Dr. Franzke

EE 364 - Spring 2026

Week 14

Supplemental Notes

Topics

① Central Limit Theorem (CLT)

- $Z_n \xrightarrow{d} Z \sim N(0,1)$ random sample
- $(1-\alpha)$ % C.I. for μ_x: $\bar{X}_n \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$$Z_n = \text{STD}(\bar{X}_n) = \text{STD}\left(\sum_{k=1}^{n} X_k\right)$$

CLT Binomial Approximation

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$\therefore X = \sum_{k=1}^{n} X_k$ independent Bernoulli X_k

$f_{X_k}(x) = p^x (1-p)^{1-x}$ for $x = 0$ or 1

$$X = \sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{CLT} \xrightarrow{d} N\left(\underbrace{np}_{\mu_x}, \underbrace{n \cdot pq}_{\sigma_x^2}\right)$$

For large n: $b(n,k,p) \approx N(np, n \cdot pq)$ if $q = 1-p$.

Rule of thumb: (1) $np \geq 5$ and (2) $nq \geq 5$

Ex: $X \sim b(10, \frac{1}{2})$

$\therefore n = 10 \quad p = \frac{1}{2} \quad \therefore np = n(1-p) = 10 \cdot \frac{1}{2} = 5.$

$\therefore$ OK to use CLT.
(but NOT Poisson Law)

$$\therefore P(3 \le X < 6) \overset{\text{continuity correction}}{\approx} P(2.5 \le X \le 5.5)$$

$$\overset{\text{standardise}}{=} P\left(\frac{2.5-np}{\sqrt{npq}} \le \frac{X-np}{\sqrt{npq}} \le \frac{5.5-np}{\sqrt{npq}}\right)$$

$$= P\left(\frac{2.5-5}{\sqrt{10/4}} \le \frac{X-5}{\sqrt{10/4}} \le \frac{5.5-5}{\sqrt{10/4}}\right)$$

$$\overset{\text{CLT}}{\approx} P(-1.58 \le Z \le 0.3162) \quad \text{for } Z \sim N(0,1)$$

$$= \phi(0.316) - \phi(-1.581)$$

$$\approx 0.5684.$$

Exact: $P(3 \le X < 6) = 0.5683$ if $X \sim b(10, \frac{1}{2})$

$\therefore |0.5684 - 0.5683| = 0.0001$

Continuity correction:

for discrete, $P(X \le x) = P(X < x+1)$

for continuous, $P(X \le x) = P(X < x)$

$\therefore$ discrete $\longrightarrow$ continuous, evaluate midpoint $\approx P\left(X \le x + \frac{1}{2}\right)$ continuity correction.

Note: $P(3 \le X \le 6) \approx P(2.5 \le X \le 5.5)$

$P(3 \le X \le 6) \approx P(2.5 \le X \le 6.5)$

$\ge, \le$ wider

$>, <$ shrink

Ex: Airline overbooking

$P(\text{no show}) = 0.04$ 126 seats but 130 bookings

Q: What is probability that more than 126 passengers show up?

(i.e. 3 or fewer no-shows)

X: # no show What is $P[X \le 3]$ i.e. X: 0, 1, 2, or 3

$n = 130$ and $p = 0.04$ ← large sample requirement

$\therefore np = 5.2$

$nq = 124.8$

$\therefore$ CLT approximation OK.

$$\therefore \sqrt{npq} = \sqrt{(130)(0.04)(0.96)} = 2.243$$

$$P[X \le 3] \overset{\text{cont. corr.}}{\le} P[X \le 3.5] = P\left[\frac{X - np}{\sqrt{npq}} \le \frac{3.5 - np}{\sqrt{npq}}\right]$$

$$= P\left[\frac{X - 5.2}{2.2343} \le \frac{3.5 - 5.2}{2.2343}\right]$$

$$= P[Z \le -0.76]$$

↓ z-table

$$= 0.2236 \quad (\approx 22\%)$$

Compare

Exact: $p = \sum_{k=0}^{3} \binom{130}{k} (0.04)^k (0.96)^{130-k} = 0.2323 \approx 23\%$ (close)

Poisson approximation: $p = 0.04 \ll 1$, $n > 100$ $\therefore \underbrace{np}_{\lambda} = 5.2$

$$P(X \le 3) \approx \sum_{k=0}^{3} \frac{e^{-5.2}(5.2)^k}{k!} = 0.23 \quad (\approx 23\%)$$

Confidence Intervals

2-sided for parameter θ:

$$\hat{\theta}_n \pm \text{M.O.E.}$$

M.O.E. = Margin of Error

usually involves $\sigma_{\hat{\theta}_n} = \sqrt{V[\hat{\theta}_n]}$ "standard error"

or its estimate $S_{\hat{\theta}_n}$

$\hat{\theta}_n$ usually involves $\bar{X}_n$, S_n^2, or S_{XY} (random sets $X: \Omega \to \mathcal{B}$).

Less confident ↓ bigger MOE

Ex: $(1-\alpha)\%$ C.I. for μ_x if σ_x^2 known and $\alpha \in [0,1]$

random sample and Large $n \geq 30$ CLT

$$\text{C.I.} = \bar{X}_n \pm \text{M.O.E.}$$

$$= \bar{X}_n \pm z_{\alpha/2} \frac{\sigma_x}{\sqrt{n}}$$

$\frac{\sigma_x}{\sqrt{n}} = \sqrt{V[\bar{X}_n]}$ standard error

unknown σ_x^2:

$$\text{C.I.} = \bar{X}_n \pm t_{\alpha/2}(n-1) \frac{S_n}{\sqrt{n}}$$

S_n estimates σ_x

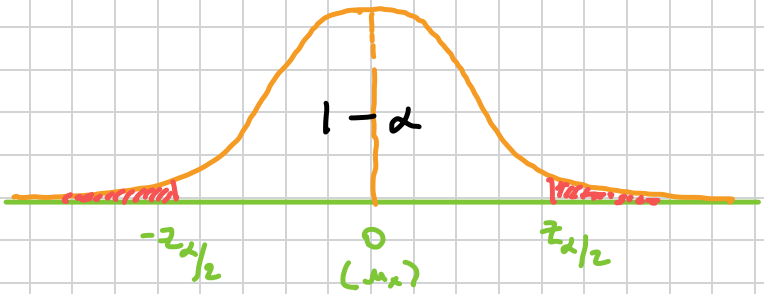
↑ wider C.I.

Thrm: $(1-\alpha)$% C.I. for μ_x if σ_x^2 known and large n:

(r.s.) → IID, $\sigma_x^2 < \infty$

$$\boxed{\bar{X}_n \pm z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}}$$ memorize

Prf:



$$Z_n \overset{D}{=} \frac{\bar{X}_n - \mu_x}{\sigma_x/\sqrt{n}}$$ since r.s.

$$\overset{d}{\approx}_{CLT} Z \sim N(0,1)$$

$$\therefore \bar{X}_n \overset{d}{\approx} N\left(\mu_x, \frac{\sigma_x^2}{n}\right)$$

$$\therefore 1-\alpha \overset{F.T.C.}{=} \boxed{P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right)}$$

constant, i.e. $\frac{\alpha}{2}$: prob $\ge z_{\alpha/2}$ (symmetric interval)

$$\overset{CLT}{\approx} P\left(-z_{\alpha/2} \le \frac{\bar{X}_n - \mu_x}{\sigma_x/\sqrt{n}} \le z_{\alpha/2}\right)$$ since n "large"

$$= P\left(\left\{\omega \in \Omega : -z_{\alpha/2} \le \frac{\bar{X}_n(\omega) - \mu_x}{\sigma_x/\sqrt{n}} \le z_{\alpha/2}\right\}\right)$$

$$= P\left(-z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}} \le \bar{X}_n - \mu_x \le z_{\alpha/2}\cdot\frac{\sigma_x}{\sqrt{n}}\right)$$

$$= P\left(\underbrace{-\bar{X}_n}_{\text{r.var}} - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}} \le \underbrace{-\mu_x}_{\text{constant}} \le \underbrace{-\bar{X}_n}_{\text{r.var}} + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}\right)$$

$$= P\left(\bar{X}_n + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}} \ge \mu_x \ge \bar{X}_n - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}\right)$$

$$= P\left(\bar{X}_n - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}} \le \mu_x \le \bar{X}_n + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}\right)$$

$\therefore \bar{X}_n \pm z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$ is the $(1-\alpha)$% CI for μ_x QED.

Picking Sample Size n

Want: interval no wider than $\bar{X}_n \pm \varepsilon$ in above case

ε = Maximal error tolerance = M.O.E. = $z_{\alpha/2} \frac{\sigma_x}{\sqrt{n}}$

$\therefore \varepsilon^2 = z_{\alpha/2}^2 \cdot \frac{\sigma_x^2}{n}$

$\therefore$ $\boxed{n = \frac{z_{\alpha/2}^2 \sigma_x^2}{\varepsilon^2}}$ memorize

Proportions: Estimate with CLT:

Q: estimate $P[\text{vote candidate}]$

$X \sim b(n, p)$ ← success/failure structure

proportion $\hat{p} \triangleq \frac{X}{n}$ ← # success / ← # trials

note: $\hat{q} = 1 - \hat{p}$

estimate for $P[\text{success}]$

(MLE of p since $X \sim b(n,p)$)

① $E[\hat{p}] = \frac{E[X]}{n} \overset{X \sim b}{=} \frac{n \cdot p}{n} = p$ $\forall n$ ∴ Unbiased

② $V[\hat{p}] = \frac{V[X]}{n^2} = \frac{n \cdot p \cdot q}{n^2} = \frac{p \cdot q}{n} \downarrow 0$ as $n \uparrow \infty$

standard error $\sigma_{\hat{\Theta}_n} = \frac{\sqrt{pq}}{\sqrt{n}}$

$MSE = V + B^2$

∴ $\hat{p}_n \xrightarrow{p} p$ by m.p.d. ∴ $\hat{p}_n$ is consistent for p.

∴ For large n: CLT $\hat{p}_n \overset{d}{\approx} N\left(p, \frac{p(1-p)}{n}\right)$

(like sample-mean)

∴ $(1-\alpha)\%$ C.I. for p:

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

Ex: Polls. (report) "$\pm 3\%$ M.O.E." in the news poll

estimate $\longrightarrow$ $P[\text{vote candidate}]$

implied $\varepsilon = 0.03$ at 95% C.I. $(\alpha = 0.05)$

problem: width depends on estimate $\hat{p}$
so take "worst case" (widest interval)

Note: $p(1-p) \leq \frac{1}{4}$ (since $p \in (0,1)$) $\therefore p = \frac{1}{2}$ maximizes $f(p) = p - p^2$

$\therefore V[\hat{p}] = \frac{p \cdot q}{n} \leq \frac{1}{4n}$

$\therefore$ Conservative estimate: $\sigma_{\hat{p}} \leq \frac{1}{2\sqrt{n}}$

$\therefore \varepsilon = \text{M.O.E.} = \frac{z_{\alpha/2}}{2\sqrt{n}}$

$\downarrow$ solve for n

$\therefore$ $\boxed{n = z_{\alpha/2}^2 / 4\varepsilon}$ pollster's sample size

$\therefore$ For 95% confidence Need $n = 1068$ samples

"$1-\alpha$", $\therefore \alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

For 99% confidence $n = 1849$ samples

$\therefore \alpha = 0.005$