Homework 3: Clarification

Problem 2: Sampling from the posterior

The data augmentation algorithm involves alternating the I-step and the P-step. The I-step requires drawing a sample from a Gaussian distribution (which is straightforward). The P-step involves sampling from an Inv-Wishart distribution, as described in Little and Rubin. But they leave out some details, which are given here.

The prior is \( p(\mu, \Sigma) \propto |\Sigma|^{-(K+1)/2} \) and the likelihood of the samples is

\[
P(x_1, \ldots, x_m) = \frac{1}{(2\pi)^{mK/2}|\Sigma|^{m/2}} e^{-\frac{1}{2} \sum_{i=1}^{m} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)}. \tag{1}
\]

So the posterior is

\[
P(\mu, \Sigma|x_1, \ldots, x_m) \propto \frac{1}{|\Sigma|^{(m+K+1)/2}} e^{-\frac{1}{2} \sum_{i=1}^{m} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)}. \tag{2}
\]

We can re-express this in terms of:

\[
P(\mu|\Sigma, x_1, \ldots, x_m) = \frac{1}{(2\pi)^{K/2}|(1/m)\Sigma|^{1/2}} e^{-\frac{1}{2} (\mu - \bar{x})^T (m\Sigma^{-1})(\mu - \bar{x})},
\]

\[
P(\Sigma|x_1, \ldots, x_m) \propto \frac{1}{|\Sigma|^{(m+K)/2}} e^{-mTr(\Sigma_x\Sigma^{-1})}, \tag{3}
\]

where \( \bar{x} = (1/m) \sum_{i=1}^{m} x_i \), \( \Sigma_x \) is the covariance matrix computed from the \( x_i \) (i.e. \( \Sigma_x = (1/m) \sum_{i=1}^{m} (x_i - \bar{x})(x_i - \bar{x})^T \)), and \( Tr(\Sigma_x\Sigma^{-1}) \) is the trace of the matrix \( \Sigma_x\Sigma^{-1} \).

\( P(\Sigma|x_1, \ldots, x_m) \) is an Inv-Wishart distribution with scale parameter \( S = m\Sigma_x \) and \( m - 1 \) degrees of freedom (see page 117 Little and Rubin).

To sample from \( P(\mu, \Sigma|x_1, \ldots, x_m) \), you first sample from \( P(\Sigma|x_1, \ldots, x_m) \) to get \( \Sigma(d) \), and then sample from \( P(\mu|\Sigma, x_1, \ldots, x_m) \) to get \( \mu(d) \).

To sample (draw) from the Inv-Wishart you use formula (6.40) and (6.41). To get \( b_{ij} \) you sample from the chi-squared distribution \( \chi^2_{n-j} \) and then take the square root (to sample from \( \chi^2_{n-j} \) you can sum \( n - j \) samples from a normal distribution with zero mean and unit variance – or use an R function). To get the Cholesky factor of \( S^{-1} \), use the R program \( \text{chol}(S^{-1}) \).

To sample (draw) from the multivariate Gaussian \( P(\mu|\Sigma, x_1, \ldots, x_m) \) you can follow the procedure in example 6.21 \( (A(d) \) can be any matrix that obeys \( A(d)^T A(d) = (1/m)\Sigma(d) \).