III. Inference

A. Confidence Intervals

1. Example: Paleontology.
   a. Goal: estimate how long ago a certain species of animal first walked or crawled the earth.
   b. You assume
      i. species population has been constant since its advent $\theta$ years ago,
      ii. the probability of finding any one of these animals is the same regardless of its age.
   c. Completely unreasonable assumptions imply that the age $X$ of a given sample $\sim \mathcal{U}(0, \theta)$.
   d. You date $n$ specimens.
      i. Recall m.l.e. $\hat{\theta}$ for $\theta$ is $\hat{\theta} = \max X_j$.
      ii. Recall that $\hat{\theta}$ is biased, but $(\max X_j)(n + 1)/n$ is unbiased.

2. Definition:
   a. How close do we believe $\hat{\theta}$ is to $\theta$? 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 \( \theta_L(X_1, \cdots, X_n) \) and \( \theta_U(X_1, \cdots, X_n) \) such that \( \theta_L(X_1, \cdots, X_n) \) will fall below the parameter and that \( \theta_U(X_1, \cdots, X_n) \) will fall above the parameter with a certain probability.

d. If such an \( \theta_L(X_1, \cdots, X_n) \) and \( \theta_U(X_1, \cdots, X_n) \) exist they are called a confidence interval c.i.

e. Satisfy mathematical statement like \( P[\theta_L \leq \theta \leq \theta_U] \geq \alpha \).

i. For concreteness, say we’re looking for confidence intervals that will hold 90% of the time; we want \( P[\theta_L \leq \theta \leq \theta_U] \geq 90\% \).

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

a. Since the m.l.e. is also a sufficient statistic, look for bounds that are functions of \( T = \hat{\theta} = \max X_j \).
Lecture 9

b. Graphical strategy:
   
   i. cumulative distribution function of $T$ is $F(t; \theta) = t^n/\theta^n$ if $t \leq \theta$.
   
   ii. Choose $t_1(\theta)$ and $t_2(\theta)$ such that $\forall \theta$, $P[T \geq t_1(\theta)] \geq .95$ and $P[T \leq t_2(\theta)] \geq .95$.
   
   iii. Here set $t_1(\theta) = F^{-1}(.05; \theta)$ and $t_2(\theta) = F^{-1}(.95; \theta)$.
   
   iv. For each potential value of $\hat{\theta}$,
      
      • draw a horizontal line between the curves.
      
      • This will be the confidence interval as a function of $\hat{\theta}$.
   
   v. If $t_2(\theta)$ and $t_1(\theta)$ are increasing in $\theta$ the vertical line above $\theta$ from $t_2(\theta)$ to $t_1(\theta)$ is the confidence interval.
   
   vi. How often will this cover $\theta$?
      
      • confidence interval covers $\theta$ if and only if the vertical and horizontal lines cross,
      
      • if and only if $\hat{\theta}$ lies between $t_2$ and $t_1$,
      
      • happens $1 - \alpha$ of the time.
   
   c. Algebraic strategy:
      
   i. Let $S = T/\theta$
      
   ii. Let $F_S(s)$ be the c.d.f. of $S$
Confidence Interval Construction for Sample of Independent Uniforms

iii. Solve $F_S(s_{.95}) = .95$ and $F(s_{.05}) = .05$. 


Lecture 10

- For example of \( n \ U[0, \theta] \) variables, \( s_{.95}^n = .95 \) or \( s_{.95} = \sqrt{.95} \) and \( s_{.05}^n = .05 \) or \( s_{.05} = \sqrt{.05} \).

iv. Hence \( P \left[ T < \theta \sqrt{n}.05 \right] = P \left[ T > \theta \sqrt{n}.95 \right] = .05 \).

v. Hence \( P \left[ \theta \sqrt{n}.05 \leq T \leq \theta \sqrt{n}.95 \right] = .90 \).

vi. Hence \( P \left[ T \sqrt{n}.05 \leq \theta \leq T \sqrt{n}.95 \right] = .90 \).

vii. Works because \( S = T/\theta \) has a distn func. ind. of what we were trying to estimate or other unknown parameters;

viii. we say \( S \) is pivotal.

ix. What we really did was: Construct intervals for the pivotal quantity, and solve for \( \theta \).

4. More Confidence Interval Examples

F: 11.2

a. Example: Normal Distribution with known variance. \( T = \bar{X} \);

\[ F_T(t; \mu) = \Phi((t - \mu)\sqrt{n\sigma^{-1}}). \]

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

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

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

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