Mathematical Statistics I

Chapter 5: Limit theorems

Jesse Wheeler

Outline

1. Convergence Concepts

Convergence Concepts

$\underline{X_1, X_2, \ldots} \longrightarrow$

Introduction

- This material comes primarily from Rice (2007, Chapter 5), but will be supplemented with material from Casella and Berger (2024, Chapter 5).
- In this chapter, we are interested in the convergence of sequences of random variables.
- For instance, we are interested in the convergence of the sample mean, $\bar{X}_n = (X_1 + X_2 + \ldots + X_n)/n$, as the number of samples n grows.
- Because $\bar{X}_n$ is itself a random variable, we have to carefully define what it means for the convergence of a random variable.
- In this class, we are mainly concerned with three types of convergence.

Introduction II

- Because convergence of random variables is a tricky topic, we will treat them in varying amounts of detail.

Convergence in Probability

- The first type of convergence is one of the weaker types, and is usually easy(ish) to verify.

Definition: Convergence in Probability

A sequence of random variables $X_1, X_2, \ldots$ converges in probability to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\big(|X_n - X| \geq \epsilon\big) = 0$$

or, equivalently,

$$\lim_{n\to\infty} P\big(|X_n - X| < \epsilon\big) = 1.$$

Convergence in Probability II

- We often use the shorthand $X_n \overset{P}{\to} X$ to denote "X_n converges in probability to X as n goes to infinity". \overset{P}{ } \rightarrow
- Note that the X_i in the definition above do *not* need to be independent and identically distributed.
- The distribution of X_n changes as the subscript changes, and each of the convergence concepts we will discuss will describe different ways in which the distribution of X_n converges to some limiting distribution as the subscript becomes large.
- A special case is when the limiting random variable X is a constant.

$$\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0$$

could be constant.

Example: The (Weak) Law of Large Numbers

Let $X_1, X_2, \ldots$ be iid random variables with $E[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then $\bar{X}_n \xrightarrow{P} \mu$.

Proof.

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

- Apply Chebychev's inequality:

$$P(|\bar{x}_n - \mu| \geq \varepsilon) = P((\bar{X}_n - \mu)^2 \geq \varepsilon^2)$$

$$\leq \frac{E[(\bar{X}_n - \mu)^2]}{\varepsilon^2}$$

$\mathrm{Var}(\bar{X}_n)$

$$= \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$$

Note:

$$E[\overline{X}_n] = E\left[\frac{1}{n}\sum X_i\right]$$
$$= \frac{1}{n}\sum E[X_i] = \frac{1}{n}(n\mu)$$
$$= \mu$$

$X_1, X_2, \ldots$ iid $\leftarrow$ independ

$$\mathrm{var}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\sum \mathrm{var}(X_i)$$

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

Convergence in Probability IV

- The WLLN is very elegant; under general conditions, the sample mean of independent random variables approaches the μ = population mean as $n \to \infty$.
- This is also used for proportions, as proportions are just means of indicator random variables.
- The WLLN can also be extended to show that the results hold even if the variance is infinite, the only condition needed is that the expectation is finite. However, the proof in this case is beyond the scope of this course.
- When a sequence of the "same" sample quantity approaches a constant, we say that the sample quantity is *consistent*.

$\{\bar{X}_n\}$ $\bar{X}_n$ is a consistent estimator of $E[X_i] = \mu$

- A natural extension of the definition of the convergence of probability, is convergence of functions of random variables: $h(X_1), h(X_2), \ldots$.

Theorem: Convergence in probability for continuous functions

Let $X_1, X_2, \ldots$ be a sequence of random variables that converges in probability to a random variable X, and let h be a continuous function.

Then, $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.

$Y_1 = h(X_1),\ Y_2 = h(X_2) \ldots \xrightarrow{P} Y = h(X)$

$|y - x| < \delta \Rightarrow |h(y) - h(x)| < \varepsilon$

$$\lim_{n\to\infty} P\left(|h(X_n) - h(X)| > \delta \right) = 0$$

$$\{(x,y) : |y-x| < \delta\} \subseteq \{(x,y) : |h(y)-h(x)| < \varepsilon\}$$

$$1 - P\left(|h(X_n) - h(X)| \leq \delta \right)$$

$$P(|h$$

$$|X_n - X| < \varepsilon \rightarrow |h(X_n) - h(X)| < \delta$$

Detailed proof below :

Proof: Using the second definition of convergence in probability, we want to show

$$(\star) \quad \lim_{n\to\infty} P\left(|h(X_n)-h(X)| < \varepsilon\right) = 1, \quad \text{for all } \varepsilon>0.$$

- Because h is a continuous function, then we fix some $\varepsilon>0$, and for all $\omega\in\Omega$, there exists a $\delta>0$ s.t. if:

$$(1) \quad |X_n(\omega)-X(\omega)| < \delta, \quad \underline{\text{then}} \quad (\Rightarrow)$$

$$(2) \quad |h(X_n(\omega)) - h(X(\omega))| < \varepsilon.$$

(next slide)

(Continued...)

Then, we can note that the set of $\omega \in \Omega$ that satisfy (1) must satisfy (2). Mathematically,

$$\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \delta\} \subseteq \{\omega \in \Omega : |h(X_n(\omega)) - h(X(\omega))| < \varepsilon\}.$$

Thus,

$$P(|X_n - X| < \delta) \leq P(|h(X_n) - h(X)| < \varepsilon).$$

However, we assume that $X_n \xrightarrow{P} X$, so taking limits,

$$\lim_{n \to \infty} P(|X_n - X| < \delta) = 1 \leq \lim_{n \to \infty} P(|h(X_n) - h(X)| < \varepsilon)$$

- Furthermore, using $p_n = P(|h(X_n) - h(X)| < \varepsilon)$ is a valid probability, $1 \le \lim_{n\to\infty} p_n \le 1$

implies

$$\lim_{n\to\infty} P(|h(X_n) - h(X)| < \varepsilon) = 1,$$

proving $h(X_n) \xrightarrow{P} h(X)$ $\square$.

Almost sure convergence

- Our next convergence concept is stronger than convergence in probability.

Definition: Almost Sure Convergence

A sequence of random variables $X_1, X_2, \ldots$ converge almost surely to a random variable X if, for every $\epsilon > 0$, Ω

$$P\big(\lim_{n\to\infty} |X_n - X| < \epsilon\big) = 1,$$

or

$$P\big(\lim_{n\to\infty} X_n = X\big) = 1.$$

- Almost sure convergence is often written as $X_n \overset{a.s.}{\to} X$.

Almost sure convergence II

- It appears similar to convergence in probability, but they are in fact very different. In particular, almost sure convergence is a stronger concept.
- One way to think about this difference is that the probability gives a weight to individual sets.
- For convergence in probability, the set where $|X_n - X| > \epsilon$ can have positive probability, but that probability converges to zero for large n.
- For almost sure convergence, the set where $|X_n - X| > \epsilon$ has probability zero. This doesn't imply that the set $|X_n - X| > \epsilon$ is empty, but it has zero probability.

Almost sure convergence III

- Almost sure convergence is very similar to pointwise convergence of a sequence of functions. This is no accident, as random variables *are* functions:

$$P\Big(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big) = 1.$$

- In the equivalent definition above, we see we must have point-wise convergence almost-everywhere, except for the possibility that for some set $N \subset \Omega$ such that $P(N) = 0$, we allow $s \in N$ to not converge: $\lim_{n\to\infty} X_n(s) \neq X(s)$.

Example: Convergence in prob, not a.s. $X_1, X_2, \ldots$

$X(s) = s$

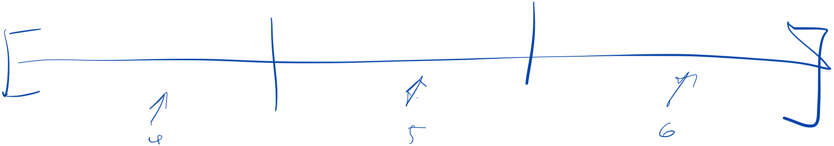
Let the sample space $\Omega = [0, 1]$, and assign the uniform probability on this interval. Define the sequence of random variables X_i as: $X_1(s) = s + 1_{[0,1]}(s)$, $X_2(s) = s + 1_{[0,\frac{1}{2}]}(s)$, $X_3(s) = s + 1_{[\frac{1}{2},1]}(s)$, $X_4(s) = s + 1_{[0,\frac{1}{3}]}(s)$, $X_5(s) = s + 1_{[\frac{1}{3},\frac{2}{3}]}(s)$, $X_6(s) = s + 1_{[\frac{2}{3},1]}(s), \ldots$, and then define $X(s) = s$. We can see that $X_n \xrightarrow{P} X$. However, X_n does not converge almost surely, because there is no values $s \in \Omega$ that satisfy $X_n(s) \to X(s)$. For every s, the value of $X_n(s)$ alternates between s and $s + 1$ infinitely often.

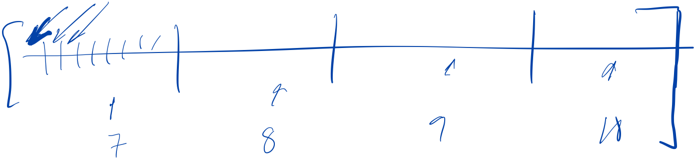
$0 \; [\qquad\qquad] \; 1$

$X_1(s) = s + \underbrace{1_{[0,1]}(s)} = s+1$

$X_2(s) = s + \mathbb{1}_{[0,1/2]}(s)$, $X_3(s) = s + \mathbb{1}_{[1/2,1]}(s)$

$X_2(s) = s + 1$, $X_3(s) = s$





$$p(\underbrace{|X_n - X|}\ > \varepsilon) \approx 1 - \frac{1}{n} \to 1$$

$$S \neq \mathbb{1}_{[1/n]}$$

Almost sure convergence V

Theorem: almost sure convergence implies convergence in probability $\{X_n\}$

If $X_1, X_2, \ldots$ are a sequence of random variables such that $X_n \overset{a.s.}{\to} X$, for some random variable X, then $X_n \overset{P}{\to} X$.

- The converse of the statement above is false. That is, convergence in probability does not imply almost sure convergence.
- A proof of the theorem above, as well as additional treatment of the connection between almost sure convergence and convergence in probability is found in Resnick (2019, Chatper 6).

- Note: As stated, the weak-law of large numbers (WLLN) can actually be shown to hold a.s., in which case we call it the strong-law of large numbers (SLLN).

$X_1, X_2, \ldots$ independent, $E[X_i] = \mu$

$$\bar{X}_n \xrightarrow{P} \mu. \qquad \text{(WLLN)}$$

$$\boxed{\bar{X}_n \xrightarrow{a.s.} \mu} \qquad \text{(SLLN)}$$

- The final form of convergence we will consider in this course is convergence in distribution.

Definition: Convergence in Distribution

A sequence of random variables $X_1, X_2, \ldots$ converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous.

- One way to think about convergence in distribution is that it's really a statement about the long-run behavior of a sequence of random variables, as it's a statement about the CDFs.

Convergence in Distribution II

- This is different from the other types of convergence, which are concerned with the random variable itself.
- A quick recap of how the different types of convergence are related:
  - a.s. convergence $\implies$ convergence in prob $\implies$ convergence in Distribution.
- In a *few* special scenarios, we can talk about more connections between the types of convergence.
- One such example is convergence in probability to a constant. Casella and Berger (Theorem 5.5.13 of 2024) shows that $X_n \xrightarrow{P} a$ for some constant a if and only if $X_n \xrightarrow{d} a$.

$$X_n \xrightarrow{P} a \iff X_n \xrightarrow{d} a$$

The Central Limit Theorem

3350, 3352

- Next we are going to introduce the Central Limit Theorem (CLT).
- The CLT is easily one of the most important theorems across all scientific disciplines, and arguably the most important result to modern science.

  "I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the [CLT]. The law would have been personified by the Greeks and deified, if they had known of it..." - Sir Francis Galton

The Central Limit Theorem II

- The theory for the CLT was developed over a period of roughly 100 years, done by some of the greatest mathematicians of the 19h and 20th centuries.
- The theorem states that, under very weak conditions, the sum of any sequence of iid random variables (with finite mean and variance) converges to a normal distribution.
- Here, we are going to work towards a proof of a simple (weak) version of the theorem.

- Collect data $\underbrace{x_1, x_2, \dots}$
- use the data to make inference about something.

- $x_1, x_2, \dots$ blood pressure $\underline{n}$ individuals.

$\bar{X}_n$ = average blood pressure

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \sim N(0, \sigma^2)$$

CLT.

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

true average blood

Review:

- Convergence in Prob. $X_n \xrightarrow{P} X$.

$\rightarrow \lim_{n\to\infty} P(|X_n - X| > \frac{1}{3}) = 0$

$\{P(|X_n - X| < \varepsilon) = 1\}$

- Almost sure (probability 1) $\leftarrow$ 2

$$\underbrace{P(\lim_{n\to\infty} |X_n - X| < \varepsilon)}_{\omega: \ |X_n(\omega) - X(\omega)| < \varepsilon} = 1$$

$X \sim N(0,1)$

$x = 2, 1, 0$

- Convergence in Dist'n:

$X_n \xrightarrow{D} X$ if $F_n(t) \to F(t)$ $\forall t$.

F_n ↑ X_n, F ↑ X

- a.s. $\Longrightarrow$ prob $\Longrightarrow$ Dist'n

$\to$ if $X_n \xrightarrow{d} a \Longrightarrow X_n \xrightarrow{P} a$

(prob $\Longleftrightarrow$ Dist'n)

- WLLN: $\bar{X}_n \xrightarrow{P} E[X_i]$, (iid) $\sigma^2 < \infty$

- SLLN: $\bar{X}_n \xrightarrow{a.s.} E[X_i]$

$F_n \to F$.

If mgf ~~exits~~ exists for all n, denote it $M_n(t)$,

if $\underset{\substack{(\downarrow)\\ X_n}}{M_n(t)} \to \underset{\substack{(\downarrow)\\ X}}{M(t)} \Rightarrow F_n(t) \to F(t) \leftarrow$

Theorem: Continuity Theorem

Let X_n be a sequence of random variables with cdf $F_n(x)$, and let X be a random variable with cdf $F(x)$. Furthermore, let $M_n(t)$ be the moment generating function of X_n, and $M(t)$ the moment generating function of X.

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 of F. That is, $X_n \xrightarrow{d} X$.

- Now, we do a brief reminder about Taylor Series and Taylor's Thereom

$X_n \xrightarrow{d} X$ iff $F_n(x) \to F(x)$

$M_n(t)$ uniquely determines $F_n(x)$

The Central Limit Theorem IV

Theorem: Taylor Series

If a function $f(x)$ has derivatives of order k, that is, $\frac{d^k}{dx^k} f(x)$ exists, then for any constant a, the Taylor Polynomial of order k, centered about a, is

$$f(x) = \sum_{n=0}^{k} \frac{f^{(n)}(a)}{n!}(x-a)^n + R_k(x),$$

where $R_k(x) = h_k(x)(x-a)^k$, for some h_k such that $\lim_{x\to a} h_k(x) = 0$.

$$\lim_{x\to a} h_k(x) = \lim_{x\to a} \frac{R_k(x)}{(x-a)^k} = 0$$

- In particular, it means that we can use a k order polynomial to approximate a differentiable function, and the remainder term $R_k(x)$ goes to zero at a rate smaller than the rate that $(x-a)^k$ goes to zero.

Theorem: The (classic) Central Limit Theorem

Let $X_1, X_2, \ldots$ be a sequence of independent and identical random variables with mean $E[X_i] = \mu$ and variance $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Assume that the mgf of X_i exists and is defined in a neighborood of zero, and denote the cdf and mgf as F and M, respectively. Then,

$$\sqrt{n}\big(\bar{X}_n - \mu\big) \xrightarrow{d} N(0, \sigma^2).$$

proof:

$Y = bX$, $\quad M_Y(t) = M_{bX}(t) = M(bt)$

$Y = \sum X_i$, $\quad M_Y(t) = M_{\sum X_i}(t) = \underbrace{\prod_{i=1}^{n} M_{X_i}(t)}_{\text{if independent.}}$

$\left.\right\}$ ch 4.

- $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$

- mgf of $\sqrt{n}(\bar{X}_n - \mu) \rightarrow$ mgf $N(0, \sigma^2)$

- let $Z_i = \dfrac{X_i - \mu}{\sigma}$.

$E[Z_i] = \frac{1}{\sigma} \underbrace{E[X_i - \mu]}_{} = 0$

$var(Z_i) = Var\left(\frac{X_i}{\sigma} - \frac{\mu}{\sigma}\right)$

$= \frac{1}{\sigma^2} Var(X_i) = 1$

$var(Z_i) = E[Z_i^2] - (\cancel{E[Z_i]})^2$

$1 = 1$

- $\sqrt{n}(\bar{X}_n - \mu) = \sqrt{n}\left(\frac{1}{n}\sum X_i - \mu\right)$

$$= \sqrt{n}\left(\frac{1}{n}\sum X_i - \frac{1}{n}\sum \mu\right)$$

$$= \sqrt{n}\left(\frac{1}{n}\sum (X_i - \mu)\right),$$

divide by σ,

$$\star \quad \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sqrt{n}}{n}\sum \frac{(X_i - \mu)}{\sigma} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i$$

$$M_{\sqrt{n}(\bar{X}_n - \mu)/\sigma}(t) = M_{\frac{1}{\sqrt{n}}\sum Z_i}(t)$$

$$= M_{\sum Z_i}(t/\sqrt{n})$$

$$= \prod_{i=1}^{n} M_{z_i}(t/\sqrt{n}) , \qquad z_i = \frac{x_i - \mu}{\sigma}.$$

X_i are iid.

$$E[e^{0x}] = 1 \qquad = \left(M_{z_i}(t/\sqrt{n}) \right)^n \qquad E[z] = 0$$

$$M_z(t/\sqrt{n}) = M_z(0) + M_z^{(1)}(0) \cdot \frac{t/\sqrt{n}}{1!}$$

$$+ M_z^{(2)}(0) \frac{(t/\sqrt{n})^2}{2!} + R_2(t/\sqrt{n}) \qquad E[z^2]$$

$$= 1 + \frac{t^2/n}{2} + R_2(t/\sqrt{n})$$

$$M_{\sqrt{n}(\bar{X}_n-\mu)/\sigma}(t) = \left(1 + \frac{t^2}{2n} + R_2(t/\sqrt{n})\right)^n \rightarrow e^{t/2}$$

★ next slide.

From Taylor's theorem,

$$\lim_{t\to 0} \frac{R_2(t/\sqrt{n})}{(t/\sqrt{n})^2} = 0$$

$n\to\infty$, $t/\sqrt{n} \to 0$

$$\lim_{n\to\infty} \frac{R_2(t/\sqrt{n})}{(t/\sqrt{n})^2} = 0 \Rightarrow$$

$$\lim_{n\to\infty} \frac{R_2(t/\sqrt{n})}{(1/\sqrt{n})^2} = \lim_{n\to\infty} nR_2(t/\sqrt{n}) = 0$$

$$M_{\sqrt{n}(\bar{x}-\mu)/\sigma}(t) = \left(1+\frac{t^2}{2n}+R_2(t/\sqrt{n})\right)^n$$

$$= \left(1+\frac{1}{n}\underbrace{\left(t^2/2+nR_2(t/\sqrt{n})\right)}_{a_n = t^2/2 + \boxed{nR_2(t/\sqrt{n})} \to 0}\right)^n$$

$$a_n \to t^2/2$$

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

$\lim_{n\to\infty} a_n = a$, then $\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n = e^a$

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$$

$$M_{\sqrt{n}(\bar{X}_n - \mu)/\sigma}(t) = \left(1 + \frac{a_n}{n}\right)^n$$

$$\lim_{n \to \infty} M_{\sqrt{n}(\bar{X}_n - \mu)/\sigma}(t) = \lim_{n \to \infty} \left(1 + \frac{a_n}{n}\right)^n = e^{t^2/2}$$

$$= M_Z(t)$$

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$$

$$\boxed{\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)}$$

$\square$

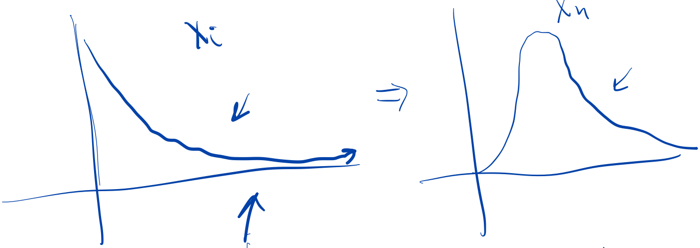
The Central Limit Theorem VI

- One practical implication of the CLT is that, for large n, we can approximate
$$\bar{X}_n \overset{d}{\approx} N\big(\mu, \sigma^2/n\big),$$
if X_i are independent and identically distributed with finite mean and variance.
- In practice, $n \approx 30$ has been found to lead to good approximations, but it depends heavily on the distribution of X_i.
- A further investigation of the CLT proof shows that the convergence towards the normal distribution happens at a rate of $\boxed{1/\sqrt{n}.}$

X_i are normally dist'd,

$\sum X_i$ is normal, $\frac{1}{n}\sum X_i$ is normal.

$n=20$



$n=30$

no extreme outliers.

- Example from 3350.

Q: Do mothers who live in poverty give birth to babies that are smaller than average?

→ collect data from sample of babies, see if they are smaller on average.

→ are differences due to random chance, or is there something there.

- National average: $\mu = 120$ oz, $\sqrt{var(X_i)} = 24$ oz.

X_i

$E[X_i] \approx \mu, \quad var(X_i) \approx \sigma^2$

- $Y_1, Y_2, \dots, Y_n$ from mothers in poverty,

$n = 100 \qquad \overline{Y}_n = 115.8$ oz

- let's assume $E[Y_i] = E[X_i] = 120$.

- If this is true, what's the prob. of observing $\overline{Y}_n = 115.8$ (or smaller)

$$\overline{Y}_n \approx N\left(120, \frac{(24)^2}{100}\right) \leftarrow$$

$P(\bar{Y}_n \leq 115.8) = 0.02$] ← P-value

→ conclude that assumption $E[Y_i] = E[X_i]$ is false.

$\alpha = 0.05$, we reject if $\boxed{\text{P-value} < 0.05}$

$$\frac{\bar{X} - \mu}{(\sigma / \sqrt{n})} \sim N \qquad \frac{\bar{X} - \mu}{(S / \sqrt{n})}$$

The Central Limit Theorem VII

- If we used a Taylor-series approximation with one additional order, we could derive a more accurate approximation under additional conditions known as the Edgeworth Expansion (See Theorem 19.3 of Keener, 2010). These are less commonly used in practice, because you need finite third moments.

Example: Binomial-Normal Approximation

Let $X \sim \text{Binomial}(n, p)$. For any $k \in \{0, 1, \ldots\}$, approximate $P(X \leq k)$.

- $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $X_1, X_2, \ldots, X_n \sim \text{Ber}(p)$

$$\sum X_i \overset{d}{=} X$$

$\bar{X} \stackrel{d}{=} \sum_i X_i$, $E[X_i] = p$, $Var(X_i) = p(1-p)$

$\bar{X}_n = \frac{1}{n} \sum X_i$. by CLT :

$$\sqrt{n}\left(\bar{X}_n - p\right) \approx N(0, p(1-p))$$

$$\bar{X}_n \approx N\left(p, \frac{p(1-p)}{n}\right)$$

$$X = n\bar{X}_n \approx N(np, np(1-p))$$

So, for any k, cdf $N(\mu, \sigma^2)$: $\Phi\left(\frac{x-\mu}{\sigma}\right)$

$$P(X \le k) = \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)$$

$$P(X \le k) \ne P(X < k)$$

$$P(Y \le k) = P(Y < k)$$

$$P(X \le k) = P(X \le k + 1/2) = P(X \le k + 3/4) \ldots$$

$$\Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right) \ne \boxed{\Phi\left(\frac{k+1/2-np}{\sqrt{np(1-p)}}\right)} \ne \ldots$$

$$P(X \le k) \approx P(Y \le k+0.5) = \Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right)$$

Slutsky's Theorem

- The following theorem is useful for our notes and supporting other ideas we will cover. However, we won't discuss the proof because it relies on other convergence concepts we don't cover in this class

Theorem: Slutsky's Theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$ for some random variable X and constant a, then

1. $Y_n X_n \xrightarrow{d} aX$.
2. $X_n + Y_n \xrightarrow{d} X + a$.

Slutsky's Theorem II

Example: CLT with estimated variance

(HW problem)? Suppose that X_i are iid $N(\mu, \sigma^2)$ random variables. By the CLT, we have

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \xrightarrow{d} N(0,1).$$

The problem with this theorem in practice is that we assume σ is known, which is often not practical. If S_n^2 is our estimate of the variance, and $S_n^2 \xrightarrow{p} \sigma^2$, then it can be shown that $\sigma / S_n \xrightarrow{p} 1$. Thus, by Slutsky's Threom:

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} = \frac{\sigma}{S_n} \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \xrightarrow{d} N(0,1)$$

Delta-Method

- The CLT is useful for determining the limiting distribution of a random variable (particularly, sums of iid random variables).
- As we have already discussed, we are often interested in functions of random variables.
- This next section gives theorems for the limiting distribution of functions of random variables.

Delta-Method II

Theorem: The Delta-Method

Let X_n be a sequence of random variables that satisfy

$$\sqrt{n}\big(X_n - \theta\big) \xrightarrow{d} N(0, \sigma^2),$$

where $\theta, \sigma^2 < \infty$. Now suppose that g is a function such that it's first derivative g' exists and $g'(\theta) \neq 0$. Then,

$$\sqrt{n}\big(g(X_n) - g(\theta)\big) \xrightarrow{d} N\big(0, \sigma^2[g'(\theta)]^2\big).$$

Delta-Method III

Theorem: Second Order Delta Method

Let X_n be a sequence of random variables that satisfies $\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma^2)$. For a given function g such that the first two derivatives of g exist and $g'(x) = 0$, $g''(x) \neq 0$, then

$$n\big(g(X_n) - g(\theta)\big) \xrightarrow{d} \sigma^2 \frac{g''(\theta)}{2} \chi_1^2,$$

where χ_1^2 is a "Chi-square" distribution with one degree of freedom.

References and Acknowledgements

Casella G, Berger R (2024). *Statistical inference*. Chapman and Hall/CRC.

Keener RW (2010). *Theoretical statistics: Topics for a core course*. Springer Science & Business Media.

Resnick S (2019). *A probability path*. Springer.

Rice JA (2007). *Mathematical statistics and data analysis*, volume 371. 3 edition. Thomson/Brooks/Cole Belmont, CA.

- Compiled on November 18, 2025 using R version 4.5.2.

References and Acknowledgements II

- We acknowledge students and instructors for previous versions of this course / slides.