Stat 8003, Homework 8

Group G

November 13, 2014

Question 8.1. Let $X_1, X_2, \cdots, X_n$ be a random sample from an exponential distribution with the density function

$$f(x|\theta) = \theta \exp(-\theta x)$$

the observed data are the following:

```
1.07 0.88 0.66 0.55 1.15 0.65 3.45 3.55 3.51 0.48
```

(a) Find an exact pivot;

(b) Use the pivot to construct the 95% confidence interval for θ;

(c) Apply your interval to this data set.

Answer:

(a) $\sum_{i=1}^n X_i$ is then Gamma distributed $S \sim Gamma(n, \frac{1}{\theta})$*
$\boxed{\theta \sum_{i=1}^n X_i}$ is then Gamma distributed with parameters $(n, 1)$. Since this distribution is independent of θ, it can be used as a pivot.

(b) $\boxed{2\theta \sum_{i=1}^n X_i}$ is then also a pivot and Gamma$(n, 2)$ distributed, or $\boxed{\chi^2_{2n}}$ distributed. We choose to construct a non-symmetric 95% confidence interval (C.I.) as follows†:

$$0.95 = P\left(\chi^2_{2n,0.025} \leq 2\theta \sum_{i=1}^n X_i \leq \chi^2_{2n,0.975}\right),$$

or (since $\sum X_i$ is positive)

$$\left(\frac{\chi^2_{2n,0.025}}{2\sum_{i=1}^n X_i} \leq \theta \leq \frac{\chi^2_{2n,0.975}}{2\sum_{i=1}^n X_i}\right).$$

*In this parametrization, the random variable with pdf $f(t|n,\theta) = \frac{1}{\Gamma(n)}\theta\,(t\theta)^{n-1}\,\mathrm{e}^{-t\theta}$ is said to be Gamma$(n, 1/\theta)$ distributed. I.e., θ is the rate parameter and $1/\theta$ is the scale parameter. This is the parametrization we used in class, only with $1/\beta$ in place of this θ.

†Other ways of constructing this C.I. are possible.

(c) Thus $2\theta\sum_{i=1}^{10} X_i$ is χ^2_{20} distributed. Applying this to the data gives

$$\left(\frac{\chi^2_{20,0.025}}{2\sum_{i=1}^n X_i} \leq \theta \leq \frac{\chi^2_{20,0.975}}{2\sum_{i=1}^n X_i}\right)$$

$$\boxed{[\,0.3, 1.07\,]}$$

Question 8.2. Consider an i.i.d. sample of random variables with density function

$$f(x|\sigma) = \frac{1}{2\sigma}\exp(-\frac{|x|}{\sigma})$$

Use the approximate pivot method to construct a $100(1-\alpha)\%$ confidence interval for σ.

Answer:

Let X be distributed according to the above `pdf`. Then[‡]:

$$\begin{aligned} \mathrm{E}\,X &= 0 \\ \mathrm{E}\,X^2 &= 2\sigma^2 \\ \mathrm{Var}(X) &= 2\sigma^2 - 0^2 \\ &= 2\sigma^2 \\ \mathrm{std}(X) &= \sqrt{2}\sigma. \end{aligned}$$

Asymptotically,

$$\frac{\bar{X} - \mathrm{E}\,\bar{X}}{\mathrm{std}(\bar{X})} \sim N(0,1),$$

so that

$$\frac{\bar{X} - \frac{1}{n}\cdot 0}{(1/\sqrt{n})\cdot\sqrt{2}\sigma} = \sqrt{\frac{n}{2}}\frac{\bar{X}}{\sigma} \sim N(0,1).$$

Then

$$\begin{aligned} -z_{\alpha/2} \leq \sqrt{\frac{n}{2}}\bar{X}\frac{1}{\sigma} &\leq z_{\alpha/2} \\ -z_{\alpha/2}\sqrt{\frac{2}{n}}\frac{1}{\bar{X}} \leq \frac{1}{\sigma} &\leq z_{\alpha/2}\sqrt{\frac{2}{n}}\frac{1}{\bar{X}} \end{aligned}$$

Therefore, the $100(1-\alpha)\%$ confidence interval for σ is given by

$$\sigma \in \left(-\infty, -\sqrt{\frac{n}{2}}\frac{\bar{X}}{z_{\alpha/2}}\right) \cup \left(\sqrt{\frac{n}{2}}\frac{\bar{X}}{z_{\alpha/2}}, \infty\right)$$

[‡]We obtained these results in homework 4

Question 8.3. A sample of students from an introductory psychology class were polled regarding the number of hours they spent studying for the last exam. All students anonymously submitted the number of hours on a 3 by 5 card. There were 24 individuals in the one section of the course polled. The data was used to make inferences regarding the other students taking the course. There data are below:

```
4.5   7.5   22    9    7    10.5   14.5  15    9
19    9     3.5   8    11    2.5    5    9     8.5
7.5   18    20   14    20    8
```

(a) Obtain a confidence interval based on central limit theorem;

(b) Obtain a confidence interval based on T-distributions;

(c) Obtain a confidence interval based on bootstrapping with $B = 10,000$.

Answer:

(a) Let X be the number of hours spent on studying for the exam. We are looking for the confidence interval for μ - the average number of hours spent on studying for the exam. Based on CLT

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathfrak{D}} N(0, 1).$$

The $100(1-\alpha)\%$ confidence interval for μ is then

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Where we use the sample standard deviation s for an estimate of σ[§].

(b) Similarly, using the t_{23}-distribution[¶],

$$\bar{X} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

(c) To use the bootstrap method we randomly select $N = 24$ elements with replacement from our sample set. We do this $B = 10,000$ times, taking the mean, μ_i^B each time. Then we can find a $(1-\alpha)$ confidence interval for μ by letting $\hat{\mu} = \bar{X}$, $V^B(\hat{\mu})$ = the sample variance of $\hat{\mu}_j^B$. Then:

$$\hat{\mu} - z_{\frac{\alpha}{2}} \sqrt{V^B(\hat{\mu})} \leq \mu \leq \hat{\mu} + z_{\frac{\alpha}{2}} \sqrt{V^B(\hat{\mu})}$$

And our CI for μ is given by

$$\left[\hat{\mu} - z_{\frac{\alpha}{2}} \sqrt{V^B(\hat{\mu})}, \hat{\mu} + z_{\frac{\alpha}{2}} \sqrt{V^B(\hat{\mu})}\right]$$

We could write our own function in R to compute this, or we could make use of the *bootstrap* library, and compute this as follows (here we let $\alpha = 0.05$):

[§] For example, if $\alpha = 0.05$, C.I. is given by $[8.676935, 13.1564]$.

[¶] For example, if $\alpha = 0.05$, C.I. is given by $[8.552726, 13.28061]$.

```r
library(bootstrap)
mu.boot <- bootstrap(x, nboot=10000,theta=mean)
conf.int = quantile(mu.boot$thetastar ,c(.025,.975))
```

For example, in one trial, we got a C.I. of $[8.79, 13.12]$ for $\alpha = 0.05$[‖].

Question 8.4. The Poisson distribution has been used by traffic engineers as a model for light traffic, based on the rationale that if the rate is approximately constant and the traffic is light (so the individual cars move independently of each other), the distribution of counts of cars in a given time interval or space area should be nearly Poisson. The following table shows the number of right turns during 300 3-min intervals at a specific intersection.

n	Frequency
0	14
1	30
2	36
3	68
4	43
5	43
6	30
7	14
8	10
9	6
10	4
11	1
12	1
13+	0

(a) Use the pivot method to construct a $(1 - \alpha)$ confidence interval of the rate;

(b) Use variance stabilization method to construct a $(1-\alpha)$ confidence interval of the rate;

(c) Plug in the data and calculate the 95% CI by both methods. Which one do you prefer?

[‖] As expected, our confidence intervals in part (c) tended to be tighter than the ones obtained in parts (a) or (b).

Answer:

Let $X_1, X_2, \cdots, X_n$ be Poisson(λ) distributed. Consider the `MLE` estimate of λ:

$$\hat{\lambda} = \frac{\sum_i X_i}{n} = \bar{X}$$

Then it's easily seen that

$$\mathrm{E}\hat{\lambda} = \lambda$$
$$\mathrm{Var}(\hat{\lambda}) = \frac{1}{n}\lambda$$

By `CLM`

$$\frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \xrightarrow{\mathfrak{D}} N(0,1).$$

This expression can be used as a pivot quantity for part (a):

(a)

$$-z_{\alpha/2} \leq \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \leq z_{\alpha/2}$$

To isolate λ, we may consider solving the resulting quadratic equation (in $\sqrt{\lambda}$).

Alternatively, we can proceed as in class (where we used the $\hat{p}$ estimate of the Bernoulli parameter to estimate the variance of a sum of Bernoullis. I.e., we used $n\hat{p}(1-\hat{p})$ instead of $np(1-p)$ for the variance.)

Thus we use $\bar{X}$ in place of λ in the denominator:

$$-z_{\alpha/2} \leq \frac{\bar{X} - \lambda}{\sqrt{\bar{X}/n}} \leq z_{\alpha/2}$$

Then our $100(1-\alpha)\%$ confidence interval is given by:

$$\bar{X} - z_{\frac{\alpha}{2}}\sqrt{\bar{X}/n} \leq \lambda \leq \bar{X} + z_{\frac{\alpha}{2}}\sqrt{\bar{X}/n}$$

(b) Let

$$\begin{aligned} g(x) &= \int_0^x \frac{1}{\sqrt{\lambda/n}}\, d\lambda \\ &= \sqrt{n}\int_0^x \frac{1}{\sqrt{\lambda}}\, d\lambda \\ &= 2\sqrt{nx} \end{aligned}$$

According to CLM

$$\sqrt{\frac{n}{\lambda}}(\hat{\lambda}-\lambda) \xrightarrow{\mathfrak{D}} N(0,1).$$

Using the above Variance Stabilizing Transform $g(\cdot)$ we have that by the delta method

$$\frac{\sqrt{n/\lambda}(g(\hat{\lambda})-g(\lambda))}{|g'(\lambda)|} \xrightarrow{\mathfrak{D}} N(0,1).$$

or

$$\frac{\sqrt{n/\lambda}(2\sqrt{n\hat{\lambda}}-2\sqrt{n\lambda})}{\sqrt{n/\lambda}} \sim N(0,1)$$

$$2\sqrt{n\hat{\lambda}}-2\sqrt{n\lambda} \sim N(0,1)$$

i.e.,

$$-z_{\alpha/2} \leq 2\sqrt{n}(\sqrt{\bar{X}}-\sqrt{\lambda}) \leq z_{\alpha/2}$$

$$\sqrt{\bar{X}}-\frac{z_{\alpha/2}}{2\sqrt{n}} \leq \sqrt{\lambda} \leq \sqrt{\bar{X}}+\frac{z_{\alpha/2}}{2\sqrt{n}}$$

Assuming $\sqrt{\bar{X}} \geq \frac{z_{\alpha/2}}{2\sqrt{n}}$ and $\lambda \geq 0$, as is the case here, we can then write this as

$$\left(\sqrt{\bar{X}}-\frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2 \leq \lambda \leq \left(\sqrt{\bar{X}}+\frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2$$

and our confidence interval is

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

(c) Here is a comparison of parts (a) and (b). The lengths of the coverage intervals are the same for both methods.

```
# given data:
n = c(0:13)
freq = c(14, 30, 36, 68, 43, 43, 30, 14, 10,  6,  4,  1,  1,  0)
alpha = 0.05

# estimate of the poisson parameter:
X.bar = sum( n * freq ) / sum(freq)   #ans:3.893333

#part(a) - C.I. based on the pivot method
lb.pivot <- X.bar - sqrt( X.bar ) * qnorm( 1-alpha/2 ) / sqrt(sum( freq ))
```

```r
ub.pivot <- X.bar + sqrt( X.bar ) * qnorm( 1-alpha/2 ) / sqrt(sum( freq ))
   ### ans: C.I.pivot = [3.670054, 4.116613]

#part(b) - C.I. based on variance stabilization
lb.vst <- (sqrt( X.bar ) - qnorm( 1-alpha/2 )/(sqrt( sum( freq ) )*2) )^2
ub.vst <- (sqrt( X.bar ) + qnorm( 1-alpha/2 )/(sqrt( sum( freq ))*2 ) )^2
   ### ans: C.I.vst = [3.673255, 4.119814]

# the lengths of the coverage intervals turn out to be the same:
ub.vst - lb.vst       #ans: 0.4465584
ub.pivot - lb.pivot   #ans: 0.4465584
```

Given that the intervals are the same size, we prefer the method which made fewer assumptions. For the pivot method, we estimated the variance based on the mean, so we prefer the VST method.