

## Statistics 582, Problem Set 3 Solutions

Wellner; 1/27/2010

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

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

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

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

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

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

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

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

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

$$l_N(p, q) = N_A \log(p^2 + 2p(1 - p - q))$$

$$+ N_B \log(q^2 + 2q(1 - p - q))$$

$$+ N_O \log(1 - p - q)^2 + N_{AB} \log(2pq).$$

Thus

$$-\frac{\partial^2}{\partial p^2} l_N(p, q) = 2N_A \left\{ \frac{1}{2p - p^2 - 2pq} + \frac{2(1 - p - q)^2}{(2p - p^2 - 2pq)^2} \right\}$$

$$+ N_B \frac{4q^2}{(2q - q^2 - 2pq)^2}$$

$$+ \frac{N_{AB}}{p^2} + \frac{2N_O}{(1 - p - q)^2},$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$Var_{\theta}(N_q|\underline{Y}) = N_B \frac{q^2}{q^2 + 2qr} \cdot \frac{2qr}{q^2 + 2qr} \equiv v_q, \quad (3)$$

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

$$Cov_{\theta}(N_p, N_q|\underline{Y}) = 0,$$

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

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

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.

2. Lehmann and Casella, TPE, Problem 4.9, page 504: Consider the following 12 observations from a bivariate normal distribution with parameters  $\mu_1 = \mu_2 = 0$ ,  $\sigma_1^2$ ,  $\sigma_2^2$ , and  $\rho$ :  $(1, 1), (1, -1), (-1, 1), (-1, -1), (2, *), (2, *), (-2, *), (-2, *), (*, 2), (*, 2), (*, -2), (*, -2)$  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. [This example is due to Murray (1977), and is discussed by Wu (1983).]

**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

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

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

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  $-\log \sigma - x^2/(2\sigma^2)$ ,  $-\log \tau - y^2/(2\tau^2)$ , 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  $l_n(\sqrt{8/3}, \sqrt{8/3}, \pm 1/2) = -15.2713\dots$  while  $l_n(\sqrt{5/2}, \sqrt{5/2}, 0) = -15.3303\dots$ . 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  $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$ .

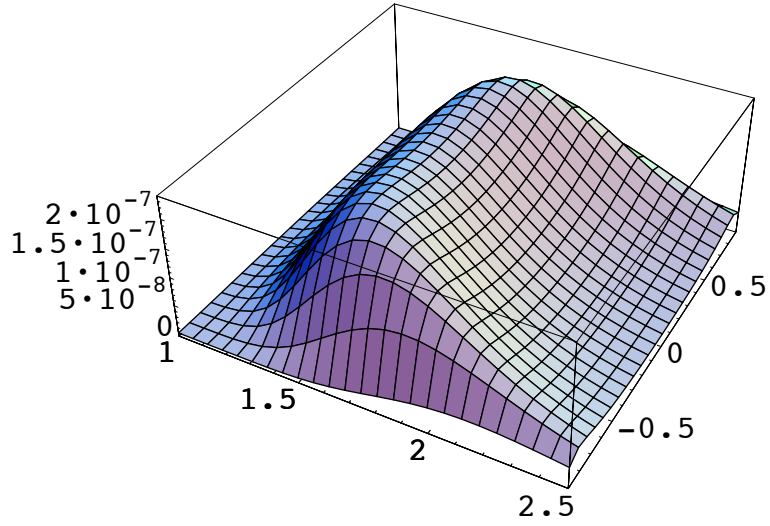


Figure 1: Plot of  $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$ .

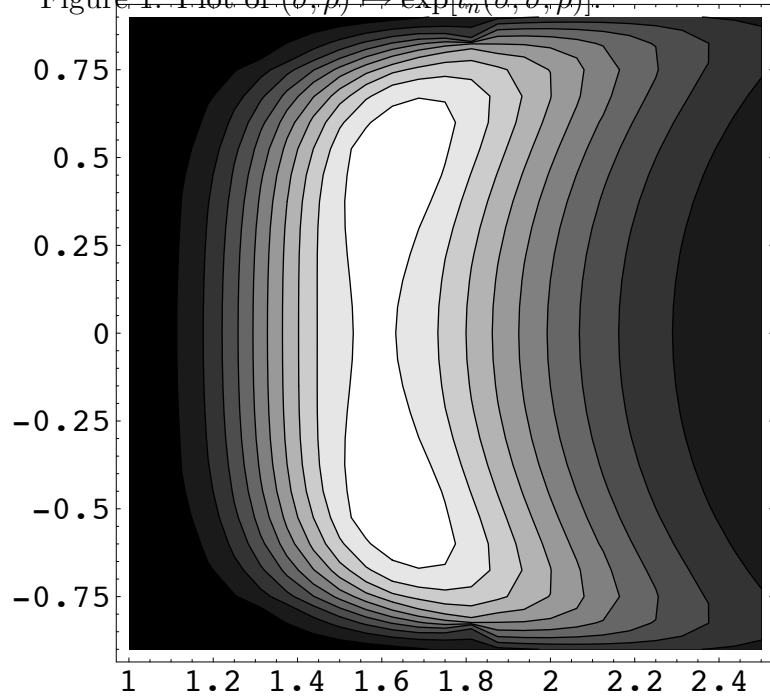


Figure 2: Contour plot of  $(\sigma, \rho) \mapsto \exp[l_n(\sigma, \sigma, \rho)]$ .

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

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

and let the incomplete data be

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

Then, since

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

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

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

where

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

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

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

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

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

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. Lehmann and Casella, TPE, Problem 4.12, page 505.

**Solution:** Suppose  $X_i, i = 1, \dots, n$  are i.i.d. with density  $p_\theta(x) = \theta g(x) + (1-\theta)h(x)$  where  $g$  and  $h$  are known densities with respect to some dominating measure  $\mu$ . Suppose that  $Z_i$  are i.i.d. Bernoulli( $\theta$ ) and that conditionally on  $Z_i$ ,  $(X_i|Z_i)$  has conditional density  $g(x)^{Z_i}h(x)^{1-Z_i}$ .

(a) It follows immediately that the joint density  $q_\theta$  of  $(X_i, Z_i)$  is

$$q_\theta(x_i, z_i) = g(x_i)^{z_i}h(x_i)^{1-z_i}\theta^{z_i}(1-\theta)^{1-z_i},$$

and the joint density of the complete data is

$$q_\theta(\underline{x}, \underline{z}) = \prod_{i=1}^n (\theta g(x_i))^{z_i} ((1-\theta)h(x_i))^{1-z_i} = \prod_{i=1}^n \{z_i g(x_i) + (1-z_i)h(x_i)\} \theta^{z_i} (1-\theta)^{1-z_i}.$$

Thus the complete data likelihood is given by

$$L(\theta|\underline{X}, \underline{Z}) = \prod_{i=1}^n (\theta g(X_i))^{Z_i} ((1-\theta)h(X_i))^{1-Z_i}.$$

(b) To show that

$$E(Z_i|\theta, X_i) = \frac{\theta g(X_i)}{\theta g(X_i) + (1-\theta)h(X_i)}, \quad (4)$$

it suffices to show that

$$E\left\{1_B(X_i) \frac{\theta g(X_i)}{\theta g(X_i) + (1-\theta)h(X_i)}\right\} = E\{1_B(X_i)Z_i\}$$

for all Borel sets  $B$ . But since the marginal density of  $X_i$  is given by  $p_\theta$ , we see that the left side in the last display is equal to  $\theta \int 1_B(x)g(x)d\mu(x)$ , and this equals the right side by direct calculation. Thus (3) holds.

For the complete data  $(X_1, Z_1), \dots, (X_n, Z_n)$ , the MLE of  $\theta$  is  $\hat{\theta}_n = \bar{Z}_n$ . Hence the EM algorithm for estimation of  $\theta$  is given by

$$\hat{\theta}_n^{(j+1)} = \frac{1}{n} \sum_{i=1}^n \hat{Z}_i(\hat{\theta}_n^{(j)}, X_i) \quad (5)$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{\hat{\theta}_n^{(j)} g(X_i)}{\hat{\theta}_n^{(j)} g(X_i) + (1-\hat{\theta}_n^{(j)})h(X_i)}. \quad (6)$$

(c) The expected complete data log-likelihood given the observed (or incomplete) data is

$$\begin{aligned} Q(\theta|\theta_0, \underline{X}) &= E_{\theta_0} \left( \sum_{i=1}^n \{Z_i(\log \theta + \log g(X_i)) + (1-Z_i)(\log(1-\theta) + \log h(X_i))\} \middle| \underline{X} \right) \\ &= \sum_{i=1}^n \left\{ \frac{\theta_0 g(X_i)}{\theta_0 g(X_i) + (1-\theta_0)h(X_i)} (\log \theta + \log g(X_i)) \right. \\ &\quad \left. + \left( 1 - \frac{\theta_0 g(X_i)}{\theta_0 g(X_i) + (1-\theta_0)h(X_i)} \right) (\log \theta + \log g(X_i)) \right\}. \end{aligned}$$

This is a continuous function of  $\theta_0$  and  $\theta \in (0, 1)$ . Thus by Theorem 4.12, Lehmann and Casella page 460, all limit points of an EM sequence  $\{\hat{\theta}^{(j)}\}$  are stationary points of  $L(\theta|\underline{X})$ , and  $L(\hat{\theta}^{(j)}|\underline{X})$  converges monotonically to  $L(\hat{\theta}|\underline{X})$  for some stationary point  $\hat{\theta}$ .

In this case, the log-likelihood function for the incomplete data is (strictly) concave: note that

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

is a concave function of  $\theta$  since the individual terms in the sum are logarithms of linear functions. Here we compute

$$\begin{aligned} \dot{l}_n(\theta) &= \sum_{i=1}^n \frac{g(X_i) - h(X_i)}{\theta g(X_i) + (1 - \theta)h(X_i)} \\ \ddot{l}_n(\theta) &= - \sum_{i=1}^n \frac{(g(X_i) - h(X_i))^2}{(\theta g(X_i) + (1 - \theta)h(X_i))^2} < 0 \end{aligned}$$

with strict inequality if  $g(X_i) \neq h(X_i)$  for some  $i$ . In the latter case, the MLE  $\hat{\theta}$  exists and is the unique stationary point of the log-likelihood. Note that if  $\hat{\theta}^{(j)}$  converges to, say  $\hat{\theta}^{(\infty)}$ , then by (5),  $\hat{\theta}^{(\infty)}$  satisfies

$$\hat{\theta}_n^{(\infty)} = \frac{1}{n} \sum_{i=1}^n \frac{\hat{\theta}_n^{(\infty)} g(X_i)}{\hat{\theta}_n^{(\infty)} g(X_i) + (1 - \hat{\theta}_n^{(\infty)})h(X_i)}.$$

But this is just another way of writing the score equation  $\dot{l}_n(\hat{\theta}^{(\infty)}) = 0$ . Thus  $\hat{\theta}^{(j)} \rightarrow \hat{\theta}$ , the MLE of  $\theta$  based on  $\underline{X}$ .

4. Lehmann and Casella, TPE, Problem 4.16, page 506, modified as follows (It seems to me that  $\zeta_i$  in the third line of the problem statement should be just  $\zeta$ .) We observe independent Bernoulli variables  $X_1, \dots, X_n$  which depend on unobservable variables  $Z_i$  distributed independently as  $N(\zeta, \sigma^2)$  where

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

Assuming that  $u_1, \dots, u_n$  are known, we are interested in obtaining Maximum Likelihood estimates of  $\zeta$  and  $\sigma^2$ .

(a) Show that the likelihood function is  $\prod_{i=1}^n p_i^{X_i} (1 - p_i)^{1 - X_i}$  where  $p_i = P(Z_i > u_i) = \Phi((\zeta - u_i)/\sigma)$ ,  $i = 1, \dots, n$ . You will need to make further appropriate changes in Lehmann and Casella parts (c)-(e) as well. (I claim that  $\zeta$  and  $\sigma^2$  are not identifiable if  $u$  is a constant as stated the problem as given.)

**Solution:** (a) Here  $Z_i \sim N(\zeta, \sigma^2)$ ,  $i = 1, \dots, n$  are i.i.d., and then  $X_i = 1_{(u_i, \infty)}(Z_i)$  for  $i = 1, \dots, n$ . 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  $\theta = (\zeta, \sigma)$ ,  $\theta_0 = (\zeta_0, \sigma_0)$ ,

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

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

$$\begin{aligned} \hat{\zeta} &= \bar{Z} \quad \text{and} \\ \hat{\sigma}^2 &= n^{-1} \sum_{i=1}^n (Z_i - \bar{Z})^2 = n^{-1} \sum_{i=1}^n Z_i^2 - \bar{Z}_n^2, \end{aligned}$$

it follows that the EM sequence is given by

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

where

$$\begin{aligned} t(\zeta, \sigma, X, u) &= E(Z | u, X, \zeta, \sigma), \\ v(\zeta, \sigma, X, u) &= E(Z^2 | u, X, \zeta, \sigma). \end{aligned}$$

(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

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

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

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

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

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

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

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

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

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

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  $\zeta$  and  $\sigma$ , note that  $t(\zeta, \sigma, X, u)$  and  $v(\zeta, \sigma, X, u)$  are continuous functions of  $\zeta$  and  $\sigma$ , 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  $\zeta$  and  $\sigma$ . Thus it seems that the  $u_i$ 's need to take on at least two distinct values in order for both  $\zeta$  and  $\sigma$  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  $Z_i$ ).

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

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

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

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

Suppose that the "incomplete data"  $Y$  consists of  $N(1)$ ,  $(N_{i.}(2) : 1 \leq i \leq r)$ ,  $(N_{.j}(3) : 1 \leq j \leq s)$ . Thus  $N(1)$  gives cell counts for a two-way table, while  $(N_{i.}(2) : 1 \leq i \leq$

$r$ ) and  $(N_{.j}(3) : 1 \leq j \leq s)$  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  $N(1)$ ,  $(N_i(2) : 1 \leq i \leq r)$  and  $(N_{.j}(3) : 1 \leq j \leq s)$ ?

B. Find the conditional distribution(s) of  $X$  given  $Y$ .

C. Suggest an EM - algorithm for estimation of  $p$ .

**Solution:** A. By elementary considerations,

$$(N_i(2) : 1 \leq i \leq r) \sim \text{Mult}_r(n_2; (p_{i.} : 1 \leq i \leq r))$$

and

$$(N_{.j}(3) : 1 \leq j \leq s) \sim \text{Mult}_s(n_3; (p_{.j} : 1 \leq j \leq s)).$$

B. First note that if

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

then

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

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

$$((N_{ij})|(N_{i.})) \sim \prod_{i=1}^r \text{Mult}_s(N_{i.}; (p_{ij}/p_{i.})). \quad (7)$$

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

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

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

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

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  $\hat{p} = (\hat{p}_{ij})$  of  $p = (p_{ij})$  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

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

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

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

where

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

Mathematica code used in problem 1C, method 1:

```

pA[p_, q_] := p^2 + 2*p*(1 - p - q)
pB[p_, q_] := q^2 + 2*q*(1 - p - q)
p0[p_, q_] := (1 - p - q)^2
pAB[p_, q_] := 2*p*q

LpA[p_, q_] := Log[pA[p, q]]
LpB[p_, q_] := Log[pB[p, q]]
Lp0[p_, q_] := Log[p0[p, q]]
LpAB[p_, q_] := Log[pAB[p, q]]
LL[p_, q_, nA_, nB_, n0_, nAB_]
    := nA*LpA[p, q] + nB*LpB[p, q] + n0*Lp0[p, q] + nAB*LpAB[p, q]

LL10[p_, q_, nA_, nB_, n0_, nAB_]
    := Derivative[1, 0, 0, 0, 0, 0][LL][p, q, nA, nB, n0, nAB]
LL01[p_, q_, nA_, nB_, n0_, nAB_]
    := Derivative[0, 1, 0, 0, 0, 0][LL][p, q, nA, nB, n0, nAB]
LL20[p_, q_, nA_, nB_, n0_, nAB_]
    := Derivative[2, 0, 0, 0, 0, 0][LL][p, q, nA, nB, n0, nAB]
LL02[p_, q_, nA_, nB_, n0_, nAB_]
    := Derivative[0, 2, 0, 0, 0, 0][LL][p, q, nA, nB, n0, nAB]
LL11[p_, q_, nA_, nB_, n0_, nAB_]
    := Derivative[1, 1, 0, 0, 0, 0][LL][p, q, nA, nB, n0, nAB]

LL10[p, q, nA, nB, n0, nAB]
LL01[p, q, nA, nB, n0, nAB]
LL20[p, q, nA, nB, n0, nAB]
LL02[p, q, nA, nB, n0, nAB]
LL11[p, q, nA, nB, n0, nAB]

EnA[p_, q_] := p^2 + 2*p*(1 - p - q)
EnB[p_, q_] := q^2 + 2*q*(1 - p - q)
En0[p_, q_] := (1 - p - q)^2
EnAB[p_, q_] := 2*p*q

ELL20[p_, q_] := LL20[p, q, EnA[p, q], EnB[p, q], En0[p, q], EnAB[p, q]]
ELL02[p_, q_] := LL02[p, q, EnA[p, q], EnB[p, q], En0[p, q], EnAB[p, q]]
ELL11[p_, q_] := LL11[p, q, EnA[p, q], EnB[p, q], En0[p, q], EnAB[p, q]]

ELL20[p, q]
ELL02[p, q]
ELL11[p, q]

```

a = .264

```

b = .093

-ELL20[a, b]
-ELL02[a, b]
-ELL11[a, b]

-ELL20[a, b]/2
-ELL02[a, b]/2
-ELL11[a, b]/2

IfM = {{-ELL20[a, b], -ELL11[a, b]}, {-ELL11[a, b], -ELL02[a, b]}}
IfMI = Inverse[IfM]

Ntot = 176 + 182 + 60 + 17
IfMIN = IfMI/Ntot

Vr = (IfMI[[1, 1]] + IfMI[[2, 2]] + 2*IfMI[[1, 2]])/Ntot
CovPR = (-IfMI[[1, 1]] - IfMI[[1, 2]])/Ntot
CovQR = (-IfMI[[2, 2]] - IfMI[[1, 2]])/Ntot

IfMINwR = {{IfMIN[[1, 1]], IfMIN[[1, 2]], CovPR},
           {IfMIN[[1, 2]], IfMIN[[2, 2]], CovQR},
           {CovPR, CovQR, Vr}}

```

**Mathematica code used in problem 1C, method 2:**

```

MinHCX[p_, q_, r_, n0_, nAA_, nAO_, nBB_, nBO_, nAB_] :=
  {{(2*nAA + nAO + nAB)/p^2 + (2*n0 + nAO + nBO)/
    r^2, (2*n0 + nAO + nBO)/r^2}, {(2*n0 + nAO + nBO)/
    r^2, (2*nBB + nBO + nAB)/q^2 + (2*n0 + nAO + nBO)/r^2}}

EnAA[nA_, p_, q_, r_] := nA*p^2/(p^2 + 2*p*r)
EnAO[nA_, p_, q_, r_] := nA*2*p*r/(p^2 + 2*p*r)
EnBB[nB_, p_, q_, r_] := nB*q^2/(q^2 + 2*q*r)
EnBO[nB_, p_, q_, r_] := nB*2*q*r/(q^2 + 2*q*r)

EMinHCY[p_, q_, r_, nA_, nB_, n0_, nAB_] :=
  MinHCX[p, q, r, n0, EnAA[nA, p, q, r], EnAO[nA, p, q, r],
    EnBB[nB, p, q, r], EnBO[nB, p, q, r], nAB]

VarNp[nA_, p_, q_, r_] := nA*2*p^3*r/(p^2 + 2*p*r)^2
VarNq[nB_, p_, q_, r_] := nB*2*q^3*r/(q^2 + 2*q*r)^2
VarNr[nA_, nB_, p_, q_, r_] :=
  nA*2*p^3*r/(p^2 + 2*p*r)^2 + nB*2*q^3*r/(q^2 + 2*q*r)^2

```

```

CovNpr[nA_, p_, q_, r_] := -nA*2*p^3*r/(p^2 + 2*p*r)^2
CovNqr[nB_, p_, q_, r_] := -nB*2*q^3*r/(q^2 + 2*q*r)^2

```

```

EMinHCXgY[p_, q_, r_, nA_, nB_, nO_,
nAB_] := {{VarNp[nA, p, q, r]*(p^(-1) + r^(-1))^2 +
  VarNq[nB, p, q, r]/r^2,
  VarNp[nA, p, q, r]*(r^(-2) + p^(-1)*r^(-1)) +
  VarNq[nB, p, q, r]*(r^(-2) + q^(-1)*r^(-1))}, {VarNp[nA, p, q,
  r]*(r^(-2) + p^(-1)*r^(-1)) +
  VarNq[nB, p, q, r]*(r^(-2) + q^(-1)*r^(-1)),
  VarNq[nB, p, q, r]*(q^(-1) + r^(-1))^2 + VarNp[nA, p, q, r]/r^2}}

```

```

InfEst2[p_, q_, r_, nA_, nB_, nO_, nAB_] :=
  EMinHCY[p, q, r, nA, nB, nO, nAB] -
  EMinHCXgY[p, q, r, nA, nB, nO, nAB]

```

```

a = .264
b = .093
c = .642
NA = 182
NO = 176
NB = 60
NAB = 17
Ntot = NA + NB + NO + NAB

```

```

NInfEst2 = InfEst2[a, b, c, NA, NB, NO, NAB]/Ntot
IfMI = Inverse[NInfEst2]

```

```

Vr = (IfMI[[1, 1]] + IfMI[[2, 2]] + 2*IfMI[[1, 2]])
CovPR = (-IfMI[[1, 1]] - IfMI[[1, 2]])
CovQR = (-IfMI[[2, 2]] - IfMI[[1, 2]])

```

```

IfMINwR = {{IfMI[[1, 1]], IfMI[[1, 2]], CovPR}, {IfMI[[1, 2]],
  IfMI[[2, 2]], CovQR}, {CovPR, CovQR, Vr}}
IfMINwR/Ntot

```

```

Out[183]=  $\frac{nAB}{p} - \frac{2 nO}{1 - p - q} + \frac{2 nA (1 - p - q)}{p^2 + 2 p (1 - p - q)} - \frac{2 nB q}{2 (1 - p - q) q + q^2}$ 
Out[184]=  $-\frac{2 nA p}{p^2 + 2 p (1 - p - q)} - \frac{2 nO}{1 - p - q} + \frac{nAB}{q} + \frac{2 nB (1 - p - q)}{2 (1 - p - q) q + q^2}$ 
Out[185]=  $-\frac{nAB}{p^2} - \frac{2 nA}{p^2 + 2 p (1 - p - q)} - \frac{2 nO}{(1 - p - q)^2} - \frac{4 nA (1 - p - q)^2}{(p^2 + 2 p (1 - p - q))^2} - \frac{4 nB q^2}{(2 (1 - p - q) q + q^2)^2}$ 
Out[186]=  $-\frac{4 nA p^2}{(p^2 + 2 p (1 - p - q))^2} - \frac{2 nO}{(1 - p - q)^2} - \frac{nAB}{q^2} - \frac{4 nB (1 - p - q)^2}{(2 (1 - p - q) q + q^2)^2} - \frac{2 nB}{2 (1 - p - q) q + q^2}$ 
Out[187]=  $-\frac{2 nA}{p^2 + 2 p (1 - p - q)} - \frac{2 nO}{(1 - p - q)^2} + \frac{4 nA p (1 - p - q)}{(p^2 + 2 p (1 - p - q))^2} + \frac{4 nB (1 - p - q) q}{(2 (1 - p - q) q + q^2)^2} - \frac{2 nB}{2 (1 - p - q) q + q^2}$ 
Out[195]=  $-4 - \frac{4 (1 - p - q)^2}{p^2 + 2 p (1 - p - q)} - \frac{2 q}{p} - \frac{4 q^2}{2 (1 - p - q) q + q^2}$ 
Out[196]=  $-4 - \frac{4 p^2}{p^2 + 2 p (1 - p - q)} - \frac{2 p}{q} - \frac{4 (1 - p - q)^2}{2 (1 - p - q) q + q^2}$ 
Out[197]=  $-6 + \frac{4 p (1 - p - q)}{p^2 + 2 p (1 - p - q)} + \frac{4 (1 - p - q) q}{2 (1 - p - q) q + q^2}$ 
Out[198]= 0.264
Out[199]= 0.093
Out[200]= 9.01584
Out[201]= 23.2541
Out[202]= 2.47553
Out[203]= 4.50792
Out[204]= 11.6271
Out[205]= 1.23776
Out[206]= {{9.01584, 2.47553}, {2.47553, 23.2541}}
Out[207]= {{0.114256, -0.0121631}, {-0.0121631, 0.044298}}
Out[208]= 435
Out[209]= {{0.000262657, -0.0000279612}, {-0.0000279612, 0.000101834}}
Out[210]= 0.000308569
Out[211]= -0.000234695
Out[212]= -0.0000738732
Out[213]= {{0.000262657, -0.0000279612, -0.000234695},
{-0.0000279612, 0.000101834, -0.0000738732}, {-0.000234695, -0.0000738732, 0.000308569}}

```

Figure 3: Output from Mathematica code, method 1

```

Out[165]= 0.264
Out[166]= 0.093
Out[167]= 0.642
Out[168]= 182
Out[169]= 176
Out[170]= 60
Out[171]= 17
Out[172]= 435
Out[173]= {{8.99267, 2.45797}, {2.45797, 23.2005}}
Out[174]= {{0.114518, -0.0121325}, {-0.0121325, 0.0443878}}
Out[175]= 0.134641
Out[176]= -0.102385
Out[177]= -0.0322553
Out[178]= {{0.114518, -0.0121325, -0.102385},
           {-0.0121325, 0.0443878, -0.0322553}, {-0.102385, -0.0322553, 0.134641}}
Out[179]= {{0.00026326, -0.0000278909, -0.000235369},
           {-0.0000278909, 0.000102041, -0.0000741501}, {-0.000235369, -0.0000741501, 0.000309519}}

```

Figure 4: Output from Mathematica code, method 2