Statistics 583, Problem Set 4 Solutions

Wellner; 4/28/2016

1. Suppose that $Z \sim N(0,1)$ and, for $\mu \in R$ and $\sigma > 0$, that $X = \mu + \sigma Z \sim P_{\mu,\sigma} = N(\mu, \sigma^2)$.
(a) Compute the likelihood ratio

$$\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(x) = \frac{\sigma^{-1}\phi((x-\mu)/\sigma)}{\sigma^{-1}\phi(x/\sigma)} \qquad \text{and} \qquad Y \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X)\,.$$

What is the distribution of Y under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$?
(b) Plot the function

$$l(\mu,\sigma;X) \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X)$$

as a function of μ.
(c) Find the maximum value of the function $l(\mu;X)$ in B (as a function of μ) and the value of $\mu \equiv \hat{\mu}$ which achieves the maximum.
(d) What is the distribution of $\hat{\mu}$ under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$? What is the distribution of $l(\hat{\mu};X)$ under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$?

Solution: (a) The likelihood ratio

$$\begin{aligned}
\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(x) &= \frac{\sigma^{-1}\phi((x-\mu)/\sigma)}{\sigma^{-1}\phi(x/\sigma)} = \frac{\exp(-(x-\mu)^2/(2\sigma^2))}{\exp(-x^2/(2\sigma^2))} \\
&= \exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2}\frac{\mu^2}{\sigma^2}\right)\,.
\end{aligned}$$

Hence

$$Y \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X) = \frac{\mu}{\sigma}\frac{X}{\sigma} - \frac{1}{2}\frac{\mu^2}{\sigma^2}\,.$$

Under $P_{0,\sigma}$ we find that $E(Y) = 0 - \frac{\mu^2}{2\sigma^2}$ and $Var(Y) = \mu^2/\sigma^2 \equiv V^2$ so that

$$Y \sim N(-\frac{1}{2}V^2, V^2) \qquad \text{under} \quad P_{0,\sigma}\,.$$

Under $P_{\mu,\sigma}$ a similar computation gives $E(Y) = \mu^2/\sigma^2 - \mu^2/(2\sigma^2) = V^2/2$ and $Var(Y) = V^2$, so

$$Y \sim N(\frac{1}{2}V^2, V^2) \qquad \text{under} \quad P_{\mu,\sigma}\,.$$

(b) and (c). The function

$$l(\mu,\sigma;X) \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X) = \frac{\mu}{\sigma}\frac{X}{\sigma} - \frac{\mu^2}{2\sigma^2} = \frac{X^2}{2\sigma^2} - \frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}$$

is quadratic in μ with maximum value $X^2/(2\sigma^2)$ which is achieved at $\mu = \hat{\mu} \equiv X$. (d) Under $P_{0,\sigma}$, $\hat{\mu} = X \sim N(0,\sigma^2)$ and $l(\hat{\mu},\sigma;X) = X^2/(2\sigma^2) \sim \chi_1^2/2$. Under $P_{\mu,\sigma}$, $\hat{\mu} = X \sim N(\mu,\sigma^2)$ and $l(\hat{\mu},\sigma;X) = X^2/(2\sigma^2) \sim \chi_1^2(\delta)/2$ with $\delta = \mu^2/\sigma^2$.

2. Suppose that $\{P_\theta : \theta \in \Theta \subset \mathbb{R}^d\}$ is a regular parametric model in the sense of satisfying Cramér's conditions A0 - A4 of the 581 Chapter 4 notes (page 5) at $\theta_0 \in \Theta$. Show that the LAN condition holds: with $\theta_n = \theta_0 + tn^{-1/2}$ for $t \in \mathbb{R}^d$,

$$\log \frac{\prod_{i=1}^n p_{\theta_n}(X_i)}{\prod_{i=1}^n p_{\theta_0}(X_i)} = l_n(\theta_n) - l_n(\theta_0) \to_d t^T Z - (1/2)t^T I(\theta_0)t \sim N_1(-(1/2)\sigma_0^2, \sigma_0^2)$$

where $Z \sim N_d(0, I(\theta_0)$ and $\sigma_0^2 = t^T I(\theta_0)t$.

Solution: In the notation used in Theorem 4.1.2,

$$l_n(\theta) - l_n(\theta_0) = \sum_{i=1}^n \log\left(\frac{p_{\theta_n}}{p_{\theta_0}}(X_i)\right).$$

Much as in the proof of part (ii) of Theorem 4.1.2, we can expand $l_n(\theta)$ about θ_0 as follows:

$$l_n(\theta) = l_n(\theta_0) + \dot{\mathbf{l}}_n(\theta_0)^T(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)^T \ddot{\mathbf{l}}_n(\theta_n^*)(\theta-\theta_0)$$

where $|\theta_n^* - \theta_0| \le |\theta - \theta_0|$. Thus with $\theta = \theta_0 + tn^{-1/2}$,

$$\begin{aligned} l_n(\theta_n) &= l_n(\theta_0) + \dot{\mathbf{l}}_n(\theta_0)^T(\theta_n-\theta_0) + \frac{1}{2}(\theta_n-\theta_0)^T \ddot{\mathbf{l}}_n(\theta_n^*)(\theta_n-\theta_0) \\ &= l_n(\theta_0) + t^T\left(\frac{1}{\sqrt{n}}\dot{\mathbf{l}}_n(\theta_0)\right) + \frac{1}{2}t^T\left(\frac{1}{n}\ddot{\mathbf{l}}_n(\theta_n^*)\right)t, \end{aligned}$$

where $|\theta_n^* - \theta_0| \le |\theta_n - \theta_0| = n^{-1/2}|t| \to 0$. Thus it follows that

$$\begin{aligned} l_n(\theta_n) - l_n(\theta_0) &= t^T\left(\frac{1}{\sqrt{n}}\dot{\mathbf{l}}_n(\theta_0)\right) - \frac{1}{2}t^T\left(-\frac{1}{n}\ddot{\mathbf{l}}_n(\theta_n^*)\right)t \\ &= t^T Z_n - -\frac{1}{2}t^T I(\theta_0)t + o_p(1) \end{aligned}$$

since

$$\left(-\frac{1}{n}\ddot{\mathbf{l}}_n(\theta_n^*)\right) \to_p I(\theta_0)$$

in the same way as in the proof of Theorem 4.1.2 part (ii).

3. Suppose that we want to model the survival of twins with a common genetic defect, but with one of the two twins receiving some treatment. Let X represent the survival time of the untreated twin and let Y represent the survival time of the treated twin. One (overly simple) preliminary model might be to assume that X and Y are independent with Exponential(η) and Exponential($\theta\eta$) distributions, respectively:

$$f_{\nu,\eta}(x,y) = \eta e^{-\eta x} \eta \nu e^{-\eta \nu y} 1_{(0,\infty)}(x) 1_{(0,\infty)}(y)$$

Suppose that we observe i.i.d. pairs (X_i, Y_i) with density given by f_{ν_0,η_0}.

(a) One crude approach to estimation in this problem is to reduce the data to $W = X/Y$, the maximal invariant for the group of scale changes $g(x,y) = (cx, cy)$ with $c > 0$. Find the distribution of W, and compute the Cramér-Rao lower bound for unbiased estimates of ν based on W.
(b) Find the information bound for estimation of ν based on observation of (X,Y) pairs when η is known and unknown.
(c) Compare the bounds you computed in (a) and (b) and discuss the pros and cons of reducing to estimation based on the W.

Solution: (a) We compute

$$\begin{aligned} P(W \le w) &= P(X/Y \le w) = EE\left(1\{Y \ge X/w\}|X\right) \\ &= E\{\exp(-\nu\eta X/w)\} = \int_0^\infty \exp(-\nu\eta x/w)\eta\exp(-\eta x)dx \\ &= \eta \int_0^\infty \exp(-(1+\nu/w)\eta x)dx \\ &= \frac{\eta}{\eta(1+\nu/w)} = \frac{w}{w+\nu} = 1 - \frac{\nu}{w+\nu}. \end{aligned}$$

Thus W has density $f_W(w,\nu) = \nu/(\nu+w)^2$ on $w \ge 0$, with $\log f_W(w,\nu) = \log\nu - 2\log(\nu+w)$. It follows that the score for ν based on observation of W is $\dot{l}_\nu(w) = \nu^{-1} - 2/(\nu+w)$. Thus $\ddot{l}_{\nu\nu}(w) = -\nu^{-2} + 2/(\nu+w)^2$ and the information for ν based on W is

$$I_W(\nu) = -E_\nu \ddot{l}_{\nu,\nu}(W) = \nu^{-2} - 2\int_0^\infty \frac{\nu}{(\nu+w)^4}dw = \nu^{-2}(1-2/3) = (1/3)\nu^{-2}.$$

(b) Based on observation of the (X,Y) pairs we have $\log f_{\nu,\eta}(x,y) = 2\log\eta + \log\nu - \eta(x+\nu y)$, and hence the scores for ν and η are given by

$$\begin{aligned} \dot{l}_\nu(x,y) &= \frac{1}{\nu} - \eta y = \frac{1}{\nu}(1-\nu\eta y), \\ \dot{l}_\eta(x,y) &= \frac{2}{\eta} - (x+\nu y) = \frac{1}{\eta}(2 - (\eta x + \nu\eta y)). \end{aligned}$$

This yields

$$\ddot{l}_{\nu\nu}(x,y) = -\nu^{-2}, \quad \ddot{l}_{\nu,\eta}(x,y) = -y, \quad \ddot{l}_{\eta\eta}(x,y) = -2\eta^{-2}.$$

and hence the information matrix for (ν,η) is given by $I(\nu,\eta) = E_{\nu,\eta}\ddot{\underline{l}}(X,Y)$

$$I(\nu,\eta) = E_{\nu,\eta}\ddot{\underline{l}}(X,Y) = \begin{pmatrix} \nu^{-2} & -(\nu\eta)^{-1} \\ -(\nu\eta)^{-1} & 2\eta^{-2} \end{pmatrix}.$$

Thus the information for ν when η is known is ν^{-2}, and the information for estimation of ν when η is unknown is $I_{11\cdot 2} = \nu^{-2} - (\nu\eta)^{-2}(\eta^2/2) = 1/(2\nu^2)$. Thus the information bound for ν when η is known is $I_{11}^{-1} = \nu^2$ (or ν^2/n), while the information bound for ν when η is unknown is $I_{11\cdot 2}^{-1} = 2\nu^2$ (or $2\nu^2/n$).
(c) From (b) we see that not knowing η increases the information bound by a factor of 2. Reduction to the $W_i = X_i/Y_i$ results in a model with no nuisance parameter η, but with an information bound greater than the information bound when η is known by a factor of 3 and greater than the information bound when η is unknown by a factor of 3/2.

4. This is a continuation of the preceding problem. A more realistic model involves assuming that the common parameter η for the two twins varies across sets of twins. There are several different ways of modeling this: one approach involves supposing that each pair of twins observed (X_i, Y_i) has its own fixed parameter η_i, $i = 1, \ldots, n$. In this model we observe (X_i, Y_i) with density f_{ν,η_i} for $i = 1, \ldots, n$; i.e.

$$f_{\nu,\eta_i}(x_i, y_i) = \eta_i e^{-\eta_i x_i} \eta_i \nu e^{-\eta_i \nu y_i} 1_{(0,\infty)}(x_i) 1_{(0,\infty)}(y_i)\,. \tag{1}$$

This is sometimes called a "functional model" (or model with incidental nuisance parameters).

Another approach is to assume that $\eta \equiv Z$ has a distribution, and that our observations are from the mixture distribution. Assuming (for simplicity) that $Z = \eta \sim \text{Gamma}(a,b)$ with density $g_{a,b}(\eta)$, it follows that the (marginal) distribution of (X,Y) is

$$\begin{aligned} p_{\nu,a,b}(x,y) &= \int_0^\infty f_{\nu,z}(x,y) g_{a,b}(z)dz \\ &= \frac{\nu}{b^2}\left(\frac{b}{b+x+\nu y}\right)^{a+2} \frac{\Gamma(a+2)}{\Gamma(a)}\,. \end{aligned} \tag{2}$$

This is sometimes called a "structural model" (or mixture model).
(a) Find the information for ν in the functional model.
(b) Find the information for ν in the structural model.
(c) Compare the information bounds you computed in (a) and (b). When is

the information for ν in the functional model larger than the information for ν in the structural model?
(d) Find the MLEs of ν in the functional model (call it $\hat{\nu}_n^f$) and in the structural model (call it $\hat{\nu}_n^s$). Are they both consistent estimators of ν?

Solution: (a) Here the parameter is $\gamma = (\nu, \eta_1, \ldots, \eta_n) \in \nu \subset R^{n+1}$. For (X_i, Y_i) the log of the density is

$$\log p_{\nu,\underline{\eta}}(x_i, y_i) = 2\log\eta_i + \log\nu - \eta_i(x_i + \nu y_i)\,,$$

so the scores are

$$\dot{l}_\nu(x_i, y_i) = \frac{1}{\nu} - \eta_i y_i \qquad \dot{l}_{\eta_i}(x_i, y_i) = \frac{2}{\eta_i} - (x_i + \nu y_i), \qquad \dot{l}_{\eta_j}(x_i, y_i) = 0, \quad j \neq i\,.$$

Thus we find that the information matrix for γ based on (X_i, Y_i) is $I_i = (I_{i,j,k})$ with $I_{i,1,1} = 1/\nu^2$, $I_{i,i,i} = 2/\eta_i^2$, $I_{i,1,i} = I_{i,i,1} = 1/(\nu\eta_i)$, and $I_{i,j,k} = 0$ otherwise (i.e. $j \neq i$ or $k \neq i$). By independence of (X_i, Y_i) for $i = 1, \ldots, n$, the information matrix for γ based on observation of all the data is

$$I_n(\gamma) \;=\; \sum_{i=1}^n I_i(\gamma) = \begin{pmatrix} n/\nu^2 & 1/(\nu\eta_1) & \cdots & \cdots & 1/(\nu\eta_n) \\ 1/(\nu\eta_1) & 2/\eta_1^2 & 0 & \cdots & 0 \\ . & 0 & . & . & 0 \\ . & 0 & . & . & 0 \\ 1/(\nu\eta_n) & 0 & \cdots & \cdots & 2/\eta_n^2 \end{pmatrix}.$$

Thus

$$\begin{aligned} I_{11\cdot 2}(\gamma) &= I_{11} - I_{12}I_{22}^{-1}I_{21} \\ &= \frac{n}{\nu^2} - \frac{1}{\nu^2}\left(\frac{1}{\eta_1}, \ldots, \frac{1}{\eta_n}\right) \begin{pmatrix} \eta_1^2/2 & 0 & . & . \\ 0 & \eta_2^2/2 & 0 & . \\ . & . & . & . \\ 0 & . & . & \eta_n^2/2 \end{pmatrix} \begin{pmatrix} \frac{1}{\eta_1} \\ . \\ . \\ \frac{1}{\eta_n} \end{pmatrix} \\ &= \frac{n}{\nu^2} - \frac{n}{2\nu^2} = \frac{n}{2\nu^2}\,. \end{aligned}$$

(b) For the structural model, first note that $\Gamma(a+2)/\Gamma(a) = a(a+1)$. Then we compute the scores:

$$\begin{aligned} \dot{l}_\nu(x, y) &= \frac{1}{\nu} - \frac{(a+2)y}{b + x + \nu y}\,, \\ \dot{l}_a(x, y) &= \frac{\Gamma'}{\Gamma}(a+2) - \frac{\Gamma'}{\Gamma}(a) + \log\left(\frac{b}{b + x + \nu y}\right) \\ &= \frac{1}{a} + \frac{1}{a+1} + \log\left(\frac{b}{b + x + \nu y}\right), \\ \dot{l}_b(x, y) &= \frac{a}{b} - \frac{a+2}{b + x + \nu y}\,. \end{aligned}$$

Furthermore, the second derivatives of the scores are:

$$\begin{aligned}
\ddot{l}_{\nu\nu}(x,y) &= -\frac{1}{\nu^2} + \frac{(a+2)y^2}{(b+x+\nu y)^2}\,, \\
\ddot{l}_{aa}(x,y) &= \psi'(a+2) - \psi'(a), \qquad \text{where} \ \ \psi(x) = \frac{\Gamma'}{\Gamma}(x) \\
&= -\frac{1}{a^2} - \frac{1}{(a+1)^2}\,, \\
\ddot{l}_{bb}(x,y) &= -\frac{a}{b^2} + \frac{a+2}{(b+x+\nu y)^2}\,, \\
\ddot{l}_{\nu a}(x,y) &= -\frac{y}{b+x+\nu y}\,, \\
\ddot{l}_{\nu b}(x,y) &= \frac{(a+2)y}{(b+x+\nu y)^2}\,, \\
\ddot{l}_{ba}(x,y) &= \frac{1}{b} - \frac{1}{b+x+\nu y}\,.
\end{aligned}$$

It follows (after some computation; I used Mathematica), that the information matrix for (ν, a, b) is:

$$I(\nu,a,b) = \begin{pmatrix} \frac{a+1}{a+3}\frac{1}{\nu^2} & \frac{1}{(a+2)\nu} & \frac{-a}{(a+3)b\nu} \\ \frac{1}{(a+2)\nu} & a^{-2}+(a+1)^{-2} & \frac{-2}{(a+2)b} \\ \frac{-a}{(a+3)b\nu} & \frac{-2}{(a+2)b} & \frac{2a}{(a+3)b^2} \end{pmatrix}. \tag{3}$$

Hence the information for ν in the structural model is, with

$$D \equiv b^2 \det(I_{22}) = \left(\frac{2a}{a+3}(a^{-2}+(a+1)^{-2}) - \frac{4}{(a+2)^2} \right),$$

$$\begin{aligned}
&I_{11\cdot 2}(\nu,a,b) = I_{11} - I_{12}I_{22}^{-1}I_{21} \\
&= \frac{a+1}{(a+3)\nu^2} \\
&\quad - \begin{pmatrix} \frac{1}{(a+2)\nu} & \frac{-a}{(a+3)b\nu} \end{pmatrix} \frac{1}{D} \begin{pmatrix} \frac{2a}{(a+3)b^2} & \frac{2}{(a+2)b} \\ \frac{2}{(a+2)b} & a^{-2}+(a+1)^{-2} \end{pmatrix} \begin{pmatrix} \frac{1}{(a+2)\nu} \\ \frac{-a}{(a+3)b\nu} \end{pmatrix} \\
&= \frac{a+1}{(a+3)\nu^2} \\
&\quad - \begin{pmatrix} \frac{1}{(a+2)\nu} & \frac{-a}{(a+3)b\nu} \end{pmatrix} \frac{1}{D} \begin{pmatrix} \frac{2a}{(a+3)b^2} & \frac{2}{(a+2)b} \\ \frac{2}{(a+2)b} & \frac{a+2}{a^2(a+1)^2} \end{pmatrix} \begin{pmatrix} \frac{1}{(a+2)\nu} \\ \frac{-a}{(a+3)b\nu} \end{pmatrix} \\
&= \frac{1}{\nu^2}\left\{ \frac{a+1}{a+3} - \frac{2a}{(a+3)(a+2)^2}\left(\frac{(a+2)}{2a^2(a+1)^2}\frac{a}{a+3}(a+2)^2 - 1 \right)\frac{1}{D} \right\} \\
&= \frac{1}{2\nu^2}\frac{2+a}{3+a}
\end{aligned}$$

after a bit of algebra (I used Mathematica again). Note that this equals $1/(3\nu^2)$ when $a = 0$, and it increases to $1/(2\nu^2)$ as $a \to \infty$.

For the semiparametric generalization of the mixture (structural) model given by (2), we have

$$p_{\nu,a,b}(x,y) = \int_0^\infty f_{\nu,z}(x,y)dG(z)$$

where G is an *arbitrary* (mixing) distribution on $(0,\infty)$. In fact, the information for ν in this larger model has the same qualitative feature as in the Gamma-mixture submodel:

$$I_{11\cdot 2}(\nu) = \frac{1}{3\nu^2} + \frac{1}{12\nu^2} I_{p_T}(scale)$$

where $I_{p_T}(scale)$ is the information for scale in for the density

$$p_T(t) \equiv t \int_0^\infty z^2 \exp(-tz)dG(z)\,.$$

It is easily seen that this information is always greater than $1/(3\nu^2)$ and always less than or equal to $1/(2\nu^2)$. See Bickel, Klaassen, Ritov, and Wellner (1993), pages 134 - 135 for this calculation. Section 4.5 of BKRW has much more on information calculations for semiparametric mixture models.

(d) In the functional model we maximize

$$l_n(\nu,\eta_1,\ldots,\eta_n) \equiv \sum_{i=1}^n \log f_{\nu,\eta_1,\ldots,\eta_n}(X_i,Y_i) = \sum_{i=1}^n \{2\log\eta_i + \log\nu - \eta_i(X_i+\nu Y_i)\}$$

over all $\eta_i > 0$ and $\nu > 0$. Thus we compute

$$\frac{\partial}{\partial\eta_i} l_n(\nu,\eta_1,\ldots,\eta_n) = \frac{2}{\eta_i} - (X_i+\nu Y_i),$$

and find that for each fixed ν and $i \in \{1,\ldots,n\}$, the log-likelihood is maximized with respect to η_i by $\hat\eta_i(\nu) = 2/(X_i+\nu Y_i)$. Thus the profile log-likelihood becomes

$$l_n^{profile}(\nu) \equiv l_n(\nu,\hat\eta_1(\nu),\ldots,\hat\eta_n(\nu)) = 2\sum_{i=1}^n \left\{\log\left(\frac{2}{X_i+\nu Y_i}\right) + \log\nu - 2\right\}.$$

Here we have

$$\frac{\partial}{\partial\nu} l_n^{profile}(\nu) = \dot{l}_{n,\nu}^{profile} = -2\sum_{i=1}^n \frac{Y_i}{X_i+\nu Y_i} + \frac{n}{\nu},$$

and hence the maximizer $\hat{\nu}_n$ of the profile likelihood satisfies

$$\frac{1}{\hat{\nu}_n} = \frac{1}{n}\sum_{i=1}^n \frac{2Y_i}{X_i + \hat{\nu}_n Y_i}.$$

Equivalently

$$\begin{aligned} \frac{1}{2} &= \frac{1}{n}\sum_{i=1}^n \frac{\hat{\nu}_n Y_i}{X_i + \hat{\nu}_n Y_i} = \frac{1}{n}\sum_{i=1}^n \frac{(\hat{\nu}_n/\nu)\nu\eta_i Y_i}{\eta_i + (\hat{\nu}_n/\nu)\nu\eta_i Y_i} \\ &\stackrel{d}{=} \frac{1}{n}\sum_{i=1}^n \frac{(\hat{\nu}_n/\nu)V_i}{U_i + (\hat{\nu}_n/\nu)V_i} \equiv \frac{1}{n}\sum_{i=1}^n \frac{\hat{\gamma}_n V_i}{U_i + \hat{\gamma}_n V_i}. \end{aligned}$$

where $(U_i, V_i) \equiv (\eta_i X_i, \eta_i \nu Y_i)$ are i.i.d. pairs of independent exponential(1) random variables and we have defined $\hat{\gamma}_n \equiv \hat{\nu}_n/\nu$. Thus we define

$$\Psi_n(\gamma) \equiv \mathbb{P}_n \psi_\gamma(U, V), \quad \text{and} \quad \Psi(\gamma) \equiv P_0 \psi_\gamma(U, V)$$

where $\psi_\gamma(u, v) = \gamma v/(u + \gamma v) - 1/2$. Note that $\psi_\gamma(u, v) \in [-1/2, 1/2]$ and that it is monotone non-decreasing and continuous from $-1/2$ to $1/2$ as γ increases from 0 to ∞ for each fixed (u, v), Thus $\Psi(\gamma)$ is monotone nondecreasing and continuous. Furthermore $\Psi(1) = P_0(U/(U + V) - 1/2) = 0$ and $\Psi_n(\widehat{\gamma}_n) = 0$. since $U/(U + V) \sim \text{Uniform}(0, 1)$. Since the class of functions $\mathcal{F} \equiv \{\psi_\gamma :\ \gamma \ge 0\}$ is a class of continuously parametrized functions with parameter set $\Gamma \subset \overline{\Gamma}\}$ where $\overline{\Gamma}$ is compact, $\mathcal{F}$ is a Glivenko-Cantelli class of functions: i.e.

$$\|\Psi_n - \Psi\|_\Gamma = \sup_{\gamma > 0} |\Psi_n(\gamma) - \Psi(\gamma)| = \|\mathbb{P}_n - P_0\|_{\mathcal{F}} \to_{a.s.} 0.$$

where $\Psi(1) = 0$ and $\dot{\Psi}(\gamma) = P_0(UV/(U + \gamma V)^2) > 0$ for all γ. It follows that $\widehat{\gamma}_n = \widehat{\nu}_n/\nu \to_{a.s.} 1$, and hence $\widehat{\nu}_n \to_{a.s.} \nu$. Thus the MLE is apparently consistent in the functional model despite the increasing number of nuisance parameters $\eta_1, \ldots, \eta_n$. Note that the score equation for the profile likelihood corresponds exactly to the score equation based on the maximal invariants $W_i = X_i/Y_i$ in problem 3, and this suggests that $\widehat{\nu}_n$ in the structural model satisfies $\sqrt{n}(\widehat{\nu}_n - \nu) \to_d N(0, 1/I_W(\nu)) = N(0, 3\nu^2)$, which apparently disagrees with our information calculation above in the functional model! (There is another version of this model in which ν enters more symmetrically and in which the MLE $\widehat{\nu}_n$ is *not consistent*, more in agreement with the Neyman - Scott example.)

The structural model in (2) is a regular parametric model; it is a type of bivariate Pareto model: see e.g.

`https://en.wikipedia.org/wiki/Multivariate_Pareto_distribution` .

Hence the MLEs should be consistent and otherwise well-behaved.