Statistics 581, Problem Set 10 Solutions

Wellner; 12/06/2018

1. 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

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

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

$$\sqrt{n}\begin{pmatrix} \overline{X}_1 - \overline{X} \\ \vdots \\ \overline{X}_k - \overline{X} \end{pmatrix} \to_d \begin{pmatrix} D_1 - \overline{D} \\ \vdots \\ D_k - \overline{D} \end{pmatrix} \sim N_k\left(0, \theta^0(I - k^{-1}\underline{1}\underline{1}^T)\right), \tag{0.1}$$

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 $2 \log \lambda_n \rightarrow_d \chi^2_{k-1}$. It is instructive to consider the natural Wald statistic W_n in this problem problem starting from (0.1) and see that we also have $W_n \rightarrow_d \chi^2_{k-1}$ 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 $t_i \neq t_{i'}$ for some $i \neq i'$, then I claim that $2 \log \lambda_n \rightarrow_d \chi^2_{k-1}(\delta)$ 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.

2. Ferguson, ACLST, page 149, problem 2 modified as follows:
(a) Find the LR test statistic of the null hypothesis $H_0 : \mu = c\theta$ for any fixed number $c > 0$, and find the asymptotic distribution of the LR statistic under H_0.
(b) Does the theory of our chapter 4 (or Ferguson's chapter 22) apply directly?
(c) Does the local asymptotic power of your test depend on c?

Solution: (b) First, allow me to slightly re-name the parameters: I will assume that $X_1, \ldots, X_n$ are i.i.d. $\exp(\lambda)$ and $Y_1, \ldots, Y_n$ are i.i.d. $\exp(\mu)$, so that $\theta = (\lambda, \mu)$. Furthermore, we can recast the problem into the context of chapter 4 by considering the pairs of observations (X_i, Y_i), $i = 1, \ldots, n$ as i.i.d. with density

$$p_\theta(x, y) = p_{(\lambda, \mu)}(x, y) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x) \mu e^{-\mu y} 1_{(0,\infty)}(y).$$

Now we are testing $H_0 : \mu = c\lambda$ versus $H_1 : \mu \neq c\lambda$. By a reparametrizaton, we can put this exactly in the setting of Section 4.2: if the original parameter is $\theta = (\lambda, \mu)$, then the new parameters $\gamma = (\gamma_1, \gamma_2)$ where $\gamma_1 \equiv \lambda$, $\gamma_2 \equiv \mu - c\lambda$. Then the null hypothesis H_0 becomes $H_0 : \gamma_2 = 0, \gamma_1 =$ anything.
(a) The MLE $\widehat{\theta}$ of $\theta = (\lambda, \mu)$ under H_1 is $\widehat{\theta} = (\widehat{\lambda}, \widehat{\mu})$ where $\widehat{\lambda} = 1/\overline{X}$ and $\widehat{\mu} = 1/\overline{Y}$. The MLE $\widehat{\theta}^0$ under H_0 is $(\widehat{\lambda}^0, c\widehat{\lambda}^0)$ where

$$\widehat{\lambda}^0 = 2/(\overline{X} + c\overline{Y}).$$

Now

$$l_n(\theta) = l_n(\lambda, \mu) = \sum_{i=1}^n \{\log \lambda - \lambda X_i + \log \mu - \mu Y_i\} = n \log \lambda + n \log \mu - n\overline{X}\lambda - n\overline{Y}\mu.$$

Thus the LR statistic for testing H_0 versus H_1 is given by

$$\begin{aligned} 2(l_n(\widehat{\theta}) - l_n(\widehat{\theta}^0)) &= 2n\left\{2\log\left(\frac{\overline{X} + c\overline{Y}}{2}\right) - \log(\overline{X}) - \log(c\overline{Y})\right\} \\ &\to_d \chi_1^2 \end{aligned}$$

under H_0.
(c) To compute the local asymptotic power of the LR test, we can reparametrize the problem by $\gamma \equiv (\gamma_1, \gamma_2)$ where $\gamma_1 \equiv \lambda$, $\gamma_2 \equiv \mu - c\lambda$. Then the null hypothesis H_0 becomes $H_0 : \gamma_2 = 0, \gamma_1 =$ anything. Then the problem fits in the context of Theorem 4.2.7: under P_{γ_n} with $\gamma_n = \gamma_0 + tn^{-1/2}$ for $\gamma_0 = (\gamma_{10}, 0)$ in the null hypothesis, we have

$$2\log\lambda_n \to_d \chi_1^2(\delta)$$

where the non-centrality parameter δ is given by $t_2^2 I_{22\cdot 1}(\gamma_0)$, and it remains only to compute $I_{22\cdot 1}$. By straightforward computation the information matrix for γ is given by

$$I(\gamma) = \begin{pmatrix} \frac{1}{\gamma_1^2} + \frac{c^2}{(c\gamma_1+\gamma_2)^2} & \frac{c}{(c\gamma_1+\gamma_2)^2} \\ \frac{c}{(c\gamma_1+\gamma_2)^2} & \frac{1}{(c\gamma_1+\gamma_2)^2} \end{pmatrix}.$$

Thus, under the null hypothesis $H_0 : \gamma_2 = 0$ we find that

$$I_{22\cdot 1}(\gamma_0) = I_{22}(\gamma_0) - I_{21}(\gamma_0)I_{11}^{-1}(\gamma_0)I_{12}(\gamma_0) = \frac{1/2}{c^2\gamma_1^2}$$

which does depend on c: the noncentrality power of the limiting distribution *decreases* as c^{-2} as c increases.

3. Ferguson, ACLST, page 118, problem 3. (See also Example 4.3.7, page 21, Chapter 4 notes.) [Neyman and Scott (1948)] Suppose we have a sample of size d from each of n normal populations with common unknown variance but possibly different unknown means $X_{i,j} \sim N(\mu_i, \sigma^2)$, $i = 1, \ldots, n$, $j = 1, \ldots, d$ where all the $X_{i,j}$ are independent.
(a) Find the maximum-likelihood estimate of σ^2.
(b) Show that for d fixed the MLE of σ^2 is not consistent as $n \to \infty$. Why don't either of Theorem 17 (Ferguson) or our Theorem 4.1.2 apply?
(c) Find a consistent estimate of σ^2.

Solution: (a) The likelihood is given by

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

and hence

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

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

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

where

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

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

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

It follows from the strong law of large numbers that

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

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

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

4. 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 a problem in an earlier problem set.)
(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 $\overline{T} = -3.0130...$ and $\tilde{S}_T = 2.0388...$ (biased variance estimator) or $S_T = 2.1295...$ (unbiased variance estimator), moment estimators of (α, β) based on (8) are given by

$$\overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}}\frac{1}{\tilde{S}_T} = .6023..., \qquad \overline{\beta}_n \equiv \frac{\pi}{\sqrt{6}}\frac{1}{S_T} = .6291...$$

and for these two estimators of β,

$$\overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 53.0588, \qquad \overline{\alpha} = \exp(-\overline{T} + \frac{\gamma}{\overline{\beta}}) = 50.9375...$$

respectively for the given data in problem 5 below.
(c) Asymptotic normality of $(\overline{\alpha}_n, \overline{\beta}_n)$ follows from joint asymptotic normality of $(\overline{T}_n, S_T^2)$ and the delta method: by the multivariate CLT and Slutsky's theorem

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

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

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

Then since $(\overline{\alpha}, \overline{\beta}) = g(\overline{T}, S_T^2)$ and $(\alpha, \beta) = g(E_\theta T, Var_\theta(T))$ where $g \equiv (g_1, g_2) : R^2 \rightarrow R^2$ is defined by

$$g_1(x, y) = \exp(\frac{\gamma\sqrt{6}}{\pi}\sqrt{y} - x),$$

$$g_2(x, y) = \frac{\pi/\sqrt{6}}{\sqrt{y}},$$

it follows by the delta method with $\tilde{\underline{Z}} \equiv (Z_1, \sqrt{2}\sigma_T^2 Z_2)$ that

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

where

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

5. (Problem 4, 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, 14, 27, 41, 55, 66, 113, 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

$$l(\gamma|\underline{X}) = n\log\beta - n\log\eta + (\beta-1)\sum_{i=1}^n \log X_i - \frac{1}{\eta}\sum_{i=1}^n X_i^\beta. \tag{0.2}$$

Thus, with $\gamma_1 \equiv \beta$, $\gamma_2 \equiv \eta$, the likelihood equations become

$$\dot{l}_1(\gamma|\underline{X}) = \frac{n}{\beta} + \sum_{i=1}^n \log X_i - \frac{1}{\eta}\sum_{i=1}^n X_i^\beta \log X_i = 0, \tag{0.3}$$

and

$$\dot{l}_2(\gamma|\underline{X}) = -\frac{n}{\eta} + \frac{1}{\eta^2}\sum_{i=1}^n X_i^\beta = 0, \tag{0.4}$$

or

$$\hat{\eta}_n = \frac{1}{n}\sum_{i=1}^n X_i^{\hat{\beta}} \tag{0.5}$$

from 0.4. Note that for each fixed β the maximizer of the log-likelihood over $\eta > 0$ is achieved at $\hat{\eta}(\beta) \equiv n^{-1}\sum_{i=1}^n X_i^\beta$, and plugging this back into $l(\gamma|\underline{X})$ in (0.2) yields the *profile log-likelihood*

$$\begin{aligned} l_n^{prof}(\beta|\underline{X}) &= l((\beta, \hat{\eta}(\beta))|\underline{X}) \\ &= n\log\beta - n\log\hat{\eta}(\beta) + (\beta-1)\sum_{i=1}^n \log X_i - n. \end{aligned}$$

Substitution of 0.5 into 0.3 yields the equation

$$\frac{\sum_i X_i^{\hat{\beta}}\log X_i}{\sum_i X_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}} = \frac{1}{n}\sum_{i=1}^n \log X_i\,, \tag{0.6}$$

or

$$h(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^n \log X_i \tag{0.7}$$

where

$$h(\beta) \equiv \frac{\sum_i X_i^\beta \log X_i}{\sum_i X_i^\beta} - \frac{1}{\beta} < \frac{\sum_i X_i^\beta \log X_i}{\sum_i X_i^\beta}$$

since $\beta > 0$. Now

$$\begin{aligned} h'(\beta) &= \frac{\sum_i X_i^\eta (\log X_i)^2}{\sum_i X_i^\beta} - (\frac{\sum_i X_i^\beta \log X_i}{\sum_i X_i^\beta})^2 + \frac{1}{\beta^2} \\ &\equiv I + II \\ &> I, \end{aligned}$$

and furthermore,

$$I = \sum a_i^2 p_i - (\sum a_i p_i)^2 = \mathrm{Var}_p(a)$$

since, with $a_i \equiv \log X_i$, $p_i \equiv X_i^\beta / \sum_j X_j^\beta \geq 0$, $\sum_i p_i = 1$. Thus $I > 0$ and hence $h'(\beta) > 0$ from (0.8) 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 $p_{(i)} \equiv X_{(i)}^\beta / \sum_j X_j^\beta$,

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

as $\beta \to \infty$.) Thus (0.7) has a unique solution $\hat{\beta}$. By taking this value of $\hat{\beta}$ in (0.5), 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 4(b) above. The one step estimator using $\hat{I}(\overline{\theta}_n) = I(\overline{\theta}_n)$ is

$$\check{\theta}_n \equiv \overline{\theta}_n + \hat{I}_n^{-1}(\overline{\theta}_n)(\frac{1}{n}\dot{l}(\overline{\theta}_n)) = (55.1538\ldots, 0.5648\ldots).$$

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

$$\check{\theta}_n = (54.2266\ldots, 0.5669\ldots),$$

(c) The maximum likelihood estimate $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n) = (54.1705..., 0.5645...)$, but note that the likelihood surface is quite flat as a function of α as shown in the plots on the following pages.

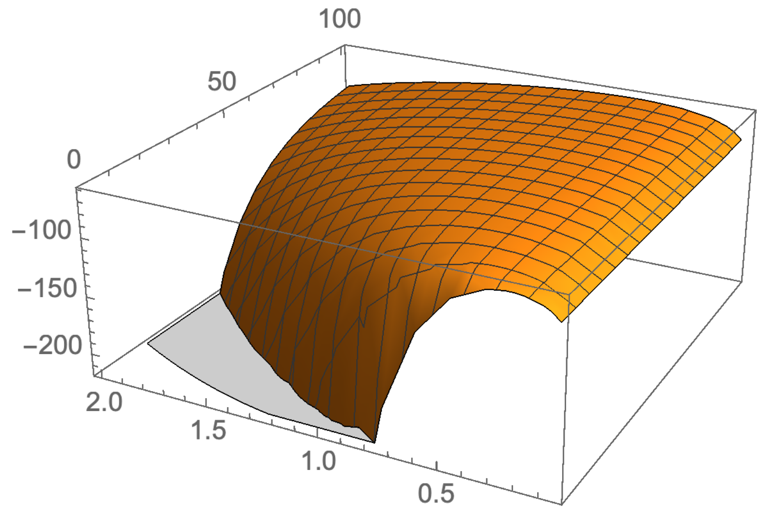

Figure 1: Weibull Likelihood

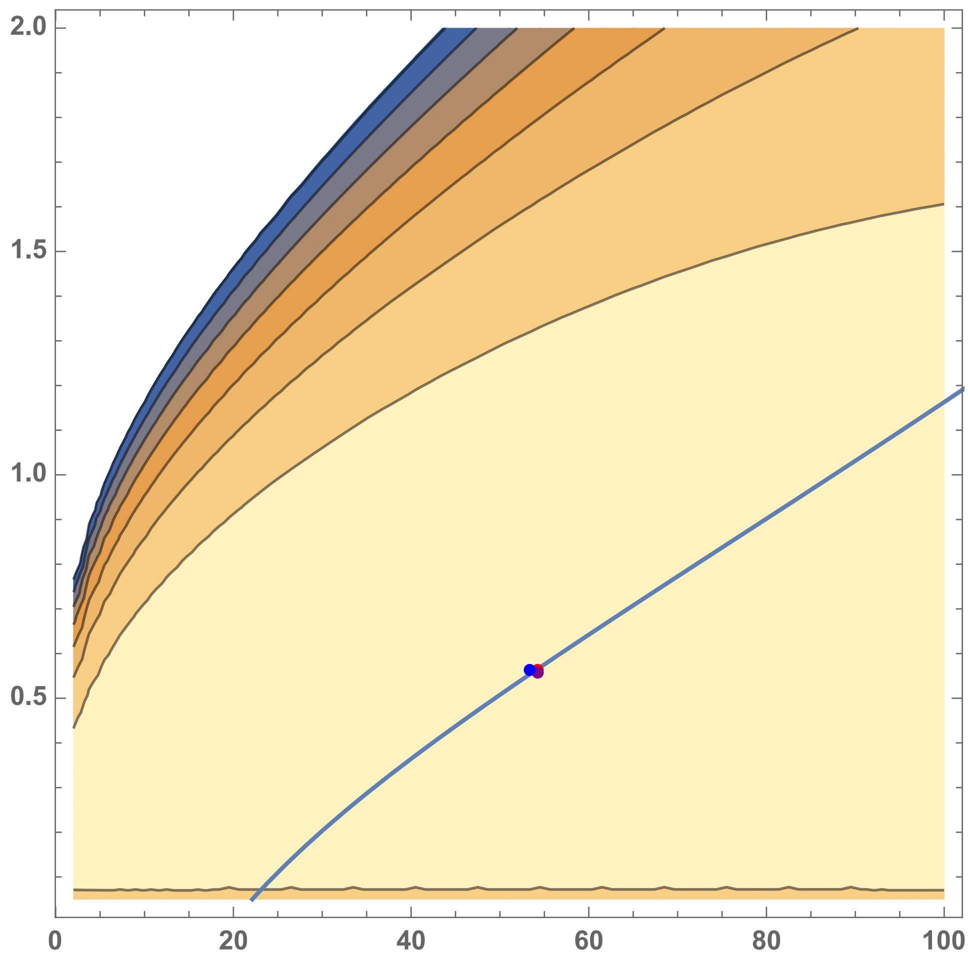

Figure 2: Contour plot Weibull Likelihood

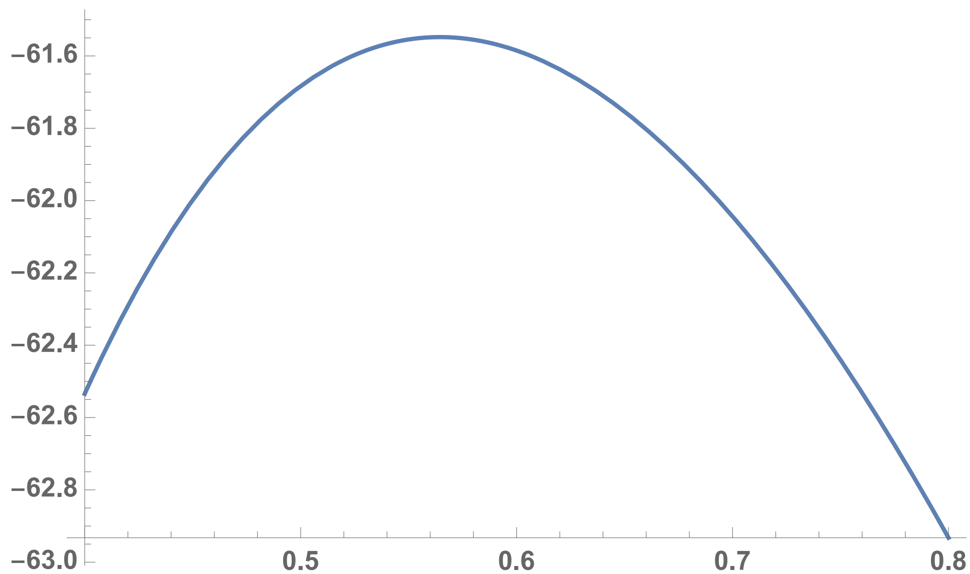

Figure 3: Weibull profile likelihood.