smfexamsoln(13).tex

STATISTICAL METHODS FOR FINANCE: EXAM SOLUTIONS 2013

Q1. *Fisher information.*

The joint density is $f = f(x_1, \ldots, x_n; \theta)$: $f = f(x; \theta)$. The *likelihood* is $L = L(\theta) := f(x; \theta)$, with the *data* here as $x = (x_1, \ldots, x_n)$ (so L is a *statistic* – can be calculated from the data). The *log-likelihood* is $\ell = \ell(\theta) := \log L(\theta)$.

The (Fisher) *score function* is $s(\theta) := \ell'(\theta)$. **[2]**

The (Fisher) *information* is defined by either $I(\theta) := E[s(\theta)^2]$ (so $I \geq 0$), or $I(\theta) := -E[s'(\theta)]$ (so I is additive) (these are equivalent – see below). **[2]**

The density integrates to 1: $\int f(x; \theta) dx = 1$: $\int f = 1$. We assume f smooth enough to differentiate under the integral sign 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 $E[g(X)] = \int g(x) f(x; \theta) dx = \int g f$, so $E[\partial \log L / \partial \theta] = 0$: $E[\partial \ell / \partial \theta] = 0$: $E[\ell'(\theta)] = 0$: $E[s(\theta)] = 0$. **[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

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

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

$$E[s(\theta)^2 + s'(\theta)] = 0.$$ **[6]**

As above, write $I(\theta) = E[s^2(\theta)] = -E[s'(\theta)]$, and call $I(\theta)$ the (Fisher) *information* on θ (in the sample $(x_1, \ldots, x_n)$). So, combining:

The score function $s(\theta) := \ell'(\theta)$ has mean 0 and variance $I(\theta)$. **[2]**

Application.

The Fisher information appears in the large-sample theory of maximum-likelihood estimation (MLE): in the regular case, with θ_0 the true value of the parameter θ, $\sqrt{n I(\theta_0)} (\hat{\theta} - \theta_0) \to \Phi = N(0,1)$ $(n \to \infty)$. **[2]**

[Seen – lectures]

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. *Sufficiency and the factorisation criterion.*

We give a Bayesian treatment of sufficiency, as this is easier than the classical one (for which see e.g. I.4, Day 2, Course website).

If $x = (x_1, x_2)$, where x_1 is informative about θ: we call x_1 *sufficient* for θ if x_2 is uninformative, i.e. x_2 cannot affect our views on θ, i.e.
(i) $f(\theta|x) = f(\theta|x_1, x_2)$ does not depend on x_2, i.e.

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

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

(ii) $f(x_2|x_1, \theta)$ does not depend on θ.
Either of (i), (ii) can be used as the definition of sufficiency in a Bayesian treatment. [Notice that (i) is essentially a Bayesian statement: it is meaningless in classical statistics, as there θ cannot have a density.] **[5]**

Now recall the classical Fisher-Neyman Factorisation Criterion for sufficiency: a statistic x_1 is *sufficient* for the parameter θ iff the likelihood $f(x|\theta)$ factorises as
(iii) $f(x|\theta)$, or $f(x_1, x_2|\theta)$, $= g(x_1, \theta)h(x_1, x_2)$,
for some functions g, h. **[5]**

Proposition. x_1 is sufficient for θ iff the Factorisation Criterion (iii) holds.

Proof. (ii) $\Rightarrow$ (iii):

$$\begin{aligned} f(x|\theta) = f(x_1, x_2|\theta) &= f(x_1|\theta)f(x_2|x_1, \theta) \qquad \text{(as in 2 above)} \\ &= f(x_1|\theta)f(x_2|x_1) \qquad \text{(by (ii))}, \end{aligned}$$

giving (iii).
(iii) $\Rightarrow$ (i): By Bayes' Theorem in the form 'posterior proportional to prior times likelihood', the factor $h(x_1, x_2)$ in (iii) can be absorbed into the constant of proportionality [which is unimportant: it can be recovered from the remaining terms, its role being merely to make these integrate to one]. Then x_2 drops out, so does not appear in the posterior, giving (i). **[10]**
[Seen – lectures]

Q4. *Regression plane.*

With two regressors u and v and response variable y, given a sample of size n of points $(u_1, v_1, y_1), \ldots, (u_n, v_n, y_n)$ we have to fit a least-squares *plane* – that is, choose parameters a, b, c to minimise the sum of squares

$$SS := \sum_{i=1}^{n} (y_i - c - au_i - bv_i)^2.$$

Taking $\partial SS/\partial c = 0$ gives

$$\sum_{i=1}^{n} (y_i - c - au_i - bv_i) = 0 : \qquad c = \bar{y} - a\bar{u} - b\bar{v}.$$

$$SS = \sum_{i=1}^{n} [(y_i - \bar{y}) - a(u_i - \bar{u}) - b(v_i - \bar{v})]^2.$$

Then $\partial SS/\partial a = 0$ and $\partial SS/\partial b = 0$ give

$$\sum_{i=1}^{n} (u_i - \bar{u})[(y_i - \bar{y}) - a(u_i - \bar{u}) - b(v_i - \bar{v})],$$

$$\sum_{i=1}^{n} (v_i - \bar{v})[(y_i - \bar{y}) - a(u_i - \bar{u}) - b(v_i - \bar{v})].$$

Multiply out, divide by n to turn the sums into averages, and re-arrange:

$$as_{uu} + bs_{uv} = s_{yu},$$

$$as_{uv} + bs_{vv} = s_{yv}.$$

These are the *normal equations* (NE) for a and b. **[10]**

Condition for non-degeneracy. The determinant is

$$s_{uu}s_{vv} - s_{uv}^2 = s_{uu}s_{vv}(1 - r_{uv}^2)$$

(as $r_{uv} := s_{uv}/(s_u.s_v)$), $\neq 0$ iff $r_{uv} \neq \pm 1$, i.e., iff the (u_i, v_i) are not collinear, and this is the condition for (NE) to have a unique solution. **[4]**

Application: Grain futures.

The two principal factors affecting grain yields (apart from the weather near harvest – unpredictable!) are *sunshine* (in hours) and rainfall (in mm) during the *spring* growing season (known in advance). Using these as *predictor* variables u, v gives a best (linear unbiased) estimator of grain yield y.

The volumes of grain traded yearly are enormous. So, the ability to predict as accurately as possible the size of the summer harvest (and so, by supply and demand, its price), given information available in the spring, is very valuable. Such predictions can be used to form trading strategies for grain futures and grain options, etc. (example, the Great Grain Steal of 1972, by the then USSR, on the USA and Canada). **[6]**

[Seen: problem sheets]

Q5. *Yule-Walker equations and AR*(2).

The $AR(p)$ model is (with (ϵ_t) white noise $WN(\sigma)$)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t. \qquad [\mathbf{2}]$$

Multiply by X_{t-k} and take E: as $E[X_{t-k}X_{t-i}] = \rho(|k-i|) = \rho(k-i)$,

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

These are the *Yule-Walker equations.* **[4]**

They give a *difference equation* of *order p*, with *characteristic polynomial*

$$\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_p = 0.$$

If the roots are $\lambda_1, \cdots, \lambda_p$, the trial solution $\rho(k) = \lambda^k$ is a solution iff λ is one of the roots λ_i. Since the equation is linear,

$$\rho(k) = c_1\lambda_1^k + \cdots + c_p\lambda_p^k$$

(for $k \geq 04$ and use $\rho(-k) = \rho(k)$ for $k < 0$) is a solution for all choices of constants c_i – the *general solution* of (YW) if all the roots λ_i are distinct. **[4]**

Example of an AR(2) *process.*

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

The Yule-Walker equations here are $\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2)$.
The characteristic polynomial is

$$\lambda^2 - \frac{1}{3}\lambda - \frac{2}{9} = 0: \quad (\lambda - 2/3)(\lambda + 1/3) = 0; \quad \lambda_1 = 2/3,\ \lambda_2 = -1/3.$$

So as the roots are distinct, the autocovariance is $\rho(k) = a\lambda_1^k + b\lambda_2^k$. **[5]**
$k = 0$: $\rho(0) = 1$ gives $a+b = 1$: $b = 1-a$. So $\rho(k) = a(2/3)^k + (1-a)(-1/3)^k$.
$k = 1$: $\rho(1) = \phi_1\rho(0) + \phi_2\rho(-1)$; as $\rho(0) = 1$ and $\rho(-1) = \rho(1)$, $\rho(1) = \phi_1/(1-\phi_2)$. As here $\phi_1 = 1/3$ and $\phi_2 = 2/9$, this gives $\rho(1) = 3/7$. So

$$\rho(1) = 3/7 = a.(2/3) + (1-a).(-1/3).$$

That is,

$$(\frac{3}{7} + \frac{1}{3}) = a.(\frac{2}{3} + \frac{1}{3}) = a:$$

$a = (9+7)/21 = 16/21$. Thus

$$\rho(k) = \frac{16}{21}(\frac{2}{3})^k + \frac{5}{21}(\frac{-1}{3})^k. \qquad [\mathbf{5}]$$

[Seen, lectures]

Q6. *The Bayes linear estimator.*

If $d(z)$ is a *linear* function, $a+b'z$, where z and b are vectors, the quadratic loss is

$$\begin{aligned} D &= E[(a+b'z-\theta)^2] \\ &= E[a^2+2ab'z+b'zz'b-2a\theta-2b'z\theta+\theta^2] \end{aligned}$$

$$= a^2+2ab'Ez+b'E(zz')b-2aE\theta-2b'E(z\theta)+E(\theta^2). \qquad \textbf{[4]}$$

Add and subtract $[E(\theta)]^2$, $(b'Ez)^2 = b'EzEz'b$ and $2b'EzE\theta$. Write $V := var\ z = E(zz') - EzEz'$ for the covariance matrix of z, $c := cov(\theta, z) = E(z\theta) - EzE\theta$ for the covariance vector between θ and the elements of the vector z.

$$D = (a+b'Ez-E\theta)^2 + b'(var z)b - 2b'cov(z,\theta) + var\theta :$$

$$D = (a+b'Ez-E\theta)^2 + b'Vb - 2b'c + var\theta. \qquad \textbf{[4]}$$

Write $b^* := V^{-1}c$, $D^* := var(\theta) - c'V^{-1}c$. Then this becomes

$$D = (a+b'Ez-E\theta)^2 + (b-b^*)'V(b-b^*) + D^* \qquad (*)$$

(the quadratic terms check as $b^{*T}Vb* = c^TV^{-1}VV^{-1}c = c^TV^{-1}c$, the linear terms as $c = Vb^*$). **[4]**

The third term on the right in (2) does not involve a, b, while the first two are non-negative (the first is a square, the second a quadratic form with matrix V, non-negative definite as V is a covariance matrix). So the expected quadratic loss D is minimised by choosing $b = b^*$, $a = -b^{*\prime}Ez + E\theta$. This choice gives

$$d(z) = E\theta + cV^{-1}(z - Ez), \qquad c := cov(z,\theta), \quad V := var(z).$$

This gives the *Bayes linear estimator* of θ based on data $z = z(x)$. **[4]**

Distributional assumptions.

The Bayes linear estimator depends only on first and second moments: $E\theta$, Ez, $c = cov(z,\theta)$, $V = var(z)$. So we do not need to know the full likelihood, just the first and second moments of $(\theta, z(x))$, the parameter and the function z in which we want the estimator to be linear. **[2]**

Application.

The Bayes linear estimator is used in the construction of the *Kalman filter* – state-space models for Time Series. **[2]**

[Seen, lectures]