Math 536 Assignment 2

Gustavo Esparza

9/18/2019

1

Suppose that $\hat{\theta}$ is the MLE for a parameter θ. Let $t(\theta)$ be a function of θ that possesses a unique inverse that is, if $\beta = t(\theta)$, then $\theta = t^{-1}(\beta)$. Show that $t(\hat{\theta})$ is the MLE of $t(\theta)$.

We will begin by letting $\beta = t(\theta)$, then $\theta = t^{-1}(\beta)$. It is assumed that $\hat{\theta}$ is the MLE, and thus $L(\hat{\theta}) \geq L(\theta)\ \forall\theta \in$ domain. Then, we have a value $\hat{\beta}$ equal to our function evaluated at $\hat{\theta}$, $t(\hat{\theta})$. Then, using our relation from θ to β, we can observe the following result

$$L(t^{-1}(\beta)) = L(\theta) \leq L(\hat{\theta}) = L(t^{-1}(\hat{\beta})) \to L(t^{-1}(\beta)) \leq L(t^{-1}(\hat{\beta}))$$

Thus, we can finally observe that $\hat{\beta}$ is the MLE of β, as it's likelihood is maximum at $\hat{\beta} = t(\hat{\theta})$. It follows that the mle of $t(\theta)$ is $t(\hat{\theta})$.

2

A random sample of n items is selected from a large number of items produced by a certain production line in one day. Find the MLE of the ratio R, the proportion of defective items divided by the proportion of good items.

We have a random sample of n items selected from a larger sample N.Each selected item, y_i is either defective or not defective with probability p. Thus, our sample of y_i values come from a binomial distribution with parameters n and p and is defined as follows:

$$Y_i = \begin{cases} 1 & \text{if} \quad \text{defective} \\ 0 & \text{if} \quad \text{Not Defective} \end{cases}$$

From this distribution, we can see that $\sum_{i=1}^{n} Y_i = D$, where D represents the total number of defective items selected from the sample. We know that the MLE of our is $\frac{\sum_{i=1}^{n} Y_i}{n} = \frac{D}{n}$.
For R, the ratio of defective items to non-defective items, we have $\frac{p}{1-p}$. Thus, by the Invariance Property of MLEs, we have the following result:

$$\hat{R}_{MLE} = \frac{\hat{p}}{1-\hat{p}} = \frac{D/n}{1-D/n} = \frac{D}{n-D}$$

3

Suppose that $X_1, \ldots, X_n$ form a random sample from a uniform distribution on the interval $(0, \theta)$. However, suppose that instead we write the pdf as follows:

$$f(x|\theta) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & otherwise \end{cases}$$

Find the MLE for θ.

From our PDF definition, we have the following Likelihood Function:

$$L(\theta) = \frac{1}{\theta^n}\ ,\ 0 < x_i < \theta$$

In order to maximize $L(\theta)$, θ must be minimized. Then, θ must be greater than the largest x_i in our sample. However, we can not be equal to $X_{(n)}$ in this case since our bounds are an open interval. Since there is an infinite amount of options for $\hat{\theta}$ being closest to x_i while still being larger, this particular distribution does not have an MLE.

4

Suppose that $X_1, \dots, X_n$ form a random sample from a uniform distribution on the interval $[\theta, \theta+1]$, where the value of the parameter θ is unknown ($\infty < \theta < \infty$), which pdf:

$$f(x|\theta) = \begin{cases} 1 & \theta \le x \le \theta+1 \\ 0 & otherwise \end{cases}$$

Find the MLE for θ.

We have $L(\theta) = 1^n = 1\ ,\ \theta \le x_i \le \theta + 1$. Rearranging our bounds, we have the following results:

$$\theta \le x_i \to \theta \le x_{(1)}$$
$$\theta + 1 \ge x_i \to \theta \ge x_i - 1 \to \theta \ge x_{(n)} - 1$$

Thus, we have a final result of
$x_{(n)} - 1 \le \hat{\theta} \le x_{(1)}$. There are multiple options for a $\hat{\theta}$ that fits this constraint, with one such option being $\frac{x_{(1)}+x_{(n)}}{2}$. The average of our highest and lowest x_i values ensures that our MLE will always fall into our inequality expression for $\hat{\theta}$.

5

Suppose that $X_1, \dots, X_n$ form a random sample from a uniform distribution on the interval $[\theta_1, \theta_2]$, where both θ_1 and θ_2 are unknown ($\infty < \theta_1 < \theta_2 < \infty$). Find the MLEs of θ_1 and θ_2. Note for clarification: $\theta_1 \le x \le \theta_2$.

For this distribution, we have the following Likelihood Function:

$$L(\theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n}\ ,\ \theta_1 \le x_i \le \theta_2$$

Here, we need $\theta_2 - \theta_1$ to be minimized for $\theta_1 \le x_{(1)} \le \dots \le x_{(n)} \le \theta_2$ in order to maximize our likelihood function. Then $(\theta_2 - \theta_1)$ is minimized when $\theta_1 = x_{(1)}$ and $\theta_2 = x_{(n)}$.
Thus, $\hat{\theta}_1 = x_{(1)}$ and $\hat{\theta}_2 = x_{(n)}$

6

Consider a distribution for which the pdf or pmf is $f(x|\theta)$, where the parameter θ is a k dimensional vector belonging to some parameter space Ω. It is said that the family of distributions indexed by the values of θ in Ω is a k parameter exponential family, or a k parameter Koopman Darmois family, if $f(x|\theta)$ can be written as follows for $\theta \in \Omega$ and all values of x

$$f(x\theta) = a(\theta)b(x)exp[\sum_{i=1}^{n} c_i(\theta)d_i(x)]$$

Here a and $c_1, \ldots, c_k$ are arbitrary functions θ, b and $d_1, \ldots, d_k$ are arbitrary functions x. Suppose now that $X_1, \ldots, X_n$ form a random sample from a distribution which belongs to a k parameter exponential family of this type, and define the kstatistics $T_1, \ldots, T_k$ as follows:

$$T_i = \sum_{i=1}^{n} d_i(x_j) \text{ for } i = 1, \ldots, k$$

Show that the statistics $T_1, \ldots, T_k$ are jointly sufficient statistics for θ.

Our Joint distribution is defined as follows:

$$\prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} b(x_i) \times (a(\theta))^n \prod_{i=1}^{n} exp[\sum_{i=1}^{n} c_i(\theta)d_i(x)]$$

Using Factorization Theorem to partition our Joint Distribution, we have the following:

$$h(x) = \prod_{i=1}^{n} b(x_i)$$
$$g(t(x), \theta) = (a(\theta))^n exp[\sum_{i=1}^{n} c_i(\theta) \times \sum_{i=1}^{n} d_i(x)]$$

From our Factorization, we can see that within $g(t(x), \theta)$, $\sum_{i=1}^{n} d_i(x)$ is the sufficient statistic for θ. Thus, for $T_1, \ldots, T_k$, we can observe that $\sum_{i=1}^{n} d_i(x)$ are jointly sufficient statistics for θ.

7

Suppose $X_1, \ldots, X_n$ is a random sample from the Binary(p) distribution.

A

Find the MLE of p.

The PDF for the binary distribution is defined as follows:

$$P(x) = p^x(1-p)^{1-x} \ , \ x = 0, 1$$

Then, the likelihood and log-likelihood are as follows:

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} l(p) = \sum_{i=1}^{n} x_i log(p) + \Big(n - \sum_{i=1}^{n} x_i\Big) log(1-p)$$

Taking our derivative with respect to p, we have the following:

$$\begin{aligned}
&\frac{\partial l}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0 \\
&\rightarrow (1-p)\sum_{i=1}^{n} x_i - np + p\sum_{i=1}^{n} x_i = 0 \\
&\rightarrow \sum_{i=1}^{n} x_i - np = 0 \\
&\rightarrow \hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}
\end{aligned}$$

B

Show that the inverse of the Fisher Information is equal to the exact variance of $\hat{p}$.

Here is the computation of our Fisher Information:

$$\begin{aligned}
&\lambda(x|p) = log(p^x(1-p)^{1-x}) = xlog(p) + (1-x)log(1-p) \\
&\lambda'(x|p) = \frac{x}{p} + \frac{x-1}{1-p} \\
&\lambda''(x|p) = \frac{-x}{p^2} + \frac{x-1}{(1-p)^2}
\end{aligned}$$

Now, we can take the expectation of our computation:

$$\begin{aligned}
I(p) &= -E\Big[\frac{-x}{p^2} + \frac{x-1}{(1-p)^2}\Big] \\
&= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} \\
&= \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)} = \frac{1}{Var(p)}
\end{aligned}$$

When considering a sample of size n, we can observe the Fisher Information function for $\hat{p}$ having a multiple of n:

$$I(\hat{p}) = \frac{n}{p(1-p)} = \frac{1}{Var(\hat{p})}$$

Thus, we have seen the inverse of the Fisher Information is equal to the exact variance of $\hat{p}$.

C

Give an approximate 99 confidence interval for p.

For our distribution, we have $E(\bar{x}) = p$ and $V(\bar{x}) = p(1-p)/n$. Thus the 99% (with a corresponding z score of 2.58) Confidence Interval for p is as follows:

$$\hat{P} \pm 2.58\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$