Statistics 581, Problem Set 10 Solutions

Wellner; 12/7/2005

1. **Continuation of problem 5, problem set 9:** Consider the Weibull family of example 3.2.5: $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset R^{+2}$ given by the (Lebesgue) densities

$$p_\theta(x) = \frac{\beta}{\alpha}(\frac{x}{\alpha})^{\beta-1}\exp(-(\frac{x}{\alpha})^\beta)1_{[0,\infty)}(x)$$

where $\theta \equiv (\alpha, \beta) \in (0,\infty)\times(0,\infty) \subset R^2$. Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with density function p_θ.
A. Does a maximum likelihood estimate of $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ exist? Is it unique? (See e.g. Lehmann and Casella, example 6.1, page 468.)
B. Compute an approximate (one - step) maximum likelihood estimate $\check{\theta}$ of θ using the method of moment estimators $\overline{\theta}_n$ as the preliminary estimators based on the following data (with $n = 19$):

0.19, 0.78, 0.96, 1.31, 2.78, 3.16, 4.15, 4.67, 4.85, 6.50,
7.35, 8.01, 8.27, 12.06, 31.75, 32.52, 33.91, 36.71, 72.89 .

[These are failure times in minutes for an insulating fluid between two electrodes subject to a voltage of 34 kV. – from Nelson, *Applied Life Data Analysis*, page 105.]
C. Compute the maximum likelihood estimator $\hat{\theta}_n$, and compare it with the one step estimator computed in B.

Solution: A. The maximum likelihood estimator exists and is unique in this model if not all the X_i's are equal (which happens with probability 1 if the model holds). The following solution is from Lehmann, TPE, page 536 (with slightly different notation).
We first reparametrize the Weibull model by writing

$$\begin{aligned} p_\theta(x) &= \frac{\beta}{\alpha}(\frac{x}{\alpha})^{\beta-1}\exp(-(\frac{x}{\alpha})^\beta)1_{(0,\infty)}(x) \\ &= \frac{\beta}{\eta}x^{\beta-1}\exp(-\frac{x^\beta}{\eta}) \\ &\equiv p_\gamma(x) \end{aligned}$$

where $\eta \equiv \alpha^\beta$ and $\gamma \equiv (\beta, \eta)$. Then

$$l(\gamma|\underline{X}) = n\log\beta - n\log\eta + (\beta-1)\sum_{i=1}^n \log X_i - \frac{1}{\eta}\sum_{i=1}^n X_i^\beta.$$

Thus, with $\gamma_1 \equiv \beta$, $\gamma_2 \equiv \eta$, the likelihood equations become

$$\dot{l}_1(\gamma|\underline{X}) = \frac{n}{\beta} + \sum_{i=1}^n \log X_i - \frac{1}{\eta}\sum_{i=1}^n X_i^\beta \log X_i = 0, \qquad (0.1)$$

and

$$\dot{l}_2(\gamma|\underline{X}) = -\frac{n}{\eta} + \frac{1}{\eta^2}\sum_{i=1}^n X_i^\beta = 0, \tag{0.2}$$

or

$$\hat{\eta}_n = \frac{1}{n}\sum_{i=1}^n X_i^{\hat{\beta}} \tag{0.3}$$

from 0.2. Substitution of 0.3 into 0.1 yields the equation

$$\frac{\sum_i X_i^{\hat{\beta}} \log X_i}{\sum_i X_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}} = \frac{1}{n}\sum_{i=1}^n \log X_i\,, \tag{0.4}$$

or

$$h(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^n \log X_i \tag{0.5}$$

where

$$h(\beta) \equiv \frac{\sum_i X_i^\beta \log X_i}{\sum_i X_i^\beta} - \frac{1}{\beta} < \frac{\sum_i X_i^\beta \log X_i}{\sum_i X_i^\beta}$$

since $\beta > 0$. Now

$$\begin{aligned} h'(\beta) &= \frac{\sum_i X_i^\eta (\log X_i)^2}{\sum_i X_i^\beta} - (\frac{\sum_i X_i^\beta \log X_i}{\sum_i X_i^\beta})^2 + \frac{1}{\beta^2} \\ &\equiv I + II \\ &> I, \end{aligned}$$

and furthermore,

$$I = \sum a_i^2 p_i - (\sum a_i p_i)^2 = \mathrm{Var}_p(a)$$

since, with $a_i \equiv \log X_i$, $p_i \equiv X_i^\beta / \sum_j X_j^\beta \ge 0$, $\sum_i p_i = 1$. Thus $I > 0$ and hence $h'(\beta) > 0$ from (0.6) while

$$-\infty = \lim_{\beta\to 0} h(\beta) < \frac{1}{n}\sum_{i=1}^n \log X_i < \log X_{(n)} = \lim_{\beta\to\infty} h(\beta).$$

[Draw the picture!] (To see this last limit, note that with $p_{(i)} \equiv X_{(i)}^\beta / \sum_j X_j^\beta$,

$$\begin{aligned} p_{(i)} &= \frac{1}{(\frac{X_{(1)}}{X_{(i)}})^\beta + \ldots + (\frac{X_{(n)}}{X_{(i)}})^\beta} \\ &\to \left\{ \begin{array}{ll} 0, & i \le n \quad\quad (\text{so } X_{(n)}/X_{(i)} > 1) \\ 1, & i = n \quad (\text{so } X_{(j)}/X_{(n)} < 1,\ j < n) \end{array} \right. \end{aligned}$$

as $\beta \to \infty$.) Thus (0.5) has a unique solution $\hat{\beta}$. By taking this value of $\hat{\beta}$ in (0.3), we see that the MLE $\hat{\gamma}$ of γ exists and is unique. Thus the unique MLE of $\theta = (\alpha, \beta)$ is $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ with $\hat{\alpha} = \hat{\eta}^{1/\hat{\beta}}$.

B. The one step estimator using $\hat{I}(\overline{\theta}_n) = I(\overline{\theta}_n)$ is

$$\check{\theta}_n \equiv \overline{\theta}_n + \hat{I}_n^{-1}(\overline{\theta}_n)(\frac{1}{n}\dot{l}(\overline{\theta}_n)) = (12.27\ldots\,,\ 0.7421\ldots).$$

The one - step estimator using $\hat{I}_n(\overline{\theta}_n) = (-n^{-1}\ddot{l}_n(\overline{\theta}_n)$ is

$$\check{\theta}_n = (11.778, 0.7633).$$

C. The maximum likelihood estimate $\hat{\theta}_n = (12.222..., 0.77082...)$; see the following pages.

Mathematica input:

```
(* Here is the data:  *)
x = {.19, .78, .96, 1.31, 2.78, 3.16, 4.15, 4.67, 4.85,
  6.50, 7.35, 8.01, 8.27, 12.06, 31.75, 32.52, 33.91,
  36.71, 72.89}
(* NSS is the sample size *)
NSS:= Length[x]
(* Some useful functions:  *)
(* f is the Weibull density function:  *)
f[t_,a_,b_] := (b/a)*(t/a)^(b-1) *Exp[-(t/a)^b] ;

(* aa and bb are the constants in the Weibull Informaton:  *)
aa := N[-(1-EulerGamma)];
bb := N[(Pi^2)/6 + aa^2 ]
(* Inf is the information matrix *)
Inf[a_,b_] := { {b^2/a^2 , aa/a}, {aa/a, bb/b^2}} ;
(* L  is the log-likelihood *)
L[a_,b_] := Sum[Log[f[x[[i]], a,b]], {i,1,NSS} ] ;
(* Sc is the vector of Scores *)
Sc[a_,b_] := Sum[ {(b/a)(((x[[i]]/a)^b) -1),
(1/b)(1-Log[(x[[i]]/a)^b]*((x[[i]]/a)^b  -1) ) },
{i,1,NSS}] ;
a[b_] := (Sum[x[[i]]^b, {i,1,NSS}]/NSS )^(1/b)

Plot3D[L[a,b], {a,4,15}, {b,.1,1.5}]
FindMinimum[-L[a,b], {a,10},  {b,2}]
ahat = 12.222
bhat = .770821
Wald[a_,b_] := NSS*({ahat,bhat} -{a,b}).Inf[ ahat,bhat].({ahat,bhat} -{a,b})
Rao[a_,b_] := Sc[a,b].Inverse[Inf[a,b]].Sc[a,b]/NSS
LR[a_,b_] := 2*(L[ahat,bhat] - L[a,b])
Wald[10,1]
Rao[10,1]
LR[10,1]
```

Mathematica output:

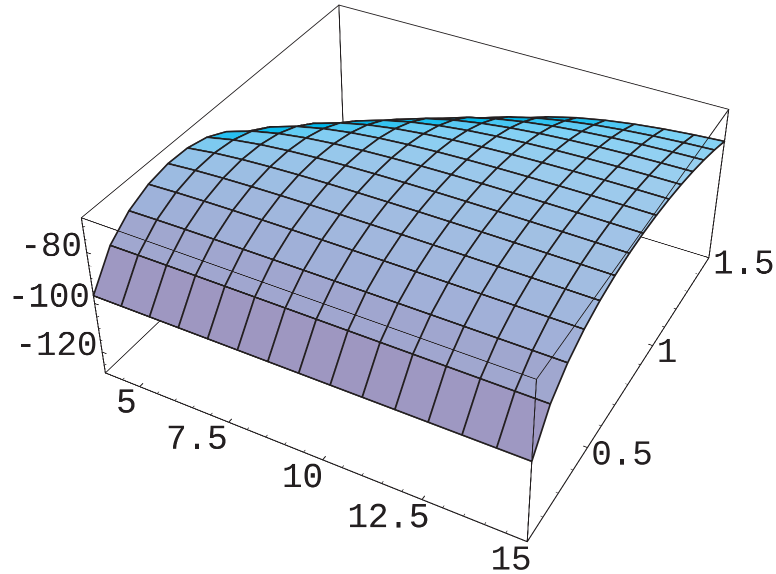

Figure 1: Weibull likelihood.

```
{68.386, {a -> 12.2222, b -> 0.770821}}
12.222
0.770821
4.10551
10.8385
5.29018
```

2. A. Ferguson, ACLST, page 139, problem 3.
   B. What if Ferguson's density $f(x|\theta)$ with $\theta \in (0,1)$ is replaced by

$$f(x|\gamma,\eta) = \{(1-\gamma)e^{-x} + \gamma\eta^2 x \exp(-\eta x)\}1_{[0,\infty)}(x)$$

with $\gamma \in (0,1)$ and $\eta > 0$?

Solution: A. First,

$$E_\theta X = (1-\theta) + \theta \int_0^\infty x^2 e^{-x} dx = (1-\theta) + \theta\Gamma(3) = 1 - \theta + 2\theta = 1 + \theta\,.$$

Thus the method of moments estimator of θ is given by $\overline{X}_n - 1$. Now

$$\begin{aligned} E_\theta(X^2) &= (1-\theta)\int_0^\infty x^2 e^{-x} dx + \theta \int_0^\infty x^3 e^{-x} dx \\ &= (1-\theta)\Gamma(3) + \theta\Gamma(4) \\ &= (1-\theta)2 + \theta 3! = (1-\theta) + 6\theta \\ &= 2 + 4\theta\,. \end{aligned}$$

Thus

$$Var_\theta(X) = 2 + 4\theta - (1+\theta)^2 = 1 + 2\theta - \theta^2\,.$$

Hence it follows by the CLT that

$$\sqrt{n}(\theta_n^* - \theta) = \sqrt{n}(\overline{X}_n - 1 - (E_\theta(X) - 1)) \rightarrow_d N(0, 1 + 2\theta - \theta^2)\,.$$

Now

$$l(\theta|X) = \log f(X|\theta) = \log[(1-\theta)e^{-x} + \theta x e^{-x}]\,,$$

and hence

$$\dot{l}_\theta(x) = \frac{xe^{-x} - e^{-x}}{(1-\theta)e^{-x} + \theta x e^{-x}} = \frac{x-1}{1+\theta(x-1)}\,.$$

Furthermore

$$\ddot{l}_{\theta\theta}(x) = -\frac{(x-1)^2}{[1+\theta(x-1)]^2}\,.$$

Hence a one-step Newton approximation to a root of the likelihood equation is given by

$$\overline{\theta}_n = \theta_n^* + \hat{I}_n(\theta_n^*)^{-1}\frac{1}{n}\sum_{i=1}^n \frac{(X_i - 1)}{1+\theta_n^*(X_i - 1)}\,,$$

where

$$\hat{I}_n(\theta_n^*) \equiv \frac{1}{n}\sum_{i=1}^n \frac{(X_i-1)^2}{[1+\theta_n^*(X_i-1)]^2}\,.$$

Note that

$$I(\theta) = -E_\theta \ddot{l}_{\theta\theta}(X) = E_\theta \frac{(X-1)^2}{[1+\theta(X-1)]^2}$$

increases from 1 at $\theta = 0$ to ∞ at $\theta = 1$, so $1/I(\theta)$ decreases from 1 at $\theta = 0$ to 0 at $\theta = 1$, while the variance of the method of moments estimator, $1 + 2\theta - \theta^2$, increases from 1 to 2 as θ increases from 0 to 1. Hence the gain in efficiency by use of the efficient one-step estimator is quite large for θ near 1. See the plot of $1/I(\theta)$ and $1 + 2\theta - \theta^2$ below.

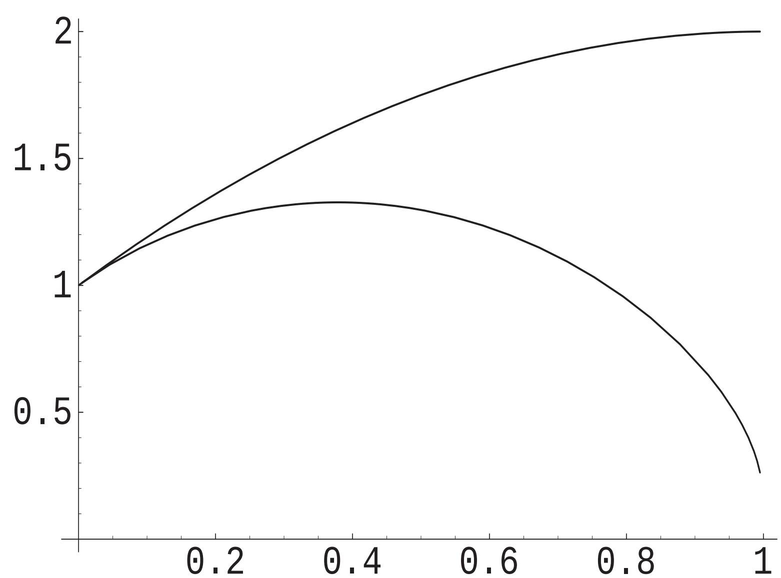

Figure 2: $1/I(\theta)$ and $1 + 2\theta - \theta^2$, $0 < \theta < 1$

B. When Ferguson's density $f(x|\theta)$ with $\theta \in (0,1)$ is replaced by

$$f(x|\gamma,\eta) = \{(1-\gamma)e^{-x} + \gamma\eta^2 x \exp(-\eta x)\} 1_{[0,\infty)}(x)$$

with $\gamma \in (0,1)$ and $\eta > 0$, the parameter to be estimated is $\theta = (\gamma, \eta)$, and we can again implement a one step procedure starting from some $n^{1/4}-$consistent preliminary esimator $\overline{\theta}_n$. One possibility for $\overline{\theta}_n$ is a a method of moments estimator. We calculate

$$\begin{aligned} E(X) &= (1-\gamma) + \gamma\frac{2}{\eta} = 1 + \gamma\left(\frac{2}{\eta} - 1\right) \\ E(X^2) &= (1-\gamma)2 + \gamma\frac{6}{\eta^2} = 2 + \gamma\left(\frac{6}{\eta^2} - 2\right) . \end{aligned}$$

For $\eta \neq 2$ this yields

$$\frac{E(X^2)-2}{E(X)-1} = \frac{6/\eta^2 - 2}{2/\eta - 1} = \frac{6 - 2\eta^2}{2\eta - \eta^2} . \tag{0.6}$$

The difficulty is that solving this for η yields two non-negative solutions in general. I have not yet found a "nice" and "simple" starting point, $\overline{\theta}_n$ for this problem. Figures 1.A and 1.B shows a plot of the right sides of (0.6).

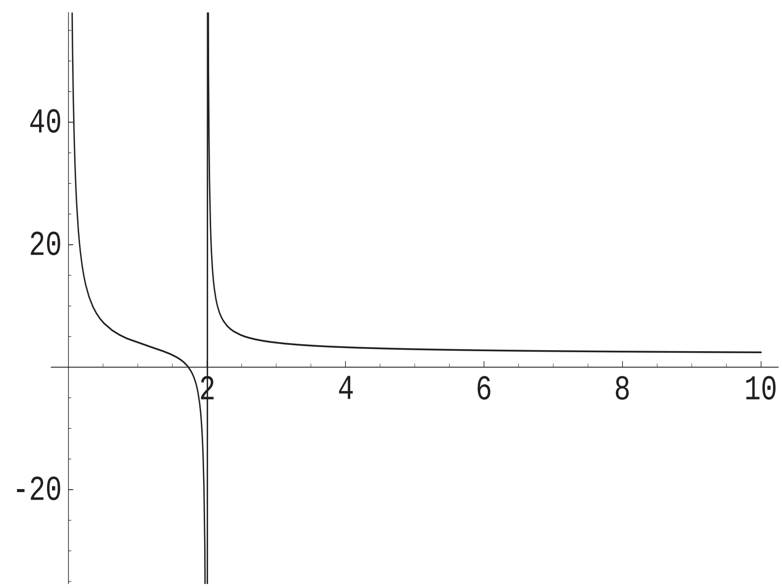

Figure 3: First function of η

But once we have found a starting point, the one-step procedure is again relatively simple: we calculate

$$\dot{\mathrm{l}}_\gamma(\theta|x) = \frac{\eta^2 x e^{-\eta x} - e^{-x}}{f(x|\gamma,\eta)} ,$$

$$
\begin{aligned}
\dot{\mathbf{l}}_\eta(\theta|x) &= \frac{2\gamma\eta x e^{-\eta x} - \gamma\eta^2 x^2 e^{-\eta x}}{f(x|\gamma,\eta)} \\
&= \frac{(2-\eta x)\gamma\eta x e^{-\eta x}}{f(x|\gamma,\eta)} \\
\ddot{\mathbf{l}}_{\gamma\gamma}(\theta|x) &= -\frac{(\eta^2 x e^{-\eta x} - e^{-x})^2}{f^2(x|\gamma,\eta)}, \\
\ddot{\mathbf{l}}_{\eta\gamma}(\theta|x) &= \frac{\eta x e^{-\eta x}(2-\eta x)}{f(x|\gamma,\eta)} - \frac{\gamma\eta x e^{-\eta x}(2-\eta x)[\eta^2 x e^{-\eta x} - e^{-x}]}{f^2(x|\gamma,\eta)}, \\
\ddot{\mathbf{l}}_{\eta\eta}(\theta|x) &= \frac{(2-\eta x)\eta x e^{-\eta x}}{f(x|\gamma,\eta)} - \frac{(2-\eta x)^2\gamma^2\eta^2 x^2 e^{-2\eta x}}{f^2(x|\gamma,\eta)}.
\end{aligned}
$$

Then

$$\check{\theta}_n = \overline{\theta}_n + \widehat{I}_n^{-1}\frac{1}{n}\dot{\mathbf{l}}_n(\overline{\theta}_n|\underline{X})$$

where

$$\dot{\mathbf{l}}_n(\overline{\theta}_n|\underline{X}) = \sum_{i=1}^n \dot{\mathbf{l}}_\theta(\overline{\theta}_n|X_i)$$

and

$$\widehat{I}_n = \frac{1}{n}\sum_{i=1}^n \ddot{\mathbf{l}}_n(\overline{\theta}_n|X_i).$$

3. Ferguson, ACLST, page 149, problem 2 modified as follows:
(a) Find the LR test statistic of the null hypothesis $H_0 : \mu = c\theta$ for any fixed number $c > 0$, and find the asymptotic distribution of the LR statistic under H_0.
(b) Does the theory of our chapter 4 (or Ferguson's chapter 22) apply directly?
(c) Does the local asymptotic power of your test depend on c?

Solution: (b) First, allow me to slightly re-name the parameters: I will assume that $X_1, \ldots, X_n$ are i.i.d. $\exp(\lambda)$ and $Y_1, \ldots, Y_n$ are i.i.d. $\exp(\mu)$, so that $\theta = (\lambda, \mu)$. Furthermore, we can recast the problem into the context of chapter 4 by considering the pairs of observations (X_i, Y_i), $i = 1, \ldots, n$ as i.i.d. with density

$$p_\theta(x,y) = p_{(\lambda,\mu)}(x,y) = \lambda e^{-\lambda x}1_{(0,\infty)}(x)\mu e^{-\mu y}1_{(0,\infty)}(y).$$

Now we are testing $H_0 : \mu = c\lambda$ versus $H_1 : \mu \neq c\lambda$. By a reparametrizaton, we can put this exactly in the setting of Section 4.2: if the original parameter is $\theta = (\lambda, \mu)$, then the new parameters $\gamma = (\gamma_1, \gamma_2)$ where $\gamma_1 \equiv \lambda$, $\gamma_2 \equiv \mu - c\lambda$. Then the null hypothesis H_0 becomes $H_0 : \gamma_2 = 0, \gamma_1 =$ anything.
(a) The MLE $\widehat{\theta}$ of $\theta = (\lambda, \mu)$ under H_1 is $\widehat{\theta} = (\widehat{\lambda}, \widehat{\mu})$ where $\widehat{\lambda} = 1/\overline{X}$ and $\widehat{\mu} = 1/\overline{Y}$. The MLE $\widehat{\theta}^0$ under H_0 is $(\widehat{\lambda}^0, c\widehat{\lambda}^0)$ where

$$\widehat{\lambda}^0 = 2/(\overline{X} + c\overline{Y}).$$

Now

$$l_n(\theta) = l_n(\lambda,\mu) = \sum_{i=1}^n \{\log\lambda - \lambda X_i + \log\mu - \mu Y_i\} = n\log\lambda + n\log\mu - n\overline{X}\lambda - n\overline{Y}\mu.$$

Thus the LR statistic for testing H_0 versus H_1 is given by

$$\begin{aligned} 2(l_n(\widehat{\theta}) - l_n(\widehat{\theta}^0)) &= 2n\left\{2\log\left(\frac{\overline{X}+c\overline{Y}}{2}\right) - \log(\overline{X}) - \log(c\overline{Y})\right\} \\ &\to_d \chi_1^2 \end{aligned}$$

under H_0.

(c) To compute the local asymptotic power of the LR test, we can reparametrize the problem by $\gamma \equiv (\gamma_1, \gamma_2)$ where $\gamma_1 \equiv \lambda$, $\gamma_2 \equiv \mu - c\lambda$. Then the null hypothesis H_0 becomes $H_0 : \gamma_2 = 0, \gamma_1 =$ anything. Then the problem fits in the context of Theorem 4.2.7: under P_{γ_n} with $\gamma_n = \gamma_0 + tn^{-1/2}$ for $\gamma_0 = (\gamma_{10}, 0)$ in the null hypothesis, we have

$$2\log\lambda_n \to_d \chi_1^2(\delta)$$

where the non-centrality parameter δ is given by $t_2^2 I_{22\cdot 1}(\gamma_0)$, and it remains only to compute $I_{22\cdot 1}$. By straightforward computation the information matrix for γ is given by

$$I(\gamma) = \begin{pmatrix} \frac{1}{\gamma_1^2} + \frac{c^2}{(c\gamma_1+\gamma_2)^2} & \frac{c}{(c\gamma_1+\gamma_2)^2} \\ \frac{c}{(c\gamma_1+\gamma_2)^2} & \frac{1}{(c\gamma_1+\gamma_2)^2} \end{pmatrix}.$$

Thus, under the null hypothesis $H_0 : \gamma_2 = 0$ we find that

$$I_{22\cdot 1}(\gamma_0) = I_{22}(\gamma_0) - I_{21}(\gamma_0)I_{11}^{-1}(\gamma_0)I_{12}(\gamma_0) = \frac{1/2}{c^2\gamma_1^2}$$

which does depend on c: the noncentrality power of the limiting distribution *decreases* as c^{-2} as c increases.

4. 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($\nu\eta$) distributions, respectively:

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

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 $W_1, \ldots, W_n$ with $W_i = X_i/Y_i$ and (X_i, Y_i) i.i.d. as (X, Y).

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 ratio $W = X/Y$.

D. Find the MLE $\hat{\nu}_n$ of ν based on (X_i, Y_i), $i = 1, \ldots, n$, and the MLE $\hat{\nu}_n^{(r)}$ of ν based on the reduced data $W_1, \ldots, W_n$. What are the limiting distributions of $\sqrt{n}(\hat{\nu}_n - \nu)$ and $\sqrt{n}(\hat{\nu}_n^{(r)} - \nu)$?

Solution: A. We compute, for $w \geq 0$,

$$\begin{aligned} P(W > w) &= P(X/Y > w) = P(X > wY) \\ &= \int_0^\infty \int_{wy}^\infty \eta^2 \nu e^{-\eta x} e^{-\eta\nu y} dx dy \\ &= \int_0^\infty \eta\nu e^{-\eta\nu y} \left(\int_{wy}^\infty \eta e^{-\eta x} dx \right) dy \\ &= \int_0^\infty \eta\nu e^{-\eta\nu y} e^{-\eta \nu y} dy \\ &= \eta\nu \int_0^\infty e^{-\eta(\nu+w)y} dy = \frac{\nu}{\nu+w} . \end{aligned}$$

[Alternatively, $\eta X \sim \text{Exp}(1)$, $\nu\eta Y \sim \text{Exp}(1)$ are independent so $2\eta X \sim \chi_2^2$, $2\nu\eta Y \sim \chi_2^2$ are independent. Thus $W/\nu = (2\eta X/2)/(2\eta\nu Y/2) \sim F_{2,2}$ with density given by (1.2.13).] Thus the density of W is given by

$$f_W(w;\nu) = \frac{\nu}{(\nu+w)^2} 1_{(0,\infty)}(w) .$$

Hence the score for ν based on observation of W is

$$\dot{\mathbf{l}}_\nu(w) = \frac{1}{\nu} - \frac{2}{\nu+w} ,$$

and the information for ν based on W is

$$\begin{aligned} I_W(\nu) &= E_\nu(\dot{\mathbf{l}}_\nu(W)^2) = -E_\nu \ddot{\mathbf{l}}_\nu \\ &= \frac{1}{\nu^2} - 2\int_0^\infty \frac{\nu}{(\nu+w)^4} dw = \frac{1}{3\nu^2} . \end{aligned}$$

Hence the information bound for estimation of ν based on observation of W is $3\nu^2$.
B. When we observe (X, Y), the scores for ν and η are given by

$$\dot{\mathbf{l}}_\nu(x,y) = \frac{1}{\nu} - \eta y , \qquad \dot{\mathbf{l}}_\eta(x,y) = \frac{2}{\eta} - (x + \nu y) ,$$

and the second derivatives are

$$\ddot{\mathbf{l}}_{\nu\nu}(x,y) = -\nu^{-2}, \qquad \ddot{\mathbf{l}}_{\eta\eta}(x,y) = -2/\eta^2 , \qquad \text{and} \qquad \ddot{\mathbf{l}}_{\nu\eta}(x,y) = -y .$$

Hence the information matrix for (ν, η) is given by

$$I(\nu,\eta) = \begin{pmatrix} 1/\nu^2 & 1/(\nu\eta) \\ 1/(\nu\eta) & 2/\eta^2 \end{pmatrix} .$$

Thus when η is known, the information for ν is $1/\nu^2$ and the information bound based on observation of (X, Y) is ν^2. When η is unknown the information for ν is

$$\begin{aligned} I_{\nu\nu\cdot\eta} &= I_{11\cdot 2} = I_{11} - I_{12} I_{22}^{-1} I_{21} \\ &= 1/\nu^2 - (\nu\eta)^{-2}\eta^2/2 = 1/(2\nu^2) , \end{aligned}$$

and the information bound for estimation of ν is $2\nu^2$. Thus lack of knowledge of η costs a factor of two in the bound.

C. Reduction to W cost a factor of 3 in the bound as compared to the bound based on (X, Y) when η is known and a factor of $3/2$ in the bound based on (X, Y) when η unknown. Thus reduction to W does not seem to be advisable. We can do better by basing estimation on *both* X and Y!

5. 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) independent with densities f_{ν,η_i} for $i = 1, \ldots, n$; i.e.

$$f_{\nu,\eta_i}(x_i, y_i) = \eta_i e^{-\eta_i x_i} \eta_i \nu e^{-\eta_i \nu y_i} 1_{(0,\infty)}(x_i) 1_{(0,\infty)}(y_i) \,. \tag{0.7}$$

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

$$\begin{aligned} p_{\nu,a,b}(x, y) &= \int_0^\infty f_{\nu,z}(x, y) g_{a,b}(z) dz \\ &= \frac{\nu}{b^2} \left(\frac{b}{b + x + \nu y} \right)^{a+2} \frac{\Gamma(a+2)}{\Gamma(a)} \,. \end{aligned} \tag{0.8}$$

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) The density of the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ is given by

$$p_{\nu,\underline{\eta}}(\underline{x}, \underline{y}) = \prod_{i=1}^n \left\{ \eta_i^2 \nu \exp(-\eta_i x_i) \exp(-\eta_i \nu y_i) \right\} = \nu^n \prod_{i=1}^n \eta_i^2 \exp(-\eta_i x_i) \exp(-\eta_i \nu y_i) \,.$$

Hence we calculate

$$\dot{\mathrm{l}}_\nu(\underline{x}, \underline{y}) = \frac{n}{\nu} - \sum_{i=1}^n \eta_i y_i \,, \qquad \dot{\mathrm{l}}_{\eta_i}(\underline{x}, \underline{y}) = \frac{2}{\eta_i} - (x_i + \nu y_i) \,,$$

$$\ddot{\mathrm{l}}_{\nu\nu}(\underline{x}, \underline{y}) = \frac{-n}{\nu^2} \,, \qquad \ddot{\mathrm{l}}_{\eta_i\eta_i}(\underline{x}, \underline{y}) = \frac{-2}{\eta_i^2} \,, \qquad \ddot{\mathrm{l}}_{\nu\eta_i}(\underline{x}, \underline{y}) = -y_i \,.$$

It follows easily that the information matrix for $(\nu, \underline{\eta})$ is given by

$$I_n(\nu, \underline{\eta}) = I_{\underline{X},\underline{Y}}(\nu, \underline{\eta}) = \begin{pmatrix} n/\nu^2 & 1/(\nu\eta_1) & \cdots & \cdots & 1/(\nu\eta_n) \\ 1/(\nu\eta_1) & 2/\eta_1^2 & 0 & \cdots & 0 \\ . & 0 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 1/(\nu\eta_n) & 0 & . & 0 & 2/\eta_n^2 \end{pmatrix},$$

Thus it follows that

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

and this is the information for ν in the presence of the nuisance parameters $\eta_1, \ldots, \eta_n$.

(b) and (c): 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:

$$I(\nu, a, b) = \begin{pmatrix} \frac{a+1}{a+3}\frac{1}{\nu^2} & \frac{1}{(a+2)\nu} & \frac{-a}{(a+3)b\nu} \\ \frac{1}{(a+2)\nu} & \frac{1}{a^2} + \frac{1}{(a+1)^2} & \frac{-2}{(a+2)b} \\ \frac{-a}{(a+3)b\nu} & \frac{-2}{(a+2)b} & \frac{2a}{(a+3)b^2} \end{pmatrix} . \tag{0.9}$$

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 a semiparametric generalization of the mixture (structural) model given by (0.8), we have

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

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, and van der Vaart (1996) gives a treatment of the limit theory for the MLE $\hat{\nu}_n$ of ν in this particular semiparametric mixture model.