STAT 346/446 Lecture 4

Central Limit Theorem, Slutsky's Theorem, and the delta method

+ continuous functions and $\xrightarrow{P}$

+ WLLN

CB Sections 5.5.3 and 5.5.4
DS Section 6.3

Central Limit Thorem (CLT)

Central Limit Theorem

Let $X_1, X_2, X_3, \ldots$ be a sequence of random variables where

- $X_1, X_2, X_3, \ldots$ are iid.
- $M_X(t)$ exists (for some t in a neighborhood of 0)

Let $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 > 0$ for all i and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then

$$\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} Z \qquad \text{where } Z \sim \mathrm{N}(0, 1)$$

Some notes on the CLT

- $X_1, X_2, X_3, \ldots$ can come from *any* distribution - with some minor conditions
- Here the condition is $M_X(t)$ exists
  - This implies that both $E(X_i)$ and $\mathrm{Var}(X_i)$ are finite
- Stronger versions of CLT: Existence of $M_X(t)$ is not necessary, but do need finite variance
  - Proof without mgfs is outside the scope of this course
- CLT is the basis for normal approximation of so many things!
- How good is the CLT approximation?
  - The CLT alone can't tell us that
  - Accuracy of the approximation depends on the actual distribution of $X_1, X_2, X_3, \ldots$

Helpful facts for proof of CLT

- Some rules for the moment generating function (mgf)

$$\frac{d^n}{dt^n}M(t)\bigg|_{t=0} = E(X^n)$$

$$M_{aX+b}(t) = e^b M_X(at)$$

$$M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t)$$

  if $X_1, X_2, \ldots, X_n$ are independent.

- If $a_1, a_2, a_3, \ldots$ is a sequence of numbers such that $\lim_{n\to\infty} a_n = a$ then

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

Helpful facts for proof of CLT

- Taylor series expansion of a function $f(x)$ around a constant a:

$$f(x) = \sum_{i=1}^{\infty} \frac{(x-a)^i}{i!} \left. \frac{d^i}{dx^i} f(x) \right|_{x=a}$$

- Define

$$\left. \frac{d^i}{dx^i} f(x) \right|_{x=a} = f^{(i)}(a)$$

- First terms of a Taylor expansion of $f(x)$ around a

$$f(x) = f(a) + (x-a)f^{(1)}(a) + \frac{(x-a)^2}{2} f^{(2)}(a) + R(x)$$

where $\lim_{x \to a} \frac{R(x)}{(x-a)^2} = 0$

Proof of CLT

Done on the board...

Example: Normal approx to the Binomial distribution

- Let $X_1, X_2, X_3, \ldots$ be iid. $\mathrm{Bernoulli}(p)$ then

$$E(X_i) = p \quad \text{and} \quad V(X_i) = p(1-p)$$

  and $Y = n\overline{X}_n \sim \mathrm{Binomial}(n, p)$
- CLT says

$$\frac{\sqrt{n}\,(\overline{X}_n - p)}{\sqrt{p\,(1-p)}} \xrightarrow{d} N(0,1)$$

  For large n we can use $N(0,1)$ as approximation for the distribution of

$$\frac{\sqrt{n}}{\sqrt{n}}\frac{\sqrt{n}\,(\overline{X}_n - p)}{\sqrt{p\,(1-p)}} = \frac{n\overline{X}_n - np}{\sqrt{np(1-p)}} = \frac{Y - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

  Or: Use $N(np, np(1-p))$ as approx. for $\mathrm{Bin}(n,p)$ for fixed p. (and n large)

Example: Normal approx to the Binomial distribution

- Say $Y \sim \text{Binomial}(400, 0.3)$ and we want to calculate

$$P(Y \leq 100) = \sum_{y=0}^{100} \binom{400}{y} 0.3^y \, 0.7^{400-y}$$

($= 0.01553$ using exact calculations in R)

- Normal approximation:

$$P(Y \leq 100) \approx P\left(Z \leq \frac{100 - 400 * 0.3}{\sqrt{400 * 0.3 * 0.7}}\right)$$

$$= \Phi(-2.1822) = 0.01455$$

Slutsky's Theorem

Slutsky's Theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$, where a is a constant, then

(a) $X_n Y_n \xrightarrow{d} aX$

(b) $X_n + Y_n \xrightarrow{d} X + a$

- Proof if outside the scope of this course
- Many of our approximate inference procedures actually rely on the the CLT + Slutsky

(and WLLN: $\bar{X}_n \xrightarrow{P} \mu$)

Example

$s_n = \sqrt{s_n^2}$ $\quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

- Show that

abuse of notation, means

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \xrightarrow{d} \mathrm{N}(0,1) \qquad \xrightarrow{d} Z$$

where $Z \sim N(0,1)$

meaning that

$$\overline{x} \pm z_{\alpha/2} s/\sqrt{n}$$

can be used as an approximate $100(1-\alpha)$ confidence interval for the population mean for any distribution

Example: approximate CI for p

- Let $X_1, X_2, X_3, \ldots$ be iid. $\mathrm{Bernoulli}(p)$. Show that

$E(X_i) = p$

$V(X_i) = p(1-p)$

$$\frac{\overline{X}_n - p}{\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}} \xrightarrow{d} \mathrm{N}(0,1)$$

$\overline{X}_n = \hat{p}_n$ i.e. the sample proportion

justifying our usual approximate $100(1-\alpha)$ confidence interval for a population proportion

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p} - p \overset{\text{aprox}}{\sim} N\left(0, \frac{p(1-p)}{n}\right) \leftarrow$ plug in $\hat{p}$ for p to get the standard error

Here we justify this approach. $\leftarrow$

Example

- Let $X_1, X_2, X_3, \ldots$ be iid. $\mathrm{Gamma}(\alpha, 1)$. Show that

$$\frac{\sqrt{n}(\overline{X}_n - \alpha)}{\sqrt{\overline{X}_n}} \xrightarrow{d} \mathrm{N}(0, 1)$$

Delta method

- We have a handle on the limiting distribution of $\overline{X}_n$ via the CLT:

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

$$\text{equivalently: } \sqrt{n}\,(\overline{X}_n - \mu) \xrightarrow{d} \mathrm{N}(0,\sigma^2)$$

- What is the limiting distribution of $g(\overline{X}_n)$?
    - Usually also Normal, but need to determine the mean and variance
    - Can approximate mean and variance of $g(\overline{X}_n)$ via Taylor expansion of $g(\cdot)$

Example: Sample odds

- Let $X_1, X_2, X_3, \ldots$ be iid. $\mathrm{Bernoulli}(p)$ and $\overline{X}_n = \hat{p}$.
- $\mathrm{odds} = \frac{p}{1-p}$
- What is the limiting distribution of the sample odds? $\widehat{\mathrm{odds}} = \frac{\hat{p}}{1-\hat{p}}$

$\triangleright$ a natural estimator of odds is $\dfrac{\hat{p}_n}{1-\hat{p}_n} = \dfrac{\overline{X}_n}{1-\overline{X}_n}$

Approx. mean and variance via Taylor expansion

- Let X be a random variable with mean μ and variance σ^2
- Let $g(x)$ be a differentiable function
- First order Taylor expansion of $g(x)$ around μ:

leaving out $R(x)$

$$g'(\mu) = \frac{d}{dx} g(x) \Big|_{x=\mu}$$

$$g(x) \approx g(\mu) + g'(\mu)\,(x-\mu)$$

$$\Rightarrow E(g(X)) \approx E\big(g(\mu) + g'(\mu)(X-\mu)\big)$$
$$= g(\mu) + g'(\mu)\big(\underbrace{E(X)}_{\mu} - \mu\big) = g(\mu)$$

$$\text{and } V(g(X)) \approx V\big(g(\mu) + g'(\mu)(X-\mu)\big)$$
$$= (g'(\mu))^2 V(X) = (g'(\mu))^2 \sigma^2$$

So, e.g $g(x) = x^2$ $g'(x) = 2x$

$E(X^2) \approx \mu^2$ but not $=$

i.e. $E(X^2) \neq \mu^2$

and $V(X^2) \approx (2\mu)^2 \sigma^2$

Example: Sample odds

$E(\bar{X}_n) = p$
$V(\bar{X}_n) = \frac{p(1-p)}{n}$

- Let $X_1, X_2, X_3, \ldots$ be iid. $\mathrm{Bernoulli}(p)$ and $\overline{X}_n = \hat{p}$.
- Let $\widehat{\mathrm{odds}} = \frac{\hat{p}}{1-\hat{p}}$

$g(x) = \frac{x}{1-x}$

$$E(\widehat{\mathrm{odds}}) \approx \frac{p}{1-p}$$

$$\frac{d}{dx} g(x) = \frac{(1-x) - x(-1)}{(1-x)^2} = \frac{1-x+x}{(1-x)^2} = \frac{1}{(1-x)^2}$$

and $V(\widehat{\mathrm{odds}}) \approx \left(\frac{1}{(1-p)^2}\right)^2 Var(\hat{p})$

$$= \frac{1}{(1-p)^4} \frac{p(1-p)}{n}$$

$$= \frac{p}{n(1-p)^3}$$

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

Delta method

Theorem: Delta Method

Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of random variables where

usually $Y_n = \bar{X}_n$

$$\sqrt{n}(Y_n - \theta) \xrightarrow{d} \mathrm{N}(0, \sigma^2) \quad \text{e.g. from CLT}$$

and θ is a constant. Let $g(y)$ be a function where $g'(\theta)$ exists and is not zero. Then

$$\sqrt{n}\,(g(Y_n) - g(\theta)) \xrightarrow{d} \mathrm{N}(0, \underbrace{\sigma^2 g'(\theta)^2}_{\text{approx var}})$$

Proof... Taylor expansion and Slutsky - see textbook

Example: Sample odds

$\text{Var}(\widehat{\text{odds}}) \approx \frac{p}{n(1-p)^3}$

- Let $X_1, X_2, X_3, \ldots$ be iid. $\mathrm{Bernoulli}(p)$ and $\overline{X}_n = \hat{p}$.
- $\mathrm{odds} = \frac{p}{1-p}$
- What is the limiting distribution of the sample odds? $\widehat{\mathrm{odds}} = \frac{\hat{p}}{1-\hat{p}}$

$\sqrt{n}(\text{odd} - \text{stuff}) \xrightarrow{d}$

CLT: $\sqrt{n}(\overline{X}_n - p) \xrightarrow{d} N(0, p(1-p))$

delta method: $g(x) = \frac{x}{1-x}$ $\quad g'(x) = \frac{1}{(1-x)^2}$

$$\sqrt{n}\left(\frac{\overline{X}_n}{1-\overline{X}_n} - \frac{p}{1-p}\right) \xrightarrow{d} N\left(0, \frac{p}{(1-p)^3}\right)$$

$$\sigma^2(g'(\theta))^2 = p(1-p)\frac{1}{(1-p)^4} = \frac{p}{(1-p)^3}$$

By slutsky: can get approx conf. int for odds:

$$\frac{\hat{p}}{1-\hat{p}} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}}{(1-\hat{p})^3 n}}$$

Second order Delta method

Theorem: Second order Delta Method

Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of random variables where

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

and θ is a constant. Let $g(y)$ be a function where $g'(\theta) = 0$, but $g''(\theta)$ exists and is not zero. Then

$$\sqrt{n}\,(g(Y_n) - g(\theta)) \xrightarrow{d} \sigma^2 \frac{g''(\theta)}{2} X$$

chi-sq distr.

where $X \sim \chi^2_1$

- Further extension: Multivariate Delta Method (skip)