

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, \dots, 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 $\bar{\theta}_n$ of $\theta = (\alpha, \beta)$.

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

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

[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 $W \equiv -\log(Y) \sim \text{Gumbel}$:

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

Thus it follows that

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

or equivalently that

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

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

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

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

where $\gamma = .577\dots$ is Euler's constant. Since $\bar{T} = -2.9934\dots$ and $\tilde{S}_T = 2.04129\dots$ (biased variance estimator) or $S_T = 2.13205\dots$ (unbiased variance estimator), moment estimators of (α, β) based on (8) are given by

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

and for these two estimators of β ,

$$\bar{\alpha} = \exp(-\bar{T} + \frac{\gamma}{\bar{\beta}}) = 52.0886, \quad \bar{\alpha} = \exp(-\bar{T} + \frac{\gamma}{\bar{\beta}}) = 50.0036\dots$$

respectively for the given data in problem 2 below.

(c) Asymptotic normality of $(\bar{\alpha}_n, \bar{\beta}_n)$ follows from joint asymptotic normality of (\bar{T}_n, S_T^2) and the delta method: by the multivariate CLT and Slutsky's theorem

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

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 $(\bar{\alpha}, \bar{\beta}) = g(\bar{T}, S_T^2)$ and $(\alpha, \beta) = g(E_\theta T, \text{Var}_\theta(T))$ where $g \equiv (g_1, g_2) : R^2 \rightarrow R^2$ is defined by

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

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

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

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

where

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

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 $\bar{\theta}_n$ 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

$$\begin{aligned} p_{\theta}(x) &= \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right) 1_{(0,\infty)}(x) \\ &= \frac{\beta}{\eta} x^{\beta-1} \exp\left(-\frac{x^{\beta}}{\eta}\right) \\ &\equiv p_{\gamma}(x) \end{aligned}$$

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

$$l_1(\gamma|\underline{X}) = \frac{n}{\beta} + \sum_{i=1}^n \log X_i - \frac{1}{\eta} \sum_{i=1}^n X_i^{\beta} \log X_i = 0, \quad (0.1)$$

and

$$l_2(\gamma|\underline{X}) = -\frac{n}{\eta} + \frac{1}{\eta^2} \sum_{i=1}^n X_i^{\beta} = 0, \quad (0.2)$$

or

$$\hat{\eta}_n = \frac{1}{n} \sum_{i=1}^n X_i^{\hat{\beta}} \quad (0.3)$$

from (0.2). Substitution of (0.3) into 0.1 yields the equation

$$\frac{\sum_i X_i^{\hat{\beta}} \log X_i}{\sum_i X_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}} = \frac{1}{n} \sum_{i=1}^n \log X_i, \quad (0.4)$$

or

$$h(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n \log X_i \quad (0.5)$$

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

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

and furthermore,

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

since, with $a_i \equiv \log X_i$, $p_i \equiv X_i^\beta / \sum_j X_j^\beta \geq 0$, $\sum_i p_i = 1$. Thus $I > 0$ and hence $h'(\beta) > 0$ from (0.6) while

$$-\infty = \lim_{\beta \rightarrow 0} h(\beta) < \frac{1}{n} \sum_{i=1}^n \log X_i < \log X_{(n)} = \lim_{\beta \rightarrow \infty} h(\beta).$$

[Draw the picture!] (To see this last limit, note that with $p_{(i)} \equiv X_{(i)}^\beta / \sum_j X_j^\beta$,

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

as $\beta \rightarrow \infty$.) 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 $\hat{I}(\bar{\theta}_n) = I(\bar{\theta}_n)$ is

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

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

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

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

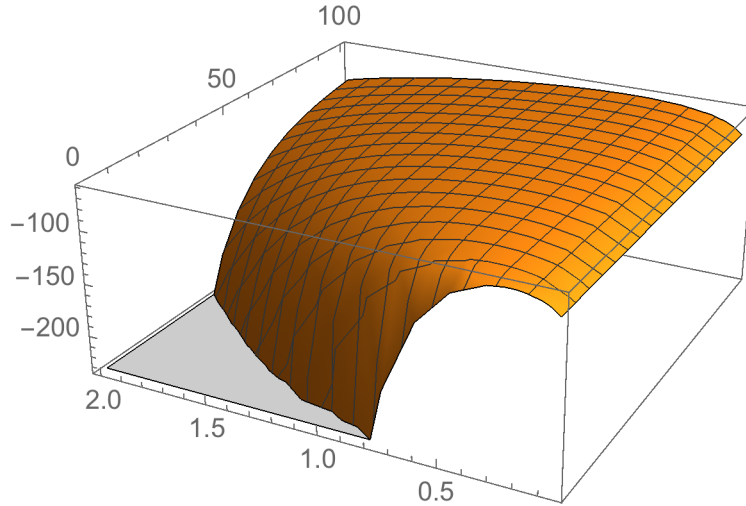


Figure 1: Weibull likelihood.

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 $\bar{\theta}_n$ of θ is given by $\bar{\theta}_n = \bar{X}_n - 1$. Now

$$\begin{aligned} E_{\theta}(X^2) &= (1 - \theta) \int_0^{\infty} x^2 e^{-x} dx + \theta \int_0^{\infty} x^3 e^{-x} dx \\ &= (1 - \theta)\Gamma(3) + \theta\Gamma(4) \\ &= (1 - \theta)2 + \theta3! = 2(1 - \theta) + 6\theta \\ &= 2 + 4\theta. \end{aligned}$$

Thus

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

Hence it follows by the CLT that

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

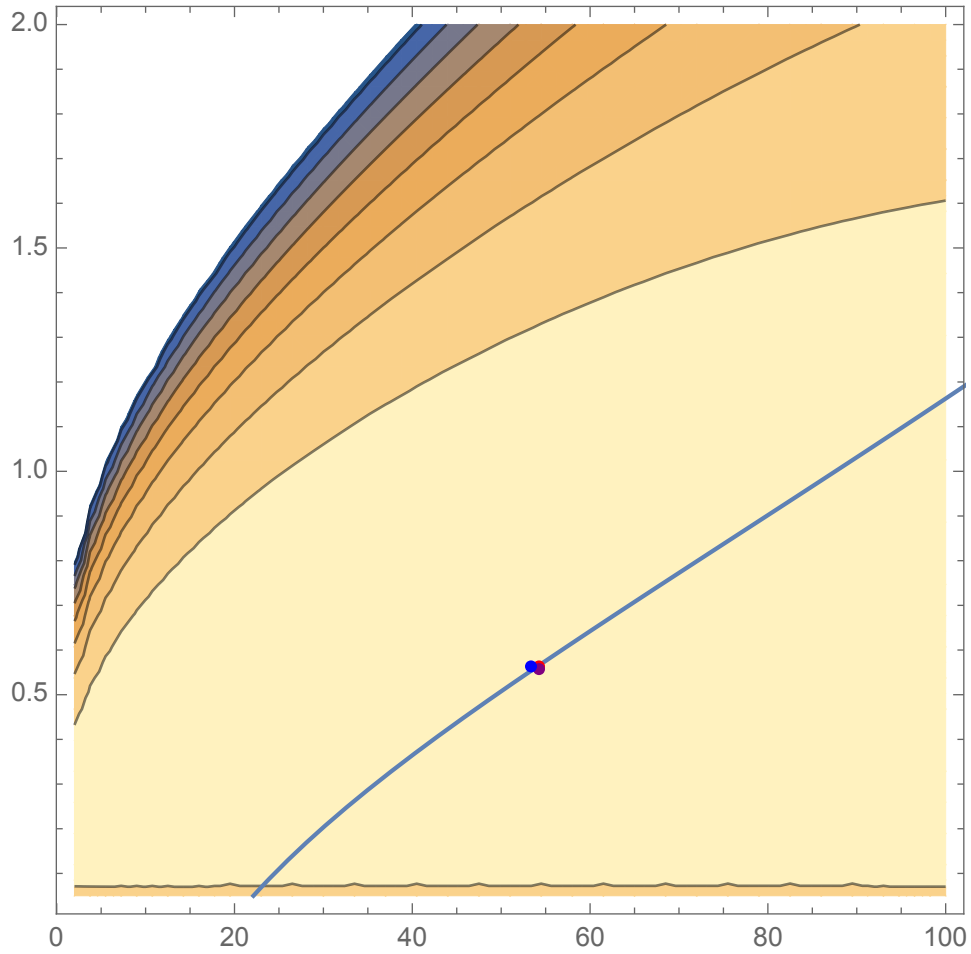


Figure 2: Contour plot Weibull likelihood/

Now

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

and hence

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

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

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

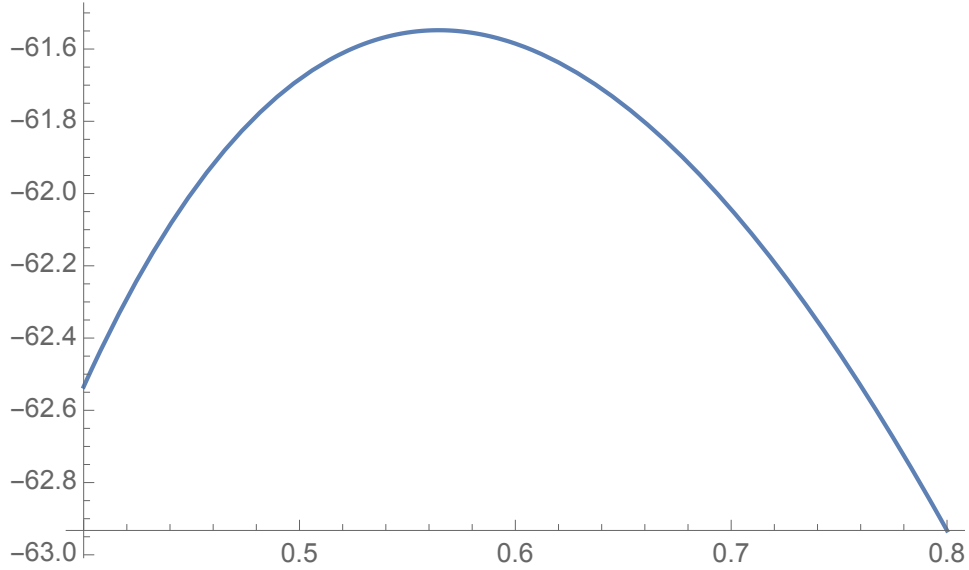


Figure 3: Profile plot Weibull likelihood

where

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

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 estimator $\bar{\theta}_n$. One possibility for $\bar{\theta}_n$ is a method of moments estimator. We calculate

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

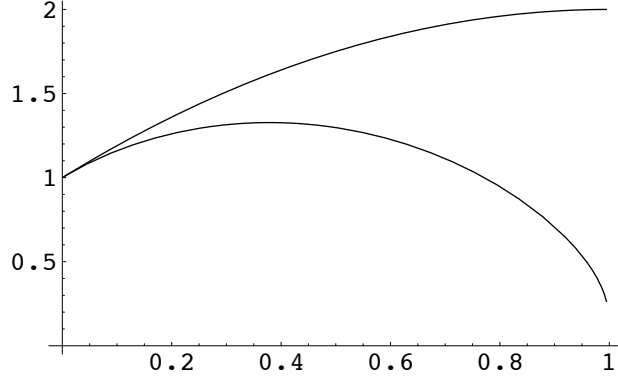


Figure 4: $1/I(\theta)$ and $1 + 2\theta - \theta^2$

For $\eta \neq 2$ this yields

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

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) $\bar{\theta}_n$ for this problem.

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

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

Then

$$\check{\theta}_n = \bar{\theta}_n + \hat{I}_n^{-1} \frac{1}{n} \dot{\mathbf{i}}_n(\bar{\theta}_n | \underline{X})$$

where

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

and

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

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 $\text{Poisson}(\theta_i)$, $i = 1, \dots, k$. 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 $\underline{X}_j = (X_{1,j}, \dots, X_{k,j})$ for $j = 1, \dots, n$ where the components $X_{i,j}$ of \underline{X}_j are independent $\text{Poisson}(\theta_i)$ random variables. Thus the \underline{X}_j random vectors are i.i.d. with (joint) probability mass function given by

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

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, \dots, n_k in the separate Poisson populations are possibly different.)

Now we calculate

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

and

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

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

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

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

$$\hat{\underline{\theta}} = (\bar{X}_1, \dots, \bar{X}_k)$$

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

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

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

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

In this case Theorem 4.1.2 applies directly and we have

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

and

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

Moreover, under the null hypothesis it is easily seen that

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

and, furthermore, that

$$\sqrt{n} \begin{pmatrix} \bar{X}_1 - \bar{X} \\ \vdots \\ \bar{X}_k - \bar{X} \end{pmatrix} \rightarrow_d \begin{pmatrix} D_1 - \bar{D} \\ \vdots \\ D_k - \bar{D} \end{pmatrix} \sim N_k(0, \theta^0 (I - k^{-1} \underline{1} \underline{1}^T)), \quad (0.7)$$

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

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

it follows that

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

and hence

$$l_n(\hat{\theta}_1, \dots, \hat{\theta}_k) = n \sum_{i=1}^k \{\bar{X}_i \log \bar{X}_i - \bar{X}_i\},$$

while

$$l_n(\hat{\theta}_1^0, \dots, \hat{\theta}_k^0) = n \sum_{i=1}^k \{\bar{X} \log \bar{X} - \bar{X}\} = n \{k\bar{X} \log \bar{X} - k\bar{X}\}.$$

Hence the log-likelihood ratio statistic is given by

$$\begin{aligned} 2 \log \lambda_n &= 2\{l_n((\hat{\theta}_1, \dots, \hat{\theta}_k) - l_n(\hat{\theta}_1^0, \dots, \hat{\theta}_k^0))\} \\ &= 2n \left\{ \sum_{i=1}^k \bar{X}_i \log \bar{X}_i - k\bar{X} \log \bar{X} \right\}. \end{aligned}$$

When the null hypothesis holds, our considerations in the i.i.d. case lead to the conclusion that $2 \log \lambda_n \rightarrow_d \chi_{k-1}^2$. It is instructive to consider the natural Wald statistic W_n in this problem starting from (0.7) and see that we also have $W_n \rightarrow_d \chi_{k-1}^2$ under the null hypothesis. If $\underline{\theta}_n = (\theta_{n,1}, \dots, \theta_{n,k}) = (\theta^0 + n^{-1/2}t_1, \dots, \theta^0 + n^{-1/2}t_k)$ where $t_i \neq t_{i'}$ for some $i \neq i'$, then I claim that $2 \log \lambda_n \rightarrow_d \chi_{k-1}^2(\delta)$ where $\delta = \sum_{i=1}^k (t_i - \bar{t})^2 / \theta^0$ 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 σ -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, \dots, 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, σ^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 $n \rightarrow \infty$.

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

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

(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

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

exactly for every n .