STAT 5430

Lec 37, M, Apr 28

- Homework 8 is assigned & due M, Apr 28 by midnight
- Homework 9 is assigned & due Sunday, May 4 but you can submit this on Monday, May 5
- Exam 2 solutions & grading key posted
- Final Exam on Tuesday, May 13, 7:30-9:30 PM
  - Comprehensive - but focus on material since Exam 2 (interval estimation)
  - Formula sheet for new material/interval & 2 formula sheets previous material (3 sheets (front/back) each total)
  - Practice Exams

see Canvas

STAT 5430: Summary to date
Where we have been
& where we are headed

- Completed
  - Introduction to Statistical Inference
  - Point Estimation
    * MME/MLE
  - Criteria for Evaluating Point Estimators
    * bias, variance, UMVUE, MSE
  - Elements of Decision Theory
    * Minimax, finding Bayes estimators
  - Sufficiency and Point Estimation
    * Factorization/Rao-Blackwell/Lehman-Scheffe Theorems
  - Hypothesis Testing
    * MP/UMP, Likelihood Ratio/Bayes Tests
  - Interval Estimation I
    * Inverting Tests/Pivotal Quantities/Asymptotic Pivots/VST
- Next: Interval Estimation II
  - MGB Method (Pivot-based)
  - Bayes Intervals
  - Evaluating Interval Estimators

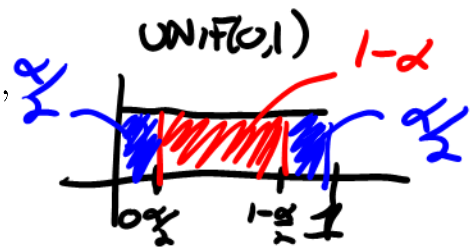
Interval Estimation II

Mood, Graybill & Boes (MGB) Interval Method

uses probability integral transform to make a certain pivot (Sec 9.2.3, C & B)

Main Idea: Suppose a statistic T is a continuous random variable having a cdf $F(t|\theta) = P(T \leq t|\theta)$, $t \in \mathbb{R}$, which depends on a real-valued parameter $\theta \in \Theta \subset \mathbb{R}$. If $\theta \in \Theta$ is the data-generating parameter, then (since T is continuous) by the probability integral transform (PIT)

$$F(T|\theta) \sim \text{Uniform}(0,1),$$



and hence $Q(T,\theta) \equiv F(T|\theta)$ is a pivotal quantity.

Then, for $Q \sim \text{Uniform}(0,1)$ and $\alpha \in (0,1)$, it holds that $P(\frac{\alpha}{2} \leq Q \leq 1-\frac{\alpha}{2}) = 1-\alpha$ and, given T, a confidence region for θ with C.C. $1-\alpha$ is

a b 1-α

$$C_T \equiv \left\{\theta \in \Theta : \frac{\alpha}{2} \leq Q(T,\theta) \leq 1 - \frac{\alpha}{2}\right\} = \left\{\theta \in \Theta : \frac{\alpha}{2} \leq F(T|\theta) \leq 1 - \frac{\alpha}{2}\right\}.$$

a b

For given a value $T=t$, if it turns out that $F(t|\theta)$ is increasing in θ or decreasing in θ, then the above confidence region based on this $T=t$ value will be an **interval**, say, $C_{T=t} = [\theta_L(t), \theta_U(t)]$ with endpoints determined by $\theta_L(t) = \min\{a(t), b(t)\}$ and $\theta_U(t) = \max\{a(t), b(t)\}$ for

$$\frac{\alpha}{2} = F(t|a(t)) = P(T \leq t|a(t)), \qquad \frac{\alpha}{2} = 1 - F(t|b(t)) = P(T \geq t|b(t));$$

$a(t), b(t) \in \Theta$ are points where, given $T=t$, the cdf $F(t|\theta)$ "crosses" $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ as a function of θ.

This is the MGB method with an extension allowing for discrete statistics T in addition to continuous ones.

Interval Estimation II

Mood, Graybill & Boes (MGB) Interval Method

Theorem: Let T be a statistic (a discrete or continuous random variable is allowed) with cdf $F(t|\theta) = P(T \le t|\theta)$, $\theta \in \Theta \subset \mathbb{R}$. Suppose $\mathcal{T}$ ="the set of all possible values of T".

for each possible value t of T ($t \in \mathcal{T}$)

Let $0 < \alpha < 1$. Suppose, for each $t \in \mathcal{T}$, there exist functions $\theta_L(t)$ and $\theta_U(t)$ such that

given

(a) if $F(t|\theta)$ is a decreasing function of θ for each t, then

$$\overbrace{P\Big(T \le t|\theta_U(t)\Big)}^{\text{lower tail}} = \frac{\alpha}{2} \quad \& \quad \overbrace{P\Big(T \ge t|\theta_L(t)\Big)}^{\text{upper tail}} = \frac{\alpha}{2}$$

given t   given t

(b) or if $F(t|\theta)$ is an increasing function of θ for each t, then

$$\overbrace{P\Big(T \le t|\theta_L(t)\Big)}^{\text{lower tail}} = \frac{\alpha}{2} \quad \& \quad \overbrace{P\Big(T \ge t|\theta_U(t)\Big)}^{\text{upper tail}} = \frac{\alpha}{2}$$

given t   given t

Then,

$$\big[\theta_L(T), \theta_U(T)\big]$$

is a CI for θ with a confidence coefficient (C.C.) satisfying C.C. $\ge 1 - \alpha$.

In particular,

when T is a continuous random variable (namely, $F(t|\theta)$ is continuous in t), then $\big[\theta_L(T), \theta_U(T)\big]$ has

$$\text{C.C.} = \min_{\theta \in \Theta} P_\theta\left(\theta \in \big[\theta_L(T), \theta_U(T)\big]\right) = 1 - \alpha.$$

T continuous, CC is exactly 1-α

when T is a discrete random variable (namely, $F(t|\theta)$ is a step function in t), then $\big[\theta_L(T), \theta_U(T)\big]$ has

not exactly 1-α for discrete T

$$\text{C.C.} = \min_{\theta \in \Theta} P_\theta\left(\theta \in \big[\theta_L(T), \theta_U(T)\big]\right) \ge 1 - \alpha.$$

Interval Estimation II

Mood, Graybill & Boes (MGB) Interval Method

Remarks:

1. A confidence interval I (a function of the data) for $\theta \in \Theta \subset \mathbb{R}$ is called a **conservative** $(1-\alpha)$ **confidence interval** if

$$\min_{\theta \in \Theta} P_\theta(\theta \in I) \geq 1 - \alpha$$

lower band on coverage

2. The theorem above says that the confidence interval $\left[\theta_L(T), \theta_U(T)\right]$ will have a C.C. of exactly $(1-\alpha)$ when the statistic T (based on the data) is a continuous random variable. But, when T is a discrete random variable, the C.C. of the interval $\left[\theta_L(T), \theta_U(T)\right]$ may not exactly equal $(1-\alpha)$, but cannot be smaller. Hence, when T is discrete, the interval $\left[\theta_L(T), \theta_U(T)\right]$ will be a conservative $(1-\alpha)$ confidence interval.

Interval Estimation II

Mood, Graybill & Boes (MGB) Interval Method: Illustration

Example 1. (Continuous random variables) Suppose $X_1, \ldots, X_n$ are iid $N(\theta, \log\theta)$, $1 < \theta < \infty$. Apply the Mood-Graybill-Boes method to obtain a $1-\alpha$ CI for θ based on $T = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = (n-1)S^2$.

Solution: first, we need to get cdf of T

Note: $\dfrac{(n-1)S^2}{\log\theta} \sim \chi^2_{n-1}$.

$$\underbrace{\log\theta}_{\text{pop. variance}} \qquad \dfrac{(n-1)S^2}{\log\theta} = \dfrac{T}{\log\theta}$$

Then,

$$F(t|\theta) = P(T \le t|\theta) = P\left(\frac{T}{\log\theta} \le \frac{t}{\log\theta}\Big|\theta\right)$$

$$= P\left(\chi^2_{n-1} \le \frac{t}{\log\theta}\right) \equiv G\left(\frac{t}{\log\theta}\right),$$

where $G(\cdot)$ is the cdf of χ^2_{n-1}

So, $\mathcal{T} \equiv$ "possible values of T" $= (0,\infty)$.

pick/fix a possible value $t \in (0,\infty)$ of T.

(Note: given t, $F(t|\theta) = G\left(\frac{t}{\log\theta}\right) \downarrow$ as $\theta \uparrow$)

Find $\theta_U \equiv \theta_U(t)$ & $\theta_L \equiv \theta_L(t)$ so that

$$\frac{\alpha}{2} = P(T \le \underset{\text{given}}{t}|\theta_U) \quad \& \quad \frac{\alpha}{2} = P(T \ge \underset{\text{given}}{t}|\theta_L)$$

$$= G\left(\boxed{\frac{t}{\log\theta_U}}\right) \qquad \xrightarrow{T \text{ continuous}} = 1 - P(T \le t|\theta_L)$$

$$= 1 - G\left(\boxed{t/\log\theta_L}\right)$$

$$\Rightarrow \boxed{\frac{t}{\log \theta_U}} = \boxed{\underbrace{\chi^2_{n-1,\frac{\alpha}{2}}}_{\frac{\alpha}{2}\text{ percentile}}} \quad \& \quad \boxed{\frac{t}{\log \theta_L}} = \boxed{\underbrace{\chi^2_{n-1,1-\frac{\alpha}{2}}}_{1-\frac{\alpha}{2}\text{ percentile}}}$$

$\Rightarrow$ solve for θ_U & θ_L & get $[\theta_L(t), \theta_U(t)]$

In summary,

$$[\theta_L(T), \theta_U(T)] = \left[\exp\left[T/\chi^2_{n-1,1-\frac{\alpha}{2}}\right], \exp\left[T/\chi^2_{n-1,\frac{\alpha}{2}}\right]\right]$$

is CI for θ with c.c $1-\alpha$.

Interval Estimation II

Mood, Graybill & Boes (MGB) Interval Method: Illustration

Example 2. (Discrete random variables) Suppose T is a geometric(θ) random variable, $0 < \theta < 1$. Apply the Mood-Graybill-Boes method to obtain a CI for θ based on T.

Solution: "range of possible values of T" $= \{1, 2, 3, \dots\}$

Fix $t \in \{1, 2, 3, \dots\}$. Then

$$F(t|\theta) = P(T \le t|\theta) = \sum_{i=1}^{t} \underbrace{\theta(1-\theta)^{i-1}}_{P_\theta(T=i)}$$

$$= \theta\{1 + (1-\theta) + \cdots + (1-\theta)^{t-1}\}\frac{[1-(1-\theta)]}{[1-(1-\theta)]}$$

$$= \cancel{\theta}\{1 + (1-\theta) + \cdots (1-\theta)^{t-1} - (1-\theta) - (1-\theta)^2 + \cdots - (1-\theta)^t\}\frac{1}{\cancel{\theta}}$$

$$= 1 - (1-\theta)^t$$ Note: as $\theta \uparrow$, $F(t|\theta) \uparrow$ increasing

Given a possible value where $t \in \{1, 2, 3 \dots\}$ of T, find $\theta_L \equiv \theta_L(t)$ & $\theta_U \equiv \theta_U(t)$ where

$$\frac{\alpha}{2} = P(T \le t|\theta_L) = 1 - (1-\theta_L)^t$$

&

$$\frac{\alpha}{2} = P(T \ge t|\theta_U) = 1 - P(T < t|\theta_U) = 1 - P(T \le t-1|\theta_U) = 1 - [1 - (1-\theta_U)^{t-1}]$$

or

$$\theta_L \equiv \theta_L(t) = 1 - \left(1 - \frac{\alpha}{2}\right)^{1/t}$$

or

$$\theta_U \equiv \theta_U(t) = 1 - \left(\frac{\alpha}{2}\right)^{\frac{1}{t-1}}$$

So, $$[\theta_L(T), \theta_U(T)] = \left[1 - \left(1 - \frac{\alpha}{2}\right)^{\frac{1}{T}}, 1 - \left(\frac{\alpha}{2}\right)^{\frac{1}{T-1}}\right]$$

is a CI for θ with C.C. $\ge 1 - \alpha$ (since T is discrete)

Note: If $T = 1$, interpret $\frac{1}{T-1} = \frac{1}{0} = \infty$ & interpret $1 - \left(\frac{\alpha}{2}\right)^{\frac{1}{T-1}} = 1 - 0 = 1$

check $P_\theta(\theta \in [\theta_L(T), \theta_U])$ for $\theta \in (0,1)$ numerically

```
theta<-0.01
M<-10000000
CC<-.90


T<-rgeom(M,theta)+1

alpha.2<-(1-CC)/2

LOWER<-1-(1-alpha.2)^{1/T}
UPPER<-1-(alpha.2)^{1/(T-1)}

COVER<-rep(1,M)
COVER[LOWER>theta]<-0
COVER[UPPER<theta]<-0
mean(COVER)
```

```
M<-10000000
CC<-.90

t<-seq(.01,1,.01)
t1<-t*0
n<-length(t)

for(i in 1:n){
theta<-t[i]
 print(i)
T<-rgeom(M,theta)+1

alpha.2<-(1-CC)/2

LOWER<-1-(1-alpha.2)^{1/T}
UPPER<-1-(alpha.2)^{1/(T-1)}

COVER<-rep(1,M)
COVER[LOWER>theta]<-0
COVER[UPPER<theta]<-0

t1[i]<-mean(COVER)
}



plot(t,t1,xlab="theta",ylab="Coverage",main="Actual Coverage
of 90% MBG Interval for Geometric Parameter theta",pch=20)
```