Statistics 582, Problem Set 2 Solutions

Wellner; 1/19/2011

1. 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.

2. 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 $N(1), (N_{i\cdot}(2) : 1 \le i \le r), (N_{\cdot j}(3) : 1 \le j \le s)$. Thus $N(1)$ gives cell counts for a two-way table, while $(N_{i\cdot}(2) : 1 \le i \le r)$ and $(N_{\cdot j}(3) : 1 \le j \le s)$ give additional information on the marginal distributions of the table. (If n_2 and n_3 are very large relative to n_1, we might regard the marginal distributions as "known".)
   A. What are the distributions of $N(1), (N_{i\cdot}(2) : 1 \le i \le r)$ and $(N_{\cdot j}(3) : 1 \le j \le s)$?
   B. Find the conditional distribution(s) of X given Y.
   C. Suggest an EM - algorithm for estimation of p.
   Note: This problem was treated by Chen and Fienberg (1974), *Biometrics* **30**, 629-642.

   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

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

then

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

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 $\text{Mult}_s(k_{i\cdot}; (p_{ij}/p_{i\cdot}))$ 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 $(N_{ij}(2) : 1 \le j \le s)|N_{i\cdot} \sim \text{Mult}_s(N_{i\cdot}; (p_{ij}/p_{i\cdot}; j = 1, \ldots, s))$. 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 $(N_{ij}(3) : 1 \le i \le r)|N_{\cdot j} \sim \text{Mult}_r(N_{\cdot j}; (p_{ij}/p_{\cdot j}; i = 1, \ldots, r))$.

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

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

3. Lehmann and Casella, TPE, Problem 4.16, page 506, modified as follows (It seems to me that ζ_i in the third line of the problem statement should be just ζ.)
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.
(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. (I claim that ζ and σ^2 are not identifiable if u is a constant as stated the problem as given.)

Solution: (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

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

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 $L(\zeta,\sigma|\underline{X})$. This is consistent with what we get by writing down the score equations for the incomplete data version of the model: In this case

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

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. 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

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

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

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

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

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. *Method 1: Direct calculation from the (incomplete) data* $\underline{Y} \equiv (N_A, N_B, N_O, N_{AB})$. 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

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

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

$$
\hat{I}(p,q) = N \begin{pmatrix} 9.01584 & 2.47553 \\ 2.47553 & 23.2541 \end{pmatrix}
$$

so that

$$
\hat{I}^{-1}(p,q) = \frac{1}{N}\begin{pmatrix} 0.114256 & -0.0121631 \\ -0.0121631 & 0.044298 \end{pmatrix} = 10^{-3}\cdot\begin{pmatrix} 0.262657 & -0.027961 \\ -0.027961 & 0.101834 \end{pmatrix}.
$$

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

$$
\begin{aligned}
Var(\hat{r}) &= Var(\hat{p}) + Var(\hat{q}) + 2Cov(\hat{p},\hat{q})\,, \\
Cov(\hat{p},\hat{r}) &= -Var(\hat{p}) - Cov(\hat{p},\hat{q})\,, \\
Cov(\hat{q},\hat{r}) &= -Var(\hat{q}) - Cov(\hat{p},\hat{q})\,;
\end{aligned}
$$

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

$$
\widehat{Cov}(\hat{p},\hat{q},\hat{r}) = 10^{-3}\cdot\begin{pmatrix} 0.262657 & -0.027961 & -0.234695 \\ -0.027961 & 0.101834 & -0.073873 \\ -0.234695 & -0.073873 & 0.308569 \end{pmatrix}.
$$

See the end of this solution set for the Mathematica code I used to carry out these calculations.

Method 2: Via Louis's formula
Louis (1982) gives the formula

$$\hat{I}_Y(\theta) = E_\theta\{-\ddot{\mathbf{l}}_{\theta\theta}(\underline{X})|\underline{Y}\} - Cov_\theta(\dot{\mathbf{l}}_\theta(\underline{X}), \dot{\mathbf{l}}_\theta(\underline{X})|\underline{Y}).$$

I will apply this to the complete data model for $\underline{X} = (N_{AA}, N_{AO}, N_{BB}, N_{BO}, N_O, N_{AB})$ parameterized by $\theta = (p, q)$, and hence $r = 1 - p - q$. Thus the scores for p and q are given by

$$\dot{\mathbf{l}}_\theta(\underline{X}) = \begin{pmatrix} \dot{\mathbf{l}}_p(\underline{X}) \\ \dot{\mathbf{l}}_q(\underline{X}) \end{pmatrix} = \begin{pmatrix} \frac{N_p}{p} - \frac{N_r}{r} \\ \frac{N_q}{q} - \frac{N_r}{r} \end{pmatrix}$$

where

$$\begin{aligned} N_p &= 2N_{AA} + N_{AO} + N_{AB}, \\ N_q &= 2N_{BB} + N_{BO} + N_{AB}, \\ N_r &= 2N_{OO} + N_{AO} + N_{BO}. \end{aligned}$$

Furthermore, minus one times the matrix of second derivatives is

$$-\ddot{\mathbf{l}}_{\theta\theta}(\underline{X}) = \begin{pmatrix} \frac{N_p}{p^2} + \frac{N_r}{r^2} & \frac{N_r}{r^2} \\ \frac{N_r}{r^2} & \frac{N_q}{q^2} + \frac{N_r}{r^2} \end{pmatrix}.$$

To compute the terms in Louis's formula we need $E_\theta(\underline{X})|\underline{Y})$ and $Cov_\theta(\underline{X}|\underline{Y})$ where $\underline{Y} = (N_A, N_B, N_O, N_{AB})$. To this end we calculate

$$\begin{aligned} E(N_{AA}|N_A) &= N_A \frac{p^2}{p^2 + 2pr}, \quad E(N_{AO}|N_A) = N_A \frac{2pr}{p^2 + 2pr}, \\ E(N_{BB}|N_A) &= N_B \frac{q^2}{q^2 + 2qr}, \quad E(N_{BO}|N_B) = N_A \frac{2qr}{q^2 + 2qr}. \end{aligned}$$

Furthermore, since the conditional distribution of $\underline{X}$ given $\underline{Y}$ is given by

$$\text{Mult}_2\left(N_A, \left(\frac{p^2}{p^2 + 2pr}, \frac{2pr}{p^2 + 2pr}\right)\right) \cdot \text{Mult}_2\left(N_A, \left(\frac{q^2}{q^2 + 2qr}, \frac{2qr}{q^2 + 2qr}\right)\right) \cdot \delta_{X_5 = Y_3} \cdot \delta_{X_6 = Y_4},$$

the conditional covariance matrix of the complete data scores given $\underline{Y}$ is given by

$$\begin{aligned} & Cov_\theta(\dot{\mathbf{l}}_\theta(\underline{X})|\underline{Y}) \qquad (2) \\ = & \begin{pmatrix} Var_\theta(N_p/p - N_r/r|\underline{Y}) & Cov_\theta(N_p/p - N_r/r, N_q/q - N_r/r|\underline{Y}) \\ Cov_\theta(N_p/p - N_r/r, N_q/q - N_r/r|\underline{Y}) & Var_\theta(N_q/q - N_r/r|\underline{Y}) \end{pmatrix} \\ = & \begin{pmatrix} v_p\left(\frac{1}{p} + \frac{1}{r}\right)^2 + \frac{1}{r^2} v_q & v_p\left(\frac{1}{r^2} + \frac{1}{rp}\right) + v_q\left(\frac{1}{r^2} + \frac{1}{rq}\right) \\ v_p\left(\frac{1}{r^2} + \frac{1}{rp}\right) + v_q\left(\frac{1}{r^2} + \frac{1}{rq}\right) & v_q\left(\frac{1}{q} + \frac{1}{r}\right)^2 + \frac{1}{r^2} v_p \end{pmatrix} \end{aligned}$$

since, by noting that $Var_\theta(N_p|\underline{Y}) = Var_\theta(-N_{AO}|\underline{Y})$ and $Var_\theta(N_q|\underline{Y}) = Var_\theta(-N_{BO}|\underline{Y})$, we have

$$Var_\theta(N_p|\underline{Y}) = N_A \frac{p^2}{p^2+2pr} \cdot \frac{2pr}{p^2+2pr} \equiv v_p, \tag{3}$$

$$Var_\theta(N_q|\underline{Y}) = N_B \frac{q^2}{q^2+2qr} \cdot \frac{2qr}{q^2+2qr} \equiv v_q, \tag{4}$$

$$Var_\theta(N_r|\underline{Y}) = Var_\theta(N_p|\underline{Y}) + Var_\theta(N_q|\underline{Y}),$$
$$Cov_\theta(N_p, N_q)|\underline{Y}) = 0,$$
$$Cov_\theta(N_p, N_r|\underline{Y}) = -Var_\theta(N_p|\underline{Y}),$$
$$Cov_\theta(N_q, N_r|\underline{Y}) = -Var_\theta(N_q|\underline{Y}).$$

Combining these pieces and computing, we find that

$$\hat{I}(p,q) = N \begin{pmatrix} 8.99267 & 2.45797 \\ 2.45797 & 23.2005 \end{pmatrix}$$

so that

$$\hat{I}^{-1}(p,q) = \frac{1}{N} \begin{pmatrix} 0.114518 & -0.0121325 \\ -0.0121325 & 0.0443878 \end{pmatrix} = 10^{-3} \cdot \begin{pmatrix} 0.26326 & -0.027891 \\ -0.027891 & 0.102041 \end{pmatrix}.$$

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

$$\widehat{Cov}(\hat{p}, \hat{q}, \hat{r}) = 10^{-3} \cdot \begin{pmatrix} 0.26326 & -0.027891 & -0.235369 \\ -0.027891 & 0.102041 & -0.074150 \\ -0.235369 & -0.074150 & 0.309095 \end{pmatrix}.$$

See the end of this solution set for the Mathematica code I used to carry out these calculations.

5. Suppose that $X_1, \ldots, X_n$ are i.i.d. exponential(θ) random variables (with density $p_\theta(x) = \theta \exp(-\theta x) 1_{(0,\infty)}(x)$ for $\theta > 0$). Suppose that $(u_1, v_1), \ldots, (u_n, v_n)$ are real numbers with $0 \le u_j < v_j < \infty$ for $j = 1, \ldots, n$ and we observe $Y_1, \ldots, Y_n$ where $Y_i \equiv 1_{[u_i, v_i]}(X_i)$ for $i = 1, \ldots, n$.
(a) Derive the EM algorithm for estimation of θ based on observation of the Y_i's, including a careful treatment of the "E" step.
(b) Give an estimator of the information $I_Y(\theta)$.

Solution: (a) Since the Y_i's are independent Bernoulli($p_i(\theta)$) random variables where

$$p_i(\theta) = P_\theta(u_i \le X_i \le v_i) = P_\theta(X_i \le v_i) - P_\theta(X_i < u_i) = e^{-\theta u_i} - e^{-\theta v_i},$$

for $i = 1, \ldots, n$, it follows that the log-likelihood function for the observed data is

$$l_n(\theta|\underline{Y}) = \sum_{i=1}^{n} \{Y_i \log p_i(\theta) + (1 - Y_i) \log(1 - p_i(\theta))\}$$

On the other hand, the log-likelihood for the complete data $X_1, \ldots, X_n$ is

$$l_n(\theta|\underline{X}) = \sum_{i=1}^{n} \{\log \theta - \theta X_i\},$$

so the MLE of θ for the complete data is $\hat{\theta}$ given by

$$\frac{1}{\hat{\theta}} = \overline{X}_n.$$

Since

$$\begin{aligned} E_{\theta_0}\left\{l_n(\theta|\underline{X})|\underline{Y}\right\} &= n \log \theta - \theta \sum_{i=1}^{n} E_{\theta_0}\{X_i|\underline{Y}\} \\ &= n \log \theta - \theta \sum_{i=1}^{n} E_{\theta_0}\{X_i|Y_i\}, \end{aligned}$$

to carry out the "E" step we just need to compute

$$\begin{aligned} E_{\theta_0}\{X_i|Y_i\} &= Y_i \frac{\int_{u_i}^{v_i} \theta_0 x \exp(-\theta_0 x) dx}{\int_{u_i}^{v_i} \theta_0 \exp(-\theta_0 x) dx} \\ &\quad + (1 - Y_i) \frac{\int_0^{u_i} x\theta_0 \exp(-\theta_0 x) dx + \int_{v_i}^{\infty} x\theta_0 \exp(-\theta_0 x) dx}{\int_0^{u_i} \theta_0 \exp(-\theta_0 x) dx + \int_{v_i}^{\infty} \theta_0 \exp(-\theta_0 x) dx} \\ &= Y_i \left\{ \frac{1}{\theta_0} + \frac{u_i \exp(-\theta_0 u_i) - v_i \exp(-\theta_0 v_i)}{\exp(-\theta_0 u_i) - \exp(-\theta_0 v_i)} \right\} \\ &\quad + (1 - Y_i) \left\{ \frac{1}{\theta_0} + \frac{v_i e^{-\theta_0 v_i} - u_i e^{-\theta_0 u_i}}{1 - e^{-\theta_0 u_i} + e^{-\theta_0 v_i}} \right\}. \end{aligned}$$

Thus if we define

$$\begin{aligned} \hat{X}_i^{(k)} &\equiv \frac{1}{\hat{\theta}^{(k)}} + Y_i \frac{u_i \exp(-\hat{\theta}^{(k)} u_i) - v_i \exp(-\hat{\theta}^{(k)} v_i)}{\exp(-\hat{\theta}^{(k)} u_i) - \exp(-\hat{\theta}^{(k)} v_i)} \\ &\quad + (1 - Y_i) \left\{ \frac{v_i \exp(-\hat{\theta}^{(k)} v_i) - u_i \exp(-\hat{\theta}^{(k)} u_i)}{1 - \exp(-\hat{\theta}^{(k)} u_i) + \exp(-\hat{\theta}^{(k)} v_i)} \right\}, \end{aligned}$$

the "M"- step of the EM algorithm is just

$$\frac{1}{\hat{\theta}^{(k+1)}} = \frac{1}{n}\sum_{i=1}^{n} \hat{X}_i^{(k)}.$$

(b) Louis's formula is:

$$\hat{I}_Y(\theta) = E\{\hat{I}_X(\theta)|\underline{Y}\} - Var_\theta(\dot{\mathbf{l}}_\theta(\underline{X})|\underline{Y}).$$

In the present case,

$$\hat{I}_X(\theta) = -\ddot{l}_n(\theta|\underline{X}) = n\theta^{-2},$$
$$\dot{l}_n(\theta|\underline{X}) = n\theta^{-1} - \sum_{i=1}^{n} X_i,$$

so

$$E\{\hat{I}_X(\theta)|\underline{Y}\} = n\theta^{-2},$$
$$Var_\theta(\dot{\mathbf{l}}_\theta(\underline{X})|\underline{Y}) = \sum_{i=1}^{n} Var_\theta(X_i|Y_i).$$

Since

$$\begin{aligned}
Var_\theta(X_i|Y_i) &= E_\theta(X_i^2|Y_i) - \{E_\theta(X_i|Y_i)\}^2 \\
&= Y_i\left\{\frac{2}{\theta^2} + \frac{e^{-u_i\theta}(2u_i\theta + u_i^2\theta) - e^{-v_i\theta}(2v_i\theta + v_i^2\theta^2)}{\theta^2(e^{-u_i\theta} - e^{-v_i\theta})}\right. \\
&\quad \left. - \left(\frac{1}{\theta} + \frac{u_i e^{-\theta u_i} - v_i e^{-\theta v_i}}{e^{-\theta u_i} - e^{-\theta v_i}}\right)^2\right\} \\
&\quad + (1-Y_i)\left\{\frac{2 + e^{-\theta u_i}(-2 - \theta u_i(2+\theta u_i)) + e^{-\theta v_i}(2 + \theta v_i(2+\theta v_i))}{\theta^2(1 - e^{-\theta u_i + e^{-\theta v_i}})}\right. \\
&\quad \left. - \frac{\left(1 - e^{-\theta u_i}(1+\theta u_i) + e^{-\theta v_i}(1+\theta v_i)\right)^2}{\theta^2\left(1 - e^{-\theta u_i} + e^{-\theta v_i}\right)^2}\right\},
\end{aligned}$$

we find after some algebra that

$$\begin{aligned}
&\frac{1}{\theta^2} - Var_\theta(X_i|Y_i) \\
&= Y_i\frac{(v_i - u_i)^2 e^{-\theta(u_i+v_i)}}{(e^{-\theta u_i} - e^{-\theta v_i})^2} \\
&\quad + (1-Y_i)\frac{e^{-\theta(u_i+v_i)}\left(e^{-\theta u_i}(1+e^{-\theta v_i})u_i^2 - 2u_i v_i e^{-\theta(u_i+v_i)} - e^{-\theta v_i}(1-e^{-\theta u_i})v_i^2\right)}{(1 - e^{-\theta u_i} + e^{-\theta v_i})^2} \\
&\equiv \widehat{Var}(\dot{l}_\theta(Y_i)).
\end{aligned}$$

It follows that

$$\hat{I}_Y(\theta) \;=\; \sum_{i=1}^{n} \widehat{Var}(\dot{l}_\theta(Y_i)).$$

Replacing θ in this expression by $\hat{\theta}^{(k)}$, the k-th iterate of the EM algorithm, yields the estimator $\hat{I}_Y(\hat{\theta}^{(k)}$ of $I_Y(\theta)$. Note that the term of $\widehat{Var}(\dot{l}_\theta(Y_i))$ involving Y_i converges to Y_i/θ^2 as $v_i \to u_i$, while the term of $\widehat{Var}(\dot{l}_\theta(Y_i))$ involving $1 - Y_i$ converges to 0 as $v_i \to u_i$, and hence in this case the information for θ grows as $\theta^{-2} \sum_{i=1}^{n} Y_i$.

A somewhat different (and perhaps more natural?) formulation of this problem is that we observe instead $Y_i \equiv (R_i, S_i, T_i) = (1_{[0,u_i]}(X_i), 1_{(u_i,v_i]}(X_i), 1_{(v_i,\infty)}(X_i))$, $i = 1, \ldots, n$. These Y_i's give somewhat more information, and the resulting EM algorithm differs in the "E" step. In this formulation

$$\begin{aligned} E_\theta(X_i|Y_i) &= R_i E_\theta(X_i|Y_i = (1,0,0)) + S_i E_\theta(X_i|Y_i = (0,1,0)) + T_i E_\theta(X_i|Y_i = (0,0,1)) \\ &= R_i \frac{\int_0^{u_i} x\theta e^{-\theta x} dx}{\int_0^{u_i} x\theta e^{-\theta x} dx} + S_i \frac{\int_{u_i}^{v_i} x\theta e^{-\theta x} dx}{\int_{u_i}^{v_i} x\theta e^{-\theta x} dx} + T_i \frac{\int_{v_i}^{\infty} x\theta e^{-\theta x} dx}{\int_{v_i}^{\infty} x\theta e^{-\theta x} dx}. \end{aligned}$$

A nonparametric version of this formulation of interval censoring (with random u_i's and v_i's) has been studied by Groeneboom and Wellner (1992), Geskus and Groeneboom (1996, 1997), and Groeneboom (1996).