**Statistics 582, Problem Set 1 Solutions**

Wellner; 1/11/2018

1. Consider the Weibull family of distributions as given in Example 3.2.5 of Chapter 3 of the course notes and as in Example 5.43 of van der Vaart's *Asymptotic Statistics* page 70. Rewrite van der Vaart's example in terms of the parametrization given in Example 3.2.5 of the course notes, including calculation of third derivatives and computation of dominating functions for the 3rd derivatives. (Are the computations simpler in one or the other of the two parametrizations?)

**Solution:** For the Weibull family, Example 3.2.5, with $\theta = (\alpha, \beta)$,

$$p_\theta(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^\beta \exp\left(-\left(\frac{x}{\alpha}\right)^\beta\right),$$

we compute

$$\begin{aligned} \dot{l}_\alpha(x) &= \frac{\beta}{\alpha}\left\{\left(\frac{x}{\alpha}\right)^\beta - 1\right\}, \\ \dot{l}_\beta(x) &= \frac{1}{\beta}\left\{1 - \left\{\left(\frac{x}{\alpha}\right)^\beta - 1\right\}\log\left(\frac{x}{\alpha}\right)^\beta\right\}. \end{aligned}$$

Thus the likelihood equations become

$$0 = \sum_{i=1}^n \dot{l}_\alpha(X_i) = \frac{\beta}{\alpha}\left\{\sum_{i=1}^n \left(\frac{X_i}{\alpha}\right)^\beta - n\right\},$$

and

$$0 = \sum_{i=1}^n \dot{l}_\beta(X_i) = \frac{1}{\beta}\left\{n - \sum_{i=1}^n \left(\left(\frac{X_i}{\alpha}\right)^\beta - 1\right)\log\left(\frac{X_i}{\alpha}\right)^\beta\right\}.$$

Equivalently,

$$\alpha = \left\{n^{-1}\sum_{i=1}^n X_i^\beta\right\}^{1/\beta},$$

and

$$0 = \frac{1}{\beta} - \frac{\sum_{i=1}^n X_i^\beta \log X_i}{\sum_{i=1}^n X_i^\beta} + \frac{1}{n}\sum_{i=1}^n \log X_i. \tag{0.1}$$

The right side of the second equation decreases from $+\infty$ at $\beta = 0$ to $\overline{\log X} - \log X_{(n)} < 0$ at $\beta = +\infty$, and hence has a unique solution, $\widehat{\beta}_n$ if $n \ge 2$ with at least two distinct values of the $X_i$'s. Then

$$\widehat{\alpha}_n = \left(n^{-1}\sum_{i=1}^n X_i^{\widehat{\beta}_n}\right)^{1/\widehat{\beta}_n}.$$

The four different third order partial derivatives are (with a little help from Mathematica),

$$
\begin{aligned}
\dddot{\mathbf{l}}_{\alpha\alpha\alpha}(x) &= \frac{\beta}{\alpha^3}\left\{2\left(\left(\frac{x}{\alpha}\right)^\beta - 1\right) + 3\beta\left(\frac{x}{\alpha}\right)^\beta + \beta^2\left(\frac{x}{\alpha}\right)^\beta\right\}, \\
\dddot{\mathbf{l}}_{\alpha\alpha\beta}(x) &= -\frac{1}{\alpha^2}\left\{(1+\beta)\left(\frac{x}{\alpha}\right)^\beta \log\left(\frac{x}{\alpha}\right)^\beta + 2\beta\left(\frac{x}{\alpha}\right)^\beta + \left(\frac{x}{\alpha}\right)^\beta - 1\right\}, \\
\dddot{\mathbf{l}}_{\alpha\beta\beta}(x) &= \frac{1}{\alpha\beta}\left\{\left(\frac{x}{\alpha}\right)^\beta \log\left(\frac{x}{\alpha}\right)^\beta \left(2 + \log\left(\frac{x}{\alpha}\right)^\beta\right)\right\}, \\
\dddot{\mathbf{l}}_{\beta\beta\beta}(x) &= \frac{1}{\beta^3}\left\{2 - \left(\frac{x}{\alpha}\right)^\beta \left(\log\left(\frac{x}{\alpha}\right)^\beta\right)^3\right\}.
\end{aligned}
$$

As in the case of the parametrization treated by van der Vaart (1998), these function are all dominated (bounded in small neighborhoods of $\theta_0 = (\alpha_0, \beta_0)$) by a function of the form

$$
M(x) = A(1 + x^B)(1 + |\log x| + |\log x|^2 + |\log x|^3)
$$

for sufficiently large $A$ and $B$. Since the Weibull distribution has an exponentially small tail, all (absolute) moments of $\dot{\mathbf{l}}_\theta$ and $\ddot{\mathbf{l}}_{\theta\theta}$ exist and $M$ is integrable. There does not seem to be any substantial simplication of the computations in either of these parametrizations.

2. This is a continuation of the previous problem concerning ML estimation in the Weibull family. (a) Using the parametrization of the course notes, find the maximizer of the log-likelihood over $\alpha$ for each fixed value of $\beta$,

$$
\widehat{\alpha}_n(\beta) \equiv \mathrm{argmax}_{\alpha>0} l_n(\alpha, \beta|\underline{X}_n).
$$

(b) The *profile likelihood function* for $\beta$ is defined by

$$
l_n^{prof}(\beta|\underline{X}) \equiv l_n(\widehat{\alpha}_n(\beta), \beta|\underline{X}).
$$

Compute $l_n^{prof}(\beta|\underline{X})$ for the Weibull family; show that it has a unique maximizer $\widehat{\beta}_n$ and that the MLE $\widehat{(\alpha, \beta)} = (\widehat{\alpha}_n(\widehat{\beta}_n), \widehat{\beta}_n)$. (This is related to Example 6.1 on page 468 of Lehmann and Casella and their problems 6.1-6.3 on page 509. 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.)
(c) Why does profile likelihood work so nicely in this example?

**Solution:** (a) As in the previous problem, for each fixed $\beta$ the score equation for $\alpha$ leads to

$$
\widehat{\alpha}_n(\beta) = \left\{n^{-1}\sum_{i=1}^n X_i^\beta\right\}^{1/\beta}.
$$

(b) Here

$$
\log p_\theta(x) = \log(\beta/\alpha) + (\beta - 1)\log(x/\alpha) - \left(\frac{x}{\alpha}\right)^\beta
$$

so the log-likelihood functions is given by

$$l_n(\alpha,\beta|\underline{X}_n) = \log p_\theta(X_i) = n\log(\beta/\alpha) + (\beta-1)\sum_{i=1}^n \log(X_i/\alpha) - \sum_{i=1}^n \left(\frac{X_i}{\alpha}\right)^\beta ,$$

and the profile log-likelihood is

$$\begin{aligned} l_n^{prof}(\beta|\underline{X}_n) &= l_n(\widehat{\alpha}_n(\beta),\beta|\underline{X}) \\ &= n\log\beta + (\beta-1)\sum_{i=1}^n \log X_i - n\log\left(\sum_{i=1}^n X_i^\beta\right) - n + n\log n \\ &= n\left\{\log\beta + (\beta-1)n^{-1}\sum_{i=1}^n \log X_i - \log\left(\sum_{i=1}^n X_i^\beta\right) - 1 + \log n\right\}. \end{aligned}$$

Thus, either by differentiation of the last line above, or by a calculation via the chain rule as follows,

$$\begin{aligned} \dot{\mathrm{l}}_{n,\beta}^{prof}(\beta) &= \dot{\mathrm{l}}_{n,\alpha}(\widehat{\alpha}_n(\beta),\beta)\frac{\partial}{\partial\beta}\widehat{\alpha}_n(\beta) + \dot{\mathrm{l}}_{n,\beta}(\widehat{\alpha}_n(\beta),\beta) \\ &= \dot{\mathrm{l}}_{n,\beta}(\widehat{\alpha}_n(\beta),\beta) \quad \text{since} \quad \dot{\mathrm{l}}_{n,\alpha}(\widehat{\alpha}_n(\beta),\beta) = 0 \\ &= n\left\{\frac{1}{\beta} - \frac{\sum_1^n X_i^\beta \log X_i}{\sum_1^n X_i^\beta} + \frac{1}{n}\sum_1^n \log X_i\right\} \end{aligned}$$

after a some algebra. Thus we see that solving the "profile score equation" obtained by setting this last display equal to zero yields the equation (0.1) obtained in problem 1 above, which has a unique solution $\widehat{\beta}_n$ as shown in problem 1 (if $n \geq 2$ and there are at least two distinct observations).
(c) Profile likelihood works well here because for fixed $\beta$ the Weibull scale family is an exponential family with an explicit MLE given by $\widehat{\alpha}_n(\beta)$.

3. van der Vaart (1998), Problem 25, page 84.
In all three parts of this problem we will suppose that $X_1,\ldots,X_n$ are i.i.d. $N(\mu_0,\sigma_0^2)$ with $\theta = (\mu,\sigma^2) \in \mathbb{R}\times\mathbb{R}^+ = (-\infty,\infty)\times(0,\infty) \equiv \Theta$. Thus the density for the family is

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

and the functions $f(x,\theta)$ as in Section 4.5 are

$$\begin{aligned} f(x,\theta) &= \log p_\theta(x) - \log p_{\theta_0}(x) = \log p_\theta(x) - M_1(x) \\ &= -(1/2)\log(2\pi) - (1/2)\log\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2} - M_1(x) \end{aligned}$$

where $M_1$ is a fixed integrable function of $x$.
(i) If we restrict to a (large) compact set $[a,b]\times[c,d] \subset \mathbb{R}\times\mathbb{R}^+$ (so that $-\infty < a < b < \infty$, then $\sup_{\theta\in K} f(x,\theta)$ is achieved from some $\theta_K = (\mu_K,\sigma_K^2) \in K$, and the supremum is measurable. Moreover, it is dominated by a function $F$ of the form $A + Bx + Cx^2$ for sufficiently large $A, B, C \in (0,\infty)$. Thus the hypotheses

of Theorem 4.5.x of the course notes hold and the MLE (over $K$) is consistent (assuming that $\theta_0 \in K$; note that since $\theta_0$ is unknown, it is not possible to assure this in advance).
(ii) When we compactify $\Theta = (-\infty, \infty) \times (0, \infty)$ we want to get an upper semi-continuous function $\bar{f}(x, \theta)$ defined on a (compact set) $\overline{\Theta}$ in a one-to-one relationship with $[-\infty, \infty] \times [0, \infty] = (\mathbb{R} \cup \{\pm\infty\}) \times ((0, \infty) \cup \{0, \infty\}$. Here we get

$$\bar{f}(x, \theta) = \begin{cases} f(x, \theta), & \theta \in \mathbb{R} \times (0, \infty), \\ -\infty, & \theta \in \{\pm\infty\} \times (0, \infty), \\ -\infty, & \theta \in \overline{\mathbb{R}} \times \{\infty\}. \end{cases}$$

but for the set $\mathbb{R} \times \{0\}$ we are in trouble since for any $\delta > 0$,

$$\begin{aligned} \sup_{\theta \in \overline{\Theta}} f(x, \theta) &\geq \sup_{\theta \in \mathbb{R} \times [0, \delta]} f(x, \theta) \\ &\geq f(x, (\mu, \delta)) = -(1/2) \log(2\pi) - (1/2) \log \delta - M_1(x) \to +\infty \end{aligned}$$

as $\delta \searrow 0$. Thus the condition of an integrable (upper) envelope fails, and we cannot apply the theorem.
(iii) [Note: The "blocking method" discussed in van der Vaart (1998), pages 48-49, was proposed by M. D. Perlman (1972), "On the strong consistency of approximate maximum likelihood estimators"; Proc. Sixth Berk. Symp. Math. Statist. Probab. **1**, 263-281.]
However, for a pair of observations we find that

$$\begin{aligned} f(x, \theta) &= \log \left( \frac{1}{2\pi\sigma^2} \exp \left( -\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2}{2\sigma^2} \right) \right) - M_2(x) \\ &= -\log(2\pi) - \log \sigma^2 - \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2}{2\sigma^2} - M_2(x). \end{aligned} \tag{0.2}$$

In this case we note that for any $\theta \in (-\infty, \infty) \times (0, \infty)$ we have

$$\begin{aligned} f(x, \theta) &\leq -\log 2\pi - \log \hat{\sigma}^2 - \frac{(x_1 - \hat{\mu}_1)^2 + (x_2 - \mu)^2}{2\hat{\sigma}^2} - M_2(x) \\ &= -\log 2\pi - \log((x_1 - x_2)^2/4) - 1 - M_2(x) \equiv F(x) \end{aligned}$$

where $M_2$ is a fixed integrable function and where the inequality is achieved at $\hat{\mu}_2 = (x_1 + x_2)/2$ and $\hat{\sigma}_2^2 = ((x_1 - \hat{\mu}_2)^2 + (x_2 - \hat{\mu}_2)^2)/2$. Here $F$ is integrable since $-M_2$ is integrable and (using $-\log x \leq 4x^{-1/4}$)

$$E\{-\log(X_1 - X_2)^2\} \leq 4E_{\theta_0}|X_1 - X_2|^{-2/4} < \infty.$$

This last claim uses the fact that under $\theta_0$ we know that $Z \equiv (X_1 - X_2)/\sqrt{2}\sigma_0 \sim N(0, 1)$ and hence $Z^2 \sim \chi_1^2$. This is slightly delicate.
It remains to check that this same envelope "works" for the boundary points. The key point here is that $X_1 \neq X_2$ with probability 1 and hence the quadratic term in the numerator on the right side of (0.2 ) is positive almost surely. Thus $\sigma^2 \searrow 0$ drives the function to $-\infty$ at all these boundary points.

Table 1:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 1 | -1 | -1 | 2 | 2 | -2 | -2 | * | * | * | * |
| $x_2$ | 1 | -1 | 1 | -1 | * | * | * | * | 2 | 2 | -2 | -2 |

4. Lehmann and Casella, TPE, Problem 4.9, page 504: Consider the following 12 observations from a bivariate normal distribution with parameters $\mu_1 = \mu_2 = 0$, $\sigma_1^2$, $\sigma_2^2$, $\rho$: where $*$ represents a missing value.
(a) Show that the likelihood function has global maxima at $\rho = \pm 1/2$, $\sigma_1^2 = \sigma_2^2 = 8/3$ and a saddlepoint at $\rho = 0$, $\sigma_1^2 = \sigma_2^2 = 5/2$.
(b) Show that if an EM sequence starts with $\rho = 0$, then it remains with $\rho = 0$ for all subsequent iterations.
(c) Show that if an EM sequence starts with $\rho$ bounded away from zero, it will converge to a maximum.

**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 $\rho$ (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

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

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

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

We compute

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

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 $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$.

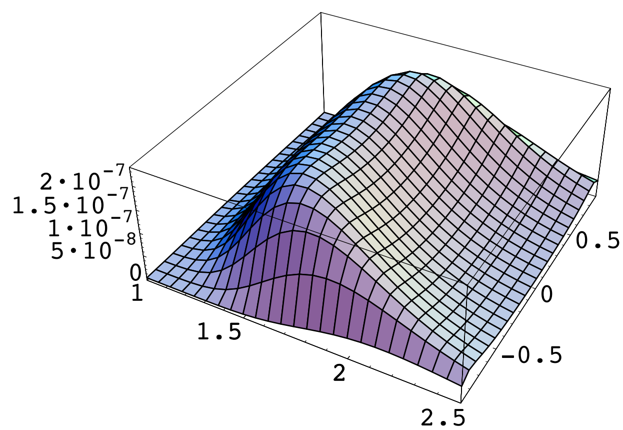


Figure 1: Plot of $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$.

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

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

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

$$E(Y_j|X_j) = \rho\tau X_j/\sigma, \qquad E(Y_j^2|X_j) = \tau^2(1-\rho^2) + (\rho\tau X_j/\sigma)^2, \qquad j = 5, \ldots, 8,$$
$$E(X_j|Y_j) = \rho\sigma Y_j/\tau, \qquad E(X_j^2|Y_j) = \sigma^2(1-\rho^2) + (\rho\sigma Y_j/\tau)^2, \qquad j = 9, \ldots, 12,$$

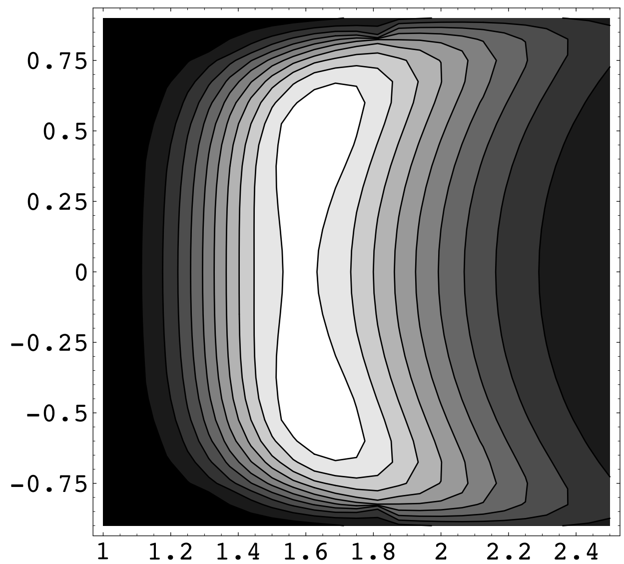


Figure 2: Contour plot of $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$.

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

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

where

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

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

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

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

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

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

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 $\hat{T}_{1,2}^{(m)} = \sum_{i=1}^{4} X_i Y_i = 0$ 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 $\rho$ 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

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

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

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

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.