Statistics 581, Problem Set 9 Solution

Wellner; 12/06/2010

1. Consider the Weibull family of example 3.2.5 and problem set #7, 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 7.)

(b) Use the result of (a) to construct method of moments estimators or quantile based estimators $\overline{\theta}_n$ of $\theta = (\alpha, \beta)$.

(c) Show that the method of moments or quantile estimators $\overline{\theta}_n$ of θ are asymptotically normal, and find the asymptotic distribution; i.e. show that

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

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

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

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

Thus it follows that

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

or equivalently that

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

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

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

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

where $\gamma = .577...$ is Euler's constant. Since $\overline{T} = -3.0011...$ and $\tilde{S}_T = 2.0423...$ (biased variance estimator) or $S_T = 2.1331...$ (unbiased variance estimator), moment estimators of (α, β) based on (8) are given by

$$\overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}}\frac{1}{\tilde{S}_T} = .62800..., \qquad \overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}}\frac{1}{S_T} = .60127...$$

and for these two estimators of β,

$$\overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 50.4097, \qquad \overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 52.5126...$$

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

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

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, 25, 46, 54, 68, 109, 319, 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 $p_{(i)} \equiv X_{(i)}^\beta / \sum_j X_j^\beta$,

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

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

$$\check{\theta}_n \equiv \overline{\theta}_n + \hat{I}_n^{-1}(\overline{\theta}_n)(\frac{1}{n}\dot{l}(\overline{\theta}_n)) = (54.6895\ldots,\ 0.562062\ldots).$$

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

$$\check{\theta}_n = (53.6888\ldots, 0.564337\ldots),$$

(c) The maximum likelihood estimate $\hat{\theta}_n = (54.5451..., 0.565953...)$, but note that the likelihood surface is quite flat as a function of α as shown in the plots on the following pages.

Mathematica input for moment and one-step estimators:

```
(* Here is the data:  *)
Print["Here is the data:"]
x = {1, 1, 2, 3, 12, 25, 46, 54, 68, 109, 319, 413 }
(* NSS is the sample size *)
NSS := Length[x]
(* First transform to -Log[x]:  *)
t = N[-Log[x]]
(* Now compute Mean and Variance of t *)
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]
```

```
(* 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: *)
Print["Moment estimator of beta, version 1:"]
betabar = N[Sqrt[VarGumbel/Stt^2]]
Print["Moment estimator of beta, version 2:"]
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"]
thetabar2 = {alphabar2, betabar2}

(* 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
Print["Information matrix estimator 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[48]:= Here is the data:

Out[50]= {1, 1, 2, 3, 12, 25, 46, 54, 68, 109, 319, 413}

Out[55]= {0., 0., -0.693147, -1.09861, -2.48491, -3.21888, -3.82864, -3.98898, \
-4.21951, -4.69135, -5.76519, -6.02345}

During evaluation of In[48]:= Mean of T = - Log[x]

Out[58]= -3.00106
```

```
During evaluation of In[48]:= Standard deviation of T

Out[60]= 2.13308

Out[61]= 13.1772

Out[62]= 4.17085

During evaluation of In[48]:= Biased estimator of std. dev

Out[64]= 2.04227

During evaluation of In[48]:= Moment estimator of beta, version 1:

Out[71]= 0.601268

During evaluation of In[48]:= Moment estimator of beta, version 2:

Out[73]= 0.628004

During evaluation of In[48]:= Moment estimator of alpha, version 1:

Out[75]= 52.5126

During evaluation of In[48]:= Moment estimator of alpha, version 2:

Out[77]= 50.4097

During evaluation of In[48]:= theta bar estimator, version 1

Out[79]= {52.5126, 0.601268}

During evaluation of In[48]:= theta bar estimator, version 2

Out[81]= {50.4097, 0.628004}

During evaluation of In[48]:= Information matrix estimator based on thetabar

Out[100]= {{0.000131102, -0.0080511}, {-0.0080511, 5.04444}}

During evaluation of In[48]:= inverse information matrix estimator based on thetab

Out[102]= {{8456.52, 13.4969}, {13.4969, 0.219779}}

During evaluation of In[48]:= vector of scores evaluated at thetabar
```

```
Out[104]= {0.000601046, -0.215297}

During evaluation of In[48]:= sample size n (NSS in the program)

Out[106]= 12

During evaluation of In[48]:= adjustment to the preliminary estimator

Out[109]= {2.17692, -0.0392055}

During evaluation of In[48]:= resulting one step estimator; based on theoretical I

Out[112]= {54.6895, 0.562062}

During evaluation of In[48]:= information matrix based on - Hessian of log-likelih

Out[116]= {{0.00014943, -0.0115159}, {-0.0115159, 5.46303}}

During evaluation of In[48]:= inverse information matrix from Hessian

Out[119]= {{7990.13, 16.8429}, {16.8429, 0.218553}}

During evaluation of In[48]:= adjustment to the preliminary estimator

Out[122]= {1.17622, -0.0369303}

During evaluation of In[48]:= resulting Hessian based version of one-step estimato

Out[125]= {53.6888, 0.564337}
```

Mathematica output for one-step estimators

Mathematica input for maximum likelihood estimators:

```
Clear[a,b,ahat,bhat]
(* Here is the data:  *)
x = {1, 1, 2, 3, 12, 25, 46, 54, 68, 109, 319, 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:

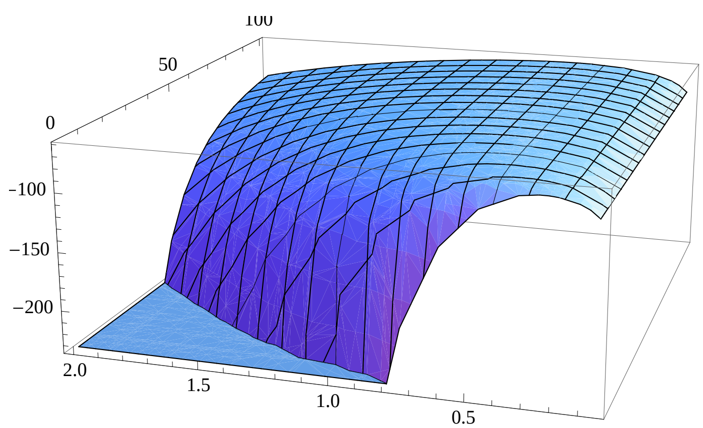

Figure 1: Weibull likelihood.

```
{1, 1, 2, 3, 12, 25, 46, 54, 68, 109, 319, 413}

12

{61.6213, {b -> 0.565953}}

{b -> 0.565953}
```

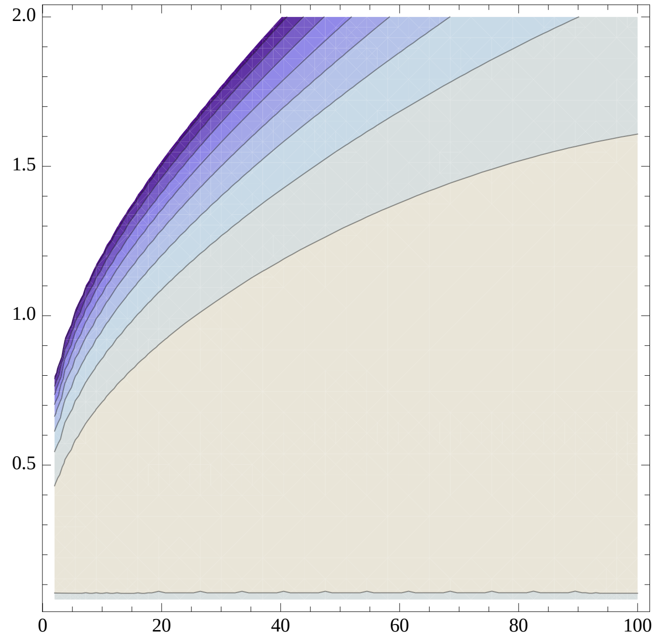

Figure 2: Contour plot Weibull likelihood

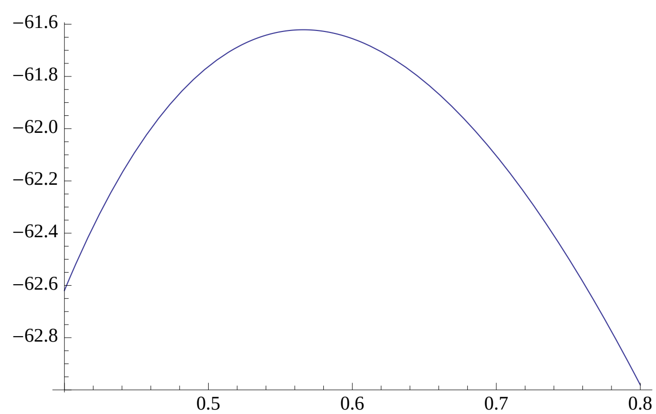

Figure 3: Profile plot Weibull likelihood

```
0.565953
```

```
54.5451
```

```
{61.8225, {a -> 47., b -> 0.62}}
```

3. (a) Ferguson, ACILST, problem 17.2, page 117: I would suggest modifiying 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

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

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

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

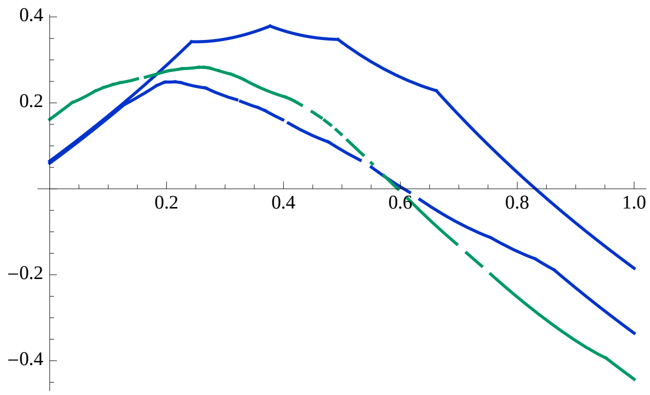

Figure 4: Three log-likelihood functions for problem 3, $n = 4,\ n = 10$, and $n = 50$

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 $\theta_0 = .2$.

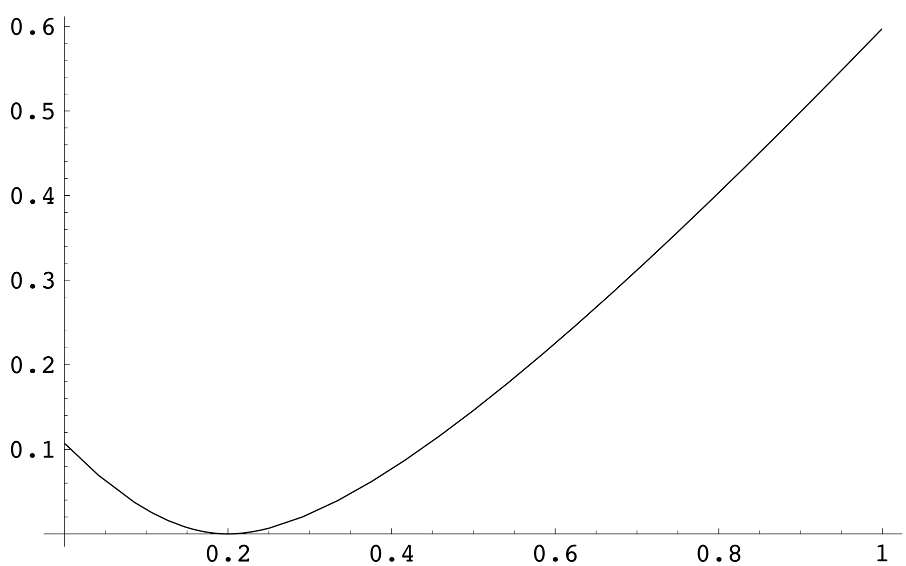

Figure 5: Kullback - Leibler function $K(P_{\theta_0}, P_\theta)$, $\theta_0 = .2$

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

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

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

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

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. 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:

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

$$= e^{2\mu}e^{\sigma^2}\left\{e^{\sigma^2} - 1\right\}.$$

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 $I^{-1}(\theta) = \text{diag}(\sigma^2, 2\sigma^4)$. Therefore by Corollary 4.1.1

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

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

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

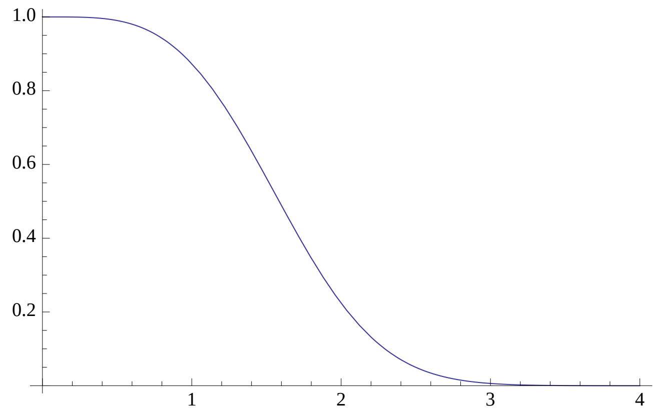

Figure 6: Ratio of asymptotic variance of MLE to variance of $\sqrt{n}(\overline{X}_n - E_\theta(X))$, lognormal mean estimation, as a function of σ.

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

5. (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)(i) For the more general case in which $X \sim \text{Binomial}(n,p)$ From (a), the MLE $\hat{p}$ of $p \in [1/3, 2/3]$ is

$$\hat{p} = \begin{cases} X/n, & \text{if } X/n \in [1/3, 2/3], \\ 1/3, & \text{if } X/n < 1/3, \\ 2/3, & \text{if } X/n > 2/3 \,. \end{cases}$$

For $n = 1$, this implies that the MLE is $1/3$ if $X = 0$ and $2/3$ if $X = 1$.
(ii) Now the estimator $\delta(X) = 1/2$ has expected squared error

$$R_1(p) \equiv E_p(\delta(X) - p)^2 = (1/2 - p)^2, \qquad 1/3 \le p \le 2/3 \,.$$

On the other hand the MLE $\hat{p}$ has expected squared error

$$\begin{aligned} R_2(p) \equiv E_p(\hat{p} - p)^2 &= p(2/3 - p)^2 + q(1/3 - p)^2 \\ &> (1/2 - p)^2 = R_1(p) \end{aligned}$$

by noting that $R_2(1/3) = 1/3^3 > 1/6^2 = R_1(1/3)$, and $R_2(1/2) = (1/2)(1/6)^2 + (1/2)(1/6)^2 = 1/6^2 > 0 = R_1(1/2)$. See the following Figure zz for a comparison of these two mean-square errors.

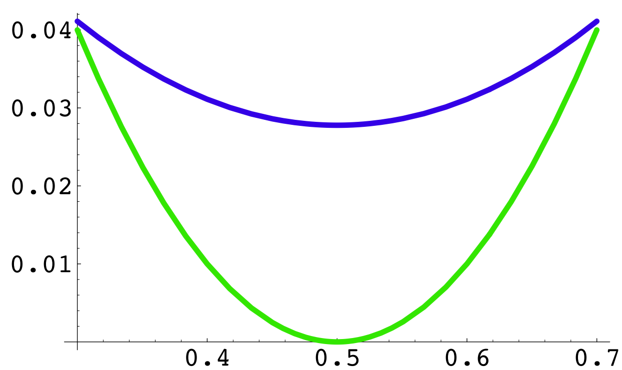

Figure 7: Mean-Square Errors.