Statistics 581, Problem Set 10 Solutions

Wellner; 12/10/2010

1. (a) Ferguson, ACLST, page 139, problem 3: Let $X_1, \ldots, X_n$ be a sample from a mixture of gamma distributions:

$$f(x|\theta) = \{(1-\theta)e^{-x} + \theta x e^{-x}\} 1_{(0,\infty)}(x)$$

where $0 < \theta < 1$. What is the estimate of θ given by the method of moments? What is its asymptotic distribution? Show how to improve this estimate by one iteration of Newton's method applied to the likelihood equation.

(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

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

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.

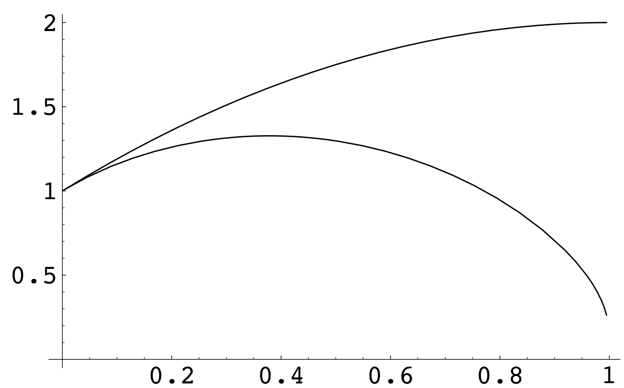

Figure 1: $1/I(\theta)$ and $1 + 2\theta - \theta^2$

(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 a method of moments estimator. We calculate

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

For $\eta \neq 2$ this yields

$$\frac{E(X^2) - 2}{E(X) - 1} = \frac{6/\eta^2 - 2}{2/\eta - 1} = \frac{6 - 2\eta^2}{2\eta - \eta^2} . \tag{0.1}$$

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

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

2. Ferguson, ACLST, page 118, problem 3. (See also Example 4.3.7, page 21, Chapter 4 notes.)

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

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

3. Lehmann and Casella, problem 6.8, page 509. If $p_\theta(x,y)$ is the bivariate normal density (with known means μ and ν equal to zero without loss of generality), the information matrix $I(\theta)$ for $\theta = (\sigma^2, \tau^2, \rho)$ is given by

$$(1-\rho^2)I(\theta) = \begin{pmatrix} \frac{2-\rho^2}{4\sigma^4} & \frac{-\rho^2}{4\sigma^2\tau^2} & \frac{-\rho}{2\sigma^2} \\ \frac{-\rho^2}{4\sigma^2\tau^2} & \frac{2-\rho^2}{4\tau^4} & \frac{-\rho}{2\tau^2} \\ \frac{-\rho}{2\sigma^2} & \frac{-\rho}{2\tau^2} & \frac{1+\rho^2}{1-\rho^2} \end{pmatrix} \tag{0.2}$$

and

$$I^{-1}(\theta) = \begin{pmatrix} 2\sigma^4 & 2\rho^2\sigma^2\tau^2 & \rho(1-\rho^2)\sigma^2 \\ 2\rho^2\sigma^2\tau^2 & 2\tau^4 & \rho(1-\rho^2)\tau^2 \\ \rho(1-\rho^2)\sigma^2 & \rho(1-\rho^2)\tau^2 & (1-\rho^2)^2 \end{pmatrix}.$$

Solution: The bivariate normal density $p_\theta(x,y)$ is given by

$$p_\theta(x,y) = \frac{1}{2\pi\sqrt{\sigma^2\tau^2(1-\rho^2)}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma^2} - 2\rho\frac{x}{\sigma}\frac{y}{\tau} + \frac{y^2}{\tau^2}\right)\right)$$

so

$$\begin{aligned} \log p_\theta(x,y) &= -\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\log(\tau^2) - \frac{1}{2}\log(1-\rho^2) \\ &\quad - \frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma^2} - 2\rho\frac{x}{\sigma}\frac{y}{\tau} + \frac{y^2}{\tau^2}\right). \end{aligned}$$

Thus we compute

$$\begin{aligned} \dot{l}_{\sigma^2}(x,y) &= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^2(1-\rho^2)}\left(\frac{x^2}{\sigma^2} - \rho\frac{x}{\sigma}\frac{y}{\tau}\right), \\ \dot{l}_{\tau^2}(x,y) &= -\frac{1}{2\tau^2} + \frac{1}{2\tau^2(1-\rho^2)}\left(\frac{y^2}{\tau^2} - \rho\frac{x}{\sigma}\frac{y}{\tau}\right), \\ \dot{l}_\rho(x,y) &= \frac{\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2}\left(\frac{x^2}{\sigma^2} - 2\rho\frac{x}{\sigma}\frac{y}{\tau} + \frac{y^2}{\tau^2}\right) + \frac{1}{1-\rho^2}\frac{x}{\sigma}\frac{y}{\sigma} \\ &= \frac{1}{1-\rho^2}\left\{\rho - \frac{\rho}{1-\rho^2}\left(\frac{x^2}{\sigma^2} + \frac{y^2}{\tau^2}\right) + \left(\frac{2\rho^2}{1-\rho^2} + 1\right)\frac{x}{\sigma}\frac{y}{\tau}\right\} \\ &= \frac{1}{(1-\rho^2)^2}\left\{\rho(1-\rho^2) - \rho\left(\frac{x^2}{\sigma^2} + \frac{y^2}{\tau^2}\right) + (1+\rho^2)\frac{x}{\sigma}\frac{y}{\tau}\right\}. \end{aligned}$$

Note that

$$E_\theta \dot{l}_{\sigma^2}(X,Y) = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^2(1-\rho^2)}\left(1-\rho^2\right) = 0,$$
$$E_\theta \dot{l}_{\tau^2}(X,Y) = -\frac{1}{2\tau^2} + \frac{1}{2\tau^2(1-\rho^2)}\left(1-\rho^2\right) = 0,$$
$$E_\theta \dot{l}_{\rho}(X,Y) = \frac{\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2}(2-2\rho^2) + \frac{\rho}{1-\rho^2} = 0.$$

Furthermore,

$$\begin{aligned}
\ddot{l}_{\sigma^2,\sigma^2}(x,y) &= \frac{1}{2\sigma^4} - \frac{1}{2\sigma^4(1-\rho^2)}\left(\frac{x^2}{\sigma^2} - \rho\frac{x}{\sigma}\frac{y}{\tau}\right) + \frac{1}{2\sigma^2(1-\rho^2)}\left(-\frac{x^2}{\sigma^4} + \frac{\rho}{2}\frac{x}{\sigma^3}\frac{y}{\tau}\right) \\
&= \frac{1}{2\sigma^4}\left\{1 - \frac{1}{1-\rho^2}\left(\frac{x^2}{\sigma^2} - \rho\frac{x}{\sigma}\frac{y}{\tau}\right) - \frac{1}{1-\rho^2}\left(\frac{x^2}{\sigma^2} - \frac{\rho}{2}\frac{x}{\sigma}\frac{y}{\tau}\right)\right\}, \\
\ddot{l}_{\tau^2,\tau^2}(x,y) &= \frac{1}{2\tau^4}\left\{1 - \frac{1}{1-\rho^2}\left(\frac{y^2}{\tau^2} - \rho\frac{x}{\sigma}\frac{y}{\tau}\right) - \frac{1}{1-\rho^2}\left(\frac{y^2}{\tau^2} - \frac{\rho}{2}\frac{x}{\sigma}\frac{y}{\tau}\right)\right\}, \\
\ddot{l}_{\rho,\rho}(x,y) &= \frac{1}{1-\rho^2}\left(1 - \left(\frac{x^2}{\sigma^2} + \frac{y^2}{\tau^2}\right) + 4\rho\frac{x}{\sigma}\frac{y}{\tau}\right) \\
&\quad + \frac{2\rho}{(1-\rho^2)^2}\left\{\rho - \rho\left(\frac{x^2}{\sigma^2} + \frac{y^2}{\tau^2}\right) + \frac{1+\rho^2}{1-\rho^2}\frac{x}{\sigma}\frac{y}{\tau}\right\}, \\
\ddot{l}_{\rho,\sigma^2}(x,y) &= \frac{1}{\sigma^2(1-\rho^2)^2}\left\{\rho\frac{x^2}{\sigma^2} - \frac{1+\rho^2}{2}\frac{x}{\sigma}\frac{y}{\tau}\right\}, \\
\ddot{l}_{\rho,\tau^2}(x,y) &= \frac{1}{\tau^2(1-\rho^2)^2}\left\{\rho\frac{y^2}{\tau^2} - \frac{1+\rho^2}{2}\frac{x}{\sigma}\frac{y}{\tau}\right\}, \\
\ddot{l}_{\sigma^2,\tau^2}(x,y) &= \frac{\rho}{4\sigma^2\tau^2(1-\rho^2)}\frac{x}{\sigma}\frac{y}{\tau}.
\end{aligned}$$

Thus we compute:

$$\begin{aligned}
-E_\theta \ddot{l}_{\sigma^2,\sigma^2}(X,Y) &= \frac{1}{2\sigma^4(1-\rho^2)}\left\{\left(1 - \frac{1}{2}\rho^2\right) + (1-\rho^2) - (1-\rho^2)\right\} = \frac{2-\rho^2}{4\sigma^4(1-\rho^2)}, \\
-E_\theta \ddot{l}_{\tau^2,\tau^2}(X,Y) &= \frac{2-\rho^2}{4\tau^4(1-\rho^2)}, \\
-E_\theta \ddot{l}_{\rho,\rho}(X,Y) &= \frac{2}{(1-\rho^2)^2}\left\{\rho(1-\rho^2) - 2\rho + \rho(1+\rho^2)\right\} \\
&\quad - \frac{1}{(1-\rho^2)^2}\left\{1 - 3\rho^2 - 2 + 2\rho^2\right\} \\
&= \frac{1+\rho^2}{(1-\rho^2)^2}, \\
-E_\theta(\ddot{l}_{\rho,\sigma^2}(X,Y) &= \frac{-\rho}{2\sigma^2(1-\rho^2)},
\end{aligned}$$

$$-E_\theta(\ddot{l}_{\rho,\tau^2}(X,Y) \;=\; \frac{-\rho}{2\tau^2(1-\rho^2)},$$
$$-E_\theta(\ddot{l}_{\sigma^2,\tau^2}(X,Y) \;=\; \frac{-\rho^2}{4\sigma^2\tau^2},$$

and hence the information matrix $I(\theta)$ is as given in (0.2).

To invert this information matrix, it is instructive to proceed via block-inversion. Let $\theta = (\theta_1, \theta_2)$ where $\theta_1 \equiv (\sigma^2, \tau^2)$, $\theta_2 = \rho$. Then we first calculate I_{11}^{-1}, the information bound for estimation of $\theta_1 = (\sigma^2, \tau^2)$ when $\theta_2 = \rho$ is known. This gives

$$\begin{aligned}
I_{11}^{-1} &= (1-\rho^2)\begin{pmatrix} \frac{2-\rho^2}{4\tau^4} & \frac{\rho^2}{4\sigma^2\tau^2} \\ \frac{\rho^2}{4\sigma^2\tau^2} & \frac{2-\rho^2}{4\sigma^4} \end{pmatrix} \frac{1}{\frac{(2-\rho^2)^2}{16\sigma^4\tau^4} - \frac{\rho^4}{16\sigma^4\tau^4}} \\
&= \frac{16\sigma^4\tau^4}{4}\begin{pmatrix} \frac{2-\rho^2}{4\tau^4} & \frac{\rho^2}{4\sigma^2\tau^2} \\ \frac{\rho^2}{4\sigma^2\tau^2} & \frac{2-\rho^2}{4\sigma^4} \end{pmatrix} \\
&= \begin{pmatrix} \sigma^4(2-\rho^2) & \rho^2\sigma^2\tau^2 \\ \rho^2\sigma^2\tau^2 & \tau^4(2-\rho^2) \end{pmatrix}.
\end{aligned}$$

Next we calculate $I_{11\cdot 2}$:

$$\begin{aligned}
I_{11\cdot 2} &= I_{11} - I_{12}I_{22}^{-1}I_{21} \\
&= \frac{1}{1-\rho^2}\left\{\begin{pmatrix} \frac{2-\rho^2}{4\sigma^4} & \frac{-\rho^2}{4\sigma^2\tau^2} \\ \frac{-\rho^2}{4\sigma^2\tau^2} & \frac{2-\rho^2}{4\tau^4} \end{pmatrix} - (1-\rho^2)\begin{pmatrix} \frac{-\rho}{2\sigma^2} \\ \frac{-\rho}{2\tau^2} \end{pmatrix} \frac{1}{1+\rho^2} \begin{pmatrix} \frac{-\rho}{2\sigma^2} & \frac{-\rho}{2\tau^2} \end{pmatrix}\right\} \\
&= \frac{1}{1-\rho^2}\left\{\begin{pmatrix} \frac{2-\rho^2}{4\sigma^4} & \frac{-\rho^2}{4\sigma^2\tau^2} \\ \frac{-\rho^2}{4\sigma^2\tau^2} & \frac{2-\rho^2}{4\tau^4} \end{pmatrix} - \frac{\rho^2(1-\rho^2)}{1+\rho^2}\begin{pmatrix} \frac{1}{4\sigma^4} & \frac{1}{4\sigma^2\tau^2} \\ \frac{1}{4\sigma^2\tau^2} & \frac{1}{4\tau^4} \end{pmatrix}\right\} \\
&= \frac{1}{2(1-\rho^2)(1+\rho^2)}\begin{pmatrix} \frac{1}{\sigma^4} & \frac{-\rho^2}{\sigma^2\tau^2} \\ \frac{-\rho^2}{\sigma^2\tau^2} & \frac{1}{\tau^4} \end{pmatrix}.
\end{aligned}$$

This yields the information bound for estimation of $\theta_1 = (\sigma^2, \tau^2)$ when $\theta_2 = \rho$ is unknown:

$$\begin{aligned}
I_{11\cdot 2}^{-1} &= 2(1-\rho^2)(1+\rho^2)\begin{pmatrix} \frac{1}{\tau^4} & \frac{\rho^2}{\sigma^2\tau^2} \\ \frac{\rho^2}{\sigma^2\tau^2} & \frac{1}{\sigma^4} \end{pmatrix} \frac{1}{\frac{1}{\sigma^4\tau^4} - \frac{\rho^4}{\sigma^4\tau^4}} \\
&= 2\begin{pmatrix} \sigma^4 & \rho^2\sigma^2\tau^2 \\ \rho^2\sigma^2\tau^2 & \tau^4 \end{pmatrix}.
\end{aligned}$$

Next we use I_{11}^{-1} to calculate $I_{22\cdot 1}$:

$$\begin{aligned}
I_{22\cdot 1} &= I_{22} - I_{21}I_{11}^{-1}I_{12} \\
&= \frac{1}{1-\rho^2}\left(\frac{1+\rho^2}{1-\rho^2} - \frac{1}{1-\rho^2}\begin{pmatrix} \frac{-\rho}{2\sigma^2} & \frac{-\rho}{2\tau^2} \end{pmatrix}\begin{pmatrix} \sigma^4(2-\rho^2) & \rho^2\sigma^2\tau^2 \\ \rho^2\sigma^2\tau^2 & \tau^4(2-\rho^2) \end{pmatrix}\begin{pmatrix} \frac{-\rho}{2\sigma^2} \\ \frac{-\rho}{2\tau^2} \end{pmatrix}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(1-\rho^2)^2}\left\{1+\rho^2-\frac{1}{4}(2\rho^2+2\rho^2)\right\} \\
&= \frac{1}{(1-\rho^2)^2}.
\end{aligned}
$$

Thus $I_{22\cdot 1}^{-1} = (1-\rho^2)^2$ as claimed. This completes the computation of the diagonal entries $I_{11\cdot 2}^{-1}$ and $I_{22\cdot 1}^{-1}$ of $I^{-1}(\theta)$. It remains to compute I^{12} or its transpose I^{21}. From our general formulas in chapter 3 we know that

$$
\begin{aligned}
I^{12} &= -I_{11.2}^{-1} I_{12} I_{22}^{-1} \\
&= -2\begin{pmatrix} \sigma^4 & \rho\sigma^2\tau^2 \\ \rho\sigma^2\tau^2 & \tau^4 \end{pmatrix} \frac{1}{1-\rho^2} \begin{pmatrix} \frac{-\rho}{2\sigma^2} \\ \frac{-\rho}{2\tau^2} \end{pmatrix} \frac{(1-\rho^2)^2}{1+\rho^2} \\
&= \begin{pmatrix} \sigma^4 & \rho\sigma^2\tau^2 \\ \rho\sigma^2\tau^2 & \tau^4 \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma^2} \\ \frac{1}{\tau^2} \end{pmatrix} \frac{\rho(1-\rho^2)}{1+\rho^2} \\
&= \rho(1-\rho^2)\begin{pmatrix} \sigma^2 \\ \tau^2 \end{pmatrix}
\end{aligned}
$$

as claimed.

4. Lehmann and Casella, problem 6.9, page 509: Suppose that (X_i, Y_i), $i = 1, \ldots, n$ are i.i.d. bivariate normal with $E(X_i) = E(Y_i) = 0$, $E(X_i^2) = E(Y_i^2) = 1$ and unknown correlation coefficient ρ. Let ρ_0 denote the true value of ρ.
(a) Show that the likelihood equation is a cubic for which the probability of a unique root tends to 1 as $n \to \infty$ [Hint: for a cubic equation $ax^3 + 3bx^2 + 3cx + d = 0$, let $G = a^2 d - 3abc + 2b^3$ and $H = ac - b^2$. Then the condition for a unique real root is $G^2 + 4H^3 > 0$.]
(b) Show that if $\hat{\rho}_n$ is a consistent solution of the likelihood equation, then it satisfies $\sqrt{n}(\hat{\rho}_n - \rho_0) \to_d N(0, (1-\rho_0^2)^2/(1+\rho_0^2))$.
(c) Show that $\delta_n \equiv n^{-1}\sum_{i=1}^n X_i Y_i$ is a consistent estimator of ρ and that $\sqrt{n}(\delta_n - \rho_0) \to_d N(0, 1+\rho_0^2)$. Hence δ_n is less efficient than the MLE.

Solution: (a) When the means and variances are known to be 0's and 1's respectively, the likelihood equation for estimation of ρ is

$$
\begin{aligned}
0 &= \sum_{i=1}^n \dot{l}_\rho(X_i, Y_i) \\
&= \sum_{i=1}^n (1-\rho^2)^{-1}\left\{\rho(1-\rho^2) - \rho(X_i^2 + Y_i^2) + (1+\rho^2)X_i Y_i\right\}, \qquad (0.3)
\end{aligned}
$$

or, equivalently,

$$
\rho(1-\rho^2) - \rho(\overline{X^2}_n + \overline{Y^2}_n) + (1+\rho^2)\overline{XY}_n = 0. \qquad (0.4)
$$

This is a cubic equation which can be rewritten as:

$$\Psi_n(\rho) \equiv \rho^3 - \overline{XY}_n\rho^2 + (\overline{X^2}_n + \overline{Y^2}_n - 1)\rho - \overline{XY}_n = 0.$$

Note that if ρ_0 is the true correlation, then $\overline{XY}_n \to_{a.s.} \rho_0$, $\overline{X^2}_n \to_{a.s.} 1$, and $\overline{Y^2} \to_{a.s.} 1$ and hence

$$\Psi_n(\rho) \to_{a.s.} \rho^3 - \rho_0\rho^2 + \rho - \rho_0 \equiv \Psi(\rho).$$

Note that $\Psi(\rho_0) = 0$. By the hint, a cubic equation $ax^3 + 3bx^2 + 3cx + d = 0$ if $G^2 + 4H^3 > 0$ where $G \equiv a^2d - 3abc + 2b^3$ and $H \equiv ac - b^2$. Thus we compute

$$\begin{aligned} G_n &= 1^2(-\overline{XY}_n - 3\cdot 1\cdot(-\frac{1}{3}\overline{XY}_n)(1/3)(\overline{X^2}_n + \overline{Y^2}_n - 1) + 2(-\frac{1}{3}\overline{XY}_n)^3 \\ &= -\overline{XY}_n + \frac{1}{3}\overline{XY}_n(\overline{X^2} + \overline{Y^2} - 1) - \frac{2}{27}\overline{XY}_n^3, \\ H_n &= 1\cdot\frac{1}{3}(\overline{X^2} + \overline{Y^2} - 1) - (\frac{-1}{3}\overline{XY}_n)^2 \\ &= \frac{1}{3}(\overline{X^2} + \overline{Y^2} - 1) - \frac{1}{9}\overline{XY}_n^2. \end{aligned}$$

Note that

$$\begin{aligned} &G_n \to_{a.s.} -\rho_0 + \frac{1}{3}\rho_0 - \frac{2}{27}\rho_0^3 = -\frac{2}{3}\rho_0(1 + \frac{1}{9}\rho_0^2) \equiv G_0, \\ &H_n \to_{a.s.} \frac{1}{3} - \frac{1}{9}\rho_0^2 \equiv H_0, \end{aligned}$$

and hence

$$\begin{aligned} G_n^2 + 4H_n^3 &\to_{a.s.} G_0^2 + 4H_0^3 = \frac{4}{9}\rho_0^2(1+\rho_0^2)^2 + \frac{4}{27}(1 - \frac{1}{3}\rho_0^2)^3 \\ &= \frac{4}{27}(1+\rho_0^2)^2 \ge \frac{4}{27} > 0 \end{aligned}$$

after a bit of algebra. Thus $\Psi(\rho) = 0$ alway has a unique real root, and, with probability converging to one, $\Psi_n(\rho) = 0$ also has a unique real root. Here is a plot of $G_0^2 + 4H_0^3$ as a function of ρ_0:

(b) If $\hat{\rho}_n$ is a consistent solution of the likelihood equation (0.3) or (0.4), then by Theorem 4.1.2 is satisfies (since conditions A0-A4 hold)

$$\sqrt{n}(\hat{\rho}_n - \rho_0) \to_d D \sim N(0, I(\rho_0)^{-1}) = N(0, (1-\rho_0^2)^2/(1+\rho_0^2))$$

where we have used the information matrix computed in problem 3(a) above.
(c) Now $\delta_n \equiv \overline{XY}_n \to_p \rho_0$ and, by the CLT,

$$\sqrt{n}(\delta_n - \rho_0) \to_d N(0, Var(XY)) = N(0, 1 + \rho_0^2)$$

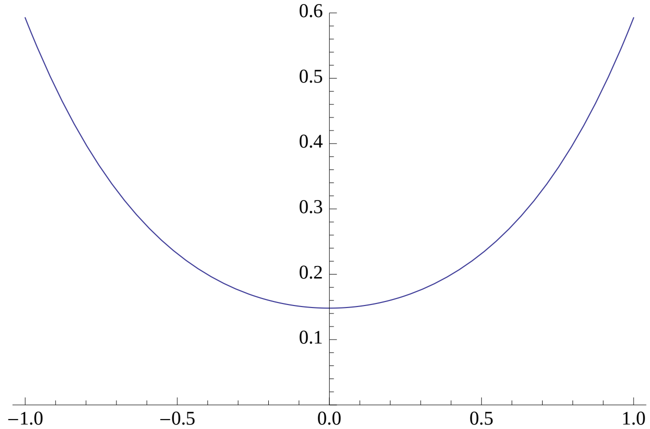

Figure 2: $G_0^2 + 4H_0^3$ as a function of ρ_0

since

$$\begin{aligned} Var(XY) &= EVar(XY|X) + Var(E(XY|X)) = E\{X^2 Var(Y|X)\} + Var\{XE(Y|X)\} \\ &= E\{X^2(1-\rho_0^2)\} + Var(X^2\rho_0) = 1 - \rho_0^2 + 2\rho_0^2 \\ &= 1 + \rho_0^2 \geq \frac{(1-\rho_0^2)^2}{1+\rho_0^2} \end{aligned}$$

with equality if and only if $\rho_0 = 0$. Here is a plot of the ratio: $I^{-1}(\rho_0)/Var_{\rho_0}(XY)$:

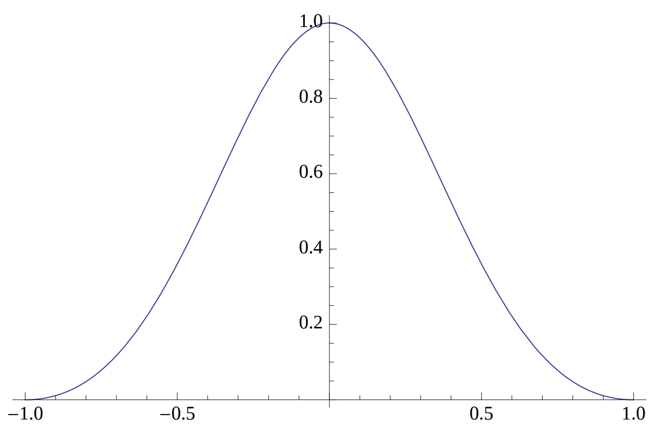

Figure 3: $I^{-1}(\rho_0)/Var_{\rho_0}(XY)$ as a function of ρ_0

This example is treated along with other problem involving multiple roots in: Small, C.G., Wang, J., and Yang, Z. (2000). Estimating multiple root problems in estimation. *Statistical Science* **15**, 313-341. A somewhat different approach to this problem and generalizations thereof is pursued in: Sampson, A. R. (1978). Simple BAN estimators of correlations for certain multivariate normal models with known variances. *J. Amer. Statist. Assoc.* **73**, 859-862.

5. 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 $H_0 : \gamma_2 = 0, \gamma_1 = \text{anything}$.
(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

$$\begin{aligned} 2(l_n(\widehat{\theta}) - l_n(\widehat{\theta}^0)) &= 2n \left\{ 2 \log \left(\frac{\overline{X} + c\overline{Y}}{2} \right) - \log(\overline{X}) - \log(c\overline{Y}) \right\} \\ &\to_d \chi^2_1 \end{aligned}$$

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 $H_0 : \gamma_2 = 0, \gamma_1 = \text{anything}$. 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

$$2 \log \lambda_n \to_d \chi^2_1(\delta)$$

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.