STA2002: Probability and Statistics II

Introduction and Preliminary

Fangda Song and Ka Wai Tsang

School of Data Science, CUHK(SZ)

September, 2025

Outline

1. Course Arrangement
2. Probability vs Statistics
3. Random variable and distribution function
4. Discrete Distributions
5. Continuous Distributions
6. Two Limit Theorems
7. Student's Theorem

Course Instructors

- Fangda Song
  - Lecture time: Tue/Thu 10:30-11:50 am
  - Office: Rm 420d, Daoyuan Building
  - Office hour: Tue 3:30-4:30 pm
- Ka Wai Tsang
  - Lecture time: Mon/Wed 8:30-9:50 am
  - Office: Rm 505b, Daoyuan Building
  - Office hour: Mon 10:30-11:30 am

Teaching Assistants

- Ruicong Wang
  - Office: Rm 311, Research Building B
  - Office hour: Mon 10:00-11:00 am
- Wendi Ren
  - Office: Rm 430, Zhi Xin Building
  - Office hour: Thu 4:00-5:00 pm
- Bokun Yu
  - Office: Rm 313, H.L.Tu Building
  - Office hour: Thu 4:00-5:00 pm
- Zhiqi Gao
  - Office hour: Mon 10:00-11:00 am
  - Tecent Meeting: 748-5967-3028

Undergraduate Student Teaching Fellow

- Frederick Khasanto
  - Email: frederickkhasanto@link.cuhk.edu.cn
- Xiaoxia Sheng
  - Email: 123090494@link.cuhk.edu.cn
- Liying Xi
  - Email: 122090594@link.cuhk.edu.cn
- Zisheng Qu
  - Email: 123090470@link.cuhk.edu.cn

Tutorials

Tutorials begin on the second week, i.e. the week of Sept 8th

T01-TUT (3841)	Session	Status
T01-TUT (3841)	1	

Days	Start	End	Room	Instructor	Dates
Mo	19:00	19:50	Teaching C Bldg 208	任文頔 (REN, Wendi), 王瑞聪 (WANG, Ruicong), 郁伯昆 (YU, Bokun), 高治淇 (GAO, Zhiqi)	2025/09/01 - 2025/12/12

T02-TUT (3842)	Session	Status
T02-TUT (3842)	1	

Days	Start	End	Room	Instructor	Dates
Mo	20:00	20:50	Teaching C Bldg 208	任文頔 (REN, Wendi), 王瑞聪 (WANG, Ruicong), 郁伯昆 (YU, Bokun), 高治淇 (GAO, Zhiqi)	2025/09/01 - 2025/12/12

T03-TUT (3843)	Session	Status
T03-TUT (3843)	1	

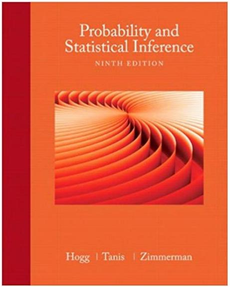
Days	Start	End	Room	Instructor	Dates
Th	19:00	19:50	Teaching C Bldg 208	任文頔 (REN, Wendi), 王瑞聪 (WANG, Ruicong), 郁伯昆 (YU, Bokun), 高治淇 (GAO, Zhiqi)	2025/09/01 - 2025/12/12

Section	Session	Status
T04-TUT (3844)	1	

Days	Start	End	Room	Instructor	Dates
Th	20:00	20:50	Teaching C Bldg 208	任文頔 (REN, Wendi), 王瑞聪 (WANG, Ruicong), 郁伯昆 (YU, Bokun), 高治淇 (GAO, Zhiqi)	2025/09/01 - 2025/12/12

Reference book

Hogg, R. V., Tanis, E. A. Zimmerman, D. L. (2015) Probability and Statistical Inference, 9th edition, Pearson



- The first two lectures will review some essential concepts from STA2001, which corresponds to Chapter 1 to 5.
- This course will focus on Chapter 6 to Chapter 9

Assessment scheme

Component/ method	% Weight
Assignments	25
Mid-Term Exam	30
Final Exam	45

- You are expected to work on all assignments on your own. **No late assignment will be accepted.** No exceptions. The lowest assignment score will be dropped.
- Both the midterm and final exams will be closed book, closed notes and no devices. However, you are allowed to bring in one cheat sheet for the midterm and up to two cheat sheets for the final. The cheat sheets should be handwritten and in A4 size. You can write on both the front and back of the cheat sheet(s).

Important notes

- A makeup exam is allowed ONLY under provable unforeseen circumstances beyond the control of the student. Only a written request for a makeup exam that is sent BEFORE the exam date and is accompanied by all relevant supporting documentation will be considered.
- Any form of academic dishonesty, which includes, but is not limited to, cheating in examinations, plagiarism and sharing of assignments with others, will NOT be tolerated. Any related offense will lead to disciplinary action including termination of studies at the University.

Tentative plan

Week	Content/ topic/ activity
1	Review of families of distributions: Gamma, Normal, Exponential, Chi-square, t-distribution, F-distribution, Bernoulli, Binomial, etc.; Moment generating function; Central Limit Theorem.
2	Maximum Likelihood Estimation; Method of moments; Unbiased estimation.
3	Confidence interval (CI) for means; CI for the difference of two means; CI for proportions.
4	Tests of statistical hypotheses, Critical region, Significance level, p-value, Power, t-test, F-test.
5	More t-test; Tests about proportions.
6	Sample size; Power of a statistical test; Order statistics; QQ plots.
7	Review Session. **Midterm Exam**.
8	Distribution-free CIs for Percentiles; The Wilcoxon tests.
9	Chi-square goodness-of-fit tests; Contingency tables.
10	Testing for homogeneity; Testing for independence; One-way ANOVA.
11	Two-way ANOVA; Introduction to regression.
12	More regression; Tests concerning regression and correlation.
13	Likelihood ratio tests.
14	Review Session. **Final Exam**.

Introduction

- Our first two lectures are warm-up lectures.
- We will review some essential concepts from STA2001, which will serve as our foundation in learning various advanced statistical concepts in STA2002 later on.
- Suggested reading: Chapter 1 to Chapter 5 in the book.

Probability vs Statistics

Probability

Given a **mathematical model** of real world, figure out **what will happen**.

Statistics

Given **data/observations** of what happened, figure out **what model or process explains what happened**.

Probability vs Statistics

Questions in Probability

When we flip a fair coin five times, what is the probability of observing the same side coming up all five times?

Questions in Statistics

- You flip the coin five times and see it landed on Heads each time. Should you conclude the coin is not a fair coin?
- You flip the coin 20 times and see Heads 14 times. What is your estimate of the probability of landing on Heads?
- What if you flip the coin 1000 times and see Heads 700 times? Do you feel equally confident about the estimates in these two scenarios?

Probability vs Statistics

In statistics,

- Compare statistical measurements with what probability predicts;
- Use statistical measurements to make inferences about the probabilistic model;
- Evaluate how much confidence we have in these inferences.

Key concepts:

- Hypothesis testing
- Parameter estimation
- Confidence interval

Random variable and distribution function

Definition

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

- Discrete: takes values in a countable set S_X;
- Continuous: described by a density.

The cumulative distribution function (CDF) of X is

$$F_X(x) = \Pr(X \le x).$$

Property: $F_X(x)$ is a non-decreasing, right-continuous, and $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$.

Support of a random variable

- Discrete: $S_X = \{x : \Pr(X = x) > 0\}$
- Continuous: $S_X = \{x : f_X(x) > 0\}$

Discrete case

Definition and properties

For a discrete X, the probability mass function (PMF) is

$$p_X(k) = \Pr(X = k), \qquad k \in S_X.$$

Properties:

- $0 \leq p_X(k) \leq 1$;
- $\displaystyle\sum_{k \in S_X} p_X(k) = 1$;

Continuous case

Definition and properties

If X is continuous, the probability density function (PDF) $f_X(x)$ satisfies

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

In other words, $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$.
When F_X is differentiable, $f_X(x) = F_X'(x)$.

Expectation

Definition

$$\mathbb{E}(X) = \begin{cases} \displaystyle\sum_{k \in S_X} k\, p_X(k), & X \text{ is discrete,} \\ \displaystyle\int_{-\infty}^{\infty} x\, f_X(x)\, dx, & X \text{ is continuous.} \end{cases}$$

Linearity of expectation

For constants a, b,

$$\mathbb{E}(a + bX) = a + b\mathbb{E}(X), \qquad \mathbb{E}\left(\sum_i b_i X_i\right) = \sum_i b_i\, \mathbb{E}(X_i).$$

Variance

Definition

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right] = \mathbb{E}[X^2] - (\mathbb{E}X)^2.$$

Property of variance

$$\mathrm{Var}(a + bX) = b^2\ \mathrm{Var}(X).$$

If X, Y have finite variances, then

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\ \mathrm{Cov}(X, Y),$$

where $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$.
If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$, hence

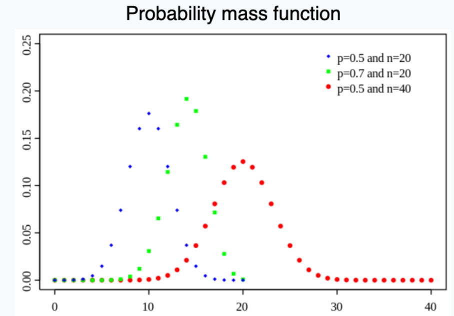
$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \quad \text{(independent)}.$$

Binomial distribution

A discrete random variable X has a binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0,1)$ if its probability mass function (PMF) is given by

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

The mean and variance of X are $\mathbb{E}(X) = np, \mathrm{Var}(X) = np(1-p)$.



Discrete Distributions

Poisson distribution

A discrete random variable X has a Poisson distribution with rate $\lambda > 0$ if its probability mass function (PMF) is

$$\Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \ldots$$

Its moment generating function (MGF) is

$$M_X(t) = \mathbb{E}\left(e^{tX}\right) = \exp\left(\lambda(e^t - 1)\right), \qquad t \in \mathbb{R}.$$

- What are the mean $\mathbb{E}(X)$ and the variance $\mathrm{Var}(X)$ of X?

Moment Generating Function

Definition

For a random variable X, the moment generating function (MGF) is

$$M_X(t) \;=\; \mathbb{E}\Big(e^{tX}\Big)\,, \qquad t \in \mathbb{R}.$$

Properties

- $M_X(0) = \mathbb{E}(e^0) = 1$.
- If $M_X(t)$ is finite on an open interval containing 0 (i.e., $t \in (-h, h)$ for some $h > 0$), then M_X **uniquely determines** the distribution of X.
- For continuous X, $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$.
- For discrete X, $M_X(t) = \sum_x e^{tx} p_X(x)$.

Moments from the MGF

Moments as derivatives

If M_X exists in a neighborhood of 0 and is differentiable there, then for $k \geq 1$,

$$M_X^{(k)}(0) = \frac{d^k}{dt^k}\mathbb{E}\left(e^{tX}\right)\Big|_{t=0} = \mathbb{E}\left(X^k\right).$$

Mean and variance

$$\mathbb{E}[X] = M_X'(0), \qquad \mathrm{Var}(X) = M_X''(0) - \left(M_X'(0)\right)^2.$$

Moments of Poisson distributions

For $X \sim \mathrm{Pois}(\lambda)$, differentiate its MGF $M_X(t) = \exp\big(\lambda(e^t - 1)\big)$ to obtain $M_X'(t) = \lambda e^t M_X(t), M_X''(t) = \lambda e^t M_X(t) + \lambda^2 e^{2t} M_X(t)$.
Let $t = 0$, and then $M_X'(0) = \lambda$, $M_X''(0) = \lambda + \lambda^2$.
Thus, $\mathbb{E}(X) = \lambda, \mathrm{Var}(X) = \lambda$.

Continuous Distributions

Exponential distribution

A random variable X has an exponential distribution with parameter $\theta > 0$ if its probability density function (PDF) is given by

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad x > 0.$$

Its moment generating function (MGF) is given by

$$M(t) = \mathbb{E}(e^{tX}) = \frac{1}{1 - \theta t}, \quad t < \frac{1}{\theta}.$$

- What is the mean $\mathbb{E}(X)$ and the variance $\mathrm{Var}(X)$ of X?

Continuous distributions

Gamma distribution

A random variable X has a gamma distribution with shape $\alpha > 0$ and scale $\theta > 0$, written as $X \sim \text{Gamma}(\alpha, \theta)$, if its PDF is given by

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, \quad x > 0.$$

Its moment generating function (MGF) is given by

$$M(t) = \mathbb{E}(e^{tX}) = \frac{1}{(1-\theta t)^\alpha}, \quad t < \frac{1}{\theta}.$$

The mean and variance of X are

$$\mathbb{E}(X) = \alpha\theta, \quad \text{Var}(X) = \alpha\theta^2.$$

Continuous distributions

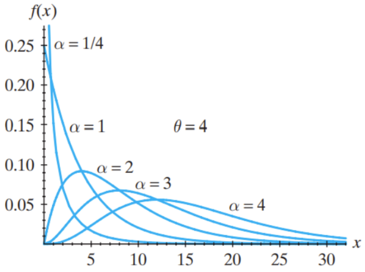


Figure 3.2-2 Gamma pdfs: $\theta = 4, \alpha = 1/4, 1, 2, 3, 4$; $\alpha = 4, \theta = 5/6, 1, 2, 3, 4$

- Connection to exponential distributions

$$X \sim \text{Exp}(\theta) \Leftrightarrow X \sim \text{Gamma}(1, \theta)$$

- Summation:

$$X_1 \sim \text{Gamma}(\alpha_1, \theta), X_2 \sim \text{Gamma}(\alpha_2, \theta)$$
$$\Rightarrow X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \theta)$$

Application Example

Waiting time

Suppose the number of customers per hour arriving at a shop follows a Poisson process with mean 30. That is, the waiting time in minutes between any two customers follows a exponential distribution with parameter $\theta = \frac{1}{30/60} = 2$. What is the probability that the shopkeeper will wait more than 5 minutes before the first two customers arrive?

Solution: Let X denotes the waiting time in minutes until the second customer arrives. Thus, X follows a gamma distribution with shape $\alpha = 2$, scale $\theta = 2$. Hence,

$$\begin{aligned}
P(X > 5) &= \int_5^\infty \frac{x^{2-1} e^{-x/2}}{\Gamma(2)\, 2^2}\, dx = \int_5^\infty \frac{xe^{-x/2}}{4}\, dx \\
&= \frac{1}{4}\left[(-2x)e^{-x/2} - 4e^{-x/2}\right]_5^\infty \\
&= \frac{7}{2} e^{-5/2} = 0.287.
\end{aligned}$$

Continuous distributions

Chi-square distribution

A random variable X has a chi-square distribution with positive integer parameter r, known as the degree of freedom, written as $X \sim \chi^2(r)$, if it is a Gamma distribution with $\alpha = r/2$ and $\theta = 2$, i.e., $X \sim \text{Gamma}(r/2, 2)$, and its PDF is given by

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2}, \quad x > 0.$$

Its moment generating function (MGF) is given by

$$M(t) = \mathbb{E}(e^{tX}) = \frac{1}{(1-2t)^{r/2}}, \quad t < \frac{1}{2}.$$

The mean and variance of X are

$$\mathbb{E}(X) = r, \quad \text{Var}(X) = 2r.$$

Continuous distributions

Normal distribution

A random variable X has a normal distribution with mean μ and variance σ^2, written as $X \sim N(\mu, \sigma^2)$, if its PDF is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty.$$

Its moment generating function (MGF) is given by

$$M(t) = E(e^{tX}) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, \quad t \in \mathbb{R}.$$

The mean and variance of X are

$$E(X) = \mu, \quad \mathrm{Var}(X) = \sigma^2.$$

A random variable Z is said to be a standard normal if $Z \sim N(0, 1)$.

Property of Normal distributions

Theorem 7 (Theorem 3.3-1 in book)

If $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$.

Theorem 8 (Theorem 3.3-2 in book)

If $X \sim N(\mu, \sigma^2)$, then $Z^2 = (X - \mu)^2/\sigma^2 \sim \chi^2(1)$.

These two theorems will be discussed more in class.

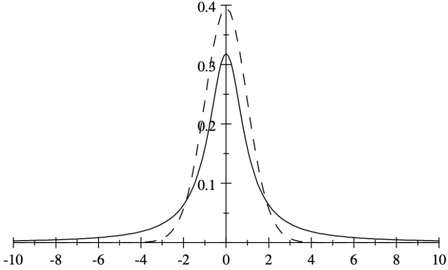
Continuous distributions

Student t-distribution

Let

$$T := \frac{Z}{\sqrt{U/r}},$$

where $Z \sim N(0,1)$, $U \sim \chi^2(r)$, and Z and U are independent. Then T follows the t-distribution with r degrees of freedom.



Plot of the $t(1)$ (solid line) and the $t(30)$ (dashed line) density functions

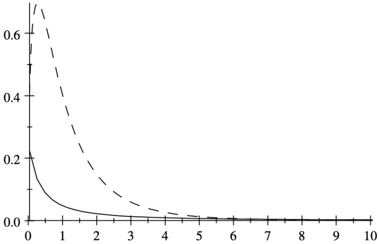
Continuous distributions

F-distribution

Let

$$F := \frac{U/r_1}{V/r_2},$$

where $U \sim \chi^2(r_1)$, $V \sim \chi^2(r_2)$, and U and V are independent. Then F follows the F-distribution with r_1 and r_2 degrees of freedom.



Plot of the $F(2, 1)$ (solid line) and the $F(3, 10)$ (dashed line) density functions.

Law of Large Numbers

Theorem 9 (Law of Large Number, Section 5.8 in textbook)

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. with common mean μ and finite variance σ^2. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\bar{X} - \mu| > \epsilon) = 0.$$

We say that the sample mean $\bar{X}$ converges in probability to the expected value μ.

The law of large numbers says that as you repeat an experiment many times, the average of the results gets closer and closer to the true expected value.

Central Limit Theorem (CLT)

Theorem 10 (Central Limit Theorem, Theorem 5.6-1 in textbook)

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. with common mean μ and finite variance σ^2. Then the distribution of

$$W = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

is $N(0, 1)$ in the limit as $n \to \infty$. We say that the random variable W converges in distribution to a standard normal distribution.

A large sum of identical and independently distributed (i.i.d.) random variables, properly normalized, will always have approximately a normal distribution.

Central Limit Theorem

Let $X_1, X_2, \ldots$ be i.i.d. random variables, each with the Bernoulli(θ) distribution with $\mathbb{E}(X_i) = \theta, \mathrm{Var}(X_i) = \theta(1-\theta)$.
The binomial Bin(n, θ) random variable is the sum of n Bernoulli variables $\Rightarrow$ we can use the CLT.

Let

$$Y_n = \sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n, \theta).$$

As $n \to \infty$,

$$\frac{Y_n - \mathbb{E}Y_n}{\sqrt{\mathrm{Var}(Y_n)}} = \frac{Y_n - n\,\mathbb{E}X_i}{\sqrt{n\,\mathrm{Var}(X_i)}} = \frac{Y_n - n\theta}{\sqrt{n\theta(1-\theta)}} \to \mathcal{N}(0,1),$$

Therefore, for any y,

$$\Pr(Y_n \leq y) = \Pr\left(\frac{Y_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{y - n\theta}{\sqrt{n\theta(1-\theta)}}\right) \approx \Phi\left(\frac{y - n\theta}{\sqrt{n\theta(1-\theta)}}\right),$$

where Φ is the CDF of $\mathcal{N}(0,1)$.

Half-unit (Continuity) Correction

When a **discrete** distribution (e.g., Binomial, Poisson) is approximated by a **continuous** Normal, probability at a single integer is represented by an interval of width 1. The half-unit correction shifts integer cutoffs by ± 0.5 for better approximation.

Rules of thumb (for X approximately $\mathcal{N}(\mu, \sigma^2)$)

$$\Pr(X \le x) \approx \Phi\left(\frac{x + 0.5 - \mu}{\sigma}\right),$$

$$\Pr(X \ge x) \approx 1 - \Phi\left(\frac{x - 0.5 - \mu}{\sigma}\right),$$

$$\Pr(a \le X \le b) \approx \Phi\left(\frac{b + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{a - 0.5 - \mu}{\sigma}\right).$$

Here Φ is the standard Normal CDF.

Central Limit Theorem

Example

Suppose we have a biased coin with head probability $\theta = 0.6$. We toss the coin $n = 1000$ times. Approximately calculate the probability of getting at least 550 heads and no more than 625 heads.

Let $Y \sim \text{Bin}(1000, 0.6)$ denote the number of heads observed.

$$\mathbb{E}(Y) = 1000 \times 0.6 = 600, \text{Var}(Y) = 1000 \times 0.6 \times 0.4 = 240.$$

Hence,

$$\begin{aligned}
\Pr(550 \leq Y \leq 625) &= \Pr\left(\frac{549.5 - 600}{\sqrt{240}} \leq \frac{Y - 600}{\sqrt{240}} \leq \frac{625.5 - 600}{\sqrt{240}}\right) \\
&\approx \Pr(-3.2598 \leq Z \leq 1.646) \\
&= \Phi(1.646) - \Phi(-3.2598) \approx 0.9496,
\end{aligned}$$

where $Z \sim \mathcal{N}(0, 1)$.

Student's Theorem

Theorem 11 (Student's Theorem)

Let $X_1, \ldots, X_n$ be i.i.d. random variables each having a normal distribution with mean μ and variance σ^2. Define the random variables

$$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Then

1. $\bar{X}$ has a $N(\mu, \frac{\sigma^2}{n})$ distribution.
2. $\bar{X}$ and S^2 are independent.
3. $(n-1)S^2/\sigma^2$ has a $\chi^2(n-1)$ distribution.
4. The random variable $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ has a Student t-distribution with $n-1$ degrees of freedom.