Statistics 582, Problem Set 2 Solutions

Wellner; 1/24/2007

1. Human beings can be classified into one of four blood groups (phenotypes) O, A, B, AB. The inheritance of blood groups is controlled by three genes, O, A, B, of which O is recessive to A and B. If r, p, q are the gene probabilities in the population of O, A, B respectively, then the probabilities of the six possible combinations (genotypes) in random mating (where two individuals drawn at random from the population contribute one gene each) are shown in the following table:

Phenotype	Genotype	probability
O	OO	r^2
A	AA	p^2
A	AO	$2rp$
B	BB	q^2
B	BO	$2rq$
AB	AB	$2pq$

We observe among N individuals the phenotype frequencies N_O, N_A, N_B, N_{AB}, and wish to estimate the gene probabilities from such data. A simple approach is to regard the observations as incomplete, the complete data set being the genotype frequencies N_{OO}, N_{AA}, N_{AO}, N_{BB}, N_{BO}, N_{AB}.
A. Derive the EM algorithm for estimation of (p, q, r).
B. Estimate (p, q, r) from $N_O = 176$, $N_A = 182$, $N_B = 60$, $N_{AB} = 17$.
C. Estimate the covariance matrix of the estimator $(\hat{p}, \hat{q}, \hat{r})$.

Solution: A. The complete data is $\underline{N} \equiv (N_{OO}, N_{AA}, N_{AO}, N_{BB}, N_{BO}, N_{AB})$ with multinomial distribution $\text{Mult}_6(N; (r^2, p^2, 2rp, q^2, 2rq, 2pq))$. Thus

$$P(\underline{N} = \underline{n}) = \frac{N!}{n_{OO}!n_{AA}!n_{AO}!n_{BB}!n_{BO}!n_{AB}!} \cdot p^{2n_{AA}+n_{AO}+n_{AB}}\, q^{2n_{BB}+n_{BO}+n_{AB}}\, r^{2n_{OO}+n_{AO}+n_{BO}}\, 2^{n_{AO}+n_{BO}+n_{AB}}.$$

This is proportional to a $\text{Mult}_3(2N; (p, q, r))$ distribution, and hence the MLE's based on the complete data are

$$(\hat{p}, \hat{q}, \hat{r}) = \frac{1}{2N}(2N_{AA} + N_{AO} + N_{AB}, 2N_{BB} + N_{BO} + N_{AB}, 2N_{OO} + N_{AO} + N_{BO}).$$

This forms the basis of the "M - step" of an E-M algorithm. The incomplete data Y is $(N_A, N_B, N_O, N_{AB}) = (N_{AA} + N_{AO}, N_{BB} + N_{BO}, N_{OO}, N_{AB})$; thus

$$(N_{AA}|Y) = (N_{AA}|N_A) \sim \text{Binomial}(N_A, \frac{p^2}{p^2 + 2rp}), \qquad E(N_{AA}|Y) = N_A \frac{p}{p + 2r},$$

$$(N_{AO}|Y) = (N_{AO}|N_A) \sim Binomial(N_A, \frac{2rp}{p^2+2rp}), \qquad E(N_{AO}|Y) = N_A\frac{2r}{p+2r},$$

$$(N_{BB}|Y) = (N_{BB}|N_B) \sim Binomial(N_B, \frac{q^2}{q^2+2rq}), \qquad E(N_{BB}|Y) = N_B\frac{q}{q+2r},$$

$$(N_{BO}|Y) = (N_{BO}|N_B) \sim Binomial(N_B, \frac{2rq}{q^2+2rq}), \qquad E(N_{BO}|Y) = N_B\frac{2r}{q+2r}.$$

This gives the basis of the "E - step" for an E - M algorithm. Hence, starting from $(\hat{p}^{(0)}, \hat{q}^{(0)}, \hat{r}^{(0)}) = (1/3, 1/3, 1/3)$ say, we take

$$(\hat{p}^{(m+1)}, \hat{q}^{(m+1)}) = \frac{1}{2N}(2\hat{N}_{AA}^{(m)} + \hat{N}_{AO}^{(m)} + N_{AB}, 2\hat{N}_{BB}^{(m)} + \hat{N}_{BO}^{(m)} + N_{AB}),$$

$$\hat{r}^{(m+1)} = 1 - \hat{p}^{(m+1)} - \hat{q}^{(m+1)}$$

where

$$\hat{N}_{AA}^{(m)} \equiv N_A\frac{\hat{p}^{(m)}}{\hat{p}^{(m)} + 2\hat{r}^{(m)}}, \qquad \hat{N}_{AO}^{(m)} \equiv N_A - \hat{N}_{AA}^{(m)},$$

$$\hat{N}_{BB}^{(m)} \equiv N_B\frac{\hat{q}^{(m)}}{\hat{q}^{(m)} + 2\hat{r}^{(m)}}, \qquad \hat{N}_{BO}^{(m)} \equiv N_B - \hat{N}_{BB}^{(m)}.$$

B. For the given data, the E - M algorithm in A yields:

Iteration	$\hat{p}^{(m)}$	$\hat{q}^{(m)}$
0	.333	.333
1	.298	.111
2	.271	.094
3	.266	.093
4	.265	.093
5	.264	.093
6	.264	.093

Thus the estimator is $(\hat{p}, \hat{q}, \hat{r}) = (.264, .093, .642)$.

C. The likelihood of the observations $(N_A, N_B, N_O, N_{AB}) = (N_{AA} + N_{AO}, N_{BB} + N_{BO}, N_{OO}, N_{AB})$ is

$$\begin{aligned} l_N(p,q) &= N_A \log(p^2 + 2p(1-p-q)) \\ &\quad + N_B \log(q^2 + 2q(1-p-q)) \\ &\quad + N_O \log(1-p-q)^2 + N_{AB}\log(2pq)\,. \end{aligned}$$

Thus

$$
\begin{aligned}
-\frac{\partial^2}{\partial p^2} l_N(p,q) &= 2N_A\{\frac{1}{2p-p^2-2pq} + \frac{2(1-p-q)^2}{(2p-p^2-2pq)^2}\} \\
&\quad + N_B \frac{4q^2}{(2q-q^2-2pq)^2} \\
&\quad + \frac{N_{AB}}{p^2} + \frac{2N_O}{(1-p-q)^2}\,,
\end{aligned}
$$

$$
\begin{aligned}
-\frac{\partial^2}{\partial p \partial q} l_N(p,q) &= 2N_A\{\frac{1}{2p-p^2-2pq} - \frac{2p(1-p-q)}{(2p-p^2-2pq)^2}\} \\
&\quad + 2N_B\{\frac{1}{2q-q^2-2pq} - \frac{4q^2}{(2q-q^2-2pq)^2}\} \\
&\quad + \frac{2N_O}{(1-p-q)^2}\,,
\end{aligned}
$$

$$
\begin{aligned}
-\frac{\partial^2}{\partial q^2} l_N(p,q) &= N_A \frac{4p^2}{(2p-p^2-2pq)^2} \\
&\quad + 2N_B\{\frac{1}{2q-q^2-2pq} + \frac{2(1-p-q)^2}{(2q-q^2-2pq)^2}\} \\
&\quad + \frac{N_{AB}}{q^2} + \frac{2N_O}{(1-p-q)^2}\,.
\end{aligned}
$$

Since

$$
\begin{gathered}
E(N_A) = N(p^2 + 2p(1-p-q))\,, \\
E(N_B) = N(2q - q^2 - 2pq)\,, \\
E(N_{AB}) = N(2pq),
\end{gathered}
$$

and

$$
E(N_O) = N(1-p-q)^2\,,
$$

it follows that

$$
\begin{aligned}
I_{11}(p,q) &= 2N\{1 + \frac{2r^2}{2p-p^2-2pq} - \frac{2q^2}{2q-q^2-2pq} + \frac{q}{p} + 1\}, \\
I_{12}(p,q) &= 2N\{2 - \frac{2p(1-p-q)}{(2p-p^2-2pq)^2} - \frac{2q(1-p-q)}{(2q-q^2-2pq)^2} + 1\}\,, \\
I_{22}(p,q) &= 2N\{1 + \frac{2r^2}{2q-q^2-2pq} - \frac{2p^2}{2p-p^2-2pq} + \frac{p}{q} + 1\}
\end{aligned}
$$

and hence the estimated Fisher information matrix is

$$
\hat{I}(p,q) = \begin{pmatrix} 5.063 & 1.793 \\ 1.793 & 12.182 \end{pmatrix}
$$

so that

$$\hat{I}^{-1}(p,q) = \frac{1}{2N}\begin{pmatrix} .208 & -.003 \\ -.003 & .087 \end{pmatrix}.$$

Furthermore, since $\hat{r} = 1 - \hat{p} - \hat{q}$,

$$Var(\hat{r}) = Var(\hat{p}) + Var(\hat{q}) + 2Cov(\hat{p},\hat{q}),$$

$$Cov(\hat{p},\hat{r}) = -Var(\hat{p}) - Cov(\hat{p},\hat{q}),$$

$$Cov(\hat{q},\hat{r}) = -Var(\hat{q}) - Cov(\hat{p},\hat{q});$$

and hence we estimate $Cov(\hat{p},\hat{q},\hat{r})$ by

$$\widehat{Cov}(\hat{p},\hat{q},\hat{r}) = \begin{pmatrix} .000240 & -.000035 & -.000205 \\ -.000035 & .000095 & -.000060 \\ -.000205 & -.000060 & .000265 \end{pmatrix}.$$

2. 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

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

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

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

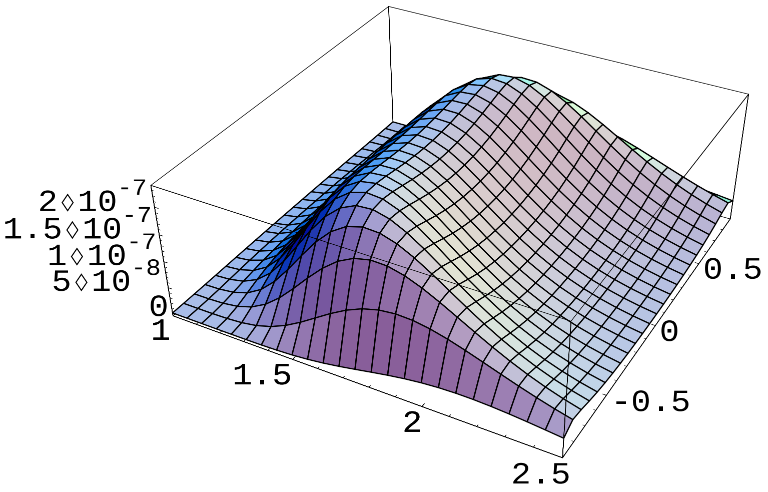

Figure 1: Plot of $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$.

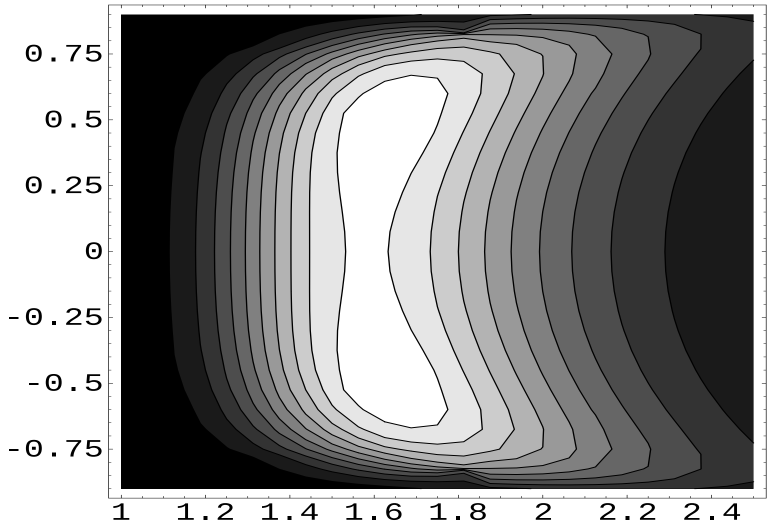

Figure 2: Contour plot of $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$.

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

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

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

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

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

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

where

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

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

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

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 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.

3. Suppose, as in Example 4.3.10, that $\underline{X}_1, \ldots, \underline{X}_n$ are i.i.d. $\text{Mult}_k(1, \underline{p})$ so that $\underline{N}_n = \sum_{i=1}^n \underline{X}_i \sim \text{Mult}_k(n, \underline{p})$.
(a) Use Jensen's inequality to show that the log-likelihood

$$l_n(\underline{p}|\underline{X}) = \sum_{j=1}^k N_j \log p_j + \sum_{i=1}^n \log\left(\frac{1!}{X_{i1}! \cdots X_{ik}!}\right)$$

is maximized by $\underline{\hat{p}} = \underline{N}_n/n$. [Hint: write the first term of $l_n(\underline{p}|\underline{X})$ as $n\sum_{j=1}^k \hat{p}_j \log p_j$.]
(b) Relate $l_n(\underline{p})$ to $K(\underline{\hat{p}}, \underline{p})$ and hence show again that the maximizing value of $\underline{p}$ is $\hat{p}$.

Solution: (a) Our goal is to show that

$$n\sum_{j=1}^k \hat{p}_j \log p_j \leq n\sum_{j=1}^k \hat{p}_j \log \hat{p}_j$$

with equality if and only if $\underline{p} = \underline{\hat{p}}$. Subtracting the right side from the the left side and dividing by n, we see that we want to show that

$$\sum_{j=1}^k \hat{p}_j \log\left(\frac{p_j}{\hat{p}_j}\right) \leq 0.$$

But since log is a concave function, Jensen's inequality yields

$$\begin{aligned}\sum_{j=1}^k \hat{p}_j \log\left(\frac{p_j}{\hat{p}_j}\right) &\leq \log\left(\sum_{j=1}^k \hat{p}_j\left(\frac{p_j}{\hat{p}_j}\right)\right)\\ &= \log\left(\sum_{j=1}^k p_j\right) = \log(1) = 0.\end{aligned}$$

(b) Note that in the above argument we have shown that

$$l_n(\underline{p}) - l_n(\underline{\hat{p}}) = -nK(\underline{\hat{p}}, \underline{p}) \leq 0$$

since $K(P,Q) \geq 0$ for all P, Q. Thus $l_n(\underline{p})$ is maximized by $\underline{p} = \underline{\hat{p}}$.

4. Suppose that the "complete data" X is given by three independent multinomial random vectors,

$$N(1) \equiv (N_{ij}(1) : i = 1, \ldots, r; j = 1, \ldots, s) \sim \text{Mult}_{rs}(n_1; p = (p_{ij}, i = 1, \ldots, r, j = 1, \ldots, s),$$

$$N(2) \equiv (N_{ij}(2) : i = 1, \ldots, r; j = 1, \ldots, s) \sim \text{Mult}_{rs}(n_2; p = (p_{ij}, i = 1, \ldots, r, j = 1, \ldots, s),$$

$$N(3) \equiv (N_{ij}(3) : i = 1, \ldots, r; j = 1, \ldots, s) \sim \text{Mult}_{rs}(n_3; p = (p_{ij}, i = 1, \ldots, r, j = 1, \ldots, s)\,.$$

Suppose that the "incomplete data" Y consists of $N(1), (N_{i\cdot}(2) : 1 \leq i \leq r), (N_{\cdot j}(3) : 1 \leq j \leq s)$.
A. What are the distributions of $N(1), (N_{i\cdot}(2) : 1 \leq i \leq r)$ and $(N_{\cdot j}(3) : 1 \leq j \leq s)$?
B. Find the conditional distribution(s) of X given Y.
C. Suggest an EM - algorithm for estimation of p.

Solution: A. By elementary considerations,

$$(N_{i\cdot}(2) : 1 \le i \le r) \sim Mult_r(n_2; (p_{i\cdot} : 1 \le i \le r))$$

and

$$(N_{\cdot j}(3) : 1 \le j \le s) \sim Mult_s(n_3; (p_{\cdot j} : 1 \le j \le s)).$$

B. First note that if

$$(N_{ij}) \sim \mathrm{Mult}_{rs}(n; (p_{ij})),$$

then

$$(N_{i\cdot}) \sim \mathrm{Mult}_r(n; (p_{i\cdot}))$$

as in A (since the components of $(N_{i\cdot})$ give the number of times outcome i occurred in n independent trials with probability $p_{i\cdot}$ on each trial). Furthermore

$$((N_{ij})|(N_{i\cdot})) \sim \prod_{i=1}^{r} Mult_s(N_{i\cdot}; (p_{ij}/p_{i\cdot}))\,. \tag{1}$$

(1) can be proved most easily by direct calculation of the conditional distribution:

$$\begin{aligned} &P(N_{ij} = k_{ij}, i = 1, \ldots, r, j = 1, \ldots, s \Big| N_{i\cdot} = k_{i\cdot}, i = 1, \ldots, r) \\ &= n! \prod_{i=1}^{r} \prod_{j=1}^{s} \frac{p_{ij}^{k_{ij}}}{k_{ij}!} / n! \prod_{i=1}^{r} \frac{p_{i\cdot}^{k_{i\cdot}}}{k_{i\cdot}!} \\ &= \prod_{i=1}^{r} \left\{ k_{i\cdot}! \prod_{j=1}^{s} \frac{(p_{ij}/p_{i\cdot})^{k_{ij}}}{k_{ij}!} \right\} \end{aligned}$$

on the set $k_{i\cdot} = \sum_{j=1}^{s} k_{ij}$, $i = 1, \ldots, r$. The terms inside the first product are just the $\mathrm{Mult}_s(k_{i\cdot}; (p_{ij}/p_{i\cdot}))$ probabilities.

Hence conditional on $(N_{i\cdot}(2) : 1 \le i \le r)$ the vectors $(N_{ij}(2) : 1 \le j \le s)$, $i = 1, \ldots, r$ are independent with $(N_{ij}(2) : 1 \le j \le s)|N_{i\cdot} \sim \mathrm{Mult}_s(N_{i\cdot}; (p_{ij}/p_{i\cdot}; j = 1, \ldots, s))$. Similarly, conditional on $(N_{\cdot j}(3) : 1 \le j \le s)$ the vectors $(N_{ij}(3) : 1 \le i \le r)$, $j = 1, \ldots, s$ are independent with $(N_{ij}(3) : 1 \le i \le r)|N_{\cdot j} \sim \mathrm{Mult}_r(N_{\cdot j}; (p_{ij}/p_{\cdot j}; i = 1, \ldots, r))$.

C. If we had the complete data $N_{ij}(1), N_{ij}(2), N_{ij}(3)$ for all i, j, then $N_{ij} \equiv N_{ij}(1) + N_{ij}(2) + N_{ij}(3)$ has a multinomial distribution with number of trials $n \equiv n_1 + n_2 + n_3$, and hence the MLE $\underline{\hat{p}} = (\hat{p}_{ij})$ of $\underline{p} = (p_{ij})$ is given by

$$\hat{p}_{ij} = \frac{N_{ij}}{n} = \frac{N_{ij}(1) + N_{ij}(2) + N_{ij}(3)}{n_1 + n_2 + n_3}.$$

This is the basis of the "M - step" of an E-M algorithm. But from B it follows that

$$E(N_{ij}(2)|N_{i\cdot}(2)) = N_{i\cdot}(2)\frac{p_{ij}}{p_{i\cdot}}\,, \qquad E(N_{ij}(3)|N_{\cdot j}(3)) = N_{\cdot j}(3)\frac{p_{ij}}{p_{\cdot j}}\,.$$

This is the basis of the "E - step" of an E-M algorithm. Thus, for some reasonable preliminary estimator like $\underline{\hat{p}}^{(0)} \equiv (\hat{p}_{ij}^{(0)}) = (N_{ij}(1)/n)$, a natural E - M algorithm is defined by

$$\hat{p}_{ij}^{(m+1)} = \frac{N_{ij}(1) + \widehat{N}_{ij}^{(m)}(2) + \widehat{N}_{ij}^{(m)}(3)}{n_1 + n_2 + n_3}$$

where

$$\widehat{N}_{ij}^{(m)}(2) \equiv N_{i\cdot}(2)\frac{\hat{p}_{ij}^{(m)}}{\hat{p}_{i\cdot}^{(m)}}\,, \qquad \widehat{N}_{ij}^{(m)}(3) \equiv N_{\cdot j}(3)\frac{\hat{p}_{ij}^{(m)}}{\hat{p}_{\cdot j}^{(m)}}\,.$$

5. (Right censored data). Suppose that $X, X_1, \ldots, X_n$ are i.i.d. survival times with unknown distribution function F, that $Y, Y_1, \ldots, Y_n$ are i.i.d. censoring times with unknown distribution function G, assumed to be independent of the X_i's, and that we can observe only the iid pairs $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$ where $Z_i \equiv X_i \wedge Y_i$ and $\delta_i \equiv 1_{[X_i \le Y_i]}$; also let $Z \equiv X \wedge Y$ and $\delta = 1_{[X \le Y]}$.

   A. Show that the joint distribution of (Z, δ) is given by

$$H^{(uc)}(z) = P(Z \le z, \delta = 1) = \int_{(0,z]} (1 - G(x-))dF(x)$$

   where $G(x-) \equiv \lim_{y \uparrow x} G(y)$, and

$$H^{(c)}(z) = P(Z \le z, \delta = 0) = \int_{(0,z]} (1 - F(y))dG(y).$$

   Furthermore, show that the survival function $1 - H(z) = P(Z > z)$ is given by $1 - H(z) = (1 - F(z))(1 - G(z))$ and also $H(z) = H^{(uc)}(z) + H^{(c)}(z)$.

   B. Suppose that the cumulative hazard function corresponding to F is defined by

$$\Lambda_F(x) = \int_{[0,x]} \frac{1}{1 - F(y-)} dF(y).$$

   Show that this can be expressed in terms of H and H_{uc} as

$$\Lambda_F(x) = \int_{[0,x]} \frac{1}{1 - H(y-)} dH^{(uc)}(y).$$

   C. If $\mathbb{H}_n^{(uc)}(z) = n^{-1}\sum_{i=1}^n \delta_i 1\{Z_i \le z\}$ and $\mathbb{H}_n(z) = n^{-1}\sum_{i=1}^n 1\{Z_i \le z\}$, suggest an estimator of Λ_F based on the observed (Z_i, δ_i)'s.

   Solution: A. First,

$$\begin{aligned} P(Z \le z, \delta = 1) &= P(X \le z, X \le Y) = E\{1_{[X \le z]} 1_{[X \le Y]}\} \\ &= E\{1_{[X \le z]} E(1_{[X \le Y]}|X)\} = E\{1_{[X \le z]}(1 - G(X-))\} \\ &= \int_{[0,z]} (1 - G(x-))dF(x). \end{aligned}$$

Similarly,

$$\begin{aligned} P(Z \le z, \delta = 0) &= P(Y \le z, Y < X) = E\{1_{[Y \le z]} 1_{[Y \le X]}\} \\ &= E\{1_{[Y \le z]} E(1_{[Y < X]} | Y)\} = E\{1_{[Y \le z]}(1 - F(Y))\} \\ &= \int_{[0,z]} (1 - F(y)) dG(y) \,. \end{aligned}$$

Also note that, using integration by parts,

$$\begin{aligned} H(z) &= P(Z \le z) = \int_{(0,z]} (1 - G(x-)) dF(x) + \int_{[0,z]} (1 - F(y)) dG(y) \\ &= (1-G)F|_{[0,z]} - \int_{[0,z]} F d(1-G) + \int_{(0,z]} (1-F) dG \\ &= (1 - G(z))F(z) + G(z) - \int_{[0,z]} (1-F) dG + \int_{[0,z]} (1-F) dG \\ &= 1 - (1-F(z))(1-G(z)). \end{aligned}$$

B. Using $H^{(uc)}(z) = \int_{[0,z]} (1 - G(x-)) dF(x)$ and $1 - H(z) = (1 - F(z))(1 - G(z))$ we compute

$$\begin{aligned} \int_{[0,x]} \frac{1}{1 - H(y-)} dH^{(uc)}(y) &= \int_{[0,x]} \frac{(1 - G(y-)) dF(y)}{(1 - F(y-))(1 - G(y-))} \\ &= \int_{[0,x]} \frac{1}{1 - F(y-)} dF(y) = \Lambda_F(x). \end{aligned}$$

C. Since the nonparametric MLE's of $H^{(uc)}$ and H are $\mathbb{H}_n^{(uc)}$ and $\mathbb{H}_n$, it follows that the nonparametric MLE of $\Lambda = \Lambda_F$ is

$$\hat{\Lambda}_n(x) = \int_{[0,x]} \frac{1}{1 - \mathbb{H}_n(t-)} d\mathbb{H}_n^{(uc)}(t).$$

As we discussed in class on 1/19 and 1/22, this is the *Nelson-Aalen* estimator of the cumulative hazard function Λ.

6. **Optional bonus problem.** Lehmann and Casella, TPE, Problem 4.15, page 506: For the one-way layout with random effects, the EM algorithm is useful for computing ML estimates (In fact, it is very useful in many mixed models; see Searle et al 1992, Chapter 8.) Suppose we have the model

$$X_{ij} = \mu + A_i + U_{ij}, \qquad j = 1, \ldots, n_i, \; j = 1, \ldots, s,$$

where A_i and U_{ij} are independent normal random variables with mean zero and known variance. To compute the ML estimates of μ, σ_A^2, and σ_U^2, it is typical to employ an EM algorithm using the unobservable A_i's as the augmented data. Write out the E-step and the M-step, and show that the EM sequence converges to the ML estimators.

Solution: First note that since $A_i \sim N(0, \sigma_A^2)$, $i = 1, \ldots, s$ and $U_{i,j} \sim N(0, \sigma_U^2)$, $j = 1, \ldots, n_i$, $j = 1, \ldots, s$ are all independent, it follows that $E(X_{i,j}) = \mu$ for all i, j and

$$\begin{aligned} Cov(X_{i,j}, X_{i',j'}) &= Cov(\mu + A_i + U_{i,j}, \mu + A_{i'} + U_{i',j'}) \\ &= Cov(A_i, A_{i'}) + Cov(U_{i,j}, U_{i',j'}) \\ &= \sigma_A^2 \delta_{i,i'} + \sigma_U^2 \delta_{i,i'} \delta_{j,j'} \\ &= \begin{cases} 0, & \text{if } i \neq i' \\ \sigma_A^2 + \sigma_U^2 \delta_{j,j'}, & \text{if } i = i'. \end{cases} \end{aligned}$$

Thus we see that the vectors $\underline{X}_i \equiv (X_{i,1}, \ldots, X_{i,n_i})$, $i = 1, \ldots, s$ are independent and that

$$\underline{X}_i \sim N_{n_i}(\mu \underline{1}, \sigma_A^2 \underline{1}\underline{1}' + \sigma_U^2 I) \equiv N_{n_i}(\underline{\mu}, \Sigma_i)$$

where $I \equiv I_{n_i}$ is the $n_i \times n_i$ identity matrix and $\underline{1} = \underline{1}_{n_i}$ is the n_i- long vector of all 1's. Thus the joint density of the $X_{i,j}$'s (the "incomplete data") is given by

$$p(\underline{x}_1, \ldots, \underline{x}_s) = \prod_{i=1}^{s} \frac{1}{(2\pi)^{n_i/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2} (\underline{x}_i - \underline{\mu})' \Sigma_i^{-1} (\underline{x}_i - \underline{\mu}) \right).$$

On the other hand, note that $Y_{i,j} \equiv X_{i,j} - A_i = \mu + U_{i,j} \sim N(\mu, \sigma_U^2)$ are i.i.d. and independent of $A_i \sim N(0, \sigma_A^2)$ which are also i.i.d. We take the "complete data" $X = (\underline{Y}, \underline{A})$ where $\underline{Y} \equiv (Y_{i,j}, j = 1, \ldots, n_i, j = 1, \ldots, s)$ and $\underline{A} \equiv (A_1, \ldots, A_s)$. The joint density of $\underline{Y}$ and $\underline{A}$ is given by

$$\begin{aligned} & p(\underline{y}, \underline{a}; \mu, \sigma_A^2, \sigma_U^2) \\ = {} & \prod_{i=1}^{s} \prod_{j=1}^{n_i} \frac{1}{\sigma_U} \phi\left(\frac{y_{ij} - \mu}{\sigma_U} \right) \prod_{i=1}^{s} \frac{1}{\sigma_A} \phi\left(\frac{a_i}{\sigma_U} \right) \\ = {} & \frac{1}{(\sqrt{2\pi}\sigma_U)^N} \exp\left(-\frac{1}{2\sigma_U^2} \sum_{i=1}^{s} \sum_{j=1}^{n_i} (y_{ij} - \mu)^2 \right) \frac{1}{(\sqrt{2\pi\sigma_A^2})^s} \exp\left(-\frac{1}{2\sigma_A^2} \sum_{i=1}^{s} a_i^2 \right) \\ = {} & \frac{1}{(\sqrt{2\pi}\sigma_U)^N} \exp\left(-\frac{1}{2\sigma_U^2} \sum\sum y_{i,j}^2 + \frac{2\mu}{2\sigma_U^2} \sum\sum y_{i,j} - \frac{N\mu^2}{2\sigma_U^2} \right) \\ & \qquad \cdot \frac{1}{(\sqrt{2\pi\sigma_A^2})^s} \exp\left(-\frac{1}{2\sigma_A^2} \sum_{i=1}^{s} a_i^2 \right) \end{aligned}$$

where $N \equiv n_1 + \cdots + n_s$. The MLE's for the complete data are easily seen to be

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{s} \sum_{j=1}^{n_i} Y_{i,j} \equiv Y_{\cdot,\cdot} \equiv \frac{1}{N} T_1,$$

$$\begin{aligned}
\hat{\sigma}_U^2 &= \frac{1}{N}\sum_{i=1}^{s}\sum_{j=1}^{n_i}(Y_{i,j}-\hat{\mu})^2 = \frac{1}{N}\sum_{i=1}^{s}\sum_{j=1}^{n_i}Y_{i,j}^2 - \hat{\mu}^2, \\
&\equiv \frac{1}{N}T_2 - \hat{\mu}^2, \\
\hat{\sigma}_A^2 &= \frac{1}{s}\sum_{i=1}^{s}A_i^2 \equiv \frac{1}{s}T_3.
\end{aligned}$$

The complete data $X = (\{Y_{i,j}\}, \{A_i\})$ is related to the incomplete data Y by

$$X_{i,j} = Y_{i,j} + A_i, \qquad j = 1, \ldots, n_i, \quad i = 1, \ldots, s.$$

To implement an EM-algorithm, we need to compute $E(A_i^2|X)$, $E(Y_{i,j}|X)$, and $E(Y_{i,j}^2|X)$. Since

$$(X_{i,1}, \ldots, X_{i,n_i}, A_i) \sim N_{n_i+1}((\underline{\mu}, 0), \tilde{\Sigma}_i)$$

where

$$\tilde{\Sigma}_i = \begin{pmatrix} \sigma_U^2 I + \sigma_A^2 \underline{1}\underline{1}' & \sigma_A^2 \underline{1} \\ \sigma_A^2 \underline{1}' & \sigma_A^2 \end{pmatrix},$$

it follows from Theorem 1.3.5 that

$$(A_i | X_{i,j}, j = 1, \ldots, n_i) \sim N_1(\sigma_A^2 \underline{1}' \Sigma_i^{-1}(\underline{X}_i - \underline{\mu}), \sigma_A^2 - \sigma_A^2 \underline{1}' \Sigma_i^{-1} \sigma_A^2 \underline{1})$$

where

$$\Sigma_i^{-1} = (\sigma_U^2 I + \sigma_A^2 \underline{1}\ \underline{1}')^{-1} = \frac{1}{\sigma_U^2}\left(I - \frac{1}{n_i + \sigma_U^2/\sigma_A^2}\underline{1}\ \underline{1}'\right) \equiv \frac{1}{\sigma_U^2}\left(I - \frac{1}{r_i}\underline{1}\ \underline{1}'\right).$$

Hence we compute

$$\begin{aligned}
E(A_i^2|Y) &= E(A_i^2 | X_{i,j}, j = 1, \ldots, n_i) \\
&= Var(A_i | X_{i,j}, j = 1, \ldots, n_i) + \{E(A_i | X_{i,j}, j = 1, \ldots, n_i)\}^2 \\
&= \sigma_A^2 \frac{\sigma_A^2/\sigma_U^2}{n_i + \sigma_A^2/\sigma_U^2} \\
&\quad + n_i \frac{\sigma_A^4}{\sigma_U^4}\left\{\frac{1}{n_i}\sum_j (X_{i,j}-\mu)^2 - \left(1 + \frac{2\sigma_U^2/\sigma_A^2}{n_i}\right)\left(\frac{\sum_j (X_{i,j}-\mu)}{n_i + \sigma_U^2/\sigma_A^2}\right)^2\right\}.
\end{aligned}$$

Similarly,

$$(X_{i,1}, \ldots, X_{i,n_i}, Y_{i,1}, \ldots, Y_{i,n_i}) \sim N_{2n_i}((\underline{\mu}, \underline{\mu}), \overline{\Sigma}_i)$$

where

$$\overline{\Sigma}_i = \begin{pmatrix} \Sigma_i & \sigma_U^2 I \\ \sigma_U^2 I & \sigma_U^2 I \end{pmatrix},$$

and hence it follows from Theorem 1.3.5 that

$$\begin{aligned}(Y_{i,j}, j = 1, \ldots, n_i | X_{i,j}, j = 1, \ldots, n_i) &\sim N_{n_i}(\underline{\mu} + \sigma_U^2 \Sigma_i^{-1}(\underline{X}_i - \underline{\mu}), \sigma_U^2 I - \sigma_U^2 \Sigma_i^{-1} \sigma_U^2) \\ &= N_{n_i}(\underline{\mu} + (I - r_i^{-1}\underline{1}\underline{1}')(\underline{X}_i - \underline{\mu}), \sigma_U^2 r_i^{-1} \underline{1}\ \underline{1}').\end{aligned}$$

Therefore,

$$\begin{aligned}E(\underline{Y}_i | \underline{X}_i) &= \mu + (I - r_i^{-1}\underline{1}\underline{1}')(\underline{X}_i - \underline{\mu}) = \underline{X}_i - r_i^{-1} \sum_{j=1}^{n_i} (X_{i,j} - \mu)\underline{1}, \\ |E(\underline{Y}_i | \underline{X}_i)|^2 &= \sum_{j=1}^{n_i} \left(X_{i,j} - r_i^{-1} \sum_{j'=1}^{n_i} (X_{i,j'} - \mu) \right)^2, \\ E(\underline{Y}_i' \underline{Y}_i | \underline{X}_i) &= \sigma_U^2 \frac{n_i}{r_i} + |E(\underline{Y}_i | \underline{X}_i)|^2 \\ &= \sigma_U^2 \frac{n_i}{r_i} + \sum_{j=1}^{n_i} \left(X_{i,j} - r_i^{-1} \sum_{j'=1}^{n_i} (X_{i,j'} - \mu) \right)^2.\end{aligned}$$

It follows that

$$\begin{aligned}E(T_1 | Y) &\equiv \hat{T}_1(Y, \mu, \sigma_A^2, \sigma_U^2) \\ &= \sum_{i=1}^{s} (\underline{X}_i - r_i^{-1} \sum_{j'=1}^{n_i} (X_{i,j'} - \mu)\underline{1}), \qquad (2) \\ E(T_2 | Y) &\equiv \hat{T}_2(Y, \mu, \sigma_A^2, \sigma_U^2) \\ &= \sum_{i=1}^{s} \sigma_U^2 \frac{n_i}{r_i} + \sum_{j=1}^{n_i} (X_{i,j} - r_i^{-1} \sum_{j'=1}^{n_i} (X_{i,j'} - \mu)^2)^2, \qquad (3) \\ E(T_3 | Y) &\equiv \hat{T}_1(Y, \mu, \sigma_A^2, \sigma_U^2) \\ &= \sum_{i=1}^{s} \frac{\sigma_A^4 / \sigma_U^2}{n_i + \sigma_A^2 / \sigma_U^2} \qquad (4) \\ &\quad + \sum_{i=1}^{s} n_i \frac{\sigma_A^4}{\sigma_U^4} \left\{ \frac{1}{n_i} \sum_j (X_{i,j} - \mu)^2 - \left(1 + \frac{2\sigma_U^2 / \sigma_A^2}{n_i} \right) \left(\frac{\sum_j (X_{i,j} - \mu)}{n_i + \sigma_U^2 / \sigma_A^2} \right)^2 \right\},\end{aligned}$$

Thus an E-M algorithm for estimation of $(\mu, \sigma_A^2, \sigma_U^2)$ proceeds as follows: start with some preliminary estimator $(\hat{\mu}^{(0)}, \hat{\sigma}_A^{(0)2}, \hat{\sigma}_U^{(0)2})$, e.g.

$$\begin{aligned}&\hat{\mu}^{(0)} = \overline{X}_{\cdot,\cdot}, \\ &\sigma_A^2 = \frac{1}{N} \sum_{i=1}^{s} n_i (\overline{X}_{i,\cdot} - \overline{X}_{\cdot,\cdot})^2, \\ &\hat{\sigma}_U^2 = \frac{1}{N} \sum_{i=1}^{s} \sum_{j=1}^{n_i} (X_{i,j} - \overline{X}_{i,\cdot})^2.\end{aligned}$$

Then for $m \geq 0$, set $\hat{T}_j^{(m)} = \hat{T}_j(Y, \hat{\mu}^{(m)}, \hat{\sigma}_A^{(m)2}, \hat{\sigma}_U^{(m)2})$, $j = 1, 2, 3$ where the $\hat{T}_j(Y, \mu, \sigma_A^2, \sigma_U^2)$ are given by (2) - (4). This is the E-step. Then then M-step is given by

$$\begin{aligned} \hat{\mu}^{(m+1)} &= N^{-1}\hat{T}_1^{(m)}, \\ \hat{\sigma}_A^{(m+1)2} &= N^{-1}\hat{T}_2^{(m)}, \\ \hat{\sigma}_U^{(m+1)2} &= N^{-1}\hat{T}_3^{(m)}. \end{aligned}$$