SMF EXAMINATION SOLUTIONS 2014-15

Q1. (i) (a) The log-likelihood is

$$\ell = -n \log 2 - \sum |x_i - \theta|.$$

To maximise this – i.e. minimise $\sum |x_i - \theta|$ – draw a graph. From this, the sum is minimised by $\theta = Med$, and increases linearly (slope $+1$ to the right, -1 to the left) on either side. So the MLE is $\hat{\mu} = Med$. **[3]**

(b) With one reading, as above, ℓ decreases with slope -1 to the right of Med, slope $+1$ to the left of Med. So $(\ell')^2 = 1$ (except at $\lambda = Med$, where the derivative is not defined – but we are going to integrate, and so can neglect null sets, e.g single points). So $I = \int (\partial \log f / \partial\theta)^2 f = \int f = 1$, as f is a density. So the CR bound is $1/n$. **[3]**

(c) We are given that Med is asymptotically normal, and that its mean is $med = \theta$, so Med is asymptotically unbiased. By symmetry, the population median is $med = \theta$, where the density is $\frac{1}{2}$. So $4f(med)^2 = 1$, and the asymptotic variance of the sample median is $1/n$, the CR bound, so Med is also asymptotically efficient. **[4]**

(ii) (a)

$$f(x;\mu) = \frac{1}{\pi(1 + (x-\mu)^2)}, \qquad \ell = \log f = c - \log[1 + (x-\mu)^2],$$

$$\ell' = \frac{2(x-\mu)}{1 + (x-\mu)^2}, \qquad \ell'(\mathbf{x};\theta) = 2\sum_1^n \frac{(x_i - \mu)}{1 + (x_i-\mu)^2}.$$

But we have efficiency iff ℓ' factorises in the form $\ell'(\mathbf{x};\theta) = A(\theta)(u(\mathbf{x}) - \theta)$. The likelihood here does not factorise, so there is no efficient estimator. **[4]**

(b) The information per reading is

$$E[(\ell')^2] = \int (\partial f/\partial\mu)^2 f = \frac{4}{\pi}\int \frac{(x-\mu)^2}{[1+(x-\mu)^2]^3} dx = \frac{4}{\pi}\int \frac{x^2}{[1+x^2]^3} dx = \frac{4}{\pi} I,$$

say. Given $I = \pi/8$ (to evaluate I by Complex Analysis: use $f(z) := z^2/[1+z^2]^3$, the contour Γ a large semicircle in the upper half-plane; f has a triple pole inside Γ of residue $-i/16$, so $I = 2\pi i\ Res = \pi/8$), so the information per reading is $\frac{1}{2}$. So the information in a sample of size n is $n/2$, and the MLE has asymptotic variance $2/n$. As in (i), $med = \mu$, $f(med) = 1/\pi$, so the sample median Med has asymptotic variance $1/(4nf(med)^2) = \pi^2/4n$. So the asymptotic efficiency is their ratio, $8/\pi^2 \sim 81\%$. **[6]**

[Seen – Problems]

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

(a)

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}. \qquad [\mathbf{3}]$$

(b)

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

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

[**3**]

(c)

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

[**4**]

(d) The parameter estimates are

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

$$\hat{a} = (2y_A + y_B - y_{B-A})/3, \quad \hat{b} = (y_A + 2y_B + y_{B-A})/3. \qquad [\mathbf{3}]$$

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

$$\begin{pmatrix} (2y_A + y_B - y_{B-A})/3 \\ (y_A + 2y_B + y_{B-A})/3 \\ (-y_A + y_B + 2y_{B-A})/3) \end{pmatrix} = \begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{b} - \hat{a} \end{pmatrix}. \qquad [\mathbf{3}]$$

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

(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, $\sum(y_i - \hat{y}_i)^2$. [**2**]

[Unseen; similar seen in lectures and problems]

Q3. $x_2|x_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$.

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

$$Ey = A.Ex = A\mu = \begin{pmatrix} \mu_1 + \mu_2 \\ \mu_2 + \mu_3 \end{pmatrix} = m, \qquad [\mathbf{3}]$$

say. The variance is

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

$$= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1+\rho & 2\rho \\ 1+\rho & 1+\rho \\ 2\rho & 1+\rho \end{pmatrix} = \begin{pmatrix} 2+2\rho & 1+3\rho \\ 1+3\rho & 2+2\rho \end{pmatrix}. \qquad [\mathbf{5}]$$

So

$$y \sim N(m, \Sigma_y), \qquad m = \begin{pmatrix} \mu_1 + \mu_2 \\ \mu_2 + \mu_3 \end{pmatrix}, \qquad \Sigma_y = \begin{pmatrix} 2+2\rho & 1+3\rho \\ 1+3\rho & 2+2\rho \end{pmatrix}. \quad [\mathbf{4}]$$

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

$$m_1 + \frac{1+3\rho}{2(1+\rho)}(y_2 - m_2) = \mu_1 + \mu_2 + \frac{1+3\rho}{2(1+\rho)}(y_2 - \mu_2 - \mu_3), \quad 2(1+\rho) - \frac{(1+3\rho)^2}{2(1+\rho)}.$$
[**4**]

So

$$y_1|y_2 \sim N(\mu_1 + \mu_2 + \frac{1+3\rho}{2(1+\rho)}(y_2 - \mu_2 - \mu_3), 2(1+\rho) - \frac{(1+3\rho)^2}{2(1+\rho)}). \qquad [\mathbf{4}]$$

[Unseen; similar seen – lectures and problems]

Q4. $ARMA(1,1)$. $X_t = \phi X_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$: $(1-\phi B)X_t = (1+\theta B)\epsilon_t$.
(i) Condition for stationarity and invertibility: $|\phi| < 1$; $|\theta| < 1$. **[2, 2]**
(ii) Assuming these:

$$X_t = (1-\phi B)^{-1}(1+\theta B)\epsilon_t = (1+\theta B)(\sum_0^\infty \phi^i B^i)\epsilon_t$$

$$= \epsilon_t + \sum_1^\infty \phi^i B^i \epsilon_t + \theta \sum_0^\infty \phi^i B^{i+1}\epsilon_t = \epsilon_t + (\theta+\phi)\sum_1^\infty \phi^{i-1}B^i\epsilon_t :$$

$$X_t = \epsilon_t + (\phi+\theta)\sum_{i=1}^\infty \phi^{i-1}\epsilon_{t-i}.$$

(a) *Variance*: lag $\tau = 0$. The ϵs are uncorrelated with variance σ^2, so

$$\gamma_0 = var X_t = E[X_t^2] = \sigma^2 + (\phi+\theta)^2 \sum_1^\infty \phi^{2(i-1)}\sigma^2$$

$$= \sigma^2 + \frac{(\phi+\theta)^2\sigma^2}{(1-\phi^2)} = \sigma^2(1-\phi^2+\phi^2+2\phi\theta+\theta^2)/(1-\phi^2) :$$

$$\gamma_0 = \sigma^2(1+(\phi+\theta)^2/(1-\phi^2)) = \sigma^2(1+2\phi\theta+\theta^2)/(1-\phi^2) \qquad [\mathbf{8}]$$

(b) *Covariance*: lag $\tau \geq 1$.

$$X_{t-\tau} = \epsilon_{t-\tau} + (\phi+\theta)\sum_{j=1}^\infty \phi^{j-1}\epsilon_{t-\tau-j}.$$

Multiply the series for X_t and $X_{t-\tau}$ and take expectations:

$$\gamma_\tau = cov(X_t, X_{t-\tau}) = E[X_t X_{t-\tau}],$$

$$= E[[\epsilon_t + (\phi+\theta)\sum_{i=1}^\infty \phi^{i-1}\epsilon_{t-i}].[\epsilon_{t-\tau} + (\phi+\theta)\sum_{j=1}^\infty \phi^{j-1}\epsilon_{t-\tau-j}]].$$

The ϵ_t-term in the first [.] gives no contribution. The i-term in the first [.] for $i = \tau$ and the $\epsilon_{t-\tau}$ in the second [.] give $(\phi+\theta)\phi^{\tau-1}\sigma^2$. The product of the i term in the first sum and the j term in the second contributes for $i = \tau + j$; for $j \geq 1$ it gives $(\phi+\theta)^2\phi^{\tau+j-1}.\phi^{j-1}.\sigma^2$. So

$$\gamma_\tau = (\phi+\theta)\phi^{\tau-1}\sigma^2 + (\phi+\theta)^2\phi^\tau\sigma^2\sum_{j=1}^\infty \phi^{2(j-1)}.$$

The geometric series is $1/(1-\phi^2)$ as before, so

$$\gamma_\tau = \sigma^2(\phi+\theta)\phi^{\tau-1} + \sigma^2\phi^\tau(\phi+\theta)^2/(1-\phi^2) \qquad (\tau \geq 1).$$

This decreases geometrically beyond the first term $\gamma_0 = 1$, and this behaviour is indicative of $ARMA(1,1)$. **[8]**
[Seen – lectures and problems]

Q5. (i) Markowitz' work of 1952 (which led on to CAPM in the 1960s) gave two key insights:
(a). *Think of risk and return together, not separately.* Now return corresponds to mean (= mean rate of return), risk corresponds to variance – hence *mean-variance* analysis, *efficient frontier*, etc. – maximise return for a given level of risk/minimise risk for a given return rate). **[2]**
(b). *Diversify* (don't 'put all your eggs in one basket'). Hold a *balanced portfolio* – a range of risky assets, with lots of *negative correlation* – so that when things change, losses on some assets will be offset by gains on others. **[2]**
Hence the vector-matrix parameter (μ, Σ) is accepted as an essential part of any model in mathematical finance.
(ii) *Elliptical distributions.*

The normal density in higher dimensions is a multiple of $\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$, where the matrices Σ, Σ^{-1} are *positive definite* (PD), so the contours $(x-\mu)^T\Sigma^{-1}(x-\mu) = \text{const.}$ are *ellipsoids.* The general *elliptically contoured* distribution has a density

$$f(x) = const.g((x-\mu)^T\Sigma^{-1}(x-\mu)).$$

This is a *semi-parametric* model, where $\theta := (\mu, \sigma)$ is the parametric part and the *density generator* g is the non-parametric part. **[4]**
(iii) *Normal (Gaussian) model*: elliptically contoured ($g(.) = e^{-\frac{1}{2}\cdot}$). Though very useful, it has various deficiencies, e.g.:
(a) It is *symmetric.* Many financial data sets show asymmetry, or *skew.* This reflects the asymmetry between profit and loss. Big profits are nice; big losses can be lethal (to the firm – bankruptcy). **[3]**
(b) It has extremely thin tails. Most financial data sets have tails that are *much fatter* than the ultra-thin normal tails. **[3]**
(iv) For asset returns (= profit/loss over initial asset price) over a period, the *return period*: matters vary dramatically with the return period.
(a) For *long* return periods (monthly, say – the Rule of Thumb is that 16 trading days suffice), the CLT applies, and asset returns are approximately *normal* ('aggregational Gaussianity). **[2]**
(ib) For *intermediate* return periods (daily, say), a commonly used model is the *generalised hyperbolic (GH)* – log-density a hyperbola, with linear asymptotes, so density decays like the exponential of a linear function). **[2]**
(c) For *high-frequency* returns ('tick data', say – every few seconds), the density typically decays like a power (as with the Student t distribution). **[2]**
[Seen – lectures.]

Q6. (i) Edgeworth's theorem says that if $x \sim N(\mu, \Sigma)$ and $K := \Sigma^{-1}$,

$$f(x) \propto \exp\{-\frac{1}{2}(x-\mu)^T K (x-\mu)\}. \qquad [\mathbf{1}]$$

$$f(x_1, x_2) \propto \exp\{-\frac{1}{2}(x_1^T - \mu_1^T, x_2^T - \mu_2^T) \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\},$$

giving (as a scalar is its own transpose, so the two cross-terms are the same)

$$\exp\{-\frac{1}{2}[(x_1^T - \mu_1^T)K_{11}(x_1-\mu_1) + 2(x_1^T - \mu_1^T)K_{12}(x_2-\mu_2) + (x_2^T - \mu_2^T)K_{22}(x_2-\mu_2)]\}.$$

So

$$f_{1|2}(x_1|x_2) = f(x_1, x_2)/f_2(x_2)$$

$$\propto \exp\{-\frac{1}{2}[(x_1^T - \mu_1^T)K_{11}(x_1-\mu_1) + 2(x_1^T - \mu_1^T)K_{12}(x_2-\mu_2)]\}, \qquad (*)$$

treating x_2 here as a constant and x_1 as the argument of $f_{1|2}$. **[4]**

By Edgeworth's theorem again, if the conditional mean of $x_1|x_2$ is ν_1,

$$f_{1|2}(x_1|x_2) \propto \exp\{-\frac{1}{2}(x_1^T - \nu_1^T)V_{11}(x_1 - \nu_1)\}, \qquad (**)$$

for some matrix V_{11}. So $x_1|x_2$ is multinormal (as may be quoted). **[3]**

Equating coefficients of the quadratic term gives the conditional concentration matrix of $x_1|x_2$ as $V_{11} = K_{11}$:

$$conc(x_1|x_2) = K_{11}. \qquad [\mathbf{3}]$$

So the conditional covariance matrix is K_{11}^{-1}. Then equating linear terms in $(*)$ and $(**)$ gives the conditional mean:

$$x_1^T K_{11}\nu_1 = x_1^T K_{11}\mu_1 - x_1^T K_{12}(x_2-\mu_2): \quad \nu_1 := E[x_1|x_2] = \mu_1 - K_{11}^{-1}K_{12}(x-\mu_2).$$

[3]

So

$$x_1|x_2 \sim N(\mu_1 - K_{11}^{-1}K_{12}(x-\mu_2), K_{11}^{-1}). \qquad [\mathbf{2}]$$

Using the quoted result for the inverse of a partitioned matrix gives

$$M = K_{11}, \qquad M^{-1} = K_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

$$K_{11}^{-1}K_{12} = M^{-1}(-MBD^{-1}) = -BD^{-1} = -\Sigma_{12}\Sigma_{22}^{-1}.$$

Combining,

$$x_1|x_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x-\mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \qquad [\mathbf{4}]$$

[Seen – Problems] NHB