Lecture 19: Pareto Example

MATH 667-01
Statistical Inference
University of Louisville

November 28, 2017

- We extend Example L16.4 and compare tests and confidence intervals based on various pivots.
- Some of the pivots are based on order statistics defined in Section 5.4 of Casella and Berger (2002)[1].

[1] Casella, G. and Berger, R. (2002). *Statistical Inference, Second Edition*. Duxbury Press, Belmont, CA.

Pivot: Maximum

- In *Example L16.4*, we saw that if $X_1, \ldots, X_n$ are iid random variables from a Pareto distribution with pdf

$$f(x|\theta) = \frac{\theta}{x^{\theta+1}} I_{(1,\infty)}(x)$$

where $\theta > 0$, then $(0, -\ln(1 - \sqrt[n]{1-\alpha})/\ln X_{(n)}]$ is a $100(1-\alpha)\%$ confidence interval for θ.

- So, if $n = 3$, then $(0, -\ln(1 - \sqrt[3]{1-\alpha})/\ln X_{(3)}]$ is a $100(1-\alpha)\%$ confidence interval for θ.

- *Example L19.1*: Suppose X_1, X_2, and X_3 are iid random variables from a Pareto distribution with pdf

$$f(x|\theta) = \frac{\theta}{x^{\theta+1}} I_{(1,\infty)}(x)$$

where $\theta > 0$.
(a) Let $Y = X_{(1)}$. Find the cdf $F_Y(y) = P(Y \leq y)$ for $y > 1$.
(b) Using (a) to obtain a pivot, find a $100(1-\alpha)\%$ confidence interval for θ with the form $(0, \theta_U]$.

- *Answer to Example L19.1*: (a) If $y \geq 1$, then the cdf of Y is

$$\begin{aligned}
F_Y(y) &= 1 - P(X_1 > y, X_2 > y, X_3 > y) \\
&= 1 - \prod_{i=1}^{3} P(X_i > y) \\
&= 1 - \left(\int_y^\infty \theta x^{-\theta-1} dx \right)^3 \\
&= 1 - \left(\left[-x^{-\theta} \right]_y^\infty \right)^3 = 1 - (y^{-\theta})^3 = 1 - y^{-3\theta}.
\end{aligned}$$

- *Answer to Example L19.1 continued*: (b) The $100(1-\alpha)\%$ confidence interval $(0, -\ln\alpha/(3\ln X_{(1)})]$ for θ can be obtained as follows.

$$\begin{aligned} 1-\alpha &= P(F_Y(Y) \leq 1-\alpha) \\ &= P(1 - Y^{-3\theta} \leq 1-\alpha) \\ &= P(Y^{-3\theta} \geq \alpha) \\ &= P(-3\theta \ln Y \geq \ln\alpha)) \\ &= P\left(\theta \leq \frac{-\ln\alpha}{3\ln Y}\right) \end{aligned}$$

- *Theorem L19.1* (Thm 5.4. on p.): Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the pdf of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)[F_X(x)]^{j-1}[1 - F_X(x)]^{n-j}.$$

- The cdf of $X_{(j)}$ is

$$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} (F_X(x))^k (1 - F_X(x))^{n-k}.$$

- *Example L19.2*: Suppose $X_1, X_2,$ and X_3 are iid random variables from a Pareto distribution with pdf

$$f(x|\theta) = \frac{\theta}{x^{\theta+1}} I_{(1,\infty)}(x)$$

where $\theta > 0$.
(a) Let $Y = X_{(2)}$. Find the cdf $F_Y(y) = P(Y \leq y)$ for $y > 1$.
(b) Using (a) to obtain a pivot, find a $100(1-\alpha)\%$ confidence interval for θ with the form $(0, \theta_U]$.

- *Answer to Example L19.2*: (a) If $y \geq 1$, then

$$\begin{aligned} F_Y(y) &= \sum_{k=2}^{3} \binom{3}{k} (1-y^{-\theta})^k (y^{-\theta})^{n-k} \\ &= 3(1-y^{-\theta})^2(y^{-\theta}) + (1-y^{-\theta})^3 \\ &= (1-y^{-\theta})^2(3y^{-\theta}+1-y^{-\theta}) \\ &= (1-y^{-\theta})^2(2y^{-\theta}+1) \\ &= 2(y^{-\theta})^3 - 3(y^{-\theta})^2 + 1. \end{aligned}$$

- (b) For any $v \in (0,1)$, $2u^3 - 3u^2 + v < 0$ if and only if $u > \frac{1}{2} + \cos(\frac{1}{3}\arccos(1-2v) - \frac{2\pi}{3})$ for all $u \in (0,1)$.

- *Answer to Example L19.2 continued*: (b) The $100(1-\alpha)\%$ confidence interval for θ can be obtained as follows.

$$
\begin{aligned}
1-\alpha &= P(F_Y(Y) \leq 1-\alpha) \\
&= P(2Y^{-3\theta} - 3Y^{-2\theta} + 1 \leq 1-\alpha) \\
&= P(2Y^{-3\theta} - 3Y^{-2\theta} + \alpha \leq 0) \\
&= P\left(Y^{-\theta} \geq \frac{1}{2} + \cos\left(\frac{1}{3}\arccos(1-2\alpha) - \frac{2\pi}{3}\right)\right) \\
&= P\left(-\theta \ln Y \geq \ln\left(\frac{1}{2} + \cos\left(\frac{1}{3}\arccos(1-2\alpha) - \frac{2\pi}{3}\right)\right)\right) \\
&= P\left(\theta \leq \frac{-\ln\left(\frac{1}{2} + \cos\left(\frac{1}{3}\arccos(1-2\alpha) - \frac{2\pi}{3}\right)\right)}{\ln Y}\right).
\end{aligned}
$$

- *Example L19.3*: Suppose X_1, X_2, and X_3 are iid random variables from a Pareto distribution with pdf

$$f(x|\theta) = \frac{\theta}{x^{\theta+1}} I_{(1,\infty)}(x)$$

where $\theta > 0$. Find a UMP level α test for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

- *Answer to Example L19.3*: The joint pdf of X_1, X_2, and X_3 is

$$f(x_1, x_2, x_3|\theta) = \theta^3 e^{-(\theta+1)\sum_{i=1}^{3} \ln x_i} I_{(1,\infty)}(x_{(1)})$$

so $\displaystyle\sum_{i=1}^{3} \ln x_i$ is sufficient for θ by *Theorem L10.2*.

UMP Test

- *Answer to Example L19.3 continued*: Since

$$\begin{aligned} P(\ln X_i \le y) &= P(X_i \le e^y) \\ &= \int_1^{e^y} \frac{\theta}{x^{\theta+1}}\, dx \\ &= \Big[-x^{-\theta}\Big]_1^{e^y} = 1 - e^{-\theta y} \end{aligned}$$

is the cdf of an exponential random variable with mean $1/\theta$, it can be shown that $T = \sum_{i=1}^{3} \ln X_i \sim$ Gamma$(3, \frac{1}{\theta})$.
This family of pdfs has a MLR since, for $\theta_1 < \theta_2$,
$\dfrac{g(t|\theta_2)}{g(t|\theta_1)} = \dfrac{\frac{1}{2}\theta_2^3 t^2 e^{-\theta_2 t}}{\frac{1}{2}\theta_1^3 t^2 e^{-\theta_1 t}} = \dfrac{\theta_2^3}{\theta_1^3} e^{-(\theta_2-\theta_1)t}$
is a nonincreasing function of t.
So, by the Karlin-Rubin Theorem, the test that rejects H_0 if and only if $T < t_0$ is a UMP level α test. (Note the change in direction from *Theorem L15.3*.)

- *Answer to Example L19.3 continued*: Now we find t_0 so that $P_{\theta_0}(T < t_0) = \alpha$.
  Using integration by parts we have

$$\begin{aligned} P_{\theta_0}(T < t_0) &= \int_0^{t_0} \frac{1}{2}\theta_0^3 t^2 e^{-\theta_0 t}\, dt \\ &= \left[\left(-\frac{1}{2}\theta_0^2 t^2 - \theta_0 t - 1\right) e^{-\theta_0 t}\right]_0^{t_0} \\ &= 1 - \left(\frac{1}{2}\theta_0^2 t_0^2 - \theta_0 t_0 - 1\right) e^{-\theta_0 t_0}. \end{aligned}$$

  So, t_0 is the value that satisfies

$$1 - \left(\frac{1}{2}\theta_0^2 t_0^2 + \theta_0 t_0 + 1\right) e^{-\theta_0 t_0} = \alpha$$

  which can be computed using the R command
  `qgamma(.05,shape=3,rate=theta0).`

- *Example L19.4*: Suppose X_1, X_2, and X_3 are iid random variables from a Pareto distribution with pdf

$$f(x|\theta) = \frac{\theta}{x^{\theta+1}} I_{(1,\infty)}(x)$$

where $\theta > 0$. If the true value of the parameter is $\theta = 5$, compare the power of the UMP level $.05$ test for testing $H_0 : \theta \leq 2$ versus $H_1 : \theta > 2$ with the power of the following three tests:

    - Test 1: Reject if $X_{(1)} \leq .95^{-1/(3\theta_0)}$
    - Test 2: Reject if $X_{(2)} \leq (\frac{1}{2} + \frac{1}{3} \cos\left(\arccos(-.9) - \frac{2\pi}{3}\right))^{-1/\theta_0}$
    - Test 3: Reject if $X_{(3)} \leq (1 - .05^{1/3})^{-1/\theta_0}$

Comparison of Tests

- *Answer to Example L19.4*: Since $\theta_0 = 2$, the UMP test rejects H_0 if $\sum_{i=1}^{3} \ln X_i \le .4088457$ (computed in R using the command `qgamma(.05,shape=3,rate=2)`.
- Since $\sum_{i=1}^{3} \ln X_i \sim$ Gamma$(3, \frac{1}{\theta})$ with $\theta = 5$, the power of the test under this alternative is

$$P\left(\sum_{i=1}^{3} \ln X_i \le .4088457\right) = \int_0^{.4088457} \frac{5^3}{\Gamma(3)} t^2 e^{-5t}\, dt \approx .335293$$

(computed in R using the command `pgamma(qgamma(.05,shape=3,rate=2),shape=3,rate=5)`).

Comparison of Tests

- *Answer to Example L19.4 continued*: We can use the cdf of each order statistic to compute the power of each of the tests:

$$
\begin{aligned}
P\left(X_{(1)} \leq .95^{-1/6}\right) &\approx P\left(X_{(1)} \leq .9914876\right) \\
&\overset{19.5}{=} 1-(.9914876)^{-15} \approx .1203518
\end{aligned}
$$

$$
P\left(X_{(2)} \leq \left(\frac{1}{2}+\frac{1}{3}\cos\left(\arccos(-.9)-\frac{2\pi}{3}\right)\right)^{-1/2}\right)
$$

$$
\begin{aligned}
\approx P\left(X_{(2)} \leq 1.075424\right) &\overset{19.9}{=} 2(1.075424)^{-15}-3(1.075424)^{-10}+1 \\
&\approx .2220945
\end{aligned}
$$

$$
\begin{aligned}
P\left(X_{(3)} \leq (1-.05^{1/3})^{-1/2}\right) &\approx P\left(X_{(3)} \leq 1.258288\right) \\
&\overset{16.20}{=} (1-1.258288^{-5})^3 \\
&\approx .3185705
\end{aligned}
$$

Simulation

Here is a simulation to check the size of the test:

```
> set.seed(345672)
> R=1000000;n=3
> theta0=2;theta=2
> X1=rep(0,R);X2=rep(0,R);X3=rep(0,R)
> sumlnX=rep(0,R)
> for (i in 1:R){
+  u=runif(3)
+  x=(1-u)^(-1/theta)
+  sx=sort(x)
+  X1[i]=sx[1];X2[i]=sx[2];X3[i]=sx[3]
+  sumlnX[i]=sum(log(x))
+ }
> mean(X1<.95^(-1/(3*theta0)))
[1] 0.049813
> mean(X2<(.5+cos(acos(-.9)/3-2*pi/3))^(-1/theta0))
[1] 0.050122
> mean(X3<(1-.05^(1/3))^(-1/theta0))
[1] 0.049606
> mean(sumlnX<qgamma(.05,shape=3,rate=theta0))
[1] 0.049739
```

Simulation

Here is a simulation to check the power of the test:

```
> set.seed(453672)
> R=1000000;n=3
> theta0=2;theta=5
> X1=rep(0,R);X2=rep(0,R);X3=rep(0,R)
> sumlnX=rep(0,R)
> for (i in 1:R){
+  u=runif(3)
+  x=(1-u)^(-1/theta)
+  sx=sort(x)
+  X1[i]=sx[1];X2[i]=sx[2];X3[i]=sx[3]
+  sumlnX[i]=sum(log(x))
+ }
> mean(X1<.95^(-1/(3*theta0)))
[1] 0.120596
> mean(X2<(.5+cos(acos(-.9)/3-2*pi/3))^(-1/theta0))
[1] 0.222173
> mean(X3<(1-.05^(1/3))^(-1/theta0))
[1] 0.317824
> mean(sumlnX<qgamma(.05,shape=3,rate=theta0))
[1] 0.335018
```