Chapter 6 Limiting Theorem

Jingrui SUN

Email: sunjr@sustech.edu.cn

Southern University of Science and Technology

September 8, 2024

Outline I

Outline

Part 6.1.1: Markov Inequality

Markov Inequality

Suppose that a r.v.X has finite kth moment, i.e. $E(|X|^k) < \infty$ ($k > 0$, but k may not be a positive integer). Then for any $\varepsilon > 0$,

$$P\{|X| \geq \varepsilon\} \leq \frac{E(|X|^k)}{\varepsilon^k}. \tag{6.1.1}$$

Proof: If X is a continuous r.v. with p.d.f. $f(x)$, then

$$
\begin{aligned}
P\{|X| \geq \varepsilon\} &= \int_{|x|\geq\varepsilon} f(x)dx \\
&\leq \int_{|x|\geq\varepsilon} \frac{|x|^k}{\varepsilon^k} f(x)dx \quad \left(\because \text{ on } |x| \geq \varepsilon \text{ we have } 1 \leq \frac{|x|^k}{\varepsilon^k}\right) \\
&= \frac{1}{\varepsilon^k} \int_{|x|\geq\varepsilon} |x|^k f(x)dx \\
&\leq \frac{1}{\varepsilon^k} \int_{-\infty}^{+\infty} |x|^k f(x)dx \quad \left[\because |x|^k f(x) \geq 0\right] \\
&= \frac{1}{\varepsilon^k} \cdot E(|X|^k).
\end{aligned}
$$

If X is discrete, then the proof is similar!

Part 6.1.2: Chebyshev's Inequality

Chebyshev's Inequality

Suppose a r.v. X satisfies the condition that $E(X^2) < \infty$, then for any $a > 0$, we have

$$P\{|X| \geq a\} \leq \frac{E(X^2)}{a^2}. \tag{6.1.2}$$

Proof: In (6.1.1), letting $k = 2$ yields (and with $\varepsilon = a$)

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

which is just (6.1.2).

Corollary 1

By (6.1.2) we can prove the following conclusion:

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2} \tag{6.1.3}$$

Proof: Let $Y = X - E(X)$, then by (6.1.2) we have (letting $\varepsilon = a > 0$)

$$P\{|Y| \geq \varepsilon\} \leq \frac{E(Y^2)}{\varepsilon^2}$$

or

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{E\left[(X - E(X))^2\right]}{\varepsilon^2} = \frac{\text{Var}\,(X)}{\varepsilon^2}.$$

As another corollary of Chebyshev's inequality, we could prove the following conclusion which we have used before.

Corollary 2

A r.v. X is almost surely a constant if and only if

$$\mathrm{Var}(X) = 0. \tag{6.1.4}$$

Note: A r.v. X is called almost surely a constant if there exists a constant C, such that

$$P\{X = C\} = 1. \tag{6.1.5}$$

In brief, essentially a constant!!

Proof: If (6.1.5) holds, then

$$\operatorname{Var}(X) = E\left[(X - E(X))^2\right] = E\left[(C - C)^2\right] = 0$$

and thus (6.1.4) is true.

Now, assume (6.1.4) is true, we try to prove (6.1.5).

By Chebyshev's inequality, for any $\varepsilon > 0$, we have

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{\operatorname{Var}(X)}{\varepsilon^2}$$

But $\operatorname{Var}(X) = 0$ by (6.1.4), and thus for any $\varepsilon > 0$ we have

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{\operatorname{Var}(X)}{\varepsilon^2} = 0. \tag{6.1.6}$$

However, $P\{|X - E(X)| \geq \varepsilon\} \geq 0$ is always true, hence (6.1.6) implies

$$P\{|X - E(X)| \geq \varepsilon\} = 0 \tag{6.1.7}$$

is true for any $\varepsilon > 0$.
In particular, let $\varepsilon = \frac{1}{n}$, where n is a positive integer then (6.1.7) is just saying that for <u>any</u> positive integer $n \geq 1$, we have

$$P\left\{|X - E(X)| \geq \frac{1}{n}\right\} = 0 \quad \forall n \geq 1.$$

Therefore, we have

$$\lim_{n\to\infty} P\left\{|X - E(X)| \geq \frac{1}{n}\right\} = 0 \tag{6.1.8}$$

Consider the event

$$A_n = \left\{\omega : |X - E(X)| \geq \frac{1}{n}\right\} \tag{6.1.9}$$

Clearly, $A_n \uparrow$ [Indeed, when n is larger, then $\frac{1}{n}$ is smaller and thus if $n > m \geq 1$, then

$$\left\{\omega : |X - E(X)| \geq \frac{1}{n}\right\} \supset \left\{\omega : |X - E(X)| \geq \frac{1}{m}\right\}$$

and thus by the continuous property of probability measure we know that if $A_n \uparrow A$, then $P\{A_n\} \uparrow P\{A\}$, where $A = \cup_{n=1}^{\infty} A_n$.

But by (6.1.9) we clearly see that

$$A = \cup_{n=1}^{\infty} A_n = \{\omega : |X - E(X)| \neq 0\}$$

and therefore, (6.1.8) yields that

$$P(A) \equiv P\{|X - E(X)| \neq 0\} = \lim_{n \to \in} P(A_n) = 0.$$

Hence,

$$P\{X = E(X)\} = 1 - P\{|X - E(X)| \neq 0\} = 1 - 0 = 1,$$

i.e., X is almost surely a constant $E(X)$. The proof is complete.

Outline

Part 6.2: Weak Law of Large Numbers

Weak Law of Large Numbers

Suppose that $\{X_1, X_2, \cdots, X_n, \cdots\}$ is a sequence of i.i.d. r.v.s with a common finite mean value $E(X_i) = \mu < \infty$ and common finite variance $\operatorname{Var}(X_i) = \sigma^2 < \infty$. Let

$$\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \tag{6.2.1}$$

be the sample mean. Then for any $\varepsilon > 0$, we have

$$\lim_{n \to \infty} P\left\{|\bar{X}_n - \mu| > \varepsilon\right\} = 0. \tag{6.2.2}$$

Proof: Easy to see

$$E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot n\mu = \mu,$$

$$\begin{aligned}
\mathrm{Var}\left(\bar{X}_n\right) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(X_i\right) \quad \text{(Independence)} \\
&= \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

Now by Chebychev's Inequality, i.e.,

$$\forall \varepsilon > 0,\ P\{|X - E(X)| \geq \varepsilon\} \leq \frac{\operatorname{Var}(X)}{\varepsilon^2},$$

we obtain (by applying the above inequality to $\bar{X}_n$) for any $\varepsilon > 0$,

$$P\left\{|\bar{X}_n - E(\bar{X}_n)| \geq \varepsilon\right\} \leq \frac{\operatorname{Var}\left(\bar{X}_n\right)}{\varepsilon^2}$$

However, $E\left(\bar{X}_n\right) = \mu, \operatorname{Var}\left(\bar{X}_n\right) = \frac{\sigma^2}{n}$ and thus for any $\varepsilon > 0$, we have

$$P\left\{|\bar{X}_n - \mu| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2} \cdot \frac{\sigma^2}{n}.$$

Hence letting $n \to \infty$ yields

$$\lim_{n\to\infty} P\left\{|\bar{X}_n - \mu| \geq \varepsilon\right\} \leq \lim_{n\to\infty} \frac{\sigma^2}{\varepsilon^2} \cdot \frac{1}{n} = 0,$$

i.e., $\lim_{n\to\infty} P\left\{|\bar{X}_n - \mu| \geq \varepsilon\right\} \leq 0$.

But $\lim_{n\to\infty} P\left\{|\bar{X}_n - \mu| \geq \varepsilon\right\} \geq 0$ is obvious and thus

$$\lim_{n\to\infty} P\left\{|\bar{X}_n - \mu| \geq \varepsilon\right\} = 0, \quad \forall \varepsilon > 0.$$

Outline

Part 6.3.1: Setting

Suppose that $\{X_1, X_2, \cdots, X_n, \cdots\}$ is a sequence of i.i.d. r.v.s with common mean $E(X_i) = \mu$ and common finite variance $\sigma^2 < \infty$.

Note: " $\sigma^2 < +\infty$ " implies that " $|\mu| < +\infty$ ".

Let $S_n = \sum_{i=1}^n X_i$, then as shown before

$$E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = n\mu, \tag{6.3.1}$$

$$\mathrm{Var}\,(S_n) = \mathrm{Var}\,\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}\,(X_i) = n\sigma^2. \tag{6.3.2}$$

Now, let

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}. \tag{6.3.3}$$

Then

$$E(Z_n) = \frac{1}{\sigma\sqrt{n}}\left[E(S_n) - n\mu\right] = \frac{n\mu - n\mu}{\sigma\sqrt{n}} = 0.$$

$$\begin{aligned}
\operatorname{Var}(Z_n) &= \operatorname{Var}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) \\
&= \frac{\operatorname{Var}(S_n - n\mu)}{\sigma^2 n} \\
&= \frac{\operatorname{Var}(S_n)}{\sigma^2 n} \\
&= 1.
\end{aligned}$$

In short: Let $\{X_1, X_2, \cdots, X_n, \cdots\}$ be i.i.d. with common mean $E(X_i) = \mu$ and common variance $\operatorname{Var}(X_i) = \sigma^2 < +\infty$, and let

$$S_n = \sum_{i=1}^{n} X_i, \quad Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}. \tag{6.3.4}$$

Then

$$E(Z_n) = 0, \tag{6.3.5}$$

$$\operatorname{Var}(Z_n) = 1. \tag{6.3.6}$$

But the distribution of Z_n might be quite complicated.

Part 6.3.2: Conclusion (Central Limit Theorem)

Theorem (Central Limit Theorem): Suppose that $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. r.v.s with common mean μ and common variance $\sigma^2 < \infty$. Let

$$S_n = \sum_{i=1}^{n} X_i, \quad Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Then Z_n converges to $N(0,1)$-r.v. in distribution as $n \to \infty$:

$$\lim_{n\to\infty} P\{Z_n \leq x\} = \Phi(x), \quad \forall x \in \mathbb{R}.$$

where $\Phi(x)$ is the c.d.f. of standard normal distribution, i.e.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy.$$

Part 6.3.3: Proof of the CLT

To prove the central limit theorem, we need the following continuity theorem.

Lemma (Continuity Theorem):

Let F_n be a sequence of c.d.f.s with the corresponding m.g.f. M_n.

Let F be a c.d.f. with corresponding m.g.f. M.

If $M_n(t) \to M(t)$ for all t in an open interval containing zero, then $F_n(x) \to F(x)$ at all continuity points x of F.

Proof of The Central Limit Theorem: Let M_i be the m.g.f. of

$$Y_i \triangleq \frac{X_i - \mu}{\sigma}, \quad i = 1, 2, \ldots$$

Then $Y_1, Y_2, \ldots$ are i.i.d. with

$$E(Y_i) = 0, \quad \mathrm{Var}(Y_i) = 1, \quad i = 1, 2, \ldots$$

Since the $(Y_i)_{i \geq 1}$ are identically distributed, M_i does not depend on i and we write M.

$$\begin{aligned} M_{Z_n}(t) &= M_{\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i}(t) = M_{\sum_{i=1}^n Y_i}\left(\frac{t}{\sqrt{n}}\right) \\ &= \prod_{j=1}^n M_i\left(\frac{t}{\sqrt{n}}\right) = \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n. \end{aligned}$$

Since $\mathrm{Var}(Y_i) = 1$, we see that M has continuous derivatives up to order two, and

$$M(0) = 1, \quad M'(0) = E(Y_i) = 0 \quad M''(0) = E(Y_i^2) = 1, \quad \forall i \geq 1.$$

Expanding M in a Taylor expansion about $t = 0$ gives

$$M(t) = 1 + 0 + \frac{t^2}{2} + t^2 h(t),$$

where $h(t) \to 0$ as $t \to 0$. Then

$$M\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + \frac{t^2}{n} h\left(\frac{t}{\sqrt{n}}\right).$$

Thus for every fixed t,

$$\lim_{n\to\infty} M\left(\frac{t}{\sqrt{n}}\right) = 1.$$

It then follows by L'Hôpital's rule that

$$
\begin{aligned}
M_{Z_n}(t) &= \lim_{n\to\infty} \left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n \\
&= \lim_{n\to\infty} e^{n\log M\left(\frac{t}{\sqrt{n}}\right)} \\
&= \lim_{n\to\infty} e^{n\log\left(1+\frac{t^2}{2n}+\frac{t^2}{n}h\left(\frac{t}{\sqrt{n}}\right)\right)} \\
&= e^{t^2/2}.
\end{aligned}
$$

Thus, by the continuity theorem we see that Z_n converges in distribution to some random variable Z with m.g.f. $M_Z(t) = e^{t^2/2}$.

By the uniqueness theorem, $Z \sim N(0,1)$.

Part 6.3.4: Applications in Statistics

In statistical problems, it is usually assume that we have a large sample of r.v.s $\{X_1, X_2, \cdots, X_n\}$ which are i.i.d r.v.s. Let

$$Y = X_1 + X_2 + \cdots + X_n. \tag{6.3.21}$$

It is usually very easy to find the mean and variance of Y (by using the mean and variance of X_i, for example), but our interests are trying to find the probabilities such as

$$P\{a < Y \leq b\} \quad \text{for constants } a < b. \tag{6.3.22}$$

If we know the c.d.f. of Y, denoted by $F_Y(y)$, then, of course

$$P\{a < Y \leq b\} = F_Y(b) - F_Y(a). \tag{6.3.23}$$

However, to find the c.d.f. of Y is usually a very hard work.

Another method to calculate the value in (6.3.22) is just to use the Central Limit Theorem.

This method is called the normal approximation method.

When the n in (6.3.21) is large, the usually procedure is as follows:

Assume the random sum Y in (6.3.21) satisfies $E(X_i) = \mu$ and $\mathrm{Var}\,(X_i) = \sigma^2 < \infty$ for all i. Then

$$
\begin{aligned}
P\{a < Y \leq b\} &= P\left\{\frac{a - n\mu}{\sigma\sqrt{n}} < \frac{Y - n\mu}{\sigma\sqrt{n}} \leq \frac{b - n\mu}{\sigma\sqrt{n}}\right\} \\
&\approx P\left\{\frac{a - n\mu}{\sigma\sqrt{n}} < Z \leq \frac{b - n\mu}{\sigma\sqrt{n}}\right\},
\end{aligned}
$$

where $Z \sim \mathrm{N}(0,1)$ (This is just the CLT), i.e.,

$$P\{a < Y \leq b\} \approx \Phi\left(\frac{b - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a - n\mu}{\sigma\sqrt{n}}\right). \tag{6.3.24}$$

When the n in (6.3.21) is large, then the value obtained in (6.3.24) is quite good!

For <u>how large of n</u>, will the normal approximation method in (6.3.24) be good? This will depend upon the p.d.f. of Y is symmetric or not.

If the p.d.f. of Y is symmetric, (such as t-distribution), then $n \geq 30$ is enough.

If the p.d.f. of Y is not symmetric, then we need more large n.

Example 1: 一复杂的系统由 100 个相互独立起作用的部件所组成, 在整个运行期间每个部件损坏的概率为 0.10. 为了使整个系统起作用, 至少必须有 85 个部件正常工作, 求整个系统起作用的概率.

Example 2: Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with a common normal distribution $N(20, 25)$. Let

$$Y = \sum_{i=1}^{6} X_i, \quad W = \sum_{i=1}^{30} X_i, \quad \bar{X}_{30} = \frac{1}{30} W.$$

(i) What distributions do Y, W and $\bar{X}_{30}$ obey?

(ii) Find the probability that the random variable $\bar{X}_{30}$ is between 19 and 21, i.e., find $P(19 < \bar{X}_{30} < 21)$. Also, find the probability that W is greater than 650, i.e., find $P(W > 650)$.

(iii) Are Y and W uncorrelated or correlated? If you think they are uncorrelated, then provide a proof. If you think they are correlated, then state whether they are positively correlated or negatively correlated. Specify your reasons clearly! Also find the correlation coefficient of Y and W, i.e., $\rho(Y, W)$ to support your conclusions.

(iv) Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Find the moment generating function of $\bar{X}_n$, denoted by $M_{\bar{X}_n}(t)$. Let $\varphi(t) = \lim_{n\to\infty} M_{\bar{X}_n}(t)$. Is $\varphi(t)$ the moment generating function of some normal random variable? If your answer is positive, specify this normal distribution. If your answer is negative, specify some reasons to support your conclusion, and state the random variable (a constant is also viewed as a random variable) whose moment generating function is $\varphi(t)$. Let $Y_n = \frac{\bar{X}_n - 20}{5/\sqrt{n}}$. Find the moment generating function of Y_n, denoted by $M_{Y_n}(t)$. Let $\psi(t) = \lim_{n\to\infty} M_{Y_n}(t)$. Answer the same questions for $\psi(t)$.