ldots, 20$. If we assume that each of the $\binom{20}{3}$ possible selections are equally likely to occur, then

$$\mathbb{P}\{X = i\} = \frac{\binom{i-1}{2}}{\binom{20}{3}} \quad \text{for } i = 3, \ldots, 20.$$

Therefore, let E be the event that we win the bet, then $E = \{X \geqslant 17\}$, and

$$\mathbb{P}(E) = \mathbb{P}\{X \geqslant 17\} = \mathbb{P}\{X = 17\} + \cdots + \mathbb{P}\{X = 20\} \approx 0.508.$$ ■

Examples

Example 11 (Coin flipping)

Independent trials consisting of the flipping of a coin having probability p of coming up heads are continually performed until either a head occurs or a total of n flips is made. If we let X denote the number of times the coin is flipped, then X is a random variable taking on one of the values $1, 2, 3, \ldots, n$ with respective probabilities

$$\mathbb{P}\{X = k\} = (1-p)^{k-1}p \quad \text{for } k = 1, \ldots, n-1,$$

and $\mathbb{P}\{X = n\} = (1-p)^{n-1}$.

Probability mass function

Probability mass function

Definition 12

Let $\mathbb{S}$ be the support of X. The probability mass function p of X is defined as $p : \mathbb{S} \to [0, 1]$,

$$p(a) = \mathbb{P}\{X = a\} \quad \text{for } a \in \mathbb{S}.$$

Proposition 13

Assume that X is a discrete random variable taking values on the set $\mathbb{S} = \{s_i, i \in J\}$, then

(i) Non-negativity: $p(s_i) \geqslant 0$ for $i \in J$ and $p(a) = 0$ for $a \in \mathbb{R} \setminus \mathbb{S}$.

(ii) Normalization: $\sum_{i \in J} p(s_i) = 1$.

Graphical format of p

It is often instructive to present the probability mass function in a graphical format by plotting $p(s_i)$ on the y-axis against s_i on the x-axis.

Example 14

If the probability mass function of Y is

$$\mathbb{P}\{Y = 0\} = \mathbb{P}\{Y = 3\} = \frac{1}{8},$$
$$\mathbb{P}\{Y = 1\} = \mathbb{P}\{Y = 2\} = \frac{3}{8},$$

then the graph is ➡

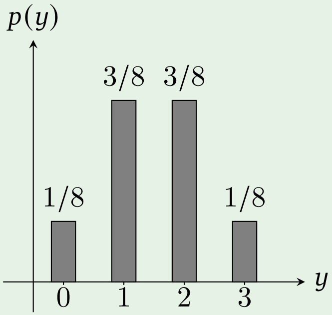


Table format of p

The probability distribution of X can also be listed in a table.

Example (Cont'd)

For example, the probability distribution of Y is the following table:

y	0	1	2	3
$p(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Example

Example 15

The probability mass function of a random variable X is given by

$$p(i) = c\frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \ldots$$

where λ is some positive value. Find

(a) the value of c,

(b) $\mathbb{P}\{X = 0\}$, and

(c) $\mathbb{P}\{X > 2\}$.

Solution.

(a) By the normalization property, we have

$$c \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} = 1$$

which implies that

$$ce^{\lambda} = 1 \implies c = e^{-\lambda}.$$

(b) We have

$$\mathbb{P}\{X = 0\} = (e^{-\lambda})\frac{\lambda^0}{0!} = e^{-\lambda}.$$

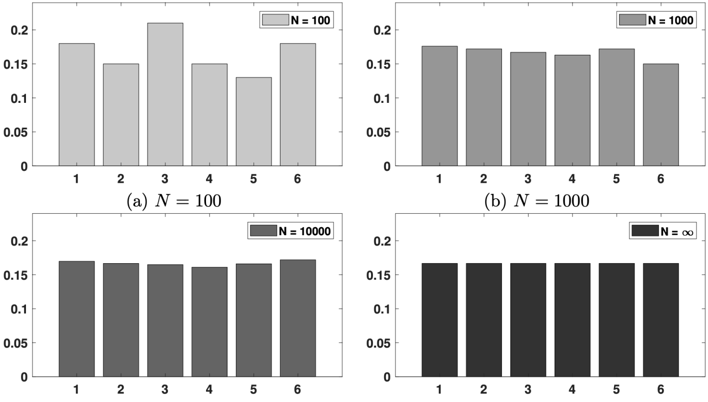
(c) Because $\{X > 2\} = (\{X = 0, 1, 2\})^c$, we have

$$\begin{aligned}
&\mathbb{P}\{X > 2\} \\
&\quad = 1 - \mathbb{P}\{X = 0\} \\
&\qquad - \mathbb{P}\{X = 1\} - \mathbb{P}\{X = 2\} \\
&\quad = 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}.
\end{aligned}$$

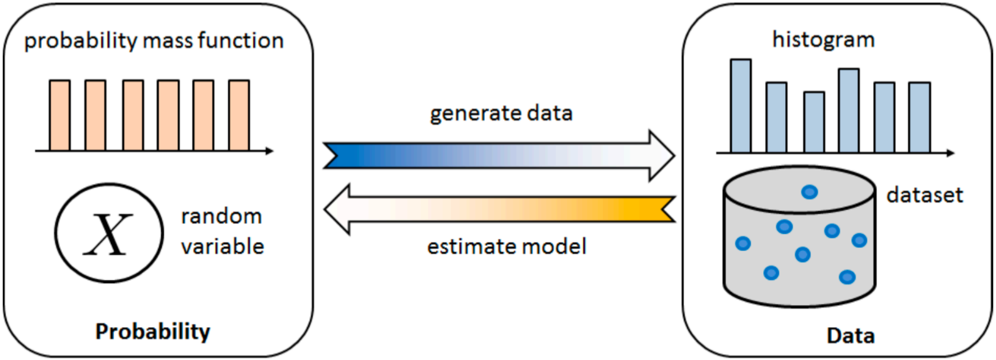
■

PMF and histograms

- A histogram is a plot that shows the frequency of a state.
- The pmf is an ideal histogram.



(a) $N = 100$ (b) $N = 1000$ (c) $N = 10000$ (d) PMF


probability mass function
random
variable
Probability
generate data
estimate model
histogram
dataset
Data

Distribution function

Distribution function

Definition 16 (Cumulative distribution function)

The cumulative distribution function F (or briefly, distribution function) of X is defined by

$$F(x) = \mathbb{P}\{X \leqslant x\}.$$

Remark

If X is a discrete random variable with support $\mathbb{S} = \{s_i, i \in J\}$ and probability mass function p, then the distribution function can be expressed by

$$F(x) = \sum_{s_i \in \mathbb{S}: s_i \leqslant x} p(s_i) = \sum_{i \in J: s_i \leqslant x} p(s_i) = \sum_{i \in J} [p(s_i)\mathbf{I}(s_i \leqslant x)].$$

Examples

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Example 17

Assume that the probability distribution of Y is

y	0	1	2	3
$p(y)$	0.125	0.375	0.375	0.125

If $x = -1$,

$$F(-1) = \mathbb{P}\{Y \leqslant -1\} = \mathbb{P}(\varnothing) = 0.$$

If $x = 1$,

$$F(1) = \mathbb{P}\{Y \leqslant 1\} = p(0) + p(1) = \frac{1}{2}.$$

If $x = 2.3$, then

$$\begin{aligned} F(2.3) &= \mathbb{P}\{Y \leqslant 2.3\} \\ &= p(0) + p(1) + p(2) = \frac{7}{8}. \end{aligned}$$

If $x = 4$, then

$$F(4) = \mathbb{P}\{Y \leqslant 4\} = \mathbb{P}(\Omega) = 1.$$

We can calculate $F(x)$ for all $x \in \mathbb{R}$, not just the support $\mathbb{S}$.

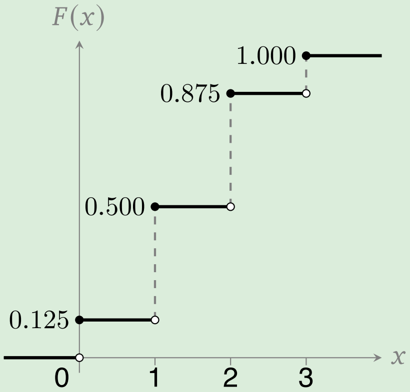
Examples

Example (Cont'd)

We have

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 0.125 & \text{if } 0 \leqslant x < 1, \\ 0.5 & \text{if } 1 \leqslant x < 2, \\ 0.875 & \text{if } 2 \leqslant x < 3, \\ 1 & \text{if } x \geqslant 3. \end{cases}$$

This function is depicted graphically as shown on the right.



Properties of distribution functions

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 18

If X is a random variable, then its CDF F has the following properties:

(i) The CDF is non-decreasing.

(ii) The CDF is right continuous.

(iii) The maximum of the CDF is $F(\infty) := \lim_{x\to\infty} F(x) = 1$.

(iv) The minimum of the CDF is $F(-\infty) := \lim_{x\to-\infty} F(x) = 0$.

Proof.

(i) If $x \leqslant y$, then $\{X \leqslant x\} \subset \{X \leqslant y\}$, and by the monotone property of probability,

$$\mathbb{P}\{X \leqslant x\} \leqslant \mathbb{P}\{X \leqslant y\} \implies F(x) \leqslant F(y).$$

(ii) For any given x_0, and let $\{x_n\}$ be an arbitrary decreasing sequence with limit x_0, and we shall prove that $F(x_n) \to F(x_0)$ as $n \to \infty$. To this end, define $E_n = \{X \leqslant x_n\}$ and $E = \{X \leqslant x_0\}$. Then, it follows that $E_n \downarrow E$. By the monotone continuity property of probability, we have

$$\mathbb{P}(E_n) \to \mathbb{P}(E) \quad \text{as } n \to \infty,$$

which is equivalent to $F(x_n) \to F(x_0)$ as $n \to \infty$.

Proof.

(iii) It is sufficient to prove that, for every increasing sequence $\{x_n, n \geqslant 1\}$ such that $\lim_{n\uparrow\infty} x_n = \infty$, we have $\lim_{n\to\infty} F(x_n) = 1$. To this end, let $E_n = \{X \leqslant x_n\}$ and $E = \{X < \infty\}$, then $E_n \uparrow E$. By the monotone continuity property of probability again, we have

$$\mathbb{P}(E_n) \to \mathbb{P}(E) = \mathbb{P}\{X < \infty\} = \mathbb{P}\{\Omega\} = 1.$$

(iv) Similar to (iii). The proof is omitted and left as an exercise. ■

Properties of distribution

Let a and b be two real numbers such that $a < b$. Then

$$\mathbb{P}\{a < X \leqslant b\} = F(b) - F(a).$$

Proof.

Define $A = \{X \leqslant a\}$ and $B = \{X \leqslant b\}$. As $a < b$, it follows that $A \subset B$ and $B \setminus A = \{a < X \leqslant b\}$. Therefore,

$$\mathbb{P}\{a < X \leqslant b\} = \mathbb{P}(B) - \mathbb{P}(A) = F(b) - F(a).$$ ■

Proposition 19

For any $b \in \mathbb{R}$, we have

$$\mathbb{P}\{X < b\} = F(b-) = \lim_{n\to\infty} F(b - \frac{1}{n}).$$

Proof.

Let $A_n = \{X \leqslant b - \frac{1}{n}\}$. Then $\{A_n, n \geqslant 1\}$ is an increasing sequence of events with limit

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\{X \leqslant b - \frac{1}{n}\} = \{X < b\}.$$

and thus

$$\mathbb{P}\{X < b\} = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} F(b - \frac{1}{n}) = F(b-).$$ ∎

Properties of distribution function

Proposition 20

For any $a \in \mathbb{R}$,

$$\mathbb{P}\{X \geqslant a\} = 1 - F(a-), \quad \mathbb{P}\{X = a\} = F(a) - F(a-).$$

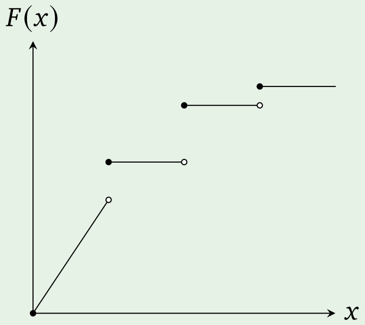
Specially, if X is a discrete random variable with pmf p and support $\mathbb{S} = \{s_j, j \in J\}$, then $F(x)$ has jumps of size $p(s_j)$ at s_j.

Example 21

The distribution function of the random variable X is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \leqslant x < 1 \\ \frac{2}{3} & 1 \leqslant x < 2 \\ \frac{11}{12} & 2 \leqslant x < 3 \\ 1 & x \geqslant 3. \end{cases}$$

Compute (a) $\mathbb{P}\{X < 3\}$, (b) $\mathbb{P}\{X > 1/2\}$, (c) $\mathbb{P}\{X = 1\}$ and (d) $\mathbb{P}\{2 < X \leqslant 4\}$.



Solution.

(a) $\mathbb{P}\{X < 3\} = F(3-) = \frac{11}{12}$.

(b) $\mathbb{P}\{X > \frac{1}{2}\} = 1 - \mathbb{P}\{X \leqslant \frac{1}{2}\} = 1 - F(\frac{1}{2}) = \frac{3}{4}$.

(c) $\mathbb{P}\{X = 1\} = F(1) - F(1-) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$.

(d) $\mathbb{P}(2 < X \leqslant 4) = F(4) - F(2) = 1 - \frac{11}{12} = \frac{1}{12}$. ■

Functions of random variables

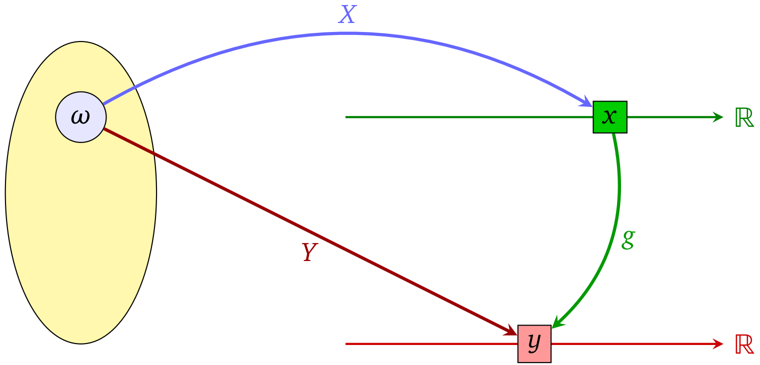
Functions of random variables

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 22

Imagine that two basketball teams (A and B) are playing a seven-game match, and let X be the number of wins for team A. Let $g(x) = 7 - x$, and let $h(x) = 1$ if $x \geqslant 4$ and $h(x) = 0$ if $x < 4$.Then $Y := g(X) = 7 - X$ is the number of wins for team B, and $Z := h(X)$ is the indicator of team A winning the majority of the games. Since X is an r.v., both $g(X)$ and $h(X)$ are also r.v.'s.

A picture



Definition 23 (Function of an random variable)

For the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, an random variable X on Ω and a function $g : \mathbb{R} \to \mathbb{R}$, $Y = g(X)$ is the random variable that maps $\omega \in \Omega$ to $Y(\omega) = g(X(\omega))$.

pmf of $g(X)$

Given a discrete random variable X with a pmf p, what is the pmf of $Y = g(X)$?

- If g is one-to-one, then

$$\mathbb{P}\{Y = y\} = \mathbb{P}\{g(X) = y\} = \mathbb{P}\{X = g^{-1}(y)\}.$$

- Example: If $g(x) = x + 2$, then $g^{-1}(y) = y - 2$.

- If g is not one-to-one, we can define the general inverse map (not necessarily a function) by

$$g^{-1}(B) = \{x \in \mathbb{R} | g(x) \in B\}, \quad \text{for any subset } B \subset \mathbb{R}.$$

- For example, if $g(x) = x^2$, then

$$g^{-1}(\{4\}) = \{-2, 2\}, \quad g^{-1}([1, 9]) = [-3, -1] \cup [1, 3].$$

- Then,

$$\mathbb{P}\{Y = y\} = \mathbb{P}\{X \in g^{-1}(\{y\})\} = \sum_{x : g(x) = y} p(x).$$

pmf of $g(X)$

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Theorem 24

Let X be a discrete random variable with support $\mathbb{S}$ and $g : \mathbb{R} \to \mathbb{R}$. Then the support of $g(X)$ is $g(\mathbb{S})$ and the pmf of $g(X)$ is

$$\mathbb{P}\{g(X) = y\} = \sum_{x:g(x)=y} \mathbb{P}\{X = x\}.$$

Example 25

Assume that X has support $\mathbb{S} = \{-2, -1, 0, 2, 3\}$ and has the following pmf: $p(-2) = \frac{1}{6}$, $p(-1) = \frac{1}{4}$, $p(0) = \frac{1}{12}$, $p(2) = \frac{1}{8}$, $p(3) = \frac{3}{8}$. Let $g(x) = x^2$. Then $Y = g(X)$ has support $g(\mathbb{S}) = \{0, 1, 4, 9\}$, and the pmf p_Y is

$$p_Y(0) = \mathbb{P}\{X = 0\} = \frac{1}{12}, \qquad p_Y(1) = \mathbb{P}\{X = -1\} = \frac{1}{4},$$

$$p_Y(4) = \mathbb{P}\{X \in \{-2, 2\}\} = \frac{7}{24}, \qquad p_Y(9) = \mathbb{P}\{X = 3\} = \frac{3}{8}.$$

Expected value

Example 26

Mary is deciding whether to book the cheaper flight home from college after her final exams, but she's unsure when her last exam will be. She thinks there is only a 20% chance that the exam will be scheduled after the last day she can get a seat on the cheaper flight. If it is and she has to cancel the flight, she will lose \$300. If she can take the cheaper flight, she will save \$200. What you will suggests her?

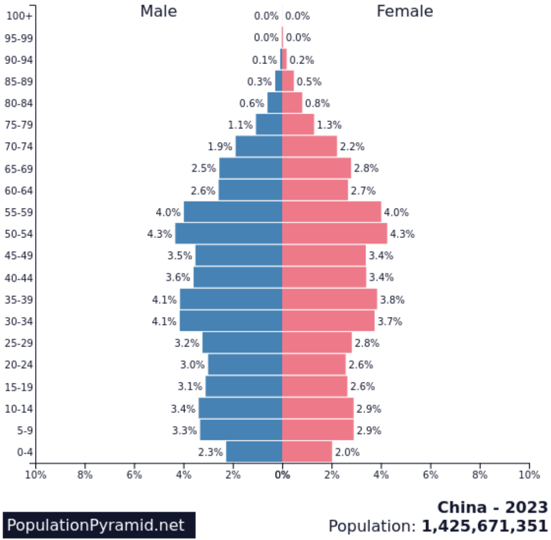
Let $\Omega = \{a, b\}$, where b means the exam will be scheduled before the last day, while a means the exam will be after the last day. Then,

$$\mathbb{P}\{a\} = 0.2, \quad \mathbb{P}\{b\} = 0.8.$$

Let X be the random variable representing the money she will gain, then

$$X(a) = -300, \quad X(b) = 200.$$

Example



Male
Female
100+
95-99
90-94
85-89
80-84
75-79
70-74
65-69
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
0-4
0.0% 0.0%
0.0% 0.0%
0.1% 0.2%
0.3% 0.5%
0.6% 0.8%
1.1% 1.3%
1.9% 2.2%
2.5% 2.8%
2.6% 2.7%
4.0% 4.0%
4.3% 4.3%
3.5% 3.4%
3.6% 3.4%
4.1% 3.8%
4.1% 3.7%
3.2% 2.8%
3.0% 2.6%
3.1% 2.6%
3.4% 2.9%
3.3% 2.9%
2.3% 2.0%
10% 8% 6% 4% 2% 0% 2% 4% 6% 8% 10%
PopulationPyramid.net
China - 2023
Population: 1,425,671,351

Definition

The expected value is one of the most important concepts in probability theory (and also in statistics and data science).

Definition 27 (Expectation)

If X is a discrete random variable having a support $\mathbb{S} = \{s_j, j \in J\}$ and a probability mass function $p(x)$, then the expectation (also called the expected value or the mean) of X, denoted by $\mathbb{E}[X]$, is defined by

$$\mathbb{E}[X] = \sum_{j \in J} s_j p(s_j) = \sum_{x \in \mathbb{S}} x p(x),$$

provided that the summation exists, say,

$$\sum_{x \in \mathbb{S}} |x| p(x) < \infty.$$

Definition

Notation

The expected value of X is sometimes denoted by μ_X, or simply, μ.

Example 28 (Coin tossing)

Consider two independent coin tosses, each with a $3/4$ probability of a head, and let X be the number of heads obtained. What is the mean of X?

Solution.

The pmf of X is given by

$$\mathbb{P}\{0\} = \frac{1}{16}, \quad \mathbb{P}\{1\} = \frac{3}{8}, \quad \mathbb{P}\{2\} = \frac{9}{16}.$$

Then, the mean of X is

$$\mathbb{E}[X] = (0)(1/16) + (1)(3/8) + (2)(9/16) = \frac{3}{2}.$$

■

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Example 29

A school class of 120 students is driven in 3 buses to a symphonic performance. There are 36 students in one of the buses, 40 in another, and 44 in the third bus. When the buses arrive, one of the 120 students is randomly chosen. Let X denote the number of students on the bus of that randomly chosen student, and find $\mathbb{E}[X]$.

Solution.

Let $\Omega = \{1, 2, 3\}$, where i represents that the selected student was on the bus i. Let $\mathscr{F}$ be the power set of Ω. Since the student is randomly chosen with equally likely chance, we can define $\mathbb{P}$ as

$$\mathbb{P}\{1\} = \frac{36}{120}, \quad \mathbb{P}\{2\} = \frac{40}{120}, \quad \mathbb{P}\{3\} = \frac{44}{120}.$$

Let $X : \Omega \to \mathbb{R}$ be defined as $X(i)$ is the number of student on the bus i, then X is a random variable from Ω to $\mathbb{R}$, and

$$\mathbb{P}\{X = 36\} = \mathbb{P}\{1\} = \frac{36}{120}, \quad \mathbb{P}\{X = 40\} = \mathbb{P}\{2\} = \frac{40}{120}, \quad \mathbb{P}\{X = 44\} = \mathbb{P}\{3\} = \frac{44}{120}.$$

By definition,

$$\mathbb{E}[X] = (36)\left(\frac{36}{120}\right) + (40)\left(\frac{40}{120}\right) + (44)\left(\frac{44}{120}\right) = 40.2667. \quad \blacksquare$$

Indicator function, probability and expectation

Example 30

We say that $\mathbf{1}_A$ is an indicator variable for the event A:

$$\mathbf{1}_A = \begin{cases} 1 & \text{if } A \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Find $\mathbb{E}[\mathbf{1}_A]$.

Whether the mean always exists?

The condition $\sum_{x\in\mathbb{S}} |x|p(x) < \infty$ guarantees that $\sum_{x\in\mathbb{S}} xp(x)$ is well-defined, especially for the case where $\mathbb{S}$ is an infinite set.

Example 31 (A counterexample)

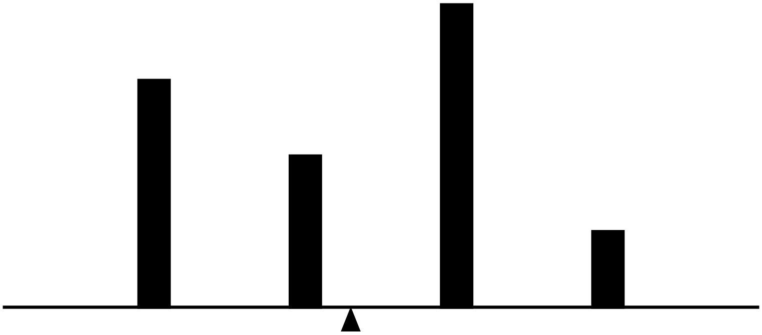
Assume that X takes the value 2^k with probability 2^{-k}, for $k = 1, 2, \ldots$. Then,

$$\sum_{k=1}^{\infty}(2^k)(2^{-k}) = \sum_{k=1}^{\infty} 1 = \infty.$$

In this case, the expectation is not well-defined.

How to understand the mean?

- It is useful to view the mean of X as a “representative” value of X, which lies somewhere in the middle of its range.
- The mean of X is the center of gravity of the pmf.



Population mean and the expectation

Assume that we consider the height of all SUSTech students as a population. Suppose that there are totally N students at SUSTech, whose height are distinct numbers $h_1, h_2, \ldots, h_N$ (世界上没有两片相同的叶子). Let H be the height of a randomly chosen student, in other words, H is a random variable taking values $h_1, \ldots h_N$ with equal probability $1/N$.
Then, the expectation of H is

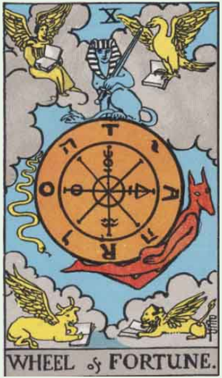
$$\mathbb{E}[H] = \frac{1}{N} \sum_{i=1}^{N} h_i,$$

which is exactly the population mean.

Gambling and expectation

Example 32 (The wheel of fortune)

The following gambling game, known as the wheel of fortune (or chuck-a-luck), is quite popular at many carnivals and gambling casinos: A player bets on one of the numbers 1 through 6. Three dice are then rolled, and if the number bet by the player appears i times, $i = 1, 2, 3,$ then the player wins i units; if the number bet by the player does not appear on any of the dice, then the player loses 1 unit. Is this game fair to the player?



Solution.

Let X denote the amount of money that the player wins. Then,

$$\mathbb{P}(X=-1)=\left(\frac{5}{6}\right)^3=\frac{125}{216},$$
$$\mathbb{P}(X=1)=(3)\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)=\frac{75}{216},$$
$$\mathbb{P}(X=2)=(3)\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)=\frac{15}{216},$$
$$\mathbb{P}(X=3)=\left(\frac{1}{6}\right)^3=\frac{1}{216}.$$

Thus, the expectation of X is

$$\mathbb{E}[X]=(-1)\left(\frac{125}{216}\right)+(1)\left(\frac{75}{216}\right)+(2)\left(\frac{15}{216}\right)+(3)\left(\frac{1}{216}\right)=-\frac{17}{216}.$$

■

Expectation and lottery

Example 33

Assume that a lottery ticket costs $10 dollors, and the distribution of prize is:

Prize	Probability
\$500,000	$\frac{5}{1,000,000}$
\$10,000	$\frac{95}{1,000,000}$
\$500	$\frac{500}{1,000,000}$
\$100	$\frac{1,000}{1,000,000}$
\$20	$\frac{10,000}{1,000,000}$
\$0	$*$

Find the expectation of the prize.

Solution.

Let X denote the prize (in dollors) of a randomly chosen lottery. Then,

$$\begin{aligned}\mathbb{E}[X] &= 500,000 \cdot \frac{5}{1,000,000} + 10,000 \cdot \frac{95}{1,000,000} + 500 \cdot \frac{500}{1,000,000} \\ &\quad + 100 \cdot \frac{1,000}{1,000,000} + 20 \cdot \frac{10,000}{1,000,000} + 0 \cdot \mathbb{P}(X = 0) \\ &= \frac{2,500,000}{1,000,000} + \frac{950,000}{1,000,000} + \frac{250,000}{1,000,000} + \frac{100,000}{1,000,000} + \frac{200,000}{1,000,000} + 0 \\ &= \frac{4,000,000}{1,000,000} \\ &= 4\end{aligned}$$

So, the expected prize is 4, which means on average, you can expect to win 4 units of the prize amount in this scenario. ■

Example 34 (Insurance)

An insurance company offers a "death and disability" policy that pays \$10,000 when the costumer die or \$5,000 if he/she is permanently disabled. It charges a premium of only \$50 a year for this benefit. Suppose that the death rate in any year is 1 out of every 1000 people, and that another 2 out of 1000 suffer some kind of disability. Is the company likely to make a profit selling such a plan?

Expectation of a function of a random variable

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- If X is a random variable and g is a real-valued function, then $g(X)$ is also a random variable.
- How to define the expectation of $g(X)$?
- Idea: find the pmf of $Y = g(X)$, and find the expected value by $\sum y p_Y(y)$?

Example 35

Let X denote a random variable having the following distribution:

$$\mathbb{P}\{X = -1\} = 0.2, \quad \mathbb{P}\{X = 0\} = 0.5, \quad \mathbb{P}\{X = 1\} = 0.3.$$

Find the expectation of $Y = X^2$.

Solution.

The pmf of Y is given by

$$\mathbb{P}\{Y=0\}=\mathbb{P}\{X=0\}=0.5, \quad \mathbb{P}\{Y=1\}=\mathbb{P}\{X=-1\}+\mathbb{P}\{X=1\}=0.5.$$

Hence,

$$\mathbb{E}[Y]=(0)(0.5)+(1)(0.5)=0.5.$$

Can we calculate the expectation of X^2 without the pmf of Y?

Note that

$$\begin{aligned}\mathbb{E}[X^2] &= 0.5 = (1)(0.2)+(0)(0.5)+(1)(0.3)\\ &= (1)(\mathbb{P}\{X=-1\})+(0)(\mathbb{P}\{X=0\})+(1)(\mathbb{P}\{X=1\})\\ &= ((-1)^2)(\mathbb{P}\{X=-1\})+(0^2)(\mathbb{P}\{X=0\})+(1^2)(\mathbb{P}\{X=1\})\\ &= \sum_{x\in\mathbb{S}} x^2 p(x).\end{aligned}$$

■

Proposition 36

If X is a discrete random variable with support $\mathbb{S}$ and pmf p, then, for any real-valued function g,

$$\mathbb{E}[g(X)] = \sum_{x\in\mathbb{S}} g(x)p(x).$$

Proof.

Let $Y = g(X)$. Note that

$$\begin{aligned}
\mathbb{E}[g(X)] &= \sum_{y} y\,\mathbb{P}\{Y = y\} \\
&= \sum_{y} \sum_{x:g(x)=y} y\,\mathbb{P}\{X = x\} \\
&= \sum_{y} \sum_{x:g(x)=y} g(x)\,\mathbb{P}\{X = x\} \\
&= \sum_{x} g(x)\,\mathbb{P}\{X = x\}.
\end{aligned}$$

■

Another expression of the expected value

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 37

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space where $\Omega = \{\omega_1, \omega_2, \dots\}$ is at most countable, and let X be a discrete random variables defined on it. Then,

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\, \mathbb{P}(\{\omega\}).$$

Example

Example 38

Number of coins.

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Example 39 (Average score)

Assume that there are 10 students in a class, whose scores are listed in the following table:

Student ID	Score
1	6
2	7
3	8
4	9
5	6
6	7
7	8
8	9
9	6
10	7

Find the average score.

Proof.

Let $\mathbb{S} = \{x_1, x_2, \ldots\}$ be the support of X, and let $A_x = \{\omega : X(\omega) = x\}$. Then, it follows that

$$\bigcup_{x \in \mathbb{S}} A_x = \Omega, \quad A_x \cap A_y = \varnothing \quad \text{for } x \neq y.$$

By definition,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in \mathbb{S}} x\, \mathbb{P}(X = x) \\
&= \sum_{x \in \mathbb{S}} x\, \mathbb{P}(A_x) \\
&= \sum_{x \in \mathbb{S}} x\, \mathbb{P}\left\{ \bigcup_{\omega \in A_x} \{\omega\} \right\} \\
&= \sum_{x \in \mathbb{S}} \sum_{\omega \in A_x} x\, \mathbb{P}(\{\omega\}) \\
&= \sum_{x \in \mathbb{S}} \sum_{\omega \in A_x} X(\omega)\, \mathbb{P}(\{\omega\}) \\
&= \sum_{\omega \in \Omega} X(\omega)\, \mathbb{P}(\{\omega\}).
\end{aligned}$$

■

Monotone property of expectation

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 40

Let X be a discrete random variable with support $\mathbb{S}$, and let $f, g : \mathbb{S} \to \mathbb{R}$ be functions such that $f(x) \leqslant g(x)$ for all $x \in \mathbb{S}$. Then,

$$\mathbb{E}[f(X)] \leqslant \mathbb{E}[g(X)].$$

Monotone property of expectation

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 41

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, where Ω is at most countable[a], and let X and Y be two discrete random variables defined on it. If $X(\omega) \leqslant Y(\omega)$ for all $\omega \in \Omega$, then

$$\mathbb{E}[X] \leqslant \mathbb{E}[Y].$$

[a] This condition can be removed.

Linearity combinations

Proposition 42

Assume that X is a discrete random variable. For any $a, b \in \mathbb{R}$ and $f, g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[af(X) + bg(X)] = a\,\mathbb{E}[f(X)] + b\,\mathbb{E}[g(X)].$$

Proof.

Assume that X has support $\mathbb{S}$ and pmf p. Then,

$$\begin{aligned}
\mathbb{E}[af(X) + bg(X)] &= \sum_{x \in \mathbb{S}} [af(x) + bg(x)]p(x) \\
&= \sum_{x \in \mathbb{S}} af(x)p(x) + \sum_{x \in \mathbb{S}} bg(x)p(x) \\
&= a \sum_{x \in \mathbb{S}} f(x)p(x) + b \sum_{x \in \mathbb{S}} g(x)p(x) \\
&= a\,\mathbb{E}[f(X)] + b\,\mathbb{E}[g(X)].
\end{aligned}$$

∎

Linearity property of expectation

Specially, if $f(x) = x$ and $g(x) = 1$, we have the following proposition, which is a corollary of the preceding proposition.

Proposition 43

For any discrete random variable X and real numbers a and b, we have

$$\mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b.$$

Linearity property of expectation

Proposition 44

Let $X_1, X_2, \ldots, X_n$ be discrete random variables, then

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

Example 45 (Fair dice)

Find the expected value of the sum obtained when n fair dice are rolled.

Example 46

For a random variable X, it is given that $\mathbb{E}[X] = 2$ and $\mathbb{E}[X^2] = 8$. Calculate the expected value of the following random variables:

$$Y = (2X - 3)^2, \quad W = X(X-1), \quad Z = X^2 + (X+1)^2.$$

Solution.

For Y,

$$\mathbb{E}[Y] = \mathbb{E}[4X^2 - 12X + 9] = 4\,\mathbb{E}[X^2] - 12\,\mathbb{E}[X] + 9 = (4)(8) - (12)(2) + 9 = 17.$$

Similarly,

$$\mathbb{E}[W] = 6, \quad \mathbb{E}[Z] = 21.$$

∎

$\mathbb{E}[g(X)]$ and $g(\mathbb{E}[X])$

- If $g(x) = ax + b$, then we have known that

$$\mathbb{E}[g(X)] = \mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b = g(\mathbb{E}[X]).$$

- Whether $\mathbb{E}[g(X)]$ and $g(\mathbb{E}[X])$ are always equal for general function g?

- If X has the following distribution:

x	-1	0	1
$p(x)$	0.2	0.5	0.3

$\Longrightarrow \mathbb{E}[X] = 0.$

- For different choice of g:

$g(x)$	$2x+1$	x^2	x^3	$\sqrt{x+1}$	e^x	$\cdots$
$\mathbb{E}[g(X)]$	1	0.5	0	0.924	1.389	$\cdots$
$g(\mathbb{E}[X])$	1	0	0	1	1	$\cdots$

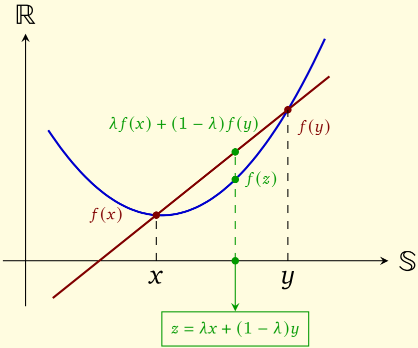
$\mathbb{E}[g(X)]$ is larger if g is convex

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Definition 47 (Convex function)

A function $f : \mathbb{S} \to \mathbb{R}$ is said to be a convex function if for any $\lambda \in (0,1)$ and $x, y \in \mathbb{S}$,

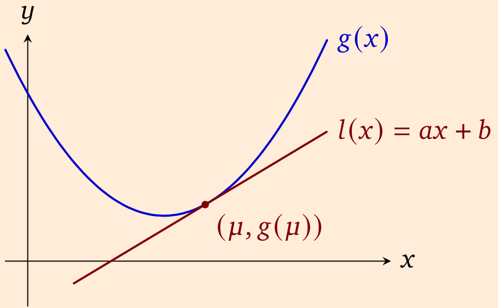
$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y).$$



Theorem 48 (Jensen's inequality)

Let X be a random variable, then for any convex function g,

$$\mathbb{E}[g(X)] \geqslant g(\mathbb{E}[X]).$$



Proof.

Let $\mathbb{S}$ be the support of X. Let $\mu = \mathbb{E}[X]$ and let $l(x) = ax + b$ be a linear function that has $l(\mu) = g(\mu)$ and $g(x) \geqslant l(x)$ for all $x \in \mathbb{S}$. To see that such a function exists, recall that the convexity implies, for any $x, x+h, x-h \in \mathbb{S}$, we have:

$$g(x) \leqslant \frac{1}{2}(g(x-h) + g(x+h)),$$

which further implies that

$$\lim_{h \downarrow 0} \frac{g(\mu) - g(\mu - h)}{h} \leqslant \lim_{h \downarrow 0} \frac{g(\mu + h) - g(\mu)}{h}.$$

Let a be any number between these two limits, $b = g(\mu) - a\mu$, and let

$$l(x) = a(x - \mu) + g(\mu) = ax + b,$$

then $l(x)$ satisfies the condition. Therefore,

$$\begin{aligned}\mathbb{E}[g(X)] \geqslant \mathbb{E}[l(X)] &= \mathbb{E}[aX + b] \\ &= a\mu + b = g(\mu),\end{aligned}$$

as desired. ∎

Example 49 (Application of Jensen's inequality)

Let X be a discrete random variable. Here are some commonly used convex functions:

- $g(x) = |x|$: $\mathbb{E}[|X|] \geqslant |\mathbb{E}[X]|$.
- $g(x) = x^2$: $\mathbb{E}[X^2] \geqslant (\mathbb{E}[X])^2$.
- $g(x) = |x|^p$ for $p \geqslant 1$: $\mathbb{E}[|X|^p] \geqslant |\mathbb{E}[X]|^p$.
- $g(x) = \max\{x, a\}$ for any $a \in \mathbb{R}$: $\mathbb{E}[\max\{X, a\}] \geqslant \max\{\mathbb{E}[X], a\}$.
- $g(x) = e^{\lambda x}$ for all $\lambda \in \mathbb{R}$: $\mathbb{E}[e^{\lambda X}] \geqslant e^{\lambda \mathbb{E}[X]}$.

kth moment and moment generating function

Definition 50

Let k be an non-negative integer. The kth moment of X is defined as $\mathbb{E}[X^k]$, while the kth absolute moment of X is defined as $\mathbb{E}[|X|^k]$.

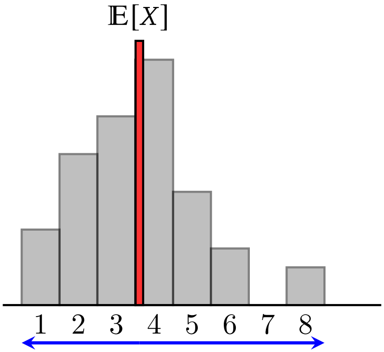
Definition 51

The moment generating function $M : \mathbb{R} \longrightarrow \mathbb{R}$ of X is defined by

$$M(t) = \mathbb{E}[e^{tX}] \quad \text{for } t \in \mathbb{R}.$$

Variance

Introduction



- Imagine that the pmf is an ideal histogram.
- The expectation represents the center of the distribution (the red bar).
- The blue arrow represents the spread of the distribution: How to measure it?

Definition

- We expect that X take on values around $\mu = \mathbb{E}[X]$, it is reasonable to measure the variation of X by the distance between X and μ, on the average.
- Here gives the definition.

Definition 52 (Variance)

The variance of X, denoted by $\mathrm{Var}(X)$, is defined by

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mu)^2].$$

Example 53

Calculate $\text{Var}(X)$ if X represents the outcome when a fair die is rolled.

Solution.

It can be shown that $\mu = \mathbb{E}[X] = \frac{7}{2}$. Also,

$$\text{Var}(X) = \sum_{i=1}^{6}(i - \frac{7}{2})^2 \cdot \frac{1}{6} = \frac{35}{12}.$$

An alternative definition

An alternative formula for $\mathrm{Var}(X)$ is given by

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

An alternative definition

Proof.

$$
\begin{aligned}
\operatorname{Var}(X) &= \mathbb{E}[(X-\mu)^2] \\
&= \sum_{x\in\mathbb{S}} (x-\mu)^2 p(x) \\
&= \sum_{x\in\mathbb{S}} (x^2 - 2\mu x + \mu^2) p(x) \\
&= \sum_{x\in\mathbb{S}} x^2 p(x) - 2\mu \sum_{x\in\mathbb{S}} x p(x) + \mu^2 \sum_{x\in\mathbb{S}} p(x) \\
&= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 \\
&= \mathbb{E}[X^2] - \mu^2.
\end{aligned}
$$

■

In words, the variance of X is equal to the expected value of X^2 minus the square of its expected value.

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 54

In the last example, we have

$$\mathbb{E}[X^2] = \sum_{i=1}^{6} i^2 \cdot \frac{1}{6} = \frac{91}{6},$$

and therefore,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Properties of variance

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 55

For any constants a and b,

$$\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X).$$

Proof.

Let $\mu = \mathbb{E}[X]$. The expected value of $aX + b$ is $a\mu + b$. Therefore,

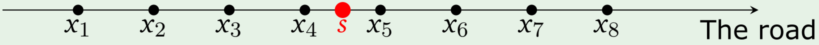
$$\begin{aligned}
\mathrm{Var}(aX + b) &= \mathbb{E}[(aX + b - (a\mu + b))^2] \\
&= \mathbb{E}[a^2(X - \mu)^2] \\
&= a^2\,\mathbb{E}[(X - \mu)^2] \\
&= a^2\,\mathrm{Var}(X).
\end{aligned}$$

∎

Loss function: connecting variance and expectation

Example 56 (Where to build the station)

In an old village people have built houses along a straight road. The travel expense is proportional to the square of the distance. Now, people wanted to build a station. Can you give a suggestion whether to build the station?



The square loss function

The square loss function is defined as

$$\ell(s) = (x - s)^2.$$

Loss function: connecting variance and expectation

Assume that the locations of the houses are randomly distributed. If the station is located at s, then the expected value of the loss is

$$L(s) = \mathbb{E}[(X - s)^2].$$

Proposition 57

Let $\mu = \mathbb{E}[X]$. We have

$$\mu = \arg\min_{s} L(s)$$

and

$$L(\mu) = \mathrm{Var}(X).$$

Variance, square of units

The variance will play an important role in statistics, but it has a problem as a measure of spread. Whatever the units of the original data are, the variance is in squared units. We want measures of spread to have the same units as the data.

Standard deviation

Definition 58

The standard deviation of X is defined by

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

Mean absolute deviation

Definition 59

The mean absolute deviation is defined as

$$\mathrm{MAD}(X) = \mathbb{E}[|X - \mu|].$$

Relation between SD and MAD

Proposition 60

We have

$$\mathrm{MAD}(X) \leqslant \mathrm{SD}(X).$$

Example 61 (Revisit Example 53)

Find the SD and MAD of X in Example 53.

Solution.

We have

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{35}{12}} \approx 1.708.$$

The MAD is

$$\mathrm{MAD}(X) = |1 - 3.5| \cdot \frac{1}{6} + \cdots + |6 - 3.5| \cdot \frac{1}{6} = 1.5.$$

∎

Median

Example 62 (Median of a sample)

The median of a finite list of numbers is the ”middle” number, when those numbers are listed in order from smallest to greatest. For example, f the data set has an odd number of observations, the middle one is selected. For example, the following list of seven numbers,

$$1, \quad 3, \quad 3, \quad 6, \quad 7, \quad 8, \quad 9$$

has the median of 6, which is the fourth value.

If the data set has an even number of observations, there is no distinct middle value and the median is usually defined to be the arithmetic mean of the two middle values.[1][2] For example, this data set of 8 numbers

$$1, \quad 2, \quad 3, \quad 4, \quad 5, \quad 6, \quad 8, \quad 9$$

has a median value of 4.5, that is, $(4+5)/2$.

Example 63 (“Bad” datasets)

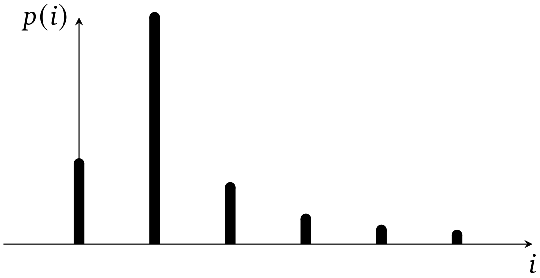
- People usually understand the mean as the center of a dataset or a distribution. However, when a dataset, or a distribution is skewed, or extreme values are not known, or outliers are untrustworthy (maybe measurement errors), then the median is more credible than the mean.
- Median income, for example, may be a better way to describe center of the income distribution because increases in the largest incomes alone have no effect on median.
- For this reason, the median is of central importance in robust statistics.

Expected value may not exist

Example 64

Consider a discrete random variable X has the following probability mass function:

$$p(i) = \begin{cases} 2 - \frac{\pi^2}{6} & i = 0, \\ \frac{1}{(1+i)^2} & i = 1, 2, \ldots. \end{cases}$$



Median: definition

Definition 65 (Median)

For any random variable X with distribution function F, the median of X, denoted by $\mathrm{med}(X)$, is defined as any real number m that satisfies the conditions

$$F(m) \geqslant \frac{1}{2}, \quad F(m-) \leqslant \frac{1}{2}.$$

Examples

Example 66

If X has the following pmf:

$$p(0) = \frac{1}{4}, \quad p(1) = \frac{1}{2}, \quad p(2) = \frac{1}{4}.$$

Find the median of X.

Example 67

Let X be the result in an experiment of tossing a fair die. Find the median of X.

Solution.

Note that for any $3 < x < 4$, we have

$$F(x) = F(x-) = \frac{1}{2}.$$

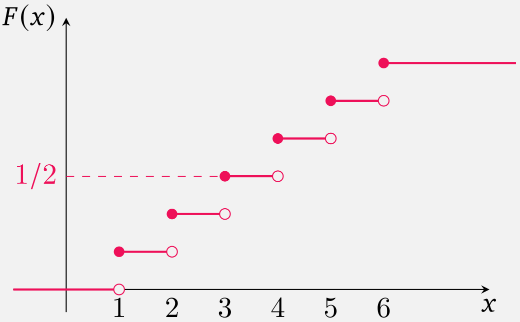
Moreover,

$$F(3) = \frac{1}{2}, \quad F(3-) = \frac{1}{3} \leqslant \frac{1}{2},$$

and

$$F(4) = \frac{2}{3} \geqslant \frac{1}{2}, \quad F(4-) = \frac{1}{2}.$$

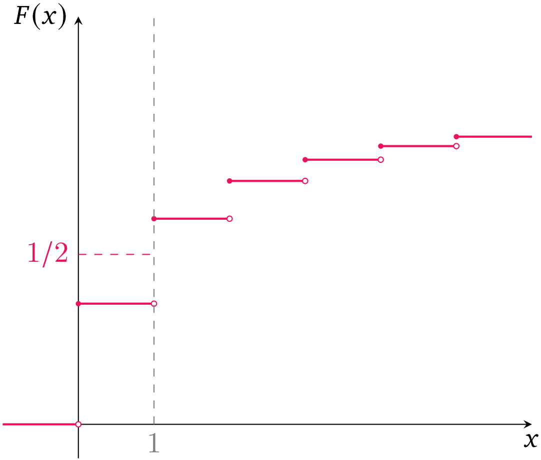
Therefore, any $x \in [3, 4]$ is a median of X.



Example

Example 68 (Revisit Example 64)

Find the median of X in Example 64.



F(x)
1/2
1
x

Commonly used discrete distributions

Bernoulli trials and Bernoulli distribution

- Suppose that a trial whose outcome can be classified as either a success or a failure is performed.
- Let $X = 1$ when the outcome is a success and $X = 0$ when it is a failure.
- Let p, $0 \leqslant p \leqslant 1$, is the probability that the trial is a success.
- Then, the probability distribution of X is given by

x	0	1
$p(x)$	$1-p$	p

Figure: Jacob Bernoulli (伯努利)

Definition 69 (Bernoulli random variable)

A random variable X is said to follow a Bernoulli distribution with parameter p if its probability distribution is given by the above table, denoted by

$$X \sim \text{Bernoulli}(p).$$

Proposition 70

If $X \sim \text{Bernoulli}(p)$, then

$$\mathbb{E}[X] = p, \quad \mathrm{Var}(X) = p(1-p).$$

Proof.

The expectation is

$$\mathbb{E}[X] = (1)(p(1)) + (0)(p(0)) = (1)(p) + (0)(1-p) = p.$$

Similarly,

$$\mathbb{E}[X^2] = (1^2)(p) + (0^2)(1-p) = p,$$

and thus

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

Remark

When $p = 1/2$, then the variance is maximized.

Binomial random variable

- Suppose now that n independent trials, each of which results in a success with probability p and in a failure with probability $1 - p$, are to be performed.

Definition 71 (Binomial distribution)

If X represents the number of successes that occur in the n trials, then X is said to be a binomial random variable with parameters (n, p), denoted by $X \sim$ Binomial(n, p) (or Bin(n, p), B(n, p) in other textbooks).

Remark

Specially, if $n = 1$, then

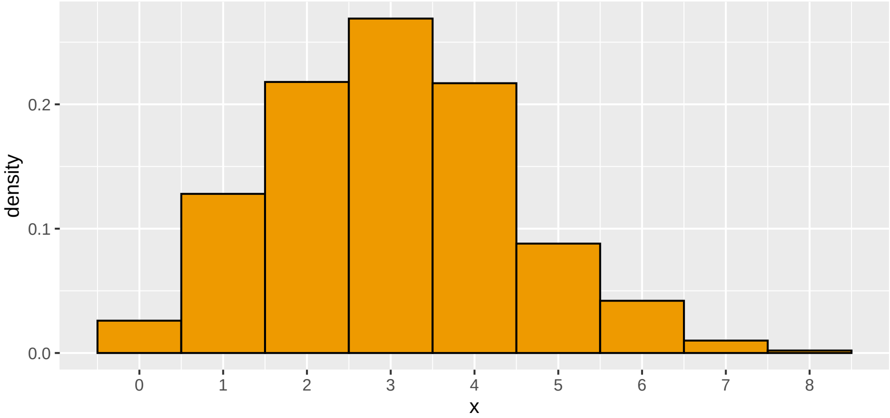
$$\text{Binomial}(1, p) \stackrel{d}{=} \text{Bernoulli}(p).$$

R codes

- Using R, we can generate Binomial random variables as follows:

```
rbinom(n=1,size=10,prob=0.3)  # size is number of trials, n is number of
    random numbers
[1] 5
rbinom(n=100,size=10,prob=0.3)
  [1] 3 4 1 3 3 3 2 2 4 5 3 3 5 3 1 1 2 4 1 4 1 2 7 3 2 5 3 3 3
 [30] 3 3 3 2 3 4 5 4 2 6 1 2 3 4 5 3 2 3 3 3 3 1 2 4 2 4 3 4 1
 [59] 2 2 6 0 3 3 2 5 3 4 1 5 2 2 3 3 2 3 3 2 3 3 1 2 3 1 2 3 1
 [88] 4 2 1 1 2 3 2 0 5 3 5 3 4
x <- rbinom(n=1000,size=10,prob=0.3)
data <- data.frame(x=x)
ggplot(data,aes(x=x,y=after_stat(density))) +
  geom_histogram(binwidth=1,fill="orange2",color="black") +
  scale_x_continuous(breaks=0:10)
```

R codes



- What is the theoretical probability

$$\mathbb{P}\{X = k\} = ?$$

pmf of Binomial(n, p)

Proposition 72

The pmf of a binomial random variable $X \sim$ Binomial(n, p) is given by

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n.$$

Remark

Here, $\binom{n}{k}$ is the number of combinations, p^k is the probability of getting k successes, and $(1-p)^{n-k}$ is the probability of getting $n-k$ failures.

Example 73 (Coins)

Five fair coins are flipped. If the outcome are assumed independent, find the probability mass function of the number of heads obtained.

Solution.

Let X be the random variable representing the number of heads that appear in these five trials. Then, $X \sim$ Binomial$(5, \frac{1}{2})$. Hence, by the pmf formula of the binomial distribution,

$$p(0) = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{1}{32} \approx 0.03125, \ldots$$

The probability distribution is given by

x	0	1	2	3	4	5
$p(x)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

R codes

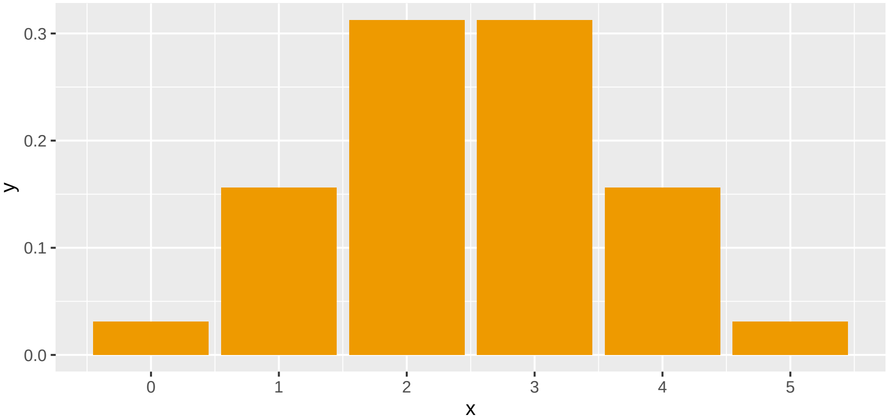
- We can use R to compute the probability mass function:

```
k <- 0:5  # all possible values
dbinom(k, size=5, prob=0.5)
[1] 0.03125 0.15625 0.31250 0.31250 0.15625 0.03125
```

- We can draw the pmf plot as follows:

```
k <- 0:5
pmf <- dbinom(k, size=5, prob=0.5)
binom_data <- data.frame(x=k,y=pmf)
ggplot(data=binom_data) +
  geom_bar(aes(x=x,y=y), stat="identity", fill="orange2") +
  scale_x_continuous(breaks=0:5)
```

R codes



SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Example 74 (Screw products)

It is known that screws produced by a certain company will be defective with probability $.01$, independently of each other. The company sells the screws in packages of 10 and offers a money-back guarantee that at most 1 of the 10 screws is defective. What proportion of packages sold must the company replace?

Solution.

If X is the number of defective screws in a package, then $X \sim$ Binomial$(10, 0.01)$. Hence, the probability that a package will have to be replaced is

$$\begin{aligned}\mathbb{P}\{X \geqslant 2\} &= 1 - \mathbb{P}\{X = 0\} - \mathbb{P}\{X = 1\} \\ &= 1 - \binom{10}{0}(0.01)^0(0.99)^{10} - \binom{10}{1}(0.01)^1(0.99)^9 \\ &\approx 0.004.\end{aligned}$$

R codes

- We can calculate $\mathbb{P}\{X \leqslant k\}$ using `pbinom(k, size=n,prob=p)`: so $\mathbb{P}\{X \leqslant 1\}$ is

```
pbinom(1, size=10, prob=0.01)
[1] 0.9957338
```

- Therefore, the desired probability is

```
1 - pbinom(1, size=10, prob=0.01)
[1] 0.0042662
```

- Alternatively, we can use the following code to compute $\mathbb{P}\{X > 1\}$:

```
pbinom(1, size=10, prob=0.01, lower=FALSE)
[1] 0.0042662
```

Properties of Binomial distribution

Proposition 75

If $X \sim \text{Binomial}(n, p)$, then

$$\mathbb{E}[X] = np, \quad \mathrm{Var}(X) = np(1 - p).$$

Proof.

By definition,

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= 0 + \sum_{k=1}^{n} k \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n \cdot (n-1)!}{(k-1)!(n-k)!} p \cdot p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \underbrace{\sum_{\ell=0}^{m} \frac{m!}{\ell!(m-\ell)!} p^{\ell} (1-p)^{m-\ell}}_{=(p+(1-p))^m=1} \qquad (\ell = k-1 \text{ and } m = n-1) \\
&= np.
\end{aligned}
$$

∎

Properties of Binomial distribution

Proposition 76

If $X \sim \text{Binomial}(n, p)$, where $0 < p < 1$, then as k goes from 0 to n, $p(k)$ first increases monotonically and then decreases monotonically, reaching its largest value when $k = \lfloor (n+1)p \rfloor$, the largest integer that is less than or equal to $(n+1)p$.

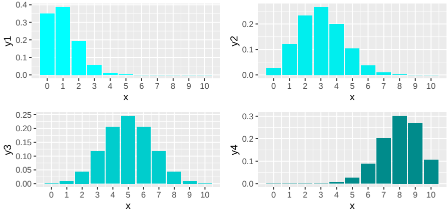


Figure: Binomial pmfs for $n = 10$ and $p = 0.1, 0.3, 0.5$ and 0.8, respectively

Proof.

Consider the ratio

$$\frac{\mathbb{P}\{X = k\}}{\mathbb{P}\{X = k-1\}} = \frac{(n-k+1)p}{k(1-p)}.$$

Hence, $\mathbb{P}\{X = k\} \geqslant \mathbb{P}\{X = k-1\}$ if and only if $(n-k+1)p \geqslant k(1-p)$, which is equivalent to

$$k \leqslant (n+1)p.$$

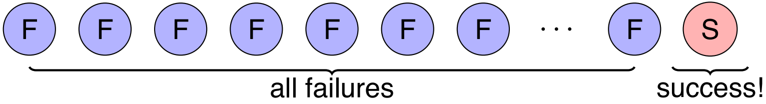
■

Geometric distribution

Suppose that independent trials, each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs.

Definition 77 (Geometric distribution)

Let X equal the number of trials required to get a success, then X is said to have a Geometric distribution with parameter p, denoted by $X \sim \text{Geometric}(p)$.



Proposition 78

The pmf p of a Geometric random variable is given by

$$p(k) = (1-p)^{k-1}p \quad k = 1, 2, \ldots$$

An alternative form of geometric distribution

In probability theory and statistics, the geometric distribution is either one of two discrete probability distributions:

- The probability distribution of the number X of Bernoulli trials needed to get one success, supported on the set $\{1, 2, \ldots\}$.
- The probability distribution of the number $Y = X - 1$ of failures before the first success, supported on the set $\{0, 1, 2, \ldots\}$.
- For the second definition, the pmf of Y is

$$p(k) = (1-p)^k p \quad k = 0, 1, 2, \ldots$$

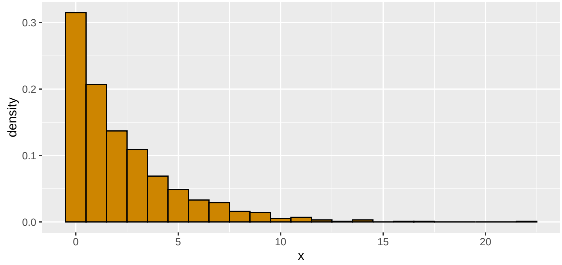
Remark

These two different geometric distributions should not be confused with each other. Often, the name shifted geometric distribution is adopted for the former one (distribution of the number X); however, to avoid ambiguity, it is considered wise to indicate which is intended, by mentioning the support explicitly.

R codes

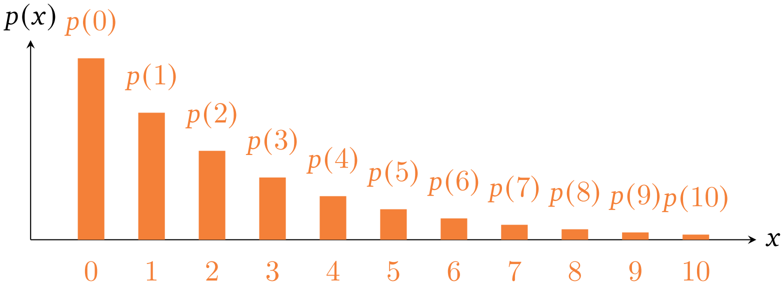
- To generate Geometric random variables, we use `rgeom()` function:

```
rgeom(100, prob=0.3)
  [1]  0  1  0  6  0  0 12  0  0  0  2  0  4  3  4  5  1  0  0
 [20]  3  3  0  0  0  4  6  3  0  0  1  2  1  0  3  1  1  8  4
 [39]  2  2  0  1  6  0  1  0  3  4  1  2  2  1  6  1  0  0  8
 [58]  1  0  1  6  1 11  1  5  5  0  1  3  0  0  1  0  0  0  4
 [77]  2  2  0  1  6  3  4  1  4  1 13  0  1  0  2  2  5  1  0
 [96]  2  1  2  1  0
```



- To calculate the pmf of Geometric(p), use `dgeom()`:

```
dgeom(0:10, prop=0.3)
 [1] 0.300000000 0.210000000 0.147000000 0.102900000 0.072030000
 [6] 0.050421000 0.035294700 0.024706290 0.017294403 0.012106082
[11] 0.008474257
```



Example 79

An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each ball selected is replaced before the next one is drawn, what is the probability that

(a) exactly n draws are needed?

(b) at least k draws are needed?

Solution.

If we let X denote the number of draws needed to select a black ball, then $X \sim$ Geometric$(\frac{M}{M+N})$. Hence,

(a)

$$\mathbb{P}\{X = n\} = \left(\frac{N}{M+N}\right)^{n-1}\left(\frac{M}{M+N}\right) = \frac{MN^{n-1}}{(M+N)^n}.$$

(b)

$$\begin{aligned}\mathbb{P}\{X \geqslant k\} &= \frac{M}{M+N}\sum_{n=k}^{\infty}\left(\frac{N}{M+N}\right)^{n-1} \\ &= \left(\frac{N}{M+N}\right)^{k-1}.\end{aligned}$$

■

Distribution function of Geometric distribution

Proposition 80

For a geometric random variable $X \sim \text{Geometric}(p)$ supported on $\{1, 2, \ldots\}$,

$$\mathbb{P}\{X \geqslant k\} = (1-p)^{k-1}, \quad k = 1, 2, \ldots$$

Proof.

As $\mathbb{P}\{X = k\} = (1-p)^{k-1}p$, it follows that

$$\mathbb{P}\{X \geqslant k\} = \sum_{n=k}^{\infty}(1-p)^{n-1}p = (1-p)^{k-1}. \quad \blacksquare$$

Expected value and variance

Proposition 81

Let $X \sim \text{Geometric}(p)$ supported on $\{1, 2, \ldots\}$. We have

(a) $\mathbb{E}[X] = \frac{1}{p}$.

(b) $\mathrm{Var}(X) = \frac{1-p}{p^2}$.

Proof.

We have

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} kp(k) = p \sum_{k=1}^{\infty} k(1-p)^{k-1} = p + 2p(1-p) + 3p(1-p)^2 + \dots$$

$$(1-p)\,\mathbb{E}[X] = \qquad p(1-p) + 2p(1-p)^2 + \dots$$

Taking the difference on both sides yields

$$p\,\mathbb{E}[X] = p + p(1-p) + p(1-p)^2 + \dots = 1 \quad \Longrightarrow \quad \mathbb{E}[X] = \frac{1}{p}.$$

To compute the variance, it suffices to prove that $\mathbb{E}[X^2] = \frac{2}{p^2} - \frac{1}{p}$:

$$\mathbb{E}[X^2] = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = p + 4p(1-p) + 9p(1-p)^2 + \dots$$

$$(1-p)\,\mathbb{E}[X^2] = p(1-p) + 4p(1-p)^2 + \dots$$

which gives

$$p\,\mathbb{E}[X^2] = \sum_{k=1}^{\infty}(2k-1)p(1-p)^{k-1} = 2 \cdot \frac{1}{p} - 1 \implies \mathbb{E}[X^2] = \frac{2}{p^2} - \frac{1}{p}. \quad \blacksquare$$

Memoryless property of Geometric distribution

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 82

The geometric distribution has the memoryless (forgetfulness) property, that is, for any nonnegative integers s and t,

$$\mathbb{P}\{X > s + t | X > s\} = \mathbb{P}\{X > t\}.$$

Proof.

Note that

$$\mathbb{P}\{X > s + t | X > s\} = \frac{\mathbb{P}\{X > s + t\}}{\mathbb{P}\{X > s\}} = \frac{(1-p)^{s+t}}{(1-p)^s}$$
$$= (1-p)^t = \mathbb{P}\{X > t\}.$$

■

Definition 83 (Poisson distribution)

A random variable X that takes values on one of the values $0, 1, 2, \ldots$ is said to be a Poisson random variable with parameter λ, $0 < \lambda < \infty$, if it has the pmf

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

denoted by $X \sim \text{Poisson}(\lambda)$.

The Poisson probability distribution was introduced by Siméon Denis Poisson in a book published in 1837.

Examples of Poisson distribution

Some examples of random variables that generally obey the Poisson probability distribution are as follows:

- The number of misprints on a page of a book.
- The number of people in a community who survive to age 100.
- The number of wrong telephone numbers that are dialed in a day.
- The number of packages of dog biscuits sold in a particular store each day.
- The number of customers entering a bank on a given day.
- The number of earthquakes all over the world in one month.

R codes

- To generate Poisson random variables, use `rpois()` function:

```
rpois(100, lambda=3)
  [1] 2 6 4 1 4 1 5 2 1 0 2 2 1 3 5 1 4 2 4 3 3 1 1 4 1 3 2 4 4
 [30] 3 4 5 7 2 2 1 2 1 0 1 0 2 4 3 2 4 2 1 4 6 2 4 2 3 1 2 2 2
 [59] 1 5 3 1 0 4 2 2 3 2 2 3 0 3 6 0 2 1 1 4 2 5 4 1 3 2 4 2 2
 [88] 4 1 5 1 5 3 4 5 2 1 1 3 2
```

- To draw the histogram:

```
poisson_data <- data.frame(x=rpois(1000, lambda=3))
ggplot(poisson_data, aes(x=x,y=after_stat(density))) +
  geom_histogram(binwidth=1,fill="orange2",color="black") +
  scale_x_continuous(breaks=0:10)
```

R codes

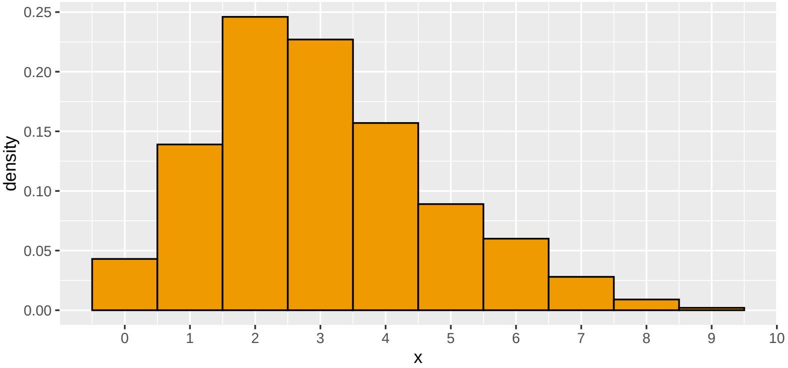
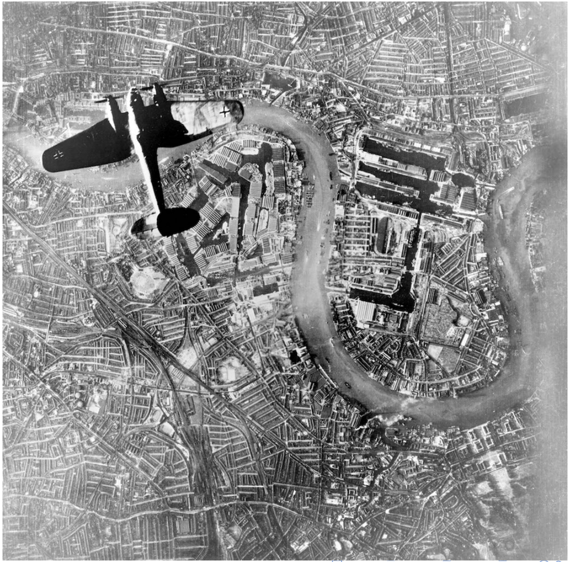


Figure: A histogram for the Poisson distribution with $\lambda = 3$

Poisson in history

Poisson in History

In his book *Gravity's Rainbow*, Thomas Pynchon describes using the Poisson to model the bombs dropping over London during World War II.



Example 84

Suppose that the number of typographical errors on a single page of a book has a Poisson distribution with parameter $\lambda = 0.5$. Calculate the probability that there is at least one error on this page.

Solution.

Let X denote the number of errors on this page. Then, $X \sim$ Poisson(0.5). We have the desired probability is

$$\mathbb{P}\{X \geqslant 1\} = \mathbb{P}\{X > 0\} = 1 - \mathbb{P}\{X = 0\} = 1 - e^{-0.5} \approx 0.393.$$

■

R codes

- If $X \sim$ Poisson(0.5).
- To calculate pmf, use `dpois()` function: for example, $\mathbb{P}\{X = 0\}$ equals

```
dpois(0, lambda=0.5)
[1] 0.6065307
```

- To calculate the cdf, use `ppois()` function: for example $\mathbb{P}\{X \leqslant 2\}$ equals

```
ppois(2, lambda=0.5)
[1] 0.9856123
```

- To calculate the upper probability, e.g., $\mathbb{P}\{X > 0\}$ equals

```
ppois(0, lambda=0.5, lower=FALSE)
[1] 0.3934693
```

R codes

- The true pmf can be shown as follows:

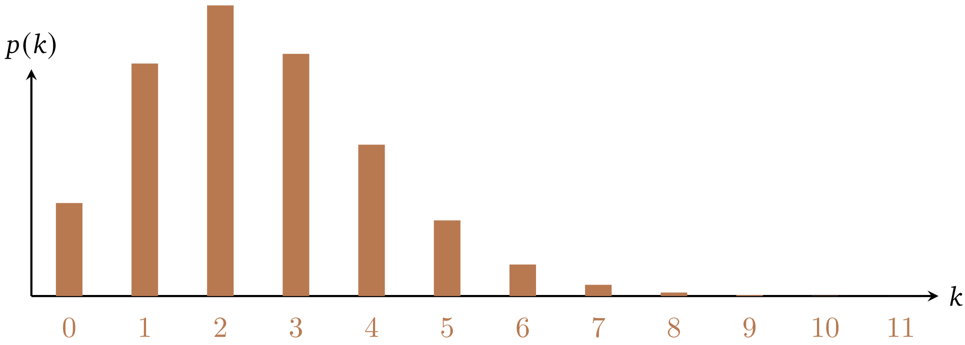


Figure: True pmf of Poisson distribution with $\lambda = 2.5$

Expected value and variance

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Proposition 85

If $X \sim \text{Poisson}(\lambda)$, then

$$\mathbb{E}[X] = \lambda, \quad \text{Var}(X) = \lambda.$$

The parameter λ is also called the intensity (强度) of the Poisson distribution.

Proof.

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} \\ &= \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} \\ &= \lambda,\end{aligned}$$

and similarly,

$$\mathbb{E}[X^2] = \lambda(\lambda + 1).$$

These results prove the proposition. ∎

Binomial & Poisson distributions

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

- Consider Bernoulli trials with rare events: p is quite small and n is very large.
- Example: In the early 1990s, a leukemia (白血病) cluster was identified at Woburn (沃本), a small town in the US, whose population was about $n = 35,000$.
- From the survey, in the United States in the early 1990s, there were about 30,800 new cases of leukemia each year and about 280,000, 000 people, giving a value for p of about 0.00011.
- Let X denote the number of more cases in Woburn in the following year. Then, $X \sim \text{Binom}(n, p)$.
- What is the mean of X?
- How to calculate $\mathbb{P}\{X \geqslant 7\}$?

Binomial & Poisson distributions

- In fact, let $\lambda = np$,

$$\begin{aligned}\mathbb{P}\{X = k\} &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{(n)_k}{n^k}\,\frac{\lambda^k}{k!}\,\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}.\end{aligned}$$

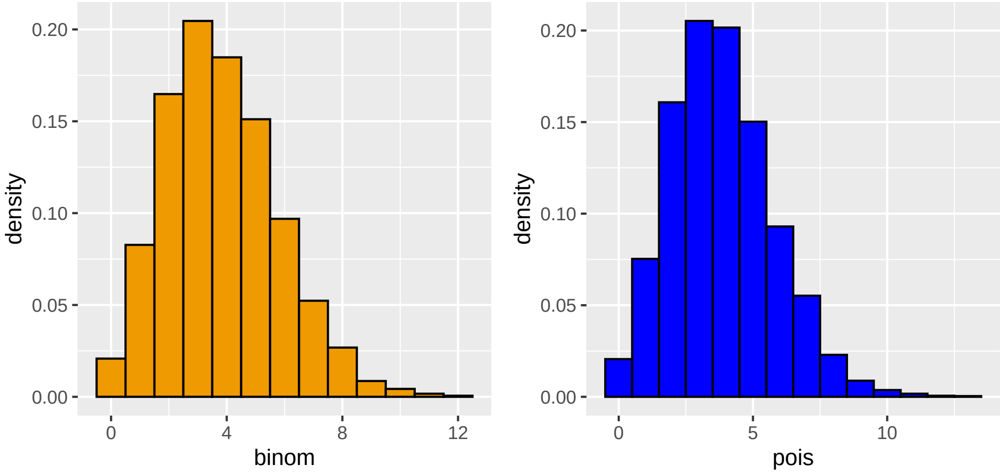
For large n and moderate λ,

$$\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \quad \frac{(n)_k}{n^k} \approx 1, \quad \left(1 - \frac{\lambda}{n}\right)^k \approx 1.$$

Therefore,

$$\mathbb{P}\{X = k\} \approx e^{-\lambda}\frac{\lambda^k}{k!}.$$

A comparison between the histograms



Examples

Example 86

Suppose that the probability that an item produced by a certain machine will be defective is 0.1. Find the probability that a sample a 10 items will contain at most 1 defective item.

Solution.

Let X denote the number of defective items. Then, X follows a Binomial distribution with parameters $(10, 0.1)$. Therefore,

$$\begin{aligned}\mathbb{P}\{X \leqslant 1\} &= \mathbb{P}\{X = 0\} + \mathbb{P}\{X = 1\} \\ &= \binom{10}{0}(0.1)^0(0.9)^{10} + \binom{10}{1}(0.1)^1(0.1)^9 = 0.7361.\end{aligned}$$

Using the Poisson approximation, $\lambda = np = 1$, and thus

$$\mathbb{P}\{X \leqslant 1\} \approx e^{-1}\frac{1^0}{0!} + e^{-1}\frac{1^1}{1!} = 2e^{-1} \approx 0.7358.$$

Example 87

Suppose that earthquakes occur in the western portion of the United States at a rate of 3 per week. Find the probability that at least 3 earth quakes occur during the next 2 weeks.

Solution.

Note that the rate of earthquakes is 3 per week, so the mean of numbers of earthquake occurring in 2 weeks is $\lambda = 3 \times 2 = 6$. Let X denote the number of earthquakes during the next 2 weeks, then it follows that $X \sim \text{Poisson}(6)$. Therefore,

$$\begin{aligned}\mathbb{P}\{X \geqslant 3\} &= 1 - \mathbb{P}\{X = 0\} - \mathbb{P}\{X = 1\} - \mathbb{P}\{X = 2\} \\ &= 1 - e^{-6} - \frac{6}{1!}e^{-6} - \frac{6^2}{2!}e^{-6} \\ &= 1 - 24e^{-6}.\end{aligned}$$

∎

Further reading

[1] Sheldon M. Ross (谢尔登·M. 罗斯).

A first course in probability (概率论基础教程): Chapter 4.

10th edition (原书第十版), 机械工业出版社

[2] Sheldon M. Ross (谢尔登·M. 罗斯).

Introduction to Probability Models (概率模型导论): Chapter 2.

12th edition (原书第十二版), 人民邮电出版社

[3] 李贤平.

概率论基础: Chapters 3 and 4.

第三版, 高等教育出版社