Chapter 3

Common Probability Laws

Contents

A probability law is a mathematical function that theoretically describes a random experiment.

Probability laws 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 contribute to a better understanding of random phenomena in the living world.

3.1 Random Variable

Definition 3.1.1. *A random variable X is a function from the sample space Ω to $\mathbb{R}$:*

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ w &\longmapsto X(w) \end{aligned}$$

There are two types of random variables:

3.1.1 Discrete Random Variable

Definition 3.1.2. *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, with possible values $X(\Omega) = \{x_1, x_2, ..., x_n\}$. The probability law of X is given by:

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

Cumulative Distribution Function (CDF)

The cumulative distribution function of the random variable X is:

$$F_X(x) = p(X \leq x) = \sum_{x_i \leq x} p(X = x_i) = \begin{cases} 0 & si & x < x_1 \\ p_1 & si & x_1 \leq x < x_2 \\ p_1 + p_2 & si & x_2 \leq x < x_3 \\ \vdots & & \\ 1 & si & x \geq x_n \end{cases}$$

Mathematical Expectation

The expectation (mean) of the random variable X is:

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

Variance and Standard Deviation

The variance of X is:

$$V(X) = E\left[\left(X - E(X)\right)^2\right] = \sum_{i=1}^{n} \left(x_i - E(X)\right)^2 p(X = x_i)$$

or

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

The standard deviation is: $\delta_x = \sqrt{V(X)}$

Example 3.1.1.

Consider the experiment of rolling a six-sided die and observing the outcome. The game is as follows:

-If the result is even, you win 2 DA.

-If the result is 1, you win 3 DA.

-If the result is 3 or 5, you lose 4 DA.

Define X as the random variable representing the gain.

❶ $\Omega = \{1,2,3,4,5,6\}$

❷ $X(\Omega) = \{-4,2,3\}$.

❸ *The probability law of X.*

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

❹ *The cumulative distribution function of X.*

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

❺ *Expectation* $E(X)$

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} x_i p(X = x_i) \\ &= -4(2/6) + 2(3/6) + 3(1/6) = 1/6 \end{aligned}$$

❻ $V(X) = E(X^2) - \left(E(X)\right)^2 = 8.8 \; et \; \delta_x = \sqrt{V(X)} = 2.97$

3.1.2 Continuous Random Variable

Definition 3.1.3. *A random variable X is said to be continuous if it can take any value within an interval.*

Probability Density Function (PDF)

Let X be a continuous random variable. A function $f : \mathbb{R} \rightarrow \mathbb{R}^{+}$ is a probability density function (PDF) of X if:

❶ $f(x) \geq 0$ for all $x \in \mathbb{R}$.

❷ $f(x)$ is continuous over $\mathbb{R}$.

❸ $\int_{-\infty}^{+\infty} f(x)dx = 1$

Cumulative Distribution Function (CDF)

The cumulative distribution function of a continuous random variable X is

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

Properties 3.1.1.

❶ $F_X(x)$ *is positive*

❷ $\lim\limits_{x \longrightarrow -\infty} F_X(x) = 0$ *and* $\lim\limits_{x \longrightarrow +\infty} F_X(x) = 1$

Remark 3.1.1.

❶ $p(X = x) = 0$

❷ $p(X \leq x) = p(X < x)$

❸ *The probability of the continuous random variable* $X \in [a, b]$ *is given by:*

$$\begin{aligned} p(a \leq X \leq b) &= p(a < X \leq b) \\ &= p(a \leq X < b) \\ &= p(a < X < b) \\ &= \int_a^b f(t)dt \\ &= F_X(b) - F_X(a) \end{aligned}$$

Mathematical Expectation

The expectation of X is

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

Variance and Standard Deviation

The variance is:

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

or

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

3.2 Common Probability Laws

3.2.1 Discrete Distributions

Bernoulli Distribution

⇝ A Bernoulli experiment is a random experiment with two possible outcomes: success and failure.

⇝ Probability Law

$$p(X=x)=\begin{cases} p & if \quad x=1 \\ 1-p & if \quad x=0 \end{cases}$$

⇝ Notation

$$X \sim \mathcal{B}(p)$$

⇝ $E(X)=p$ ⇝ $V(X)=p(1-p)$ ⇝ $\delta_X = \sqrt{V(X)}$

Example 3.2.1.

We flip a fair coin in the air, if it lands on heads, we have a success.

$\rightsquigarrow$ *The experiment is a Bernoulli trial*

$$\Omega = \{heads, tails\}, \quad A = \{heads\}, \quad \bar{A} = \{tails\}$$

$\rightsquigarrow$ *Probability Law*

The probability of success $p = p(A) = \frac{1}{2}$

The probability of failure $p(\bar{A}) = \frac{1}{2}$

$$p(X = x) = \begin{cases} \dfrac{1}{2} & si \quad x = 1 \\ \dfrac{1}{2} & si \quad x = 0 \end{cases}$$

$\rightsquigarrow E(X) = p = \dfrac{1}{2}$ $\rightsquigarrow V(X) = p(1-p) = \dfrac{1}{4}$ $\rightsquigarrow \delta_X = \sqrt{V(X)} = \dfrac{1}{2}$

Binomial Distribution

$\rightsquigarrow$ The binomial distribution with parameters n and p models the number of successes obtained when repeating n identical and independent Bernoulli trials.

$\rightsquigarrow$ Probability Law

$$p(X = k) = C_n^k p^k (1-p)^{n-k}; \ \ k = 1, 2, ..., n. \ \ C_n^k = \frac{n!}{k!(n-k)!}$$

$\rightsquigarrow$ Notation

$$X \sim \mathcal{B}(n, p)$$

$\rightsquigarrow E(X) = np$

$\rightsquigarrow V(X) = np(1-p)$

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

Example 3.2.2.

A fair die is rolled 5 times, and we are interested in the outcome "getting the number 2".

❶ *What is the probability of obtaining the number 2 exactly twice?*

❷ *What is the probability of obtaining the number 2 at least three times?*

❸ *Determine $E(X), V(X)$ and δ_X*

Solution:

The random variable X represents the number of times the number 2 appears

$$X \sim \mathcal{B}(5, p)$$

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

Probability of success: $p = p(A) = \frac{1}{6}$

Probability of failure: $p(\bar{A}) = \frac{5}{6}$

❶ $p(X = 2) = C_n^k p^k (1-p)^{n-k} = C_5^2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{5-2} = 0.16$

❷

$$\begin{aligned} p(X \geq 3) &= 1 - p(X < 3) \\ &= 1 - \Big(p(X=0) + p(X=1) + p(X=2)\Big) \\ &= 1 - \Big(0.4 + 0.4 + 0.16\Big) = 0.036 \end{aligned}$$

❸ $E(X) = np = \dfrac{5}{6}, \quad V(X) = np(1-p) = \dfrac{25}{36}, \quad \delta_X = \sqrt{V(X)} = \dfrac{5}{6}$

Poisson Distribution

⇝ The Poisson distribution is the probability distribution of rare events, meaning events that have a low probability of occurring.

⇝ Probability Law

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

⇝ Notation

$$X \sim \mathcal{P}(\lambda)$$

⇝ $E(X) = \lambda$

⇝ $V(X) = \lambda$

⇝ $\delta_X = \sqrt{\lambda}$

Example 3.2.3.

A telephone exchange receives an average of 5 calls per minute.

What is the probability that the exchange receives exactly two calls in one minute?

The random variable X: represents the number of calls received: $X \sim \mathcal{P}(5)$

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

3.2.2 Continuous Distributions

Normal Distribution

⇝ A random variable X follows a normal distribution or Gaussian 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 normal distribution with parameters $N(0, 1)$.

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

⇝ If $X \sim N(m, \delta)$ then $Y = \dfrac{X-m}{\delta} \sim N(0, 1)$

Example 3.2.4. *Standard Normal Table and Probability Calculation*

• $p(X \leq 1.25)$ *and* $p(X \leq 0.67)$

$$x = 1.25 \Longrightarrow \quad row = 1.2 \;\; and \;\; column = 0.05 \Longrightarrow p(X \leq 1.25) = F(1.25) = 0.8944$$

$$x = 0.67 \Longrightarrow \quad row = 0.6 \;\; and \;\; column = 0.07 \Longrightarrow p(X \leq 0.67) = F(0.67) = 0.7486$$

• $p(X \geq 0.87)$ *and* $p(X \leq 0.74)$

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

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

• $p(X \leq -x) = 1 - p(X \leq x)$

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

• $p(X \geq -x) = p(X \leq x)$

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

• $p(a \leq X \leq b) = F(b) - F(a)$

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

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

• $p(-a \leq X \leq b) = F(b) + F(a) - 1$

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

• $p(-a \leq X \leq a) = 2F(a) - 1$

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

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

.

Chi-Square Distribution

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

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

follows a Chi-Square distribution with n degrees of freedom..

$$f(y) = \frac{y^{\frac{n}{2}-1}e^{-y/2}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \quad \text{where} \quad \Gamma(n) = \int_0^{+\infty} e^x x^{n-1} dx$$

⇝ Notation

$$Y \sim \chi_n^2$$

⇝ $E(Y) = n$

⇝ $V(Y) = 2n$

Example 3.2.5. *Chi-Square Table and Probability Calculation*

• *Given* $Y \sim \chi_{18}^2$ *calculate* $p(Y \leq 28.87)$

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

• *Given*$Y \sim \chi_{10}^2$ *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* $Y \sim \chi_7^2$

$$p(Y \leq y) = 1 - p(Y \geq y) = 1 - 0.99 = 0.01 \quad \textit{then} \quad y = 1.239$$

Student's Distribution

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

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

follows a Student's distribution with n degrees of freedom.

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

⇝ Notation

$$T \sim t_n$$

⇝ $E(T) = 0, \quad n > 1$

⇝ $V(T) = \frac{n}{n-2}, \quad n > 2$

Example 3.2.6. *Student's Table and Probability Calculation*

- *Given $T \sim t_9$ calculate $p(T \geq 2.2622)$ and $p(T \geq 1.3830)$*

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

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

- *Given $T \sim t_{16}$ calculate $p(T \leq 1.746)$*

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