H. Confidence Intervals

1. Definition:
   a. How close do we believe our estimator is to the true value? What range can we be pretty sure of seeing the true value in?
   b. We saw that on the one hand, the probability of hitting the true value on the head is zero, and on the other hand in order to get a range of possible values that will always include the true value, we’d have to take the whole parameter domain.
   c. Compromise solution is to look for bounds \( L(X_1, \cdots, X_n) \) and \( U(X_1, \cdots, X_n) \) s.t. \( L(X_1, \cdots, X_n) \) will fall below the parameter and that \( U(X_1, \cdots, X_n) \) will fall above the parameter with a certain probability.
   d. If such an \( L(X_1, \cdots, X_n) \) and \( U(X_1, \cdots, X_n) \) exist they are called a confidence interval (c.i.)
   e. Satisfy mathematical statement like \( P(L \leq \theta \leq U) \geq \alpha \).
      i. For concreteness, say we’re looking for c.i.s that will hold 90% of the time; we want \( P(L \leq \theta \leq U) \geq 90% \).

2. Strategies:
a. Algebraic strategy:

i. Manipulate a probability statement about the parameter of interest and a statistic that the interval end points are likely to be a function of.

ii. We solve \( F(t_{.95}) = .95 \) giving \((t/\theta)^n = .95\) or \( t = \theta n^{\sqrt{.95}} \) and \( F(t_{.05}) = .05 \) giving \((t/\theta)^n = .05\) or \( t = \theta n^{\sqrt{.05}} \).

iii. Hence \( P(\max_j X_j < \theta n^{\sqrt{.05}}) = P(\max_j X_j > \theta n^{\sqrt{.95}}) = .05 \). See Fig. 3.

iv. Hence \( P(\theta n^{\sqrt{.05}} \leq \max_j X_j \leq \theta n^{\sqrt{.95}}) = .90 \).

v. Hence \( P(T^{-n^{\sqrt{.05}}} \leq \theta \leq T^{-n^{\sqrt{.95}}}) = .90 \).

vi. Works because

- \( T/\theta \) has a dist'n func. ind. of what we were trying to estimate or other unknown parameters;
- we say it is **pivotal**.
- solve for \( \theta \).

b. Example: Normal Distribution with known variance.

\[ F_{\bar{X}}(x; \mu) = \Phi((x - \mu)\sqrt{n\sigma^{-1}}). \]

i. Hence \( (\bar{X} - \mu)\sqrt{n\sigma^{-1}} \) is pivotal.

ii. \( P(z_{\alpha/2} \leq (\bar{X} - \mu)\sqrt{n\sigma^{-1}} \leq z_{1-\alpha/2}) = 1 - \alpha \)
iii. \[ P \left( \frac{\sigma z_{\alpha/2}}{\sqrt{n}} \leq \bar{X} - \mu \leq \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} \right) = 1 - \alpha \]

iv. \[ P \left( -\frac{\sigma z_{\alpha/2}}{\sqrt{n}} \geq \mu - \bar{X} \geq -\frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} \right) = 1 - \alpha \]

v. \[ P \left( \bar{X} + \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} \geq \mu \geq \bar{X} - \frac{\sigma z_{\alpha/2}}{\sqrt{n}} \right) = 1 - \alpha \]

c. Example: Normal Distribution with unknown variance.

i. \[ P \left( -t_{0.025, n-1} \leq \frac{\bar{X} - \mu}{s/\sqrt{n}} \leq -t_{0.25, n-1} \right) = .95 . \]

ii. \( t_{0.025, n-1} \) is value such that \( t \) distribution with \( n - 1 \) df has .025 probability above it.

iii. \[ P \left( -s/\sqrt{nt_{0.25, n-1}} \leq \bar{X} - \mu \leq -s/\sqrt{nt_{0.025, n-1}} \right) = .95 . \]
Sample size is 10. 94 intervals cover true value.