SAMPLE EXAM QUESTIONS - SOLUTION

As you might have gathered if you attempted these problems, they are quite long relative to the 24 minutes you have available to attempt similar questions in the exam; I am aware of this. However, these questions were designed to cover as many of the topics we studied in the course.

SAMPLE EXAM QUESTION 1 - SOLUTION

(a) *State Cramer's result (also known as the Delta Method) on the asymptotic normal distribution of a (scalar) random variable Y defined in terms of random variable X via the transformation $Y = g(X)$, where X is asymptotically normally distributed*

$$X \sim AN\left(\mu, \sigma^2\right).$$

This is bookwork. If the derivative, $\dot{g}$, of g is non-zero, then

$$Y \sim AN(g(\mu), \{\dot{g}(\mu)\}^2 \sigma^2)$$

We saw this result in a slightly different form in the lectures, stated as

$$\sqrt{n}(X_n - \mu) \xrightarrow{\mathcal{L}} N(0, \sigma^2) \qquad \Longrightarrow \qquad \sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{\mathcal{L}} N(0, \{\dot{g}(\mu)\}^2\sigma^2)$$

and these two results are equivalent.

[4 MARKS]

(b) *Suppose that $X_1, ..., X_n$ are independent and identically distributed Poisson(λ) random variables. Find the maximum likelihood (ML) estimator, and an asymptotic normal distribution for the estimator, of the following parameters*

(i) λ,

(ii) $\exp\{-\lambda\}$.

(i) By standard theory, log-likelihood is

$$l(\lambda) = -n\lambda + s_n \log \lambda - \log\left(\prod_{i=1}^{n} x_i!\right)$$

where $s_n = \sum_{i=1}^{n} x_i$ (with corresponding random variable $S_n = \sum_{i=1}^{n} X_i$) so

$$\dot{l}(\lambda) = -n + \frac{s_n}{\lambda}$$

so equating this to zero yields $\hat{\lambda} = s_n/n = \bar{x}$. Using the Central Limit Theorem,

$$\frac{(S_n - n\lambda)}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0, \lambda)$$

as $\mathrm{E}[X] = \mathrm{Var}[X] = \lambda$. Hence

$$S_n \sim AN(n\lambda, n\lambda) \qquad \Longrightarrow \qquad \bar{X} = \frac{S_n}{n} \sim AN(\lambda, \lambda/n)$$

[3 MARKS]

(ii) By invariance of ML estimators to reparameterization, or from first principles, the ML estimator of $\phi = \exp(-\lambda)$ is $\hat{\phi} = \exp(-\bar{X}) = T_n$, say.

For Cramer's Theorem (Delta Method), let $g(t) = \exp(-t)$, so that $\dot{g}(t) = -\exp(-t)$. Thus

$$T_n \sim AN(\exp(-\lambda), \exp(-2\lambda)\lambda/n)$$

[3 MARKS]

(c) *Suppose that, rather than observing the random variables in (b) precisely, only the events*

$$X_i = 0 \qquad \textit{or} \qquad X_i > 0$$

for $i = 1, ..., n$ *are observed.*

(i) *Find the ML estimator of* λ *under this new observation scheme.*

(ii) *In this new scheme, when does the ML estimator not exist (at a finite value in the parameter space) ? Justify your answer.*

(iii) *Compute the probability that the ML estimator does not exist for a finite sample of size* n*, assuming that the true value of* λ *is* λ_0*.*

(iv) *Construct a modified estimator that is consistent for* λ *.*

(i) We now effectively have a Bernoulli sampling model; let Y_i be a random variable taking the value 0 if $X_i = 0$, and 1 otherwise; note that $P[Y_i = 0] = P[X_i = 0] = \exp(-\lambda) = \theta$, say, so that the log likelihood is

$$l(\theta) = (n - m)\log\theta + m\log(1-\theta)$$

where $m = \sum_{i=1}^{n} y_i$, the number of times that Y_i, and hence X_i, is greater than zero. From this likelihood, the ML estimate of θ is $\hat{\theta} = (n-m)/n$, and hence the ML estimate of λ is

$$\hat{\lambda} = -\log(\hat{\theta}) = -\log((n-m)/n)$$

and the estimator is $T_n = -\log(n^{-1}\sum_{i=1}^{n} Y_i)$

[2 MARKS]

(ii) This estimate is not finite if $m = n$, that is, if we never observe $X_i=0$ in the sample, so that $m = \sum_{i=1}^{n} y_i = n$.

[2 MARKS]

(iii) The event of interest from (ii) occurs with the following probability:

$$P\left[\sum_{i=1}^{n} Y_i = n\right] = \prod_{i=1}^{n} P[Y_i = 1] = \prod_{i=1}^{n}[1 - \exp(-\lambda_0)] = (1 - \exp(-\lambda_0))^n$$

which, if λ_0 is not large, can be appreciable. Thus, for a finite value of n, there is a non-zero probability that the estimator is not finite.

[3 MARKS]

(iv) Consistency (weak or strong) for λ will follow from the consistency of the estimator of θ, as we have, from the Strong Law

$$\frac{\sum_{i=1}^{n} Y_i}{n} \xrightarrow{a.s.} \theta$$

The only slight practical problem is that raised in (ii) and (iii), the finiteness of the estimator. We can overcome this by defining the estimator as follows; estimate λ by

$$T_n' = \begin{cases} -\log(n^{-1}\sum_{i=1}^{n} Y_i) & \text{if } \max\{Y_1, \ldots, Y_n\} > 0 \\ \\ k & \text{if } \max\{Y_1, \ldots, Y_n\} = 0 \end{cases}$$

where k is some constant value. As the event $(\max\{Y_1, \ldots, Y_n\} = 0)$ occurs with probability $(1 - \exp(-\lambda_0))^n$ which converges to 0 as $n \longrightarrow \infty$, this adjustment does not disrupt the strong convergence. Note that we could choose $k = 1$, or $k = 10^{10^6}$, and consistency would be preserved.

[3 MARKS]

SAMPLE EXAM QUESTION 2 - SOLUTION

(a) *Suppose that $X_{(1)} < \ldots < X_{(n)}$ are the order statistics from a random sample of size n from a distribution F_X with continuous density f_X on $\mathbb{R}$. Suppose $0 < p_1 < p_2 < 1$, and denote the quantiles of F_X corresponding to p_1 and p_2 by x_{p_1} and x_{p_2} respectively.*

Regarding x_{p_1} and x_{p_2} as unknown parameters, natural estimators of these quantities are $X_{(\lceil np_1 \rceil)}$ and $X_{(\lceil np_2 \rceil)}$ respectively, where $\lceil x \rceil$ is the smallest integer not less than x. Show that

$$\sqrt{n}\begin{pmatrix} X_{(\lceil np_1 \rceil)} - x_{p_1} \\ X_{(\lceil np_2 \rceil)} - x_{p_2} \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \Sigma)$$

where

$$\Sigma = \begin{bmatrix} \dfrac{p_1(1-p_1)}{\{f_X(x_{p_1})\}^2} & \dfrac{p_1(1-p_2)}{f_X(x_{p_1})f_X(x_{p_2})} \\ \\ \dfrac{p_1(1-p_2)}{f_X(x_{p_1})f_X(x_{p_2})} & \dfrac{p_2(1-p_2)}{\{f_X(x_{p_2})\}^2} \end{bmatrix}$$

State the equivalent result for a single quantile x_p corresponding to probability p.

This is bookwork, from the handout that I gave out in lectures. In solving the problem, it is legitimate to state without proof some of the elementary parts; in terms of the handout, after describing the set up, you would be allowed to quote without proof Results 1 through 3, and would only need to give the full details for the final parts.

For the final result, for a single quantile x_p, we have that

$$\sqrt{n}\left(X_{(\lceil np \rceil)} - x_p\right) \xrightarrow{\mathcal{L}} N\left(0, \frac{p(1-p)}{\{f_X(x_p)\}^2}\right)$$

[10 MARKS]

(b) *Using the results in (a), find the asymptotic distribution of*

(i) *The sample median estimator of the median F_X (corresponding to $p = 0.5$), if F_X is a Normal distribution with parameters μ and σ^2.*

(ii) *The upper and lower quartile estimators (corresponding to $p_1 = 0.25$ and $p_2 = 0.75$) if F_X is an Exponential distribution with parameter λ*

(i) Here we have $p = 0.5$, and $x_p = \mu$, as the Normal distribution is symmetric about μ.

$$\sqrt{n}\left(X_{(\lceil n/2 \rceil)} - \mu\right) \xrightarrow{\mathcal{L}} N\left(0, \frac{(1/2)(1/2)}{\{\phi(0)\}^2}\right) \equiv N\left(0, \frac{\pi\sigma^2}{2}\right)$$

as $\phi(0) = 1/\sqrt{2\pi\sigma^2}$ and hence $X_{(\lceil n/2 \rceil)} \sim AN(\mu, \pi\sigma^2/2n)$.

[3 MARKS]

(ii) For probability p the corresponding quantile is given by

$$p = F_X(x;\lambda) = 1 - e^{-\lambda x_p} \qquad \Longrightarrow \qquad x_p = -\log(1-p)/\lambda$$

and $f_X(x;\lambda) = \lambda e^{-\lambda x}$. Let $p_1 = 1/4, p_2 = 3/4$, and $c_1 = -\log(1-1/4)/\lambda$ and $c_2 = -\log(1-3/4)/\lambda$. Then the key asymptotic covariance matrix is

$$\Sigma = \begin{bmatrix} \dfrac{(1/4)(3/4)}{\lambda^2 e^{-2\lambda c_1}} & \dfrac{(1/4)(1/4)}{\lambda^2 e^{-\lambda(c_1+c_2)}} \\ \\ \dfrac{(1/4)(1/4)}{\lambda^2 e^{-\lambda(c_1+c_2)}} & \dfrac{(3/4)(1/4)}{\lambda^2 e^{-2\lambda c_2}} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{3\lambda^2} & \dfrac{1}{9\lambda^2} \\ \\ \dfrac{1}{9\lambda^2} & \dfrac{3}{\lambda^2} \end{bmatrix}$$

which gives that

$$\begin{pmatrix} X_{(\lceil n/4 \rceil)} \\ X_{(\lceil 3n/4 \rceil)} \end{pmatrix} \sim AN\left(\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \begin{bmatrix} \dfrac{1}{3n\lambda^2} & \dfrac{1}{9n\lambda^2} \\ \\ \dfrac{1}{9n\lambda^2} & \dfrac{3}{n\lambda^2} \end{bmatrix} \right)$$

[3 MARKS]

(c) *The results in (a) and (b) describe convergence in law for the estimators concerned. Show how the form of convergence may be strengthened using the Strong Law for any specific quantile x_p.*

The standard Strong Law result says, effectively, that for i.i.d. random variables $X_1, X_2, \ldots$, for arbitrary function G

$$\frac{1}{n}\sum_{i=1}^{n} G(X_i;\theta) \xrightarrow{a.s.} E_{X|\theta}[G(X)].$$

So, here, if we define $G(X_i;\theta) = 1$ if $X_i \leq x_p$, and zero otherwise, then

$$U_n = \frac{1}{n}\sum_{i=1}^{n} G(X_i;\theta) \xrightarrow{a.s.} E_{X|\theta}[G(X)] = P[X \leq x_p] = p$$

and we have strong convergence of the statistic on the left-hand side to p. Now F_X^{-1} is a continuous, monotone increasing function, so we can map both sides of the last result by F_X^{-1} to obtain the result

$$F_X^{-1}(U_n) \xrightarrow{a.s.} F_X^{-1}(p) = x_p.$$

[4 MARKS]

SAMPLE EXAM QUESTION 3 : SOLUTION

(a) (i) *State (without proof) Wald's Theorem on the strong consistency of maximum likelihood (ML) estimators, listing the five conditions under which this theorem holds.*

Bookwork (although we focussed less on strong consistency of the MLE this year, and studied weak consistency in more detail): Let $X_1, \ldots, X_n$ be i.i.d. with pdf $f_X(x|\theta)$ (with respect to measure ν), let Θ denote the parameter space, and let θ_0 denote the true value of the parameter θ. Suppose θ is 1-dimensional. Then, if

(1) Θ is compact,

(2) $f_X(x|\theta)$ is upper semi-continuous (USC) in θ on Θ for all x, that is for all $\theta \in \Theta$ and any sequence $\{\theta_n\}$ such that $\theta_n \longrightarrow \theta$

$$\limsup_{n \longrightarrow \infty} f_X(x|\theta_n) \leq f_X(x|\theta)$$

or equivalently for all $\theta \in \Theta$

$$\sup_{|\theta'-\theta|<\delta} f_X(x|\theta') \longrightarrow f(x|\theta) \qquad \text{as} \qquad \delta \longrightarrow 0$$

for all x,

(3) there exists a function $M(x)$ with $E_{f_{X|\theta_0}}[M(x)] < \infty$ and

$$U(x,\theta) = \log f_X(x|\theta) - \log f_X(x|\theta_0) \leq M(x)$$

for all x and θ,

(4) for all $\theta \in \Theta$, and sufficiently small $\delta > 0$,

$$\sup_{|\theta'-\theta|<\delta} f_X(x|\theta')$$

is measurable (wrt ν) in x.

(5) If $f_X(x|\theta) = f_X(x|\theta_0)$ almost everywhere wrt ν in x, then $\theta = \theta_0$; this is the **identifiability** condition,

any sequence of ML estimators $\{\hat{\theta}_n\}$ of θ is strongly consistent for θ, that is

$$\hat{\theta}_n \xrightarrow{a.s.} \theta_0$$

as $n \longrightarrow \infty$.

[5 MARKS]

(ii) *Verify that the conditions of the theorem hold when random variables $X_1, \ldots, X_n$ correspond to independent observations from the Uniform density on $(0, \theta)$*

$$f_X(x|\theta) = \frac{1}{\theta} \qquad 0 \leq x \leq \theta$$

and zero otherwise, for parameter $\theta \in \Theta \equiv [a,b]$, where $[a,b]$ is the closed interval from a to b, $0 < a < b < \infty$.

[Hint: for $x \in \mathbb{R}$, consider the function

$$M(x) = \max_{\theta \in \Theta} \frac{f_X(x|\theta)}{f_X(x|\theta_0)}$$

where $\theta_0 \in \Theta$ is the true value of θ.]

The governing measure here is Lebesgue measure.

(1) $\Theta \equiv [a, b]$ is closed and bounded, therefore compact.

(2) Need to consider all possible (fixed) $x \in \mathbb{R}$. Now, when $x \leq a$,

$$f_X(x|\theta) = \frac{1}{\theta}$$

which is continuous in θ. From the second USC definition, it is easy to see that continuous functions are USC. Similarly, when $x > b$,

$$f_X(x|\theta) = 0$$

which is continuous in θ, and hence USC. Finally, for $a < x \leq b$,

$$f_X(x|\theta) = \begin{cases} 0 & \theta < x \\ \frac{1}{\theta} & \theta \geq x \end{cases}$$

which is not continuous at $x = \theta$, but is USC, as for all $\delta > 0$,

$$\sup_{|\theta' - \theta| < \delta} f_X(x|\theta') = \frac{1}{\theta} = f_X(x|\theta)$$

($|\theta' - \theta| < \delta$ defines an interval centered at θ, to the left of θ the function is zero, to the right of θ the function is $1/\theta$, so the supremum over the interval is always $1/\theta$.)

(3) If

$$M(x) = \max_{\theta \in \Theta} \frac{f_X(x|\theta)}{f_X(x|\theta_0)}$$

then

$$M(x) = \begin{cases} \frac{\theta_0}{a} & x \leq a \\ \frac{\theta_0}{x} & a < x \leq \theta_0 \\ \infty & x > \theta_0 \end{cases}$$

The expectation of $M(X)$, when $\theta = \theta_0$, is finite as the third case is excluded ($P[X > \theta_0] = 0$).

(4) f_X is measurable (by definition), and supremum operations preserve measurability.

(5) Identifiability is assured, as different θ values yield densities with different supports.

[5 MARKS]

(b) *Wald's Theorem relates to one form of consistency; the remainder of the question focuses on another form.*

We are now dealing with weak consistency ...

Suppose that random variables $X_1, \ldots, X_n$ correspond independent observations from density (wrt Lebesgue measure) $f_X(x|\theta)$, and for $\theta \in \Theta$, this family of densities have common support $\mathbb{X}$. Let the true value of θ be denoted θ_0, and let $L_n(\theta)$ denote the likelihood for θ

$$L_n(\theta) = \prod_{i=1}^{n} f_X(x_i|\theta).$$

(i) *Using Jensen's inequality for the function $g(x) = -\log x$, and an appropriate law of large numbers, show that*

$$P_{\theta_0}\left[L_n(\theta_0) > L_n(\theta)\right] \longrightarrow 1 \qquad \text{as} \qquad n \longrightarrow \infty$$

for any ***fixed*** *$\theta \neq \theta_0$, where P_{θ_0} denotes probability under the true model, indexed by θ_0.*

This follows in a similar fashion to the proof of the positivity of the Kullback-Liebler (K) divergence;

$$L_n(\theta_0) > L_n(\theta) \Leftrightarrow \frac{L_n(\theta_0)}{L_n(\theta)} > 1 \Leftrightarrow \log\frac{L_n(\theta_0)}{L_n(\theta)} > 0 \Leftrightarrow \sum_{i=1}^{n} \log\frac{f_X(X_i|\theta_0)}{f_X(X_i|\theta)} > 0 \qquad (1)$$

Now, by the weak law of large numbers

$$T_n(\theta_0,\theta) = \frac{1}{n}\sum_{i=1}^{n} \log\frac{f_X(X_i|\theta_0)}{f_X(X_i|\theta)} \xrightarrow{p} E_{f_{X|\theta_0}}\left[\log\frac{f_X(X|\theta_0)}{f_X(X|\theta)}\right] = K(\theta_0,\theta) \qquad (2)$$

To finish the proof we use the Kullback-Liebler proof method; from Jensen's inequality

$$\begin{aligned} E_{f_{X|\theta_0}}\left[\log\frac{f_X(X|\theta_0)}{f_X(X|\theta)}\right] &= -E_{f_{X|\theta_0}}\left[\log\frac{f_X(X|\theta)}{f_X(X|\theta_0)}\right] \geq -\log E_{f_{X|\theta_0}}\left[\frac{f_X(X|\theta)}{f_X(X|\theta_0)}\right] \\ &= -\log\int \frac{f_X(x|\theta)}{f_X(x|\theta_0)} f_X(x|\theta_0)d\nu \\ &= -\log\int f_X(x|\theta)d\nu \geq -\log 1 = 0. \end{aligned}$$

with equality if and only if $\theta = \theta_0$. Thus, by (1) and (2)

$$T_n(\theta_0,\theta) \xrightarrow{p} K(\theta_0,\theta) > 0$$

so that

$$P_{\theta_0}[L_n(\theta_0) > L_n(\theta)] = P_{\theta_0}[T_n(\theta_0,\theta) > 0] \longrightarrow 1$$

as $n \longrightarrow \infty$.

Which other condition from (a)(i) needs to be assumed in order for the result to hold ? Identifiability; the strictness of the inequality relies on $\theta \neq \theta_0$.

[5 MARKS]

(ii) *Suppose that, in addition to the conditions listed in (b), parameter space* Θ *is finite, that is,* $\Theta \equiv \{t_1, \ldots, t_p\}$ *for some positive integer* p.

Show that, in this case, the ML estimator $\hat{\theta}_n$ *exists, and is weakly consistent for* θ_0.

This follows from the result in (b)(i); the standardized log-likelihood

$$\frac{1}{n} l(\theta; x) = \frac{1}{n} \log L(\theta; x) = \frac{1}{n} \sum_{i=1}^{n} \log f_X(X_i|\theta)$$

evaluated at $\theta = \theta_0 = t^\star \in \Theta$, say, is greater than the log-likelihood evaluated at any other value $t \in \Theta$ with probability 1, as $n \longrightarrow \infty$. Thus the sequence of ML estimators $\left\{\hat{\theta}_n\right\}$ is such that

$$\lim_{n \to \infty} P[\hat{\theta}_n \neq \theta_0] = 0$$

which is the definition for weak consistency. Note that existence of the ML estimator (as a finite value in the parameter space) is guaranteed for every n, as Θ is finite, and uniqueness of the ML estimator is also guaranteed, with probability tending to 1, as $n \to \infty$.

[5 MARKS]

SAMPLE EXAM QUESTION 4 : SOLUTION

(a) (i) *Give definitions for the following modes of stochastic convergence, summarizing the relationships between the various modes;*

- *convergence in law (convergence in distribution)*
- *convergence almost surely*
- *convergence in r^{th} mean*

Bookwork: For a sequence of rvs $X_1, X_2, \ldots$ with distribution functions $F_{X_1}, F_{X_2}, \ldots$ with and common governing probability measure P on space Ω with associated sigma algebra $\mathcal{A}$;

(a) Convergence in Law

$$X_n \xrightarrow{\mathcal{L}} X \qquad \Longleftrightarrow \qquad F_{X_n}(x) \longrightarrow F_X(x)$$

for all $x \in \mathbb{R}$ at which F_X is continuous, where F_X is a valid cdf on $\mathbb{R}$.

(b) Convergence almost surely

$$X_n \xrightarrow{a.s.} X \qquad \Longleftrightarrow \qquad P\left[\lim_{n\to\infty} X_n(\omega) = X(\omega)\right] = 1$$

almost everywhere with respect to P (that is, for all $\omega \in \Omega$ except in sets $A \in \mathcal{A}$ such that $P(A) = 0$. Equivalently,

$$X_n \xrightarrow{a.s.} X \qquad \Longleftrightarrow \qquad P\left[\lim_{n\to\infty} |X_n(\omega) - X(\omega)| < \epsilon\right] = 1, \ \forall \epsilon > 0, \ \text{a.e. } P.$$

Also equivalently,

$$X_n \xrightarrow{a.s.} X \qquad \Longleftrightarrow \qquad P\left[|X_n(\omega) - X(\omega)| < \epsilon \text{ i.o.}\right] = 1, \ \forall \epsilon > 0, \ \text{a.e. } P.$$

where i.o. means infinitely often.

(c) Convergence in r^{th} mean

$$X_n \xrightarrow{r} X \qquad \Longleftrightarrow \qquad E\left[|X_n - X|^r\right] \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

for some positive integer r.

In summary, convergence in law is implied by both convergence a.s. and convergence in r^{th} mean, but there are no general relations between the latter two modes.

[6 MARKS]

(ii) *Consider the sequence of random variables $X_1, X_2, \ldots$ defined by*

$$X_n(Z) = nI_{[0,n)}(Z)$$

where Z is a single random variable having an Exponential distribution with parameter 1. Under which modes of convergence does the sequence $\{X_n\}$ converge ? Justify your answer.

In this form, this question is rather boring, as

$$P[X_n = 0] = \exp\{-n\} \to 0$$

so the sequence converges almost surely to infinity, as $X_n = n$ for infinitely many n; let A_n be the event $A_n > M$ for any finite M. Then $P(A_n \text{ occurs i.o.}) = 1$. In fact X_n converges to infinity under all modes.

A more interesting question defines X_n as follows:

$$X_n(Z) = nI_{[n,\infty)}(Z) = n(1 - I_{[0,n)}(Z))$$

in which case

$$P[X_n = 0] = P[Z \leq n] = 1 - \exp\{-n\} \to 1$$

as $n \to \infty$, which makes things more interesting. Direct from the definition, we have $X_n \xrightarrow{a.s.} 0$, as

$$P\left[\lim_{n\to\infty} |X_n| < \epsilon\right] = 1$$

or equivalently

$$\lim_{n\to\infty} P\left[|X_k| < \epsilon, \forall k \geq n\right] = 1.$$

To see this, for some n, n_0 say, $Z \in [0, n_0)$, and thus for all $k > n_0$, $Z \in [0, k)$ also, so $|X_k| = 0 < \epsilon$.

Note that this result follows because we are considering a **single** Z that is used to define the sequence $\{X_n\}$, so that the $\{X_n\}$ are **dependent** random variables. If the $\{X_n\}$ were generated **independently**, using a sequence of independent rvs $\{Z_n\}$, then assessment of convergence would need use of, say, the Borel-Cantelli Lemma (b).

For convergence in r^{th} mean for the new variable: note that

$$E[|X_n|^r] = n^r P[X_n = n] + 0^r P[X_n = 0] = n^r \exp\{-n\} + 0(1 - \exp\{-n\}) \to 0$$

as $n \to \infty$, so $X_n \xrightarrow{r} 0$ for all $r > 0$.

[5 MARKS]

(b) *Suppose that $X_1, X_2, \ldots$ are independent, identically distributed random variables defined on $\mathbb{R}$, with common distribution function F_X for which $F_X(x) < 1$ for all finite x. Let M_n be the maximum random variable defined for finite n by*

$$M_n = \max\{X_1, X_2, \ldots, X_n\}$$

(i) *Show that the sequence of random variables $\{M_n\}$ converges almost surely to infinity, that is*

$$M_n \xrightarrow{a.s.} \infty$$

as $n \to \infty$.

Hint: use the Borel-Cantelli lemma.

(ii) *Now suppose that $F_X(x_U) = 1$ for some $x_U < \infty$. Find the almost sure limiting random variable for the sequence $\{M_n\}$.*

From M2S1 or first principles, the cdf of M_n is

$$F_{M_n}(x) = P[X_i \leq x, \ \forall i] = \prod_{i=1}^{n} P[X_i \leq x] = \{F_X(x)\}^n$$

(i) Now, note that, for any finite x,

$$\sum_{n=1}^{\infty} P[M_n \leq x] = \sum_{n=1}^{\infty} \{F_X(x)\}^n = \frac{F_X(x)}{1 - F_X(x)} < \infty$$

as $F_X(x) < 1$, so we just have a geometric progression in $F_X(x)$. Now let

$$A_x = \limsup_{n\to\infty} (M_n \leq x) = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} (M_k \leq x).$$

By the Borel-Cantelli Lemma (a),

$$P[M_n \leq x \text{ i.o.}] = P\left[\limsup_{n\to\infty}(M_n \leq x)\right] = 0.$$

Thus $P(A_x) = 0$, for any finite x, and thus $P\left(\bigcup_x A_x\right) = P(B) = 0$ also, where the union corresponds to an arbitrary interval in $\mathbb{R}$. To complete the proof, we need to examine $P(B')$, and demonstrate that if $\omega \in B'$, then

$$\lim_{n\to\infty} M_n(\omega) = \infty$$

that is, for all x, there exists $n_0 = n_0(\omega, x)$ such that if $n \geq n_0$ then $M_n(\omega) \geq x$. Now,

$$A'_x = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} (M_k > x).$$

that is, if $\omega \in A'_x$ then there exists an n such that for $k \geq n$, $M_n(\omega) > x$. Thus $B' = \bigcap_x A'_x$ has probability 1 under P, so that for all ω in sets of probability 1,

$$\lim_{n\to\infty} M_n(\omega) = \infty.$$

[5 MARKS]

(ii) We demonstrate that $M_n \xrightarrow{a.s.} x_U$. Fix $\epsilon > 0$. Let $E_n \equiv (M_n < x_U - \epsilon)$. Then

$$\begin{aligned} P[\limsup_{n\to\infty} E_n] = P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} E_k\right) &\leq P\left(\bigcup_{n=1}^{\infty} E_n\right) \\ &\leq \sum_{n=1}^{\infty} P(E_n) \\ &= \sum_{n=1}^{\infty} F_X(x_U - \epsilon)^n \\ &< \infty \end{aligned}$$

so by the Borel-Cantelli Lemma (a),

$$P[E_n \text{ occurs i.o.}] = P[|M_n - x_U| > \epsilon \text{ i.o.}] = 0$$

and thus $M_n \xrightarrow{a.s.} x_U$.

[5 MARKS]

SAMPLE EXAM QUESTION 5 : SOLUTION

(a) *Suppose that $X_1, \ldots, X_n$ are an independent and identically distributed sample from distribution with density $f_X(x|\theta)$, for vector parameter $\theta \in \Theta \subseteq \mathbb{R}^k$. Suppose that f_X is twice differentiable with respect to the elements of θ, and let the true value of θ be denoted θ_0.*

Define

(i) *The Score Statistic (or Score function), $S(X;\theta)$.*

(ii) *The Fisher Information for a single random variable, $I(\theta)$*

(iii) *The Fisher Information for the sample of size n, $I_n(\theta)$.*

(iv) *The Estimated or Observed Fisher Information, $\hat{I}_n(\theta)$.*

[$I(\theta)$ is sometimes called the unit Fisher Information; $\hat{I}_n(\theta)$ is the estimator of $I(\theta)$]

Give the asymptotic Normal distribution of the score statistic under standard regularity conditions, when the data are distributed as a Normal distribution with mean zero and variance $1/\theta$.

Bookwork: Let X and x denote the vector of random variables/observations, let L denote the likelihood, l denote the log likelihood, and let partial differentiation be denoted by dots.

(i) **Score function**:

$$S(X;\theta) = \dot{l}(\theta;X) = \frac{\partial}{\partial\theta}\log L(\theta;X) = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f_{X|\theta}(X_i;\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_{X|\theta}(X_i;\theta)$$

where, by convention the partial differentiation yields a $k \times 1$ column vector.

(ii) **Unit Fisher Information**:

$$I(\theta) = -E_{X_1|\theta}\left[\frac{\partial}{\partial\theta}\left\{\frac{\partial}{\partial\theta}\log f_{X|\theta}(X_1|\theta)\right\}\right]$$

where twice partial differentiation returns a $k \times k$ symmetric matrix. It can be shown that

$$I(\theta) = E_{X_1|\theta}\left[S(X_1;\theta)S(X_1;\theta)^T\right]$$

where $S(X_1,\theta)$ is the score function derived from X_1 only.

(iii) **Fisher Information for** $X_1, \ldots, X_n$:

$$I_n(\theta) = -E_{X|\theta}\left[\frac{\partial}{\partial\theta}\left\{\frac{\partial}{\partial\theta}\log L(X;\theta)\right\}\right] = -E_{X|\theta}\left[\sum_{i=1}^{n}\left\{\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta}\log f_{X|\theta}(X_i;\theta)\right\}\right]$$

so it follows that

$$I_n(\theta) = nI(\theta) = E_{X|\theta}\left[S(X;\theta)S(X;\theta)^T\right]$$

(iv) **Estimated/Observed Fisher Information**:

$$\hat{I}_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta}\log f_{X|\theta}(x_i;\theta)\right\} = \frac{1}{n}\sum_{i=1}^{n} S(x;\theta)S(x;\theta)^T$$

where θ may be replaced by estimate $\widehat{\theta}$ if such an estimate is available. The estimated information from n data points is $n\hat{I}_n(\theta)$.

[8 MARKS]

If $X_1 \dots X_n \sim N(0, \theta^{-1})$ iid, then

$$\begin{aligned}
f_{X|\theta}(X_1|\theta) &= \left(\frac{\theta}{2\pi}\right)^{1/2} \exp\left\{-\frac{\theta X_i^2}{2}\right\} \\
\log f_{X|\theta}(X_1|\theta) &= \frac{1}{2}\log\theta - \frac{1}{2}\log(2\pi) - \frac{\theta X_1^2}{2} \\
\frac{\partial}{\partial\theta}\log f_{X|\theta}(X_1|\theta) &= \frac{1}{2\theta} - \frac{X_1^2}{2} \\
\frac{\partial^2}{\partial\theta^2}\log f_{X|\theta}(X_1|\theta) &= -\frac{1}{2\theta^2}
\end{aligned}$$

Thus, as the expectation of the score function is zero, then

$$S(X;\theta) \sim AN(0, I_n(\theta)) \equiv AN(0, 2n/\theta^2)$$

where AN means asymptotically normal.

[2 MARKS]

(b) *One class of estimating procedures for parameter θ involves solution of equations of the form*

$$G_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} G_i(X_i;\theta) = 0 \qquad (3)$$

for suitably defined functions $G_i, i = 1, \dots, n$.

(i) *Show that maximum likelihood (ML) estimation falls into this class of estimating procedures.* For ML estimation, we find estimator $\widehat{\theta}$, where

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} L(X;\theta)$$

by, typically differentiating $l(X;\theta)$ partially in turn with respect to each component of θ, and then setting the resulting derivative equations equal to zero, that is, we solve the system of k equations

$$\frac{\partial}{\partial\theta}\log l(X;\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f_{X|\theta}(X_i|\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_{X|\theta}(X_i|\theta) = 0$$

which is of the same form as equation (3) with

$$G_i(X_i;\theta) \equiv n\frac{\partial}{\partial\theta}\log f_{X|\theta}(X_i|\theta)$$

(Note that ML estimation does not always coincide with solving these equations, as sometimes the $\arg\max$ of the likelihood lies on the boundary of Θ).

[4 MARKS]

(ii) *Suppose that $\hat{\theta}_n$ is a solution to (3) which is weakly consistent for θ.*

Using a "one-step" approximation to G (motivated by a Taylor expansion of G around θ_0) of the form

$$G_n(\hat{\theta}_n) = G_n(\theta_0) + (\widehat{\theta}_n - \theta_0)\dot{G}_n(\theta_0),$$

where $\dot{G}_n$ is the first partial derivative vector wrt the k components of θ, find an asymptotic normal distribution of $\hat{\theta}_n$.

State precisely the assumptions made in order to obtain the asymptotic Normal distribution.

Apologies, some lax notation here; this is a vector problem, and θ, θ_0, $\widehat{\theta}_n$ and G are conventionally $k \times 1$ (column) vectors, and $\dot{G}_n$ is a $k \times k$ matrix, so it makes more sense to write

$$G_n(\widehat{\theta}_n) = G_n(\theta_0) + \dot{G}_n(\theta_0)(\widehat{\theta}_n - \theta_0) \tag{4}$$

although working through with the form given, assuming row rather than column vectors, is OK. Anyway, proceeding with column vectors:

Now, $\widehat{\theta}_n$ is a solution to equation (3) by definition of the estimator, so rearranging equation (4) after setting the LHS to zero and multiplying through by $\sqrt{n}$ yields

$$\sqrt{n}\, G_n(\theta_0) = -\sqrt{n}\, \dot{G}_n(\theta_0)(\widehat{\theta}_n - \theta_0). \tag{5}$$

But also, by the Central Limit Theorem, **under the assumption** that

$$E_{X|\theta_0}[G_n(\theta_0)] = 0$$

(that is, the usual "unbiasedness" assumption made for score equations), we have

$$\sqrt{n} G_n(\theta_0) \xrightarrow{\mathcal{L}} Z \sim N(0, V_G(\theta_0))$$

where

$$V_G(\theta_0) = Var_{X|\theta_0}[G_n(\theta_0)]$$

But, by analogy with the standard likelihood case, a natural **assumption** (that can be proved formally) is that

$$-\dot{G}_n(\theta_0) \xrightarrow{a.s.} V_G(\theta_0)$$

akin to the likelihood result that says the Fisher Information is minus one times the expectation of the log likelihood second derivative matrix. Thus, from equation (5), we have by rearrangement (formally, using Slutsky's Theorem) that

$$\sqrt{n}\, (\widehat{\theta}_n - \theta_0) = (-\dot{G}_n(\theta_0))^{-1}\sqrt{n}\, G_n(\theta_0) \xrightarrow{\mathcal{L}} V_G(\theta_0)^{-1} Z \sim N(0, V_G(\theta_0)^{-1})$$

under the assumption that $(-\dot{G}_n(\theta_0))^{-1}$ exists.

This result follows in the same fashion as in the Cramer's Theorem from lectures.

[6 MARKS]

SAMPLE EXAM QUESTION 6 SOLUTION

(a) *Suppose that $X_1,\ldots,X_n$ are a finitely exchangeable sequence of random variables with (De Finetti) representation*

$$p(X_1,\ldots,X_n) = \int_{-\infty}^{\infty} \prod_{i=1}^{n} f_{X|\theta}(X_i|\theta) p_\theta(\theta) d\theta$$

In the following cases, find the joint probability distribution $p(X_1,\ldots,X_n)$, and give an interpretation of the parameter θ in terms of a strong law limiting quantity.

(i)

$$\begin{aligned} f_{X|\theta}(X_i|\theta) &= Normal(\theta, 1) \\ p_\theta(\theta) &= Normal(0, \tau^2) \end{aligned}$$

for parameter $\tau > 0$.

We have

$$\prod_{i=1}^{n} f_{X|\theta}(X_i|\theta) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}(X_i-\theta)^2\right\} = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2\right\}$$

Now, in the usual decomposition

$$\sum_{i=1}^{n}(X_i-\theta)^2 = n(\bar{X}-\theta)^2 + \sum_{i=1}^{n}(X_i-\bar{X})^2$$

so

$$\begin{aligned} \prod_{i=1}^{n} f_{X|\theta}(X_i|\theta) &= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\left[n(\bar{X}-\theta)^2 + \sum_{i=1}^{n}(X_i-\bar{X})^2\right]\right\} \\ &= K_1(X,n) \exp\left\{-\frac{n}{2}(\bar{X}-\theta)^2\right\} \end{aligned}$$

where

$$K_1(X,n) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\bar{X})^2\right\}.$$

Now

$$p_\theta(\theta) = \frac{1}{(2\pi\tau^2)^{1/2}} \exp\left\{-\frac{1}{2\tau^2}\theta^2\right\}$$

so

$$\prod_{i=1}^{n} f_{X|\theta}(X_i|\theta) p_\theta(\theta) = K_1(X,n) \exp\left\{-\frac{n}{2}(\bar{X}-\theta)^2\right\} \frac{1}{(2\pi\tau^2)^{1/2}} \exp\left\{-\frac{1}{2\tau^2}\theta^2\right\}$$

and combining the terms in the exponents, completing the square, we have

$$n(\bar{X}-\theta)^2 + \theta^2/\tau^2 = (n+1/\tau^2)\left(\theta - \frac{n\bar{X}}{n+1/\tau^2}\right)^2 + \frac{n/\tau^2}{n+1/\tau^2}(\bar{X})^2$$

This uses the general (and useful) completing the square identity

$$A(x-a)^2 + B(x-b)^2 = (A+B)(x-(Aa+Bb)/(A+B))^2 + ((AB)/(A+B))(a-b)^2$$

with $A = n$, $a = \bar{X}$, $B = 1/\tau^2$ and $b = 0$. Thus

$$\prod_{i=1}^{n} f_{X|\theta}(X_i|\theta)p_\theta(\theta) = K_2(X, n, \tau^2) \exp\left\{-\frac{\eta_n}{2}(\theta - \mu_n)^2\right\}$$

where

$$\begin{aligned}
K_2(X, n, \tau^2) &= \frac{K_1(X, n)}{(2\pi\tau^2)^{1/2}} \exp\left\{-\frac{n/\tau^2}{2(n + 1/\tau^2)}(\bar{X})^2\right\} \\
\mu_n &= \frac{n\bar{X}}{n + 1/\tau^2} \\
\eta_n &= (n + 1/\tau^2)
\end{aligned}$$

and thus

$$\begin{aligned}
\int_{-\infty}^{\infty} \prod_{i=1}^{n} f_{X|\theta}(X_i|\theta)p_\theta(\theta)d\theta &= \int_{-\infty}^{\infty} K_2(X, n, \tau^2) \exp\left\{-\frac{\eta_n}{2}(\theta - \mu_n)^2\right\} d\theta \\
&= K_2(X, n, \tau^2)\sqrt{2\pi/\eta_n}
\end{aligned}$$

as the integrand is proportional to a Normal pdf.

The parameter θ in the conditional distribution for the X_i is the expectation. Thus, θ has the interpretation

$$\bar{X} \xrightarrow{a.s.} \theta$$

as $n \to \infty$. To see this more formally, we have the posterior distribution for θ from above as

$$p_{\theta|X}(\theta|X = x) \propto \prod_{i=1}^{n} f_{X|\theta}(X_i|\theta)p_\theta(\theta) \propto \exp\left\{-\frac{\eta_n}{2}(\theta - \mu_n)^2\right\}$$

so $p_{\theta|X}(\theta|X = x) \equiv N(\mu_n, 1/\eta_n)$. As $n \to \infty$,

$$\mu_n = \frac{n\bar{X}}{n + 1/\tau^2} \xrightarrow{a.s.} E[X_i]$$

and $1/\eta_n \to 0$.

(ii)

$$\begin{aligned}
f_{X|\theta}(X_i|\theta) &= \mathit{Exponential}(\theta) \\
p_\theta(\theta) &= \mathit{Gamma}(\alpha, \beta)
\end{aligned}$$

for parameters $\alpha, \beta > 0$.

We have

$$\prod_{i=1}^{n} f_{X|\theta}(X_i|\theta) = \prod_{i=1}^{n} \theta \exp\{-\theta X_i\} = \theta^n \exp\left\{-\theta \sum_{i=1}^{n} X_i\right\}$$

and

$$p_\theta(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp\{-\beta\theta\}$$

so

$$\begin{aligned}\prod_{i=1}^{n} f_{X|\theta}(X_i|\theta) p_\theta(\theta) &= \theta^n \exp\left\{-\theta \sum_{i=1}^{n} X_i\right\} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp\{-\beta\theta\} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{n+\alpha-1} \exp\left\{-\theta \left(\sum_{i=1}^{n} X_i + \beta\right)\right\}\end{aligned}$$

which yields

$$\int_0^\infty \prod_{i=1}^{n} f_{X|\theta}(X_i|\theta) p_\theta(\theta) d\theta = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(n+\alpha)}{\left(\sum_{i=1}^{n} X_i + \beta\right)^{n+\alpha}}$$

as the integrand is proportional to a Gamma pdf.

Now, as

$$p_{\theta|X}(\theta|X=x) \propto \prod_{i=1}^{n} f_{X|\theta}(X_i|\theta) p_\theta(\theta) \propto \theta^{n+\alpha-1} \exp\left\{-\theta \left(\sum_{i=1}^{n} X_i + \beta\right)\right\}$$

so

$$p_{\theta|X}(\theta|X=x) \equiv Ga\left(n+\alpha, \sum_{i=1}^{n} X_i + \beta\right).$$

As $n \to \infty$, this distribution becomes degenerate at

$$\lim_{n\to\infty} \frac{n}{\sum_{i=1}^{n} X_i} = \lim_{n\to\infty} \frac{1}{\bar{X}} = \frac{1}{E[X_i]}$$

so θ is interpreted as the strong law limit of the reciprocal of the expected value of the X_i.

[5 MARKS]

(b) *In each of the two cases of part (a), compute the posterior predictive distribution*

$$p(X_{m+1},\ldots,X_{m+n}|X_1,\ldots,X_m)$$

for $0 < n, m$ *where* $X_1,\ldots,X_{m+n}$ *are a finitely exchangeable sequence.*

Find in each case the limiting posterior predictive distribution as $n \longrightarrow \infty$.

[5 MARKS each]

We compute

$$p(X_{m+1},\ldots,X_{m+n}|X_1,\ldots,X_m) = \int_{-\infty}^{\infty} \prod_{i=m+1}^{m+n} f_{X|\theta}(X_i|\theta) p_{\theta|X^{(1)}}(\theta|X^{(1)} = x^{(1)}) d\theta$$

where $X^{(1)} = (X_1,\ldots,X_m)$.

In the first example;

$$\begin{aligned} p_{\theta|X^{(1)}}(\theta|X^{(1)} = x^{(1)}) &\equiv N(\mu^{(1)}, 1/\eta^{(1)}). \\ \prod_{i=m+1}^{m+n} f_{X|\theta}(X_i|\theta) &= K_1(X^{(2)}, n) \exp\left\{-\frac{n}{2}(\bar{X}^{(2)} - \theta)^2\right\} \end{aligned}$$

where $\mu^{(1)}$ and $\eta^{(1)}$ are as defined earlier, computed for $X^{(1)}$. The posterior predictive is computed in a fashion similar to earlier, completing the square in θ to facilitate the integral; here we have by the previous identity

$$n(\bar{X}^{(2)} - \theta)^2 + \eta^{(1)}(\theta - \mu^{(1)})^2 = (n + \eta^{(1)}) \left(\theta - \frac{n\bar{X}^{(2)} + \eta^{(1)}\mu^{(1)}}{n + \eta^{(1)}}\right)^2 + \frac{n\eta^{(1)}}{n + \eta^{(1)}}(\bar{X}^{(2)} - \mu^{(1)})^2$$

Thus, on integrating out θ, and cancelling terms, we obtain the posterior predictive as

$$K_1(X^{(2)}, n) \exp\left\{-\frac{n\eta^{(1)}}{2(n + \eta^{(1)})}(\bar{X}^{(2)} - \mu^{(1)})^2\right\} \left(\frac{\eta^{(1)}}{n + \eta^{(1)}}\right)^{1/2}$$

In the second example;

$$\begin{aligned} p_{\theta|X^{(1)}}(\theta|X^{(1)} = x^{(1)}) &\equiv Ga\left(m + \alpha, S^{(1)} + \beta\right). \\ \prod_{i=m+1}^{m+n} f_{X|\theta}(X_i|\theta) &= \theta^n \exp\left\{-\theta S^{(2)}\right\} \end{aligned}$$

where

$$S^{(1)} = \sum_{i=1}^{m} X_i \qquad S^{(2)} = \sum_{i=m+1}^{n} X_i$$

Thus

$$\prod_{i=m+1}^{m+n} f_{X|\theta}(X_i|\theta) p_{\theta|X^{(1)}}(\theta|X^{(1)} = x^{(1)}) = \theta^n \exp\left\{-\theta S^{(2)}\right\} \frac{\left(S^{(1)}+\beta\right)^{m+\alpha}}{\Gamma(m+\alpha)} \theta^{m+\alpha-1} \exp\{-\theta(S^{(1)}+\beta)\}$$

$$= \frac{\left(S^{(1)}+\beta\right)^{m+\alpha}}{\Gamma(m+\alpha)} \theta^{n+m+\alpha-1} \exp\{-\theta(S^{(1)}+S^{(2)}+\beta)\}$$

and on integrating out θ, as this form is proportional to a Gamma pdf, we obtain the posterior predictive as

$$\frac{\left(S^{(1)}+\beta\right)^{m+\alpha}}{\Gamma(m+\alpha)} \frac{\Gamma(n+m+\alpha)}{\left(S^{(1)}+S^{(2)}+\beta\right)^{n+m+\alpha}}$$

In both cases, by the general theorem from lecture notes, the limiting posterior predictive when $n \to \infty$ is merely the posterior distribution based on the sample $X_1 \ldots, X_m$.

[5 MARKS each]