Statistics 581, Problem Set 10 Solutions

Wellner; 12/9/2009

1. Suppose that (as in Lemma 5.2, page 38, Chapter 3 Notes) P and Q are two probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$ with densities p and q with respect to a $\sigma-$finite dominating measure μ, and P^n and Q^n denote the corresponding product measures on $(\mathcal{X}^n, \mathcal{A}_n)$ (of $X_1, \ldots, X_n$ i.i.d. as P or Q respectively).
(a) What is the relationship between $K(P^n, Q^n)$ and $K(P, Q)$, if any?
(b) If P is the Normal$(0, \sigma^2)$ distribution and Q is the Normal(μ, σ^2) distribution, compute $K(P, Q)$, $\rho(P, Q) = \int \sqrt{pq} d\mu$, and $H^2(P, Q)$.
(c) Use the results of (a) and (b) together with Lemma 5.2 to calculate $K(P^n, Q^n)$, $\rho(P^n, Q^n)$, and $H^2(P^n, Q^n)$ when P and Q are as in (b).
(d) Find a sequence μ_n so that, with Q_n being the Normal distribution with mean μ_n, the quantities $K(P^n, Q_n^n)$, $\rho(P^n, Q_n^n)$, and $H^2(P^n, Q_n^n)$ converge to finite limits as $n \to \infty$.

Solution: (a) Let $X_1, \ldots, X_n$ be i.i.d. P. Then

$$\begin{aligned} K(P^n, Q^n) &= E_{P^n} \log \prod_{i=1}^n \frac{p(X_i)}{q(X_i)} = E_{P^n} \sum_{i=1}^n \log \frac{p(X_i)}{q(X_i)} \\ &= \sum_{i=1}^n E_{P^n} \log \frac{p(X_i)}{q(X_i)} = n E_P \log \frac{p(X_1)}{q(X_1)} \\ &= nK(P, Q). \end{aligned}$$

(b) If $P = N(0, \sigma^2)$ and $Q = N(\mu, \sigma^2)$, then $p(x) = \phi(x/\sigma)/\sigma$, $q(x) = \phi((x - \mu)/\sigma)/\sigma$, and hence

$$\log \left(\frac{p(x)}{q(x)} \right) = -\frac{\mu x}{\sigma^2} + \frac{\mu^2}{2\sigma^2}.$$

Therefore

$$K(P, Q) = E_P \log \frac{p(X)}{q(X)} = -\frac{\mu}{\sigma^2} E_P(X) + \frac{\mu^2}{2\sigma^2} = \frac{\mu^2}{2\sigma^2}.$$

Also, using $E \exp(tX) = \exp(\sigma^2 t^2/2)$ if $X \sim N(0, \sigma^2)$,

$$\begin{aligned} \rho(P, Q) &= \int \sqrt{pq} d\mu = \int \sqrt{\frac{1}{\sigma} \phi(x/\sigma) \frac{1}{\sigma} \phi((x - \mu)/\sigma)} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int \exp(-x^2/(4\sigma^2)) \exp(-(x - \mu)^2/(4\sigma^2)) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int \exp(-x^2/(2\sigma^2)) \exp(\mu x/(2\sigma^2) - \mu^2/(4\sigma^2)) dx \\ &= \exp(-\mu^2/(4\sigma^2)) \exp(\mu^2/(8\sigma^2)) \\ &= \exp(-\mu^2/(8\sigma^2)). \end{aligned}$$

This implies that $H^2(P,Q) = 1 - \rho(P,Q) = 1 - \exp(-\mu^2/(8\sigma^2))$.
(c) It follows from the results of (a) and (b) that

$$\begin{aligned}
K(P^n, Q^n) &= nK(P,Q) = \frac{n\mu^2}{2\sigma^2}, \\
\rho(P^n, Q^n) &= \rho(P,Q)^n = \exp(-n\mu^2/(8\sigma^2)), \\
H^2(P^n, Q^n) &= 1 - \rho(P^n, Q^n) = 1 - \exp(-n\mu^2/(8\sigma^2)).
\end{aligned}$$

(d) If $\mu_n = c/\sqrt{n}$ with $c \in \mathbb{R}$, then it follows from (c) that

$$\begin{aligned}
K(P^n, Q_n^n) &= \frac{n\mu_n^2}{2\sigma^2} = \frac{c^2}{2\sigma^2}, \\
\rho(P^n, Q_n^n) &= \exp\left(-\frac{n\mu_n^2}{8\sigma^2}\right) = \exp\left(-\frac{c^2}{8\sigma^2}\right), \\
H^2(P^n, Q_n^n) &= 1 - \exp\left(-\frac{c^2}{8\sigma^2}\right),
\end{aligned}$$

which are all constant functions of n.

2. 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) If $X \sim P_\theta \in \mathcal{P}$, show that the distributions of $-\log X$ form a location and scale family from a Gumbel (extreme value) density on R.
(b) Use the result of (a) to construct method of moments estimators or quantile based estimators $\overline{\theta}_n$ of $\theta = (\alpha, \beta)$.
(c) Show that the method of moments or quantile estimators $\overline{\theta}_n$ of θ are asymptotically normal, and find the asymptotic distribution; i.e. show that

$$\sqrt{n}(\overline{\theta}_n - \theta) \rightarrow_d N_2(0, \Sigma) \qquad \text{for some} \qquad \Sigma.$$

[We will use these estimators as "starting points" approximate (or one-step) maximum likelihood estimators in the next problem .]

Solution: (a) Recall that $Y \equiv (X/\alpha)^\beta \sim \exp(1)$, and that $W \equiv -\log(Y) \sim$ Gumbel:

$$P(W \le w) = P(-\log(Y) \le w) = P(Y \ge e^{-w}) = \exp(-e^{-w}).$$

Thus it follows that

$$W = -\log(Y) = \beta\{-\log(X) + \log(\alpha)\},$$

or equivalently that

$$T \equiv -\log(X) = \frac{1}{\beta}W - \log(\alpha).$$

Thus the distributions of $T \equiv -\log(X)$ form a location - scale family of the Gumbel (extreme value) distribution with d.f. $\exp(-\exp(-x))$.

(b) Now $T = -\log X$ has

$$E(T) = \frac{\gamma}{\beta} - \log \alpha, \qquad Var(T) = \frac{1}{\beta^2}\frac{\pi^2}{6}$$

where $\gamma = .577...$ is Euler's constant. Since $\overline{T} = -2.9518...$ and $\tilde{S}_T = 2.01103...$ (biased variance estimator) or $S_T = 1.92541...$ (unbiased variance estimator), moment estimators of (α, β) based on (8) are given by

$$\overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}}\frac{1}{\tilde{S}_T} = .66612..., \qquad \overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}}\frac{1}{S_T} = .63775...$$

and for these two estimators of β,

$$\overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 45.5285, \qquad \overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 47.317...$$

respectively for the given data in problem 3 below.

(c) Asymptotic normality of $(\overline{\alpha}_n, \overline{\beta}_n)$ follows from joint asymptotic normality of $(\overline{T}_n, S_T^2)$ and the delta method: by the multivariate CLT and Slutsky's theorem

$$\left(\begin{array}{c} \sqrt{n}(\overline{T} - ET)/\sigma \\ \sqrt{n}(S_T^2 - \sigma_T^2)/(\sqrt{2}\sigma_T^2) \end{array} \right) \rightarrow_d \underline{Z} \sim N_2(0, \Sigma)$$

where, with $\gamma_1 \equiv E(T - E(T))^3/\sigma_T^3$, $\gamma_2 \equiv E(T - ET)^4/\sigma_T^4 - 3$,

$$\Sigma = \left(\begin{array}{cc} 1 & \gamma_1/\sqrt{2} \\ \gamma_1/\sqrt{2} & 1 + \gamma_2/2 \end{array} \right).$$

Then since $(\overline{\alpha}, \overline{\beta}) = g(\overline{T}, S_T^2)$ and $(\alpha, \beta) = g(E_\theta T, Var_\theta(T))$ where $g \equiv (g_1, g_2) : R^2 \rightarrow R^2$ is defined by

$$g_1(x, y) = \exp(\frac{\gamma\sqrt{6}}{\pi}\sqrt{y} - x),$$

$$g_2(x, y) = \frac{\pi/\sqrt{6}}{\sqrt{y}},$$

it follows by the delta method with $\underline{\tilde{Z}} \equiv (Z_1, \sqrt{2}\sigma_T^2 Z_2)$ that

$$\sqrt{n}\left((\overline{\alpha}_n, \overline{\beta}_n)^T - (\alpha, \beta)^T\right) \rightarrow_d \nabla g \underline{\tilde{Z}}$$

where

$$\nabla g \equiv \nabla g(E_\theta T, \mathrm{Var}_\theta T) = \left(\begin{array}{cc} -\alpha & (3\gamma/\pi^2)\alpha\beta \\ 0 & -3\beta^3/\pi^2 \end{array} \right).$$

3. (Problem #2, continued).
(a) Does a maximum likelihood estimate of $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ exist? Is it unique? (See 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 (or quantile) estimators $\overline{\theta}_n$ as the preliminary estimators based on the following data (with $n = 12$):

$$1,\ 1.3,\ 1.7,\ 3.2,\ 10.7,\ 24.3,\ 51.2,\ 77.1,\ 93.7,\ 105,\ 111,\ 305.$$

[These are failure times in seconds for "breakdown" of an insulating fluid between two electrodes subject to a voltage of 40 kV. – from Nelson, *Applied Life Data Analysis*, page 252, but with some modifications or "recording errors".]
(c) Compute the maximum likelihood estimator $\hat{\theta}_n$, and compare it with the one step estimator computed in E.

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

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

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 $a_i \equiv \log X_i$, $p_i \equiv X_i^\beta / \sum_j X_j^\beta \geq 0$, $\sum_i p_i = 1$. 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 $p_{(i)} \equiv X_{(i)}^\beta / \sum_j X_j^\beta$,

$$\begin{aligned} p_{(i)} &= \frac{1}{(\frac{X_{(1)}}{X_{(i)}})^\beta + \ldots + (\frac{X_{(n)}}{X_{(i)}})^\beta} \\ &\to \begin{cases} 0, & i \leq n \quad \text{(so } X_{(n)}/X_{(i)} > 1) \\ 1, & i = n \quad \text{(so } X_{(j)}/X_{(n)} < 1,\ j < n) \end{cases} \end{aligned}$$

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 method of moment estimators were computed in 3(b) above. The one step estimator using $\hat{I}(\overline{\theta}_n) = I(\overline{\theta}_n)$ is

$$\check{\theta}_n \equiv \overline{\theta}_n + \hat{I}_n^{-1}(\overline{\theta}_n)(\frac{1}{n}\dot{l}(\overline{\theta}_n)) = (48.2169\ldots, 0.63632\ldots).$$

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

$$\check{\theta}_n = (48.1805\ldots, 0.636249\ldots),$$

(c) The maximum likelihood estimate $\hat{\theta}_n = (48.2162..., 0.636229...)$, but note that the likelihood surface is quite flat as a function of α as shown in the plots on the following pages.

Mathematica input for moment and one-step estimators:

```
(* Here is the data:  *)
x={1,1.3,1.7,3.2,10.7,24.3,51.2,77.1,93.7,105,111,305.}
(* NSS is the sample size *)
NSS := Length[x]
(* First transform to -Log[x]:  *)
t := -Log[x]
(* Now compute Mean and Variance of y *)
tbar := Sum[t[[i]], {i,1,NSS}]/NSS
tsquaredbar := Sum[t[[i]]^2 ,{i,1,NSS}]/NSS
Stt := tsquaredbar - tbar^2
tbar
Sqrt[Stt]
(* For the Method of Moment Estimators,  *)
(* compute mean and variance of standard Gumbel *)
VarGumbel := (Pi^2)/6
MeanGumbel :=  EulerGamma
(* Then the Moment estimators of beta and alpha are: *)
betabar = N[Sqrt[VarGumbel/Stt]]
alphabar = N[Exp[-tbar + MeanGumbel/betabar]]
thetabar = {alphabar, betabar}

(* Now for the One-Step Estmators of Theta = (a,b) :  *)
(*   We compute the One-Step Based on Two Estimators  *)
(*    of the information matrix  I( theta )           *)
(* 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 *)
(*   and Infminus1 is the inverse informaton matrix *)
Inf[a_,b_] := { {b^2/a^2 , aa/a}, {aa/a, bb/b^2}} ;
Infminus1[a_,b_] := Inverse[Inf[a,b]]

(* L  is the log-likelihood *)
L[a_,b_] := Sum[Log[f[x[[i]], a,b]], {i,1,NSS} ] ;

(* Sc is the vector of Scores *)
(*   for all the data /sample size *)

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}]/NSS
Inf[alphabar,betabar]
Infminus1[alphabar,betabar]
```

```
Sc[alphabar,betabar]
NSS
Delta1 := Infminus1[alphabar,betabar].Sc[alphabar,betabar]
Delta1
thetaCaret1 :=
{alphabar,betabar}  + {Delta1[[1]],Delta1[[2]]}
thetaCaret1

LDotDot[a_,b_] :=
Sum[{{(-b/(a^2))(((x[[i]]/a)^b)*(1+b) -1),
  (1/a)*(((x[[i]]/a)^b)*
        (1 + Log[(x[[i]]/a)^b]) - 1 )},
 {(1/a)*(((x[[i]]/a)^b)*
        (1 + Log[(x[[i]]/a)^b]) - 1 ),
  (-1/(b^2))*(1 + ((x[[i]]/a)^b)*(Log[(x[[i]]/a)^b])^2)
 }
}, {i,1,NSS}]/NSS
Inf2[a_,b_] := - LDotDot[a,b]
Inf2[alphabar,betabar]
Infminus2[a_,b_] := Inverse[Inf2[a,b]]
Infminus2[alphabar,betabar]
Delta2 := Infminus2[alphabar,betabar].Sc[alphabar,betabar]
Delta2
thetaCaret2 :=
{alphabar,betabar}  + {Delta2[[1]],Delta2[[2]]}
thetaCaret2
```

Mathematica output for one-step estimators

```
During evaluation of In[1]:= Here is the data:

Out[3]= {1, 1.3, 1.7, 3.2, 10.7, 24.3, 51.2, 77.1, 93.7, 105, 111, 305.}

Out[7]= {0., -0.262364, -0.530628, -1.16315, -2.37024, -3.19048, -3.93574, -4.3451
-4.5401, -4.65396, -4.70953, -5.72031}

During evaluation of In[1]:= Mean of T = - Log[x]

Out[10]= -2.9518

During evaluation of In[1]:= Standard deviation of T

Out[12]= 2.01103

Out[13]= 12.4203
```

```
Out[14]= 3.7072

During evaluation of In[1]:= Biased estimator of std. dev

Out[16]= 1.92541

During evaluation of In[1]:= Moment estimator of beta, version 1:

Out[23]= 0.637759

During evaluation of In[1]:= Moment estimator of beta, version 2:

Out[25]= 0.666118

During evaluation of In[1]:= Moment estimator of alpha, version 1:

Out[27]= 47.317

During evaluation of In[1]:= Moment estimator of alpha, version 2:

Out[29]= 45.5285

During evaluation of In[1]:= theta bar estimator, version 1

Out[31]= {47.317, 0.637759}

During evaluation of In[1]:= theta bar estimator, version 2

Out[33]= {45.5285, 0.666118}

During evaluation of In[1]:= Information matrix estimator based on thetabar

Out[52]= {{0.000181669, -0.00893515}, {-0.00893515, 4.48369}}

During evaluation of In[1]:= inverse information matrix estimator based on thetaba

Out[54]= {{6102.68, 12.1615}, {12.1615, 0.247266}}

During evaluation of In[1]:= vector of scores evaluated at thetabar

Out[56]= {0.000176355, -0.0144948}

During evaluation of In[1]:= sample size n (NSS in the program)

Out[58]= 12

During evaluation of In[1]:= adjustment to the preliminary estimator
```

```
Out[61]= {0.899961, -0.00143933}

During evaluation of In[1]:= resulting one step estimator; based on theoretical In

Out[64]= {48.2169, 0.63632}

During evaluation of In[1]:= information matrix based on - Hessian of log-likeliho

Out[68]= {{0.000187773, -0.00940704}, {-0.00940704, 4.21775}}

During evaluation of In[1]:= inverse information matrix from Hessian

Out[71]= {{5995.5, 13.3721}, {13.3721, 0.266918}}

During evaluation of In[1]:= adjustment to the preliminary estimator

Out[74]= {0.863513, -0.00151069}

During evaluation of In[1]:= resulting Hessian based version of one-step estimator

Out[77]= {48.1805, 0.636249}
```

Mathematica input for maximum likelihood estimators:

```
Clear[a,b,ahat,bhat]
(* Here is the data:  *)
x={1,1.3,1.7,3.2,10.7,24.3,51.2,77.1,93.7,105,111,305.}

(* 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}];
```

```
aprof[b_] := (Sum[x[[i]]^b, {i,1,NSS}]/NSS )^(1/b)

Plot3D[L[a,b], {a,2,100}, {b,.05,2.0}]
ContourPlot[L[a,b],{a,2,100},{b,.05,2.0}]

Plot[L[aprof[b],b],{b,.4,.8}]
FindMinimum[-L[aprof[b],b],{b,.63}]
FindMinimum[-L[aprof[b],b],{b,.63}][[2]]
bhat=Replace[b,FindMinimum[-L[aprof[b],b],{b,.63}][[2]]]
ahat=aprof[bhat]

MinL=FindMinimum[-L[a,b], {a,47},{b,.62}]
MinL[[1]]
MinL[[2]]
```

Mathematica output:

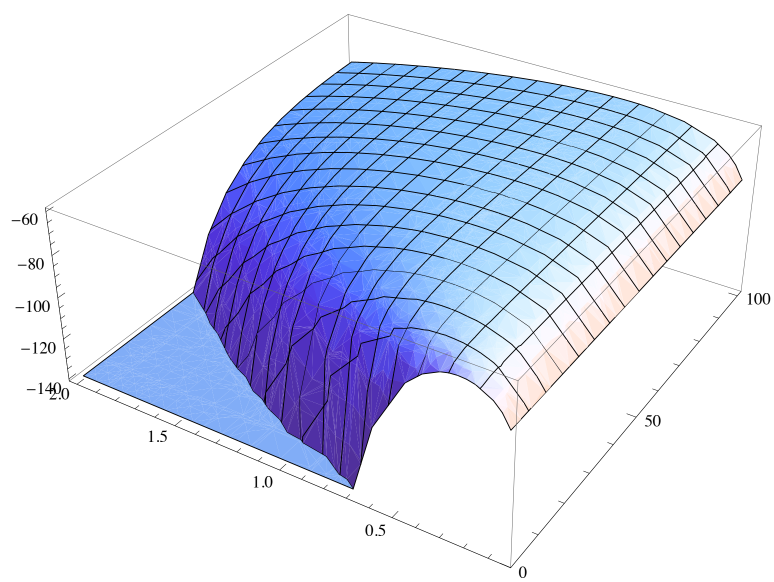

Figure 1: Weibull likelihood.

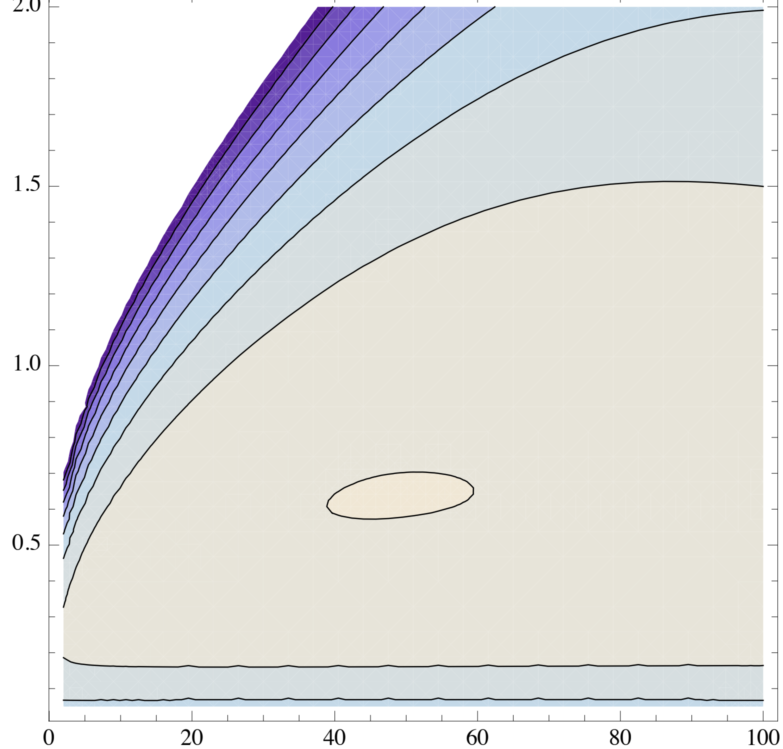

Figure 2: Contour plot Weibull likelihood.

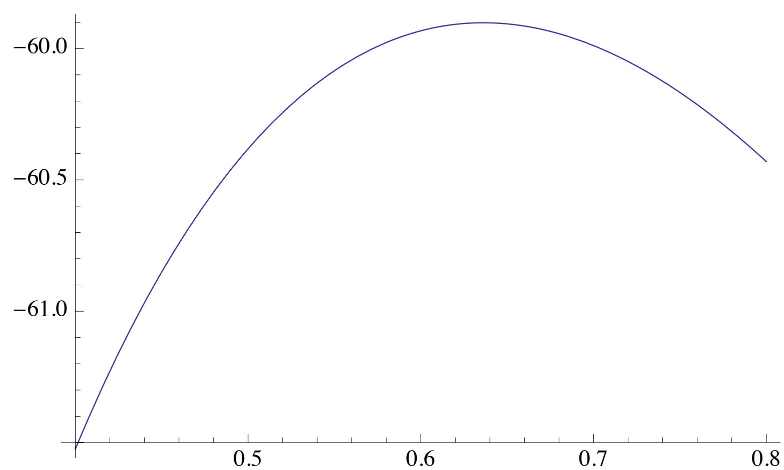

Figure 3: Weibull profile likelihood.

```
{1, 1.3, 1.7, 3.2, 10.7, 24.3, 51.2, 77.1, 93.7, 105, 111, 305.}

12

{59.9017, {b -> 0.636229}}
```

```
{b -> 0.636229}

0.636229

48.2162

{59.9079, {a -> 47., b -> 0.62}}

59.9079

{a -> 47., b -> 0.62}
```

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 $w \geq 0$,

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

[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

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

and the information for ν based on W is

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

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

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

and the second derivatives are

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

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. If the paired exponential model is true, then 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 if we believe the exponential model. We can do better by basing estimation on *both* X and Y! On the other hand, if we suppose the true density is not of the exponential form hypothesized, say

$$f_{\nu,\eta}(x,y) = \eta g_0(\eta x)\eta\nu g_0(\eta\nu y) 1_{(0,\infty)}(x) 1_{(0,\infty)}(y)$$

where g_0 is some other density function on $[0,\infty)$ (for example $\Gamma(2,1)$ or half-normal, or Pareto, or ...), then it could easily be the case that our estimate of ν based on the W_i's would continue to be a consistent estimator of ν and the increase in variance might not be of primary concern.

D. The MLE $\hat{\theta}_n = (\hat{\nu}_n, \hat{\eta}_n)$ of $\theta = (\nu, \eta)$ based on the (X_i, Y_i) pairs, $i = 1, \ldots, n$ solves the score equations

$$0 = \sum_{i=1}^{n} \dot{l}_\nu(X_i, Y_i) = \sum_{i=1}^{n} \left(\frac{1}{\hat{\nu}_n} - \hat{\eta}_n Y_i \right),$$
$$0 = \sum_{i=1}^{n} \dot{l}_\eta(X_i, Y_i) = \sum_{i=1}^{n} \left(\frac{2}{\hat{\eta}_n} - (X_i + \hat{\nu}_n Y_i) \right).$$

Equivalently

$$\frac{1}{\hat{\nu}_n} = \hat{\eta}_n \overline{Y}_n, \qquad \text{and} \qquad \frac{1}{\hat{\eta}_n} = \frac{1}{2}(\overline{X}_n + \hat{\nu}_n \overline{Y}_n);$$

and hence

$$\frac{1}{\hat{\eta}_n} = \hat{\nu}_n \overline{Y}_n = \frac{1}{2}(\overline{X}_n + \hat{\nu}_n \overline{Y}_n).$$

Solving the last equality for $\hat{\nu}_n$ yields $\hat{\nu}_n = \overline{X}_n / \overline{Y}_n$, and then we find

$$\frac{1}{\hat{\eta}_n} = \overline{X}_n, \qquad \text{or} \qquad \hat{\eta}_n = \frac{1}{\overline{X}_n}.$$

From our general theory

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d N_2(0, I(\theta)^{-1}))$$

where

$$I^{-1}(\theta) = \begin{pmatrix} I_{11\cdot 2}^{-1} & -I_{11\cdot 2}^{-1} I_{12} I_{22} \\ -I_{22\cdot}^{-1} & I_{22\cdot 1}^{-1} \end{pmatrix} = \begin{pmatrix} 2\nu^2 & -\nu\eta \\ -\nu\eta & \eta^2 \end{pmatrix};$$

In particular, $\sqrt{n}(\hat{\nu}_n - \nu) \to_d N(0, 2\nu^2)$. (This can be easily checked via joint asymptotic normality of $\sqrt{n}(\overline{X}_n - 1/\eta\ ,\ \overline{Y}_n - 1/(\nu\eta))$ and the delta - method with $g(u, v) = (u/v, 1/u)$.)

The estimator $\hat{\nu}_n^{(r)}$ of ν based on the reduced data $W_i = X_i / Y_i$, $i = 1, \ldots, n$ is the solution of

$$0 = \sum_{i=1}^{n} \dot{l}_\nu(W_i) = \sum_{i=1}^{n} \left(\frac{1}{\hat{\nu}^{(r)}} - \frac{2}{\hat{\nu}_n^{(r)} + W_i} \right) = \frac{n}{\hat{\nu}^{(r)}} - 2 \sum_{i=1}^{n} \frac{1}{\hat{\nu}_n^{(r)} + W_i};$$

or, in other words,

$$\frac{1}{\hat{\nu}_n^{(r)}} = \frac{2}{n} \sum_{i=1}^{n} \frac{1}{\hat{\nu}_n^{(r)} + W_i}.$$

This apparently does not have a closed form solution. On the other hand it is clear that the solution is unique. The equation can be written as the solution of the equality

$$1 = \frac{2}{n} \sum_{i=1}^{n} \frac{\nu}{\nu + W_i}$$

where the right side is a monotone increasing function of ν which is 0 at $\nu = 0$, and which converges to 2 as $\nu \to \infty$.

Again from our general theory it follows that

$$\sqrt{n}(\hat{\nu}_n^{(r)} - \nu) \to_d N(0, 3\nu^2)$$

as $n \to \infty$.