SMF EXAMINATION SOLUTIONS 2016-17

Q1. With $\ell(\theta)$ the log-likelihood, the *score function* is

$$s := \ell'; \qquad [\mathbf{2}]$$

the *information per reading* is

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

In the example given, write $v := \sigma^2$.

$$\ell(v) = \log f = const - \frac{1}{2}\log v - \frac{1}{2}(X-\mu)^2/v,$$

$$s(v) := \ell'(v) = -\frac{1}{2v} + \frac{(X-\mu)^2}{2v^2},$$

$$s'(v) = \frac{1}{2v^2} - \frac{(X-\mu)^2}{v^4}.$$

The information per reading is

$$I = I(v) = E[-s'(v)] = -\frac{1}{2v^2} + \frac{E[(X-\mu)^2]}{v^3} = -\frac{1}{2v^2} + \frac{v}{v^3} = \frac{1}{2v^2}. \qquad [\mathbf{8}]$$

The CR bound is

$$1/(nI) = 2v^2/n. \qquad [\mathbf{2}]$$

Write

$$S_0^2 := \frac{1}{n}\sum_1^n (X_i - \mu)^2. \qquad [\mathbf{2}]$$

Then

$$nS_0^2/\sigma^2 \sim \chi^2(n)$$

(definition of $\chi^2(n)$), so has mean n and variance $2n$ – because $\chi^2(1)$ has mean 1 ('normal variance') and variance 2 (by an MGF calculation or from memory). So S_0^2 has mean σ^2 (so is unbiased for σ^2), and variance $2n.\sigma^4/n^2 = 2v^2/n$, the CR bound above, so is efficient for $v = \sigma^2$. [**4**]
Seen – lectures (bookwork) and problems (example).

Q2. *Lognormal distribution; normal means.*

X has the *log-normal* distribution with parameters μ and σ, $X \sim LN(\mu, \sigma)$, if $Y := \log X \sim N(\mu, \sigma^2)$. **[2]**

The MGF of Y is $M_Y(t) := E[e^{tY}] = \exp\{\mu t + \frac{1}{2}\sigma^2 t^2\}$: $M_Y(1) = E[e^Y] = \exp\{\mu + \frac{1}{2}\sigma^2\}$.

But $e^Y = X$: $E[X] = \exp\{\mu + \frac{1}{2}\sigma^2\}$: $LN(\mu, \sigma)$ has mean $\exp\{\mu + \frac{1}{2}\sigma^2\}$. **[3]**

In *geometric Brownian motion (GBM)*, as in the *Black-Scholes model*, the price process $S = (S_t)$ of a risky asset is driven by the SDE

$$dS_t/S_t = \mu dt + \sigma dW_t, \qquad (GBM)$$

with $W = (W_t)$ Brownian motion/the Wiener process. This has solution

$$S_t = S_0 \exp\{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\} :$$

$\log S_t$ is lognormally distributed. **[5]**

For a normal population $N(\mu, \sigma)$ with σ known: to test $H_0 : \mu = \mu_0$ v. $H_1 : \mu < \mu_0$. First, take any $\mu_1 < \mu_0$. To test H_0 v. $\mu = \mu_1$, by the Neyman-Pearson Lemma (NP), the best (most powerful) test uses test statistic the likelihood ratio (LR) $\lambda := L_0/L_1 = L(\mu_0)/L(\mu_1)$, where with data $x_1, \ldots, x_n$

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

and critical region R of the form $\lambda \leq$ const: *reject* H_0 if λ is *too small.* Here $\lambda = \exp\{-\frac{1}{2}[\sum(x_i - \mu_0)^2 - \sum(x_i - \mu_1)^2]\}$. Forming the LR λ, the constants cancel, so R has the form $\log \lambda \leq$ const, or $-2\log\lambda \geq$ const. Expanding the squares, the $\sum x_i^2$ terms cancel, so (as $\sum x_i = n\bar{x}$) this is

$$-2\mu_0 n\bar{x} + n\mu_0^2 + 2\mu_1 n\bar{x} - n\mu_1^2 \geq \text{const} : \quad 2(\mu_1 - \mu_0)\bar{x} + (\mu_0^2 - \mu_1^2) \geq \text{const}.$$

As $\mu_1 < \mu_0$, this is $\bar{x} \leq c$. At significance level α, c is the lower α-point of the distribution of $\bar{x}$ under H_0. Then $\bar{x} \sim N(\mu_0, \sigma^2/n)$, so $Z := (\bar{x} - \mu_0)\sqrt{n}/\sigma \sim \Phi = N(0,1)$. If c_α is the lower σ-point of $\Phi = N(0,1)$, i.e. of $Z := (\bar{x} - \mu_0)\sqrt{n}/\sigma$, $c_\alpha = (c - \mu_0)\sqrt{n}/\sigma$: $c = \mu_0 + \sigma c_\alpha/\sqrt{n}$. **[7]**

But this holds for *all* $\mu_1 < \mu_0$. So R is *uniformly most powerful* (UMP) for $H_0 : \mu = \mu_0$ (simple null) v. $H_1 : \mu < \mu_0$ (composite alternative). **[3]**

[Seen – lectures]

Q3. (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.

Q4 (*Sufficiency for the multivariate normal*). Given a sample $x_1, \dots, x_n$ from a multivariate distribution, form the *sample mean* (vector) and the *sample covariance matrix* as in the one-dimensional case:

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad S := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T. \qquad \mathbf{[2, 2]}$$

(i) The *multivariate normal distribution* (in d dimensions) $N(\mu, \Sigma)$ (μ a d-vector, Σ an $d \times d$ symmetric positive definite matrix) has density (Edgeworth's Theorem)

$$f(\mathbf{x}) := \frac{1}{(2\pi)^{\frac{1}{2}d}|\mathbf{\Sigma}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x} - \mu)\}.$$

The likelihood for a sample of size 1 is

$$L(x|\mu, \Sigma) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\},$$

so the likelihood for a sample of size n is

$$L = (2\pi)^{-np/2}|\Sigma|^{-n/2} \exp\{-\frac{1}{2}\sum_{1}^{n}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)\}.$$

Writing $x_i - \mu = (x_i - \bar{x}) - (\mu - \bar{x})$,

$$\sum_{1}^{n}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu) = \sum_{1}^{n}(x_i - \bar{x})^T \Sigma^{-1}(x_i - \bar{x}) + n(\bar{x} - \mu)^T \Sigma^{-1}(\bar{x} - \mu)$$

(the cross-terms cancel as $\sum(x_i - \bar{x}) = 0$). The summand in the first term on the right is a scalar, so is its own trace. Since $trace(AB) = trace(BA)$ and $trace(A + B) = trace(B + A)$,

$$trace(\sum_{1}^{n}(x_i - \bar{x})^T \Sigma^{-1}(x_i - \bar{x})) = trace(\Sigma^{-1}\sum_{1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T)$$

$$= trace(\Sigma^{-1}.nS) = n \; trace(\Sigma^{-1}S).$$

Combining,

$$L = (2\pi)^{-np/2}|\Sigma|^{-n/2} \exp\{-\frac{1}{2}n[trace(\Sigma^{-1}S) + (\bar{x} - \mu)^T \Sigma^{-1}(\bar{x} - \mu)]\}. \quad \mathbf{[12]}$$

So by the Fisher-Neyman Theorem, $(\bar{X}, S)$ is sufficient for (μ, Σ). **[4]**
(Seen – lectures)

Q5. $ARMA(1,1)$.

$$X_t = \phi X_{t-1} + \epsilon_t + \theta\epsilon_{t-1} : \qquad (1-\phi B)X_t = (1+\theta B)\epsilon_t.$$

Condition for stationarity and invertibility: $|\phi| < 1$; $|\theta| < 1$. **[2, 2]**

Assuming these:

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

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

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

Variance: lag $\tau = 0$. Square and take expectations. The ϵs are uncorrelated with variance σ^2, so

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

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

$$\gamma_0 = \sigma^2(1+2\phi\theta+\theta^2)/(1-\phi^2)$$ **[8]**

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:

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

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

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 for $\tau \geq 1$

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

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

Q6 *Poisson with Gamma prior.*

Data: $x = (x_1, \cdots, x_n)$, x_i independent, Poisson $P(\theta)$:

$$f(x|\theta) = \Pi_1^n f(x_i|\theta) = \theta^{x_1+\cdots+x_n} e^{-n\theta}/x_1! \cdots x_n! = \theta^{n\bar{x}} e^{-n\theta}/\Pi x_i!,$$

where $\bar{x} := \frac{1}{n}\Sigma x_i$ is the sample mean.
Prior: the Gamma density $\Gamma(a, b) \quad (a, b > 0)$:

$$f(\theta) = \frac{a^b \theta^{b-1}}{\Gamma(b)} e^{-a\theta} \qquad (\theta > 0):$$

$$f(x|\theta) f(\theta) = \frac{a^b}{\Gamma(b)\Pi x_i!} \theta^{n\bar{x}+b-1} e^{-(n+a)\theta},$$

$$f(\theta|x) \propto f(x|\theta) f(\theta) = const. \theta^{n\bar{x}+b-1} e^{-(n+a)\theta}.$$

This has the form of a Gamma density. So, it *is* a Gamma density, $\Gamma(n + a, n\bar{x} + b)$: the posterior density is

$$f(\theta|x) = \frac{(n+a)^{n\bar{x}+b}}{\Gamma(n\bar{x}+b)} . \theta^{n\bar{x}+b-1} e^{-(n+a)\theta} \qquad (\theta > 0). \qquad \textbf{[8]}$$

Means. For $\Gamma(a, b)$, the mean is

$$E\theta = \int_0^\infty \theta f(\theta) d\theta = \frac{a^b}{\Gamma(b)} . \int_0^\infty \theta^b e^{-a\theta} d\theta, = b/a$$

(with $t := a\theta$, the integral is $\Gamma(b + 1)/a^{b+1}, = b\Gamma(b)/a^{b+1}$ as $\Gamma(x + 1) = x\Gamma(x)$). **[4]**

So by above, the prior mean is b/a; the posterior mean is $(n\bar{x}+b)/(n+a)$; the data mean is $\bar{x}$. Write

$$\lambda := a/(n + a), \qquad \text{so } 1 - \lambda = n/(n + a): \qquad \text{since}$$

$$\frac{n\bar{x} + b}{n + a} = \frac{a}{n + a} . \frac{b}{a} + \frac{n}{n + a} . \bar{x},$$

posterior mean $(n\bar{x}+b)/(n+a) = \lambda$. prior mean $b/a+(1-\lambda)$. sample mean $\bar{x}$.

Again, this is a weighted average, with weights proportional to n and a. Now n is the sample size, a measure of the precision of the data, and a is the rate of decay of the Gamma density, a measure of the precision of the prior information. **[8]**

[Seen – lectures]