SMF MOCK EXAMINATION SOLUTIONS. 18.2.2011

Q1. (i) Spectral decomposition states that for a real symmetric matrix A, A may be diagonalised by an orthogonal transformation:

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

where $\Lambda = diag(\lambda_i)$ is a diagonal matrix of eigenvalues λ_i, $\Gamma = (\gamma_1, \ldots, \gamma_n)$ is an orthogonal matrix whose columns γ_i are standardised eigenvectors. **[4]**
(ii) The matrix square root $A^{\frac{1}{2}}$ is defined as

$$A^{\frac{1}{2}} := \Gamma \Lambda^{\frac{1}{2}} \Gamma^T; \qquad [\mathbf{2}]$$

the inverse square root is defined by

$$A^{\frac{1}{2}} := \Gamma \Lambda^{\frac{1}{2}} \Gamma^T. \qquad [\mathbf{2}]$$

(iii) If $x_1, \ldots, x_n$ are independent $N(0, \sigma^2)$, the joint density is

$$\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2} x_i^2/\sigma^2\} = \frac{1}{\sigma^n (2\pi)^{\frac{1}{2}n}} \exp\{-\frac{1}{2} \sum_{i=1}^{n} x_i^2/\sigma^2\} = \frac{1}{\sigma^n (2\pi)^{\frac{1}{2}n}} \exp\{-\frac{1}{2} \|x\|^2/\sigma^2\}.$$

If $y = Ox$ with O orthogonal, the Jacobian of the transformation is $|O| = 1$, and $\|y\| = \|x\|$, so the joint density of the y_i is

$$\frac{1}{\sigma^n (2\pi)^{\frac{1}{2}n}} \exp\{-\frac{1}{2} \|y\|^2/\sigma^2\}.$$

So $y_1, \ldots, y_n$ are independent $N(0, \sigma^2)$. **[5]**
(iv) The quadratic form $Q := x^T A x$ is thus $x^T \Gamma \Lambda \Gamma^T x$ by spectral decomposition. Writing $y := \Gamma^T x$, Γ^T is orthogonal as Γ is, so by above $Q = y^T \Lambda y$ is now a quadratic form in the independent $N(0, \sigma^2)$ random variables y_i with diagonal matrix Λ. **[4]**
(v) $A^2 = AA = \Gamma \Lambda \Gamma^T \Gamma \Lambda \Gamma^T, = \Gamma \Lambda \Lambda \Gamma^T = \Gamma \Lambda^2 \Gamma^T$, as Γ is orthogonal. So A is idempotent ($A^2 = A$) iff $\Lambda^2 = \Lambda$, i.e. iff each $\lambda_i^2 = \lambda_i$, i.e. iff each $\lambda_i = 0$ or 1. **[4]**
(vi) If P is a symmetric projection, P is symmetric and idempotent, so by above all its eigenvalues are 0 or 1. The trace is the sum of the eigenvalues; the rank is the number of non-zero eigenvalues. These are the same when, as here, all the eigenvalues are 0 or 1. **[4]**

Q2. (i) x is *multivariate normal* if all linear combinations of its components are univariate normal. Then if x has mean vector μ and covariance matrix Σ, $x \sim N(\mu, \Sigma)$. **[2]**

(ii) If x is multivariate normal, and y is an affine transformation of x, all linear combinations of components of y are (to within constants) linear combinations of (linear combinations of) components of x. So y is multivariate normal. **[2]**

$Ey = E[Ax + b] = aE[x] + b = A\mu + b$. **[2]**

$cov(y_i, y_j) = E[\sum_r a_{ir}(x_r - \mu_r) \sum_s a_{js}(x_s - \mu_s)] = \sum_{rs} a_{ir}\sigma_{rs}a_{js} = (A\Sigma A^T)_{ij}$: y has covariance matrix $A\Sigma A^T$. **[3]**

(iii) A subvector can always be obtained from x by applying a suitable matgrix A of 0s and 1s. So by (i), any subvector of a multinormal vector is multinormal. **[2]**

(iv)$x_1|x_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$. **[2]**

(v)

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & 0 \\ \rho^2 & 0 & 1 \end{pmatrix}, \quad \Sigma_{11} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \Sigma_{21} = (\rho^2, 0) = \Sigma_{12}^T, \quad \Sigma_{22} = 1.$$

So the conditional mean is

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \rho^2 \\ 0 \end{pmatrix} .1.(x_3 - \mu_3) = \begin{pmatrix} \mu_1 + \rho^2(x_3 - \mu_3) \\ \mu_2 \end{pmatrix}. \qquad [\mathbf{4}]$$

The conditional variance is

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \begin{pmatrix} \rho^2 \\ 0 \end{pmatrix} .1.(\rho^2 \ 0) = \begin{pmatrix} 1 - \rho^4 & \rho \\ \rho & 1 \end{pmatrix}. \qquad [\mathbf{6}]$$

Combining,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} |x_3 \sim N(\mu, \Sigma), \quad \text{with} \quad \mu = \begin{pmatrix} \mu_1 + \rho^2(x_3 - \mu_3) \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 - \rho^4 & \rho \\ \rho & 1 \end{pmatrix}.$$

[2]

Q3. (i) With time t discrete: if $X = (X_t)$ has $M := \sup_t E[|X_t|] < \infty$ and $\psi = (\psi_j) \in \ell_1$, i.e. $\|\psi\|_1 := \sum_{-\infty}^{\infty} |\psi_j| < \infty$ – then

$$E[\sum_j |\psi_j||X_{t-j}|] = \sum_j |\psi_j| E[|X_{t-j}|] \leq M\|\psi\|_1 < \infty$$

(interchanging E and $\sum$ by Fubini's theorem), so $\sum_j |\psi_j||X_{t-j}| < \infty$ a.s.: $\sum \psi_j X_{t-j}$ is a.s. absolutely convergent, to S say. **[4]**

Then

$$|\sum_{|j|>n} \psi_j X_{t-j}| \leq M|\sum_{|j|>n} \psi_j| \to 0 \quad (n \to \infty)$$

(tail of a convergent series), so $\sum \psi_i X_{t-j}$ converges to S in ℓ_1 also. **[4]**

(ii) If $\psi \in \ell_1$, $\sum |\psi_j| < \infty$. So $\psi_j \to 0$, so is bounded: $|\psi_j| \leq K$ say. Then $\sum_j |\psi_j|^2 \leq C \sum_j |\psi_j| = K\|\psi\|_1 < \infty$, i.e. $\psi \in \ell_2$. So $\ell_1 \subset \ell_2$. **[4]**

(iii) If $C := \sup_t E[|X_t|^2] < \infty$: take $n > m > 0$; then

$$E[|\sum_{m<|j|\leq n} \psi_i X_{t-j}|^2] = \sum_{m<j\leq n} \sum_{m<k\leq n} \psi_j \overline{\psi}_k E[X_{t-j}\overline{X_{t-k}}].$$

Now $|E[X_{t-j}\overline{X_{t-k}}]| \leq \sqrt{E[|X_{t-j}|^2].E[|X_{t-k}|^2]} \leq C$, by the Cauchy-Schwarz inequality. So the RHS

$$\leq C \sum_{m<j\leq n} \sum_{m<k\leq n} \psi_j \overline{\psi}_k = C|\sum_{m<j\leq n} \psi_j|^2 \to 0 \quad (m, n \to \infty),$$

as $\psi \in \ell_1$. So by completeness of ℓ_2, $\sum \psi_j X_{t-j}$ converges in ℓ_2 (that is, in mean square) – to S', say. **[8]**

Then by Fatou's Lemma

$$E[|S' - \sum_j \psi_j X_{t-j}|^2] = E[\liminf_n |S' - \sum_{-n}^{n} \psi_j X_{t-j}|^2] \leq \liminf_n E[|S' - \sum_{-n}^{n} \psi_j X_{t-j}|^2] = 0,$$

as $\sum \psi_j X_{t-j}$ converges to S in ℓ_2. So $S' = \sum_j \psi_j X_{t-j} = S$ a.s.: the a.s., ℓ_1 and ℓ_2 limits coincide. // **[5]**

Note. The a.s. convergence also follows from Kolmogorov's theorem on random series: $\psi_j X_{t-j}$ has variance

$$var(\psi_j X_{t-j}) = \psi_j^2 var(X_{t-j}) \leq \psi_j^2 \sup_t E[X_t^2] = C\psi_j^2,$$

so $\sum_j var(\psi_j X_{t-j})$ converges as $\psi \in \ell_2$. The same bound also gives ℓ_2-convergence, by dominated convergence.

Q4. In principal components analysis (PCA), we seek a dimension reduction, say from p to k. The covariance (or correlation) matrix Σ can be written by Spectral Decomposition as

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

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 $y_1, \ldots, k_{k-1}$. 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 in k dimensions of the data in p dimensions. **[12]**

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. **[3]**

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. **[5]**

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. **[5]**

Q5. (i)

$$\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = (1+2\rho) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix};$$

$$\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = (1-\rho) \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}; \quad \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = (1-\rho) \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

So the eigenvalues are $1+2\rho$ (single) and $1-\rho$ (double), with the eigenvectors as above, as required. **[4]**

(ii) The matrix, being a covariance matrix, must be non-negative definite, and so have non-negative eigenvalues. Now the correlation ρ lies in $[-1, 1]$, by the Cauchy-Schwarz inequality (and in $(-1, 1)$ in the non-degenerate case). This imposes no restriction on the eigenvalue $1 - \rho$, but forces $1 + 2\rho \geq 0$, i.e. $\rho \geq -1/2$, on the eigenvalue $1 + 2\rho$. **[5]**

The corresponding restriction in the positive-definite case $\rho > -1/2$. **[2]** (In the $n \times n$ case, the restriction is $\rho > -1/(n-1)$. The relevant theory from Linear Algebra is that of *circulant* matrices and determinants. Correlation matrices of this form arise in Statistics in the *intraclass correlation model.*)

(iii) If $x_i \sim N(\mu, \Sigma)$ are independent, $y_1 := x_1 + x_2$, $y_2 := x_2 + x_3$, $y = Ax$, 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 vector is

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

say. **[4]**

The covariance matrix is

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

$$= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1+\rho & 2\rho \\ 1+\rho & 1+\rho \\ 2\rho & 1+\rho \end{pmatrix} = \begin{pmatrix} 2+2\rho & 1+3\rho \\ 1+3\rho & 2+2\rho \end{pmatrix}. \qquad \textbf{[10]}$$