Chapter 2: Preliminary

STAT2602A Probability and statistics II
(2024-2025 1st Semester)

Contents

2.1 Moment generating function

- **Introduction** Let r be a positive integer. The r-th moment about the origin of a random variable X is defined as $\mu_r = \mathrm{E}(X^r)$. In order to calculate μ_r, we can make use of the moment generating function (m.g.f.).
- **Definition:** (Moment Generating Function) A moment generating function of X is a function of $t \in \mathcal{R}$ defined by $M_X(t) = \mathrm{E}(e^{tX})$ if exists.
- **Property 2.1** Suppose $M_X(t)$ exists. Then,

  (1) $M_X(t) = \sum_{r=0}^{\infty} \mu_r \left(\frac{t^r}{r!}\right)$;

  (2) $\mu_r = M_X^{(r)}(0)$ for $r = 1, 2, \ldots$;

  (3) For constants a and b, $M_{aX+b}(t) = e^{bt} M_X(at)$.

(Please refer to lecture notes for detailed proof of Property 2.1)

2.1 Moment generating function

- **Property 2.2** If $M_X(t)$ exists, there is a one-to-one correspondence between $M_X(t)$ and the p.d.f. $f(x)$ (or c.d.f. $F(x)$). (Proof of Property 2.2 is omitted)

Remark The above property shows that the distribution of X can be obtained by calculating its m.g.f.

- *Example 2.1* The m.g.f. of $\mathrm{N}(\mu, \sigma^2)$ is

$$
\begin{aligned}
\mathrm{E}(e^{tX}) &= \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-2\sigma^2 tx + x^2 - 2\mu x + \mu^2}{-2\sigma^2}} \mathrm{d}x \\
&= e^{\frac{-2\mu\sigma^2 t - \sigma^4 t^2}{-2\sigma^2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu-\sigma^2 t)^2}{-2\sigma^2}} \mathrm{d}x \\
&= e^{\mu t + \frac{1}{2}\sigma^2 t^2},
\end{aligned}
$$

because $\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu-\sigma^2 t)^2}{-2\sigma^2}}$ is the density function of $\mathrm{N}(\mu + \sigma^2 t, \sigma^2)$.

2.1 Moment generating function

- *Example 2.2* Find the m.g.f. of a random variable X following a Poisson distribution with mean λ.
  Solution.

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{\infty} e^{tx}\mathrm{P}(X=x) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}. \end{aligned}$$

$\square$

- *Example 2.3* Find the m.g.f. of a random variable which has a (probability) density function given by

$$f(x) = \begin{cases} e^{-x}, & \text{for } x > 0; \\ 0, & \text{otherwise,} \end{cases}$$

and then use it to find μ_1, μ_2, and μ_3.

Solution of example 2.3.

$$M_X(t) = \mathrm{E}(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f(x)\mathrm{d}x$$

$$= \int_0^{+\infty} e^{tx} e^{-x}\mathrm{d}x = \begin{cases} \left.\dfrac{e^{(t-1)x}}{t-1}\right|_0^{+\infty} = \dfrac{1}{1-t}, & \text{for } t < 1; \\ \text{does not exist}, & \text{for } t \geq 1. \end{cases}$$

Then,

$$\begin{aligned}
\mu_1 &= M_X^{(1)}(0) = \left.\frac{1}{(1-t)^2}\right|_{t=0} = 1, \\
\mu_2 &= M_X^{(2)}(0) = \left.\frac{2}{(1-t)^3}\right|_{t=0} = 2, \\
\mu_3 &= M_X^{(3)}(0) = \left.\frac{2 \times 3}{(1-t)^4}\right|_{t=0} = 3!.
\end{aligned}$$

2.1 Moment generating function

- **Property 2.3** If $X_1, X_2, \dots, X_n$ are independent random variables, $M_{X_i}(t)$ exists for $i = 1, 2, \cdots, n$, and $Y = X_1 + X_2 + \cdots + X_n$, then $M_Y(t)$ exists and

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

Proof of Property 2.3 is left as an exercise.

- *Example 2.4* Find the distribution of the sum of n independent random variables $X_1, X_2, \dots, X_n$ following Poisson distributions with means $\lambda_1, \lambda_2, \dots, \lambda_n$ respectively.
  Solution. Let $Y = X_1 + X_2 + \cdots + X_n$. Then,

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} e^{\lambda_i(e^t-1)} = e^{(e^t-1)\sum_{i=1}^{n}\lambda_i},$$

  which is the m.d.f. of Poisson random variable with mean $\sum_{i=1}^{n}\lambda_i$. Hence, by Example 2.2 and Property 2.3, $Y \sim$ Poisson distribution with mean $\sum_{i=1}^{n}\lambda_i$. $\square$

2.1 Moment generating function

- *Example 2.5* For positive numbers α and λ, find the moment generating function of a gamma distribution $\mathrm{Gamma}(\alpha, \lambda)$ of which the density function is given by

$$f(x) = \begin{cases} \dfrac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & \text{for } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

- *Example 2.6* Find the distribution of the sum of n independent random variables $X_1, X_2, \ldots, X_n$ where X_i follows $\mathrm{Gamma}(\alpha_i, \lambda)$, $i = 1, 2, \ldots, n$, with the p.d.f. given by

$$f(x) = \begin{cases} \dfrac{\lambda^{\alpha_i} x^{\alpha_i-1} e^{-\lambda x}}{\Gamma(\alpha_i)}, & \text{for } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

2.1 Moment generating function

- *Example 2.7* Prove that the sum of n independent random variables $X_1, X_2, \ldots, X_n$ each following a Bernoulli distribution with parameter p follows $\mathrm{B}(n, p)$, the binomial distribution with parameters n and p.
- *Example 2.8* Let $X_1, X_2, \ldots, X_n$ be independent $\mathrm{N}(0, 1)$ random variables. Show that $Y = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$.

Detailed answers of example 2.5-2.8 can be found in the lecture notes from page 13-15

2.2 Convergence

- **Brief Introduction**
  - **Statistics:** functions of random sample $\mathbf{X} = \{X_1, \cdots, X_n\}$
  - **Important Statistics:** sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$; sample variance $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

  Although in a particular sample, say $x_1, \cdots, x_n$, we observe definite values of these statistics, $\overline{x}$ and s^2, we should recognize that each value is only one observation of the respective random variables $\overline{X}$ and S^2. That is, each $\overline{X}$ or S^2 is also a random variable with its own distribution.
  Suppose that the random sample $\mathbf{X}$ from a distribution $F(x)$ with mean $\mu = \mathrm{E}(X)$ and variance $\sigma^2 = \mathrm{Var}(X)$. When n is large, Theorem 1.4 shows that $F(x)$ can be well approximated by $F_n(x)$.
  Meanwhile, we can easily show that $\overline{X}$ and S^2 are the mean and variance of a random variable from a distribution $F_n(x)$. Therefore, it is expected that when n is large, μ and σ^2 can be well approximated by $\overline{X}$ and S^2, respectively.

2.2 Convergence

- **Definition** (*Convergence in probability*) Let $(Z_n; n \geq 1)$ be a sequence of random variables. We say the sequence Z_n converges in probability to Z if, for any $\epsilon > 0$,

$$\mathrm{P}(|Z_n - Z| > \epsilon) \to 0 \quad \text{as } n \to \infty.$$

For brevity, this is often written as $Z_n \to_p Z$.

Remark 2.1

1. For a deterministic sequence $\{a_n\}$,

$$\begin{aligned} a_n \to a \text{ as } n \to \infty \iff & \text{ for any } \epsilon > 0, \text{ there exists an integer} \\ & N(\epsilon) > 0, \text{such that when } n \geq N, \\ & |a_n - a| < \epsilon \text{ (for sure!)} \end{aligned}$$

2. For a random sequence $\{Z_n\}$,

$$\begin{aligned} Z_n \to_p Z \text{ as } n \to \infty \iff & \text{ for any } \epsilon > 0, \text{ there exists an integer} \\ & N(\epsilon) > 0, \text{such that when } n \geq N, \\ & \mathrm{P}(|Z_n - Z| < \epsilon) \text{ is very close} \\ & \text{to one (but not for sure!)} \end{aligned}$$

2.2 Convergence

Remark 2.1 (con't)

3. ϵ specifies the accuracy of the convergence, which can be achieved for large $n (\geq N)$.

- **Theorem 2.1** (Weak law of large numbers (LLN)) Let $(X_i; i \geq 1)$ be a sequence of independent random variables having the same finite mean and variance, $\mu = \mathrm{E}(X_1)$ and $\sigma^2 = \mathrm{Var}(X_1)$. Then, as $n \to \infty$,

$$\overline{X} \to_p \mu.$$

  It is customary to write $S_n = \sum_{i=1}^{n} X_i$ for the partial sums of the X_i.

(Detailed proof of LLN is in lecture notes page 16)

2.2 Convergence

- **Property 2.4** (Chebyshov's inequality) Suppose that $\mathrm{E}(X^2) < \infty$. Then, for any constant $a > 0$,

$$\mathrm{P}(|X| \geq a) \leq \frac{\mathrm{E}(X^2)}{a^2}.$$

(Proof of property 2.4 is left as an exercise.)

- **Property 2.5** If $X_n \to_p \mu$ and $Y_n \to_p \nu$, then (i) $X_n + Y_n \to_p \mu + \nu$; (ii) $X_n Y_n \to_p \mu\nu$; (iii) $X_n / Y_n \to \mu/\nu$ if $Y_n \neq 0$ and $\nu \neq 0$; (iv) $g(X_n) \to_p g(\mu)$ for a continuous function $g(\cdot)$.

(Proof of property 2.5 is omitted.)

2.2 Convergence

- **Example 2.9** Let $(X_i; i \geq 1)$ be a sequence of independent random variables having the same finite mean $\mu = \mathrm{E}(X_1)$, finite variance $\sigma^2 = \mathrm{Var}(X_1)$, and finite fourth moment $\mu_4 = E(X_1^4)$. Show that

$$S^2 \to_p \mathrm{Var}(X_1).$$

(Hint: $S^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2$)

2.2 Convergence

- **Definition:** (*Convergence in distribution*) Let $(Z_n; n \geq 1)$ be a sequence of random variables. We say the sequence Z_n converges in distribution to Z if, as $n \to \infty$,

$$G_n(x) \to G(x) \quad \text{whereever } G(x) \text{ is continuous.}$$

  Here, $G_n(x)$ and $G(x)$ are the c.d.f. of Z_n and Z, respectively.

- **Theorem 2.2** (Central limit theorem (CLT)) Let $(X_i; i \geq 1)$ be a sequence of independent random variables having the same finite mean and variance, $\mu = \mathrm{E}(X_1)$ and $\sigma^2 = \mathrm{Var}(X_1)$. Then, as $n \to \infty$,

$$\frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{Var}(\overline{X})}} = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \to_d \mathrm{N}(0, 1).$$

Central limit theorem shows that $\overline{X} \sim \mathrm{N}\big(\mathrm{E}(\overline{X}), \mathrm{Var}(\overline{X})\big)$, and hence it tells us the distribution of $\overline{X}$ when the sample size n is large.

2.2 Convergence - Simulation study of CLT

(1) Generate a realization $\{x_1, x_2, \cdots, x_n\}$ of the independent random sample $\{X_1, X_2, \cdots, X_n\}$ from $\mathrm{N}(0, 1)$;

(2) Calculate $z_n = \sqrt{n}(\overline{x} - \mu)/\sigma$ with $\mu = 0$ and $\sigma = 1$;

(3) Repeat (1)-(2) J times to get $\{z_n^{(1)}, z_n^{(2)}, \cdots, z_n^{(J)}\}$, which is a sequence of realizations of Z_n, where $Z_n = \sqrt{n}(\overline{X} - \mu)/\sigma$;

(4) Plot the (relative frequency) histogram of $\{z_n^{(1)}, z_n^{(2)}, \cdots, z_n^{(J)}\}$.

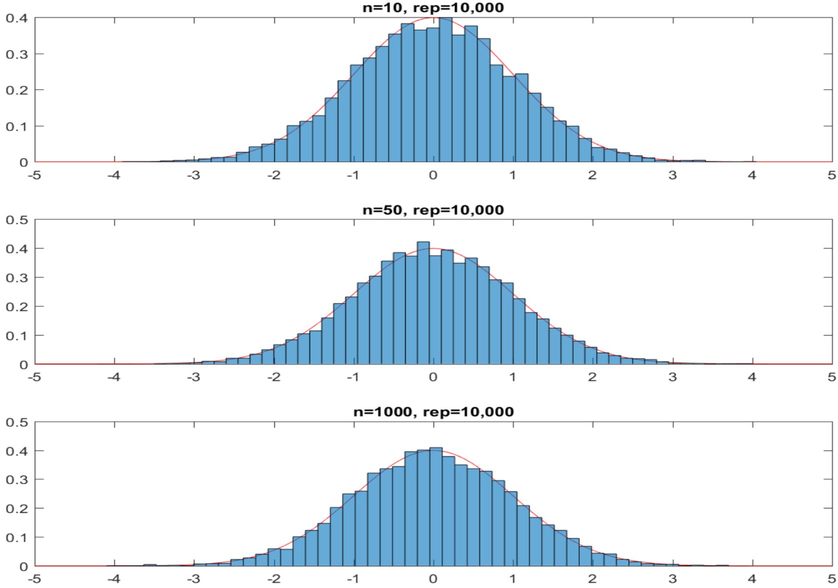


Figure 1: The histogram of $z_n^{(1)}, z_n^{(2)}, \ldots, z_n^{(J)}$ with $J = 10000$.

2.2 Convergence

To prove the above central limit theorem, we need the following lemma:

- **Lemma 2.1** If
  1. $M_{Z_n}(t)$, the moment generating function of Z_n, exists, $n = 1, 2, \ldots,$
  2. $\lim_{n\to\infty} M_{Z_n}(t)$ exists and equals the moment generating function of a random variable Z,

  then

  $$\lim_{n\to\infty} G_{Z_n}(x) = G_Z(x) \qquad \text{for all } x \text{ at which } G_Z(x) \text{ is continuous,}$$

  where $G_{Z_n}(x)$ is the c.d.f. of Z_n, $n = 1, 2, \ldots,$ and $G_Z(x)$ is the c.d.f. of Z.

(Proof of lemma 2.1 can be found in lecture notes page 17-19)

- *Example 2.10* Suppose that $Y \sim \chi^2(50)$. Approximate $\mathrm{P}(40 < Y < 60)$.
  Solution. By Example 2.8, $Y \sim \sum_{i=1}^{50} X_i^2$, where $X_1, X_2, \cdots, X_{50}$ are independent $\mathrm{N}(0,1)$ random variables. Let $\overline{Y} = \frac{1}{50}\sum_{i=1}^{50} X_i^2$. Hence,

$$
\begin{aligned}
\mathrm{P}(40 < Y < 60) &= \mathrm{P}(40 < 50\overline{Y} < 60) \\
&= \mathrm{P}(\frac{4}{5} < \overline{Y} < \frac{6}{5}) \\
&= \mathrm{P}\left(\frac{\sqrt{50}(\frac{4}{5}-\mu)}{\sigma} < \frac{\sqrt{50}(\overline{Y}-\mu)}{\sigma} < \frac{\sqrt{50}(\frac{6}{5}-\mu)}{\sigma}\right) \\
&\approx \Phi(\frac{\sqrt{50}(\frac{6}{5}-\mu)}{\sigma}) - \Phi(\frac{\sqrt{50}(\frac{4}{5}-\mu)}{\sigma}) \quad \text{(by CLT)} \\
&= \Phi(1) - \Phi(-1) \approx 0.68,
\end{aligned}
$$

where $\mu = \mathrm{E}X_i^2 = 1$, $\sigma^2 = \mathrm{Var}X_i^2 = 2$, and $\Phi(\cdot)$ is the c.d.f. of $\mathrm{N}(0,1)$.

2.3 Resampling

- **Realization of the empirical distribution:** Suppose $\{X_1, \cdots, X_n\}$ be a random sample from one population with an unknown c.d.f. $F(\cdot)$. Let $\{x_1, \cdots, x_n\}$ be one realization of $\{X_1, \cdots, X_n\}$. Based on $\{x_1, \cdots, x_n\}$, we have a realization of the empirical distribution:

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathrm{I}(x_k \leq x).$$

By Theorem 1.4,

$$F(x) \approx F_n(x). \tag{3.1}$$

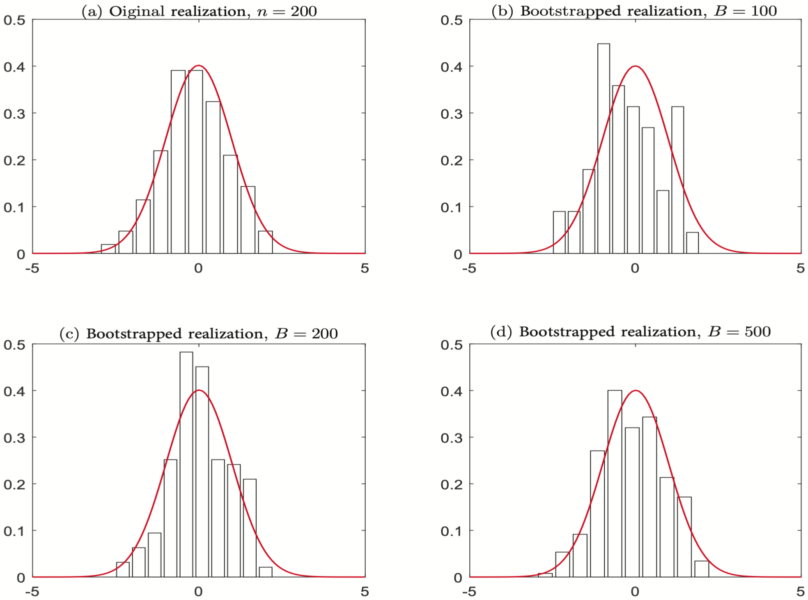
2.3 Resampling

- **Randomly draw samples:** Since $F_n(x)$ is a discrete c.d.f, we can draw a random sample $\{X_1^*, X_2^*, \cdots, X_B^*\}$ from $F_n(x)$, and it is expected that the (relative frequency) histogram of $\{X_1^*, X_2^*, \cdots, X_B^*\}$ should be close to $f(x)$. Here, $X_i^* \sim X^* \sim F_n(x)$ is a discrete random variable such that

$$\mathrm{P}(X^* = x_j) = \frac{1}{n} \text{ for } j = 1, 2, \cdots, n.$$

Remark: Conventionally, $\{X_1^*, X_2^*, \cdots, X_B^*\}$ is called the bootstrap (resampling) random sample, and B is the bootstrap sample size.

Example 2.11 Let $\{x_i\}_{i=1}^{200}$ be a realization from $N(0,1)$. Below Figure plots the histogram of original realization of $\{x_i\}_{i=1}^{200}$ and bootstrapped realizations $\{x_i^*\}_{i=1}^{100}$, $\{x_i^*\}_{i=1}^{200}$, $\{x_i^*\}_{i=1}^{500}$.



(a) Oiginal realization, $n = 200$
(b) Bootstrapped realization, $B = 100$
(c) Bootstrapped realization, $B = 200$
(d) Bootstrapped realization, $B = 500$
0.5
0.4
0.3
0.2
0.1
0
-5
0
5

2.3 Resampling - Bootstrap procedures

How to use bootstrap method to approximate the distribution of a statistic $T = g(X_1, X_2, \ldots, X_n)$, where $g(\cdot)$ is a given function.

(1) Generate a bootstrapped realization $\{x_1^*, x_2^*, \cdots, x_n^*\}$ from the distribution $F_n(\cdot)$;

(2) Calculate $t^* = g(x_1^*, x_2^*, \cdots, x_n^*)$, which is a realization of T^*;

(3) Repeat (1)-(2) J times to get $\{t^{*(1)}, t^{*(2)}, \cdots, t^{*(J)}\}$, which is a sequence of realizations of T^*;

(4) Plot the (relative frequency) histogram of $\{t^{*(1)}, t^{*(2)}, \cdots, t^{*(J)}\}$.

Since the histogram of $\{t^{*(1)}, t^{*(2)}, \cdots, t^{*(J)}\}$ is close to the p.d.f. of T^*, it is also close to the p.d.f. of T. Clearly, this bootstrap method provides us an easy way to calculate the percentile of the distribution of T, which is important in many applications.