Problem Set 6

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

(Due March 29)

[1] Let $X_1, \ldots, X_n$ represent a random sample from each of the distributions having the following pdfs:

a)

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1,\ 0 < \theta < \infty \\ 0, & \text{otherwise} \end{cases}$$

b)

$$f(x;\theta) = \begin{cases} e^{-(x-\theta)}, & \theta \le x < \infty,\ -\infty < \theta < \infty \\ 0, & \text{otherwise} \end{cases}$$

In each case find the MLE $\hat{\theta}$ of θ.

[2] Suppose $X_1, \ldots, X_n$ are iid with pdf

$$f(x;\theta) = \begin{cases} \frac{2x}{\theta^2}, & 0 < x \le \theta \\ 0, & \text{otherwise} \end{cases}$$

a) Find the MLE $\hat{\theta}$ for θ.
b) Find the constant c so that $E[c\hat{\theta}] = \theta$.
c) The MLE for the median of the distribution.
d) Find the MLE of $P(X \le 2)$.

[3] Let the table

x	0	1	2	3	4	5
Frequency	7	14	12	13	6	3

represent a summary of a random sample of size 55 from a Poisson distribution. Find the maximum likelihood estimate of $P(X = 2)$.

[4] Let X be $N(0,\theta)$, $0 < \theta < \infty$.

a) Find the Fisher information $I(\theta)$.
b) If $X_1, \ldots, X_n$ is a random sample from this distribution, show that the mle of θ is an efficient estimator of θ.

c) What is the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$?

[5] Let $X_1, \ldots, X_n$ be a random sample from a $N(0, \theta)$ distribution. Suppose we want to estimate the standard deviation $\sqrt{\theta}$. Find the constant c so that $Y = c\sum_{i=1}^n |X_i|$ is an unbiased estimator of $\sqrt{\theta}$ and determine its efficiency.

[6] In this problem you are asked to prove Corollary 7 in the lecture notes. Suppose that the assumptions of Theorem 6 hold.

a) Suppose $g(x)$ is a continuous function of x which is differentiable at θ_0 such that $g'(\theta_0) \neq 0$. Then, show that

$$\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta_0)\right) \xrightarrow{d} N\left(0, \frac{[g'(\theta_0)]^2}{I(\theta_0)}\right)$$

b) Show that the following asymptotic representation of $\hat{\theta}$ holds

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{I(\theta_0)}\frac{1}{\sqrt{n}}\sum \frac{\partial \log f(x_i; \theta_0)}{\partial \theta} + R_n$$

where $R_n \xrightarrow{p} 0$. (Hint: This result follows immediately from the equations we used in proving Theorem 6)

c) For a given $\alpha \in (0, 1)$, show that the following interval is an approximate $(1 - \alpha)100\%$ confidence interval for θ

$$\left(\hat{\theta}_n - z_{\alpha/2}\frac{1}{\sqrt{n \times I(\hat{\theta}_n)}}, \quad \hat{\theta}_n + z_{\alpha/2}\frac{1}{\sqrt{n \times I(\hat{\theta}_n)}}\right)$$

Therefore, Theorem 6 is also a practical result for it gives us a way of doing inference.

[7] Consider the following linear model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2)$ is a vector of unknown parameters, and x_i is a one-dimensional observable variable. We have a sample of $i = 1, \ldots, N$ independent observations and assume that the error terms ε_i are iid $N(0, \sigma^2)$, independent of all x_i. The density function of y_i (for a given x_i) is then given by

$$f(y_i|\beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^2}\right\}$$

a) Give an expression for the log-likelihood contribution of observation i, $\log L_i(\boldsymbol{\beta}, \sigma^2)$. Explain why the log-likelihood function of the entire sample is given by

$$\log L(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{N} \log L_i(\boldsymbol{\beta}, \sigma^2)$$

b) Determine expressions for the two elements in $\partial \log L_i(\boldsymbol{\beta}, \sigma^2)/\partial \boldsymbol{\beta}$ and show that both have expectation zero for the true parameter values.

c) Derive an expression for $\partial \log L_i(\boldsymbol{\beta}, \sigma^2)/\partial \sigma^2$ and show that it also has expectation zero for the true parameter values.

Suppose that x_i is a dummy variable equal to 1 for males and 0 for females, such that $x_i = 1$ for $i = 1, \ldots, N_1$ (the first N_1 observations) and $x_i = 0$ for $i = N_1 + 1, \ldots, N$.

(d) Derive the first-order conditions for maximum likelihood. Show that the maximum likelihood estimators for $\boldsymbol{\beta}$ are given by

$$\hat{\beta}_1 = \frac{1}{N - N_1} \sum_{i=N_1+1}^{N} y_i$$

$$\hat{\beta}_2 = \frac{1}{N_1} \sum_{i=1}^{N_1} y_i - \hat{\beta}_1$$

What is the interpretation of these two estimators? What is the interpretation of the true parameter values β_1 and β_2?

(e) Show that

$$\frac{\partial^2 \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \beta \partial \sigma^2} = \frac{\partial^2 \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2 \beta \partial}$$

and show that it has expectation zero. What are the implications of this for the asymptotic covariance matrix of the mle of $(\beta_1, \beta_2, \sigma^2)$?

(f) Compute $I(\beta_1, \beta_2, \sigma^2)$.

[8] In this exercise we are going to derive a likelihood ratio test for the exponential distribution. Suppose $X_1, \ldots, X_n$ are iid with pdf $f(x;\theta) = (1/\theta)e^{-x/\theta}$, for $x, \theta > 0$. Let the hypotheses given by

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0$$

a) Show that the likelihood function simplifies to

$$L(\theta) = \theta^{-n} \exp\{-(n/\theta)\bar{x}\}$$

b) Show that the likelihood ratio test statistic simplifies to

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = e^n \left(\frac{\bar{x}}{\theta_0}\right)^n \exp\{-n\bar{x}/\theta_0\}$$

(Hint: use the mle of exponential distribution)

c) The decision rule is to reject H_0 if $\Lambda \leq c$. Show that

$$\Lambda \leq c \iff \frac{\bar{x}}{\theta_0} \leq c_1 \text{ or } \frac{\bar{x}}{\theta_0} \geq c_2$$

for some c_1 and c_2. (Hint: Define a variable $t = \bar{x}/\theta_0$, substitute in Λ and analyse the resulting function)

d) Show that for a significance level α the decision rule can be expressed as follow

$$\text{Reject } H_0 \text{ if } (2/\theta_0)\sum_{i=1}^{n} X_i \leq \chi^2_{1-\alpha/2}(2n) \text{ or } (2/\theta_0)\sum_{i=1}^{n} X_i \geq \chi^2_{\alpha/2}(2n)$$

where $\chi^2_{1-\alpha/2}(2n)$ is the lower $\alpha/2$ quantile of a χ^2 distribution with $2n$ degrees of freedom and $\chi^2_{\alpha/2}(2n)$ is the upper $\alpha/2$ quantile of a χ^2 distribution with $2n$ degrees of freedom. (Hint: Note that under the null hypothesis, H_0, the statistic $(2/\theta)\sum_{i=1}^{n} X_i$ has a χ^2 distribution with $2n$ degrees of freedom.)

[9] This problem provides another example how we can use Newton's method to find the mle numerically. This problem is **OPTIONAL**: you do not have to return it.

Suppose that $X_1, \ldots, X_n$ are iid random variables following the Cauchy distribution with density function

$$f(x;\theta) = \frac{1}{\pi\left(1 + (x-\theta)^2\right)}$$

The log likelihood function in this case is

$$\ell(\theta) = -\sum_{i=1}^{n} \log\left[1 + (x_i - \theta)^2\right] - n\log\pi$$

This is differentiable, and so if $\hat{\theta}$ is a maximum of $\ell(\theta)$, it must satisfy $\ell'(\hat{\theta}) = 0$, or equivalently,

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

The equation above cannot be explicitly solved to readily give us the form of the maximum likelihood estimator as an explicit function of the data. Therefore, the estimator remains implicitly defined. We will need to solve the equation by some iterative/approximate solution method in order to get the numerical value of the maximum likelihood estimate.

There are several numerical methods that one can employ in order to calculate the value of the maximum likelihood estimator in a specific sample (that is, in order to calculate the estimate). Among these, chief are the Newton-Raphson method, the method of bisection, the method of gradient descent and the EM-algorithm. Which one is most appropriate depends on the specific example. What is common to all of them is that they are iterative: they start at a given input value and iterate some operation until a convergence criterion is attained. Since the function ℓ' might not be monotone (and so may have multiple roots) it is important that the starting input value $\hat{\theta}^{(0)}$ be within a reasonable distance of the true maximum; otherwise, the algorithm may converge to a root that does not correspond to the maximum.

How can we find a reasonable starting value $\hat{\theta}^{(0)}$? In some cases, reasonable starting values may be found by direct inspection. Notice that the density $f(x;\theta)$ is symmetric about θ thus a potential starting value for θ is the median of the sample.

Implement Newton's method in `R` for this problem. You basically need to adapt the code we saw in the lecture to this problem.