Statistics 581, Problem Set 9 Solutions

Wellner; 12/6/2006

1. Lehmann and Casella, TPE, problem 6.3.22, page 503, reworded as follows. (In other words, prove (vi) of theorem 1.5, page 5, chapter 4 notes). Suppose that $X_1, \ldots, X_n$ are i.i.d. with density p_θ, $\theta \in \Theta \subset R^k$, satisfying the hypotheses of theorem 4.1, page 429 (the Cramér conditions given in (A) - (D) on page 429). Show that the following Local Asymptotic Normality (LAN) result holds for the (local) log- likelihood ratios: with

$$L_n(\theta) \equiv \log(\prod_{i=1}^n p_\theta(X_i)) = \sum_{i=1}^n \log p_\theta(X_i),$$

for a fixed $\theta_0 \in \Theta$,

$$\begin{aligned} L_n(\theta_0 + n^{-1/2}\underline{t}) - L_n(\theta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \underline{t}^T \underline{\dot{l}}_\theta(X_i) - \frac{1}{2}\underline{t}^T I(\theta_0)\underline{t} + o_p(1) \\ &\to_d N(0, t^T I(\theta_0)t) - \frac{1}{2}t^T I(\theta_0)t \\ &\stackrel{d}{=} N(-\sigma^2/2, \sigma^2) \end{aligned}$$

under P_{θ_0} where $\sigma^2 \equiv \underline{t}^T I(\theta_0)\underline{t}$. (The convergence in the last display actually holds under the considerably weaker hypothesis of Hellinger differentiability of p_θ at θ_0, as stated in Corollary 3 of section 3.3, page 28, of the Chapter 3 notes.)

Proof. 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

$$\left(-\frac{1}{n}\ddot{\mathbb{I}}_n(\theta_n^*)\right) \rightarrow_p I(\theta_0)$$

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

2. (a) Exercise 2.1.6, page 10, chapter 2 notes; i.e. show that $d_{TV}(P,Q) = 1 - \int p \wedge q \, d\mu$.
(b) Exercise 2.1.7, page 10, chapter 2 notes; i.e. show that

$$H^2(P,Q) \le d_{TV}(P,Q) \le H(P,Q)\{1+\rho(P,Q)\}^{1/2} \le \sqrt{2}H(P,Q).$$

Solution: (a) From the proof of proposition 1.13, chapter 2 notes, page 9, we see that

$$\begin{aligned}
d_{TV}(P,Q) &= \frac{1}{2}\int |p-q|d\mu = \int_{[p\ge q]} (p-q)\ d\mu = \int_{[p\ge q]} pd\mu - \int_{[p\ge q]} p\wedge q\ d\mu \\
&= \int_{[p\ge q]} pd\mu + \int_{[p<q]} pd\mu - \int_{[p\ge q]} p\wedge q\ d\mu - \int_{[p<q]} p\ d\mu \\
&= \int pd\mu - \int_{[p\ge q]} p\wedge q\ d\mu - \int_{[p<q]} p\wedge q\ d\mu \\
&= 1 - \int p\wedge q\ d\mu \equiv 1 - \eta(P,Q).
\end{aligned}$$

Alternatively, use the identity $|a-b| = a+b-2(a\wedge b)$ for all $a,b\in\mathbb{R}$ to deduce that

$$|p(x)-q(x)| = p(x)+q(x)-2p(x)\wedge q(x)$$

for each fixed x, and hence

$$\begin{aligned}
d_{TV}(P,Q) &= \frac{1}{2}\int|p-q|d\mu = \frac{1}{2}\left(\int pd\mu + \int qd\mu - 2\int p\wedge qd\mu\right) \\
&= 1 - \int p\wedge qd\mu \equiv 1-\eta(P,Q).
\end{aligned}$$

(b) To see the first inequality, note that $H^2(P,Q) = 1-\rho(P,Q)$ where

$$\rho(P,Q) = \int \sqrt{pq}\ d\mu \ge \int p\wedge q\ d\mu \equiv \eta(P,Q)$$

since $\sqrt{p(x)q(x)} \ge p(x)\wedge q(x)$ for all x. Thus we have

$$H^2(P,Q) \;=\; 1-\rho(P,Q) \le 1-\eta(P,Q) = d_TV(P,Q).$$

For the second inequality, write $|p-q| = |(\sqrt{p}-\sqrt{q})(\sqrt{p}+\sqrt{q})|$ and then apply the Cauchy-Schwarz inequality: thus

$$2d_{TV}(P,Q) \;=\; \int|p-q|d\mu = \int|(\sqrt{p}-\sqrt{q})(\sqrt{p}+\sqrt{q})|d\mu$$

$$
\begin{aligned}
&\leq \left(\int |\sqrt{p}-\sqrt{q}|^2 d\mu\right)^{1/2} \left(\int |\sqrt{p}+\sqrt{q}|^2 d\mu\right)^{1/2} \\
&= \sqrt{2}H(P,Q)\left\{\int (p+2\sqrt{pq}+q)d\mu\right\}^{1/2} \\
&= \sqrt{2}H(P,Q)\left\{2+2\rho(P,Q)\right\}^{1/2} \\
&= 2H(P,Q)\left\{1+\rho(P,Q)\right\}^{1/2},
\end{aligned}
$$

and this yields the claimed inequality. The third inequality is easy since $\rho(P,Q) \leq 1$ by Cauchy-Schwarz again.

3. 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:

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

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 (don't confuse this with the γ above!). For the data in 4(b) below, $\overline{T} = -3.00593$ and $S_T = 2.04635$, and hence the method of moment estimators of (α, β) based on (8) are given by

$$\overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}} \frac{1}{S_T} = 0.626751,$$

$$\overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 50.749$$

for the given data.
(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

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

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

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

4. (Problem #3, 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 $n = 12$):

$$1,\ 1,\ 2,\ 3,\ 12,\ 25,\ 46,\ 56,\ 68,\ 109,\ 323,\ 417.$$

[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.]
(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

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

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

$$\check{\theta}_n \equiv \overline{\theta}_n + \hat{I}_n^{-1}(\overline{\theta}_n)(\frac{1}{n}\dot{l}(\overline{\theta}_n)) = (55.181\ldots, 0.551\ldots).$$

The one - step estimator using $\hat{I}_n(\overline{\theta}_n) = (-n^{-1}\ddot{l}_n(\overline{\theta}_n)$ gives the (not sensible!) result

$$\check{\theta}_n = (1.478\ldots, -0.0671\ldots),$$

(c) The maximum likelihood estimate $\hat{\theta}_n = (54.922..., 0.5648...)$; 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

Here is the data: Out[301]=1,1,2,3,12,25,46,56,68,109,323,417

Out[305]=0.,0.,-0.693147,-1.09861,-2.48491,-3.21888,-3.82864,-4.02535, -4.21951,-4.69135,-5.77765,-6.03309

Mean of T = - Log[x] Out[308]=-3.00593

Standard deviation of T Out[310]=2.13734

Out[311]=13.2231 Out[312]=4.18754

Biased estimator of std. dev Out[314]=2.04635

Moment estimator of beta, version 1: Out[321]=0.600068

Moment estimator of beta, version 2: Out[323]=0.626751

Moment estimator of alpha, version 1: Out[325]=52.8704

Moment estimator of alpha, version 2: Out[327]=50.749

theta bar estimator, version 1 Out[329]=52.8704,0.600068

theta bar estimator, version 2 50.749,0.626751

Information matrix estimator based on thetabar 0.000128818,-0.00799661,-0.00799661,5.06463

inverse information matrix estimator based on thetabar 8606.47,13.5889,13.5889,0.218903

vector of scores evaluated at thetabar 0.000595617,-0.215567

sample size n (NSS in the program) Out[356]=12

adjustment to the preliminary estimator Out[359]= 2.19684,-0.0390947

resulting one step estimator; based on theoretical Inform matrix Out[362]=55.0673,0.560973

information matrix based on - Hessian of log-likelihood Out[366]=0.000146843,-0.0114358,-0.0114358,5.48394

inverse information matrix from Hessian Out[369]=8130.36,16.9545,16.9545,0.217707

adjustment to the preliminary estimator Out[372]=1.18774,-0.036832

resulting second version of one-step estimator Out[375]= 54.0582,0.563236

Mathematica input for maximum likelihood estimators:

```
Clear[a,b,ahat,bhat]
(* 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]
(* 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:

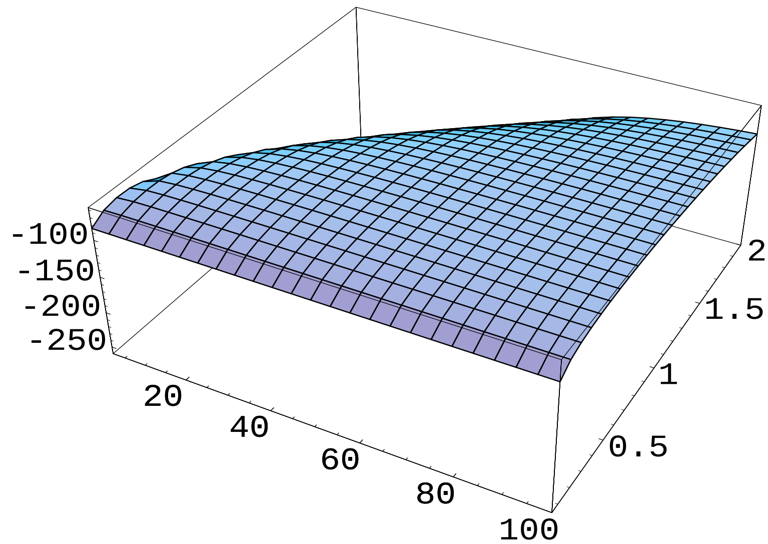

Figure 1: Weibull likelihood.

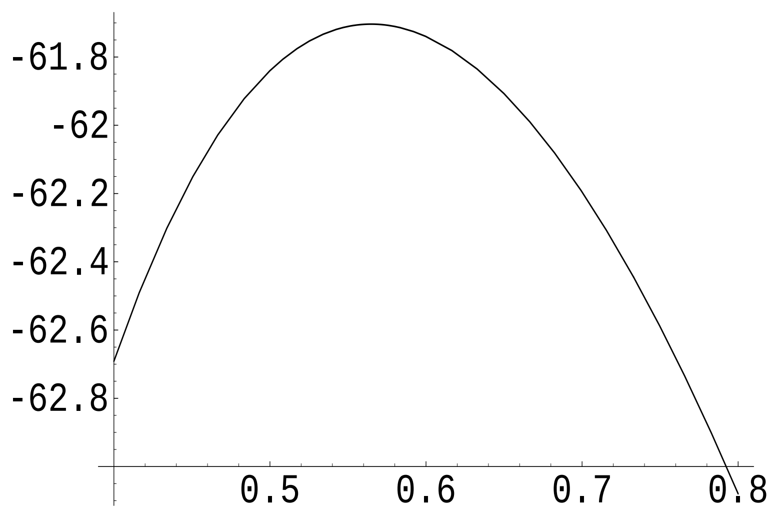

Figure 2: Weibull profile likelihood.

```
{61.7036,{b \mapsto 0.564847}}
{b\mapsto 0.564847}
0.564847
54.9218
```

5. (a) Ferguson, ACILST, problem 17.2, page 117.

   (i) Show that for $X_{(k)} \leq \theta \leq X_{(k+1)}$ the likelihood function is decreasing if

$\theta < k/n$ and increasing if $\theta > k/n$.

(ii) Conclude that the maximum likelihood estimate is equal to one of the order statistics $X_{(k)}$ for which $(k-1)/n \le X_{(k)} \le k/n$.

(b) Do our hypotheses A0-A2 hold in this example?

(c) Compute $K(P_{\theta_0}, P_\theta)$ where P_θ has density as given in this problem.

(d) Do our hypotheses A3 and A4 hold in this example? Why or why not?

(e) Does there exist an estimator $\overline{\theta}_n$ of θ which is $n^{1/2}-$consistent?

Solution: (a) (i) Since the density function p_θ is given by

$$p_\theta(x) = 2\left\{\frac{x}{\theta}1_{[0,\theta]}(x) + \frac{1-x}{1-\theta}1_{(\theta,1]}(x)\right\}, \tag{0.6}$$

it follows that for $X_{(k)} < \theta < X_{(k+1)}$ the likelihood is given by

$$L_n(\theta) = 2^k\theta^{-k}\prod_{i\le k} X_{(i)}(1-\theta)^{-(n-k)}\prod_{i>k}(1-X_{(i)}).$$

(This corrects the expression on page 215 of Ferguson in several respects: it changes Ferguson's $+$ to $\cdot$, and it changes the second product from $\prod X_{(i)}$ to $\prod(1-X_{(i)})$.) Thus for $X_{(k)} < \theta < X_{(k+1)}$ we compute

$$\dot{\mathrm{l}}_n(\theta) = -\frac{k}{\theta} + \frac{n-k}{1-\theta},$$

which is < 0 if $\theta < k/n$ and > 0 if $\theta > k/n$.

(a) (ii) Similarly, for $X_{(k)} < \theta < X_{(k+1)}$,

$$\ddot{\mathrm{l}}_n(\theta) = \frac{k}{\theta^2} + \frac{n-k}{(1-\theta)^2} > 0,$$

so the zeroes of the likelihood equation correspond to *local minima* of the (log-)likelihood, and any local maxima of the log-likelihood occur at the order statistics $X_{(k)}$. It is easily seen that a local maximum occurring at an observation $X_{(k)}$ must correspond to a cusp in the (log-)likelihood: i.e. a point at which $\dot{\mathrm{l}}_n(\theta)$ is positive to the left of $X_{(k)}$ and negative to the right of $X_{(k)}$. Therefore if $\theta = X_{(k)}$ yields a local maximum we have

$$\lim_{\theta\nearrow X_{(k)}}\left\{-\frac{(k-1)}{\theta} + \frac{n-k+1}{1-\theta}\right\} = -\frac{k-1}{X_{(k)}} + \frac{n-k+1}{1-X_{(k)}} > 0,$$

and

$$\lim_{\theta\searrow X_{(k)}}\left\{-\frac{k}{\theta} + \frac{n-k}{1-\theta}\right\} = -\frac{k}{X_{(k)}} + \frac{n-k}{1-X_{(k)}} < 0.$$

But these two inequalities imply that

$$\frac{k-1}{n} < X_{(k)} < \frac{k}{n} \qquad \text{or} \qquad \frac{k-1}{n} < \mathbb{F}_n^{-1}(k/n) < \frac{k}{n}.$$

(b) A0 - A2 all hold in this example: If $\theta \ne \theta^*$, then $p_\theta \ne p_{\theta^*}$ and hence $P_\theta \ne P_{\theta^*}$. The set $A = \{x : \ p_\theta(x) > 0\} = (0,1)$ for all θ, and hence does not depend on θ;

thus A1 holds. A2 holds with μ given by Lebesgue measure on $[0,1]$.
(c) Suppose that $\theta_0 < \theta$. Then the Kullback-Leibler information $K(P_{\theta_0}, P_\theta)$ is given by

$$\begin{aligned}
K(P_{\theta_0}, P_\theta) &= \int_0^{\theta_0} p_{\theta_0}(x)\log(\theta/\theta_0)dx + \int_{\theta_0}^{\theta} p_{\theta_0}(x)\log\left(\frac{1-x}{1-\theta_0}\frac{\theta}{x}\right)dx \\
&\quad + \int_\theta^1 p_{\theta_0}(x)\log\frac{1-\theta}{1-\theta_0}dx \\
&= \theta_0\log(\theta/\theta_0) + \frac{(1-\theta)^2}{1-\theta_0}\log\frac{1-\theta}{1-\theta_0} \\
&\quad + \frac{1}{1-\theta_0}\{(1-\theta_0)^2-(1-\theta)^2\}\log\left(\frac{\theta}{1-\theta_0}\right) \\
&\quad + \frac{2}{1-\theta_0}\int_{\theta_0}^{\theta}(1-x)\log\left(\frac{1-x}{x}\right)dx.
\end{aligned}$$

Similarly, if $\theta_0 > \theta$, then

$$\begin{aligned}
K(P_{\theta_0}, P_\theta) &= \int_0^{\theta} p_{\theta_0}(x)\log(\theta/\theta_0)dx + \int_{\theta}^{\theta_0} p_{\theta_0}(x)\log\left(\frac{x}{\theta_0}\frac{1-\theta}{1-x}\right)dx \\
&\quad + \int_{\theta_0}^1 p_{\theta_0}(x)\log\frac{1-\theta}{1-\theta_0}dx \\
&= \frac{\theta^2}{\theta_0}\log(\theta/\theta_0) + (1-\theta_0)\log\frac{1-\theta}{1-\theta_0} \\
&\quad + \frac{1}{\theta_0}\{\theta_0^2-\theta^2\}\log\left(\frac{1-\theta}{\theta_0}\right) + \frac{2}{\theta_0}\int_\theta^{\theta_0} x\log\left(\frac{x}{1-x}\right)dx.
\end{aligned}$$

Here is a plot of $\theta \mapsto K(P_{\theta_0}, P_\theta)$ for $\theta_0 = .2$.

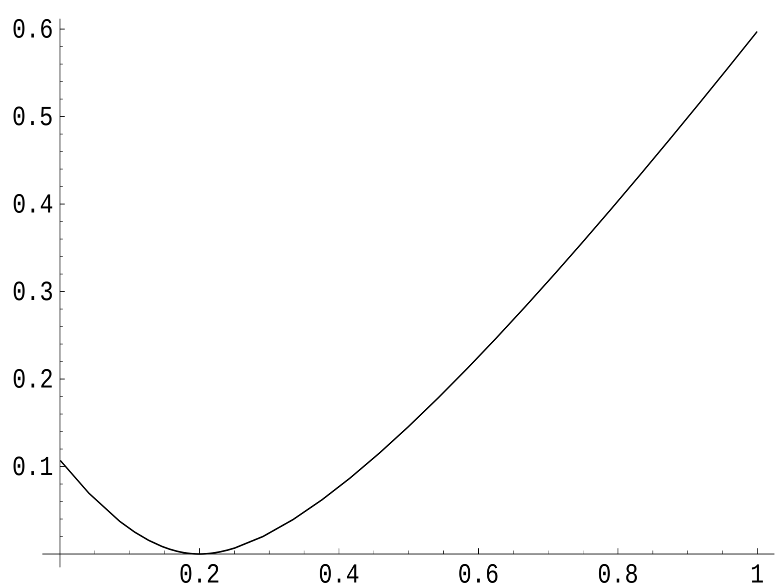

Figure 3: Kullback - Leibler function $K(P_{\theta_0}, P_\theta)$, $\theta_0 = .2$

(d) Since p_θ is given by (0.6),

$$\log p_\theta(x) = \begin{cases} \log 2 + \log x - \log\theta, & \text{if } \ x \le \theta, \\ \log 2 + \log(1-x) - \log(1-\theta), & \text{if } \ x > \theta, \end{cases}$$

so

$$\dot{\mathbf{l}}_\theta(x) = -\frac{1}{\theta} 1_{[x<\theta]} + \frac{1}{1-\theta} 1_{[x>1-\theta]},$$

but the derivative does not exist at $\theta = x$ (since the left and right derivatives are different). Similarly

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

but the second derivative does not exist at $\theta = x$. Note that $\dot{\mathbf{l}}_\theta$ is a discontinuous function of θ for every $0 < x < 1$. Although

$$E_\theta \dot{\mathbf{l}}_\theta(X) = -\frac{1}{\theta}\theta + \frac{1}{1-\theta}(1-\theta) = 0,$$

and

$$E_\theta \dot{\mathbf{l}}_\theta^2(X) = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)},$$

we also have

$$-E_\theta \ddot{\mathbf{l}}_{\theta\theta}(X) = -\frac{1}{\theta} - \frac{1}{1-\theta} = -\frac{1}{\theta(1-\theta)} \ne E_\theta \dot{\mathbf{l}}_\theta^2(X).$$

Thus A3 and A4(iii) fail, while A4(i) and A4(ii) hold.

(e) First a $\sqrt{n}-$consistent estimator of θ via moments: note that

$$\begin{aligned}
E_\theta X &= 2\int_0^\theta \frac{x^2}{\theta} dx + 2\int_\theta^1 \frac{x(1-x)}{1-\theta} dx \\
&= \frac{2}{3}\theta^2 + \frac{2}{1-\theta}(\frac{1}{2}x^2 - \frac{1}{3}x^3|_\theta^1 \\
&= \frac{2}{3}\theta^2 + \frac{2}{1-\theta}\left\{\frac{1}{6} - \frac{1}{2}\theta^2 + \frac{1}{3}\theta^3\right\} \\
&= \frac{2}{1-\theta}\left\{\frac{1}{3}\theta^2(1-\theta) + \frac{1}{6} - \frac{1}{2}\theta^2 + \frac{1}{3}\theta^3\right\} \\
&= \frac{2}{1-\theta}\left\{\frac{1}{6} - \frac{1}{6}\theta^2\right\} = \frac{1}{3}(1+\theta).
\end{aligned}$$

Since $\overline{X}_n \to_p E_\theta X = (1+\theta)/3$, it follows by continuous mapping that $3\overline{X}_n - 1 \to_p \theta$. Thus with $g(x) \equiv 3x - 1$ we have

$$\sqrt{n}(g(\overline{X}_n) - \theta) \to_d g'(\theta)\sigma(\theta) Z$$

where $g'(x) = 3$, $\sigma^2(\theta) = Var_\theta(X) = (1 - \theta + \theta^2)/18$, and $Z \sim N(0,1)$. Thus it follows that

$$\sqrt{n}(3\overline{X}_n - 1 - \theta) \to_d N(0, (1-\theta+\theta^2)/2).$$

Thus the estimator $\overline{\theta}_n \equiv 3\overline{X}_n - 1$ is a $\sqrt{n}-$consistent estimator of θ.

Now for an estimator of θ based on the median. The distribution function F_θ corresponding to p_θ is

$$F_\theta(x) = \frac{x^2}{\theta} 1_{[0,\theta]}(x) + \left(1 - \frac{(1-x)^2}{1-\theta}\right) 1_{(\theta,1]}(x),$$

and the corresponding quantile function is

$$F_\theta^{-1}(u) = \sqrt{\theta u} 1_{[u<\theta]} + (1 - \sqrt{(1-\theta)(1-u)}) 1_{[u \ge \theta]}.$$

Thus the median is

$$F_\theta^{-1}(1/2) = \sqrt{\theta/2} 1_{[1/2<\theta]} + (1 - \sqrt{(1-\theta)/2}) 1_{[1/2>\theta]} \equiv g(\theta),$$

which has inverse function

$$g^{-1}(x) = 2x^2 1_{[x\ge 1/2]} + (1 - 2(1-x)^2) 1_{[x<1/2]} \equiv h(x)$$

Note that $g^{-1}(1/2+) = g^{-1}(1/2-) = 1/2$, so g^{-1} is continuous at $1/2$, and

$$\frac{d}{dx} g^{-1}(x) = \frac{d}{dx} h(x) = 4x 1_{[x\ge 1/2]} + 4(1-x) 1_{[x<1/2]},$$

so the derivative of g^{-1} is also continuous at $x = 1/2$. It follows that $g^{-1}(\mathbb{F}_n^{-1}(1/2)) = h(\mathbb{F}_n^{-1})$ is a consistent and asymptotically normal estimator of θ:

$$g^{-1}(\mathbb{F}_n^{-1}(1/2)) \to_{a.s.} g^{-1}(F_\theta^{-1}(1/2)) = g^{-1}(g(\theta)) = \theta,$$

and

$$\begin{aligned} \sqrt{n}(g^{-1}(\mathbb{F}_n^{-1}(1/2)) - g^{-1}(F_\theta^{-1}(1/2))) &= \sqrt{n}(h(\mathbb{F}_n^{-1}(1/2)) - h(F_\theta^{-1}(1/2))) \\ &\to_d h'(F_\theta^{-1})\{-Q'(1/2)\mathbb{U}(1/2)\} \sim N(0, \sigma^2(\theta)) \end{aligned}$$

where

$$\sigma^2(\theta) = \{h'(F_\theta^{-1}(1/2))\}^2 \cdot Q'(1/2)^2 \cdot (1/4).$$

There are many other $\sqrt{n}-$consistent estimators of θ in this example and, in fact, the MLE is consistent, $\sqrt{n}-$consistent, and asymptotically efficient. We will return to this example in Stat 582.