**Statistics 581, Problem Set 10 Solutions**

Wellner; 12/3/2008

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 $\mu$, 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$(\mu, \sigma^2)$ 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 $\mu_n$ so that, with $Q_n$ being the Normal distribution with mean $\mu_n$, 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

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

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 $p_\theta$.
(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 $\theta$ are asymptotically normal, and find the asymptotic distribution; i.e. show that

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

[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 $\overline{T} = -3.02984...$ and $\tilde{S}_T = 2.06378...$ (biased variance estimator) or $S_T = 2.1555...$ (unbiased variance estimator), moment estimators of $(\alpha, \beta)$ based on (8) are given by

$$\overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}}\frac{1}{\tilde{S}_T} = .62145..., \qquad \overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}}\frac{1}{S_T} = .5949...$$

and for these two estimators of $\beta$,

$$\overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 52.3865, \qquad \overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 54.59...$$

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

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

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

$$\Sigma = \begin{pmatrix} 1 & \gamma_1/\sqrt{2} \\ \gamma_1/\sqrt{2} & 1 + \gamma_2/2 \end{pmatrix}.$$

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 $\tilde{\underline{Z}} \equiv (Z_1, \sqrt{2}\sigma_T^2 Z_2)$ that

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

where

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

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 $\theta$ using

the method of moment (or quantile) estimators $\overline{\theta}_n$ as the preliminary estimators based on the following data (with $n = 12$):

$$1,\ 1,\ 2,\ 3,\ 12,\ 25,\ 46,\ 56,\ 79,\ 125,\ 323,\ 417.$$

[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 $\gamma$ 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

$$\check{\theta}_n \equiv \overline{\theta}_n + \hat{I}_n^{-1}(\overline{\theta}_n)(\frac{1}{n}\dot{l}(\overline{\theta}_n)) = (55.1837\ldots, 0.550762\ldots)$$

or, using the unbiased estimator of variance in the preliminary estimator

$$\check{\theta}_n \equiv \overline{\theta}_n + \hat{I}_n^{-1}(\overline{\theta}_n)(\frac{1}{n}\dot{l}(\overline{\theta}_n)) = (56.777\ldots, 0.5615\ldots)\ .$$

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

$$\check{\theta}_n = (55.903\ldots, 0.5628\ldots),$$

(c) The maximum likelihood estimate $\hat{\theta}_n = (56.6464..., 0.564124...)$; see the following pages.

**Mathematica input for moment and one-step estimators:**

```
(* Here is the data:  *)
x={1,1,2,3,12,25,46,56,68,109,323,417 }
(* 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, 2, 3, 12, 25, 46, 56, 79, 125, 323, 417}

Out[7]= {0., 0., -0.693147, -1.09861, -2.48491, -3.21888, -3.82864, -4.02535, \
-4.36945, -4.82831, -5.77765, -6.03309}

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

Out[10]= -3.02984

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

Out[12]= 2.15555
```

```
Out[13]= 13.4391

Out[14]= 4.25921

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

Out[16]= 2.06378

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

Out[23]= 0.594998

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

Out[25]= 0.621455

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

Out[27]= 54.5954

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

Out[29]= 52.3865

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

Out[31]= {54.5954, 0.594998}

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

Out[33]= {52.3865, 0.621455}

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

Out[52]= {{0.000118773, -0.00774396}, {-0.00774396, 5.15131}}

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

Out[54]= {{9334.28, 14.0322}, {14.0322, 0.21522}}

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

Out[56]= {0.000518316, -0.189292}

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]= {2.18192, -0.0334664}

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

Out[64]= {56.7773, 0.561532}

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

Out[68]= {{0.000133916, -0.010678}, {-0.010678, 5.45659}}

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

Out[71]= {{8847.99, 17.3147}, {17.3147, 0.217148}}

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

Out[74]= {1.30852, -0.0321299}

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

Out[77]= {55.9039, 0.562868}
```

Here is the data:

Out[3]= 1, 1, 2, 3, 12, 25, 46, 56, 79, 125, 323, 417

Out[7]= 0., 0., -0.693147, -1.09861, -2.48491, -3.21888, -3.82864, -4.02535, -4.36945, -4.82831, -5.77765, -6.03309

Mean of T = - Log[x] Out[10]= -3.02984

Standard deviation of T Out[12]= 2.15555

Out[13]= 13.4391 Out[14]= 4.25921

Biased estimator of std. dev Out[16]= 2.06378

Moment estimator of beta, version 1: Out[23]= 0.594998

Moment estimator of beta, version 2: Out[25]= 0.621455

Moment estimator of alpha, version 1: Out[27]= 54.5954

Moment estimator of alpha, version 2: Out[29]= 52.3865

theta bar estimator, version 1 Out[31]= 54.5954, 0.594998

theta bar estimator, version 2 Out[33]= 52.3865, 0.621455

Information matrix estimator based on thetabar Out[52]= 0.000118773, -0.00774396, -0.00774396, 5.15131 inverse information matrix estimator based on thetabar Out[54]= 9334.28, 14.0322, 14.0322, 0.21522

vector of scores evaluated at thetabar Out[56]= 0.000518316, -0.189292

sample size n (NSS in the program) Out[58]= 12

adjustment to the preliminary estimator Out[61]= 2.18192, -0.0334664

resulting one step estimator; based on theoretical Information matrix Out[64]= 56.7773, 0.561532

information matrix based on - Hessian of log-likelihood Out[68]= 0.000133916, -0.010678, -0.010678, 5.45659

inverse information matrix from Hessian Out[71]= 8847.99, 17.3147, 17.3147, 0.217148

adjustment to the preliminary estimator Out[74]= 1.30852, -0.0321299

resulting second version of one-step estimator Out[77]= 55.9039, 0.562868

**Mathematica input for maximum likelihood estimators:**

```
Clear[a,b,ahat,bhat]
(* Here is the data:  *)
x={1,1,2,3,12,25,46,56,79,125,323,417 }
(* 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}]

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]
FindMinimum[-L[a,b], {a,50},{b,1}]
```

**Mathematica output:**

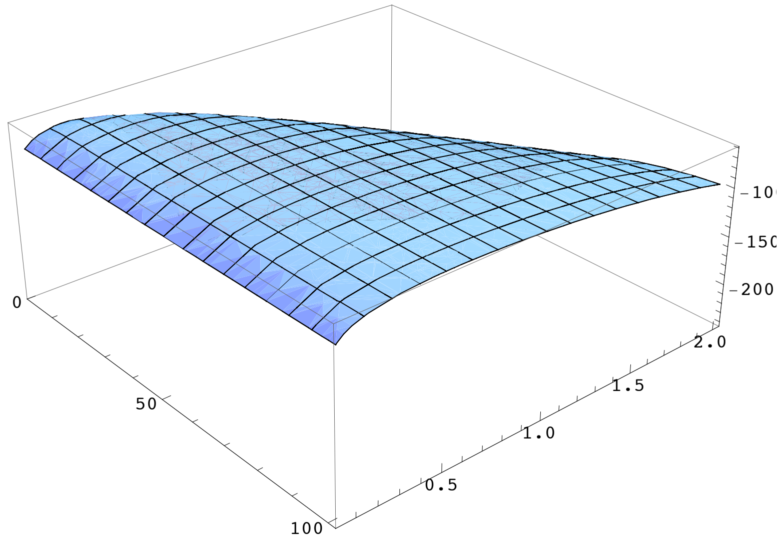


Figure 1: Weibull likelihood.

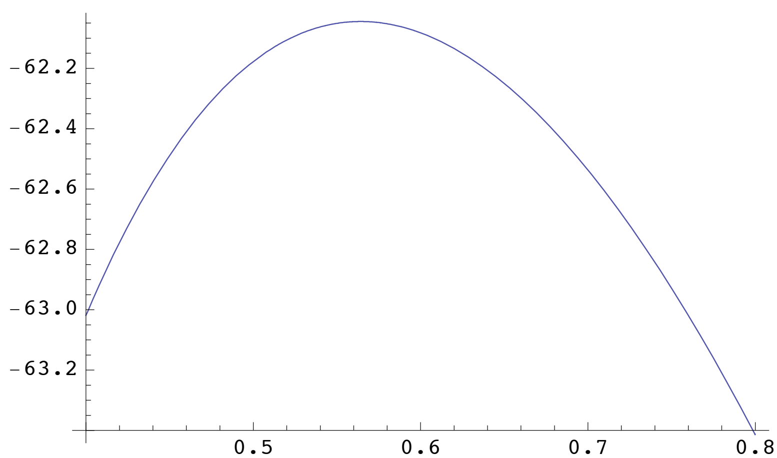


Figure 2: Weibull profile likelihood.

```
{62.0446, {b \mapsto 0.564124}}
%{61.7036,{b \mapsto 0.564847}}
{b \mapsto 0.564124}
%{b\mapsto 0.564847}
```

```
0.564124
%0.564847
56.6464
%54.9218
{62.0446, {a \mapsto 56.6464, b \mapsto 0.564124}}
```

4. (a) Ferguson, ACLST, page 139, problem 3.
(b) What if Ferguson's density $f(x|\theta)$ with $\theta \in (0,1)$ is replaced by $\theta = (\gamma, \eta) \in (0,1) \times (0,\infty)$ and

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

Can you estimate $\gamma$ and $\eta$ by the method of moments? Can you improve method of moment estimators via one-step estimators?

**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 a method of moments estimator $\overline{\theta}_n$ of $\theta$ is given by $\overline{X}_n - 1$. Now

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

Thus

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

Hence it follows by the CLT that

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

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

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

where

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

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 $\infty$ 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 $\theta$ increases from 0 to 1. Hence the gain in efficiency by use of the efficient one-step estimator is quite large for $\theta$ near 1. See the plot of $1/I(\theta)$ and $1 + 2\theta - \theta^2$ below.

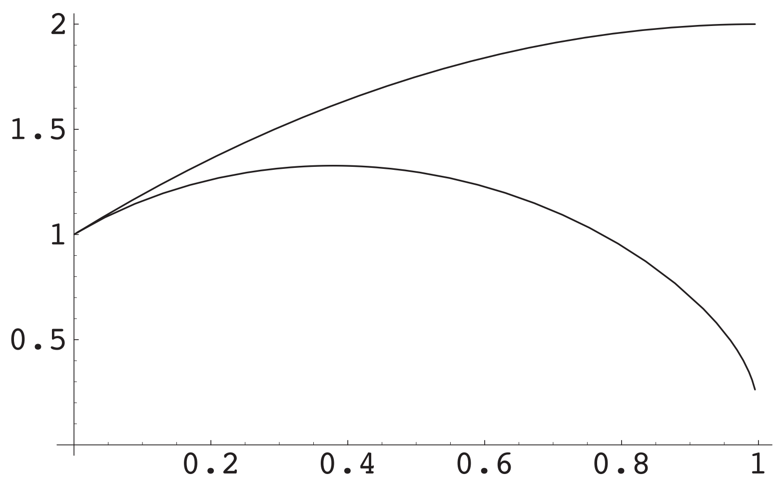


Figure 3: $1/I(\theta)$ and $1 + 2\theta - \theta^2$, $0 < \theta < 1$

(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 an estimator based on empirical estimators of the distribution function at two points as follows: note that

$$S_\theta(x) \equiv 1 - F_\theta(x) = (1-\gamma)e^{-x} + \gamma(1+\eta x)e^{-\eta x},$$

where we can estimate $S_\theta(x)$ by the empirical survival function $\mathbb{S}_n(x) \equiv 1 - \mathbb{F}_n(x)$. Thus we fix two points $x$ and $y$ with $0 < x < y < \infty$. Then

$$S_1 \equiv 1 - F_\theta(x) = (1-\gamma)e^{-x} + \gamma(1+\eta x)e^{-\eta x},$$
$$S_2 \equiv 1 - F_\theta(y) = (1-\gamma)e^{-y} + \gamma(1+ \ etay)e^{-\eta y},$$

give two equations in the two unknown parameters $\gamma$ and $\eta$. With $\hat{S}_1 \equiv \mathbb{S}_n(x)$ and $\hat{S}_2 \equiv \mathbb{S}_n(y)$, replacing $S_1$ and $S_2$ on the left side, these yield a preliminary estimator $\overline{\theta}_n = (\overline{\gamma}_n, \overline{\eta}_n)$. It is easily seen that the solution $\overline{\gamma}_n$ is given by

$$\overline{\gamma}_n = \frac{e^x \hat{S}_1 - 1}{B_1(\overline{\eta}_n) - 1} = \frac{e^y \hat{S}_2 - 1}{B_2(\overline{\eta}) - 1} \tag{0.6}$$

where

$$B_1(\eta) \equiv e^x e^{-\eta x}(1+\eta x), \qquad B_2(\eta) \equiv e^y e^{-\eta y}(1+\eta y)$$

and where $\overline{\eta}_n$ is the value of $\eta$ which makes the two expressions on the right side of (0.6) equal. Figure 4 shows a plot of the difference of the two functions of $\eta$ on the right side of (0.6) for hypothetical values of $x$, $y$, $\hat{S}_1$, and $\hat{S}_2$.

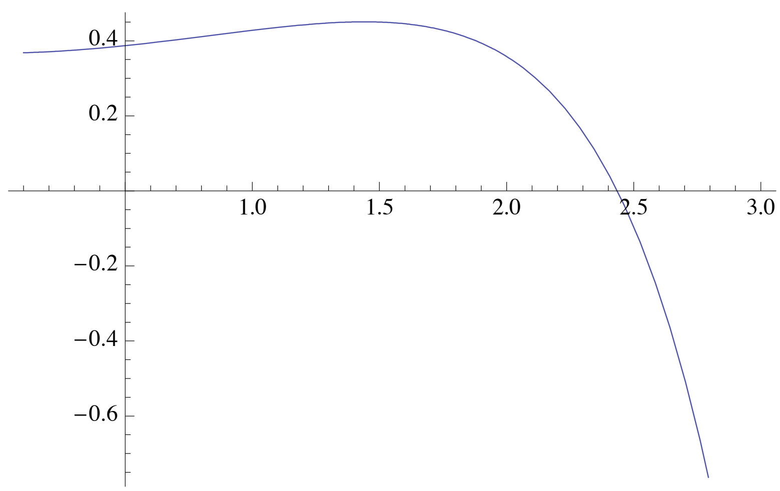


Figure 4: First minus second function of $\eta$

Once we have found a preliminary estimator $\overline{\theta}_n$, the one-step procedure is again relatively simple: we calculate

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

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

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

5. Ferguson, ACLST, page 118, problem 3. [Also see Lehmann and Casella, Example 7.9, page 482.]

**Solution:** (a) The likelihood is given by

$$\begin{aligned} L(\underline{\mu}, \sigma^2) &= \prod_{j=1}^{d} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left( -\frac{(X_{ij} - \mu_i)^2}{2\sigma^2} \right) \\ &= \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^{nd} \exp\left( -\frac{1}{2\sigma^2} \sum_{j=1}^{d} \sum_{i=1}^{n} (X_{ij} - \mu_i)^2 \right) \end{aligned}$$

and hence

$$\begin{aligned} l(\underline{\mu}, \sigma^2) &= -\frac{nd}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{d} \sum_{i=1}^{n} (X_{ij} - \mu_i)^2 + \text{constant} \\ &= -\frac{nd}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \left\{ \sum_{j=1}^{d} \sum_{i=1}^{n} (X_{ij} - \hat{\mu}_i)^2 + d \sum_{i=1}^{n} (\hat{\mu}_i - \mu_i)^2 \right\} + \text{constant.} \end{aligned}$$

where $\hat{\mu}_i = d^{-1} \sum_{j=1}^{d} X_{i,j}$ for $i = 1, \ldots, n$. This is easily seen to be maximized by

$$\begin{aligned} &\mu_i = \hat{\mu}_i, \qquad i = 1, \ldots, n, \\ &\sigma^2 = \hat{\sigma}^2 = \frac{1}{nd} \sum_{j=1}^{d} \sum_{i=1}^{n} (X_{ij} - \hat{\mu}_i)^2 = \frac{1}{n} \sum_{i=1}^{n} S_i^2 \end{aligned}$$

where

$$S_i^2 = \frac{1}{d} \sum_{j=1}^{d} (X_{i,j} - \hat{\mu}_i)^2.$$

(b) Note that the random variables $\{S_i^2\}_{i=1}^n$ defined in (a) are i.i.d. and $dS_i^2/\sigma^2 \sim \chi^2_{d-1}$. Therefore

$$E(S_1^2) = \frac{d-1}{d} \sigma^2$$

It follows from the strong law of large numbers that

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} S_i^2 \to_{a.s.} \frac{d-1}{d} \sigma^2$$

as $n \to \infty$. Our Theorem 4.1.2 on consistent roots of the likelihood equations does not apply because, in the current problem, the dimension of the parameter space $\Theta = \mathbb{R}^n \times \mathbb{R}^+$ is $n+1$, which grows with the sample size $n$.
(c) A consistent estimator of $\sigma^2$ is given by

$$\tilde{\sigma}_n^2 \equiv \frac{d}{d-1}\hat{\sigma}^2 = \frac{1}{(d-1)n}\sum_{j=1}^{d}\sum_{i=1}^{n}(X_{i,j} - \hat{\mu}_i)^2.$$