COURSE

2 Probability Laws

Probability laws are essential in biology for quantifying and predicting variability in various biological processes. They allow biologists to analyze data, formulate hypotheses, and make informed decisions in fields such as genetics, evolution, and ecology. These mathematical laws thus contribute to a better understanding of random phenomena in the living world.

1 General Concepts

Definition

A probability is a mapping $P : \Omega \to [0,1]$ such that: for every $A \in \Omega$, we have

1. $P(A) \geq 0$,
2. $P(\Omega) = 1$,
3. If $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$.

Remark

The triplet $(\Omega, P(\Omega), P)$ is called a probabilistic space.

Basic Properties

1. $P(\overline{A}) = 1 - P(A)$
2. $P(\varnothing) = 0$
3. $A \subset B \Rightarrow P(A) \leq P(B)$
4. $0 \leq P(A) \leq 1$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Remark

$P(A) = 0$ does not mean that A is an impossible event.

Conditional Probability

Let A and B be two events such that $P(B) \neq 0$. The conditional probability of A given B, denoted $P(A \backslash B)$ or $P_B(A)$ (probability of A given B), is defined by:

$$P(A\backslash B) = \frac{P(A \cap B)}{P(B)}$$

Definition

Let E be a set. $A_1, A_2, ..., A_n$ form a partition of E if:

1. $\forall\ i \in 1, ..., n;\ A_i \neq \varnothing$
2. $\forall\ i \neq j;\ A_i \cap A_j = \varnothing$
3. $A_1 \cup A_2 \cup ... \cup A_n = E$

Total Probability Formula

If the events $B_1, B_2, ..., B_n$ form a partition of Ω, then

$$P(A) = P(B_1)P(A\backslash B_1) + P(B_2)P(A\backslash B_2) + ... + P(B_n)P(A\backslash B_n) = \sum_{i=1}^{n} P(B_i)P(A\backslash B_i)$$

Bayes' Formula

If $P(A) \neq 0$ and $P(B) \neq 0$ then

$$P(A\backslash B) = \frac{P(A)P(B\backslash A)}{P(B)}$$

If the events $A_1, A_2, ..., A_n$ form a partition of Ω, then

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

Example

A systematic screening test is established to detect a disease M. The risk of having this disease is 0.001. The test gives false positives with probability 0.1 and false negatives with probability 0.3. An individual takes the test, which turns out negative. What is the probability that the individual is actually sick?

Independence

Two events A and B are said to be independent if and only if

$$P(A \cap B) = P(A)P(B)$$

Remarks

If A and B are independent, then

$$P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B)$$

Be careful not to confuse independence with incompatibility.

2 Discrete Probability Laws

2.2.1 Bernoulli Law

Definition

The Bernoulli distribution with parameter p is the law of a discrete random variable X which takes the value 1 with probability p and the value 0 with probability $1 - p$. The associated experiment is called a **Bernoulli trial**.

Notation

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

Probability Function

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

Expectation and Variance

$$E(X) = \sum_{i=1}^{2} x_i p_i = p$$

$$Var(X) = \sum_{i=1}^{2} x_i^2 p_i - E(X)^2 = p(1 - p)$$

Example

Heads or Tails

2.2.2 Binomial Law

Definition

The binomial distribution with parameters n and p is the law of the sum X of n independent random variables Y_i such that $Y_i \sim \mathcal{B}(p)$.

Notation

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

Probability Function

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

Expectation and Variance

$$E(X) = \sum_{i=1}^{n} E(Y_i) = np$$

$$Var(X) = \sum_{i=1}^{n} Var(Y_i) = np(1-p)$$

Example

Counting the number of successes in n Bernoulli trials.

3 Continuous Probability Laws

2.3.1 Normal Law

Definition

A random variable X follows a normal law (or Gaussian law) with parameters μ and σ^2 if:

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

We denote:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Probability Function

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx$$

There is no analytical form for the cumulative distribution function F.

Expectation and Variance

If X is a real random variable such that $X \sim \mathcal{N}(\mu, \sigma^2)$, then:

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

Stability under Linear Combinations

1. If X is a random variable such that $X \sim \mathcal{N}(\mu, \sigma^2)$, then:

$$aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

2. If X and Y are two independent random variables such that $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then:

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

2.3.2 Standard Normal Law

Definition

The standard normal law is denoted by $\mathcal{N}(0, 1)$.

Property

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then: $Y = \dfrac{X - \mu}{\sqrt{\sigma^2}} \sim \mathcal{N}(0, 1)$

Notations

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

Properties

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

4 Laws Derived from the Normal Distribution

2.4.1 Chi-squared Law χ^2_n

Definition

Let $X_1, ..., X_n$ be n independent and identically distributed random variables following the standard normal distribution. The random variable $Y = X_1^2 + ... + X_n^2$ follows a continuous distribution called the Chi-squared distribution with n degrees of freedom:

$$Y = \sum_{i=1}^{n} X_i^2 \sim \chi^2_n$$

Properties

1. If $Y_1 \sim \chi^2_{n_1}$ and $Y_2 \sim \chi^2_{n_2}$ with $Y_1 \perp\!\!\!\perp Y_2$, then $Y = Y_1 + Y_2 \sim \chi^2_{n_1+n_2}$
2. If $Y \sim \chi^2_n$, then $E(Y) = n$ and $Var(Y) = 2n$

Density

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

2.4.2 Student's T Law

Definition

Let X and Y be two independent random variables such that $X \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2_n$. The random variable $T = X/\sqrt{Y/n}$ follows a continuous distribution called Student's T distribution with n degrees of freedom:

$$T = X\sqrt{Y/n} \sim t_n$$

df

1. $E(T) = 0$
2. $Var(T) = \dfrac{n}{n-2}$ if $n > 2$

Density

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