Probability Theory

沈威宇

February 7, 2025

Contents

1 Probability Theory (機率論)

I Experiment or Trial

- (Random) Experiment or Trial (（隨機）試驗): A process that can be repeated and the results may be different each time. It has repeatability, that is, the test can be repeated under the same conditions, and randomness, that is, the results of each test may be different, with randomness and uncertainty.
- Sample space (樣本空間): The set of all possible outcomes of an experiment. For example, the sample space for flipping a coin is $\{\text{head}, \text{tail}\}$.
- Event (事件): A subset of the sample space. For example, the event of rolling a die and getting an even number is $\{2, 4, 6\}$.
- Impossible event (空事件): An event with zero probability.
- Sure event (全事件): An event with one probability.
- Sum event (和事件): The sum event of event A and event B is $A \cup B$.
- Product event (積事件): The product event of event A and event B is $A \cap B$.
- Complement event (餘事件): The complement event of event A in sample space S is $A' = S \setminus A$.
- Partitions (分割或劃分): If $\{A_i\}_{i \in I}$ is a set of partitions of the sample space Ω then the following conditions are met:

$$\begin{aligned}
&\forall i \in I: \ A_i \subseteq \Omega, \\
&\forall i, j \in I \wedge i \neq j: \ A_i \cap A_j = \emptyset, \\
&\bigcup_{i \in I} A_i = \Omega.
\end{aligned}$$

- Repeated trials: A trial that consists of repeated independent trials. The probability of an event A where events A_i occured in each trial i is $\prod_i P(A_i)$.
- Objective probability (客觀機率) or Frequency probability (頻率機率): An objective probability value obtained based on past experience or statistical data, usually the frequency of past events or repeated experiments to obtain the probability of an event occurring.
- Subjective probability (主觀機率): A probability value that is not supported by statistical data.

II Probability Rules

- Probability (機率): The likelihood of an event occurring, a number between 0 and 1. The closer the probability is to 1, the more likely the event is to occur.

- Classical probability (古典機率): If the number of all possible outcomes of an event is finite or countable infinite, and the chance of each outcome occurring in the sample space is equal, then the probability of the event occurring can be calculated by:

$$P(A) = \frac{\text{The number of outcomes of event A}}{\text{The number of all possible outcomes}}$$

- Conditional probability (條件機率): The probability of another event occurring given that a certain event has occurred. Usually expressed as $P(A|B)$, which is the probability of event A occurring given that event B has occurred. $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.

- Independent events (獨立事件): Two or more events have no effect on each other, that is,

$$\left(\forall J \neq \varnothing \subseteq \{a \mid a \in \mathbb{N} \wedge 1 \leq a \leq n\} : P\left(\bigcap_{j \in J} A_j \right) = \prod_{j \in J} P\left(A_j\right) \right) \iff \left(\left(A_1, A_2, \ldots, A_n\right) \text{ are independent ev}\right.$$

- Mutually exclusive events (互斥事件): Two or more events cannot occur at the same time, that is,

$$\left(\forall J \neq \varnothing \subseteq \{a \mid a \in \mathbb{N} \wedge 1 \leq a \leq n\} : P\left(\bigcap_{j \in J} A_j \right) = 0 \text{ are mutually exclusive events.} \right)$$

- Bayes' Theorem (貝葉斯定理或貝氏定理): If $\{A_i\}_{i \in I}$ is a set of partitions of a sample space Ω, then,

$$\forall 1 \leq j \leq |I| : P(A_j|B) = \frac{P\left(A_j\right) \times P\left(B \middle| A_j\right)}{\sum_{k=1}^{|I|} P\left(A_k\right) \times P\left(B \middle| A_k\right)}$$

III Random Variable (隨機變數)

The random variable is a measurable function of which the domain is the sample space and the range is typically a subset of real numbers. If the range of it is finite or countable infinite, it is called a discrete random variable; otherwise, it is called a continuous random variable. Given ramdom variable X and sample space Ω, $X = x$ means that event $\{\omega \in \Omega \mid X(\omega) = x\}$ occured.

IV Probability Mass Function (PMF) (機率質量函數) or Probability Function (機率函數) of Discrete Random Variable

The probability mass function, denoted as $P(x)$, is a function of which the domain is the range of a discrete random variable and the codomain is $[0, 1]$, indicating the probability of events, such that the probability sum of the probability mass function of all elements in its domain is 1.

V Probability Density Function (PDF) (機率分布函數) or Probability Function (機率函數) of Continuous Random Variable

The probability density function, denoted as $f(x)$, is a function of which the domain is the range of a continuous random variable and the codomain is $[0, 1]$, indicating the probability of events, such that the probability sum of the probability density function of all elements in its domain is 1.

VI Cumulative Distribution Function (CDF)

Cumulative distribution function is a function $F : \mathbb{R} \to [0, 1]$ satisfying

$$\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to\infty} F(x) = 1.$$

For a discrete random variable X with probability mass function $P(x)$, the cumulative distribution function F is given by:

$$F(x) = P(X \le x) = \sum_{k\le x} P(X = k).$$

For a continuous random variable X with probability density function $f(x)$, the cumulative distribution function F is given by:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$

VII Probability space (機率空間)

A probability space (Ω, Σ, μ) is a measure space of which Ω is the sample space of an experiment, μ is called the probability measure, indicating the probability of events, such that $\mu : \Sigma \to [0, 1]$, $\mu(\Omega) = 1$. The composition function of the random variable and the probability function of an experiment is the probability measure of that experiment.

VIII (Mathematical) Expected Value (期望值), Expectation, or Expectancy

For a discrete random variable X with probability mass function $P(x)$, the expected value is given by:

$$E[X] = \mu_X = \sum_{x\in\mathrm{range}(X)} x \cdot P(x).$$

For a continuous random variable X with probability density function $f(x)$, the expected value is given by:

$$E[X] = \mu_X = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x.$$

IX Variance (變異數)

For a discrete random variable X with probability mass function $P(x)$, the variance is given by:

$$\mathrm{Var}(X) = {\sigma_X}^2 = E[(X - E[X])^2] = \sum_{x\in\mathrm{range}(X)} (x - E[X])^2 \cdot P(x).$$

For a continuous random variable X with probability density function $f(x)$, the variance is given by:

$$\mathrm{Var}(X) = {\sigma_X}^2 = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f(x)\,\mathrm{d}x.$$

X Standard Deviation (標準差)

The standard deviation is the positive square root of the variance.

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

XI Affine Transformation

There are two random variable X and $Y = aX + b$ where $a, b \in \mathbb{R}$, then,

$$E[X] = aE[X] + b,$$

$$\text{Var}(Y) = a^2\text{Var}(X),$$

$$\text{SD}(Y) = |a|\text{SD}(X).$$

XII Bernoulli Trial or Binomial Trial

i Bernoulli Trial (伯努力試驗) or Binomial Trial (二項試驗)

A random experiment with exactly two possible outcomes, ”success” and ”failure”, in which the probability of success is the same every time the experiment is conducted.

ii Binomial Distribution

If the random variable X follows the binomial distribution with number of Bernoulli trials $n \in \mathbb{N}$ and probability of success p, we write $X \sim B(n, p)$. The probability of getting exactly k successes in n independent Bernoulli trials (with the same success probability p) is given by the probability mass function:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

$P(X = k)$ when $X \sim B(n, p)$ has the following property:

- If $(n+1)p \in \mathbb{N} \wedge p \neq 1$, it has two maxima, which are also absolute maxima, at $k = (n+1)p$ and $k = (n+1)p - 1$; if $p = 1$, it has one maximum, which is also absolute maximum, at $k = (n+1)p - 1$; otherwise, it has one maximum, which is also absolute maximum, at $\lfloor (n+1)p \rfloor$.

  Proof.

  $$K = \arg \max_{k \in \mathbb{Z}, 0 \le k \le n} \left(\binom{n}{k} p^k (1-p)^{n-k} \right), \quad n \in \mathbb{N}, p > 0.$$

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

  $$g'(k) = \frac{(-1-n)p}{(k+1)^2(1-p)} < 0.$$

  $$K - 1 = \{\min\left(k \text{ s.t. } \frac{n-k}{k+1} > \frac{1-p}{p}\right), \min\left(k \text{ s.t. } \frac{n-k}{k+1} \ge \frac{1-p}{p}\right)\}.$$

  Solve:

  $$\frac{n - (k^* + 1)}{(k^* + 1) + 1} = \frac{1-p}{p}$$

  for k^*.

  $$np - (k^* + 1)p = (k^* + 1) + 1 - (k^* + 1) - p.$$

  $$k^* = (n+1)p.$$

  □

- The expected value of it is np.

  Proof.

$$\begin{aligned}
E[X] =& \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
=& \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k(1-p)^{n-k} \\
=& np\sum_{k=1}^{n} \binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} \\
=& np(p+1-p)^{n-1} \\
=& np.
\end{aligned}$$

□

- The variance of it is $np(1-p)$.

  Proof.

$$\begin{aligned}
E[X^2] =& \sum_{k=0}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k} \\
=& np\sum_{k=1}^{n} k\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} \\
=& np\sum_{k=1}^{n-1} k\binom{n-1}{k}p^{k}(1-p)^{n-1-k} + np \\
=& np(n-1)p + np \\
=& n^2p^2 - np^2 + np
\end{aligned}$$

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p).$$

□

iii Geometric Distribution

If the random variable X follows the geometric distribution with probability of success p, we write $X \sim G(p)$. The number of Bernoulli trials (with the same success probability p) needed to get one success is given by the probability mass function:

$$P(X = k) = p(1-p)^{k-1}.$$

$P(X = k)$ when $X \sim G(p)$ has the following property:

- Decreasing.

  Proof.

$$g(k) := p(1-p)^{k-1}.$$

$$g'(k) = p\ln(1-p)(1-p)^{k-1} < 0.$$

□

- $k > j \in \mathbb{N}$,

$$P(X = k | X > j) = P(X = k - j).$$

Proof.

$$\frac{p(1-p)^k}{(1-p)^j} = p(1-p)^{k-j}.$$

$\square$

- The expected value of it is $\frac{1}{p}$.

Proof.

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

$$(1-p)E[X] = \sum_{k=1}^{\infty} kp(1-p)^k$$

$$E[X] - (1-p)E[X] = pE[X] = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p\frac{1}{1-(1-p)} = 1.$$

$\square$

- The variance of it is $\frac{1-p}{p^2}$.

Proof.

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

$$(1-p)E[X^2] = \sum_{k=1}^{\infty} k^2 p(1-p)^k.$$

$$E[X^2] - 2E[X] + 1 - (1-p)E[X^2] = pE[X^2] - 2E[X] + 1 = 0.$$

$$E[X^2] = \frac{2E[X]-1}{p} = \frac{2-p}{p^2}.$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{1-p}{p^2}.$$

$\square$