S620 - Introduction To Statistical Theory - Homework 8

Enrique Areyan
May 6, 2014

1. Let $X_1, X_2, \ldots \sim Pois(\mu)$, where $\mu \in (0, \infty)$. Then, for $k = 0, 1, 2, \ldots$ we have that $P(X_i = j) = \dfrac{e^{-\mu}\mu^j}{j!}$. Show that the MLE of μ is the sample mean.

   Solution: We wish to maximize the likelihood function, i.e.,

$$L(\mu) = L_x(\mu) = f(x;\mu) = \left(\frac{e^{-\mu}\mu^{x_1}}{x_1!}\right)\left(\frac{e^{-\mu}\mu^{x_2}}{x_2!}\right)\cdots\left(\frac{e^{-\mu}\mu^{x_n}}{x_n!}\right) = \frac{e^{-n\mu}\mu^{\sum\limits_{i=1}^{n} x_i}}{x_1!x_2!\cdots x_n!}$$

   Taking the derivative we get:

$$\frac{dL(\mu)}{d\mu} = \frac{d}{d\mu}\left(\frac{e^{-n\mu}\mu^{\sum\limits_{i=1}^{n} x_i}}{x_1!x_2!\cdots x_n!}\right) = (x_1!x_2!\cdots x_n!)^{-1}\left(-ne^{-n\mu}\mu^{\sum\limits_{i=1}^{n} x_i} + e^{-n\mu}\mu^{\sum\limits_{i=1}^{n} x_i - 1}\sum_{i=1}^{n} x_i\right)$$

   Setting the derivative equal to zero we get the critical point $\hat{\mu}$:

$$\begin{aligned}
\frac{dL(\hat{\mu})}{d\hat{\mu}} = 0 \quad &\iff \quad (x_1!x_2!\cdots x_n!)^{-1}\left(-ne^{-n\hat{\mu}}\hat{\mu}^{\sum\limits_{i=1}^{n} x_i} + e^{-n\hat{\mu}}\hat{\mu}^{\sum\limits_{i=1}^{n} x_i - 1}\sum_{i=1}^{n} x_i\right) = 0 \\
&\iff \quad ne^{-n\hat{\mu}}\hat{\mu}^{\sum\limits_{i=1}^{n} x_i} = e^{-n\hat{\mu}}\hat{\mu}^{\sum\limits_{i=1}^{n} x_i - 1}\sum_{i=1}^{n} x_i \\
&\iff \quad n = \hat{\mu}^{-1}\sum_{i=1}^{n} x_i \\
&\iff \quad \hat{\mu} = \left(\sum_{i=1}^{n} x_i\right)/n, \text{ i.e., the sample mean}
\end{aligned}$$

   An argument using second derivative shows that this is indeed the MLE.

2. Consider the following data:

Partition	J	O_j
E_1	0	57
E_2	1	203
E_3	2	383
E_4	3	525
E_5	4	532
E_6	5	408
E_7	6	273
E_8	7	139
E_9	8	45
E_{10}	9	27
E_{11}	10	10
E_{11}	11	4
E_{11}	12	0
E_{11}	13	1
E_{11}	14	1

   Where $E_{11} = \{j : j \geq 10\}$

   Test the null hypothesis that the data were drawn from a Poisson distribution, i.e.:

$$\Theta = \{\theta \in \mathbb{R}^{11} : \theta_j \geq 0, \sum_{j=1}^{11}\theta_j = 1\}, \qquad \Theta_0 = \{\theta \in \Theta : \theta_j = \frac{\mu^j e^{-\mu}}{j!}, j = 1, \ldots, 10, \mu \in (0, \infty)\}$$

   Compute the MLE of μ

Solution: The following table summarizes all the data needed to test the given null hypothesis: ($n = 2608$)

Partition	J	o_j	o_j/n	$o_j * j$	$\bar{\theta}_j$	$\bar{e}_j = \bar{\theta}_j * n$	$log(o_j/\bar{e}_j)$
E_1	0	57	0.021855828	0	0.020826083	54.31442488	2.750901199
E_2	1	203	0.077837423	203	0.080629203	210.2809617	-7.153429752
E_3	2	383	0.146855828	766	0.156079959	407.0565318	-23.33124373
E_4	3	525	0.201303681	1575	0.20142374	525.3131137	-0.313020406
E_5	4	532	0.20398773	2128	0.194955474	508.4438755	24.09356317
E_6	5	408	0.156441718	2040	0.150955937	393.6930837	14.56378199
E_7	6	273	0.104677914	1638	0.097405553	254.0336826	19.65734539
E_8	7	139	0.053297546	973	0.053872911	140.5005529	-1.492511221
E_9	8	45	0.017254601	360	0.026071453	67.99434828	-18.57429443
E_{10}	9	27	0.010352761	243	0.011215212	29.24927295	-2.160481972
E_{11}	10	10	0.006134969	100	0.006564475	17.12015192	-5.376711513
E_{11}	11	4		44			
E_{11}	12	0		0	$\mathbf{T_n} = 2\sum log(o_j/\bar{e}_j)$	=	5.327797435
E_{11}	13	1		13			
E_{11}	14	1		14			
			$\sum o_j * j =$	10097			
			$\hat{\mu} = \sum o_j * j/2608 =$	3.87154908			

Where $\bar{\theta}_j$ are the probabilities under the null hypothesis, i.e.: $\bar{\theta}_j = \dfrac{e^{-\hat{\mu}}\hat{\mu}^j}{j!}$, where $\hat{\mu} = 3.87154908$

Also, $\bar{\theta}_{10} = 1 - \sum_{i=0}^{9} \bar{\theta}_j$ and $o_{10}/n = (\sum_{i=10}^{14} o_j)/n$

Since $dim(\Theta) = 11 - 1 = 10$ and $dim(\Theta_0) = 1$, the appropriate degrees of freedoms are $10 - 1 = 9$. Assuming a significance level $\alpha = 0.05$, we would not reject the null hypothesis:

$$P(\chi^2(9) \leq \mathbf{T_n}) = P(\chi^2(9) \leq 5.327797435) = 0.195151435 > \alpha$$

8.1 (a) Let $X = (X_1, \ldots, X_n)$ be a random sample of size $n \geq 3$ from the exponential distribution of mean $1/\theta$. Find a sufficient statistic $T(X)$ for θ and write down its density. Obtain a maximum likelihood estimator $\hat{\theta}_n$ based on the sample of size n for θ and show that it is biased, but that a multiple of it is not.

Solution: The density of a single exponential random variable with parameter θ is $P(X_i = x_i) = \theta e^{-\theta x_i}$. Let $X_1, \ldots, X_n \sim Exp(\theta)$. The joint distribution is:

$$f(X;\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta\left(\sum\limits_{i=1}^{n} x_i\right)}$$

Then $T(X) = \sum\limits_{i=1}^{n} x_i$ is a sufficient statistic for θ by the factorization theorem, letting $h(x) = \theta^n$.

Now, let us find the distribution of $T(X)$. To this end, consider the following facts:

i) If $X \sim Gamma(1, \beta)$ then $X \sim Exp(\beta)$. In other words $Gamma(1, \beta) = Exp(\beta)$

ii) If $X_1 \sim Gamma(\alpha_1, \beta)$ and $X_2 \sim Gamma(\alpha_2, \beta)$, X_1 independet of X_2, then

$$X_1 + X_2 \sim Gamma(\alpha_1 + \alpha_2, \beta)$$

It follows from $i)$ and $ii)$ that $T(X) \sim Gamma(n, \theta)$, i.e., $f(T(X);\theta) = \dfrac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}$

Finally, let us find a maximum likelihood estimator $\hat{\theta}_n$ based on the sample of size n for θ by solving the likelihood equation:

$$\frac{d}{d\theta} l(\hat{\theta}) = 0 \iff \frac{d}{d\theta} log\left(\hat{\theta}^n e^{-\hat{\theta}\left(\sum\limits_{i=1}^{n} x_i\right)}\right) = 0 \iff \frac{d}{d\theta}\left(nlog(\hat{\theta}) + log(e^{-\hat{\theta}\left(\sum\limits_{i=1}^{n} x_i\right)})\right) = 0 \iff$$

$$\frac{d}{d\theta}\left(nlog(\hat{\theta}) - \hat{\theta}\left(\sum_{i=1}^{n} x_i\right)\right) = 0 \iff \frac{n}{\hat{\theta}} - \sum_{i=1}^{n} x_i = 0 \iff \hat{\theta} = n/\sum_{i=1}^{n} x_i \iff \hat{\theta} = n/T(X)$$

This estimator is biased:

$$E_\theta[\hat{\theta}] = E\left[n/T(X)\right] = \int_0^\infty \frac{n}{t}\frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)} dt = \int_0^\infty \frac{n\theta}{n-1}\frac{\theta^{n-1} t^{(n-1)-1} e^{-\theta t}}{\Gamma(n-1)} dt = \frac{n\theta}{n-1}\int_0^\infty \frac{\theta^{n-1} t^{(n-1)-1} e^{-\theta t}}{\Gamma(n-1)} dt = \frac{n}{n-1}\theta$$

Hence, we can correct this estimator to obtain an unbiased estimator, i.e., the estimator $\frac{n-1}{n}\hat{\theta} = \frac{n-1}{\sum_{i=1}^{n} x_i}$ is unbiased for θ. Check: $E[\frac{n-1}{n}\hat{\theta}] = \frac{n-1}{n}E[\hat{\theta}] = \frac{n-1}{n}\frac{n}{n-1}\theta = \theta$

(b) Calculate the Cramer-Rao Lower Bound for the variance of an unbiased estimator, and explain why you would not expect the bound to be attained in this example. Confirm this by calculating the variance of your unbiased estimator and comment on its behavior as $n \to \infty$.

Solution: Let us compute the fisher information $i(\theta)$ for our unbiased estimator. First note that $i_1(\theta) = E_\theta[-\frac{d^2}{d\theta^2}log(f_1(x_1;\theta))] = E_\theta\left[\frac{1}{\theta^2}\right]$, i.e., the fisher information for a single exponential random variable. Since we have n in pendent i.i.d exponential r.v.s, it follows:

$$i(\theta) = E_\theta\left[\frac{n}{\theta^2}\right] = E_\theta\left[\frac{n}{\left(\frac{(n-1)}{T}\right)^2}\right] = \frac{n}{(n-1)^2}E_\theta[T^2]$$

Recall that $T \sim Gamma(n,\theta)$. We can compute its second moment:

$$E[T^2] = \int_0^\infty t^2\frac{\theta^n t^{n-1}e^{-\theta t}}{\Gamma(n)}dt = \int_o^\infty \frac{\theta^n t^{n+1}e^{-\theta t}}{\Gamma(n)}dt = \frac{(n+1)n}{\theta^2}\int_0^\infty \frac{\theta^{n+2}t^{n+1}e^{-\theta t}}{\Gamma(n+2)}dt = \frac{(n+1)n}{\theta^2}$$

Hence,

$$i(\theta) = \frac{n}{(n-1)^2}\frac{(n+1)n}{\theta^2} = \frac{n^2(n+1)}{(n-1)^2\theta^2} \Longrightarrow \text{Cramer-Rao lower bound is } \frac{1}{i(\theta)} = \boxed{\frac{(n-1)^2}{n^2(n+1)}\theta^2}$$

Next, let us calculate the variance of the unbiased estimator $\frac{n-1}{T}$:

$$Var\left[\frac{n-1}{T}\right] = E\left[\left(\frac{n-1}{T}\right)^2\right] - E\left[\frac{n-1}{T}\right]^2 = E\left[\left(\frac{n-1}{T}\right)^2\right] - \theta^2$$

where,

$$E\left[\left(\frac{n-1}{T}\right)^2\right] = \int_0^\infty \left(\frac{n-1}{t}\right)^2\frac{\theta^n t^{n-1}e^{-\theta t}}{\Gamma(n)}dt = \frac{(n-1)^2\theta^2}{(n-1)(n-2)}\int_0^\infty \frac{\theta^{n-2}t^{(n-2)-1}e^{-\theta t}}{\Gamma(n-2)}dt = \frac{n-1}{n-2}\theta^2$$

Hence,

$$Var\left[\frac{n-1}{T}\right] = \frac{n-1}{n-2}\theta^2 - \theta^2 = \theta^2\left[\frac{n-1}{n-2} - 1\right] = \boxed{\frac{1}{n-2}\theta^2}$$

Therefore, the variance of the unbiased estimators approaches 0 as n approaches infinity, which means that we would not expect the Cramer-Rao Lower Bound to be attained in this example

8.2 You are given a coin, which you are going to test for fairness. Let the probability of a head be p, and consider testing $H_0 : p = 1/2$ against $H_1 : p > 1/2$.

(i) You toss the coin 12 times and observe nine heads, three tails. Do you reject H_0 in a test of size $\alpha = 0.05$?

Solution: Let X = number of heads in 12 coin tosses. Then, under H_0, $X \sim Bin(12, 1/2)$.

$$\begin{aligned}
P(X \geq 9) = \sum_{i=9}^{12}\binom{12}{i}\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^{12-i} = \left(\frac{1}{2}\right)^{12}\sum_{i=9}^{12}\binom{12}{i} &= \left(\frac{1}{2}\right)^{12}\left(\binom{12}{9}+\binom{12}{10}+\binom{12}{11}+\binom{12}{12}\right) \\
&= \frac{1}{4096}(220+66+12+1) \\
&= \frac{299}{4096} \\
&= \boxed{0.07299804687}
\end{aligned}$$

Hence we **fail** to reject H_0 at $\alpha = 0.05$

(ii) You toss the coin until you observe the third head, and note that this occurs on the 12th toss. Do you reject H_0 in a test of size $\alpha = 0.05$?

Solution: Let Y = number of tosses until third head. Then, under H_0, $Y \sim NegativeBin(3, 1/2)$.

$$\begin{aligned}
P(Y \geq 9) = 1 - P(Y \leq 8) = 1 - \left(\sum_{i=0}^{8} P(Y = i)\right) &= 1 - \left(\sum_{i=0}^{8} \binom{i+2}{i}\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^3\right) \\
& 1 - \left(\sum_{i=0}^{8} \binom{i+2}{i}\left(\frac{1}{2}\right)^{i+3}\right) \\
&= 1 - \left(\frac{1}{2^3} + \frac{3}{2^4} + \frac{6}{2^5} + \frac{10}{2^6} + \frac{15}{2^7} + \frac{21}{2^8} + \frac{28}{2^9} + \frac{36}{2^{10}} + \frac{45}{2^{11}}\right) \\
&= 1 - \frac{256 + 384 + 384 + 320 + 240 + 168 + 112 + 72 + 45}{2048} \\
&= 1 - \frac{1981}{2048} \\
&= \frac{67}{2048} \\
&= \boxed{0.032714844}
\end{aligned}$$

Hence we **reject** H_0 at $\alpha = 0.05$

Note that even though the data is the same in case (i) and (ii), the classical approach changes with the experimental setup.

In (i), the number of heads has a binomial distribution, while in (ii) the number of tosses performed has a negative binomial distribution. What is the likelihood function for p in the two cases? The likelihood principle would demand that identical inference about p would be drawn in the two cases. Comment.

Solution: The likelihood function for p in the two cases is:

(i) $L_x(p) = \binom{12}{x}p^x(1-p)^{12-x}$, for $x = 9$ we get, $L_9(p) = \binom{12}{9}p^9(1-p)^3 = 220p^9(1-p)^3$

(ii) $L_y(\theta) = \binom{y+2}{y}p^y(1-p)^{11-y+1}$, for $y = 9$ we get, $L_9(p) = \binom{11}{9}p^9(1-p)^3 = 55p^9(1-p)^3$

These two functions are proportional to each other. According to the likelihood principle, identical conclusions regarding p should be drawn from x and y. For example, if we were to use a likelihood ratio test, then L_1/L_0 would be the same likelihood ratio function as the following calculations show:

(i) $L_1(\hat{p}) = f(X; \hat{p}) = \binom{12}{k}\hat{p}^k(1-\hat{p})^{12-k}$, and $L_0(1/2) = f(X; 1/2) = \binom{12}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{12-k} = \binom{12}{k}\left(\frac{1}{2}\right)^{12}$.

Thus,

$$L_1/L_0 = \frac{\binom{12}{k}\hat{p}^k(1-\hat{p})^{12-k}}{\binom{12}{k}\left(\frac{1}{2}\right)^{12}} = 2^{12}\hat{p}^k(1-\hat{p})^{12-k}$$

(ii) $L_1(\hat{p}) = f(Y; \hat{p}) = \binom{k+2}{k}\hat{p}^k(1-\hat{p})^{11-k+1} = \binom{k+2}{k}\hat{p}^k(1-\hat{p})^{12-k}$, and $L_0(1/2) = f(Y; 1/2) = \binom{k+2}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{11-k+1} = \binom{k+2}{k}\left(\frac{1}{2}\right)^{12}$.

Thus,

$$L_1/L_0 = \frac{\binom{k+2}{k}\hat{p}^k(1-\hat{p})^{12-k}}{\binom{k+2}{k}\left(\frac{1}{2}\right)^{12}} = 2^{12}\hat{p}^k(1-\hat{p})^{12-k}$$

In either case $\boxed{L_1/L_0 = 2^{12}\hat{p}^k(1-\hat{p})^{12-k}}$, so inferences using this function will yield the same results as required by the likelihood principle.