Statistics 581, Problem Set 9 Solutions

Wellner; 11/26/2014

1. Consider the Weibull family of example 3.2.5 and problem set #6, problem 1: $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ 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. (This amounts to a rephrasing of the statement of problem 1 in problem set 6.)
(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

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

[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 $\overline{T} = -2.9822...$ and $\tilde{S}_T = 2.12897...$ (biased variance estimator) or $S_T = 2.0383...$ (unbiased variance estimator), moment estimators of (α, β) based on (8) are given by

$$\overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}}\frac{1}{\tilde{S}_T} = .60243.., \qquad \overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}}\frac{1}{S_T} = .62921...$$

and for these two estimators of β,

$$\overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 51.439, \qquad \overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 49.383...$$

respectively for the given data in problem 2 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

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

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 $g \equiv (g_1, g_2) : R^2 \to R^2$ 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

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

where

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

2. (Problem #1, 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 $n = 12$):

$$1, \ 1, \ 2, \ 3, \ 12, \ 21, \ 46, \ 54, \ 65, \ 109, \ 317, \ 413.$$

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

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

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 1(b) above. The one step estimator using $\hat I(\overline\theta_n) = I(\overline\theta_n)$ is

$$\check\theta_n \equiv \overline\theta_n + \hat I_n^{-1}(\overline\theta_n)(\frac{1}{n}\dot l(\overline\theta_n)) = (53.7305\ldots,\ 0.55833\ldots).$$

The one - step estimator using $\hat I_n(\overline\theta_n) = (-n^{-1}\ddot l_n(\overline\theta_n)$ gives the result

$$\check\theta_n = (52.5638\ldots, 0.561407\ldots),$$

(c) The maximum likelihood estimate is $\hat\theta_n$ = (53.5665..., 0.563395...) (by profile methods), 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:

```
Print["Here is the data:"]
x = {1, 1, 2, 3, 12, 21, 46, 54, 65, 109, 317, 413}
(* NSS is the sample size *)
NSS = Length[x]

(* First transform to -Log[x]:  *)
Print["Here is the transformed \
data, -Log[x]"]
t = N[-Log[x]]
(* Now compute Mean and Variance of y *)
Print["Mean of T = -Log(x"]
tbar = Mean[t]
Print["Standard deviation of T"]
Stt = Sqrt[Variance[t]]
tsquaredbar = Sum[t[[i]]^2, {i, 1, NSS}]/NSS
Stt1 = tsquaredbar - tbar^2
```

```
Print["Biased estimator of std dev"]
Stt2 = Sqrt[Stt1]

(*compute mean and variance of standard Gumbel*)
VarGumbel := (Pi^2)/
  6
MeanGumbel := EulerGamma

(*The the Method of Moment Estimators of beta and alpha are:*)
\
Print["Moment estimator of beta, version 1:"]
betabar = N[Sqrt[VarGumbel/Stt^2]]
Print["Moment estimator of beta, version 1:"]
betabar2 = N[Sqrt[VarGumbel/Stt2^2]]
Print["Moment estimator of alpha, version 1:"]
alphabar = N[Exp[-tbar + MeanGumbel/betabar]]
Print["Moment estimator of alpha, version 2:"]
alphabar2 = N[Exp[-tbar + MeanGumbel/betabar2]]
Print["theta bar estimator, version 1"]
thetabar = {alphabar, betabar}
Print["theta bar estimator, version 2"]
thetabar = {alphabar2, betabar2}

(*f is the Weibull density function:*)
f[t_, a_, b_] := (b/a)*(t/a)^(b - 1)*Exp[-(t/a)^b];
(*L is the log-likelihood*)

L[a_, b_] := Sum[Log[f[x[[i]], a, b]], {i, 1, NSS}];



(*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 (theoretical) 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]]

(*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
Print["Information matrix based on thetabar"]
Inf[alphabar, betabar]
Print["Inverse information matrix estimator based on thetabar"]
Infminus1[alphabar, betabar]
Print["vector of scores evaluated at thetabar"]
Sc[alphabar, betabar]
Print["sample size n (NSS in the program)"]
NSS
Delta1 := Infminus1[alphabar, betabar].Sc[alphabar, betabar]

Print["adjustment to the preliminary estimator"]
Delta1
thetaCaret1 := {alphabar, betabar} + {Delta1[[1]], Delta1[[2]]}
Print["resulting one step estimator; based on theoretical inform \
matrix"]
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]
Print["information matrix based on -Hessian of log-likelihood"]
Inf2[alphabar, betabar]
Print["inverse information matrix from Hessian"]
Infminus2[a_, b_] := Inverse[Inf2[a, b]]
Infminus2[alphabar, betabar]
Print["adjustment to the preliminary estimator"]
Delta2 := Infminus2[alphabar, betabar].Sc[alphabar, betabar]
Delta2
Print["resulting hessian based version of one-step estimator"]
thetaCaret2 := {alphabar, betabar} + {Delta2[[1]], Delta2[[2]]}
thetaCaret2
```

Mathematica output for one-step estimators:

```
During evaluation of In[958]:= Here is the data:

Out[959]= {1, 1, 2, 3, 12, 21, 46, 54, 65, 109, 317, 413}

Out[960]= 12

During evaluation of In[958]:= Here is the transformed data, -Log[x]

Out[962]= {0., 0., -0.693147, -1.09861, -2.48491, -3.04452, -3.82864, \
-3.98898, -4.17439, -4.69135, -5.7589, -6.02345}

During evaluation of In[958]:= Mean of T = -Log(x

Out[964]= -2.98224

During evaluation of In[958]:= Standard deviation of T

Out[966]= 2.12897

Out[967]= 13.0486

Out[968]= 4.15481

During evaluation of In[958]:= Biased estimator of std dev

Out[970]= 2.03834

During evaluation of In[958]:= Moment estimator of beta, version 1:

Out[974]= 0.602427

During evaluation of In[958]:= Moment estimator of beta, version 1:

Out[976]= 0.629214

During evaluation of In[958]:= Moment estimator of alpha, version 1:

Out[978]= 51.4388

During evaluation of In[958]:= Moment estimator of alpha, version 2:

Out[980]= 49.3827

During evaluation of In[958]:= theta bar estimator, version 1
```

```
Out[982]= {51.4388, 0.602427}

During evaluation of In[958]:= theta bar estimator, version 2

Out[984]= {49.3827, 0.629214}

During evaluation of In[958]:= Information matrix based on thetabar

Out[994]= {{0.00013716, -0.00821918}, {-0.00821918, 5.02505}}

During evaluation of In[958]:= Inverse information matrix estimator based on theta

Out[996]= {{8083., 13.2209}, {13.2209, 0.220628}}

During evaluation of In[958]:= vector of scores evaluated at thetabar

Out[998]= {0.000676756, -0.240412}

During evaluation of In[958]:= sample size n (NSS in the program)

Out[1000]= 12

During evaluation of In[958]:= adjustment to the preliminary estimator

Out[1003]= {2.29176, -0.0440942}

During evaluation of In[958]:= resulting one step estimator; based on theoretical

Out[1006]= {53.7305, 0.558333}

During evaluation of In[958]:= information matrix based on -Hessian of log-likelih

Out[1010]= {{0.000158242, -0.0121582}, {-0.0121582, 5.52737}}

During evaluation of In[958]:= inverse information matrix from Hessian

Out[1013]= {{7604.62, 16.7273}, {16.7273, 0.217712}}

During evaluation of In[958]:= adjustment to the preliminary estimator

Out[1016]= {1.12503, -0.0410201}

During evaluation of In[958]:= resulting hessian based version of one-step estimat

Out[1019]= {52.5638, 0.561407}
```

Mathematica input for maximum likelihood estimators:

```
Clear[a,b,ahat,bhat]
(* 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}
  *)
(*  2008 data
x={1,1,2,3,12,25,46,56,79,125,323,417 }
*)
(*  2009 data
x={1,1.3,1.7,3.2,10.7,24.3,51.2,77.1,93.7,105,111,305.}
*)
(*
x = {1, 1, 2, 3, 12, 25, 46, 54, 68, 109, 319, 413 }
*)
x = {1, 1, 2, 3, 12, 21, 46, 54, 65, 109, 317, 413 }


(* 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:

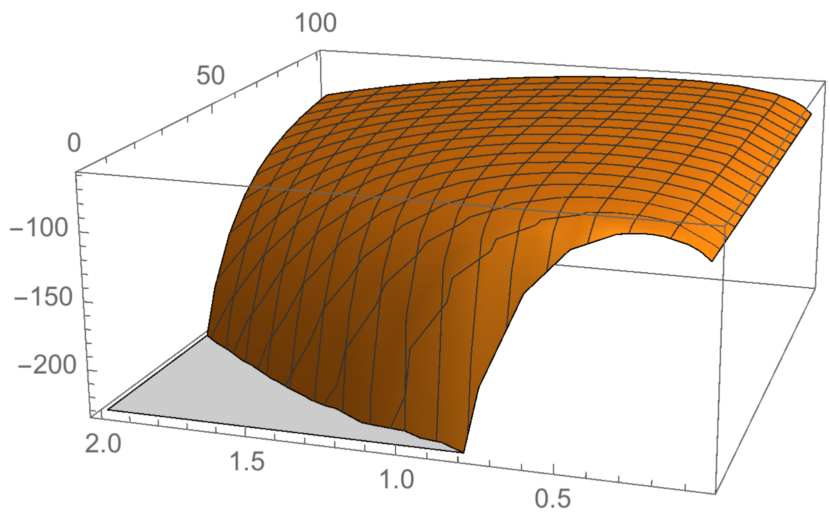

Figure 1: Weibull likelihood.

```
{b -> 0.563395}
0.563395
53.5665
{61.62, {a -> 47., b -> 0.62}}
61.62
{a -> 47., b -> 0.62}
```

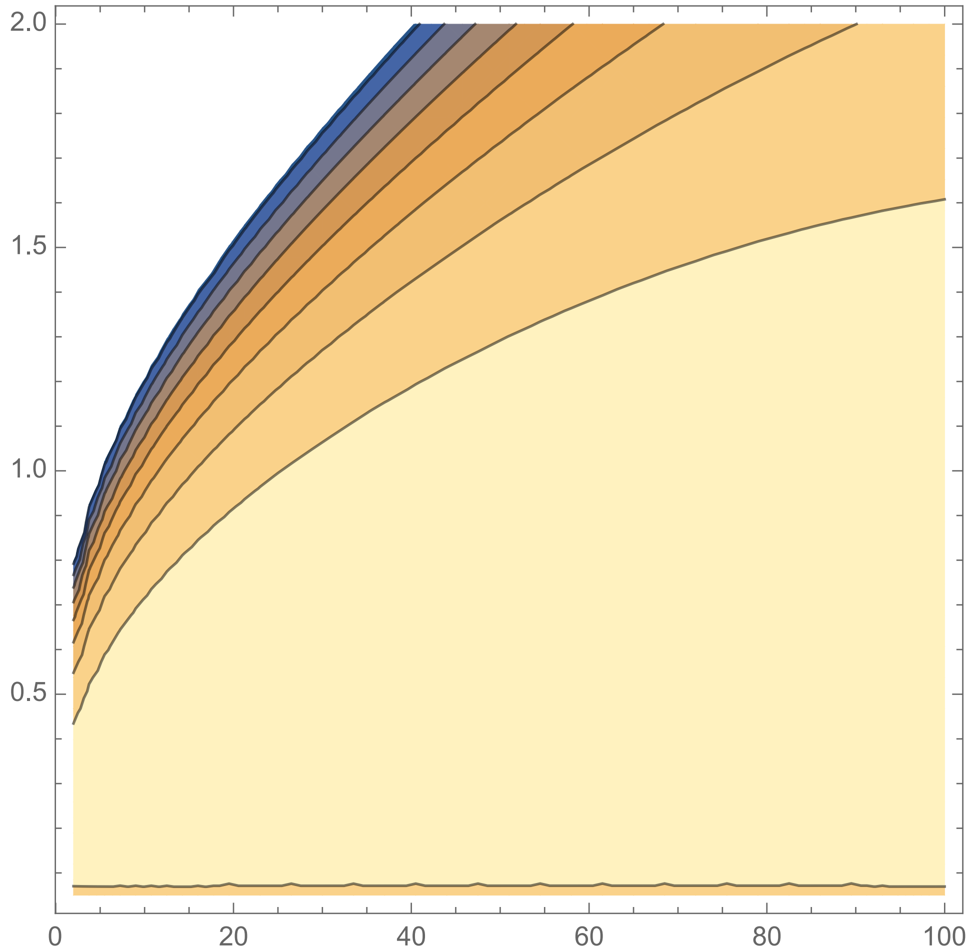

Figure 2: Contour plot Weibull likelihood

3. (a) Ferguson, ACILST, problem 17.2, page 117: I would suggest modifying Ferguson's definition of the density to:

$$p_\theta(x) \equiv f(x|\theta) = 2\left\{\frac{x}{\theta}1_{[0,\theta]}(x) + \frac{1-x}{1-\theta}1_{(\theta,1]}(x)\right\}.$$

(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) Let $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ denote the order statistics of the sample $X_1, \ldots, X_n$. Now

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

so the likelihood function is

$$L_n(\theta|\underline{X}) = \prod_{i=1}^n 2\left\{\frac{X_i}{\theta}1_{[0,\theta]}(X_i) + \frac{1-X_i}{1-\theta}1_{(\theta,1]}(X_i)\right\}$$

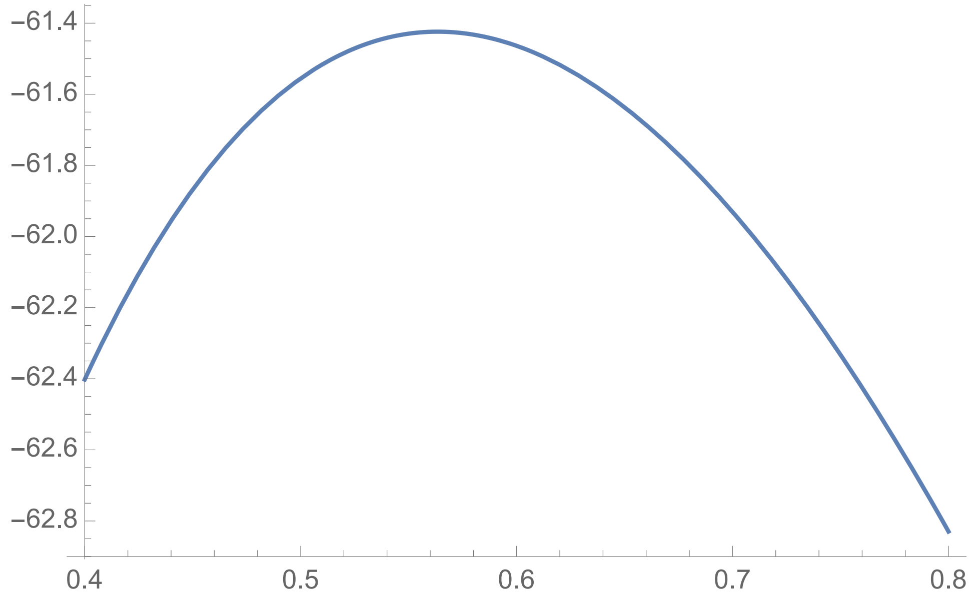

Figure 3: Profile plot Weibull likelihood

$$
\begin{aligned}
&= \prod_{i=1}^{n} 2\left\{\frac{X_{(i)}}{\theta}1_{[0,\theta]}(X_{(i)}) + \frac{1-X_{(i)}}{1-\theta}1_{(\theta,1]}(X_{(i)})\right\} \\
&= \left(\frac{2}{\theta}\right)^k \prod_{j=1}^{k} X_{(j)} \cdot \left(\frac{2}{1-\theta}\right)^{n-k} \prod_{j=k+1}^{n} (1-X_{(j)}) \ \text{ if } \ X_{(k)} \le \theta < X_{(k+1)}.
\end{aligned}
$$

Thus

$$
l_n(\theta|\underline{X}) = \log L_n(\theta|\underline{X}) = -k\log\theta - (n-k)\log(1-\theta) + \text{const. in } \theta,
$$

for $X_{(k)} < \theta < X_{(k+1)}$, and on this interval

$$
\dot{l}_{n,\theta}(\theta|\underline{X}) = -\frac{k}{\theta} + \frac{n-k}{1-\theta} = \frac{n\theta - k}{\theta(1-\theta)} \begin{cases} > 0, & \text{if } \theta > k/n, \\ < 0, & \text{if } \theta < k/n, \end{cases}
$$

so l_n and L_n are decreasing on $X_{(k)} \le \theta < X_{(k+1)} \wedge k/n$ and increasing on $X_{(k)} \vee k/n \le \theta < X_{(k+1)}$.

(b) First note that $L_n(\theta)$ is continuous. Moreover, by the computation in (a), it can only have local *minima* at the solutions of the likelihood equations (which can only occur at the points k/n, $k = 1, \ldots, n$), and hence the local maxima of the likelihood occur only at the order statistics. Furthermore, if $(k-1)/n < X_{(k)} < k/n$, then the log-likelihood $l_n(\theta)$ and the likelihood function $L_n(\theta)$ has a local maximum at $X_{(k)}$: if $k/n > \theta > X_{(k)}$ then $\dot{l}_{n,\theta}(\theta|\underline{X}) < 0$ from (a), while if $(k-1)/n < \theta < X_{(k)}$, then $\dot{l}_{n,\theta}(\theta|\underline{X}) > 0$ also by (a).

Here is a plot of the two examples of the likelihood function for samples of size $n = 4, 10$, and $n = 50$.

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

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

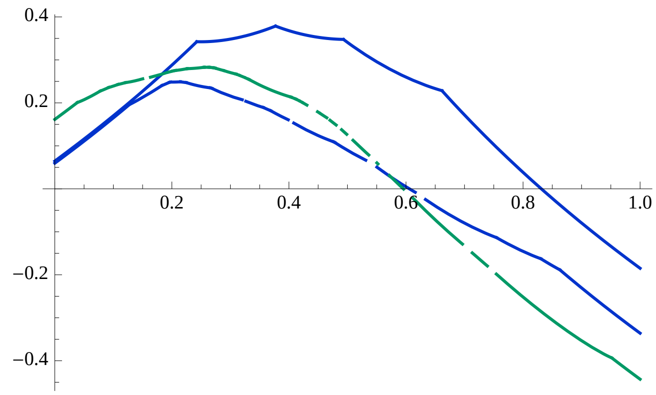

Figure 4: Three log-likelihood functions for problem 3, $n = 4$, $n = 10$, and $n = 50$

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 roots (or zeros) 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 $\theta_0 = .2$.

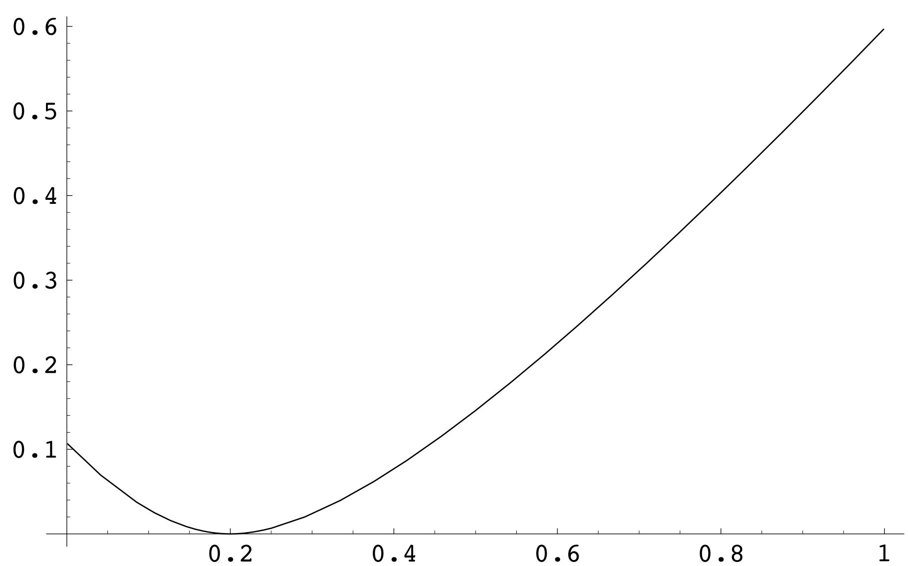

Figure 5: Kullback - Leibler function $K(P_{\theta_0}, P_\theta)$, $\theta_0 = .2$

(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

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

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.

4. (a) Lehmann and Casella, problem 6.3.1, page 501.
   (b) Lehmann and Casella, problem 6.3.2, page 501.
   (c) Lehmann and Casella, problem 6.3.4, page 501.

Solution: (a)(i) Since $\log P_p(X = x) = x \log p + (n-x)\log(1-p)$, we have $l(p|X) = X \log p + (n-X)\log(1-p)$; differentiating this with respect to p yields

$$\dot{l}(p|X) = \frac{X}{p} - \frac{n-X}{1-p} = \frac{X(1-p) - (n-X)p}{p(1-p)}$$

and this equals 0 if $p = \hat{p} \equiv X/n$. Since the second derivative is

$$\ddot{l}(p|X) = -\frac{X}{p^2} - \frac{n-X}{(1-p)^2} < 0$$

it follows that $\hat{p} = X/n$ is the MLE of $p \in [0,1]$.
(a)(ii) Since $(\prod_{i=1}^n y_i)^{1/n} \le n^{-1}(y_1 + \cdots + y_n)$ for any numbers $y_i \ge 0$, it follows, with $y_i \equiv np/X$ for $i = 1, \ldots, X$, and $y_i \equiv nq/(n-X)$, $i = X+1, \ldots, n$, that

$$\left\{ \left(\frac{np}{X}\right)^X \left(\frac{nq}{n-X}\right)^{n-X} \right\}^{1/n} \le n^{-1} \left\{ X \frac{np}{X} + (n-X)\frac{nq}{n-X} \right\} = 1\,,$$

or, equivalently,

$$p^X(1-p)^{n-X} \le \left(\frac{X}{n}\right)^X \left(\frac{n-X}{n}\right)^{n-X},$$

with equality if and only if $p = X/n \equiv \hat{p}$. Thus $\hat{p} = X/n$ is the MLE of $p \in [0,1]$.
(b) When the closed interval $[0,1]$ is replaced by the open interval $(0,1)$, then the MLE exists if $0 < X < n$ and is $\hat{p} = X/n \in (0,1)$ in this case. If $X = 0$, then the log-likelihood equals $n\log(1-p)$, so $\sup_{p\in(0,1)} l(p) = 0$, but this supremum is not achieved (in the set $(0,1)$). Thus the MLE does not exist in this case. Similarly, if $X = n$, the the log-likelihood equals $n\log p$, so $\sup_{p\in(0,1)} l(p) = 0$, but this supremum is not achieved (in the set $(0,1)$).
(c) When $X_1,\ldots,X_n$ are i.i.d. $N(\xi,1)$ with $\xi > 0$, the log-likelihood is

$$\begin{aligned} l_n(\xi) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n (X_i-\xi)^2 \\ &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n (X_i-\overline{X}_n)^2 - \frac{n}{2}(\overline{X}_n-\xi)^2. \end{aligned}$$

When $\overline{X}_n > 0$, $l_n(\xi)$ achieves its supremum over $(0,\infty)$ at $\xi = \overline{X}_n$; hence the MLE is $\hat{\xi}_n = \overline{X}_n$ in this case. On the other hand when $\overline{X}_n \le 0$, $l_n(\xi)$ fails to achieve its supremum over $(0,\infty)$, and hence the MLE does not exist. (A "reasonable" estimator is $\tilde{\eta}_n \equiv \overline{X}_n 1_{[0,\infty)}(\overline{X}_n)$, but it is not the MLE.)

5. **Optional bonus problem 1:** Suppose that $X_1,\ldots,X_n$ are i.i.d. log-normal(μ,σ^2) with density

$$p_\theta(x) = \frac{1}{x\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)1_{(0,\infty)}(x).$$

Here $\theta = (\mu,\sigma^2) \in \mathbb{R}\times\mathbb{R}^+ \equiv (-\infty,\infty)\times(0,\infty)$.
(a) Find the MLE $\hat{\theta} = (\hat{\mu},\hat{\sigma}^2)$ of $\theta = (\mu,\sigma^2)$.
(b) Show that $\log X \stackrel{d}{=} \mu + \sigma Z \sim N(\mu,\sigma^2)$ where $Z \sim N(0,1)$.
(c) Suppose that $\nu(P_\theta) = q(\theta) = E_\theta(X)$. Express $q(\theta)$ explicitly as a function of θ.
(d) Suggest a natural nonparametric estimator $\overline{\nu}_n$ of $E_\theta(X)$.
(e) Find the asymptotic variance of $\sqrt{n}(\overline{\nu}_n - \nu(P_\theta))$ for the estimator $\overline{\nu}_n$ you proposed in (d).
(f) What is the MLE $\hat{\nu}_n$ of $\nu = \nu(P_\theta)$ assuming that the log-normal model is true? What do our results in chapter 3 say about the asymptotic distribution of $\sqrt{n}(\hat{\nu}_n - \nu(P_\theta))$ (assuming that the model holds)?
(g) Compare the variances you found in (e) and (f). Which estimator do you prefer if the log-normal model holds?

Solution: First part (b): Note that with $Y \equiv \log X$ it follows that the density f_Y of Y is given by

$$\begin{aligned} f_Y(y) &= f_X(e^y)e^y = p_\theta(e^y)e^y \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right); \end{aligned}$$

i.e. $Y = \log X \sim N(\mu, \sigma^2)$.
(a) Since $Y_i \equiv \log X_i$, $i = 1, \ldots, n$ are i.i.d. $N(\mu, \sigma^2)$,

$$\hat{\mu} = n^{-1} \sum_{i=1}^{n} Y_i = n^{-1} \sum_{1}^{n} \log X_i$$

and

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2 = n^{-1} \sum_{i=1}^{n} (\log X_i - \overline{\log(X)})^2$$

are the MLE's of μ and σ^2.
(c) Since $Y = \log X \stackrel{d}{=} \mu + \sigma Z$ where $Z \sim N(0,1)$ and the moment generating function of Z is $Ee^{tZ} = \exp(t^2/2)$, it follows that

$$\begin{aligned} \nu(P_\theta) &= q(\theta) \equiv E_\theta(X) = E_\theta \exp(Y) \\ &= E_\theta \exp(\mu + \sigma Z) = e^{\mu} \cdot \exp(\sigma^2/2). \end{aligned}$$

(d) A natural nonparametric estimator of $q(\theta) = E_\theta(X)$ is $\overline{v}_n = \nu(\mathbb{P}_n) = \int x d\mathbb{P}_n(x) = \overline{X}_n$.
(e) Note that the variance of X is:

$$\begin{aligned} Var_\theta(X) &= E_\theta X^2 - (E_\theta X)^2 = E \exp(2\mu + 2\sigma Z) - e^{2\mu + \sigma^2} \\ &= e^{2\mu} \left\{ e^{2\sigma^2} - e^{\sigma^2} \right\} = e^{2\mu} e^{\sigma^2} \left\{ e^{\sigma^2} - 1 \right\}. \end{aligned}$$

Thus it follows from the central limit theorem that

$$\begin{aligned} \sqrt{n}(\overline{\nu}_n - \nu(P_\theta)) &= \sqrt{n}(\overline{X}_n - E_\theta(X)) \\ &\to N(0, Var_\theta(X)) = N(0, e^{2\mu} e^{\sigma^2} \{e^{\sigma^2} - 1\}). \end{aligned}$$

(f) From our theory in chapter 4 it follows that

$$\sqrt{n}((\hat{\mu}, \hat{\sigma}^2) - (\mu, \sigma^2)) \to_d D \equiv N_2(0, I^{-1}(\theta))$$

where $I^{-1}(\theta) = \mathrm{diag}(\sigma^2, 2\sigma^4)$. Therefore by Corollary 4.1.1 with $\dot{q}(\theta) = (1, 2^{-1}) \exp(\mu + \sigma^2/2)$,

$$\begin{aligned} \sqrt{n}(\hat{\nu} - \nu(P_\theta)) &\to_d \dot{q}(\theta)^T I(\theta)^{-1} Z \\ &\sim N(0, \dot{q}(\theta)^T I(\theta)^{-1} q(\theta)) = N(0, e^{2\mu + \sigma^2}(\sigma^2 + \sigma^4/2)). \end{aligned}$$

(g) The ratio of the asymptotic variances in (e) and (f) is

$$\begin{aligned} \frac{e^{2\mu + \sigma^2}(\sigma^2 + \sigma^4/2)}{e^{\mu + \sigma^2}(e^{\sigma^2} - 1)} &= \frac{\sigma^2(1 + \sigma^2/2)}{e^{\sigma^2} - 1} \\ &< 1 \quad \text{for all} \quad \mu, \sigma^2 \end{aligned}$$

$e^x > 1 + x + x^2/2$ for all $x > 0$. Note that this ratio become very small as σ^2 grows; see Figure yy.

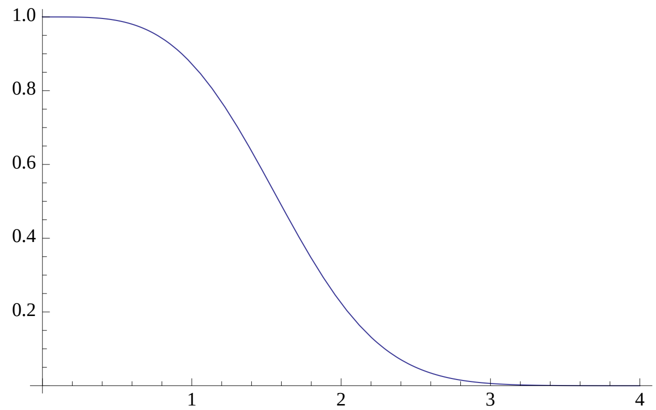

Figure 6: Ratio of asymptotic variance of MLE to variance of $\sqrt{n}(\overline{X}_n - E_\theta(X))$, lognormal mean estimation, as a function of σ.

6. **Optional bonus problem 2:** Lehmann and Casella, TPE, problem 6.3.22, page 503, reworded as follows. (In other words, prove (vi) of theorem 1.2, pages 5-6, 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 463 (the Cramér conditions given in (A) - (D) on pages 462-463). Show that the following Local Asymptotic Normality (LAN) result holds for the (local) log- likelihood ratios: (in the following I have changed notation slightly to agree with the course notes) 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 $\theta_0 \in \Theta$ we have

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

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.)

Solution: This follows easily from the same expansion method we used earlier in the proof of Theorem 1.2:

$$\begin{aligned} l_n(\theta_0 + n^{-1/2}\underline{t}) - l_n(\theta_0) &= \frac{t^T}{\sqrt{n}} \sum_{i=1}^n \dot{\mathbf{l}}_\theta(\theta_0|X_i) + \frac{1}{2}\frac{t^T}{\sqrt{n}} \ddot{\mathbf{l}}_{n,\theta\theta}(\theta_n^*)\frac{t}{\sqrt{n}} \\ &= t^T Z_n - \frac{1}{2} t^T(-n^{-1}\ddot{\mathbf{l}}_{n,\theta\theta}(\theta_n^*) t^T \\ &= t^T Z_n - \frac{1}{2} t^T I(\theta_0) t^T + o_p(1) \end{aligned}$$

where

$$\underline{Z}_n \equiv n^{-1/2} \sum_{i=1}^n \dot{\mathbf{l}}_\theta(\theta_0|X_i).$$

7. **Optional bonus problem 3:** Suppose that $(Y|Z) \sim$ Poisson$(\lambda e^{\gamma Z})$, and $Z \sim$ *Bernoulli*(η), and $\theta = (\lambda, \gamma, \eta)$. Let $X = (Y, Z)$, and suppose that we observe $X_1, \ldots, X_n$ i.i.d. as X.
(a) Find the score equations for estimation of θ.
(b) Give conditions on the data $X_1, \ldots, X_n = (Y_1, Z_1), \ldots, (Y_n, Z_n)$ guaranteeing that the score equations have a unique solution which maximizes the likelihood. Call the resulting estimators $\widehat{\theta}_n = (\widehat{\lambda}_n, \widehat{\gamma}_n, \widehat{\eta}_n)$.
(c) What does theorem 4.1.2 (Chapter 4, page 5), say about the asymptotic distribution of $\sqrt{n}(\widehat{\theta} - \theta_0)$ when the distribution of the data is given by P_{θ_0}?
(d) Suppose that $\theta_1 \neq \theta_0$ is the "true" value of the parameter θ, and we consider the likelihood ratio $L_n(\theta_1)/L_n(\theta_0)$ where $L_n(\theta) \equiv \prod_{i=1}^n p_\theta(X_i)$. Show that $n^{-1}\log(L_n(\theta_1)/L_n(\theta_0) \rightarrow_p$ some constant, and identify the constant explicitly in terms of θ_1, θ_0.

Solution: (a) The density of $X = (Y, Z)$ is

$$p_\theta(y, z) = f_{\lambda,\gamma}(y|z) g_\eta(z) = e^{-\lambda e^{\gamma z}} \frac{(\lambda e^{\gamma z})^y}{y!} \eta^z (1-\eta)^{1-z},$$

for $y \in \{0, 1, \ldots\}$, $z \in \{0, 1\}$, and hence

$$\log p_\theta(y, z) = y \log(\lambda e^{\gamma z}) - \lambda e^{\gamma z} - \log y! + z \log \eta + (1 - z) \log(1 - \eta).$$

From this we calculate the scores $\dot{l}_\lambda$, $\dot{l}_\gamma$, and $\dot{l}_\eta$:

$$\begin{aligned}
\dot{l}_\lambda(y, z) &= \frac{y}{\lambda} - e^{\gamma z} = \frac{1}{\lambda}(y - \lambda e^{\gamma z}), \\
\dot{l}_\gamma(y, z) &= yz - \lambda e^{\gamma z} z = z(y - \lambda e^{\gamma z}), \\
\dot{l}_\eta(y, z) &= \frac{z}{\eta} - \frac{1-z}{1-\eta}.
\end{aligned}$$

Thus the score equations for $\theta = (\lambda, \gamma, \eta)$ are

$$\begin{aligned}
0 &= \sum_{i=1}^n \dot{l}_\lambda(Y_i, Z_i) = \frac{1}{\lambda}\sum_{i=1}^n (Y_i - \lambda e^{\gamma Z_i}), \\
0 &= \sum_{i=1}^n \dot{l}_\gamma(Y_i, Z_i) = \sum_{i=1}^n Z_i(Y_i - \lambda e^{\gamma Z_i}), \\
0 &= \sum_{i=1}^n \dot{l}_\eta(Y_i, Z_i) = \sum_{i=1}^n \left\{ \frac{Z_i}{\eta} - \frac{1 - Z_i}{1-\eta} \right\} = (\eta(1-\eta))^{-1} \sum_{i=1}^n (Z_i - \eta).
\end{aligned}$$

(b) The third equation always has the unique solution $\hat{\eta} = \overline{Z}_n$. It is clear that the score equation for λ has only the solution $\lambda = 0$ if all $Y_i = 0$, and in this case there is clearly no unique solution for γ. So as least one Y_i must be non-zero in order to get a unique solution. If all the Z_i's are equal to 1, then the two equations agree, and it is clear that all we can estimate is the product λe^{γ}. Similarly, if all the Z_i's are equal to 0, then the score equation for γ is trivially satisfied (for any γ), and the score equation for λ gives just an estimator of λ. For the first and second equations, we compute

$$\begin{aligned}
\ddot{l}_{n,\lambda\lambda} &= -\frac{1}{\lambda^2}\sum_{i=1}^n Y_i\,, \\
\ddot{l}_{n,\lambda\gamma} &= -\sum_{i=1}^n Z_i e^{\gamma Z_i} = \ddot{l}_{n,\gamma\lambda}\,, \\
\ddot{l}_{n,\gamma\gamma} &= -\lambda\sum_{i=1}^n Z_i^2 e^{\gamma Z_i}\,.
\end{aligned}$$

Thus the matrix of second partial derivatives (the Hessian) fails to be negative definite if $\sum_1^n Y_i = 0$ or if

$$\left(\lambda\sum_{i=1}^n Z_i^2 e^{\gamma Z_i}\right)\left(\lambda^{-2}\sum_{i=1}^n Y_i\right) \le \left(\sum_{i=1}^n Z_i e^{\gamma Z_i}\right)^2 .$$

Because the Z_i's are Bernoulli,

$$\sum_{i=1}^n Z_i^2 e^{\gamma Z_i} = \sum_{i=1}^n Z_i e^{\gamma Z_i} = e^{\gamma}\sum_{i=1}^n Z_i\,.$$

Thus the above inequality holds if and only if

$$\sum_{i=1}^n Y_i \le \lambda e^{\gamma}\sum_{i=1}^n Z_i\,. \tag{0.7}$$

(Note that upon division by n the right side converges in probability to $\eta_0\lambda e^{\gamma}$ while the left side (divided by n) converges to $\eta_0\lambda_0 e^{\gamma_0} + \lambda_0(1-\eta_0)$.) At the MLE, the inequality (0.7) holds if and only if

$$\sum_{i=1}^n Y_i \le \sum_{i=1}^n Y_i Z_i\,.$$

This inequality can occur only if $Z_i = 1$ whenever $Y_i > 0$. Thus it becomes clear that the equations will have a unique solution yielding a maximum of the likelihood if at least one $Z_i = 0$ and $Y_i \ge 1$.

(c) Theorem 4.1.5 says that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N_3(0, I(\theta_0)^{-1})$$

where

$$I(\theta) = \begin{pmatrix} \lambda^{-1}E(e^{\gamma Z}) & E(Ze^{\gamma Z}) & 0 \\ E(Ze^{\gamma Z}) & \lambda E(Z^2 e^{\gamma Z}) & 0 \\ 0 & 0 & (\eta(1-\eta))^{-1} \end{pmatrix}.$$

(d) When θ_1 is true,

$$\begin{aligned}
n^{-1} \log \frac{L_n(\theta_1)}{L_n(\theta_0)} &= n^{-1} \sum_{i=1}^{n} \log \frac{p_{\theta_1}}{p_{\theta_0}}(X_i) \\
&\to_p E_{\theta_1} \log \frac{p_{\theta_1}}{p_{\theta_0}}(X_1) \\
&= K(P_{\theta_1}, P_{\theta_0})\,.
\end{aligned}$$

Here

$$\begin{aligned}
\log \frac{p_{\theta_1}}{p_{\theta_0}}(x) &= y \log \left(\frac{\lambda_1 e^{\gamma_1 z}}{\lambda_0 e^{\gamma_0 z}} \right) - \lambda_1 e^{\gamma_1 z} + \lambda_0 e^{\gamma_0 z} \\
&\quad + z \log \frac{\eta_1}{\eta_0} + (1 - z) \log \frac{1 - \eta_1}{1 - \eta_0}\,,
\end{aligned}$$

so

$$\begin{aligned}
K(P_{\theta_1}, P_{\theta_0}) &= E_{\theta_1} \log \frac{p_{\theta_1}}{p_{\theta_0}}(X_1) \\
&= E_{\theta_1} \left(Y \log \frac{\lambda_1 e^{\gamma_1 Z}}{\lambda_0 e^{\gamma_0 Z}} \right) + \eta_1 (\lambda_0 e^{\gamma_0} - \lambda_1 e^{\gamma_1}) + (1 - \eta_1)(\lambda_0 - \lambda_1) \\
&\quad + \eta_1 \log \frac{\eta_1}{\eta_0} + (1 - \eta_1) \log \frac{1 - \eta_1}{1 - \eta_0} \\
&= E_{\theta_1} \left(\lambda_1 e^{\gamma_1 Z} \log \left(\frac{\lambda_1 e^{\gamma_1 Z}}{\lambda_0 e^{\gamma_0 Z}} \right) \right) + \eta_1 (\lambda_0 e^{\gamma_0} - \lambda_1 e^{\gamma_1}) + (1 - \eta_1)(\lambda_0 - \lambda_1) \\
&\quad + \eta_1 \log \frac{\eta_1}{\eta_0} + (1 - \eta_1) \log \frac{1 - \eta_1}{1 - \eta_0} \\
&= \eta_1 \left(\lambda_1 e^{\gamma_1} \log \left(\frac{\lambda_1 e^{\gamma_1}}{\lambda_0 e^{\gamma_0}} \right) \right) + (1 - \eta_1) \left(\lambda_1 \log \left(\frac{\lambda_1}{\lambda_0} \right) \right) \\
&\quad + \eta_1 (\lambda_0 e^{\gamma_0} - \lambda_1 e^{\gamma_1}) + (1 - \eta_1)(\lambda_0 - \lambda_1) \\
&\quad + \eta_1 \log \frac{\eta_1}{\eta_0} + (1 - \eta_1) \log \frac{1 - \eta_1}{1 - \eta_0}\,.
\end{aligned}$$