SMF SOLUTIONS TO MOCK EXAMINATION. 2012

Q1. (i) With $\ell := \log L$ the log-likelihood, the *score function* is

$$s(\theta) := \ell'(\theta). \qquad [\mathbf{2}]$$

The *information* is

$$I(\theta) := E[s(\theta)^2] = -E[s'(\theta)] \qquad [\mathbf{2}]$$

(we shall see below that these are equal – either will do here).

We have a joint density $f = f(x_1, \ldots, x_n; \theta)$, which we write as $f = f(x; \theta)$. This integrates to 1: $\int f(x;\theta)dx = 1$ (where dx is n-dimensional Lebesgue measure), which we abbreviate to $\int f = 1$. We assume throughout that $f(x;\theta)$ is smooth enough for use to differentiate under the integral sign (w.r.t. dx, understood) w.r.t. θ, twice. Then

$$\int \frac{\partial f}{\partial \theta} = \frac{\partial}{\partial \theta}\int f = \frac{\partial}{\partial \theta} 1 = 0: \quad \int \Big(\frac{1}{f}\frac{\partial f}{\partial \theta}\Big).f = 0: \quad \int \Big(\frac{\partial}{\partial \theta}\log f\Big).f = 0.$$

Now $E[g(X)] = \int g(x)f(x;\theta)dx = \int gf$, so this says

$$E\Big[\frac{\partial \log L}{\partial \theta}\Big] = 0: \quad E\Big[\frac{\partial \ell}{\partial \theta}\Big] = 0: \quad E[\ell'(\theta)] = 0: \quad E[s(\theta)] = 0. \qquad (a) \quad [\mathbf{6}]$$

Differentiate under the integral sign wrt θ again:

$$\frac{\partial}{\partial \theta}\int \Big(\frac{1}{f}\frac{\partial f}{\partial \theta}\Big).f = 0, \qquad \int \frac{\partial}{\partial \theta}\Big[\Big(\frac{1}{f}\frac{\partial f}{\partial \theta}\Big).f\Big] = 0:$$

$$\int \Big[\Big(\frac{1}{f}\frac{\partial f}{\partial \theta}\Big)\frac{\partial f}{\partial \theta} + f\frac{\partial}{\partial \theta}\Big(\frac{1}{f}\frac{\partial f}{\partial \theta}\Big)\Big] = 0.$$

As the bracket in the second term is $\partial \log f/\partial \theta$, this says

$$\int \Big[\Big(\frac{1}{f}\frac{\partial f}{\partial \theta}\Big)^2 + \frac{\partial}{\partial \theta}\Big(\frac{\partial \log f}{\partial \theta}\Big)\Big]f = 0, \quad \int \Big[\Big(\frac{\partial \log f}{\partial \theta}\Big)^2 + \frac{\partial^2}{\partial \theta^2}(\log f)\Big]f = 0,$$

$$E\Big[\Big(\frac{\partial}{\partial \theta}\log L\Big)^2 + \frac{\partial^2}{\partial \theta^2}\log L\Big] = 0: \ E[\{\ell'(\theta)\}^2 + \ell''(\theta)] = 0: \ E[s(\theta)^2 + s'(\theta)] = 0.$$

So with

$$I(\theta) := E[\{\ell'(\theta)\}^2] = -E[\ell''(\theta)]: \qquad I(\theta) = E[s^2(\theta)] = -E[s'(\theta)], \quad (b) \quad [\mathbf{6}]$$

giving the equivalence of the two definitions above. By (a) and (b):

The score function $s(\theta) := \ell'(\theta)$ has mean 0 and variance $I(\theta)$. **[4]**

(Seen – lectures)

Q2. *Sufficiency.* We choose the Bayesian framework, as it is easier (for the non-Bayesian approach, see lectures, Day 2).
(i) If $x = (x_1, x_2)$, we call x_1 *sufficient* for θ if x_2 is uninformative about θ, i.e. does not affect our views on θ, that is,
(a) $f(\theta|x) = f(\theta|x_1, x_2)$ does not depend on x_2, i.e.

$$f(\theta|x_1, x_2) = f(\theta|x_1), \qquad \text{or} \qquad \frac{f(\theta, x_1, x_2)}{f(x_1, x_2)} = \frac{f(\theta, x_1)}{f(x_1)} :$$

$$\frac{f(\theta, x_1, x_2)}{f(\theta, x_1)} = \frac{f(x_1, x_2)}{f(x_1)}, \qquad \text{i.e.} \qquad f(x_2|x_1, \theta) = f(x_2|x_1) :$$

(b) $f(x_2|x_1, \theta)$ does not depend on θ.
Either of (a), (b), which are equivalent, can be used as the definition of sufficiency in a Bayesian treatment. [Notice that (a) is essentially a Bayesian statement: it is meaningless in classical statistics, as there θ cannot have a density.] **[6]**

The Fisher-Neyman Factorisation Criterion for sufficiency is that the likelihood $f(x|\theta)$ factorises as
(c) $f(x|\theta)$, or $f(x_1, x_2|\theta)$, $= g(x_1, \theta)h(x_1, x_2)$,
for some functions g, h. The Fisher-Neyman Theorem is that this is necessary and sufficient: x_1 is sufficient for θ iff the Factorisation Criterion (c) holds. **[6]**

Proof. (b) $\Rightarrow$ (c):

$$\begin{aligned} f(x|\theta) = f(x_1, x_2|\theta) &= \frac{f(x_1, x_2, \theta)}{f(\theta)} \\ &= \frac{f(x_1, \theta)}{f(\theta)} . \frac{f(x_1, x_2, \theta)}{f(x_1, \theta)} \\ &= f(x_1|\theta) f(x_2|x_1, \theta) \\ &= f(x_1|\theta) f(x_2|x_1) \qquad \text{(by (b))}, \end{aligned}$$

giving (c).
(c) $\Rightarrow$ (a): By Bayes' Theorem, posterior is proportional to prior times likelihood. The factor $h(x_1, x_2)$ in (c) can be absorbed into the constant of proportionality. Then x_2 disappears, so does not appear in the posterior, giving (a). // **[8]**
(Seen – lectures)

Q3. *Normal means* $N(\mu, \sigma^2)$, σ *unknown.*

The *likelihood ratio test* (LRT) for H_0 v. H_1, where $H_0 \subset H_1$, is to let $\lambda_i := \sup_{H_i} L$, define the *likelihood ratio* statistic (LR) as $\lambda := L_0/L_1$, and *reject* H_0 if λ is *too small.* **[2]**

$$H_0: \quad \mu = \mu_0 \qquad v. \quad H_1: \quad \mu \text{ unrestricted.}$$

$$L = \frac{1}{\sigma^n (2\pi)^{n/2}} . \exp\{-\frac{1}{2} \sum_1^n (x_i - \mu)^2 / \sigma^2\},$$

$$L_0 = \frac{1}{\sigma^n (2\pi)^{n/2}} . \exp\{-\frac{1}{2} \sum_1^n (x_i - \mu_0)^2 / \sigma^2\} = \frac{1}{\sigma^n (2\pi)^{n/2}} . \exp\{-\frac{1}{2} n S_0^2 / \sigma^2\},$$

in an obvious notation. The MLEs under H_1 are $\hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = S^2$, as usual (and as in lectures). Similarly (though more simply), under H_0, we obtain $\sigma = S_0$. So

$$L_1 = \frac{e^{-\frac{1}{2}n}}{S^n (2\pi)^{\frac{1}{2}n}}; \qquad L_0 = \frac{e^{-\frac{1}{2}n}}{S_0^n (2\pi)^{\frac{1}{2}n}}.$$

So

$$\lambda := L_0/L_1 = S^n / S_0^n,$$

and the LR test is: reject if λ is too small. **[8]**

Now

$$n S_0^2 = \sum_1^n (X_i - \mu_0)^2 = \sum [(X_i - \bar{X}) + (\bar{X} - \mu_0)]^2$$

$$= \sum (X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2 = n S^2 + n(\bar{X} - \mu_0)^2$$

(as $\sum (X_i - \bar{X}) = 0$):

$$\frac{S_0^2}{S^2} = 1 + \frac{(\bar{X} - \mu_0)^2}{S^2}.$$

But $t := (\bar{X} - \mu_0)\sqrt{n-1}/S$ has the Student t-distribution $t(n-1)$ with n df under H_0, so

$$S_0^2 / S^2 = 1 + t^2/(n-1).$$

The LR test is: reject if

$\lambda = (S/S_0)^n$ too small;

$S_0^2/S^2 = 1 + t^2/(n-1)$ too big;

t^2 too big: $|t|$ too big, which is the Student t-test:

The LR test here is the Student t-test. **[10]**

(Seen – lectures)

Q4. $AR(2)$.

$$X_t = \frac{1}{3}X_{t-1} + \frac{2}{9}X_{t-2} + \epsilon_t, \qquad (\epsilon_t) \quad WN. \tag{1}$$

Moving-average representation. Let the MA representation of (X_t) be

$$X_t = \sum\nolimits_{i=0}^{\infty} \psi_i \epsilon_{t-i}. \tag{2}$$

Substitute (2) into (1):

$$\begin{aligned}\sum\nolimits_0^{\infty} \psi_i \epsilon_{t-i} &= \frac{1}{3}\sum\nolimits_0^{\infty} \psi_i \epsilon_{t-i-1} + \frac{2}{9}\sum\nolimits_0^{\infty} \psi_i \epsilon_{t-2-i} + \epsilon_t \\ &= \frac{1}{3}\sum\nolimits_1^{\infty} \psi_{i-1} \epsilon_{t-i} + \frac{2}{9}\sum\nolimits_2^{\infty} \psi_{i-2} \epsilon_{t-i} + \epsilon_t.\end{aligned}$$

Equate coefficients of ϵ_{t-i}:
$i = 0$ gives $\psi_0 = 1$; $i = 1$ gives $\psi_1 = \frac{1}{3}\psi_0 = 1/3$; $i \geq 2$ gives

$$\psi_i = \frac{1}{3}\psi_{i-1} + \frac{2}{9}\psi_{i-2}.$$

This is again a difference equation, which we solve as above. The characteristic polynomial is

$$\lambda^2 - \frac{1}{3}\lambda - \frac{2}{9} = 0, \qquad \text{or} \qquad (\lambda - \frac{2}{3})(\lambda + \frac{1}{3}) = 0,$$

with roots $\lambda_1 = 2/3$ and $\lambda_2 = -l/3$. The general solution of the difference equation is thus $\psi_i = c_1\lambda_1^i + c_2\lambda_2^i = c_1(2/3)^i + c_2(-1/3)^i$. We can find c_1, c_2 from the values of ψ_0, ψ_1, found above:
$i = 0$ gives $c_1 + c_2 = 0$, or $c_2 = 1 - c_1$.
$i = 1$ gives $c_1.(2/3) + (1 - c_1)(-1/3) = \psi_1 = 1/3$: $2c_1 - (1 - c_1) = 1$: $c_1 = 2/3$, $c_2 = 1/3$. So

$$\psi_i = \frac{2}{3}(\frac{2}{3})^i + \frac{1}{3}(\frac{-1}{3})^i = (\frac{2}{3})^{i+1} - (\frac{-1}{3})^{i+1},$$

and

$$X_t = \sum\nolimits_0^{\infty} [(\frac{2}{3})^{i+1} - (\frac{-1}{3})^{i+1}]\epsilon_{t-i},$$

giving the MA representation. **[14]**

The Yule-Walker equations are

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2).$$

We solve this difference equation as above, obtaining $\rho(k) = a\lambda_1^k + b\lambda_2^k$, and find a, b from the first few values, again as above. **[3, 3]**
(Seen – lectures)

Q5. (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. 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. // **[8]**
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.
(Seen – 2011 Mock Exam)

Q6. (i) (*Rank-one matrices*). If C is the zero matrix, it has rank 0 – a trivial case, which we exclude.

If C has rank one, the range of C is one-dimensional (this is one of several equivalent definitions of rank). If the domain and range of C have bases e_i, f_j, Ce_i is non-zero for some i (or C would be zero) – w.l.o.g., $Ce_1 = \sum_j c_{1j} f_j \neq 0$. Write $b_j := c_{1j}$: $Ce_1 = \sum_j b_j f_j \neq 0$. As the range of C is one-dimensional, for each i, Ce_i is a multiple $a_i Ce_1$ of Ce_1: $Ce_i = \sum_j a_i b_j f_j$. This says that the linear transformation represented by C has matrix $C = (c_{ij}) = (a_i b_j)$ w.r.t. the bases e_i and f_j.

Conversely, if $C = (a_i b_j)$ is not the zero matrix: at least one $a_i b_j \neq 0$; w.l.o.g., by re-ordering rows and columns, take $a_1, b_1 \neq 0$. Then column j is the multiple b_j/b_1 of column 1, and row i is the multiple a_i/a_1 of row 1. So C has column-rank 1 (only 1 linearly independent column), and row-rank 1, so has rank 1. **[10]**

(ii) With $\mathbf{a}$ the column-vector $(a, 1, 1)^T$, $A = \mathbf{a}\mathbf{a}^T$. So A has *rank* 1, and as it is 3×3, it has one non-zero eigenvalues and two 0 eigenvalues.

$$A\mathbf{a} = \mathbf{a}\mathbf{a}^T\mathbf{a} = (a^2 + 2)\mathbf{a},$$

as $\mathbf{a}^T\mathbf{a} = a^2 + 2$. This says that A has eigenvalue $a^2 + 2$ with eigenvector $\mathbf{a}$. The other two eigenvalues are 0, with eigenvectors $\mathbf{x}$, $\mathbf{y}$ say. The eigenequation for $\mathbf{x} = (x_1, x_2, x_3)^T$ is

$$ax_1 + x_2 + x_3 = 0,$$

three times (check). Taking $x_3 = 0$, we can take $x_1 = 1$, $x_2 = -a$; taking $x_1 = 0$, we can take $x_2 = 1$, $x_3 = 1$, giving

$$\mathbf{x} = \begin{pmatrix} 1 \\ -a \\ 0 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}. \qquad [\mathbf{10}]$$

(Seen – Problems)

NHB