Statistics 581, Problem Set 10 Solutions

Wellner; 12/07/2017

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

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

where $\theta \equiv (\alpha, \beta) \in (0, \infty) \times (0, \infty) \subset R^2$. Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with density function p_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 $\overline{T} = -2.9934...$ and $\tilde{S}_T = 2.04129...$ (biased variance estimator) or $S_T = 2.13205...$ (unbiased variance estimator), moment estimators of (α, β) based on (8) are given by

$$\overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}} \frac{1}{\tilde{S}_T} = .628305..., \qquad \overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}} \frac{1}{S_T} = .601556...$$

and for these two estimators of β,

$$\overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 52.0886, \qquad \overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 50.0036...$$

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

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

Mathematica Input: Weibull-moment-estimators.Fall.2017.nb

```
Print["Here is the data:"]
x = {1, 1, 2, 3, 12, 23, 46, 55, 66, 109, 320, 413}
(*  data 2014
x = {1, 1, 2, 3, 12, 21, 46, 54, 65, 109, 317, 413}
*)
```

```
(*  data 2010
x = {1, 1, 2, 3, 12, 25, 46, 54, 68, 109, 319, 413 }
*)
(* data 2009
x={1,1.3,1.7,3.2,10.7,24.3,51.2,77.1,93.7,105,111,305.}
*)
(* data 2008
x={1,1,2,3,12,25,46,56,79,125,323,417 }
*)

(* 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 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"]
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: Weibull-moment-estimators.Fall.2017.nb

```
Here is the data:
{1, 1, 2, 3, 12, 23, 46, 55, 66, 109, 320, 413}

12

Here is the transformed data, -Log[x]
{0., 0., -0.693147, -1.09861, -2.48491, -3.13549, -3.82864, -4.00733, \
-4.18965, -4.69135, -5.76832, -6.02345}

Mean of T = -Log(x)
-2.99341

Standard deviation of T:
2.13205

13.1273
4.16685
```

```
Biased estimator of std dev
2.04129

Moment estimator of beta, version 1:
0.601556

Moment estimator of beta, version 2:
0.628305

Moment estimator of alpha, version 1:
52.0886

Moment estimator of alpha, version 2:
50.0036

theta bar estimator, version 1
{52.0886, 0.601556}

theta bar estimator, version 2
{50.0036, 0.628305}

Information matrix based on thetabar
{{0.000133373, -0.00811664}, {-0.00811664, 5.0396}}

Inverse information matrix estimator based on thetabar
{{8312.53, 13.3879}, {13.3879, 0.219991}}

vector of scores evaluated at thetabar
{0.000634659, -0.227063}

sample size n (NSS in the program)
12

adjustment to the preliminary estimator
{2.23572, -0.041455}

resulting one step estimator; based on theoretical inform matrix
{54.3243, 0.560101}

information matrix based on -Hessian of log-likelihood
{{0.000152887, -0.011794}, {-0.011794, 5.49736}}
```

```
inverse information matrix from Hessian
{{7837.98, 16.8155}, {16.8155, 0.217981}}

adjustment to the preliminary estimator
{1.15626, -0.0388235}

resulting hessian based version of one-step estimator
{53.2449, 0.562733}
```

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 $n = 12$):

$$1,\ 1,\ 2,\ 3,\ 12,\ 23,\ 46,\ 55,\ 66,\ 109,\ 320,\ 413.$$

[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 (**??**). Substitution of (**??**) into **??** 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 (**??**) 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 (**??**) has a unique solution $\hat{\beta}$. By taking this value of $\hat{\beta}$ in (**??**), 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

$$\check{\theta}_n \equiv \overline{\theta}_n + \hat{I}_n^{-1}(\overline{\theta}_n)(\frac{1}{n}\dot{l}(\overline{\theta}_n)) = (54.3243\ldots\,,\ 0.560101\ldots).$$

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

$$\check{\theta}_n = (53.2449\ldots, 0.562733\ldots),$$

(c) The maximum likelihood estimate is $\hat{\theta}_n = (54.1705...\,,\ 0.56452...)$ (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.

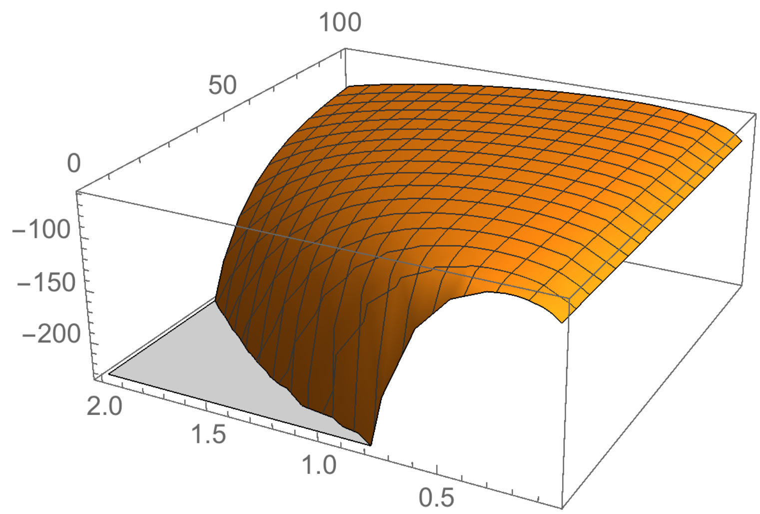

Figure 1: Weibull likelihood.

Mathematica Input: Weibull.Likelihood.17.nb

```
Clear[a,b,ahat,bhat]
(* Here is the data:  *)

(*  2017 data *)
x = {1,1,2,3,12,23,46,55,66,109,320,413}

(*  2014 data)
x = {1, 1, 2, 3, 12, 21, 46, 54, 65, 109, 317, 413 }
```

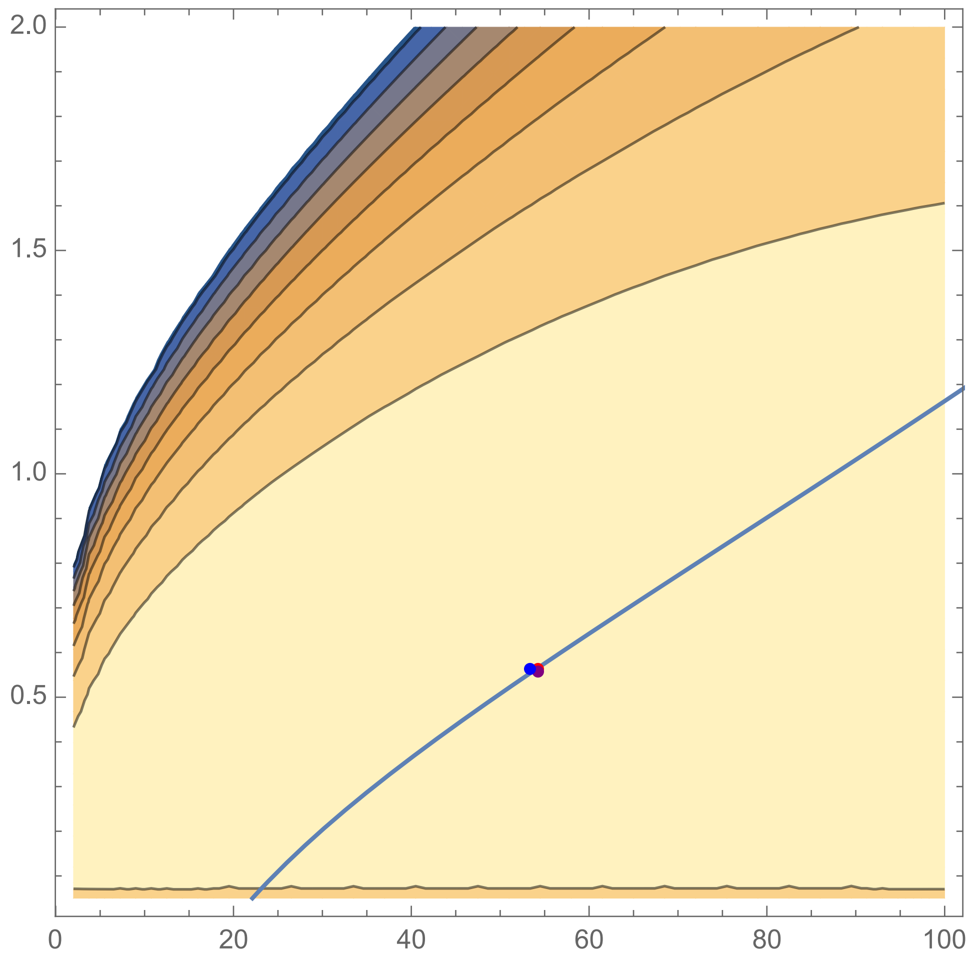

Figure 2: Contour plot Weibull likelihood/

```
*)
(* 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 }
*)
```

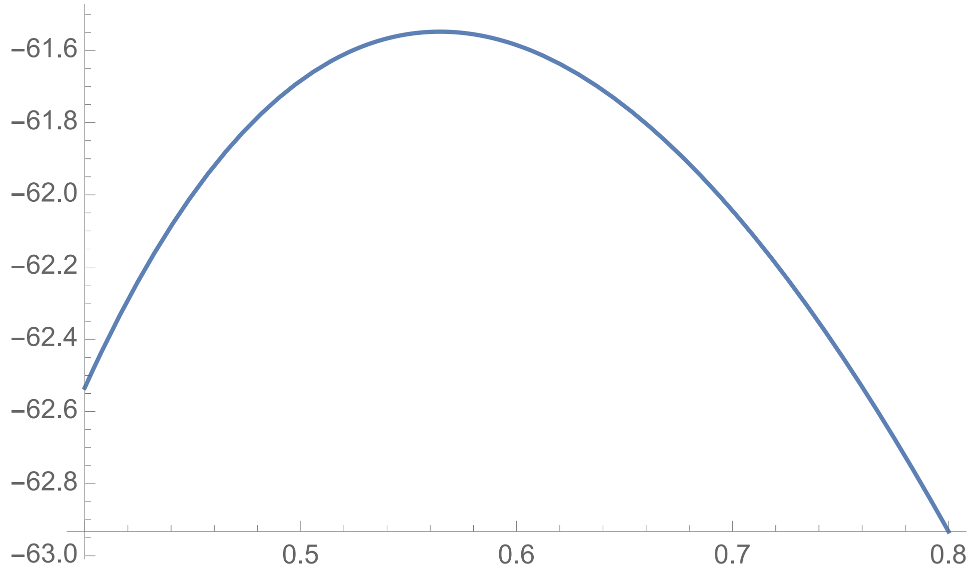

Figure 3: Profile plot Weibull likelihood

```
(* 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)

Print["3D plot of log-likelihood for alpha , beta"]
P2=Plot3D[L[a,b], {a,2,100}, {b,.05,2.0}]
Print["contour plot of log-likelihood for alpha , beta"]
```

```
P1=ContourPlot[L[a,b],{a,2,100},{b,.05,2.0}]

P3=ParametricPlot[{aprof[b],b},{b,.05,2.0}] (*PlotRange->{{2,100},{.05,2.0}}]*)

Print["profile likelihood as a function of beta"]
Plot[L[aprof[b],b],{b,.4,.8}]
Print["maximum of the profile likelihood"]
FindMinimum[-L[aprof[b],b],{b,.63}]
FindMinimum[-L[aprof[b],b],{b,.63}][[2]]

Print["MLE beta hat of beta"]
bhat=Replace[b,FindMinimum[-L[aprof[b],b],{b,.63}][[2]]]
Print["MLE alpha hat of alpha"]
ahat=aprof[bhat]

ostep1={54.3243,.56010}
ostep2={53.2449,.56273}
p1= Graphics[{PointSize[Medium],Red,Point[{ahat,bhat}]}]
p2= Graphics[{PointSize[Medium],Purple,Point[ostep1]}]
p3= Graphics[{PointSize[Medium],Blue,Point[ostep2]}]

q1= Graphics[{PointSize[Medium],Red,Point[{ahat,bhat,-61.5479}]}]

(*
q2= Graphics[{PointSize[Medium],Purple,Point[ostep1]}]
q3= Graphics[{PointSize[Medium],Blue,Point[ostep2]}]
*)


Show[P1,{p1,p2,p3}]
Show[P1,P3,{p1,p2,p3}]
Show[P2,{q1}]

Print["direct minimization of log-likelihood (no profiling)"]
MinL=FindMinimum[-L[a,b], {a,47},{b,.62}]
MinL[[1]]
MinL[[2]]


Wald[a_,b_] := NSS*({ahat,bhat} -{a,b}).Inf[ahat,bhat].({ahat,bhat} -{a,b})
Rao[a_,b_] := Sc[a,b].Inverse[Inf[a,b]].Sc[a,b]/NSS
RaoMod[a_,b_]:=Sc[a,b].Inverse[Inf[ahat,bhat]].Sc[a,b]/NSS
```

```
LR[a_,b_] := 2*(L[ahat,bhat] - L[a,b])
Print["Wald statistic for testing (alpha,beta) = (50,1)"]
Wald[50,1]
Print["Rao statistic for testing (alpha,beta) = (50,1)"]
Rao[50,1]
Print["Modified Rao statistic for testing (alpha,beta) = (50,1)"]
RaoMod[50,1]
Print["LR statistic for testing (alpha,beta) = (50,1)"]
LR[50,1]
```

Mathematica Output:

```
Out[795]= {1, 1, 2, 3, 12, 23, 46, 55, 66, 109, 320, 413}
Out[805]= 12
During evaluation of In[791]:= 3D plot of log-likelihood for alpha , beta

During evaluation of In[791]:= maximum of the profile likelihood
Out[835]= {61.5479, {b -> 0.564517}}
Out[836]= {b -> 0.564517}
During evaluation of In[791]:= MLE beta hat of beta
Out[839]= 0.564517
During evaluation of In[791]:= MLE alpha hat of alpha
Out[841]= 54.1705
Out[843]= {54.3243, 0.5601}
Out[844]= {53.2449, 0.56273}
```

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

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

Can you estimate γ and η by the method of moments? Can you improve method of moment estimators via one-step estimators?

Solution: (a) First,

$$E_\theta X = (1-\theta) + \theta \int_0^\infty x^2 e^{-x} dx = (1-\theta) + \theta\Gamma(3) = 1 - \theta + 2\theta = 1 + \theta\,.$$

Thus the method of moments estimator $\overline{\theta}_n$ of θ is given by $\overline{\theta}_n = \overline{X}_n - 1$. Now

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

Thus

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

Hence it follows by the CLT that

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

Now

$$l(\theta|X) = \log f(X|\theta) = \log[(1-\theta)e^{-x} + \theta x e^{-x}]\,,$$

and hence

$$\dot{l}_\theta(x) = \frac{xe^{-x} - e^{-x}}{(1-\theta)e^{-x} + \theta x e^{-x}} = \frac{x-1}{1+\theta(x-1)}\,.$$

Furthermore

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

Hence a one-step Newton approximation to a root of the likelihood equation is given by

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

where

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

Note that

$$I(\theta) = -E_\theta \ddot{l}_{\theta\theta}(X) = E_\theta \frac{(X-1)^2}{[1+\theta(X-1)]^2}$$

increases from 1 at $\theta = 0$ to ∞ at $\theta = 1$, so $1/I(\theta)$ decreases from 1 at $\theta = 0$ to 0 at $\theta = 1$, while the variance of the method of moments estimator, $1 + 2\theta - \theta^2$, increases from 1 to 2 as θ increases from 0 to 1. Hence the gain in efficiency by use of the efficient one-step estimator is quite large for θ near 1. See the plot of $1/I(\theta)$ and $1 + 2\theta - \theta^2$ below.

(b) When Ferguson's density $f(x|\theta)$ with $\theta \in (0,1)$ is replaced by

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

with $\gamma \in (0,1)$ and $\eta > 0$, the parameter to be estimated is $\theta = (\gamma, \eta)$, and we can again implement a one step procedure starting from some $n^{1/4}-$consistent preliminary esimator $\overline{\theta}_n$. One possibility for $\overline{\theta}_n$ is a method of moments estimator. We calculate

$$\begin{aligned} E(X) &= (1-\gamma) + \gamma\frac{2}{\eta} = 1 + \gamma\left(\frac{2}{\eta} - 1\right) \\ E(X^2) &= (1-\gamma)2 + \gamma\frac{6}{\eta^2} = 2 + \gamma\left(\frac{6}{\eta^2} - 2\right)\,. \end{aligned}$$

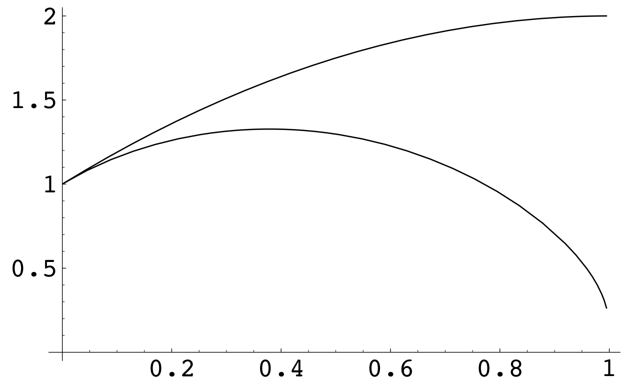

Figure 4: $1/I(\theta)$ and $1 + 2\theta - \theta^2$

For $\eta \neq 2$ this yields

$$\frac{E(X^2) - 2}{E(X) - 1} = \frac{6/\eta^2 - 2}{2/\eta - 1} = \frac{6 - 2\eta^2}{2\eta - \eta^2}\,. \tag{0.6}$$

The difficulty is that solving this for η yields two non-negative solutions in general. I have not yet found a "nice" starting point (preliminary estimator) $\overline{\theta}_n$ for this problem.

But once we have found a starting point, the one-step procedure is again relatively simple: we calculate

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

Then

$$\check{\theta}_n = \overline{\theta}_n + \widehat{I}_n^{-1}\frac{1}{n}\dot{\mathbf{l}}_n(\overline{\theta}_n|\underline{X})$$

where

$$\dot{\mathbf{l}}_n(\overline{\theta}_n|\underline{X}) = \sum_{i=1}^{n} \dot{\mathbf{l}}_\theta(\overline{\theta}_n|X_i)$$

and

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

4. Ferguson, ACLST, page 150, problem 3. Does the theory in our chapter 4 (or Ferguson's chapter 22) apply directly? Does the local asymptotic power of your test depend on the common value of θ_j in the null hypothesis?

   Solution: The theory in chapter 4 of the course notes does not apply directly since the data is *not* i.i.d., at least in the form given in Ferguson. The difficulty is that the distribution of the data in the general (unconstrained) setting is not that of i.i.d. random variables from one distribution, but that of k independent samples from from different distributions, namely Poisson(θ_i), $i = 1, \ldots, k$. On the other hand, in this special case with all the sample sizes equal to n we can consider the data as consisting of the vectors $\underline{X}_j = (X_{1,j}, \ldots, X_{k,j})$ for $j = 1, \ldots, n$ where the components $X_{i,j}$ of $\underline{X}_j$ are independent Poisson(θ_i) random variables. Thus the $\underline{X}_j$ random vectors are i.i.d. with (joint) probability mass function given by

$$p_{\underline{\theta}}(\underline{x}) = \prod_{i=1}^{k} \exp(-\theta_i)\frac{\theta_i^{x_i}}{x_i!}.$$

   In this way the setting in section 4.1 of the course notes does apply. (Note that this apparently breaks down if the sample sizes $n_1, \ldots, n_k$ in the separate Poisson populations are possibly different.)

   Now we calculate

$$\log p_{\underline{\theta}}(\underline{x}) = \sum_{i=1}^{k} \{x_i \log\theta_i - \theta_i - \log(x_i!)\}$$

   and

$$\dot{\mathbf{l}}_\theta(\underline{x}) = \left(\frac{x_1}{\theta_1} - 1, \ldots, \frac{x_k}{\theta_k} - 1\right)^T,$$

   so that we have, by independence of the coordinates of $\underline{X}$,

$$I(\underline{\theta}) = \begin{pmatrix} \theta_1^{-1} & 0 & \ldots & 0 \\ 0 & \theta_2^{-1} & \ldots & 0 \\ \vdots & 0 & \ldots & 0 \\ 0 & \ldots & 0 & \theta_k^{-1} \end{pmatrix} = \mathrm{diag}(\underline{\theta}^{-1}).$$

Thus the (unrestricted) MLE of $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ is given by

$$\hat{\underline{\theta}} = (\overline{X}_1, \ldots, \overline{X}_k)$$

where $\overline{X}_i = n^{-1} \sum_{j=1}^{n} X_{i,j}$ for $i = 1, \ldots, k$, and it follows from Theorem 4.1.2 that

$$\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}) \to_d N_k(0, I^{-1}(\underline{\theta})) = N_k(0, \mathrm{diag}(\underline{\theta})).$$

Under the null hypothesis that all the θ_i's are equal, all the $X_{i,j}$'s are i.i.d. Poisson(θ) and the MLE of $\underline{\theta} = \theta\underline{1}$ is

$$\hat{\underline{\theta}}^0 = \frac{1}{nk} \sum_{i=1}^{k} \sum_{j=1}^{n} X_{i,j}\underline{1} \equiv \overline{X}\underline{1}.$$

In this case Theorem 4.1.2 applies directly and we have

$$\sqrt{n}\left(\hat{\underline{\theta}}^0 - \underline{\theta}^0\right) = \sqrt{n}(\overline{X}_n - \theta^0)\underline{1} \to D_0\underline{1} \sim N_1(0, k^{-1}\theta^0)\underline{1} \sim N_k(0, k^{-1}\theta^0\underline{1}\underline{1}^T).$$

and

$$\sqrt{n}\left(\overline{X} - \theta^0\right) = \sqrt{n}\left(k^{-1}\sum_{i=1}^{k}\overline{X}_i - \theta^0\right) \to k^{-1/2}D_0 \sim N(0, k^{-1}\theta^0).$$

Moreover, under the null hypothesis it is easily seen that

$$\sqrt{n}\begin{pmatrix} \overline{X}_1 - \theta^0 \\ \vdots \\ \overline{X}_k - \theta_0 \\ k^{-1}\sum_{i=1}^{k}\overline{X}_i - \theta^0 \end{pmatrix} \to_d \begin{pmatrix} \underline{D} \\ \overline{D} \end{pmatrix} \sim N_{k+1}\left(0, \theta^0\begin{pmatrix} I_{k\times k} & k^{-1}\underline{1} \\ k^{-1}\underline{1}^T & k^{-1}1 \end{pmatrix}\right),$$

and, furthermore, that

$$\sqrt{n}\begin{pmatrix} \overline{X}_1 - \overline{X} \\ \vdots \\ \overline{X}_k - \overline{X} \end{pmatrix} \to_d \begin{pmatrix} D_1 - \overline{D} \\ \vdots \\ D_k - \overline{D} \end{pmatrix} \sim N_k\left(0, \theta^0(I - k^{-1}\underline{1}\underline{1}^T)\right), \tag{0.7}$$

Note that $\dim(\Theta) = k$ and $\dim(\Theta_0) = 1$. Since

$$L_n(\theta_1, \ldots, \theta_k) = \prod_{i=1}^{k} \exp(-n\theta_i)\frac{\theta_i^{\sum_{j=1}^{n} X_{ij}}}{\prod_{j=1}^{n} X_{i,j}!},$$

it follows that

$$l_n(\theta_1, \ldots, \theta_k) = \sum_{i=1}^{k}\left\{\sum_{j=1}^{n} X_{i,j}\log\theta_i - n\theta_i\right\}$$

and hence

$$l_n(\hat\theta_1, \ldots, \hat\theta_k) = n \sum_{i=1}^k \left\{ \overline{X}_i \log \overline{X_i} - \overline{X_i} \right\},$$

while

$$l_n(\hat\theta_1^0, \ldots, \hat\theta_k^0) = n \sum_{i=1}^k \left\{ \overline{X} \log \overline{X} - \overline{X} \right\} = n \left\{ k\overline{X} \log \overline{X} - k\overline{X} \right\}.$$

Hence the log-likelihood ratio statistic is given by

$$\begin{aligned} 2\log\lambda_n &= 2\{l_n((\hat\theta_1, \ldots, \hat\theta_k) - l_n(\hat\theta_1^0, \ldots, \hat\theta_k^0)\} \\ &= 2n \left\{ \sum_{i=1}^k \overline{X}_i \log \overline{X}_i - k\overline{X} \log \overline{X} \right\}. \end{aligned}$$

When the null hypothesis holds, our considerations in the i.i.d. case lead to the conclusion that $2\log\lambda_n \to_d \chi^2_{k-1}$. It is instructive to consider the natural Wald statistic W_n in this problem problem starting from (**??**) and see that we also have $W_n \to_d \chi^2_{k-1}$ under the null hypothesis. If $\underline{\theta}_n = (\theta_{n,1}, \ldots, \theta_{n,k}) = (\theta^0 + n^{-1/2}t_1, \ldots, \theta^0 + n^{-1/2}t_k)$ where $t_i \ne t_{i'}$ for some $i \ne i'$, then I claim that $2\log\lambda_n \to_d \chi^2_{k-1}(\delta)$ where $\delta = \sum_{i=1}^k (t_i - \overline{t})^2/\theta^0$ and similarly for W_n. Thus the noncentrality parameter δ depends inversely on θ^0.

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

Solution: (a) Note $K(P^n, Q^n) = E_{P^n} \log(p_n/q_n)$ where $p_n(\underline{x}) = \prod_{i=1}^n p(x_i)$ and $q_n(\underline{x}) = \prod_{i=1}^n q(x_i)$, so

$$\begin{aligned} K(P^n, Q^n) &= E_{P^n} \log(p_n/q_n)(\underline{X}) = E_{P^n} \sum_{i=1}^n \log \frac{p(X_i)}{q(X_i)} \\ &= \sum_{i=1}^n E_{P^n} \log \frac{p(X_i)}{q(X_i)} = nE_P \log \frac{p(X_1)}{q(X_1)} \\ &= nK(P, Q). \end{aligned}$$

(b) If P is $N(0, \sigma^2)$ and $Q = N(\mu, \sigma^2)$, then

$$\frac{q}{p}(x) = \frac{\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}x^2\right)} = \exp\left(\frac{\mu x}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right),$$

so

$$\begin{aligned} K(P,Q) &= E_P\left\{-\log\frac{q}{p}(X)\right\} = E_P\left\{-\left(\frac{\mu X}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right)\right\} \\ &= \frac{\mu^2}{2\sigma^2}, \end{aligned}$$

$$\begin{aligned} \rho(P,Q) &= \int \sqrt{pq}\,d\lambda = \int \sqrt{\frac{q}{p}}\,p\,d\lambda \\ &= E_P \exp\left(\frac{\mu X}{2\sigma^2} - \frac{\mu^2}{4\sigma^2}\right) \\ &= \exp\left(\left(\frac{\mu}{2\sigma^2}\right)^2 \frac{\sigma^2}{2} - \frac{\mu^2}{4\sigma^2}\right) \\ &= \exp\left(-\frac{\mu^2}{8\sigma^2}\right), \end{aligned}$$

and hence

$$H^2(P,Q) = 1 - \rho(P,Q) = 1 - \exp\left(-\frac{\mu^2}{8\sigma^2}\right).$$

(c) From (a) we have $K(P^n, Q^n) = nK(P,Q) = n\mu^2/(2\sigma^2)$. From Lemma 2.5.2 it follows that

$$\rho(P^n, Q^n) = \rho(P,Q)^n = \exp\left(-\frac{n\mu^2}{8\sigma^2}\right),$$

and hence

$$H^2(P^n, Q^n) = 1 - \exp\left(-\frac{n\mu^2}{8\sigma^2}\right).$$

(d) When $\mu_n = c/\sqrt{n}$ for some $c \in \mathbb{R}$ we see that

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

exactly for every n.