ECON 709 - PS 6

Alex von Hafften*

10/18/2020

1. Let X be distributed Bernoulli $P(X = 1) = p$ and $P(X = 0) = 1 - p$ for some unknown parameter $0 < p < 1$.

(a) Verify the probability mass function can be written as $f(x) = p^x(1-p)^{(1-x)}$.

$$f(1) = p^1(1-p)^{(1-1)} = p = P(X = 1)$$
$$f(0) = p^0(1-p)^{(1-0)} = 1 - p = P(X = 0)$$

(b) Find the log-likelihood function $\ell_n(\theta)$.

$$\begin{aligned}
\ell_n(\theta) &= \sum_{i=1}^n \ln(f(x_i|\theta)) \\
&= \sum_{i=1}^n \ln(p^{x_i}(1-p)^{(1-x_i)}) \\
&= \sum_{i=1}^n [x_i \ln(p) + (1-x_i)\ln(1-p)] \\
&= \ln(p)\sum_{i=1}^n x_i + \ln(1-p)\Big(n - \sum_{i=1}^n x_i\Big)
\end{aligned}$$

(c) Find the MLE $\hat{p}$ for p.

$$\begin{aligned}
\frac{\partial \ell_n}{\partial p} &= 0 \\
\frac{\partial}{\partial p}\left[\ln(p)\sum_{i=1}^n x_i + \ln(1-p)\Big(n - \sum_{i=1}^n x_i\Big)\right] &= 0 \\
\frac{\sum_{i=1}^n x_i}{p} - \frac{\Big(n - \sum_{i=1}^n x_i\Big)}{1-p} &= 0 \\
\sum_{i=1}^n x_i &= pn - p\sum_{i=1}^n x_i + p\sum_{i=1}^n x_i \\
\hat{p} &= \frac{1}{n}\sum_{i=1}^n x_i \\
\hat{p} &= \bar{X}_n
\end{aligned}$$

*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Emily Case, Sarah Bass, and Danny Edgel.

2. Let X be distributed Pareto with density $f(x) = \frac{\alpha}{x^{1+\alpha}}$ for $x \geq 1$. The unknown parameter is $\alpha > 0$.

(a) Find the log-likelihood function $\ell_n(\alpha)$.

$$\begin{aligned}
\ell_n(\alpha) &= \sum_{i=1}^{n} \ln(f(x_i|\alpha)) \\
&= \sum_{i=1}^{n} \ln\left(\frac{\alpha}{x_i^{1+\alpha}}\right) \\
&= \sum_{i=1}^{n} \ln\alpha - \sum_{i=1}^{n} \ln x_i^{1+\alpha} \\
&= n\ln\alpha - (1+\alpha)\sum_{i=1}^{n} \ln x_i
\end{aligned}$$

(b) Find the MLE $\hat{\alpha}_n$ for α.

$$\frac{\partial \ell_n}{\partial \alpha} = 0 \implies \frac{n}{\hat{\alpha}_n} - \sum_{i=1}^{n} \ln x_i = 0 \implies \hat{\alpha}_n = \frac{n}{\sum_{i=1}^{n} \ln x_i}$$

3. Let X be distributed Cauchy with density $f(x) = \frac{1}{\pi(1+(x-\theta)^2)}$ for $x \in \mathbb{R}$. The unknown parameter is θ.

(a) Find the log-likelihood function $\ell_n(\theta)$.

$$\begin{aligned}
\ell_n(\theta) &= \sum_{i=1}^{n} \ln(f(x_i|\theta)) \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{\pi(1+(x_i-\theta)^2)}\right) \\
&= -\sum_{i=1}^{n} \ln(\pi) - \sum_{i=1}^{n} \ln\left(1+(x_i-\theta)^2\right) \\
&= -n\ln(\pi) - \sum_{i=1}^{n} \ln\left(1+(x_i-\theta)^2\right)
\end{aligned}$$

(b) Find the first-order condition for the MLE $\hat{\theta}$ for θ. You will not be able to solve for $\hat{\theta}$.

$$\frac{\partial \ell_n}{\partial \theta} = 0 \implies 0 - \sum_{i=1}^{n} \frac{2(x_i-\theta)(-1)}{1+(x_i-\theta)^2} \implies \sum_{i=1}^{n} \frac{2(x_i-\theta)}{1+(x_i-\theta)^2}$$

4. Let X be distributed double exponential (or Laplace) with density $f(x) = \frac{1}{2}\exp(-|x-\theta|)$ for $x \in \mathbb{R}$. The unknown parameter is θ.

(a) Find the log-likelihood function $\ell_n(\theta)$.

$$\begin{aligned}
\ell_n(\theta) &= \sum_{i=1}^{n} \ln(f(x_i|\theta)) \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{2}\exp(-|x_i-\theta|)\right) \\
&= -\sum_{i=1}^{n} \ln(2) + \sum_{i=1}^{n} \ln(\exp(-|x_i-\theta|)) \\
&= -n\ln(2) - \sum_{i=1}^{n} |x_i-\theta|
\end{aligned}$$

(b) Extra challenge: Find the MLE $\hat{\theta}_n$ for θ. This is challenging as it is not simply solving the FOC due to the nondifferentiability of the density function.

I consider the median x_i as a candidate for the MLE $\hat{\theta}_n$. Without loss of generality, let us consider an ordered sample $x_1 < x_2 < ... < x_{n-1} < x_n$. Consider even n:

$$\ell_n(\theta) = -n\ln(2) - \sum_{i=1}^{n} |x_i-\theta| = -n\ln(2) - \sum_{i=1}^{n} ((x_i-\theta)^2)^{1/2}$$

$\ell_n(\theta)$ is differentiable at $\theta \neq x_i$ for all $i = 1, ..., n$. In particular, it is differentiable at the median, defined as any point strictly between $x_{\lfloor n/2 \rfloor}$ and $x_{\lceil n/2 \rceil}$.

$$\frac{\partial \ell_n}{\partial \theta} = -(1/2)\sum_{i=1}^{n} ((x_i-\theta)^2)^{-1/2}(2(x_i-\theta))(-1) = \sum_{i=1}^{n} \frac{x_i-\theta}{|x_i-\theta|}$$

If $x_i > \theta$, $\frac{x_i-\theta}{|x_i-\theta|} = 1$ and if $x_i < \theta$, $\frac{x_i-\theta}{|x_i-\theta|} = -1$. If θ is any point between $x_{\lfloor n/2 \rfloor}$ and $x_{\lceil n/2 \rceil}$, then there is an equal number of x_i less than $\hat{\theta}_n$ and x_i larger than $\hat{\theta}_n$, so $\frac{\partial \ell_n}{\partial \theta} = 0$. Thus, the median is the MLE $\hat{\theta}_n$.

Consider case when n is odd. Since the median equals $x_{(n+1)/2}$, the $\ell_n(\theta)$ is not differentiable at proposed MLE estimator. Construct a new sample $\{y_1, ..., y_{n-1}\}$ when $y_i = x_i$ and $y_j = x_{j+1}$ for $i = 1, ..., (n-1)/2$ and $j = (n+1)/2, ..., n-1$. This sample omits the median observation $x_{(n+1)/2}$. By the logic above, $\ell_{n-1}(\theta)$ is maximized at any point between $y_{(n-1)/2} = x_{(n-1)/2}$ and $y_{(n+1)/2} = x_{(n+3)/2}$ including $x_{(n+1)/2}$. Now, consider the sample with $x_{(n+1)/2}$. Notice that $\ell_n(\theta) = \ell_{n-1}(\theta) - \ln(2) - |x_{(n+1)/2} - \hat{\theta}_n|$. For any $\hat{\theta}_n \neq x_{(n+1)/2}$, $|x_{(n+1)/2} - \hat{\theta}_n| > 0$, so it reduces the log-likelihood function. If $\hat{\theta}_n = x_{(n+1)/2}$, $|x_{(n+1)/2\rceil} - \hat{\theta}_n| = 0$. Thus, the median is the MLE $\hat{\theta}_n$.

5. Take the Pareto model $f(x) = \alpha x^{-1-\alpha}, x \geq 1$. Calculate the information for α using the second derivative.

The information for α is

$$
\begin{aligned}
I_0 &= -E\left[\frac{\partial^2}{\partial^2\alpha}\log(\alpha X^{-1-\alpha})\bigg|_{\alpha=\alpha_0}\right] \\
&= -E\left[\frac{\partial^2}{\partial^2\alpha}(\log\alpha + (-1-\alpha)\log X)\bigg|_{\alpha=\alpha_0}\right] \\
&= -E\left[\frac{\partial^2}{\partial^2\alpha}(\log\alpha - \log X - \alpha\log X)\bigg|_{\alpha=\alpha_0}\right] \\
&= -E\left[\frac{\partial}{\partial\alpha}(\alpha^{-1} - \log X)\bigg|_{\alpha=\alpha_0}\right] \\
&= -E\left[(-1)\alpha^{-2}\bigg|_{\alpha=\alpha_0}\right] \\
&= \alpha^{-2}
\end{aligned}
$$

6. Take the model $f(x) = \theta\exp(-\theta x), x \geq 0, \theta > 0$.

(a) Find the Cramer-Rao lower bound for θ.

$$
\begin{aligned}
I_0 &= -E\left[\frac{\partial^2}{\partial^2\theta}\log(\theta\exp(-\theta X))\bigg|_{\theta=\theta_0}\right] \\
&= -E\left[\frac{\partial^2}{\partial^2\theta}\log(\theta) - \theta X\bigg|_{\theta=\theta_0}\right] \\
&= -E\left[\frac{\partial}{\partial\theta}\frac{1}{\theta} - \theta\bigg|_{\theta=\theta_0}\right] \\
&= -E\left[-\frac{1}{\theta^2}\bigg|_{\theta=\theta_0}\right] \\
&= \frac{1}{\theta_0^2}
\end{aligned}
$$

Thus, the Cramer-Rao lower bound is $(nI_0)^{-1} = (n\theta_0^{-2})^{-1} = \theta_0^2/n$.

(b) Find the MLE $\hat{\theta}_n$ for θ. Notice that this is a function of the sample mean. Use this formula and the delta method to find the asymptotic distribution for $\hat{\theta}_n$.

The log-likelihood function $\ell_n(\theta)$ is

$$\begin{aligned}
\ell_n(\theta) &= \sum_{i=1}^{n} \ln(f(x_i|\theta)) \\
&= \sum_{i=1}^{n} \ln(\theta \exp(-\theta x_i)) \\
&= \sum_{i=1}^{n} (\ln(\theta) - \theta x_i) \\
&= n\ln(\theta) - \theta \sum_{i=1}^{n} x_i \\
&= n\ln(\theta) - n\theta \bar{X}_n
\end{aligned}$$

Thus, $\hat{\theta}_n$ is

$$\frac{\partial \ell_n}{\partial \theta} = 0 \implies 0 = \frac{n}{\hat{\theta}_n} - n\bar{X}_n \implies \hat{\theta}_n = \frac{1}{\bar{X}_n}$$

By the delta method, $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, V)$ where $V = \left((-1)\left(\frac{1}{\theta_0}\right)^{-2}\right)^2 \sigma^2 = \sigma^2\theta_0^4$ and $\sigma^2 = Var(X) = \frac{1}{\theta_0^2}$. Therefore, $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, \theta_0^2)$

(c) Find the asymptotic distribution for $\hat{\theta}_n$ using the general formula for the asymptotic distribution of MLE introduced in Section 6. Do you find the same answer as in part (b)?

From Section 6, we have that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, I_0^{-1})$$

The information of θ is

$$\begin{aligned}
I_0 &= -E\left[\frac{\partial^2}{\partial^2\theta} \log(\theta \exp(-\theta X))\bigg|_{\theta=\theta_0}\right] \\
&= -E\left[\frac{\partial^2}{\partial^2\theta} \Big(\log(\theta) - \theta X\Big)\bigg|_{\theta=\theta_0}\right] \\
&= -E\left[\frac{\partial}{\partial\theta} \Big(\theta^{-1} - X\Big)\bigg|_{\theta=\theta_0}\right] \\
&= -E\left[-\theta^{-2}\bigg|_{\theta=\theta_0}\right] \\
&= \theta_0^{-2}
\end{aligned}$$

Therefore, we get the same asymptotic distribution:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, \theta_0^2)$$

7. In the Bernoulli model, you found the asymptotic distribution of the MLE in Problem 1(c).

(a) Propose an estimator of V, the asymptotic variance.

In 1(c), we found that $\hat{p} = \bar{X}_n$. So, by the CLT, we know that $\sqrt{n}(\hat{p} - p) \to_d N(0, \sigma^2)$ where $\sigma^2 = Var(X)$. Thus, V should be an estimator for $Var(X)$. Since X is Bernoulli, consider $\bar{X}_n(1 - \bar{X}_n)$.

(b) Show that this estimator is consistent for V as $n \to \infty$.

By the WLLN, $\bar{X}_n \to_p p$. Define g as $g(x) = x(1-x)$. By the continuous mapping theorem, $\bar{X}_n(1 - \bar{X}_n) = g(\bar{X}_n) \to_p g(p) = p(1-p) = Var(X)$.

(c) Propose a standard error $s(\hat{p}_n)$ for the MLE $\hat{p}_n$.[1]

Based on (a) and (b), consider $s(\hat{p}_n) = \frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}$.

8. Consider the MLE for the upper bound of the uniform distribution in the Uniform Boundary example in Section 3. Assume that $\{X_1, ..., X_n\}$ is a random sample from $Uniform[0, \theta]$. The general asymptotic distribution formula in Section 6 does not apply here because $\ell_n(\theta)$ is not differentiable at the MLE. But you can derive the asymptotic distribution using the definition of convergences in distribution. Do so by following the steps below.

(a) Let F_X denote the CDF of $Uniform[0, \theta]$. Calculate $F_X(c)$ for all $c \in \mathbb{R}$ based on the PDF of $Uniform[0, \theta]$.

$$F_X(c) = \int_{\infty}^{c} f_X(x)dx = \begin{cases} 0, c < 0 \\ c/\theta, 0 \le c < \theta \\ 1, \theta \le c \end{cases}$$

Because $\int_0^c \frac{1}{\theta} dx = \frac{c}{\theta}$.

(b) Show that the CDF of $n(\hat{\theta}_n - \theta)$: $F_{n(\hat{\theta}_n - \theta)}(x) = \Pr(\max_{i=1,\dots,n}(n(X_i - \theta)) \le x) = (F_X(\theta + \frac{x}{n}))^n$.

In Section 3, we found that $\hat{\theta}_n = \max_{i=1,\dots,n} X_i$. Because $n(\hat{\theta}_n - \theta) = n(\max_{i=1,\dots,n}(X_i) - \theta) = \max_{i=1,\dots,n}(n(X_i - \theta))$. Thus, $F_{n(\hat{\theta}_n - \theta)}(x) = \Pr(\max_{i=1,\dots,n}(n(X_i - \theta)) \le x)$. Furthermore,

$$\begin{aligned}
\Pr\left(\max_{i=1,\dots,n}(n(X_i - \theta)) \le x\right) &= \Pr(n(X_i - \theta) \le x \; \forall i = 1, ..., n) \\
&= \Pi_{i=1}^n \Pr(n(X_i - \theta) \le x) \\
&= \Pr(n(X_i - \theta) \le x)^n \\
&= \Pr\left(X_i \le \frac{x}{n} + \theta\right)^n \\
&= \left(F_X\left(\theta + \frac{x}{n}\right)\right)^n
\end{aligned}$$

[1] Recall that the standard error is supposed to approximate the variance of $\hat{p}_n$, not that of the variance of $\sqrt{n}(\hat{p}_n - p)$. What would be a reasonable approximation of the variance of $\hat{p}_n$ once you have a reasonable approximation of the variance of $\sqrt{n}(\hat{p}_n - p)$ from part (b)?

(c) Recall that $\lim_{n\to\infty}(1+\frac{y}{n})^n = e^y$ for any $y \in \mathbb{R}$. Derive the limit of $F_{n(\hat{\theta}_n-\theta)}(x)$ for all fixed $x \in \mathbb{R}$.[2]

Fix $x \in \mathbb{R}$. If $x < 0$,

$$\begin{aligned}
\lim_{n\to\infty} F_{n(\hat{\theta}_n-\theta)}(x) &= \lim_{n\to\infty}\left(F_X\left(\theta+\frac{x}{n}\right)\right)^n \\
&= \lim_{n\to\infty}\left(\theta^{-1}\left(\theta+\frac{x}{n}\right)\right)^n \\
&= \lim_{n\to\infty}\left(\theta^{-1}\left(\theta\left(1+\frac{x/\theta}{n}\right)\right)\right)^n \\
&= \lim_{n\to\infty}\left(1+\frac{x/\theta}{n}\right)^n \\
&= e^{x/\theta}
\end{aligned}$$

If $x \geq 0$,

$$\begin{aligned}
\lim_{n\to\infty} F_{n(\hat{\theta}_n-\theta)}(x) &= \lim_{n\to\infty}\left(F_X\left(\theta+\frac{x}{n}\right)\right)^n \\
&= \lim_{n\to\infty}(1)^n \\
&= 1
\end{aligned}$$

(d) Conclude that $n(\hat{\theta}_n - \theta) \to_d -Z$ for Z being an exponential distribution with parameter θ.

Since $F_{n(\hat{\theta}_n-\theta)}(x) \to e^{x/\theta}$, $f_{n(\hat{\theta}_n-\theta)}(x) = \frac{\partial}{\partial x}F_{n(\hat{\theta}_n-\theta)}(x) \to \frac{\partial}{\partial x}e^{x/\theta} = \frac{e^{x/\theta}}{\theta}$. Thus, $f_{n(\hat{\theta}_n-\theta)}(-x) = \frac{e^{-x/\theta}}{\theta}$, which the density function of an exponential distribution with parameter θ. So $n(\hat{\theta}_n - \theta) \to_d -Z$ for Z being an exponential distribution with parameter θ.

9. Take the model $X \sim N(\mu, \sigma^2)$. Propose a test for $H_0 : \mu = 1$ against $H_1 : \mu \neq 1$.

Assuming that σ^2 is unknown, we can use a two-sided t-test by constructing the following t-statistic:

$$T = \frac{|\sqrt{n}(\bar{X}_n - 1)|}{S_X}$$

where $S_X^2 = \frac{1}{n-1}(X_i - \bar{X}_n)^2$. Under the $H_0 : \mu = 1$, $T \sim |t_{n-1}|$. Therefore, $\phi_n(\alpha) = 1(T > t_{\alpha/2,n-1})$ where $t_{\alpha/2,n-1}$ is the $(1-\alpha/2)$ quantile of $t_{n=1}$.

If σ^2 is known, we can use a z-test by replacing S_X with σ in the test statistic:

$$T = \frac{|\sqrt{n}(\bar{X}_n - 1)|}{\sigma}$$

Under the $H_0 : \mu = 1$, $T \sim |N(0,1)|$. Therefore, $\phi_n(\alpha) = 1(T > z_{\alpha/2})$ where $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of a standard normal.

[2]Hint: consider the case where $x < 0$ and the case where $x \geq 0$ separately.

10. Take the model $X \sim N(\mu, 1)$. Consider testing $H_0 : \mu \in \{0, 1\}$ against $H_1 : \mu \notin \{0, 1\}$. Consider the test statistic $T = \min\{|\sqrt{n}\bar{X}_n|, |\sqrt{n}(\bar{X}_n - 1)|\}$ Let the critical value be the $1 - \alpha$ quantile of the random variable $\min\{|Z|, |Z - \sqrt{n}|\}$, where $Z \sim N(0, 1)$. Show that $\Pr(T > c|\mu = 0) = \Pr(T > c|\mu = 1) = \alpha$. Conclude that the size of the test $\phi_n = 1(T > c)$ is α.[3]

Assuming that $\mu = 0$, $X \sim N(0, 1) \implies \sqrt{n}\bar{X}_n \sim N(0, 1)$ by the CLT $\implies |\sqrt{n}\bar{X}_n| \sim |N(0, 1)|$. In addition,

$$\sqrt{n}\bar{X}_n \sim N(0, 1) \implies \sqrt{n}\bar{X}_n - \sqrt{n} \sim N(-\sqrt{n}, 1) \implies |\sqrt{n}(\bar{X}_n - 1)| \sim |N(-\sqrt{n}, 1)|$$

Therefore, for $Z \sim N(0, 1)$,

$$\begin{aligned}
&\Pr(T > c|\mu = 0) \\
&\qquad = \Pr(\min\{|\sqrt{n}\bar{X}_n|, |\sqrt{n}(\bar{X}_n - 1)|\} > c) \\
&\qquad = \Pr(\min\{|Z|, |Z - \sqrt{n}|\} > c) \\
&\qquad = \alpha
\end{aligned}$$

by definition of c. Assuming that $\mu = 1$, $X \sim N(1, 1) \implies \sqrt{n}(\bar{X}_n - 1) \sim N(0, 1)$ by the CLT $\implies |\sqrt{n}(\bar{X}_n - 1)| \sim |N(0, 1)|$. In addition,

$$\sqrt{n}(\bar{X}_n - 1) \sim N(0, 1) \implies \sqrt{n}\bar{X}_n \sim N(\sqrt{n}, 1) \implies |\sqrt{n}\bar{X}_n| \sim |N(\sqrt{n}, 1)|$$

Therefore, because Z is symmetric (Z and $-Z$ have the same distribution),

$$\begin{aligned}
\Pr(T > c|\mu = 1) &= \Pr(\min\{|\sqrt{n}\bar{X}_n|, |\sqrt{n}(\bar{X}_n - 1)|\} > c) \\
&= \Pr(\min\{|Z|, |Z + \sqrt{n}|\} > c) \\
&= \Pr(\min\{|Z|, | - Z + \sqrt{n}|\} > c) \\
&= \Pr(\min\{|Z|, |(-1)(Z - \sqrt{n})|\} > c) \\
&= \Pr(\min\{|Z|, |Z - \sqrt{n}|\} > c) \\
&= \alpha
\end{aligned}$$

by definition of c. Thus, the size of the test $\phi_n = 1(T > c)$ is α.

[3] Use the fact that Z and $-Z$ have the same distribution. This is an example where the null distribution is the same under different points in a composite null. The test $\phi_n = 1(T > c)$ is called a similar test because $\inf_{\theta_0 \in \Theta_0} \Pr(T > c|\theta = \theta_0) = \sup_{\theta_0 \in \Theta_0} \Pr(T > c|\theta = \theta_0)$.