

Section 27

The Central Limit Theorem

Po-Ning Chen, Professor

Institute of Communications Engineering

National Chiao Tung University

Hsin Chu, Taiwan 30010, R.O.C.

Identically distributed summands

27-1

- **Central limit theorem:** The sum of many independent random variables will be approximately *normally distributed*, if each summand has high probability of being small.

Theorem 27.1 (Lindeberg-Lévy theorem) Suppose that $\{X_n\}_{n=1}^{\infty}$ is an independent sequence of random variables having the same distribution with mean c and finite positive variance σ^2 . Then

$$\frac{S_n - nc}{\sigma\sqrt{n}} \Rightarrow N,$$

where N is standard normal distributed, and $S_n = X_1 + \cdots + X_n$.

Idea behind the proof.

27-2

Proof:

- Without loss of generality, we can assume that X_n has zero-mean and unit variance.
- The characteristic function of S_n/\sqrt{n} is then given by:

$$\begin{aligned} E \left[e^{itS_n/\sqrt{n}} \right] &= E \left[e^{it(X_1+\dots+X_n)/\sqrt{n}} \right] \\ &= E^n \left[e^{itX/\sqrt{n}} \right] \\ &= \varphi_X^n \left(\frac{t}{\sqrt{n}} \right). \end{aligned}$$

- By the continuity theorem,

Theorem 26.3 (Continuity theorem)

$X_n \Rightarrow X$ if, and only if $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ for every $t \in \mathfrak{R}$.

we need to show that

$$\lim_{n \rightarrow \infty} \varphi_X^n \left(\frac{t}{\sqrt{n}} \right) = e^{-t^2/2},$$

or equivalently,

$$\lim_{n \rightarrow \infty} n \log \left[\varphi_X \left(\frac{t}{\sqrt{n}} \right) \right] = -\frac{t^2}{2}.$$

Idea behind the proof.

27-3

Lemma If $E[|X^k|] < \infty$, then

$$\varphi^{(k)}(0) = i^k E[X^k].$$

- By noting that $\varphi_X(0) = 1$, $\varphi'_X(0) = 0$ and $\varphi''_X(0) = -1$,

$$\begin{aligned} \lim_{n \rightarrow \infty} n \log \varphi_X(t/\sqrt{n}) &= \lim_{s \rightarrow 0} \frac{\log \varphi_X(ts)}{s^2} \quad (s = 1/\sqrt{n}) \\ &= \lim_{s \rightarrow 0} \frac{t \frac{\varphi'_X(ts)}{\varphi_X(ts)}}{2s} \quad (\text{L'Hospital's Rule}) \\ &= \lim_{s \rightarrow 0} \frac{t^2 \left(\frac{\varphi''_X(ts) \varphi_X(ts) - [\varphi'(ts)]^2}{\varphi_X^2(ts)} \right)}{2} \\ &= -\frac{t^2}{2}. \end{aligned}$$

□

Generalization of Theorem 27.1

27-4

- Theorem 27.1 requires the distribution of each X_n being **identical**.
- Can we relax this requirement for the central limit theorem to hold?
Yes, but different proof technique must be developed.

Illustration of idea behind alternative proof

27-5

Lemma If X has a moment of order n , then

$$\begin{aligned} \left| \varphi(t) - \sum_{k=0}^n \frac{(it)^k}{k!} E[X^k] \right| &\leq E \left[\min \left\{ \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right\} \right] \\ &\leq \min \left\{ \frac{|t|^{n+1} E[|X|^{n+1}]}{(n+1)!}, \frac{2|t|^n E[|X|^n]}{n!} \right\}. \end{aligned}$$

(Notably, this inequality is valid even if $E[|X|^{n+1}] = \infty$.)

Suppose that $E[|X|^3] < \infty$. (The assumption is simply made for showing the idea behind an alternative proof).

With $E[X] = 0$ and $E[X^2] = 1$,

$$\left| \varphi_X \left(\frac{t}{\sqrt{n}} \right) - \left(1 - \frac{t^2}{2n} \right) \right| \leq E \left[\min \left\{ \frac{|tX|^3}{6n^{3/2}}, \frac{t^2}{n} X^2 \right\} \right] \leq \frac{|t|^3}{n^{3/2}} \frac{E[|X|^3]}{6}$$

Illustration of idea behind alternative proof

27-6

Lemma 1 Let z_1, \dots, z_m and w_1, \dots, w_m be complex numbers of modulus at most 1; then

$$|z_1 \times z_2 \times \cdots \times z_m - w_1 \times w_2 \times \cdots \times w_m| \leq \sum_{k=1}^m |z_k - w_k|.$$

Proof: As

$$\begin{aligned} & z_1 \times z_2 \times \cdots \times z_m - w_1 \times w_2 \times \cdots \times w_m \\ &= z_1 \times z_2 \times \cdots \times z_m - w_1 \times z_2 \times \cdots \times z_m + w_1 \times z_2 \times \cdots \times z_m - w_1 \times w_2 \times \cdots \times w_m \\ &= (z_1 - w_1)(z_2 \times \cdots \times z_m) + w_1(z_2 \times \cdots \times z_m - w_2 \times \cdots \times w_m), \end{aligned}$$

we obtain:

$$\begin{aligned} & |z_1 \times z_2 \times \cdots \times z_m - w_1 \times w_2 \times \cdots \times w_m| \\ &= |(z_1 - w_1)(z_2 \times \cdots \times z_m) + w_1(z_2 \times \cdots \times z_m - w_2 \times \cdots \times w_m)| \\ &\leq |(z_1 - w_1)(z_2 \times \cdots \times z_m)| + |w_1(z_2 \times \cdots \times z_m - w_2 \times \cdots \times w_m)| \\ &\leq |z_1 - w_1| + |z_2 \times \cdots \times z_m - w_2 \times \cdots \times w_m|. \end{aligned}$$

The lemma therefore can be proved by induction. ■

Illustration of idea behind alternative proof

27-7

With the lemma and considering those n satisfying $n \geq t^2/4$,

$$\begin{aligned} \left| \varphi_X^n \left(\frac{t}{\sqrt{n}} \right) - \left(1 - \frac{t^2}{2n} \right)^n \right| &\leq n \left| \varphi_X \left(\frac{t}{\sqrt{n}} \right) - \left(1 - \frac{t^2}{2n} \right) \right| \\ &\leq n \times \frac{|t|^3 E[|X|^3]}{n^{3/2} \cdot 6} \\ &= \frac{|t|^3 E[|X|^3]}{\sqrt{n} \cdot 6} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

The central limit statement can then be validated by:

$$\left(1 + \frac{-t^2/2}{n} \right)^n \rightarrow e^{-t^2/2}.$$

Exemplified application of central limit theorem

27-8

Example 27.2 Suppose one wants to estimate the parameter α of an exponential distribution on the basis of independent samples X_1, \dots, X_n .

By law of large numbers,

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n) \text{ converges to mean } \frac{1}{\alpha} \text{ with probability 1.}$$

As the variance of exponential distribution is $1/\alpha^2$, Lindeberg-Lévy theorem gives that

$$\frac{X_1 + X_2 + \dots + X_n - n/\alpha}{\sqrt{n}/\alpha} = \alpha\sqrt{n} (\bar{X}_n - 1/\alpha) \Rightarrow N.$$

Equivalently,

$$\lim_{n \rightarrow \infty} \Pr \left[\alpha\sqrt{n} \left(\bar{X}_n - \frac{1}{\alpha} \right) \leq x \right] = \Phi(x),$$

where $\Phi(\cdot)$ represents the standard normal cdf.

Roughly speaking, \bar{X}_n is approximately Gaussian distributed with mean $1/\alpha$ and variance $1/(\alpha^2 n)$. Notably, this statement exactly indicates that

$$\lim_{n \rightarrow \infty} \Pr \left[\frac{\bar{X}_n - (1/\alpha)}{1/(\alpha\sqrt{n})} \leq x \right] = \Phi(x)$$

Exemplified application of central limit theorem

27-9

Theorem 25.6 (Skorohod's theorem) Suppose μ_n and μ are probability measures on $(\mathfrak{R}, \mathcal{B})$, and $\mu_n \Rightarrow \mu$. Then there exist random variables Y_n and Y such that:

1. they are both defined on common probability space (Ω, \mathcal{F}, P) ;
2. $\Pr[Y_n \leq y] = \mu_n(-\infty, y]$ for every y ;
3. $\Pr[Y \leq y] = \mu(-\infty, y]$ for every y ;
4. $\lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega)$ for every $\omega \in \Omega$.

By Skorohod's Theorem, there exist $\bar{Y}_n : \Omega \rightarrow \mathfrak{R}$ and $Y : \Omega \rightarrow \mathfrak{R}$ such that

$$\lim_{n \rightarrow \infty} \alpha \sqrt{n} (\bar{Y}_n(\omega) - 1/\alpha) = Y(\omega) \text{ for every } \omega \in \Omega,$$

and \bar{Y}_n and Y have the same distributions as \bar{X}_n and N , respectively.

As $P(\{\omega \in \Omega : \lim_{n \rightarrow \infty} \bar{Y}_n(\omega) = 1/\alpha\}) = 1$,

$$\frac{\sqrt{n}}{\alpha} \left(\frac{1}{\bar{Y}_n(\omega)} - \alpha \right) = \frac{-\alpha \sqrt{n} (\bar{Y}_n(\omega) - \alpha^{-1})}{\alpha \bar{Y}_n(\omega)} \xrightarrow{n \rightarrow \infty} \frac{-Y(\omega)}{\alpha \cdot (1/\alpha)} = -Y(\omega),$$

where $-Y$ is also standard normal distributed.

Exemplified application of central limit theorem

27-10

This concludes to

$$\frac{\sqrt{n}}{\alpha} \left(\frac{1}{\bar{X}_n} - \alpha \right) \Rightarrow N.$$

In words, $1/\bar{X}_n$ is approximately Gaussian distributed with mean α and variance α^2/n .

Lindeberg Theorem

27-11

Definition A triangular array of random variables is

$$\begin{array}{ccccccc} X_{1,1} & \cdots & X_{1,r_1} & & & & \\ X_{2,1} & X_{2,2} & \cdots & X_{2,r_2-1} & X_{2,r_2} & & \\ X_{3,1} & X_{3,2} & X_{3,3} & \cdots & X_{3,r_3-2} & X_{3,r_3-1} & X_{3,r_3} \\ & & & \vdots & & & \end{array}$$

where the probability space, on which each sequence $X_{n,1}, \dots, X_{n,r_n}$ is commonly defined, may change (since we do not care about the dependence across sequences.)

- A sequence of random variables is just a special case of a triangular array of random variables with $r_n = n$ and $X_{n,k} = X_k$.

Lindeberg's condition For an array of independent zero-mean random variables $X_{n,1}, \dots, X_{n,r_n}$,

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|x| \geq \varepsilon s_n} x^2 dF_{X_{n,k}}(x) = \lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^2} E \left[X_{n,k}^2 I_{\{|X_{n,k}| \geq \varepsilon s_n\}} \right] = 0,$$

where $s_n^2 = \sum_{k=1}^{r_n} E \left[X_{n,k}^2 \right]$.

Lindeberg Theorem

27-12

Theorem 27.2 (Lindeberg theorem) For an array of independent zero-mean random variables $X_{n,1}, \dots, X_{n,r_n}$, if Lindeberg's condition holds for all positive ε , then

$$\frac{S_n}{s_n} \Rightarrow N,$$

where $S_n = X_{n,1} + \dots + X_{n,r_n}$.

Discussions

- Theorem 27.1 (Lindeberg-Lévy theorem) is a special case of Theorem 27.2.
- Specifically, (i) $r_n = n$, (ii) $X_{n,k} = X_k$, (iii) finite variance, and (iv) identically distributed assumption give:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{n\sigma^2} E [X_k^2 I_{[|X_k| \geq \varepsilon \sigma \sqrt{n}]}] = \lim_{n \rightarrow \infty} \frac{1}{\sigma^2} E [X_1^2 I_{[|X_1| \geq \varepsilon \sigma \sqrt{n}]}] = 0,$$

where $\sigma^2 = E [X_1^2]$.

Lindeberg Theorem

27-13

Proof:

- Without loss of generality, assume $s_n = 1$ (since we can replace each $X_{n,k}$ by $X_{n,k}/s_n$).

Hence, Lindeberg's condition reduces to:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \int_{|x| \geq \varepsilon} x^2 dF_{X_{n,k}}(x) = \lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} E \left[X_{n,k}^2 I_{[|X_{n,k}| \geq \varepsilon]} \right] = 0.$$

- Since $E[X_{n,k}] = 0$,

$$\left| \varphi_{X_{n,k}}(t) - \left(1 - \frac{1}{2} t^2 E[X_{n,k}^2] \right) \right| \leq E \left[\min \{ |tX_{n,k}|^3, |tX_{n,k}|^2 \} \right] < \infty.$$

Lemma If X has a moment of order n , then

$$\left| \varphi(t) - \sum_{k=0}^n \frac{(it)^k}{k!} E[X^k] \right| \leq E \left[\min \left\{ \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right\} \right] \\ \left(\leq E \left[\min \{ |tX|^{n+1}, |tX|^n \} \right] \text{ for } n \geq 2 \right).$$

Lindeberg Theorem

27-14

- Observe that:

$$\begin{aligned} E \left[\min \left\{ |tX_{n,k}|^3, |tX_{n,k}|^2 \right\} \right] &= \int_{|x| < \varepsilon} \min \left\{ |tx|^3, |tx|^2 \right\} dF_{X_{n,k}}(x) \\ &+ \int_{|x| \geq \varepsilon} \min \left\{ |tx|^3, |tx|^2 \right\} dF_{X_{n,k}}(x) \\ &\leq \int_{|x| < \varepsilon} |tx|^3 dF_{X_{n,k}}(x) + \int_{|x| \geq \varepsilon} |tx|^2 dF_{X_{n,k}}(x) \\ &\leq |t\varepsilon| \int_{|x| < \varepsilon} |tx|^2 dF_{X_{n,k}}(x) + \int_{|x| \geq \varepsilon} |tx|^2 dF_{X_{n,k}}(x) \\ &\leq \varepsilon |t|^3 E[X_{n,k}^2] + t^2 \int_{|x| \geq \varepsilon} x^2 dF_{X_{n,k}}(x), \end{aligned}$$

which implies that:

$$\begin{aligned} \sum_{k=1}^{r_n} \left| \varphi_{X_{n,k}}(t) - \left(1 - \frac{1}{2} t^2 E[X_{n,k}^2] \right) \right| &\leq \varepsilon |t|^3 \sum_{k=1}^{r_n} E[X_{n,k}^2] + t^2 \sum_{k=1}^{r_n} \int_{|x| \geq \varepsilon} x^2 dF_{X_{n,k}}(x) \\ &= \varepsilon |t|^3 + t^2 \sum_{k=1}^{r_n} \int_{|x| \geq \varepsilon} x^2 dF_{X_{n,k}}(x). \end{aligned}$$

Lindeberg Theorem

27-15

As ε can be made arbitrarily small,

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \left| \varphi_{X_{n,k}}(t) - \left(1 - \frac{1}{2} t^2 E[X_{n,k}^2] \right) \right| = 0.$$

• By

$$\left| \prod_{k=1}^{r_n} \varphi_{X_{n,k}}(t) - \prod_{k=1}^{r_n} \left(1 - \frac{1}{2} t^2 E[X_{n,k}^2] \right) \right| \leq \sum_{k=1}^{r_n} \left| \varphi_{X_{n,k}}(t) - \left(1 - \frac{1}{2} t^2 E[X_{n,k}^2] \right) \right|,$$

we get:

$$\lim_{n \rightarrow \infty} \left| \prod_{k=1}^{r_n} \varphi_{X_{n,k}}(t) - \prod_{k=1}^{r_n} \left(1 - \frac{1}{2} t^2 E[X_{n,k}^2] \right) \right| = 0.$$

(Hint: Is this correct? See the end of the proof!)

It remains to show that

$$\lim_{n \rightarrow \infty} \prod_{k=1}^{r_n} \left(1 - \frac{1}{2} t^2 E[X_{n,k}^2] \right) = e^{-t^2/2}.$$

Lindeberg Theorem

27-16

•

$$\begin{aligned} \left| e^{-t^2/2} - \prod_{k=1}^{r_n} \left(1 - \frac{1}{2}t^2 E[X_{n,k}^2] \right) \right| &= \left| \prod_{k=1}^{r_n} e^{-t^2 E[X_{n,k}^2]/2} - \prod_{k=1}^{r_n} \left(1 - \frac{1}{2}t^2 E[X_{n,k}^2] \right) \right| \\ &\leq \sum_{k=1}^{r_n} \left| e^{-t^2 E[X_{n,k}^2]/2} - \left(1 - \frac{1}{2}t^2 E[X_{n,k}^2] \right) \right| \end{aligned}$$

For each complex z ,

$$|e^z - 1 - z| \leq |z|^2 \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{k!} \leq |z|^2 \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{(k-2)!} = |z|^2 e^{|z|}.$$

$$\begin{aligned} &\leq \sum_{k=1}^{r_n} \frac{1}{4} t^4 E^2[X_{n,k}^2] e^{t^2 E[X_{n,k}^2]/2} \quad (\text{Take } z = -\frac{1}{2}t^2 E[X_{n,k}^2]) \\ &\leq \frac{1}{4} t^4 e^{t^2/2} \sum_{k=1}^{r_n} E^2[X_{n,k}^2] \quad (\text{by } E[X_{n,k}^2] \leq s_n^2 = 1) \\ &\leq \frac{1}{4} t^4 e^{t^2/2} \left(\max_{1 \leq k \leq r_n} E[X_{n,k}^2] \right) \sum_{k=1}^{r_n} E[X_{n,k}^2] \\ &= \frac{1}{4} t^4 e^{t^2/2} \left(\max_{1 \leq k \leq r_n} E[X_{n,k}^2] \right). \end{aligned}$$

Lindeberg Theorem

27-17

- Finally,

$$E[X_{n,k}^2] \leq \varepsilon^2 + \int_{[|x|^2 \geq \varepsilon^2]} x^2 dF_{X_{n,k}}(x),$$

which implies that

$$\begin{aligned} \max_{1 \leq k \leq r_n} E[X_{n,k}^2] &\leq \varepsilon^2 + \max_{1 \leq k \leq r_n} \int_{[|x| \geq \varepsilon]} x^2 dF_{X_{n,k}}(x) \\ &\leq \varepsilon^2 + \sum_{1 \leq k \leq r_n} \int_{[|x| \geq \varepsilon]} x^2 dF_{X_{n,k}}(x). \end{aligned}$$

The proof is then completed by taking arbitrarily small ε and Lindeberg's condition. \square

$(1 - \frac{1}{2}t^2 E[X_{n,k}^2])$ is a complex number of modulus at most 1 for $t^2 \leq 4/E[X_{n,k}^2] \uparrow \infty$.

Converse to Lindeberg theorem

- Give an array of independent zero-mean random variables $X_{n,1}, \dots, X_{n,r_n}$.
If $S_n/s_n \Rightarrow N$, then Lindeberg's condition holds, provided that

$$\max_{1 \leq k \leq r_n} E[X_{n,k}^2]/s_n^2 \rightarrow 0.$$

Lindeberg Theorem

27-18

- Without the extra condition of

$$\max_{1 \leq k \leq r_n} E[X_{n,k}^2]/s_n^2 \rightarrow 0,$$

the converse of Lindeberg theorem may not be valid.

Counterexample Let $X_{n,k}$ be Gaussian distributed with mean 0 and variance 1 for $1 \leq k < r_n = n$, and $X_{n,n}$ be Gaussian distributed with mean 0 and variance $(n - 1)$.

Thus,

$$s_n^2 = \sum_{k=1}^n E[X_{n,k}^2] = (n - 1) + (n - 1) = 2(n - 1).$$

Lindeberg Theorem

27-19

Then $S_n/s_n = (X_{n,1} + X_{n,2} + \cdots + X_{n,n})/s_n$ is Gaussian distributed with mean 0 and variance 1, but

$$\begin{aligned} \sum_{k=1}^n \frac{1}{s_n^2} \int_{\{|x| \geq \varepsilon s_n\}} x^2 dF_{X_{n,k}}(x) &\geq \frac{1}{s_n^2} \int_{\{|x| \geq \varepsilon s_n\}} x^2 dF_{X_{n,n}}(x) \\ &= \frac{1}{s_n^2} \int_{\{|x| \geq \varepsilon s_n\}} x^2 \frac{1}{\sqrt{2\pi(n-1)}} e^{-x^2/(2(n-1))} dx \\ &= \frac{1}{2(n-1)} \int_{\{|y| \geq \varepsilon \sqrt{2(n-1)}/\sqrt{n-1}\}} (n-1)y^2 \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \\ &\quad \text{where } x = \sqrt{n-1}y \\ &= \frac{1}{2} \int_{\{|y| \geq \varepsilon \sqrt{2}\}} y^2 \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \end{aligned}$$

and hence Lindeberg's condition fails.

Notably,

$$\max_{1 \leq k \leq n} \frac{E[X_{n,k}^2]}{s_n^2} = \max_{1 \leq k \leq n} \frac{(n-1)}{2(n-1)} = \frac{1}{2}$$

does not converge to zero.

Goncharov's theorem

27-20

Observation If $X_{n,k}$ is uniformly bounded for each n and k , and $s_n \rightarrow \infty$ as $n \rightarrow \infty$, then Lindeberg's condition holds.

Proof: Let M be the bound for $X_{n,k}$, namely $\Pr[|X_{n,k}| \leq M] = 1$ for each k and n . Then for any $\varepsilon > 0$, εs_n will ultimately exceed M , and therefore $\int_{|x| \geq \varepsilon s_n} x^2 dF_{X_{n,k}}(x) = 0$ for every n and k , which implies

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|x| \geq \varepsilon s_n} x^2 dF_{X_{n,k}}(x) = 0.$$

□

Goncharov's theorem

27-21

Example 27.3 Let $\Pr[Y_{n,k} = 1] = \frac{1}{k}$ and $\Pr[Y_{n,k} = 0] = 1 - \frac{1}{k}$.

Let $X_{n,k} = Y_{n,k} - E[Y_{n,k}] = Y_{n,k} - \frac{1}{k}$ for $1 \leq k \leq n$.

Then $E[X_{n,k}] = 0$, $E[X_{n,k}^2] = \frac{k-1}{k^2}$, and $s_n^2 = \sum_{k=1}^n \frac{k-1}{k^2}$

Since $|X_{n,k}|$ is bounded by 1 with probability 1, and $s_n^2 \xrightarrow{n \rightarrow \infty} \infty$, Lindeberg's condition holds.

The Lindeberg theorem thus concludes:

$$\frac{X_{n,1} + \cdots + X_{n,n}}{s_n} = \frac{Y_{n,1} + \cdots + Y_{n,n} - \sum_{k=1}^n (1/k)}{\sqrt{\sum_{k=1}^n (1/k) - \sum_{k=1}^n (1/k^2)}} \Rightarrow N.$$

Goncharov's theorem

27-22

Goncharov's theorem:

$$\frac{Y_{n,1} + \cdots + Y_{n,n} - \log(n)}{\sqrt{\log(n)}} \Rightarrow N.$$

Theorem 25.4 If $X_n \Rightarrow X$ and $X_n - Y_n \Rightarrow 0$, then $Y_n \Rightarrow X$.

Proof: Goncharov's theorem can be easily proved by:

$$\left(\frac{Y_{n,1} + \cdots + Y_{n,n} - \sum_{k=1}^n (1/k)}{\sqrt{\sum_{k=1}^n (1/k) - \sum_{k=1}^n (1/k^2)}} \right) \Rightarrow N.$$

and

$$\left(\frac{Y_{n,1} + \cdots + Y_{n,n} - \sum_{k=1}^n (1/k)}{\sqrt{\sum_{k=1}^n (1/k) - \sum_{k=1}^n (1/k^2)}} \right) - \left(\frac{Y_{n,1} + \cdots + Y_{n,n} - \log(n)}{\sqrt{\log(n)}} \right) \Rightarrow 0.$$

□

Lyapounov's condition

27-23

Discussions

- Lindeberg's condition is quite satisfiable in the sense that it is the *sufficient* and *necessary* condition for normalized row sum of independent array random variables to converge to standard normal, provided that the variances are uniformly and asymptotically negligible.
- It however may not be easy to examine the validity of Lindeberg's condition.
- A useful sufficiency for Lindeberg's condition to hold is Lyapounov's condition, which is often easier to verify than Lindeberg's condition (since only moments are involved in the computation).

Lyapounov's condition Fix an array of independent zero-mean random variables $X_{n,1}, \dots, X_{n,r_n}$. For some $\delta > 0$,

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} E [|X_{n,k}|^{2+\delta}] = 0,$$

where $s_n^2 = \sum_{k=1}^{r_n} E [X_{n,k}^2]$.

Lyapounov's condition

27-24

Observation Lyapounov's condition implies Lindeberg's condition.

Proof:

$$\begin{aligned} \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{[|x| \geq \varepsilon s_n]} x^2 dF_{X_{n,k}}(x) &\leq \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{[|x| \geq \varepsilon s_n]} x^2 \left(\frac{|x|^\delta}{(\varepsilon s_n)^\delta} \right) dF_{X_{n,k}}(x) \\ &= \frac{1}{\varepsilon^\delta} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} \int_{[|x| \geq \varepsilon s_n]} |x|^{2+\delta} dF_{X_{n,k}}(x) \\ &\leq \frac{1}{\varepsilon^\delta} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} \int_{\mathfrak{R}} |x|^{2+\delta} dF_{X_{n,k}}(x). \end{aligned}$$

□

Theorem 27.3 For an array of independent zero-mean random variables $X_{n,1}, \dots, X_{n,r_n}$, if Lyapounov's condition holds for some positive δ , then

$$\frac{S_n}{s_n} \Rightarrow N,$$

where $S_n = X_{n,1} + \dots + X_{n,r_n}$.

Lyapounov's condition

27-25

Example 27.4 The Lyapounov's condition is always valid for i.i.d. sequence with bounded $(2 + \delta)$ th absolute moment.

Proof:

$$\begin{aligned}\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} E [|X_{n,k}|^{2+\delta}] &= \lim_{n \rightarrow \infty} \frac{r_n E [|X_1|^{2+\delta}]}{r_n^{1+\delta/2} E^{1+\delta/2}[X_1^2]} \\ &= \frac{E [|X_1|^{2+\delta}]}{E^{1+\delta/2}[X_1^2]} \lim_{n \rightarrow \infty} r_n^{-\delta/2} \\ &= 0\end{aligned}$$

□

Problem of coupon collector

27-26

Example 27.5 (Problem of coupon collector) A coupon collector has to collect r_n distinct coupons to exchange for some free gift. Each purchase will give him one coupon, randomly and with replacement. The statistical behavior of this collection can be described as follows.

- Coupons are drawn from a coupon population of size n , randomly and with replacement, until the number of distinct coupons that have been collected is r_n , where $1 \leq r_n \leq n$.
- Let S_n be the number of purchases required for this collection.
- Assume that $r_n/n \rightarrow \rho > 0$.

What is the approximation distribution of $(S_n - E[S_n])/\sqrt{\text{Var}[S_n]}$?

We can also apply this problem to, say, that r_n out of n pieces need to be collected in order to recover the original information.

Problem of coupon collector

27-27

Solution:

- If $(k-1)$ distinct coupons have thus far been collected, the number of purchases until the next distinct one enters is distributed as X_k , where

$$\Pr[X_k = j] = (1 - p_k)^{j-1} p_k \quad \text{for } j = 1, 2, 3, \dots$$

$$\text{where } p_k = \frac{n - (k - 1)}{n} = 1 - \frac{k - 1}{n}.$$

(Notably, we wish to see any one of the remaining $n - (k - 1)$ coupons to appear.)

- $S_n = X_1 + X_2 + \dots + X_{r_n}$
- $E[X_k] = \frac{1}{p_k}$ and $\text{Var}[X_k] = \frac{1 - p_k}{p_k^2}$.

Problem of coupon collector

27-28

- Hence,

$$n \int_{-1/n}^{r_n/n-1/n} \frac{1}{1-x} dx \leq E[S_n] = \sum_{k=1}^{r_n} \frac{1}{p_k} = \sum_{k=1}^{r_n} \frac{1}{1 - \frac{k-1}{n}} \leq n \int_0^{r_n/n} \frac{1}{1-x} dx,$$

which implies that

$$\lim_{n \rightarrow \infty} \frac{E[S_n]}{n \log[1/(1-\rho)]} = 1.$$

-

$$\text{Var}[S_n] = \sum_{k=1}^{r_n} \text{Var}[X_k] = \sum_{k=1}^{r_n} \frac{1-p_k}{p_k^2} = \sum_{k=1}^{r_n} \frac{\frac{k-1}{n}}{\left(1 - \frac{k-1}{n}\right)^2},$$

which implies that

$$\lim_{n \rightarrow \infty} \frac{\text{Var}[S_n]}{n \int_0^\rho x/(1-x)^2 dx} = \lim_{n \rightarrow \infty} \frac{\text{Var}[S_n]}{n[\rho/(1-\rho) + \log(1-\rho)]} = 1.$$

Problem of coupon collector

27-29

- Therefore, since Lyapounov's's condition holds for $\delta = 2$, i.e.,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^4} E [|X_k - E[X_k]|^4] &\leq \lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^4} E [|X_k|^4] \\ &= \lim_{n \rightarrow \infty} \frac{1}{s_n^4} \sum_{k=1}^{r_n} \frac{(2 - p_k)(12 - 12p_k + p_k^2)}{p_k^4} \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{s_n^4} \sum_{k=1}^{r_n} \frac{24}{p_k^4} \\ &= \lim_{n \rightarrow \infty} \frac{1}{s_n^4} \sum_{k=1}^{r_n} \frac{24}{p_k^4} \\ &\leq \lim_{n \rightarrow \infty} \frac{24n}{s_n^4} \int_0^{r_n/n} \frac{1}{(1-x)^4} dx = 0, \end{aligned}$$

we obtain:

$$\frac{S_n - n \log[1/(1 - \rho)]}{\sqrt{n[\rho/(1 - \rho) + \log(1 - \rho)]}} \Rightarrow N.$$

□

Lindeberg condition $\not\Rightarrow$ Lyapounov condition

27-30

Counterexample Lindeberg's condition does not necessarily imply Lyapounov's condition.

Consider a sequence of random variables that are independent, and each has density function

$$f(x) = \frac{c}{x^3(\log(x))^2} \text{ for } x \geq e \approx 2.71828 \dots,$$

where $c = e^{-2} - 2 \int_2^\infty t^{-1}e^{-t}dt \approx 0.0375343 \dots$

It can be verified that the i.i.d. sequence has finite marginal second moment

$$\int_e^\infty \frac{c}{x(\log(x))^2} dx = \int_1^\infty \frac{c}{u^2} du = c; \quad (u = \log(x))$$

hence, Lindeberg's condition holds.

However,

$$E[|X_1|^{2+\delta}] = \int_e^\infty \frac{c}{x^{1-\delta}(\log(x))^2} dx = c \int_1^\infty \frac{e^{\delta u}}{u^2} du = \infty,$$

where $u = \log(x)$. Accordingly, Lyapounov's condition does not hold!

Dependent variables

27-31

- Can we extend the central limit theorem to sequence of dependent variables?

Definition (α -mixing) A sequence of random variables is said to be α -mixing, if there exists a non-negative sequence $\alpha_1, \alpha_2, \alpha_3, \dots$ such that

$$\lim_{n \rightarrow \infty} \alpha_n = 0$$

and

$$\sup_{k \geq 1} \sup_{\mathcal{H} \in \mathfrak{R}^k \wedge \mathcal{G} \in \mathfrak{R}^\infty} \left| \Pr [(X_1^k \in \mathcal{H}) \wedge (X_{n+k}^\infty \in \mathcal{G})] - \Pr [X_1^k \in \mathcal{H}] \Pr [X_{n+k}^\infty \in \mathcal{G}] \right| \leq \alpha_n.$$

- Operational meaning: By α -mixing, we mean that $X_1^k = (X_1, X_2, \dots, X_k)$ and $X_{n+k}^\infty = (X_{n+k}, X_{n+k+1}, \dots)$ are approximately independent, when n is large.
- An independent sequence is α -mixing with $\alpha_k = 0$ for all $k = 1, 2, 3, \dots$

m -dependent

27-32

Definition A sequence of random variables is said to be *m -dependent*, if

$$(X_i, \dots, X_{i+k}) \quad \text{and} \quad (X_{i+k+n}, \dots, X_{i+k+n+l})$$

are independent whenever $n > m$.

- An independent sequence is 0-dependent.
- A m -dependent sequence is α -mixing with $\alpha_n = 0$ for $n > m$.

Example 27.7 Let Y_1, Y_2, \dots be i.i.d. sequence.

Define $X_n = f(Y_n, Y_{n+1}, \dots, Y_{n+m})$ for a real Borel-measurable function on \mathfrak{R}^{m+1} .

Then X_1, X_2, \dots is stationary and m -dependent.

α -mixing and m -dependent

27-33

Example 27.6 Let Y_1, Y_2, \dots be a first-order finite-state Markov chain with positive transition probabilities.

Suppose the initial probability used is the one that makes Y_1, Y_2, \dots stationary.

Then

$$\begin{aligned} & \Pr[Y_1 = i_1, \dots, Y_k = i_k, Y_{k+n} = j_0, \dots, Y_{k+n+l} = j_\ell] \\ &= (p_{i_1 i_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k}) p_{i_k j_0}^{(n)} (p_{j_0 j_1} \cdots p_{j_{\ell-1} j_\ell}), \end{aligned}$$

where $p_{ij} = P_{Y_n|Y_{n-1}}(j|i)$ and $p_{ij}^{(k)} = P_{Y_n|Y_{n-k}}(j|i)$.

Also,

$$\begin{aligned} & \Pr[Y_1 = i_1, \dots, Y_k = i_k] \Pr[Y_{k+n} = j_0, \dots, Y_{k+n+l} = j_\ell] \\ &= (p_{i_1 i_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k}) p_{j_0} (p_{j_0 j_1} \cdots p_{j_{\ell-1} j_\ell}). \end{aligned}$$

α -mixing and m -dependent

27-34

Theorem 8.9 There exists a stationary distribution $\{\pi_i\}$ for a *finite-state, irreducible, aperiodic, first-order* Markov chain such that

$$\left| p_{ij}^{(n)} - \pi_j \right| \leq A\rho^n,$$

where $A \geq 0$ and $0 \leq \rho < 1$.

A first-order Markov chain is *aperiodic*, if the greatest common divisor of the integers in the set $\{n \in \mathbb{N} : p_{jj}^{(n)} > 0\}$ is 1 for every j .

A Markov chain is *irreducible*, if for every i and j , $p_{ij}^{(n)} > 0$ for some n .

$$\begin{aligned}
 & \left| \Pr [(Y_1^k \in \mathcal{H}) \wedge (Y_{n+k}^{n+k+\ell} \in \mathcal{G})] - \Pr [Y_1^k \in \mathcal{H}] \Pr [Y_{n+k}^{n+k+\ell} \in \mathcal{G}] \right| \\
 &= \left| \sum_{i_1^k \in \mathcal{H}} \sum_{j_0^\ell \in \mathcal{G}} \left(\Pr[Y_1 = i_1, \dots, Y_k = i_k, Y_{k+n} = j_0, \dots, Y_{k+n+\ell} = j_\ell] \right. \right. \\
 &\quad \left. \left. - \Pr[Y_1 = i_1, \dots, Y_k = i_k] \Pr[Y_{k+n} = j_0, \dots, Y_{k+n+\ell} = j_\ell] \right) \right| \\
 &= \left| \sum_{i_1^k \in \mathcal{H}} \sum_{j_0^\ell \in \mathcal{G}} (p_{i_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k}) \left(p_{i_k j_0}^{(n)} - p_{j_0} \right) (p_{j_0 j_1} \cdots p_{j_{\ell-1} j_\ell}) \right| \\
 &\leq \sum_{i_1^k \in \mathcal{H}} \sum_{j_0^\ell \in \mathcal{G}} (p_{i_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k}) \left| p_{i_k j_0}^{(n)} - p_{j_0} \right| (p_{j_0 j_1} \cdots p_{j_{\ell-1} j_\ell}) \\
 &\leq A \rho^n \sum_{i_1^k \in \mathcal{H}} \sum_{j_0^\ell \in \mathcal{G}} (p_{i_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k}) (p_{j_0 j_1} \cdots p_{j_{\ell-1} j_\ell}) \\
 &\leq A \rho^n \sum_{i_1^k \in \mathcal{Y}^k} \sum_{j_0^\ell \in \mathcal{Y}^{\ell+1}} (p_{i_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k}) (p_{j_0 j_1} \cdots p_{j_{\ell-1} j_\ell}) \\
 &= A \rho^n \sum_{j_0 \in \mathcal{Y}} 1 = A |\mathcal{Y}| \rho^n,
 \end{aligned}$$

for some $A \geq 0$ and $0 \leq \rho < 1$.

α -mixing and m -dependent

27-36

Since the upper bound is independent of $k \geq 1$, $\ell \geq 0$, $\mathcal{H} \in \mathfrak{R}^k$ and $\mathcal{G} \in \mathfrak{R}^{\ell+1}$, we obtain that Y_1, Y_2, \dots is α -mixing with $\alpha_n = A|\mathcal{Y}|\rho^n$. ■

Hence, a *finite-state, irreducible, aperiodic, first-order* Markov chain is α -mixing with exponentially decaying α_n .

Central limit theorem for α -mixing sequence

27-37

Theorem 27.4 Suppose that X_1, X_2, \dots is zero-mean, stationary and α -mixing with $\alpha_n = O(n^{-5})$ as $n \rightarrow \infty$.

Assume that $E[X_n^2] < \infty$. Then

1. $\frac{1}{n} \text{Var}[S_n] \xrightarrow{n \rightarrow \infty} \sigma^2 = E[X_1^2] + 2 \sum_{k=1}^{\infty} E[X_1 X_{1+k}]$.

2.

$$\frac{S_n}{\sigma \sqrt{n}} \Rightarrow N. \quad \left(\frac{S_n}{\sqrt{\text{Var}[S_n]}} \Rightarrow N. \right)$$

Proof: No proof is provided.

□

Central limit theorem for α -mixing sequence

27-38

Example 27.8 Let Y_1, Y_2, \dots be a first-order finite-state Markov chain with positive transition probabilities.

Suppose the initial probability used is the one that makes Y_1, Y_2, \dots stationary.

Example 27.6 already proves that Y_1, Y_2, \dots is α -mixing with $\alpha_n = A|\mathcal{Y}|\rho^n$.

Thus, $X_1 = Y_1 - E[Y_1], X_2 = Y_2 - E[Y_2], \dots$ is also α -mixing with $\alpha_n = A|\mathcal{Y}|\rho^n$.

Furthermore, the finite state assumption indicates that all moments of X_n are bounded.

Accordingly, Theorem 27.4 holds, namely,

$$\frac{X_1 + \dots + X_n}{\sqrt{\text{Var}[X_1 + \dots + X_n]}} \Rightarrow N.$$

Central limit theorem for α -mixing sequence

27-39

More specifically, let

$$Y_n \in \{-1, +1\}$$

and

$$P_{Y_n|Y_{n-1}}(+1|-1) = P_{Y_n|Y_{n-1}}(-1|+1) = \varepsilon.$$

Since the probability transition matrix can be expressed as:

$$\begin{bmatrix} 1 - \varepsilon & \varepsilon \\ \varepsilon & 1 - \varepsilon \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 - 2\varepsilon \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T,$$

we have:

$$\begin{bmatrix} 1 - \varepsilon & \varepsilon \\ \varepsilon & 1 - \varepsilon \end{bmatrix}^n = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1 - 2\varepsilon)^n \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T = \begin{bmatrix} \frac{1}{2} + \frac{1}{2}(1 - 2\varepsilon)^n & \frac{1}{2} - \frac{1}{2}(1 - 2\varepsilon)^n \\ \frac{1}{2} - \frac{1}{2}(1 - 2\varepsilon)^n & \frac{1}{2} + \frac{1}{2}(1 - 2\varepsilon)^n \end{bmatrix},$$

and Y_1, Y_2, \dots are α -mixing with $\alpha_n = 2(1/2)|1 - 2\varepsilon|^n = |1 - 2\varepsilon|^n$.

Central limit theorem for α -mixing sequence

27-40

Theorem 27.4 then gives that:

$$\begin{aligned} \frac{\text{Var}[Y_1 + \cdots + Y_n]}{n} &\xrightarrow{n \rightarrow \infty} E[Y_1^2] + 2 \sum_{k=1}^{\infty} E[Y_1 Y_{1+k}] \\ &= \sum_{y_1 \in \{-1,1\}} y_1^2 P_{Y_1}(y_1) \\ &\quad + 2 \sum_{k=1}^{\infty} \sum_{y_1 \in \{-1,1\}} \sum_{y_{1+k} \in \{-1,1\}} y_1 y_{1+k} P_{Y_{1+k}|Y_1}(y_{1+k}|y_1) P_{Y_1}(y_1) \\ &= 1 + 2 \sum_{k=1}^{\infty} \sum_{y_1 \in \{-1,1\}} \sum_{y_2 \in \{-1,1\}} \frac{1}{2} y_1 y_{1+k} P_{Y_{1+k}|Y_1}(y_{1+k}|y_1) \\ &= 1 + 2 \sum_{k=1}^{\infty} (1 - 2\varepsilon)^k = \frac{1}{\varepsilon} - 1, \end{aligned}$$

and

$$\frac{Y_1 + \cdots + Y_n}{\sqrt{n \left(\frac{1}{\varepsilon} - 1\right)}} \Rightarrow N.$$

■