SMF SOLUTIONS TO EXAMINATION. 2011

Q1. (i) The joint moment-generating function (MGF) is

$$M(u,v) := E\exp\{u^T Ax + v^T Bx\} = E\exp\{(A^T u + B^T v)^T x\}.$$

This is the MGF of x at argument $t = A^T u + B^T v$, so

$$M(u,v) = \exp\{(u^T A + v^T B)\mu + \frac{1}{2}[u^T A\Sigma A^T u + u^T A\Sigma B^T v + v^T B\Sigma A^T u + v^T B\Sigma B^T v]\}.$$

This factorises into a product of a function of u and a function of v iff the two cross-terms in u and v vanish, that is, iff $A\Sigma B^T = 0$ and $B\Sigma A^T = 0$; by symmetry of Σ, the two are equivalent. //
(ii)
(a) With $P = A$, $\Sigma = \sigma^2 I$, $I - P = B$, $A\Sigma B^T = \sigma^2 P(I - P) = 0$, as $P(I-P) = P - P^2 = P - P = 0$ as P is a projection. So independence of the linear forms follows by (i).
(b) $X^T PX = X^T PPX = X^T P^T PX = (PX)^T(PX)$, as P is a symmetric projection, and similarly $X^T(I-P)X = ((I-P)X)^T((I-P)X)$. So independence of the quadratic forms follows by (a).
(iii) As P is real and symmetric, it can be diagonalised by an orthogonal transformation: $P = B^T DB$ with B orthogonal and D diagonal; the elements of D are the eigenvalues of (D and) P. As $X \sim N(0, \sigma^2 I)$ and B is orthogonal, $Y := BX \sim N(0, \sigma^2 I)$ also. So $X^T PX = X^T B^T DBX = (BX)^T D(BX) = Y^T DX$; as $Y =_d X$, w.l.o.g. take $P = D$ orthogonal. As P, or D, is idempotent, its eigenvalues are 0 or 1. Its trace is the sum of the eigenvalues, = the number of 1s, = the number of non-zero eigenvalues, = the rank, = k, given. So the quadratic form is σ^2 times the sum of the squares of k independent $N(0,1)$s, which is $\chi^2(k)$. Similarly for $I - P$ and $n - k$:

$$X^T PX/\sigma^2 \sim \chi^2(k), \qquad X^T(I-P)X/\sigma^2 \sim \chi^2(n-k).$$

Q2. $y_A = \alpha + \epsilon_1$, $y_B = \beta + \epsilon_2$, $y_{B-A} = -\alpha + \beta + \epsilon_3$. So the regression model is $y = A\vec{\beta} + \epsilon$, where y is the 3-vector of readings, ϵ is the 3-vector of errors, $\vec{\beta}$ is the 2-vector of parameters and A is the design matrix.

(a)

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}.$$

(b)

$$C := A^T A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}; \quad |C| = 3.$$

So

$$C^{-1} = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, C^{-1}A^T = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \end{pmatrix}.$$

(c)

$$P = \frac{1}{3}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}, \quad I - P = \frac{1}{3}\begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}.$$

(d) The parameter estimates are

$$C^{-1}A^T y = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \end{pmatrix}\begin{pmatrix} y_A \\ y_B \\ y_{B-A} \end{pmatrix} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}:$$

$$\hat{\alpha} = (2y_A + y_B - y_{B-A})/3, \quad \hat{\beta} = (y_A + 2y_B + y_{B-A})/3.$$

(e) The fitted values are $\hat{y} = Py$, which can be written in two ways:

$$\begin{pmatrix} (2y_A + y_B - y_{B-A})/3 \\ (y_A + 2y_B + y_{B-A})/3 \\ (-y_A + y_B + 2y_{B-A})/3) \end{pmatrix} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\beta} - \hat{\alpha} \end{pmatrix}.$$

(f) As $n = 3$, $p = 2$ here, the ranks of P, $I - P$ are 2 and 1.

(g) Write SSE, the *sum of squares for error*, for $y^T(I-P)y$. As $n - p = 1$, $\hat{\sigma}^2 = SSE/(n-p) = y^T(I-P)y$, which can be calculated by above. With numerical data, SSE is the sum of squared residuals, $\sum(y_i - \hat{y}_i)^2$.

Q3. (i) The $AR(p)$ model is

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t, \qquad (*)$$

with (ϵ_t) white noise. In terms of the lag operator B,

$$X_t - \phi_1 B X_t - \cdots - \phi_p B^p X_t = \epsilon_t.$$

Write $\phi(\lambda) := 1 - \phi\lambda - \cdots - \phi_p \lambda^p$: $\phi(B)X_t = \epsilon_t$: $X_t = \phi(B)^{-1}\epsilon_t$. So if

$$1/\phi(\lambda) \equiv 1 + \psi_1\lambda + \cdots + \psi_n\lambda^n + \cdots, \quad X_t = \sum_{i=0}^{\infty} \psi_i B^i \epsilon_t = \sum_0^{\infty} \psi_i \epsilon_{t-i}, \ (MA)$$

giving the *moving-average representation.* **[5]**

(ii) Multiply $(*)$ through by X_{t-k} and take expectations. Since $E[X_{t-k}X_{t-i}] = \rho(|k-i|) = \rho(k-i)$, this gives the *Yule-Walker equations*

$$\rho(k) = \phi_1 \rho(k-1) + \cdots + \phi_p \rho(k-p) \qquad (k > 0), \qquad (YW)$$

a difference equation of order p with characteristic polynomial $\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p$. If $\lambda_1, \cdots, \lambda_p$ are the roots of this characteristic polynomial, the general solution is $\rho(k) = c_1 \lambda_1^k + \cdots + c_p \lambda_p^k$ (for $k \geq 0$, and use $\rho(-k) = \rho(k)$ for $k < 0$) if all the roots λ_i are distinct, with appropriate modifications for repeated roots (if $\lambda_1 = \lambda_2$, use $c_1\lambda_1^k + c_2 k \lambda_1^k$, etc.). **[5]**

(iii)

$$X_t = X_{t-1} - \frac{1}{4} X_{t-2} + \epsilon_t, \qquad (\epsilon_t) \quad WN. \qquad (*)$$

Substitute (MA) into $(*)$:

$$\sum_0^{\infty} \psi_i \epsilon_{t-i} = \sum_0^{\infty} \psi_i \epsilon_{t-i-1} - \frac{1}{4} \sum_0^{\infty} \psi_i \epsilon_{t-2-i} + \epsilon_t = \sum_1^{\infty} \psi_{i-1} \epsilon_{t-i} - \frac{1}{4} \sum_2^{\infty} \psi_{i-2} \epsilon_{t-i} + \epsilon_t.$$

Equate coefficients of ϵ_{t-i}: $i = 0$ gives $\psi_0 = 1$; $i = 1$ gives $\psi_1 = 1$; $i \geq 2$ gives

$$\psi_i = \psi_{i-1} - \frac{1}{4}\psi_{i-2}.$$

This is a difference equation, with characteristic polynomial $\lambda^2 - \lambda + 1/4 = 0$, or $(\lambda - 1/2)^2 = 0$, with a double root $\lambda_1 = 1/2$. The general solution of the difference equation is thus $\psi_i = a/2^i + b.i/2^i$. As $\psi_0 = 1$, $a = 1$; as $\psi_1 = 1$, $(1+b)/2 = 1$, $b = 1$. So $\psi_i = (1+i)/2^i$. The moving-average representaton is thus

$$X_t = \sum_{i=0}^{\infty} (1+i)\epsilon_i / 2^i. \qquad \textbf{[10]}$$

(iv) This converges a.s. and in mean, as $\psi = (\psi_i) \in \ell_1$, and in mean square, as $\psi \in \ell_2$. **[5]**

Q4. $x_2|x_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$.

If $x_1, x_2, x_3 \sim N(\mu, \Sigma)$ are independent, $y_1 := x_1 + x_2$, $y_2 := x_2 + x_3$, then $y = Ax$, where

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

So the mean is

$$Ey = A.Ex = A\mu = \begin{pmatrix} \mu_1 + \mu_2 \\ \mu_2 + \mu_3 \end{pmatrix} = m,$$

say. The variance is

$$var(y) = A\Sigma A^T = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1+\rho & 2\rho \\ 1+\rho & 1+\rho \\ 2\rho & 1+\rho \end{pmatrix} = \begin{pmatrix} 2+2\rho & 1+3\rho \\ 1+3\rho & 2+2\rho \end{pmatrix}.$$

So

$$y \sim N(m, \Sigma_y), \qquad m = \begin{pmatrix} \mu_1 + \mu_2 \\ \mu_2 + \mu_3 \end{pmatrix}, \qquad \Sigma_y = \begin{pmatrix} 2+2\rho & 1+3\rho \\ 1+3\rho & 2+2\rho \end{pmatrix}.$$

(ii) So on conditioning, the four partitioned submatrices are the four components; the conditional mean and conditional variance of $y_1|y_2$ are

$$m_1 + \frac{1+3\rho}{2(1+\rho)}(y_2 - m_2) = \mu_1 + \mu_2 + \frac{1+3\rho}{2(1+\rho)}(y_2 - \mu_2 - \mu_3), \quad 2(1+\rho) - \frac{(1+3\rho)^2}{2(1+\rho)}.$$

So

$$y_1|y_2 \sim N(\mu_1 + \mu_2 + \frac{1+3\rho}{2(1+\rho)}(y_2 - \mu_2 - \mu_3), 2(1+\rho) - \frac{(1+3\rho)^2}{2(1+\rho)}).$$

Q5. In principal components analysis (PCA), we seek a dimension reduction, say from p to k. The covariance (or correlation) matrix Σ, being real and symmetric, can be diagonalised by an orthogonal transformation:

$$\Sigma = \Gamma \Lambda \Gamma^T,$$

where $\Lambda = diag(\lambda_i)$ with $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$ are the eigenvalues of Σ and Γ is an orthogonal matrix of corresponding normalised eigenvectors. Then $y_1 := \gamma_1^T(x - \mu)$ is the standardised linear combination (SLC – sums of squares of coefficients = 1) of x with largest variance (λ_1), ...,

$$y_k := \gamma_k^T(x - \mu)$$

the SLC of largest variance (λ_k) uncorrelated with $y_1, \ldots, y_{k-1}$. Then the proportion of the total variability explained by the first k principal components is $(\lambda_1 + \ldots + \lambda_k)/(\lambda_1 + \ldots + \lambda_p)$. We continue to retain PCs until we are satisfied that this fraction is acceptably high. We then use these k PCs as a parsimonious summarisation of the data in k rather than p dimensions. **[6]**

We need to choose, *before* doing PCA, whether to work with covariances or with correlations. One prefers covariances when the units in which the data are measured are meaningful, correlations otherwise. **[6]**

Examples with correlations. Typically, data are given in terms of prices, and these are meaningful – they are expressed directly in terms of money. But what matters to an investor now is whether the stock will appreciate or depreciate. The actual amounts he cares about are the amounts he will *invest* in the various candidate stocks, and the number of stocks he holds in the company is simply the ratio of his stake to the stock price. Similarly, with foreign exchange, the units of currency in different countries may be of different orders of magnitude. Similarly for an investor dividing his holdings between different economic sectors: what counts here is proportions.

Examples with covariances. Examples where the units are meaningful include the internal accounts of a company, where different departments, or activities, contribute to the overall company accounts and balance sheet: all entries are in terms of money, and relate directly to profit and loss.

Empirical evidence suggests that in managing a portfolio of a range of stocks (that should be balanced – include lots of negative correlation – by Markowitzian diversification), covariances are better than correlations. **[13]**

N. H. Bingham