ST 790, Homework 2 Solutions
Spring 2017

1. (a) Let $U_{gh} = \sum_{i=1}^{N} I(Y_{i1} = g, Y_{i2} = h)$. The joint density of the U_{gh}, $g = 1, \ldots, G$, $h = 1, \ldots, H$, is the multinomial density

$$\frac{N!}{\prod_{g=1}^{G}\prod_{h=1}^{H} U_{gh}!}\prod_{g=1}^{G}\prod_{h=1}^{H}\theta_{gh}^{U_{gh}}.$$

Taking logarithms and ignoring the constant term, we thus want to maximize

$$\sum_{g=g}^{G}\sum_{h=1}^{H} U_{gh}\log(\theta_{gh}) \quad \text{subject to} \quad \sum_{g=g}^{G}\sum_{h=1}^{H}\theta_{gh} = 1.$$

Most of you imposed the constraint by using a Lagrange multiplier approach. Letting

$$L(\theta) = \sum_{g=g}^{G}\sum_{h=1}^{H} U_{gh}\log(\theta_{gh}) - \lambda\left(\sum_{g=g}^{G}\sum_{h=1}^{H}\theta_{gh} - 1\right),$$

taking derivatives of $L(\theta)$ with respect to each θ_{gh} and λ and setting equal to zero yields

$$\theta_{gh} = \frac{U_{gh}}{\lambda}, \quad g = 1, \ldots, G, h = 1, \ldots, H,$$

and

$$\sum_{g=g}^{G}\sum_{h=1}^{H}\theta_{gh} = 1.$$

Using the fact that

$$\sum_{g=g}^{G}\sum_{h=1}^{H} U_{gh} = N,$$

we get

$$\lambda = N,$$

from whence it follows that

$$\widehat{\theta}_{gh} = \frac{U_{gh}}{N}, \quad g = 1, \ldots, G, h = 1, \ldots, H.$$

(b) Clearly, from the full data loglikelihood, $U_{gh} = \sum_{i=1}^{N} I(Y_{i1} = g, Y_{i2} = h)$ is a sufficient statistic for θ_{gh}, $g = 1, \ldots, G$, $h = 1, \ldots, H$. Accordingly, from pages 68-69 of the notes, the E-step involves finding for each $g = 1, \ldots, G$, $h = 1, \ldots, H$

$$E\{I(Y_1 = g, Y_2 = h)|R, Z_{(R)}\}.$$

Here, R takes on values $r = (1,1)^T$, $(1,0)^T$, or $(0,1)^T$. Trivially, when $R = (1,1)^T$, $Z_{(R)} = (Y_1, Y_2)$, and

$$E\{I(Y_1 = g, Y_2 = h)|R, Z_{(R)}\} = I(Y_1 = g, Y_2 = h).$$

When $R = (1,0)^T$, $Z_{(R)} = Y_1$, and

$$
\begin{aligned}
E\{I(Y_1 = g, Y_2 = h)|R, Z_{(R)}\} &= E\{I(Y_1 = g)I(Y_2 = h)|Y_1 = g)\} \\
&= I(Y_1 = g)E\{I(Y_2 = h)|Y_1 = g)\} = I(Y_1 = g)\mathrm{pr}(Y_2 = h|Y_1 = g) \\
&= I(Y_1 = g)\frac{\theta_{gh}}{\theta_{g\cdot}},
\end{aligned}
$$

using $\theta_{gh} = \mathrm{pr}(Y_1 = g, Y_2 = h)$ and $\mathrm{pr}(Y_1 = g) = \theta_{g\cdot} = \sum_{h=1}^{H} \theta_{gh}$. Analogously, when $R = (0,1)^T$,

$$
E\{I(Y_1 = g, Y_2 = h)|R, Z_{(R)}\} = I(Y_2 = h)E\{I(Y_1 = g)|Y_2 = h\} = I(Y_2 = h)\frac{\theta_{gh}}{\theta_{\cdot h}},
$$

where $\theta_{\cdot h} = \sum_{g=1}^{G} \theta_{gh}$. Thus, given the tth iterate $\theta_{gh}^{(t)}$ for $g = 1, \dots, G,$, $h = 1, \dots, H$, and defining $\theta_{g\cdot}^{(t)}$ and $\theta_{\cdot h}^{(t)}$ in the obvious way, the E-step involves calculating for each $i = 1, \dots, N$

$$
\theta_{gh,i}^{(t+1)} = R_{i1}R_{i2}I(Y_{i1} = g, Y_{i2} = h) + R_{i1}(1 - R_{i2})I(Y_{i1} = g)\frac{\theta_{gh}^{(t)}}{\theta_{g\cdot}^{(t)}} + (1 - R_{i1})R_{i2}I(Y_{i2} = h)\frac{\theta_{gh}^{(t)}}{\theta_{\cdot h}^{(t)}}.
$$

The M-step is then

$$
\theta_{gh}^{(t+1)} = N^{-1}\sum_{i=1}^{N} \theta_{gh,i}^{(t+1)}, \quad g = 1, \dots, G, h = 1, \dots, H.
$$

(c) See attached R program and output for my inefficient implementation.

2. We wish to show that

$$
E_{\theta'}\Big[\log\{p_Z(Z_i;\theta)\}|R_i, Z_{(R_i)i}\Big] = \sum_r I(R_i = r)\, E_{\theta'}\Big[\log\{p_Z(Z_i;\theta)\}|Z_{(r)i}\Big].
$$

By definition,

$$
E_{\theta'}\Big[\log\{p_Z(Z_i;\theta)\}|R_i, Z_{(R_i)i}\Big] = \sum_r I(R_i = r)\, E_{\theta'}\Big[\log\{p_Z(Z_i;\theta)\}|R_i = r, Z_{(r)i}\Big]. \tag{1}
$$

The expectation of a summand of the right hand side of (1), with fixed r and suppressing i, is equal to

$$
\begin{aligned}
&\int \log\{p_Z(Z_{(r)}, z_{(\bar r)}; \theta)\}\, p_{Z|R,Z_{(R)}}(Z_{(r)}, z_{(\bar r)}|r, Z_{(r)}; \theta', \psi)\, d\nu(z_{(\bar r)}) \\
&= \int \log\{p_Z(Z_{(r)}, z_{(\bar r)}; \theta)\} \frac{p_{R,Z}(r, Z_{(r)}, z_{(\bar r)}; \theta', \psi)}{\int p_{R,Z}(r, Z_{(r)}, z_{(\bar r)}; \theta', \psi)\, d\nu(z_{(\bar r)})}\, d\nu(z_{(\bar r)}) \\
&= \int \log\{p_Z(Z_{(r)}, z_{(\bar r)}; \theta)\} \frac{p_{R|Z}(r|Z_{(r)}, z_{(\bar r)}; \psi) p_Z(Z_{(r)}, z_{(\bar r)}; \theta')}{\int p_{R|Z}(r|Z_{(r)}, z_{(\bar r)}; \psi) p_Z(Z_{(r)}, z_{(\bar r)}; \theta')\, d\nu(z_{(\bar r)})}\, d\nu(z_{(\bar r)}) \\
&= \int \log\{p_Z(Z_{(r)}, z_{(\bar r)}; \theta)\} \frac{p_Z(Z_{(r)}, z_{(\bar r)}; \theta')}{\int p_Z(Z_{(r)}, z_{(\bar r)}; \theta')\, d\nu(z_{(\bar r)})}\, d\nu(z_{(\bar r)}) \\
&= \int \log\{p_Z(Z_{(r)}, z_{(\bar r)}; \theta)\} \int p_Z(Z_{(r)}, z_{(\bar r)}; \theta')\, d\nu(z_{(\bar r)})\, d\nu(z_{(\bar r)}) \\
&= \int \log\{p_Z(Z_{(r)}, z_{(\bar r)}; \theta)\}\, p_{Z|Z_{(r)}}(Z_{(r)}, z_{(\bar r)}|Z_{(r)}; \theta')\, d\nu(z_{(\bar r)}) \\
&= E_{\theta'}\Big[\log\{p_Z(Z; \theta)\} | Z_{(r)} \Big]
\end{aligned}
$$

as required.

3. See the attached SAS and R programs and output.

   (a) To get the summary of missingness patterns in R, I used the `md.pattern` function in the `mice` package. In SAS, `proc mi` prints out a summary automatically. You might have used something different to summarize/visualize the missingness patterns.

   (b)-(d) It is necessary to massage and reconfigure the data set to use SAS `proc mixed` or R `gls` to fit the model as parameterized in (4), namely

$$
Y_{ij} = \mu_{0j} + \beta_j A_i + \epsilon_{ij}, \quad \epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{i5})^T \sim \mathcal{N}(0, \Sigma).
$$

   Here, the model is parameterized in terms of the means for treatment 0 (placebo) at each of the 5 time points (μ_{0j}, $j = 1, \dots, 5$) and the differences in means between treatments 1 and 0 (active and placebo) at each of the time points (β_j, $j = 1, \dots, 5$). See the programs/output for the calls to SAS `proc mixed` and R `gls`. Recall as we discussed in class that `corSymm` specifies the unstructured correlation structure. `gls` assumes by default that all outcomes have the *same variance*, so to generalize this to have different variances at each week/time, we also need to use the `weights = varIdent` option as shown in the program.

   The results for (b)-(d) are in the attached.

   (e) The estimates of β_5 from the three analyses (SAS) are

	Est	SE	P-value
Available	-5.624	2.534	0.03
Complete Case	-4.710	2.700	0.08
LOCF	-3.698	2.426	0.13

   Note: The standard errors from `proc mixed` and `gls()` are slightly different. This is due to a bug in `gls()` that has never been corrected. The standard errors from SAS are correct. On the `Examples` link on the Longitudinal Data Analysis course you will find sample R programs using `gls()` that include a function I wrote to calculate the correct ("model-based") standard errors. The discrepancy is sufficiently minor to not worry about it here.

Of course, none of the standard errors reported here are reliable. As in Section 3.5, the standard errors for the available case analysis are based on the expected rather than observed information matrix, so they may or may not be a reasonable reflection of the true sampling variation. For the complete case and LOCF analyses, the standard errors are most likely meaningless. Accordingly, getting too worked up about the resulting test statistics and p-values may not be advisable.

That said, for the available case analysis based on all observed data, which is the appropriate observed data likelihood analysis *under the assumptions of MAR and multivariate normality* there is evidence supporting a mean difference at 52 weeks (level 0.05 test) if the standard errors are not "too terrible." The p-values for the complete case and LOCF analyses are likely fairly meaningless; these naive analyses do yield point estimates that are attenuated relative to that from the available observed data analysis.

(f) See attached.

(g) See attached.

(h) The estimates from the available observed data likelihood analyses for each treatment group obtained from `gls` or `proc mixed` are virtually identical. This is entirely expected, and we'd suspect some sort of error if this were not the case. Both analyses are doing the same thing – maximizing the observed data likelihood. `gls` and `proc mixed` are carrying out the maximization using standard optimization techniques. The EM algorithm is just another numerical technique to maximize the same likelihood.

4. (a) Defining $C = 1$ if $R = (1, 1)$ and $C = 0$ if $R = (1, 0)$ as in Section 3.5, the observed data loglikelihood is as in (3.72),

$$\ell = (1/2) \sum_{i=1}^{N} \Big[I(C_i = 1)\{-\log|\Sigma| - (Y_i - \mu)^T \Sigma^{-1} (Y_i - \mu)\} + I(C_i = 0)\{-\log \sigma_1^2 - (Y_{i1} - \mu_1)^2/\sigma_1^2\} \Big].$$

By direct differentiation, we obtain

$$\frac{\partial \ell}{\partial \mu} = \sum_{i=1}^{N} \left\{ I(C_i = 1)\Sigma^{-1}(Y_i - \mu) + I(C_i = 0) \begin{pmatrix} (Y_{i1} - \mu_1)/\sigma_1^2 \\ 0 \end{pmatrix} \right\}. \tag{2}$$

Accordingly, we have

$$S_\mu(R, Z_{(R)}) = I(C = 1)\Sigma^{-1}(Y - \mu) + I(C = 0) \begin{pmatrix} (Y_1 - \mu_1)/\sigma_1^2 \\ 0 \end{pmatrix}. \tag{3}$$

(b) The full data loglikelihood involves N terms like the first one in the observed data loglikelihood, so it follows immediately that

$$S_\mu(Z) = \Sigma^{-1}(Y - \mu). \tag{4}$$

We thus want to show that the expectation of (4) conditional on $(R, Z_{(R)})$ is equal to (3). Now by definition

$$E\{S_\mu(Z)|R, Z_{(R)}\} = I(C = 1)E\{\Sigma^{-1}(Y - \mu)|C = 1, Y\} + I(C = 0)E\{\Sigma^{-1}(Y - \mu)|C = 0, Y_1\} \tag{5}$$

The first term on the right hand side of (5) is immediately equal to

$$I(C=1)\Sigma^{-1}(Y-\mu),$$

so is equal to the first term on the right hand side of (3). So consider the second term in (5), which can be written

$$\begin{aligned} & I(C=0)\Sigma^{-1}\begin{pmatrix} Y_1-\mu_1 \\ E(Y_2-\mu_2|C=0,Y_1) \end{pmatrix} \\ &= I(C=0)\Sigma^{-1}\begin{pmatrix} Y_1-\mu_1 \\ E(Y_2-\mu_2|Y_1) \end{pmatrix} && (6) \\ &= I(C=0)\Sigma^{-1}\begin{pmatrix} Y_1-\mu_1 \\ \frac{\sigma_{12}}{\sigma_1^2}(Y_1-\mu_1) \end{pmatrix}, && (7) \end{aligned}$$

where (6) follows by MAR ($C \perp\!\!\!\perp Y_2|Y_1$) and (7) follows by the standard result that

$$E(Y_2|Y_1) = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(Y_1-\mu_1)$$

for a bivariate normal. Using the fact that

$$\Sigma^{-1} = (\sigma_1^2\sigma_2^2-\sigma_{12}^2)^{-1}\begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix},$$

it follows by straightforward matrix multiplication that

$$\begin{aligned} \Sigma^{-1}\begin{pmatrix} Y_1-\mu_1 \\ \frac{\sigma_{12}}{\sigma_1^2}(Y_1-\mu_1) \end{pmatrix} &= (\sigma_1^2\sigma_2^2-\sigma_{12}^2)^{-1}\begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}\begin{pmatrix} 1 \\ \frac{\sigma_{12}}{\sigma_1^2} \end{pmatrix}(Y_1-\mu_1) \\ &= \begin{pmatrix} \frac{Y_1-\mu_1}{\sigma_1^2} \\ 0 \end{pmatrix}, && (8) \end{aligned}$$

so that the second term on the right hand side of (5) is equal to the second term on the right hand side of (3), as required.

(c) (i) Differentiating (2) with respect to σ_{12} gives

$$-\frac{\partial^2\ell}{\partial\mu\partial\sigma_{12}} = \sum_{i=1}^{N} I(C_i=1)\frac{\partial\Sigma^{-1}}{\partial\sigma_{12}}(Y_i-\mu).$$

Using the standard matrix differentiation result that

$$\frac{\partial\Sigma^{-1}}{\partial\sigma_{12}} = -\Sigma^{-1}\frac{\partial\Sigma}{\partial\sigma_{12}}\Sigma^{-1},$$

where

$$\frac{\partial\Sigma}{\partial\sigma_{12}}$$

is the (2×2) matrix of partial derivatives of the elements of Σ with respect to σ_{12}, which is clearly equal to

$$E_{12} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

we obtain (3.77).

For (ii), the expectation of a summand of (3.77) is

$$\Sigma^{-1}E_{12}\Sigma^{-1}E\left\{I(C=1)\begin{pmatrix} Y_1-\mu_1 \\ Y_2-\mu_2 \end{pmatrix}\right\} = \Sigma^{-1}E_{12}\Sigma^{-1}E\left[E\left\{I(C=1)\begin{pmatrix} Y_1-\mu_1 \\ Y_2-\mu_2 \end{pmatrix}\middle| C\right\}\right]. \tag{9}$$

The expectation in (9) is, by definition

$$\begin{aligned} & E\left\{I(C=1)\begin{pmatrix} Y_1-\mu_1 \\ Y_2-\mu_2 \end{pmatrix}\middle| C=1\right\}\pi + E\left\{I(C=1)\begin{pmatrix} Y_1-\mu_1 \\ Y_2-\mu_2 \end{pmatrix}\middle| C=0\right\}(1-\pi) \\ & = E\left\{\begin{pmatrix} Y_1-\mu_1 \\ Y_2-\mu_2 \end{pmatrix}\middle| C=1\right\}\pi \end{aligned} \tag{10}$$

because the second term is equal to zero. Thus, the expectation of (3.77) is

$$\pi\Sigma^{-1}E_{12}\Sigma^{-1}\begin{pmatrix} E(Y_1|C=1)-\mu_1 \\ E(Y_2|C=1)-\mu_2 \end{pmatrix} \tag{11}$$

This depends on $E(Y_1|C=1)-\mu_1 = E(Y_1|C=1)-E(Y_1)$ and $E(Y_2|C=1)-\mu_2 = E(Y_2|C=1)-E(Y_2)$. Under MCAR, $C \perp\!\!\!\perp Y$, so that $E(Y_2|C=1)=\mu_2$ and $E(Y_1|C=1)=\mu_1$ and thus (11) is equal to zero. To go the other direction and argue that if (11) is equal to zero the missingness mechanism must be MCAR, it is tempting to simply say that (11) will be zero if $E(Y_1|C=1)-\mu_1=0$ and $E(Y_2|C=1)-\mu_2=0$, in which case because $\mu_1 = E(Y_1|C=1)\pi + E(Y_1|C=0)(1-\pi)$ it must be that $E(Y_1|C=0)-\mu_1=0$ and similarly $E(Y_2|C=0)-\mu_2=0$, so that $C \perp\!\!\!\perp Y$, i.e., MCAR. However, we'd better check that there is no other way that (11) can equal zero; e.g., by some fortuitous property of a normal distribution or if $\sigma_{12}=0$.

It is straightforward to observe that

$$E(Y_2|C) = E\{E(Y_2|Y_1,C)|C\} = E\{E(Y_2|Y_1)|C)$$

by MAR, from whence it follows by standard properties of the bivariate normal distribution that

$$E(Y_2|C) = E\left\{\mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(Y_1-\mu_1)\middle| C\right\},$$

so that

$$E(Y_2|C)-\mu_2 = \frac{\sigma_{12}}{\sigma_1^2}\{E(Y_1|C)-\mu_1\}$$

and thus

$$E(Y_2|C=1)-\mu_2 = \frac{\sigma_{12}}{\sigma_1^2}\{E(Y_1|C=1)-\mu_1\}. \tag{12}$$

Thus, we can rewrite (11) as

$$\pi\Sigma^{-1}E_{12}\Sigma^{-1}\begin{pmatrix} 1 \\ \sigma_{12}/\sigma_1^2 \end{pmatrix}\{E(Y_1|C=1)-\mu_1\} \tag{13}$$

or, if $\sigma_{12}\neq 0$, as

$$\pi\Sigma^{-1}E_{12}\Sigma^{-1}\begin{pmatrix} \sigma_1^2/\sigma_{12} \\ 1 \end{pmatrix}\{E(Y_2|C=1)-\mu_2\}. \tag{14}$$

Evaluating $\Sigma^{-1}E_{12}\Sigma^{-1}$ and simplifying, (13) and (14) become

$$\frac{\pi}{\sigma_1^2\sigma_2^2-\sigma_{12}^2}\begin{pmatrix} -\sigma_{12}/\sigma_1^2 \\ 1 \end{pmatrix}\{E(Y_1|C=1)-\mu_1\} \tag{15}$$

$$=\frac{\pi}{\sigma_1^2\sigma_2^2-\sigma_{12}^2}\begin{pmatrix} -1 \\ \sigma_1^2/\sigma_{12} \end{pmatrix}\{E(Y_2|C=1)-\mu_2\} \tag{16}$$

The expressions (15) and (16) make it clear that, if $\sigma_{12} \neq 0$, these expressions can only be zero if $E(Y_1|C=1)-\mu_1=0$ and $E(Y_2|C=1)-\mu_2=0$, which implies MCAR. If $\sigma_{12}=0$, (16) is erroneous, but in this case from (12) $E(Y_2|C=1)-\mu_2=0$ and by reversing the roles of Y_1 and Y_2 in (12) $E(Y_1|C=1)-\mu_1=0$, which again implies MCAR.