Teaching Notes on Ch05 (Limiting Theorems)

I. Some Inequality

1. Markov Inequality

① Suppose a random variable X has finite k'th moment, i.e.

$$E[|X|^k] < \infty \quad (k>0,\ k \text{ may not be a positive integer.})$$

Then for any $\varepsilon > 0$

$$P\{|X| \geq \varepsilon\} \leq \frac{E[|X|^k]}{\varepsilon^k} \qquad (1.1)$$

② Proof. Assume that X is a continuous r.v. with pdf $f(x)$ and cdf $F(x)$ 「If X is discrete, then the proof is similar!」 then

$$P\{|X| \geq \varepsilon\} = \int_{|x| \geq \varepsilon} f(x)dx \leq \int_{|x|\geq\varepsilon} \frac{|x|^k}{\varepsilon^k} f(x)dx \quad \left(\because \text{on } |x| \geq \varepsilon \text{ we have } 1 \leq \frac{|x|^k}{\varepsilon^k}\right)$$

$$\leq \frac{1}{\varepsilon^k}\int_{|x|\geq\varepsilon} |x|^k f(x)dx \leq \frac{1}{\varepsilon^k}\int_{-\infty}^{+\infty} |x|^k f(x)dx \quad \left[\because |x|^k f(x) \geq 0\right]$$

$$= \frac{1}{\varepsilon^k}\cdot E[|X|^k]$$

(1.1) is proven.

2. Chebyshev's Inequality

① Conclusion: Suppose a random variable X satisfies the condition that $E(X^2) < \infty$, then for any $a > 0$, we have

$$P\{|X| \geq a\} \leq \frac{E[X^2]}{a^2} \qquad (1.2).$$

② Proof. ~~Inequalit~~ In (1.1), letting $k = 2$ yields (and with $\varepsilon = a$)

$$P\{|X| \geq a\} \leq \frac{E[X^2]}{a^2}$$

which is (1.2).

③ (1.2) can also be written as

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2} \qquad (1.3)$$

Let $Y = X - E(X)$, then by (1.2) we have (letting $\varepsilon = a > 0$)

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

or $$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{E[|X - E(X)|^2]}{\varepsilon^2} = \frac{E[(X - E(X))^2]}{\varepsilon^2}$$

$$= \frac{Var(X)}{\varepsilon^2}$$

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

④ Corollary. A random variable X is almost surely a constant if and only if

$$\mathrm{Var}(X) = 0 \tag{1.4}$$

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

$$P\{X = C\} = 1 \tag{1.5}$$

「In brief, essentially a constant !!」

If (1.5) holds, then $\mathrm{Var}(X) = E[(X - E(X))^2] = E((C-C)^2) = 0$ and thus (1.4) is true.

Now, assume (1.4) is true, we try to prove (1.5).

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

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

But $\mathrm{Var}(X) = 0$ (by (1.4)) and thus for any $\varepsilon > 0$ we have

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{\mathrm{Var}(X)}{\varepsilon^2} = 0. \tag{1.6}$$

However, $P\{|X - E(X)| \geq \varepsilon\} \geq 0$ is always true, hence (1.6) implies

$$P\{|X - E(X)| \geq \varepsilon\} = 0 \quad \text{is true for any } \varepsilon > 0. \tag{1.7}$$

In particular, let $\varepsilon = \frac{1}{n}$, where n is a positive integer. Then (1.7) reads

for any positive integer $n \geq 1$, we have

$$P\{|X - E(X)| \geq \frac{1}{n}\} = 0 \qquad (\forall n \geq 1)$$

Therefore, we, of course, have

$$\lim_{n \to \infty} P\{|X - E(X)| \geq \frac{1}{n}\} = 0 \qquad (1.8)$$

Consider the event

$$A_n = \{w;\ |X - E(X)| \geq \frac{1}{n}\} \qquad (1.9)$$

Clearly, $A_n \uparrow$ [Indeed, when n is larger then $\frac{1}{n}$ is smaller and thus if $n > m \geq 1$, then $\{w;\ |X - E(X)| \geq \frac{1}{n}\} \supset \{w;\ |X - E(X)| \geq \frac{1}{m}\}$]

and thus by the property of probability that

if $A_n \uparrow A$ then $P\{A_n\} \uparrow P\{A\}$

But by (1.9) we clearly see that

$$A = \lim_{n\to\infty} A_n = \{w;\ |X - E(X)| \neq 0\}$$

and therefore, (1.8) yields that

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

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.

II. Weak Law of Large Numbers

3. Conclusions: Suppose $\{X_1, X_2, \cdots, X_n, \cdots\}$ is a sequence of independent identically distributed random variables with a common finite mean value $E(X_i) = \mu < \infty$ and common finite variance $Var(X_i) = \sigma^2 < \infty$ $(\forall i)$.

Let
$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \qquad (2.1)$$

be the sample mean.

Then for any $\varepsilon > 0$, we have

$$\lim_{n\to\infty} P\{|\overline{X}_n - \mu| > \varepsilon\} = 0 \qquad (2.2)$$

4. Proof. Easy to see

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

$$Var(\overline{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \qquad \text{(Independence)}$$

$$= \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Now by the Chebyshev's Inequality (proved in (1.3))

$$\left(\text{i.e. } \forall \varepsilon > 0, \quad P\{|X - E(X)| \geq \varepsilon\} \leq \frac{Var(X)}{\varepsilon^2} \right)$$

we obtain (by applying the above inequality to $\bar{X}_n$)

for any $\varepsilon > 0$,

$$P\{|\bar{X}_n - E(\bar{X}_n)| \geq \varepsilon\} \leq \frac{Var(\bar{X}_n)}{\varepsilon^2} \tag{2.3}$$

However, $E(\bar{X}_n) = \mu$, $Var(\bar{X}_n) = \dfrac{\sigma^2}{n}$

and thus for any $\varepsilon > 0$, we have

$$P\{|\bar{X}_n - \mu| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} \cdot \frac{\sigma^2}{n} \tag{2.4}$$

Hence letting $n \to \infty$ in (2.4) yields

$$\lim_{n\to\infty} P\{|\bar{X}_n - \mu| \geq \varepsilon\} \leq \lim_{n\to\infty} \frac{\sigma^2}{\varepsilon^2} \cdot \frac{1}{n} = 0 \quad (\because \sigma^2 < \infty)$$

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

Hence

$$\lim_{n\to\infty} P\{|\bar{X}_n - \mu| \geq \varepsilon\} = 0 \quad (\forall \varepsilon > 0)$$

III. The Central Limit Theorem

1. Setting: Suppose $\{X_1, X_2, \cdots, X_n, \cdots\}$ is a sequence of i.i.d r.vs (with common distribution) with common mean $E(X_i) = \mu$ $(\forall i)$ and common finite variance $Var(X_i) = \sigma^2 < +\infty$ $(\forall i)$

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

Let
$$S_n = \sum_{i=1}^{n} X_i,$$

then as shown before

$$E(S_n) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = n\mu \quad (3.1)$$

$$Var(S_n) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = n\sigma^2 \quad (3.2)$$

Now, let

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad (3.3)$$

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

$$Var(Z_n) = Var\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) = \frac{Var(S_n - n\mu)}{\sigma^2 n} = \frac{Var(S_n)}{\sigma^2 n}$$ ($\because$ $n\mu$ is a constant)

i.e. $Var(Z_n) = \frac{Var(S_n)}{\sigma^2 n}$.

But $Var(S_n) = n\sigma^2$ (see (3.2))

which implies that $Var(Z_n) = \frac{Var(S_n)}{\sigma^2 n} = \frac{n\sigma^2}{\sigma^2 n} = 1$

In short:

If $\{X_1, X_2 \cdots X_n \cdots\}$ are iid with common mean $E(X_i) = \mu < +\infty$ and common variance $Var(X_i) = \sigma^2 < +\infty$ $(\forall i)$. Then

Let $S_n = \sum_{i=1}^{n} X_i$

and $$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad (3.4)$$

Then
$$\boxed{\begin{aligned} E(Z_n) &= 0 \quad (3.5) \\ Var(Z_n) &= 1 \quad (3.6) \end{aligned}}$$

But the distribution of Z_n might be quite complicated.

2. Conclusion (Central Limit Theorem)

If $\{X_1, X_2, \cdots X_n, \cdots\}$ are independent, identically distributed radom variables with common mean $E(X_i) = \mu < +\infty$ $(\forall i)$ and common variance $Var(X_i) = \sigma^2 < \infty$ $(\forall i)$, ~~then~~

Let $S_n = \sum_{n=1}^{\infty} X_i$

and $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$

Then for any $x \in R$ $(-\infty < x < +\infty)$ we have

$$\lim_{n\to\infty} P\{Z_n \le x\} = \Phi(x) \qquad (3.7)$$

「

i.e. $$\lim_{n\to\infty} P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right\} = \Phi(x) \qquad (3.8)$$

」

where $\Phi(x)$ is the cdf of standard normal distribution.

i.e. $$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \qquad (3.9)$$

3. Idea of the proof:

Since $\{X_1, X_2, \cdots X_n, \cdots\}$ are i.i.d with common mean $E(X_i) = \mu < \infty$ and common variance $Var(X_i) = \sigma^2 < \infty$, and thus the random variables $\{Y_1, Y_2, \cdots Y_n \cdots\}$ are also i.i.d.
where

$$Y_i = \frac{X_i - \mu}{\sigma} \quad (\forall i) \qquad (3.10)$$

Also, since $E(X_i) = \mu < +\infty$ and $Var(X_i) = \sigma^2 < \infty$
we could easily get

$$E(Y_i) = 0 \quad (\forall i \geq 1) \qquad (3.11)$$

and

$$Var(Y_i) = 1 \quad (\forall i \geq 1) \qquad (3.12)$$

It follows that all the $\{Y_i ; i \geq 1\}$ have the same MGF ($\because$ Y_i are i.i.d), denoted by $M_Y(t)$, say. We expand the $M_Y(t)$ as

$$M_Y(t) = M_Y(0) + \frac{M_Y'(0)}{1!} t + \frac{M_Y''(0)}{2!} t^2 + o(t^2) \qquad (3.13)$$

where in (3.13), $o(t^2)$ means

$$\lim_{t \to 0} \frac{o(t^2)}{t^2} = 0 \qquad (3.14)$$

「Note that $Var(Y_i) < +\infty$ guarantees that the form (3.13) holds at least for a small interval $0 \leq t < \varepsilon$.」

But we know $E(Y_i) = 0$ and $Var(Y_i) = 1$ $(\forall i \geq 1)$ (See (3.11) and (3.12))

and since (i) $M_Y(0) = 1$ (This is true for any MGF)

(ii) $M_Y'(0) = E(Y_i) = 0 \quad (\forall i)$

(iii) $M_Y''(0) = E(Y_i^2) = Var(Y_i) + (E(Y_i))^2$

$= 1 + 0^2 = 1$

and hence the form (3.13) takes the simple form as

$$M_Y(t) = 1 + \frac{1}{2}t^2 + o(t^2) \qquad (3.15)$$

where

$$\lim_{t \to 0} \frac{o(t^2)}{t^2} = 0 \qquad (3.16) = (3.14)$$

By (3.16) (i.e. $o(t^2)$), we may assume, without loss of generality, that

$$o(t^2) = t^3 \cdot R(t)$$

where $R(t)$ is a continuous, bounded function of t near $t=0$.

Therefore, the MGF of $\{Y_i\}$ $(\forall i \geq 1)$, $M_Y(t)$ can be written as

$$M_Y(t) = 1 + \frac{1}{2}t^2 + t^3 R(t) \qquad (3.17)$$

Now by (3.4) we know that

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

$$= \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i - \mu}{\sigma}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i$$

Since $\{Y_1, Y_2, \cdots Y_n\}$ are independent, the MGF of Z_n, denoted by $M_{Z_n}(t)$, is

$$M_{Z_n}(t) = \left(M_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n \quad \text{(Think why here !!!)}$$

But $M_Y(t) = 1 + \frac{1}{2}t^2 + t^3 R(t)$

and hence

$$M_{Z_n}(t) = \left(1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + \left(\frac{t}{\sqrt{n}}\right)^3 \cdot R\left(\frac{t}{\sqrt{n}}\right)\right)^n \tag{3.18}$$

Consider the term $\left(\frac{t}{\sqrt{n}}\right)^3 \cdot R\left(\frac{t}{\sqrt{n}}\right)$ in (3.18).

For any fixed t, $\lim_{n\to\infty} t^3 \cdot R\left(\frac{t}{\sqrt{n}}\right) = t^3 \cdot R(0)$ which is bounded. But $\lim_{n\to\infty}\left(\frac{1}{\sqrt{n}}\right)^3 = 0$

In short, when considering the large n, for any fixed t, this term is very "small" comparing with $\frac{t^2}{n}$ for both n and t.

Hence, by (3.18), we actually have

$$M_{Z_n}(t) \approx \left(1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2\right)^n$$

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

We could write, by ignoring the "small error", that for large n

$$M_{Z_n}(t) \doteq \left(1 + \frac{t^2}{2n}\right)^n \qquad (3.19)$$

But for any $x \in \mathbb{R}$, we know $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$

and hence

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left(1 + \frac{\left(\frac{t^2}{2}\right)}{n}\right)^n = e^{\frac{t^2}{2}} \qquad (3.20).$$

(Just $x = \frac{t^2}{2}$!!)

In short,

$$\lim_{n\to\infty} M_{Z_n}(t) = e^{\frac{t^2}{2}}$$

But $e^{\frac{t^2}{2}}$ is the MGF of $N(0,1)$. Hence the MGF of Z_n tends to the MGF of $N(0,1)$ when $n \to \infty$.

That is that the cdf of Z_n tends to the cdf of standard normal distribution (Here, the strict mathematical details are necessary, but we shall ignore the details here!)

That is that $\forall x \in R$, (by denoting the cdf of $N(0,1)$ as $\Phi(x)$)

$$\lim_{n\to\infty} P\{Z_n \le x\} = \Phi(x) \quad \text{or}$$

$$\lim_{n\to\infty} P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right\} = \Phi(x).$$

※

4. Applications in Statistics.

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

Let $\quad X = X_1 + X_2 + \cdots + X_n \qquad (3.21)$

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

$$P\{a < X \le b\} \quad \text{for constants } a < b. \qquad (3.22)$$

If we know the cdf of X, denoted by $F_X(x)$ then, of course

$$P\{a < X \le b\} = F_X(b) - F_X(a) \qquad (3.23)$$

However to find the cdf of X is a usually very hard work. Another method to calculate the value in (3.22) is just to use the Central Limit Theorem. This method is called the normal approximation method.

When the n in (3.21) is large, the usually procedure is as follows.

Assume the random sum ξ in (3.21) satisfies $E(\xi) = \mu < \infty$ and $Var(\xi) = \sigma^2 < \infty$. then

$$P\{a < \xi \le b\} = P\left\{\frac{a-\mu}{\sigma} < \frac{\xi-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right\}$$

$$\approx P\left\{\frac{a-\mu}{\sigma} < Z \le \frac{b-\mu}{\sigma}\right\}$$

(where $Z \sim N(0,1)$ This is just the CLT)

i.e.

$$P\{a < \xi \le b\} \approx \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \quad (3.24)$$

When the n in (3.21) is large, then the value obtained in (3.24) is quite good!

For how large of n, will the normal approximation method in (3.24) be good? This will depends upon the ~~cdf~~ pdf of ξ is symmetric or not. If the pdf of ξ is symmetric, (such as t-distribution), then $n \gtrsim 30$ is enough. If the pdf of ξ is not symmetric, then we need more large n.