Confidence Intervals

Suppose $X_1, X_2, X_3, \ldots, X_n$ sampled from a distribution with parameter θ.

Point estimator

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n).$$

Question: How close is $\hat{\theta}$ to θ, and some measure of this?

Confidence interval. (Interval estimators).

Goal is to find two estimators

$$\hat{\theta}_L = \hat{\theta}_L(X_1, X_2, \ldots, X_n) \text{ - low estimator}$$

$$\hat{\theta}_H = \hat{\theta}_H(X_1, X_2, \ldots, X_n) \text{ - high estimator.}$$

Gives a random interval $[\hat{\theta}_L, \hat{\theta}_H]$

such that $P\left(\theta \in [\hat{\theta}_L, \hat{\theta}_H]\right) \geq \underline{\underline{1-\alpha}}$

$\parallel$ random

$P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_H)$

deterministic

confidence level.

Ex $\alpha = .05$, confidence level $= 1-\alpha = .95$

$= 95\%$

$[\hat{\theta}_L, \hat{\theta}_H]$ is 95% confidence interval for θ.

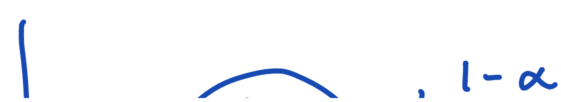
How to do this?

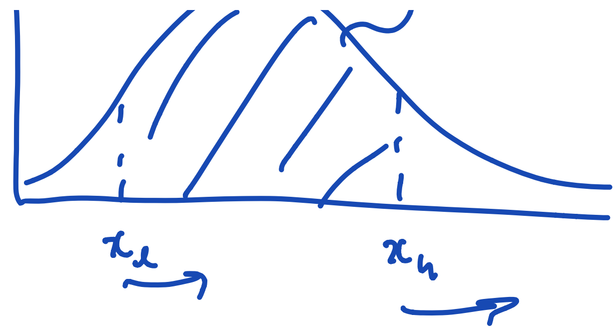
Consider instead: X, continuous RV.

$F_X(x)$ - cdf.

$$P(x_\ell \leq X \leq x_h) = 1-\alpha \quad ?$$

What are x_ℓ, x_h? (Not unique)





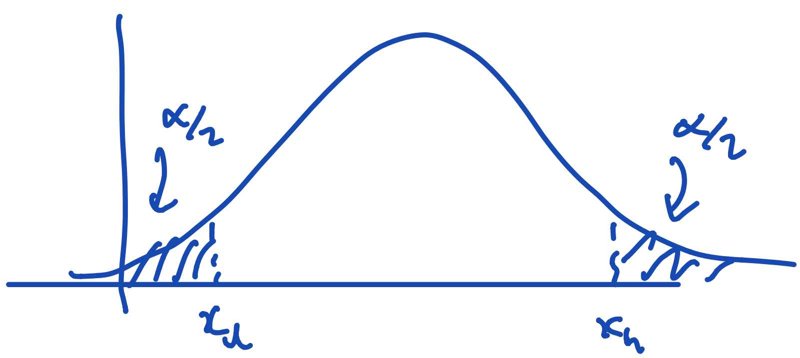
$$P(x_l \le X \le x_u) = F_X(x_u) - F_X(x_l)$$

choose x_u, x_l so that.

$$F_X(x_l) = 1 - \alpha/2 \quad , \quad F_X(x_u) = \alpha/2.$$

$$\hookrightarrow \quad F_X(x_l) - F_X(x_u) = 1 - \alpha/2 - (\alpha/2)$$
$$= 1 - \alpha.$$

We are choosing x_l, x_u so that



Choose $x_l = F_X^{-1}(\alpha/2)$, $x_u = F_X^{-1}(1-\alpha/2)$

so that X has probability $1-\alpha$ of being in $[x_l, x_u]$.

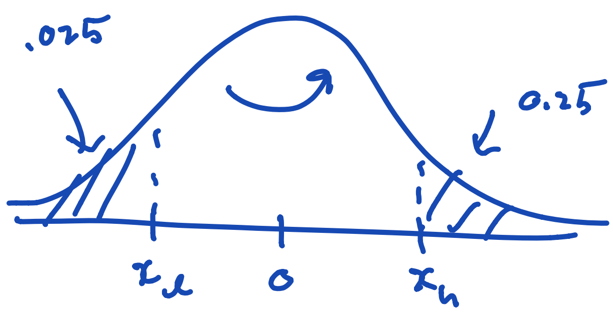
<u>Example</u>: $Z \sim N(0,1)$ find x_l, x_u

$$P(x_l \le Z \le x_u) = .95. \quad , \alpha = .05$$

z-table

$$x_l = \Phi^{-1}\left(\frac{.05}{2}\right) = \Phi^{-1}(.025) \approx -1.96.$$

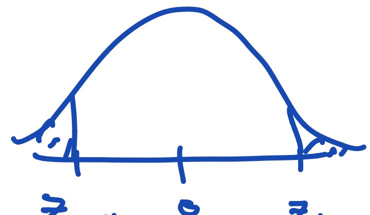
$$x_u = \Phi\left(1-\frac{.05}{2}\right) = \Phi^{-1}(1-.025) \approx 1.96$$



By symmetry $x_u = -x_l$.

How about any α?

Define. $z_{\alpha/2} := \Phi^{-1}(1-\alpha/2)$



symmetry $\boxed{z_{\alpha/2} = -z_{1-\alpha/2}.}$

$$\boxed{P(z_{1-\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha.}$$

How do we find a confidence interval $\{\hat{\theta}_L, \hat{\theta}_H\}$?

<u>Example</u> $X_1, X_2, \dots, X_n$ sample of $N(\theta, 1)$

unknown.

Find a 95% confidence interval for θ.

Let $\hat{\theta} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Since $X_i \sim N(0,1)$ and independent.

then $\underline{X_1 + X_2 + \dots + X_n}$ is normal.

and so

$$\bar{X} \sim N(\theta, 1/n)$$

since

$$\mathbb{E}\bar{X} = \theta \quad , \quad \mathrm{Var}(\bar{X}) = \frac{\overset{\sim 1}{\mathrm{Var}(X_i)}}{n}$$

Standardize

$$Z = \frac{\bar{X} - \mathbb{E}(\bar{X})}{\sqrt{\mathrm{Var}(\bar{X})}} = \frac{\bar{X} - \theta}{1/\sqrt{n}}$$

$$= \sqrt{n}(\bar{X} - \theta).$$

therefore

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

and so

$$P(-1.96 \le Z \le 1.96) = .95.$$

$$-1.96 \le Z \le 1.96 \iff -1.96 \le \sqrt{n}(\bar{X}-\theta) \le 1.96.$$

$$\iff -1.96 \le \sqrt{n}(\bar{X}-\theta)$$

and $\sqrt{n}(\bar{X}-\theta) \le 1.96.$ $\iff$ $\theta \le \bar{X} + \frac{1.96}{\sqrt{n}}$

$\theta \ge \bar{X} - \frac{1.96}{\sqrt{n}}.$

solve for θ.

$$-1.96 \le \sqrt{n}(\bar{X}-\theta) \le 1.96$$

$$\Updownarrow$$

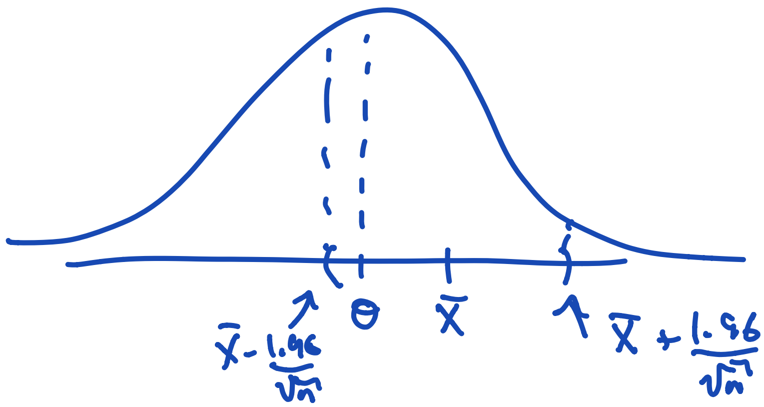
$$\bar{X} - \frac{1.96}{\sqrt{n}} \le \theta \le \bar{X} + \frac{1.96}{\sqrt{n}}.$$

Therefore

$$P\left(\bar{X} - \frac{1.96}{\sqrt{n}} \le \theta \le \bar{X} + \frac{1.96}{\sqrt{n}}\right) = .95.$$

so

$$[\hat{\theta}_L, \hat{\theta}_H] = \left[\bar{X} - \frac{1.96}{\sqrt{n}}, \bar{X} + \frac{1.96}{\sqrt{n}}\right].$$



This is an example of the <u>pivotal method</u>

We were able to find a quantity

$$Z = \sqrt{n}\,(\bar{X} - \theta) \sim N(0,1)$$

whose distribution doesn't depend on θ.

<u>Pivotal Method</u>

Let $X_1, X_2, \ldots, X_n$ be a sample from a distribution with parameter θ.

A pivotal quantity Q, is a random variable such that

① $Q = Q(X_1, X_2, \dots, X_n, \theta)$

function of the data <u>and</u> θ

② The distribution of Q <u>does not</u> depend on θ.

<u>The method.</u>

① Find a pivotal quantity. $Q(X_1, X_2, \dots, X_n, \theta)$.

② Find q_L, q_H such that.

$$P(q_L \le Q \le q_H) = 1 - \alpha.$$

do not depend on θ

③ Use algebra to find. $\hat{\theta}_L$, $\hat{\theta}_H$

$$P(\hat{\theta}_L \le \theta \le \hat{\theta}_H) = 1 - \alpha.$$

What do we do if X_i are not normal?

Large sample size.

Ex $X_1, X_2, \ldots, X_n$, $Var(X_i) = \sigma^2$

Find an $1-\alpha$ confidence interval for

$$\theta = E X_i .$$

Estimate θ by

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Since n is large,

$$Q = \frac{\bar{X} - \theta}{\sigma / \sqrt{n}} \approx N(0,1).$$

So $Q \approx$ pivotal quantity.

Find confidence interval for Q.

$$P\left(-z_{\alpha/2} \leq Q \leq z_{\alpha/2}\right) \approx 1-\alpha.$$

$$\underline{z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)}$$

<u>therefore</u>

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X}-\theta}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) \approx 1-\alpha.$$

is equivalent to.

$$P\left(\bar{X} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \leq \theta \leq \bar{X} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

therefore

$$\left[\bar{X} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}}, \bar{X} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right]$$

is an $1-\alpha$ confidence interval.

What if we don't know the variance?

σ is not known.

Two options

① Find a bound for σ

$$\sigma \leq \sigma_{max} \leftarrow \text{known.}$$

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma_{max}}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma_{max}}{\sqrt{n}}\right].$$

Is still $(1-\alpha)$ confidence interval.

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma_{max}}{\sqrt{n}} < \theta < \bar{X} + z_{\alpha/2}\frac{\sigma_{max}}{\sqrt{n}}\right) \geq 1-\alpha.$$

② Estimate σ^2. Finding another estimator σ^2. A little harder.

Example (Opinion polling).

Want to estimate the fraction of people who plan to vote for candidate A.

θ - proportion of people who plan to vote for A.

$\hookrightarrow$ want to estimate

Sample the population:

$$X_i = \begin{cases} 1 & \text{if person } i \text{ plans to vote for candidate } A. \\ 0 & \text{otherwise.} \end{cases}$$

$X_i \sim \text{Bernoulli}(\theta)$

Random sample $X_1, X_2, \ldots, X_n$

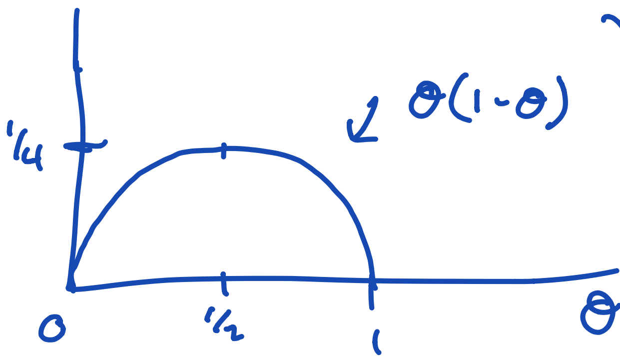
Find a $1-\alpha$ confidence interval for θ if n is large.

$\mathbb{E}X_i = \theta$, $Var(X_i) = \theta(1-\theta) = \sigma^2$

We don't know $\sigma^2 = \theta(1-\theta)$.

However $\theta(1-\theta) \le 1/4$, $\theta \in (0,1)$.

$\qquad\qquad\qquad\qquad \| \ \sigma^2_{max}$.



Therefore $\sigma_{max} = 1/2$

By our previous Example (using the CLT)

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma_{max}}{\sqrt{n}}, \ \bar{X} + z_{\alpha/2}\frac{\sigma_{max}}{\sqrt{n}}\right].$$

$$= \left[\bar{X} - z_{\alpha/2}\frac{1}{2\sqrt{n}}, \ \bar{X} + z_{\alpha/2}\frac{1}{2\sqrt{n}}\right]$$

is a $(1-\alpha)$ confidence interval for Θ.

Ex Suppose we want a 3% margin of error on our opinion poll. How big do we need to take n so that we are 95% confident that our answer is within 3% of the true value?

Confidence interval was.

$$\left[\bar{X} - z_{\alpha/2}\frac{1}{2\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{1}{2\sqrt{n}}\right].$$

$$\text{Margin.} = z_{\alpha/2}\frac{1}{2\sqrt{n}} - \left(-z_{\alpha/2}\frac{1}{2\sqrt{n}}\right)$$

$$= \frac{z_{\alpha/2}}{\sqrt{n}} = \frac{1.96}{\sqrt{n}} \quad (\alpha = .05)$$

$$\le .03.$$

$$n \ge \left(\frac{1.96}{.03}\right)^2$$