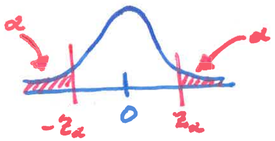
§7.2: Large Sample Confidence Intervals

Recall: $z_\alpha = \text{qnorm}(\alpha, 0, 1)$

is "# std deviations away from mean for Prob $= \alpha$"

→ Textbook doesn't like z_α to be negative so it uses $z_\alpha = -\text{qnorm}(\alpha, 0, 1)$

$P(Z > z_\alpha) = \alpha$



Basic plan for confidence intervals for θ:

$$\theta = \hat{\theta} \pm z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$$

Point estimate for θ

standard error of $\hat{\theta}$

(Pretending that $\hat{\theta}$ is Normal)

Previously: If we know $\sigma^2 = \text{Var}[X]$ then $\text{Var}[\bar{X}] = \sigma^2/n$ so

$$\mu = \bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}.$$

Now: What if we don't know $\text{Var}[X]$?

$\text{Var}[X]$ unknown $\Longrightarrow$ Must use point estimate for it too...

$\boxed{\text{Plan: Replace } \sigma^2 \text{ by } S^2}$

Thm: If n big (like... $n > 40$) then

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \approx \text{Normal}(0, 1)$$

So confidence interval is

$$\mu = \bar{x} \pm z_{\alpha/2} \cdot S/\sqrt{n}$$

!!!

(Seriously... at some point you've got to start worrying that the world is too nice...)

We can apply this to the **Binomial** distribution.

Recall: (§4.3, p166)

$$\text{Binomial}(n, p) \approx \text{Normal}(np, \sqrt{npq})$$

if n big & $np \geq 10$

$nq \geq 10$

$q = 1-p$

X = # occurrences in population of size n

↪ $X \sim \text{Binomial}$

Also $\hat{p} = \frac{X}{n}$ is point estimator for p.

Confidence Interval for $\hat{p}$?

$$E[\hat{p}] = E[X/n] = np/n = p \quad \text{(unbiased)}$$

$$Var[\hat{p}] = Var[X/n] = \frac{1}{n^2}(npq) = pq/n$$

"std. error" $\sigma_{\hat{p}} = \sqrt{pq/n}$

Rough confidence interval:

$$p = \hat{p} \pm z_{\alpha/2} \cdot \sigma_{\hat{p}} = X/n \pm z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$$

$\hat{p} = X/n$, $\hat{q} = 1 - \hat{p} = \frac{n-X}{n}$

Note: Textbook computes a more precise interval which works for all values of p (not just $np \geq 10$)

$$P\left(\frac{|\hat{p}-p|}{\sqrt{pq/n}} < z_{\alpha/2}\right) \approx 1-\alpha$$

2 roots

$$(\hat{p}-p)^2 = \left(z_{\alpha/2}\cdot\sqrt{pq/n}\right)^2$$

(want to solve for p...)

$$\hat{p} - 2\hat{p}p + p^2 = z_{\alpha/2}^2 \frac{p(1-p)}{n}$$

$$\left[\left(1 + z_{\alpha/2}^2/n\right)p - \left(\hat{p} + z_{\alpha/2}^2/2n\right)\right]^2 = z_{\alpha/2}^2\left[\hat{p}(1-\hat{p})/n + z_{\alpha/2}^2/4n^2\right]$$

$$p = \underbrace{\frac{\hat{p} + z_{\alpha/2}^2/2n}{1 + z_{\alpha/2}^2/n}}_{\approx \hat{p}} \pm z_{\alpha/2}\underbrace{\frac{\sqrt{\hat{p}(1-\hat{p})/n + z_{\alpha/2}^2/4n^2}}{1 + z_{\alpha/2}^2/n}}_{\approx \sqrt{\hat{p}\hat{q}/n}}$$

please don't try to memorize this...

(So this can get ugly if you insist on things being precise....)

One-Sided Intervals

If you only want to compute intervals on a single side of $\hat{\theta}$ then you can:

Two-Sided

$$\theta = \hat{\theta} \pm z_{\alpha/2}\cdot\sigma_{\hat{\theta}}$$

One Sided

left: $(\hat{\theta} - z_{\alpha}\cdot\sigma_{\hat{\theta}},\ \hat{\theta})$

right: $(\hat{\theta},\ \hat{\theta} + z_{\alpha}\cdot\sigma_{\hat{\theta}})$