Chapter 3
Usual Probability Distributions

Contents

1 Random Variable **2**
1.1 Discrete Random Variable . . . 2
1.2 Continuous Random Variable . . . 4

2 3.2 Usual Probability Distributions **6**
2.1 3.2.1 Discrete Distributions . . . 6
2.2 3.2.2 Continuous Distributions . . . 9

Introduction

A probability distribution is a mathematical function that theoretically describes a random experiment.
Probability distributions are essential in biology to quantify and predict variability in various biological processes. They allow biologists to analyze data, formulate hypotheses and make decisions.
These mathematical laws thus contribute to a better understanding of random phenomena in the living world.

1 Random Variable

Definition 1

A random variable X is a mapping from the sample space Ω into $\mathbb{R}$

$$X : \Omega \longrightarrow \mathbb{R}$$

$$\omega \longmapsto X(\omega)$$

We distinguish two types of random variables:

1.1 Discrete Random Variable

Definition 2

A random variable X is said to be discrete if it can take a finite number of isolated values.

Probability Law

Let X be a random variable on Ω, $X(\Omega) = \{x_1, x_2, ..., x_n\}$, the probability law of X is given by:

x_i	x_1	x_2	$\cdots$	x_n
$p(X = x_i)$	p_1	p_2	$\cdots$	p_n

Distribution Function

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

$$F_X(x) = \begin{cases} 0 & \text{if } x < x_1 \\ p_1 & \text{if } x_1 \leq x < x_2 \\ p_1 + p_2 & \text{if } x_2 \leq x < x_3 \\ \vdots & \\ 1 & \text{if } x \geq x_n \end{cases}$$

Mathematical Expectation

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$$

Variance and Standard Deviation

$$V(X) = E[(X - E(X))^2]$$

$$V(X) = \sum_{i=1}^{n} (x_i - E(X))^2 P(X = x_i)$$

or

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

$$= \sum_{i=1}^{n} x_i^2 P(X = x_i) - \left(\sum_{i=1}^{n} x_i P(X = x_i) \right)^2$$

$$\delta_X = \sqrt{V(X)}$$

Example 3.1

Consider the random experiment: "a six-faced die is thrown and we observe the result".
We consider the following game:
- If the result is even we win 2 DA - If the result is 1 we win 3 DA - If the result is 3 or 5 we lose 4 DA
We define a random variable X that gives the gain of this game.
$\Omega = \{1, 2, 3, 4, 5, 6\}$
$X(\Omega) = \{-4, 2, 3\}$
Probability law:

x_i	-4	2	3
$P(X = x_i)$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{1}{6}$

Distribution function:

$$F_X(x) = \begin{cases} 0 & x < -4 \\ \frac{2}{6} & -4 \leq x < 2 \\ \frac{5}{6} & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

$$E(X) = -4\frac{2}{6} + 2\frac{3}{6} + 3\frac{1}{6} = \frac{1}{6}$$

$$V(X) = E(X^2) - (E(X))^2 = 8.8$$

$$\delta_X = \sqrt{V(X)} = 2.97$$

1.2 Continuous Random Variable

Definition 3

A random variable X is said to be continuous if it can take all values within an interval.

Probability Density

Let X be a continuous random variable. We say that

$$f : \mathbb{R} \longrightarrow \mathbb{R}^+$$

is a probability density of X if

- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
- f is continuous on $\mathbb{R}$
- $$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Distribution Function

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

Property

Let $F_X(x)$ be a cumulative distribution function. Then:

- $F_X(x)$ is positive.
- $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to +\infty} F_X(x) = 1$.

Mathematical Expectation

The mathematical expectation of the continuous random variable X is given by:

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

Variance and Standard Deviation

The variance of the continuous random variable X is given by:

$$V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)\, dx$$

or equivalently

$$V(X) = E(X^2) - [E(X)]^2 = \int_{-\infty}^{+\infty} x^2 f(x)\, dx - [E(X)]^2$$

The standard deviation is defined by:

$$\sigma_X = \sqrt{V(X)}$$

Remark

- $P(X = x) = 0$
- $P(X \leq x) = P(X < x)$
- $P(a \leq X \leq b) = \int_a^b f(t)dt = F_X(b) - F_X(a)$

2 3.2 Usual Probability Distributions

2.1 3.2.1 Discrete Distributions

Bernoulli Distribution

Any random experiment having two possible outcomes success and failure is called a Bernoulli experiment.
Probability law
The random variable $X = 1$ in case of success with probability p and $X = 0$ in case of failure with probability $q = 1 - p$

$$P(X = x) = \begin{cases} p & \text{if } x = 1 \\ q & \text{if } x = 0 \end{cases}$$

Notation

$$X \sim B(p)$$

$E(X) = p$
$V(X) = pq$
$\delta_X = \sqrt{V(X)}$

Example 3.2

A fair coin is tossed in the air, if we obtain head we have success.
It is a Bernoulli experiment

$$\Omega = \{face, pile\}, \quad A = \{face\}, \quad \bar{A} = \{pile\}$$

Probability law
Probability of success:

$$p = P(A) = \frac{1}{2}$$

Probability of failure:

$$q = P(\bar{A}) = \frac{1}{2}$$

$$P(X = x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = 0 \end{cases}$$

$E(X) = p = \frac{1}{2}$
$V(X) = pq = \frac{1}{4}$
$\delta_X = \sqrt{V(X)} = \frac{1}{2}$

Binomial Distribution

The Binomial distribution with parameters n and p models the number of successes obtained during the repetition of n Bernoulli experiments in an identical and independent manner.
Probability law

$$P(X = k) = C_n^k p^k q^{n-k}, \quad k = 1, 2, ..., n$$

$$C_n^k = \frac{n!}{k!(n-k)!}$$

Notation

$$X \sim B(n, p)$$

$E(X) = np$
$V(X) = npq$
$\delta_X = \sqrt{V(X)}$

Example 3.3

A fair die is thrown 5 times and we are interested in the result "obtaining the number 02".
What is the probability of obtaining the number 02 two times?
What is the probability of obtaining the number 02 at least three times?
Determine $E(X), V(X)$ and δ_X
Solution:
The random variable X: "obtaining the number 02", $X \sim B(5, p)$

$$\Omega = \{1,2,3,4,5,6\}, \quad A = \{2\}, \quad \bar{A} = \{1,3,4,5,6\}$$

Probability of success:

$$p = P(A) = \frac{1}{6}$$

Probability of failure:

$$q = P(\bar{A}) = \frac{5}{6}$$

$$P(X = 2) = C_5^2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{5-2} = 0.16$$

$$P(X \geq 3) = 1 - P(X < 3)$$

$$= 1 - (P(X = 0) + P(X = 1) + P(X = 2))$$

$$= 1 - (0.4 + 0.4 + 0.16) = 0.036$$

$$E(X) = np = \frac{5}{6}$$

$$V(X) = npq = \frac{25}{36}$$

$$\delta_X = \sqrt{V(X)} = \frac{5}{6}$$

Poisson Distribution

The Poisson distribution is the distribution of rare events, that is to say events having a low probability of occurrence.
Probability law

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad \lambda > 0$$

Notation

$$X \sim P(\lambda)$$

$E(X) = \lambda$
$V(X) = \lambda$
$\delta_X = \sqrt{\lambda}$

Example 3.4

A telephone center receives on average 5 calls per minute.
What is the probability that the center receives exactly two calls?
The random variable X: number of calls in the center, $X \sim P(5)$

$$P(X = 2) = e^{-5}\frac{5^2}{2!} = 0.084$$

2.2 3.2.2 Continuous Distributions

Normal Distribution

A random variable X follows a normal distribution or Gauss-Laplace distribution with parameters m and δ if:

$$f(x) = \frac{1}{\sqrt{2\pi}\delta}e^{-\frac{1}{2}\left(\frac{x-m}{\delta}\right)^2}$$

Notation

$$X \sim N(m, \delta)$$

Log-normal Distribution

A random variable X follows a log-normal distribution if $\ln(X)$ follows a normal distribution $N(m, \delta)$:

$$f(x) = \frac{1}{x\sqrt{2\pi}\delta}e^{-\frac{1}{2}\left(\frac{\ln(x)-m}{\delta}\right)^2}, \quad x > 0$$

Notation

$$X \sim LN(m, \delta)$$

Standard Normal Distribution

The standard normal distribution is the distribution $N(0,1)$.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

If $X \sim N(m, \delta)$ then

$$Y = \frac{X - m}{\delta} \sim N(0,1)$$

Notations

By convention, F denotes the density of $\mathcal{N}(0,1)$ and F its cumulative distribution function.

Properties

1. F is an even function.
2. $F(x) = 1 - F(-x)$
3. $P(|X| \leq x) = P(-x \leq X \leq x) = 2(F(x) - 1/2)$
4. $P(|X| \geq x) = P((X \leq -x) \cap (X \geq x)) = 2(1 - F(x))$

Example 3.5: Standard Normal Table and Probability Calculation

- $X \sim N(0,1)$ calculate $P(X \leq 1.25)$ and $P(X \leq 0.67)$

$$x = 1.25 \Rightarrow \text{row} = 1.2 \text{ and column} = 0.05 \Rightarrow P(X \leq 1.25) = F(1.25) = 0.8944$$

$$x = 0.67 \Rightarrow \text{row} = 0.6 \text{ and column} = 0.07 \Rightarrow P(X \leq 0.67) = F(0.67) = 0.7486$$

- $X \sim N(0,1)$ calculate $P(X \geq 0.87)$ and $P(X \leq 0.74)$

$$P(X \geq 0.87) = 1 - P(X \leq 0.87) = 1 - F(0.87) = 1 - 0.8078 = 0.1922$$

$$P(X \geq 0.74) = 1 - P(X \leq 0.74) = 1 - F(0.74) = 1 - 0.7704 = 0.2296$$

- $P(X \leq -x) = 1 - P(X \leq x)$ calculate $P(X \leq -1.87)$

$$P(X \leq -1.87) = 1 - P(X \leq 1.87) = 1 - F(1.87) = 1 - 0.9693 = 0.0407$$

- $P(X \geq -x) = P(X \leq x)$ calculate $P(X \geq -0.74)$

$$P(X \geq -0.74) = P(X \leq 0.74) = 0.7704$$

- $P(a \leq X \leq b) = F(b) - F(a)$ calculate $P(1.15 \leq X \leq 2.25)$ and $P(-0.58 \leq X \leq -0.14)$

$$P(1.15 \leq X \leq 2.25) = F(2.25) - F(1.15) = 0.9878 - 0.8749 = 0.1129$$

$$P(-0.58 \leq X \leq -0.14) = F(0.58) - F(0.14) = 0.7190 - 0.5557 = 0.1633$$

- $P(-a \leq X \leq b) = F(b) + F(a) - 1$ calculate $P(-1.14 \leq X \leq 2.58)$

$$P(-1.14 \leq X \leq 2.58) = F(2.58) + F(1.14) - 1 = 0.9951 + 0.8729 - 1 = 0.8679$$

- $P(-a \leq X \leq a) = 2F(a) - 1$ calculate $P(-1 \leq X \leq 1)$ and $P(-1.96 \leq X \leq 1.96)$

$$P(-1 \leq X \leq 1) = 2F(1) - 1 = 2(0.8413) - 1 = 0.6827$$

$$P(-1.96 \leq X \leq 1.96) = 2F(1.96) - 1 = 2(0.976) - 1 = 0.95$$

Chi-square Distribution

Chi-square Distribution

Let $X_1, X_2, ..., X_n$ be n independent random variables $\sim N(0,1)$. Then

$$Y = X_1^2 + X_2^2 + ... + X_n^2$$

follows a distribution called chi-square distribution with n degrees of freedom.

$$f(y) = \frac{y^{\frac{n}{2}-1}e^{-y/2}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}$$

with

$$\Gamma(n) = \int_0^{+\infty} e^{-x}x^{n-1}dx$$

Notation

$$Y \sim \chi_n^2$$

$E(Y) = n$
$V(Y) = 2n$

Example 3.6: Chi-square Table and Probability Calculation

- $Y \sim \chi_{18}^2$ calculate $P(Y \leq 28.87)$

$$P(Y \leq 28.87) = 0.95$$

- $Y \sim \chi_{10}^2$ calculate $P(Y \geq 23.209)$

$$P(Y \geq 23.209) = 1 - P(Y \leq 23.209) = 1 - 0.99 = 0.01$$

- Find y such that $P(Y \leq y) = 0.975$ and $Y \sim \chi_{22}^2$

$$y = 36.781$$

- Find y such that $P(Y \geq y) = 0.99$ and $Y \sim \chi_7^2$

$$P(Y \leq y) = 1 - P(Y \geq y) = 1 - 0.99 = 0.01$$

$$y = 1.239$$

Student Distribution

Student Distribution

Let X and Y be two independent random variables such that $X \sim N(0,1)$ and $Y \sim \chi_n^2$. Then

$$T = \frac{X}{\sqrt{Y/n}}$$

follows a distribution called Student distribution with n degrees of freedom.

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

Notation

$$T \sim t_n$$

$E(T) = 0$, $n > 1$
$V(T) = \frac{n}{n-2}$, $n > 2$

Example 3.7: Student Table and Probability Calculation

- $T \sim t_9$ calculate $P(T \geq 2.2622)$ and $P(T \geq 1.3830)$

$$P(T \geq 2.2622) = \frac{0.05}{2} = 0.025$$

$$P(T \geq 1.3830) = \frac{0.2}{2} = 0.1$$

- $T \sim t_{16}$ calculate $P(T \leq 1.746)$

$$P(T \leq 1.746) = 1 - P(T \geq 1.746) = 1 - \frac{0.1}{2} = 0.95$$