Statistics 582, Problem Set 1 Solutions

Wellner; 1/14/2015

1. Lehmann and Casella, TPE, Problem 4.9, page 504.

Solution: (a) The density of a bivariate normal random vector (X, Y) with $\mu_1 = \mu_2 = 0$, variances $\sigma_1^2 \equiv \sigma^2$, $\sigma_2^2 \equiv \tau^2$, and correlation ρ (so that $\theta = (\sigma, \tau, \rho)$) is given by

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

and the marginal densities of X and Y respectively are given by

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

Thus the contributions to the log-likelihood are of the form

$$\log p_\theta(x, y) = -\log\sigma - \log\tau - \frac{1}{2}\log(1-\rho^2) - \frac{\frac{x^2}{\sigma^2} - \frac{2\rho xy}{\sigma\tau} + \frac{y^2}{\tau^2}}{2(1-\rho^2)},$$

and $-\log\sigma - x^2/(2\sigma^2)$, $-\log\tau - x^2/(2\tau^2)$, respectively. Thus for the given data the log-likelihood is given by

$$\begin{aligned}
l_n(\theta) &= -4\log\sigma - 4\log\tau - 2\log(1-\rho^2) \\
&\quad -\frac{1}{2(1-\rho^2)} \left\{ \frac{1}{\sigma^2} - \frac{2\rho}{\sigma\tau} + \frac{1}{\tau^2} + \frac{1}{\sigma^2} + \frac{2\rho}{\sigma\tau} + \frac{1}{\tau^2} \right. \\
&\quad \left. + \frac{1}{\sigma^2} + \frac{2\rho}{\sigma\tau} + \frac{1}{\tau^2} + \frac{1}{\sigma^2} - \frac{2\rho}{\sigma\tau} + \frac{1}{\tau^2} \right\} \\
&\quad - 4\log\sigma - 4\log\tau - \frac{8}{\sigma^2} - \frac{8}{\tau^2} \\
&= -8\log\sigma - 8\log\tau - 2\log(1-\rho^2) - \frac{1}{1-\rho^2}\left\{ \frac{2}{\sigma^2} + \frac{2}{\tau^2} \right\} - \frac{8}{\sigma^2} - \frac{8}{\tau^2}.
\end{aligned}$$

We compute

$$\begin{aligned}
\frac{\partial}{\partial\sigma} l_n(\theta) &= -\frac{8}{\sigma} + \frac{4}{(1-\rho^2)\sigma^3} + \frac{16}{\sigma^3} = -\frac{1}{\sigma}\left\{ 8 - \frac{4}{(1-\rho^2)\sigma^2} - \frac{16}{\sigma^2} \right\}, \\
\frac{\partial}{\partial\tau} l_n(\theta) &= -\frac{8}{\tau} + \frac{4}{(1-\rho^2)\tau^3} + \frac{16}{\tau^3} = -\frac{1}{\tau}\left\{ 8 - \frac{4}{(1-\rho^2)\tau^2} - \frac{16}{\tau^2} \right\}, \\
\frac{\partial}{\partial\rho} l_n(\theta) &= \frac{4\rho}{1-\rho^2} - \frac{2\rho}{(1-\rho^2)^2}\left\{ \frac{2}{\sigma^2} + \frac{2}{\tau^2} \right\} = \frac{2\rho}{(1-\rho^2)}\left\{ 2 - \frac{1}{1-\rho^2}\left\{ \frac{2}{\sigma^2} + \frac{2}{\tau^2} \right\} \right\}.
\end{aligned}$$

It is easily seen that these scores are zero at both $\theta = (\sqrt{8/3}, \sqrt{8/3}, \pm 1/2)$ and at $\theta = (\sqrt{5/2}, \sqrt{5/2}, 0)$. Furthermore $l_n(\sqrt{8/3}, \sqrt{8/3}, \pm 1/2) = -15.2713...$ while $l_n(\sqrt{5/2}, \sqrt{5/2}, 0) = -15.3303....$ Thus it seems that the first pair of points, $\theta = (\sqrt{8/3}, \sqrt{8/3}, \pm 1/2)$, yield a (non-unique) maximum, and that $\theta = (\sqrt{5/2}, \sqrt{5/2}, 0)$ corresponds to a saddle point. The plot below shows the (exponential of the) likelihood function $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$.

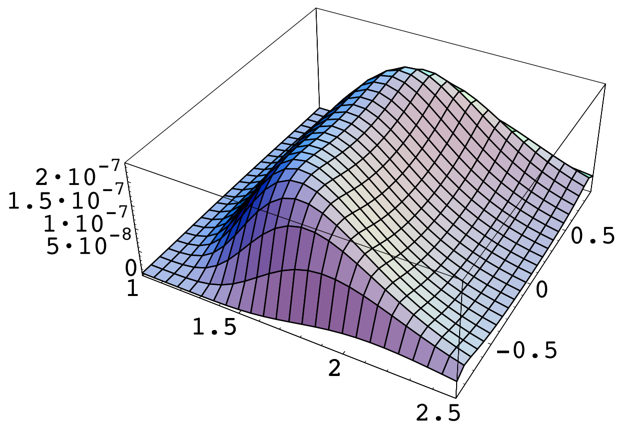

Figure 1: Plot of $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$.

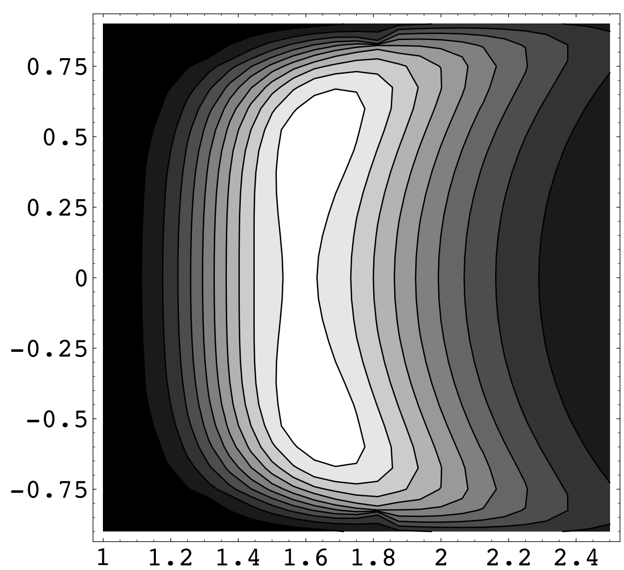

Figure 2: Contour plot of $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$.

(b) A natural EM - algorithm for estimation of θ proceeds as follows. Let the complete data X be

$$X = ((X_1, Y_1), \dots (X_n, Y_n)) \qquad \text{with} \quad n = 12,$$

and let the incomplete data be

$$Y = ((X_1, Y_1), \ldots, (X_4, Y_4), X_5, \ldots, X_8, Y_9, \ldots, Y_{12}).$$

Then, since

$$\begin{array}{lll} E(Y_j|X_j) = \rho\tau X_j/\sigma, & E(Y_j^2|X_j) = \tau^2(1-\rho^2) + (\rho\tau X_j/\sigma)^2, & j = 5, \ldots, 8, \\ E(X_j|Y_j) = \rho\sigma Y_j/\tau, & E(X_j^2|Y_j) = \sigma^2(1-\rho^2) + (\rho\sigma Y_j/\tau)^2, & j = 9, \ldots, 12, \end{array}$$

the conditional expectation of the complete data log-likelihood given Y is given by

$$\begin{aligned} & E\{\log p_\theta(X)|Y\} \\ &= -12\log\{\sigma\tau(1-\rho^2)^{1/2}\} \\ &\qquad - \frac{1}{2(1-\rho^2)}\left\{\frac{E(\sum_1^{12} X_i^2|Y)}{\sigma^2} - \frac{2\rho E(\sum_1^{12} X_iY_i|Y)}{\sigma\tau} + \frac{E(\sum_1^{12} Y_i^2|Y)}{\tau^2}\right\} \\ &= -12\log\{\sigma\tau(1-\rho^2)^{1/2}\} - \frac{1}{2(1-\rho^2)}\left\{\frac{\hat{T}_{1,1}(Y)}{\sigma^2} - \frac{2\rho\hat{T}_{1,2}(Y)}{\sigma\tau} + \frac{\hat{T}_{2,2}(Y)}{\tau^2}\right\} \end{aligned}$$

where

$$\begin{aligned} \hat{T}_{1,1}(Y) &\equiv \hat{T}_{1,1}(Y,\theta) \equiv E\left(\sum_1^{12} X_i^2|Y\right) \\ &= \sum_{i=1}^{8} X_i^2 + \sum_{i=9}^{12} E(X_i^2|Y_i) \\ &= \sum_{i=1}^{8} X_i^2 + \sum_{i=9}^{12}\{\sigma^2(1-\rho^2) + (\rho\sigma Y_i/\tau)^2\}, \\ \hat{T}_{1,2}(Y) &\equiv \hat{T}_{1,2}(Y,\theta) \equiv E\left(\sum_1^{12} X_iY_i|Y\right) \\ &= \sum_{i=1}^{4} X_iY_i + \sum_{i=5}^{8} X_iE(Y_i|X_i) + \sum_{i=9}^{12} Y_iE(X_i|Y_i) \\ &= \sum_{i=1}^{4} X_iY_i + \sum_{i=5}^{8} X_i(\rho\tau X_i/\sigma) + \sum_{i=9}^{12} Y_i(\rho\sigma Y_i/\tau), \\ \hat{T}_{2,2}(Y) &\equiv \hat{T}_{2,2}(Y,\theta) \equiv E(\sum_1^{12} Y_i^2|Y) \\ &= \sum_{i=1}^{4} Y_i^2 + \sum_{i=5}^{8} E(Y_i^2|X_i) + \sum_{i=9}^{12} Y_i^2 \\ &= \sum_{i=1}^{4} Y_i^2 + \sum_{i=5}^{8}\{\tau^2(1-\rho^2) + (\rho\tau X_i/\sigma)^2\} + \sum_{i=9}^{12} Y_i^2. \end{aligned}$$

Furthermore, the MLE's $\hat{\theta} \equiv \hat{\theta}(X) = (\hat{\sigma}, \hat{\tau}, \hat{\rho})$ of $\theta = (\sigma, \tau, \rho)$ for the complete data are given by

$$\hat{\sigma}^2 = n^{-1}T_{1,1}(X) \equiv n^{-1}\sum_{i=1}^{n} X_i^2,$$

$$\hat{\tau}^2 = n^{-1}T_{2,2}(X) \equiv n^{-1}\sum_{i=1}^{n} Y_i^2,$$

$$\hat{\rho} = n^{-1}T_{1,2}(X)/(\hat{\sigma}\hat{\tau}) \equiv n^{-1}\sum_{i=1}^{n} X_iY_i/(\hat{\sigma}\hat{\tau}).$$

We find that the $E-$step of an EM - algorithm is given by

$$\hat{T}^{(m)} \equiv (\hat{T}_{1,1}(Y,\hat{\theta}^{(m)}), \hat{T}_{1,2}(Y,\hat{\theta}^{(m)}), \hat{T}_{2,2}(Y,\hat{\theta}^{(m)})) \equiv (\hat{T}_{1,1}^{(m)}, \hat{T}_{1,2}^{(m)}, \hat{T}_{2,2}^{(m)}).$$

Here $\hat{\theta}^{(0)} = (\hat{\sigma}^{(0)}, \hat{\tau}^{(0)}, \hat{\rho}^{(0)})$ is an initial point to start the algorithm, and, for $m \geq 0$,

$$\hat{\theta}^{(m+1)} = \hat{\theta}(\hat{T}^{(m)}) \equiv \left(n^{-1}\hat{T}_{1,1}^{(m)}, n^{-1}\hat{T}_{2,2}^{(m)}, n^{-1}\hat{T}_{1,2}^{(m)}/(\hat{\sigma}^{(m)}\hat{\tau}^{(m)})\right)$$

gives the M-step.
Note that when $\hat{\rho}^{(0)} = 0$ we have $\hat{T}_{1,2}^{(m)} = \sum_{i=1}^{4} X_iY_i = 0$ for all $m \geq 1$, and hence $\hat{\rho}^{(m)} = 0$ for all $m \geq 0$.
(c) To show that if an EM sequence starts with ρ bounded away from zero, it converges to one of the two maximizing points $\hat{\theta}_{\pm}^{(\infty)} \equiv (\sqrt{8/3}, \sqrt{8/3}, \pm 1/2)$, note that if we start with $\hat{\rho}^{(0)} > 0$, then the sequence $\hat{\rho}^{(m)}$ stays positive for all m. This follows because

$$\hat{T}_{1,2}^{(m)} = 0 + 16\hat{\rho}^{(m)}\frac{\hat{\tau}^{(m)}}{\hat{\sigma}^{(m)}} + 16\hat{\rho}^{(m)}\frac{\hat{\sigma}^{(m)}}{\hat{\tau}^{(m)}} > 0.$$

Furthermore, if we start at $\hat{\theta}^{(0)}$ with $\hat{\sigma}^{(0)} = \hat{\tau}^{(0)}$, then by symmetry of the data, the whole sequence $\hat{\theta}^{(m)}$ satisfies $\hat{\sigma}^{(m)} = \hat{\tau}^{(m)}$. Thus

$$\begin{aligned}\hat{T}_{1,1}^{(m)} &= 20 + 4(\hat{\sigma}^{(m)})^2(1-(\hat{\rho}^{(m)})^2) + 16(\hat{\rho}^{(m)})^2 \\ &= 20 + 4(\hat{\tau}^{(m)})^2(1-(\hat{\rho}^{(m)})^2) + 16(\hat{\rho}^{(m)})^2 = \hat{T}_{2,2}^{(m)}\end{aligned}$$

and

$$\hat{T}_{1,2} = 16\hat{\rho}^{(m)} + 16\hat{\rho}^{(m)} = 32\hat{\rho}^{(m)}.$$

Since

$$\begin{aligned}\hat{\rho}^{(m+1)} &= \frac{32\hat{\rho}^{(m)}/12}{\hat{\sigma}^{(m)}\hat{\tau}^{(m)}} = \frac{32\hat{\rho}^{(m)}/12}{(\hat{\sigma}^{(m)})^2} \\ (\hat{\sigma}^{(m+1)})^2 &= \frac{20 + 4(\hat{\sigma}^{(m)})^2(1-(\hat{\rho}^{(m)})^2) + 16(\hat{\rho}^{(m)})^2}{12},\end{aligned}$$

it follows that any limiting point $(\sigma_\infty, \tau_\infty, \rho_\infty)$ must satisfy

$$\begin{aligned}\rho_\infty &= \frac{32}{12}\frac{\rho_\infty}{\sigma_\infty^2}, \qquad \text{and} \\ \sigma_\infty^2 &= \frac{20}{12} + \frac{1}{3}\sigma_\infty^2(1-\rho_\infty^2) + \frac{4}{3}\rho_\infty^2.\end{aligned}$$

The first of these implies that $\sigma_\infty^2 = 8/3$, and plugging this into the second relation we find that $\rho_\infty^2 = 1/4$, or $\rho_\infty = \pm 1/2$. The resulting two points $\theta_{\pm}^{(\infty)} = (\sqrt{8/3}, \sqrt{8/3}, \pm 1/2)$ are exactly the points of maximum of the incomplete data log-likelihood. This argument extends to the case in which $\hat{\sigma}^{(0)} \neq \hat{\tau}^{(0)}$.

It is straightforward to implement the algorithm in Mathematica or R, and numerical experimentation confirms these conclusions.

2. Compare the explanation of the EM algorithm in Lehmann and Casella TPE, pages 458-459 with the explanation given in Groeneboom's notes, pages 1 - 3 and 10 - 12. Correct the expressions given in (4.24) of TPE, page 459.

Solution: In (4.24) on page 459 of Lehmann and Casella, the expression for $Q(\theta|\theta_0, z)$ should be

$$Q(\theta|\theta_0, \mathbf{y}) = \int \log\{L(\theta|\mathbf{y}, \mathbf{z})\} k(\mathbf{z}|\theta_0, \mathbf{y}) d\mathbf{z},$$

and the expression for $H(\theta|\theta_0, \mathbf{y})$ should be

$$H(\theta|\theta_0, \mathbf{y}) = \int k(\mathbf{z}|\theta_0, \mathbf{y}) \log\left(k(\mathbf{z}|\theta_0, \mathbf{y})\right) d\mathbf{z}.$$

the difference $Q(\theta|\theta_0, \mathbf{y}) - H(\theta|\theta_0, \mathbf{y})$ corresponds to the two terms in (1.20) on page 11 of Groeneboom's notes.

3. Lehmann and Casella, TPE, Problem 4.16, page 506, modified as follows: First, it seems to me that ζ_i in the third line of the problem statement should be just ζ. (For alternative formulations involving different ζ_i's, see TPE section 3.6.)
We observe independent Bernoulli variables $X_1, \ldots, X_n$ which depend on unobservable variables Z_i distributed independently as $N(\zeta, \sigma^2)$ where

$$X_i = \begin{cases} 0, & \text{if } Z_i \le u_i, \\ 1, & \text{if } Z_i > u_i. \end{cases}$$

Assuming that $u_1, \ldots, u_n$ are known, we are interested in obtaining maximum likelihood estimates of ζ and σ^2 based on $X_1, \ldots, X_n$.
(aa) Show that if $u_1 = u_2 = \cdots = u_n \equiv u$ and $\theta = (\zeta, \sigma)$, then the distribution P_θ of X_1 is *not identifiable*.
(a) Show that the likelihood function is $\prod_{i=1}^n p_i^{X_i}(1-p_i)^{1-X_i}$ where $p_i = P(Z_i > u_i) = \Phi((\zeta - u_i)/\sigma)$, $i = 1, \ldots, n$. You will need to make further appropriate changes in Lehmann and Casella parts (c)-(e) as well.

Solution: (aa) If all the u_i's are equal to one fixed u, then the X_i's are Bernoulli (p) where

$$p \equiv p(\zeta, \sigma) = P(X > u) = 1 - \Phi((u - \zeta)/\sigma) = \Phi((\zeta - u)/\sigma).$$

Thus $\sum_{i=1}^n X_i \sim \text{Binomial}(n, p)$ and we can estimate $p(\theta) = p(\zeta, \sigma)$, but not $\theta = (\zeta, \sigma)$: $\theta = (\zeta, \sigma)$ is not identifiable: if $\theta_1 = (\zeta_1, \sigma_1) \neq (\zeta_2, \sigma_2) = \theta_2$ satisfy $\zeta_1\sigma_2 - \zeta_2\sigma_1 = u(\sigma_2 - \sigma_1)$, then $(\zeta_1 - u)/\sigma_1 = (\zeta_2 - u)/\sigma_2$ and hence $p(\theta_1) = p(\theta_2)$.

(a) Here $Z_i \sim N(\zeta, \sigma^2)$, $i = 1, \ldots, n$ are i.i.d., and then $X_i = 1_{(u_i, \infty)}(Z_i)$ for $i = 1, \ldots, n$. Thus $X_i \sim \text{Bernoulli}(p_i)$ with

$$\begin{aligned} p_i &\equiv p_i(\zeta, \sigma) = P_{\zeta,\sigma}(Z > u_i) = 1 - \Phi\left(\frac{u_i - \zeta}{\sigma}\right) \\ &= \Phi\left(\frac{\zeta - u_i}{\sigma}\right). \end{aligned}$$

(a) Thus the likelihood of the X_i's (the incomplete data) is

$$L(\zeta, \sigma | \underline{X}) = \prod_{i=1}^{n} p_i^{X_i} (1 - p_i)^{1-X_i}.$$

(b) The likelihood of the Z_i's (the complete data) is

$$L(\zeta, \sigma | \underline{Z}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} (Z_i - \zeta)^2 \right),$$

and the expected complete data log-likelihood given the observed (or incomplete data) is, with $\theta = (\zeta, \sigma)$, $\theta_0 = (\zeta_0, \sigma_0)$,

$$Q(\theta | \theta_0, \underline{X}) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{1}^{n} \left\{ E_{\theta_0}(Z_i^2 | X_i) - 2\zeta E_{\theta_0}(Z_i | X_i) + \zeta^2 \right\}.$$

(c) Since the MLE's for the complete data (the Z_i's) are the usual

$$\hat{\zeta} = \overline{Z} \qquad \text{and}$$
$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z})^2 = n^{-1} \sum_{i=1}^{n} Z_i^2 - \overline{Z}_n^2,$$

it follows that the EM sequence is given by

$$\hat{\zeta}^{(j+1)} = \frac{1}{n} \sum_{i=1}^{n} t(\hat{\zeta}^{(j)}, \hat{\sigma}^{(j)}, X_i, u_i),$$
$$(\hat{\sigma}^{(j+1)})^2 = \frac{1}{n} \sum_{i=1}^{n} v(\hat{\zeta}^{(j)}, \hat{\sigma}^{(j)}, X_i, u_i) - (\hat{\zeta}^{(j+1)})^2,$$

where

$$t(\zeta, \sigma, X, u) = E(Z | u, X, \zeta, \sigma),$$
$$v(\zeta, \sigma, X, u) = E(Z^2 | u, X, \zeta, \sigma).$$

(d) To find explicit expressions for the conditional expectations in the last display, we proceed via the following two claims:
Claim 1. Let $W \sim N(0, 1)$ and $Y \equiv 1_{(t,\infty)}(W)$. Then

$$E(W|Y) = Y \frac{\phi(t)}{1 - \Phi(t)} + (1 - Y) \frac{-\phi(t)}{\Phi(t)} \equiv H(t, Y),$$
$$E(W^2|Y) = 1 + tE(W|Y) = 1 + tH(t, Y).$$

Claim 2. If $Z \sim N(\zeta, \sigma^2)$ and $X \equiv 1_{(u,\infty)}(Z)$, then

$$E(Z|u, X, \zeta, \sigma^2) = \zeta + \sigma H\left(\frac{u - \zeta}{\sigma}, X \right),$$
$$E(Z^2|u, X, \zeta, \sigma^2) = \zeta^2 + \sigma^2 + \sigma(u + \zeta) H\left(\frac{u - \zeta}{\sigma}, X \right).$$

Proof of Claim 1: To prove the first part of Claim 1 we need to show that

$$E\left\{1_B(Y)\left(Y\frac{\phi(t)}{1-\Phi(t)}+(1-Y)\frac{-\phi(t)}{\Phi(t)}\right)\right\}=E\{1_B(Y)W\}$$

for all Borel sets B; see e.g. Lehmann and Romano, TSH, pages 36 - 37, or Shorack, *Probability for Statisticians*, page 158. Since Y takes values in $\{0,1\}$, it suffices to show this for $B=\{0\}$ and for $B=\{1\}$. For $B=\{1\}$, the left side equals

$$E\left\{Y\frac{\phi(t)}{1-\Phi(t)}\right\}=\phi(t),$$

while the right side equals

$$E\{1_B(Y)W\}=E\{W1_{[W>t]}\}=\int_t^\infty z\phi(z)dz=-\int_t^\infty \phi'(z)dz=\phi(t),$$

so the required identity holds. For $B=\{0\}$ the left side equals

$$E\left\{(1-Y)\frac{-\phi(t)}{\Phi(t)}\right\}=-\phi(t),$$

and the right side equals

$$E\{W1_{[W\le t]}\}=\int_{-\infty}^t z\phi(z)dz=-\int_{-\infty}^t \phi'(z)dz=-\phi(t),$$

so the identity holds in this case as well, and this completes the proof of the first part of the claim. To prove the second part of claim 1, we need to show that

$$E\left\{1_B(Y)(1+tE(W|Y))\right\}=E\{1_B(Y)W^2\}$$

for all Borel sets B. As before, since Y takes values in $\{0,1\}$ it suffices to consider $B=\{0\}$ and $B=\{1\}$. For $B=\{1\}$, we use the calculations above to see that the left side equals

$$p+t\phi(t)$$

while the right side equals

$$\begin{aligned}E(W^2 1_{[W>t]}) &= \int_t^\infty z^2\phi(z)dz=-\int_t^\infty z\phi'(z)dz\equiv-\int_t^\infty udv\\ &= -\left\{uv|_t^\infty-\int_t^\infty vdu\right\}\qquad\text{with } u=z,\quad v=\phi(z),\\ &= t\phi(t)+1-\Phi(t)=t\phi(t)+p,\end{aligned}$$

so the required identity holds. The verification for $B=\{0\}$ is similar, and this completes the proof of Claim 1.

Proof of Claim 2: This proceeds by reduction to Claim 1: note that $(Z-\zeta)/\sigma =_d W\sim N(0,1)$, and $X=1_{[Z>u]}=1_{[(Z-\zeta)/\sigma>(u-\zeta)/\sigma]}=_d Y$ with $t=(u-\zeta)/\sigma$. Thus

$$\begin{aligned}t(\zeta,\sigma,X,u) &\equiv E(Z|u,X,\zeta,\sigma)\\ &= \zeta+\sigma E\left(\frac{Z-\zeta}{\sigma}|X=1\{(Z-\zeta)/\sigma>(u-\zeta)/\sigma\}\right)\\ &= \zeta+\sigma H\left(\frac{u-\zeta}{\sigma},X\right),\end{aligned}$$

and (with $t = (u-\zeta)/\sigma$),

$$
\begin{aligned}
v(\zeta,\sigma,X,u) &\equiv E(Z^2|u,X,\zeta,\sigma) = E((\zeta+\sigma W)^2|Y) \\
&= \zeta^2 + 2\sigma\zeta E(W|Y) + \sigma^2 E(W^2|Y) \\
&= \zeta^2 + 2\sigma\zeta H\left(\frac{u-\zeta}{\sigma}, X\right) + \sigma^2\left(1 + \frac{u-\zeta}{\sigma} H\left(\frac{u-\zeta}{\sigma}, X\right)\right) \\
&= \zeta^2 + \sigma^2 + \sigma(u+\zeta) H\left(\frac{u-\zeta}{\sigma}, X\right).
\end{aligned}
$$

(e) To see that the EM iterates converge to the ML estimates $\hat{\zeta}$ and $\hat{\sigma}$ of ζ and σ, note that $t(\zeta,\sigma,X,u)$ and $v(\zeta,\sigma,X,u)$ are continuous functions of ζ and σ, and hence the expected complete data log-likelihood $Q(\theta|\theta_0,\underline{X})$ is continuous in both $\theta = (\zeta,\sigma)$ and $\theta_0 = (\zeta_0,\sigma_0)$. It follows from Theorem 4.12 of TPE page 460 that all the limit points of an EM iteration sequence are stationary points of $L(\zeta,\sigma|\underline{X})$. I have not yet succeeded in showing that the log-likelihood is concave with a unique stationary point yielding the maximum. Although the functions $g(z) = \log(\Phi(z))$ and $h(z) = \log(1-\Phi(z))$ are both concave (which follows from log-concavity of the standard normal density ϕ), the functions $r(\zeta,\sigma) = (\zeta - u_i)/\sigma$ are apparently not jointly concave, as can be seen by computing the Hessian. Although I can show that the log-likelihood is concave in each parameter separately, I have not yet succeeded in proving joint concavity. Thus I still lack a proof of uniqueness of the global maximum. [It is tempting to to try to use preservation properties of concave and convex functions as given, for example, in Boyd and Vandenberghe (2004), pages 79 - 85.] This is consistent with what we get by writing down the score equations for the incomplete data version of the model: In this case

$$
l(\zeta,\sigma|\underline{X}) = \sum_1^n \left\{X_i \log p_i(\zeta,\sigma) + (1-X_i)\log(1-p_i(\zeta,\sigma))\right\},
$$

so

$$
\begin{aligned}
\dot{l}_\zeta(\zeta,\sigma|\underline{X}) &= \sum_{i=1}^n \left\{\frac{X_i}{p_i(\zeta,\sigma)} - \frac{1-X_i}{1-p_i(\zeta,\sigma)}\right\}\frac{\partial p_i}{\partial \zeta} = \sum_{i=1}^n (X_i - p_i(\zeta,\sigma))\frac{\partial p_i/\partial\zeta}{p_i(1-p_i)}, \\
\dot{l}_\sigma(\zeta,\sigma|\underline{X}) &= \sum_{i=1}^n \left\{\frac{X_i}{p_i(\zeta,\sigma)} - \frac{1-X_i}{1-p_i(\zeta,\sigma)}\right\}\frac{\partial p_i}{\partial \sigma} = \sum_{i=1}^n (X_i - p_i(\zeta,\sigma))\frac{\partial p_i/\partial\sigma}{p_i(1-p_i)}.
\end{aligned}
$$

Note that if all the u_i's are equal, then the ratios

$$
\frac{\partial p_i/\partial\zeta}{p_i(1-p_i)}, \quad \text{and} \quad \frac{\partial p_i/\partial\sigma}{p_i(1-p_i)}
$$

are constant in i and the two score equations above degenerate to the same equation and yield $\hat{p} = n^{-1}\sum_{i=1}^n X_i$. But this is not enough to be able to estimate both ζ and σ. Thus it seems that the u_i's need to take on at least two distinct values in order for both ζ and σ to be identifiable. Also note that this problem is a parametric version of the model discussed in Example 4.6.5, page 40, Chapter 4 notes (with Y_i there being the current (deterministic) u_i, and X_i there being the current Z_i).

4. Suppose that $X \sim F_\theta =$ exponential(θ) with density $f_\theta(x) = \theta e^{-\theta x} 1_{(0,\infty)}(x)$ and $Y \sim G_\eta$ independent of X with densities $\{g_\eta : \eta \in R^+\}$, a regular parametric model on $(0,\infty)$. In (optional) problem xx of problem set 8, Statistics 581, we considered the following three scenarios for observation of X or functions of X:
(a) Uncensored: we observe X and Y.
(b) Right-censored: we observe $T(X,Y) = (X \wedge Y, 1\{X \le Y\}) \equiv (\min\{X,Y\}, 1\{X \le Y\}) \equiv (Z,\Delta)$.
(c) Interval-censored (case 1): we observe $S(X,Y) = (Y, 1\{X \le Y\}) \equiv (Y,\Delta)$.
In that problem, in each of the three scenarios (a), (b), (c), we computed: (i) The joint density of (X,Y) and joint distributions of $T(X,Y)$ and $S(X,Y)$.
(ii) The scores for θ and η. (Let $(\partial/\partial\eta)\log g_\eta(y) \equiv a(y)$ with $a \in L_2^0(G_\eta)$.)
(iii) And we compared $I_{X,Y}(\theta)$, $I_{T(X,Y)}(\theta)$, and $I_{S(X,Y)}(\theta)$.
Here we will additionally assume that $Y \sim$ exponential(η) (independent of X). In scenarios (b) and (c), find EM algorithms for computation of the MLE's $\hat\theta_n$ of θ based on $T_1,\ldots,T_n$ and $S_1,\ldots,S_n$ i.i.d. as $T(X,Y)$ and $S(X,Y)$ respectively. How would you estimate the variance of $\hat\theta$ in each case?

Solution: The complete data is (X_i, Y_i), $i = 1,\ldots,n$ where $X_i \sim \exp(\theta)$ and $Y_i \sim \exp(\eta)$. Hence the complete data log-likelihood is given by

$$\begin{aligned} l_n(\theta,\eta|\underline{X},\underline{Y}) &= \sum_{i=1} Y^n \{\log\theta - \theta X_i + \log\eta - \eta Y_i\} \\ &= n\log\theta - \theta\sum_{i=1}^n X_i + n\log\eta - \eta\sum_{i=1}^n Y_i. \end{aligned}$$

(b) In the right-censoring scenario we observe $T_i \equiv T(X_i,Y_i) = (Z_i,\Delta_i) = (X_i \wedge Y_i, 1\{X_i \le Y_i\}$ for $i = 1,\ldots,n$, and for the E-step of an EM-algorithm we want to compute

$$\begin{aligned} &E_\theta\left\{l_n(\theta,\eta|\underline{X},\underline{Y})\Big|\underline{T}\right\} \\ &= n\log\theta - \theta\sum_{i=1}^n E(X_i|T_i) + n\log\eta - \sum_{i=1}^n E(Y_i|T_i). \end{aligned}$$

Here

$$\begin{aligned} E(X|T) &= E(X|Z,\Delta) = \Delta Z + (1-\Delta)E(X|Z=Y, 1\{X>Y\}) \\ &= \Delta Z + (1-\Delta)\frac{\int_Z^\infty x\,dF(x)}{1-F(Z)} \\ &= \Delta Z + (1-\Delta)\frac{\int_Z^\infty x\theta e^{-\theta x}dx}{1-\exp(-\theta Z)} \\ &= \Delta Z + (1-\Delta)\left(\theta^{-1} + Z\right) \end{aligned}$$

where the last equality can be easily seen from the lack of memory property of the exponential distribution of X. Thus for estimation of θ, the EM algorithm consists of an E-step given by

$$\hat X_i^{(m)} = \Delta_i Z_i + (1-\Delta_i)\left(\frac{1}{\hat\theta_i^{(m)}} + Z_i\right), \quad i = 1,\ldots,n,$$

followed by an M-step which yields

$$\hat{\theta}^{(m)} = \left(\frac{1}{n} \sum_{i=1}^{n} \hat{X}_i^{(m)} \right)^{-1} .$$

(c) In the interval censoring (or current status) scenario we observe $S_i \equiv S(X_i, Y_i) = (Y_i, \Delta_i) = (Y_i, 1\{X_i \le Y_i\}$ for $i = 1, \ldots, n$, and for the E-step of an EM-algorithm we want to compute

$$\begin{aligned} & E_\theta \left\{ l_n(\theta, \eta | \underline{X}, \underline{Y}) \Big| \underline{S} \right\} \\ & = \quad n \log \theta - \theta \sum_{i=1}^{n} E(X_i | S_i) + n \log \eta - \sum_{i=1}^{n} E(Y_i | S_i). \end{aligned}$$

Here

$$\begin{aligned} E(X|S) &= E(X|Y, \Delta) = \Delta E(X|Y, 1\{X \le Y\} + (1 - \Delta) E(X|Y, 1\{X > Y\}) \\ &= \Delta \frac{\int_0^Y x dF(x)}{F(Y)} + (1 - \Delta) \frac{\int_Y^\infty x dF(x)}{1 - F(Y)} \\ &= \Delta \frac{\int_0^Y x\theta \exp(-\theta x) dx}{F(Y)} + (1 - \Delta) \frac{\int_Y^\infty x \theta e^{-\theta x} dx}{1 - \exp(-\theta Y)} \\ &= \Delta \frac{1 - e^{-\theta Y}(1 + \theta Y)}{\theta(1 - e^{-\theta} Y)} + (1 - \Delta) \left(\theta^{-1} + Y \right) \end{aligned}$$

where the last equality can again be seen from the lack of memory property of the exponential distribution of X. Thus for estimation of θ, the EM algorithm consists of an E-step given by

$$\hat{X}_i^{(m)} = \Delta_i \frac{1 - e^{-\hat{\theta}^{(m)} Y_i}(1 + \hat{\theta}^{(m)} Y_i)}{1 - e^{-\hat{\theta}^{(m)}}} + (1 - \Delta_i) \left(\frac{1}{\hat{\theta}_i^{(m)}} + Y_i \right), \quad i = 1, \ldots, n,$$

followed by an M-step which yields

$$\hat{\theta}^{(m)} = \left(\frac{1}{n} \sum_{i=1}^{n} \hat{X}_i^{(m)} \right)^{-1} .$$

5. **Optional bonus problem 1:** (Profile likelihood) [For nice plots to accompany this exercise, see pages 41 - 43 of Cox, D. R. and Oakes, D. (1984); *Analysis of Survival Data*, Chapman and Hall.] Consider the Weibull family of example 3.2.5 (581 Course Notes) $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^{+2}$ given by the (Lebesgue) densities

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

where $\theta \equiv (\alpha, \beta) \in (0, \infty) \times (0, \infty) \subset \mathbb{R}^2$.
(a) For a sample of n observations from p_θ, we know that, for each fixed value of β the value of α which maximizes the likelihood as a function of α is

$$\widehat{\alpha}(\beta) = \{\frac{1}{n} \sum_{i=1}^{n} X_i^\beta\}^{1/\beta} .$$

Use this to compute the *profile likelihood* $l_{\text{profile}}(\beta) = l_{\text{profile}}(\beta|\underline{X})$ defined by

$$l_{\text{profile}}(\beta) = l(\widehat{\alpha}(\beta), \beta) = l(\widehat{\alpha}(\beta), \beta|\underline{X})\,.$$

(b) Use what we know from Statistics 581 problem 9.2 to show that the profile likelihood is strictly concave and hence has a unique maximum. Show that maximizing the profile likelihood as a function of β yields the maximum likelihood estimate: i.e. that $(\widehat{\alpha}, \widehat{\beta}) = (\widehat{\alpha}(\widehat{\beta}_{\text{profile}}), \widehat{\beta}_{\text{profile}})$.

Solution: (a) The log-likelihood is

$$l(\alpha, \beta) = n\log(\beta/\alpha) + (\beta - 1)\sum_{i=1}^{n}\log(\frac{X_i}{\alpha}) - \sum_{i=1}^{n}(\frac{X_i}{\alpha})^{\beta}$$

and for fixed β the value of α which maximizes this is

$$\hat{\alpha}(\beta) = (\frac{1}{n}\sum_{i=1}^{n} X_i^{\beta})^{1/\beta}.$$

Thus the profile log-likelihood is

$$l_{profile}(\beta) \;=\; l(\hat{\alpha}(\beta), \beta) = n\log\beta - n\log(\sum_{i=1}^{n} X_i^{\beta}) + (\beta - 1)\sum_{i=1}^{n}\log X_i + n\log n - n.$$

(b) It follows that the score function for β corresponding to the profile log-likelihood is

$$\dot{\mathbf{l}}_{\text{profile},\beta}(\underline{X}) = \frac{n}{\beta} - n\frac{\sum_{i=1}^{n} X_i^{\beta}\log X_i}{\sum_{i=1}^{n} X_i^{\beta}} + \sum_{i=1}^{n}\log X_i,$$

and the observed information is

$$\begin{aligned} -\ddot{\mathbf{l}}_{profile,\beta}(\underline{X}) &= \frac{n}{\beta^2} + n\left\{\frac{\sum_{i=1}^{n} X_i^{\beta}(\log X_i)^2}{\sum_{i=1}^{n} X_i^{\beta}} - \left(\frac{\sum_{i=1}^{n} X_i^{\beta}\log X_i}{\sum_{i=1}^{n} X_i^{\beta}}\right)^2\right\} \\ &> 0 \end{aligned}$$

since the term in brackets is a variance, and hence is positive. Thus the profile likelihood is strictly concave and its maximum is unique.

Let $l^{\#}(\beta) = l_{profile}(\beta|\underline{X})$. Then, by the chain rule,

$$\dot{\mathbf{l}}^{\#}_{\beta}(\underline{X}) = \dot{\mathbf{l}}_{n\alpha}|_{\hat{\alpha}(\beta)}\dot{\alpha}(\beta) + \dot{\mathbf{l}}_{n\beta}|_{\hat{\alpha}(\beta)} = \dot{\mathbf{l}}_{n\beta}|_{\hat{\alpha}(\beta)} \tag{0.1}$$

since

$$\dot{\mathbf{l}}_{n\alpha}|_{\hat{\alpha}(\beta)} = 0. \tag{0.2}$$

Hence solving the profile score equation $\dot{\mathbf{l}}^{\#}_{\beta}(\underline{X}) = 0$ yields a solution of the likelihood equations $\dot{\mathbf{l}}_{n\alpha} = 0$ and $\dot{\mathbf{l}}_{n\beta} = 0$.

6. **Optional bonus problem 2.** This is a continuation of problem 5 above.
(a) What is the relationship of the score function for β from the profile likelihood, $\dot{l}_{\beta,\text{profile}}$ to the (efficient) score for β from the full likelihood? Prove or disprove my claim: the profile score for β (based on n observations) is asymptotically equivalent to the sum of efficient scores for β over the sample in the sense that their difference divided by $\sqrt{n}$ converges to 0 in probability.
(b) What is the relationship of the observed information from the profile likelihood $-\ddot{l}_{\beta\beta,\text{profile}}$ to information quantities from the full likelihood?

7. **Optional bonus problem 3.** Lehmann and Casella, problem 3.15, page 502: note that part (a) is false in the modern understanding of "log-concave" since all uniform distributions $f(x) = (b-a)^{-1}1_{[a,b]}(x)$ are log-concave, but the mode is not unique. (See Dharmadhikari and Jogdeo (1988), *Unimodality, Convexity, and Applications*, Section 1.4 and page 23.)