STUDENT SOLUTION MANUAL

2013 THEORY SECTION, PART I

BYRON[1]

July 9, 2015

NOTATION

Symbol	**Meaning**
$\mathcal{V}(Y)$	*Variance of* Y
$\mathcal{V}(\mathbf{Y})$	*Covariance Matrix of* $\mathbf{Y}$
$\mathcal{V}(X, Y)$	*Covariance of* X *and* Y
$I\{X\}$	*Identity function for event* X
$\partial_x(y)$	$\partial y / \partial x$
$\partial_{x,z}(y)$	$\partial^2 y / \partial x \partial z$
$\partial_{x^2,z}(y)$	$\partial^3 y / \partial x \partial x \partial z$
$\mathcal{L}(\theta)$	*Likelihood function of*θ
$\mathfrak{S}(\theta)$	*Score Function of* θ
$\mathfrak{H}(\theta)$	*Hessian Matrix of* θ
$\mathfrak{I}(\theta)$	*Fisher Information Matrix of* θ

[1]If you have comments or suggestions, you can reach me at *byron.jaeger@gmail.com*

1 PROBLEM # 1

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ *be an i.i.d. sample of n pairs of random variables, each pair having joint density*

$$f(x, y; \alpha) = \alpha(\alpha+1)(1+x+y)^{-(\alpha+2)} \tag{1}$$

for parameter $0 < \alpha < \infty$ *and* $x, y > 0$.

(A) *Show that the maximum likelihood estimator for* α, $\widehat{\alpha}_n$, *has the following properties*

1. $\widehat{\alpha}_n$ *exists, and is unique and has the form* $g^{-1}(\widehat{\mu}_n)$, *where* $\widehat{\mu}_n = n^{-1}\sum_{i=1}^{n} \log(1+X_i+Y_i)$ *and* g^{-1} *is the inverse of some function* g. *Give the form of* g *and show* g^{-1} *exists.*
2. $\widehat{\alpha}_n \to \alpha_0$, *where* α_0 *is the true value of* α.
3. $\sqrt{n}(\widehat{\alpha}_n - \alpha_0)$ *is asymptotically normal with mean zero and variance*

$$\sigma_1^2 = \frac{\alpha_0^2(\alpha_0+1)^2}{\alpha_0^2 + (\alpha_0+1)^2}$$

(B) *Suppose now that* $(X_1, \ldots, X_n)$ *are fixed and known, i.e.* $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$, *and we observe the sample of independent observations* $Y_1, \ldots, Y_n$, *where, for* $i = 1, \ldots, n$, Y_i *is drawn from the conditional distribution of* $Y_i \mid X_i = x_i$, *and where the unconditional joint density of* (Y_i, X_i) *is given above. Show that for* $i = 1, \ldots, n$ *the density of* $Y_i \mid X_i = x_i$ *is given by*

$$\widetilde{f}_i(y_i; \alpha) = (\alpha+1)(1+x_i)^{-1}\left(1 + \frac{y_i}{1+x_i}\right)^{-(\alpha+2)} \tag{2}$$

(C) *In the setting of (b), verify that the maximum likelihood estimator,* $\widetilde{\alpha}_n$, *has the following properties:*

1. $\widetilde{\alpha}_n$ *exists, is unique, and can be expressed in explicit closed form.*
2. $\widetilde{\alpha}_n \xrightarrow{a.s.} \alpha_0$
3. $\sqrt{n}(\widetilde{\alpha}_n - \alpha_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_2^2 = h(\alpha_0)\right)$, *please give the form of* h

(D) *What is the asmyptotic relative efficiency of* $\widetilde{\alpha}_n$ *to* $\widehat{\alpha}_n$*?*

Solution:

(A) The joint log likelihood of the sample is

$$\ell_n(\alpha) = \sum_{i=1}^{n} \ell(x_i, y_i; \alpha) = \sum_{i=1}^{n} \log(\alpha) + \log(\alpha+1) - (\alpha+2)\cdot\log(1+x_i+y_i) \tag{3}$$

So the score function is

$$\mathcal{S}(\alpha) = \frac{n}{\alpha} + \frac{n}{\alpha+1} - \sum_{i=1}^{n} \log(1+X_i+Y_i). \tag{4}$$

Setting the score equal to zero gives us

$$g(\widehat{\alpha}_n) = \frac{1}{\widehat{\alpha}_n} + \frac{1}{\widehat{\alpha}_n + 1} = \frac{1}{n}\sum_{i=1}^{n} \log(1 + X_i + Y_i) = \widehat{\mu}_n \tag{5}$$

In other words, $\widehat{\alpha}_n = g^{-1}(\widehat{\mu}_n)$. Now taking note of the fact that $g(x) = x^{-1} + (x+1)^{-1}$ is monotone decreasing and $g(x) \in \mathbb{R}$, it follows that g is one-to-one and onto, which in turn tells us that g^{-1} exists. Now by the strong law of large numbers,

$$g(\widehat{\alpha}_n) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + X_i + Y_i) \xrightarrow{\text{a.s.}} \mathbb{E}\left[\log(1 + X_i + Y_i)\right]$$

Now define $Z_i = \log(1 + X_i + Y_i)$. and $V_i = Y_i$. Therefore, $X_i = e^{Z_i} - V_i - 1$ and $Y_i = V_i$. The Jacobian matrix for this transformation is

$$\frac{\partial(x_i, y_i)}{\partial(z_i, v_i)} = \begin{vmatrix} \frac{\partial x_i}{\partial z_i} & \frac{\partial x_i}{\partial v_i} \\ \frac{\partial y_i}{\partial z_i} & \frac{\partial y_i}{\partial v_i} \end{vmatrix} = e^{Z_i}$$

and the joint distribution is

$$\begin{aligned} f(z_i, v_i) &= \exp\{Z_i\} \cdot (\alpha)(1+\alpha)(1 + \exp\{Z_i\} - V_i - 1 + V_i)^{-(\alpha+2)} \\ &= \exp\{-Z_i(\alpha+1)\} \cdot \alpha(\alpha+1) \end{aligned}$$

and the marginal distribution of Z_i is

$$\begin{aligned} f(z_i) &= \int_0^{e^{z_i}-1} \exp\{-z_i(\alpha+1)\} \cdot \alpha(\alpha+1) dy \\ &= (\exp\{z_i\} - 1) \cdot \exp\{-z_i(\alpha+1)\} \cdot \alpha(\alpha+1) \\ &= (\exp\{-z_i\alpha\} - \exp\{-z_i\alpha - z_i\}) \cdot \alpha(\alpha+1) \\ &= \exp\{-z_i\alpha\}(1 - \exp\{-z_i\}) \cdot \alpha(\alpha+1) \end{aligned}$$

Now we can find the value that $g(\widehat{\alpha}_n)$ converges to,

$$\begin{aligned} \mathbb{E}[Z_i] &= \alpha(\alpha+1) \cdot \int_0^{\infty} z_i(\exp\{-z_i\alpha\} - \exp\{-z_i(\alpha+1)\}) \\ &= \frac{\alpha(\alpha+1)}{\alpha^2}\left[\int_0^{\infty} \frac{\alpha^2}{\Gamma(2)} z_i^{2-1} \exp\{-z_i\alpha\}\, dz_i\right] \\ &\quad - \frac{\alpha(\alpha+1)}{(\alpha+1)^2}\left[\int_0^{\infty} \frac{(\alpha+1)^2}{\Gamma(2)} z_i^{2-1} \exp\{-z_i(\alpha+1)\}\, dz_i\right] \end{aligned}$$

$$= \frac{\alpha+1}{\alpha} - \frac{\alpha}{\alpha+1}$$

$$= \frac{2}{\alpha+1} + \frac{1}{\alpha} - \frac{1}{\alpha+1}$$

$$= \frac{1}{\alpha+1} + \frac{1}{\alpha} = g(\alpha).$$

So $g^{-1}(\mathbb{E}[Z_i]) = \alpha$, and 2. follows. Now we use properties associated with maximum likelihood estimators to satisfy 3., namely,

$$\sqrt{n}(\widehat{\alpha}_n - \alpha_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{V}(\widehat{\alpha}_n))$$

where $\mathcal{V}(\widehat{\alpha}_n) = \lim_{n\to\infty} \frac{1}{n} \mathcal{I}^{-1}(\alpha)$ and $\mathfrak{I}(\widehat{\alpha}_n) = -\mathbb{E}\left[\frac{\partial^2 \ell_n(\alpha)}{\partial\alpha\partial\alpha}\right] = n \cdot \frac{\alpha_0^2 + (1+\alpha_0)^2}{\alpha_0^2(1+\alpha_0)^2}$

(B) This follows from the standard definition of conditional distributions:

$$\widetilde{f}_i(y;\alpha) = \frac{f(Y_i, X_i)}{f(X_i)}$$

$$= \frac{(1+x_i+y_i)^{-(\alpha+2)}}{\int_0^\infty (1+x_i+y_i)^{-(\alpha+2)} dy}$$

$$= \frac{(1+x_i+y_i)^{-(\alpha+2)}}{\frac{(1+x_i)^{-(\alpha+1)}}{\alpha+1}}$$

$$= (\alpha+1)(1+x_i)^{\alpha+1} \cdot (1+x_i)^{-(\alpha+2)} \cdot \left(1 + \frac{y_i}{1+x_i}\right)^{-(\alpha+2)}$$

$$= (\alpha+1)(1+x_i)^{-1} \cdot \left(1 + \frac{y_i}{1+x_i}\right)^{-(\alpha+2)}$$

(C) 1. The joint log likelihood is

$$\ell_n(\alpha) = n \cdot \log(\alpha+1) - \sum_{i=1}^n \log(1+x_i) + (\alpha+2)\log\left(1 + \frac{y_i}{1+x_i}\right)$$

so the score function is

$$\mathcal{S}(\alpha) = \frac{n}{\alpha+1} - \sum_{i=1}^n \log\left(1 + \frac{y_i}{1+x_i}\right)$$

Which leads to the maximum likelihood estimate

$$\widetilde{\alpha}_n = \frac{n - \sum_{i=1}^n \log\left(1 + \frac{y_i}{1+x_i}\right)}{\sum_{i=1}^n \log\left(1 + \frac{y_i}{1+x_i}\right)}$$

Since $\ell_n(\alpha)$ is identifiable, it follows that $\widetilde{\alpha}_n$ exists and is unique.

2. Let $U_i = \log\left(1 + \frac{Y_i}{1+x_i}\right)$ so that

$$Y_i = (\exp\{U_i\} - 1) \cdot (1 + x_i)$$

and $\partial_{U_i}(Y_i) = \exp\{U_i\} \cdot (1+x_i)$. The density of U_i is

$$f(u_i) = (\alpha+1) \cdot \exp\{-U_i(\alpha+1)\}$$

Now by the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} U_i \xrightarrow{\text{a.s.}} \mathbb{E}[U_i]$$

where

$$\begin{aligned}
\mathbb{E}[U_i] &= \int_0^\infty u_i \cdot (\alpha+1)\exp\{-u_i(\alpha+1)\} \\
&= \frac{1}{\alpha+1}\int_0^\infty \frac{(\alpha+1)^2}{\Gamma(2)} u_i^{2-1} \exp\{-u_i(\alpha+1)\} \\
&= \frac{1}{\alpha+1}
\end{aligned}$$

Hence,

$$\widetilde{\alpha}_n \xrightarrow{\text{a.s.}} \frac{1 - \frac{1}{\alpha+1}}{\frac{1}{\alpha+1}} = \alpha$$

3. Using the same approach as we did in part a., $\sqrt{n}(\widetilde{\alpha}_n - \alpha_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{V}(\widetilde{\alpha}))$, where

$$\mathcal{V}(\widetilde{\alpha})^{-1} = \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}\left[\frac{\partial^2 \ell_n(\alpha)}{\partial\alpha\partial\alpha}\right] = \frac{1}{(\alpha+1)^2}$$

Thus, $\sigma_2^2 = h(\alpha_0) = (\alpha_0+1)^2$

(D) To find the asymptotic relative efficiency, we divide the variances of the estimators,

$$\frac{\mathcal{V}(\widetilde{\alpha})}{\mathcal{V}(\widehat{\alpha})} = \frac{\alpha_0^2}{(\alpha_0+1)^2 + \alpha_0^2} < 1$$

In other words, conditioning on $X_1, \ldots, X_n$ reduces the variance in our estimation of α. This result is intuitive since fixing X removes the variability corresponding to the random variable.

2 PROBLEM # 3

Consider an r-sided coin and suppose that on each flip of the coin exactly one of the sides appears: side i with probability P_i, $\sum_{i=1}^{r} P_i = 1$. *For given positive integers* $n_1, \ldots, n_r$, *let* N_i *denote the number of flips required until side* i *has appeared for the* n_i *time,* $i = 1, \ldots, r$ *and let* $N = \min_{i=1,\ldots,r} N_i$. *Thus, N is the number of flips required until some side* i *has appeared* n_i *times, for* $i = 1, \ldots, r$.

(A) *Derive the marginal distribution of* N_i, *for* $i = 1, \ldots, r$

(B) *Prove whether or not* $N_i, i = 1, \ldots, r$ *are independent random variables*

Now, suppose that the flips are performed at random times generated by a Poisson process with rate $\lambda = 1$. *Let* T_i *denote the time until side* i *has appeared for the* n_i *time,* $i = 1, \ldots, r$, *and let* $T = \min_{i=1,\ldots,r} T_i$.

(C) *Derive the marginal distribution of* T_i, *for* $i = 1, \ldots, r$.

(D) *Prove whether or not* $T_i, i = 1, \ldots, r$ *are independent random variables.*

(E) *Derive the density of* T.

(F) *Obtain an expression for* $\mathbb{E}[N]$ *as a function of* $\mathbb{E}[T]$.

Solution:

(A) Denote $N_i = \sum_{j=1}^{n_i} N_{ij}$, where $N_{i1}, \ldots, N_{in_i}$ are identically distributed random variables denoting the number of times until side i appears 1 time. Note also that $N_{ij} \perp\!\!\!\perp N_{ik}$ for $j \neq k$ if we assume flipping the coin until side i appears 1 time does not impact the coin's chances of landing on side i. Consider

$$\mathbb{P}(N_{i1} = n) = p_i(1-p_i)^{n-1} \sim \text{Geom}(p_i)$$

Therefore, $N_i \sim \text{NegBin}(p_i, n_i)$ since N_i is the sum of n_i i.i.d. geometric random variables with probability parameter p_i.

(B) $N_i, i = 1, \ldots, r$ are not independent random variables. Consider the following counterexample: for sides 1 and 2 of the coin, $\mathbb{P}(N_1 = 1) > 0$ and $\mathbb{P}(N_2 = 1) > 0$, however,

$$\mathbb{P}(N_1 = 1 \mid N_2 = 1) = 0 \neq \mathbb{P}(N_1 = 1).$$

Thus, the random variables are not independent.

(C) Let X_n denote the time passing between the $(n-1)$st and n^{th} coin flip so that $T_i = \sum_{j=1}^{N_i} X_j$. Figure 1 illustrates the problem set up. Note that $X_1, \ldots, X_{N_i}$ are i.i.d. and

$$\mathbb{P}(X_1 > t) = \mathbb{P}(\text{ No coin flips in } (0, t)).$$

Then $\mathbb{P}(X_1 > t) = \exp\{-1 \cdot t\} \frac{(1 \cdot t)^0}{0!} = \exp\{-t\} \Rightarrow X_1 \sim \exp(1)$. Now define $S_n = \sum_{i=1}^{n} X_i$. By definition, $S_n \sim \text{Gamma}(n, 1)$,

$$\begin{aligned}
\mathbb{P}(T_i < t) &= \mathbb{P}(T_i < t \mid N_i = n) \cdot \mathbb{P}(N_i = n) \\
&= \mathbb{P}(S_n < t) \cdot \mathbb{P}(N_i = n) \\
&= \int_0^t \frac{1^n}{\Gamma(n)} s^{n-1} \exp\{-s\} \cdot \binom{n-1}{n_i} p_i^{n_1} (1-p_i)^{n-n_i}
\end{aligned}$$

where $s < t$.

Figure 1: Illustrative example of notation