MATH 2411 T1B Tutorial 4

HOU Zhen

zhouah@connect.ust.hk

March 10, 2026

Random Variables

Population Variance

The (population) variance of a discrete random variable X is defined as

$$\mathrm{Var}(X) = \sum_{x \in R_X} (x - \mathbb{E}[X])^2 p_X(x) = \mathbb{E}[(X - \mathbb{E}[X])^2],$$

where $p_X(x)$ is the probability mass function of X, if $\sum_{x \in R_X} x^2 p_X(x) < +\infty$ which guarantees that $\mathbb{E}[X]$ and $\mathrm{Var}(X)$ exist.

Population standard deviation $\mathrm{SD}(X) := \sqrt{\mathrm{Var}(X)}$.

Properties of Mean and Variance

1. $\mathbb{E}[g(X)] = \sum_{x \in R_X} g(x) p_X(x)$
2. $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
3. $\text{Var}(aX + b) = a^2 \text{Var}(X)$
4. $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Sample Mean and Sample Variance

Let the data values be $x_1, x_2, \ldots, x_n$, then the sample mean and sample variance are defined as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

and

$$s_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

respectively.

The sample standard deviation is defined as

$$s_{n-1} = \sqrt{s_{n-1}^2}$$

Exercise

Prove that $s_{n-1}^2 = \frac{1}{n-1}\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right)$.

Continuous Random Variables

Definition

When the range of a random variable X consists of intervals on the real line, we call it a continuous random variable.

Probability density function

A non-negative function $f(x) \geq 0$ is called a probability density function (pdf) of a rv X, if for any $a \leq b$ (can also be $\pm\infty$),

$$P(a < X \leq b) = \int_a^b f(x)\, dx.$$

Properties of pdf

1. $f_X(x) \geq 0$ for all x in $\mathbb{R}$.
2. $\int_{-\infty}^{+\infty} f_X(x)dx = 1$.
3. $\mathbb{P}(X = a) = 0$ for all a in $\mathbb{R}$.
4. $\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$.
5. $F_X(a) = \int_{-\infty}^a f_X(x)dx$.

Exercise

Express the cdf of $X^+ = \max(X, 0)$ in terms of F_X (the cdf of X).

Population Mean and Population Variance

The (population) mean of a continuous random variable X is defined as

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x) dx,$$

if $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$.

The (population) variance of a continuous random variable X is defined as

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 f_X(x) dx,$$

if $\int_{-\infty}^{\infty} x^2 f_X(x) dx < \infty$.

The standard deviation $\mathrm{SD}(X) := \sqrt{\mathrm{Var}(X)}$.

Properties of Mean and Variance

1. $\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$
2. $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
3. $\text{Var}(aX + b) = a^2 \text{Var}(X)$
4. $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Exercise

X is a continuous random variable with pdf $f_X(x) = ce^{-|x|}$ for all $x \in \mathbb{R}$.
Find the mean of X.

Joint distribution

Discrete case For two discrete r.v.s X and Y with range R_X and R_Y. Their joint pmf is defined as $p_{X,Y}(x,y) = P(\{X = x\} \cap \{Y = y\}), x \in R_X, y \in R_Y$.

- $0 \le p_{X,Y}(x,y) \le 1$.
- $\sum_{x \in R_X, y \in R_Y} p_{X,Y}(x,y) = 1$.
- $\sum_{y \in R_Y} p_{X,Y}(x,y) = p_X(x)$.

Joint distribution (continued)

Continuous case For two continuous r.v.s X and Y, their joint pdf $f_{X,Y}(x,y)$ satisfies $\int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x,y)dydx = P(X \leq a, Y \leq b)$, for any a, b.

- $f_{X,Y}(x,y) \geq 0$.
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx = 1$.
- $\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = f_X(x)$.

We can also define the joint distribution of more than two r.v.s in a similar way.

Expectation and Variance of Functions

With joint distribution, we can define the expectation and variance of a function of several r.v.s. Take two r.v.s X and Y as an example.

Expectation $\mathbb{E}[g(X, Y)] =$

- (Discrete Case) $\sum_{x \in R_X, y \in R_Y} g(x, y) p_{X,Y}(x, y)$,
- (Continuous Case) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dy dx$.

Variance $\mathrm{Var}[g(X, Y)] =$

- (Discrete Case) $\sum_{x \in R_X, y \in R_Y} [g(x, y) - \mathbb{E}[g(X, Y)]]^2 p_{X,Y}(x, y)$,
- (Continuous Case) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [g(x, y) - \mathbb{E}[g(X, Y)]]^2 f_{X,Y}(x, y) dy dx$.

Independence

Two r.v.s X and Y are independent if

- (Discrete Case) $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ for all $x \in R_X, y \in R_Y$.
- (Continuous Case) $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x, y.

Recall that for two events A and B, they are independent if $P(A \cap B) = P(A)P(B)$.

Generalization: For r.v.s $X_1, \ldots, X_n$, they are independent if

- (Discrete) $P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$
- (Continuous) $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$.

Example

Find $X_1, \ldots, X_n$ such that they are pairwise independent (any two of them are independent) but not independent.

Properties

- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
- $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$ if X and Y are independent.
- $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ if X and Y are independent and f, g are continuous. In particular, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Exercise

Given independently and identically distributed (i.i.d.) r.v.s $X_1, \ldots, X_n$ with mean μ and variance σ^2.

(1) Find the mean and variance of $\bar{X} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$.

(2) Find the mean of $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$.

Remark: Connections between Probability and Statistics.