IMPERIAL COLLEGE LONDON

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)
MAY–JUNE 2005

M3S3/M4S3 (SOLUTIONS)

STATISTICAL THEORY II

1. (a) (i) Let Ω be a set, and $\mathcal{F}$ be a set of subsets of Ω such that

- $\Omega \in \mathcal{F}$
- $F \in \mathcal{F} \Longrightarrow F' \in \mathcal{F}$ (closed under complementation)
- If $\{F_n\}$ is a countable collection of elements of $\mathcal{F}$, then $\bigcup\limits_n F_n \in \mathcal{F}$ (closed under countable union)

so that $\mathcal{F}$ is a *sigma-algebra*. A non-negative (set) function υ acting on $\mathcal{F}$ is a *measure* if it has the following property: for any countable collection of elements of $\mathcal{F}, \{F_n\}$, we have

$$\upsilon\left(\bigcup_n F_n\right) \leq \sum_n \upsilon\left(F_n\right)$$

with equality if $\{F_n\}$ are disjoint sets. Then

- the pair $(\Omega, \mathcal{F})$ is a *measurable space*
- the triple $(\Omega, \mathcal{F}, \upsilon)$ is a *measure space*

Finally, if $\upsilon(\Omega) = 1$, then υ is a *probability measure*, $(\Omega, \mathcal{F}, \upsilon)$ is a *probability space*.

4 MARKS

(ii) Let $(\Omega, \mathcal{F}, \upsilon)$ denote the measure space. If ψ is a simple function then it takes the following form: for $\omega \in \Omega$

$$\psi(\omega) = \sum_{i=1}^{k} a_i I_{A_i}(\omega)$$

where k is a non-negative integer, $a_1, ..., a_k$ are constants, and $A_1, ..., A_k$ are (measurable) disjoint subsets of Ω, that is, they are elements $\mathcal{F}$.

2 MARKS

The Lebesgue-Stieltjes integral of ψ with respect to υ is denoted and defined by

$$\int \psi d\upsilon = \sum_{i=1}^{k} a_i \upsilon(A_i).$$

2 MARKS

Finally let $\mathcal{S}_f$ denote the set of simple functions defined by

$$\mathcal{S}_f = \{\psi : 0 \leq \psi(\omega) \leq f(\omega) \text{ for all } \omega \in \Omega\}.$$

Then

$$\int f d\upsilon = \sup_{\psi \in \mathcal{S}_f} \int \psi d\upsilon$$

2 MARKS

SEEN

(b) The Wald theorem proves the strong consistency of the MLE, whereas the Cramer theorem proves the asymptotic normality of the MLE (or indeed any sequence of consistent solutions to the likelihood equation), that is, if θ_0 is the true value of the parameter θ in the probability model $f_X(x;\theta)$, then

$$\widetilde{\theta}_n \overset{a.s.}{\rightarrow} \theta_0 \qquad \text{gives} \qquad \sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right) \overset{\mathcal{L}}{\rightarrow} Z \sim N\left(0, \left[I(\theta_0)\right]^{-1}\right)$$

For the Wald Theorem, regularity conditions (for the cases seen by the students) include the compactness of the parameter space Θ, the (upper-semi) continuity of the density in θ for all x, the boundedness of the function

$$U(x,\theta) = \log f_X(x;\theta) - \log f_X(x;\theta_0)$$

the uniform measurability of the density with respect to x on an open neighbourhood of any $\theta \in \Theta$, and the identifiability of the density with respect to θ. For the Cramer theorem, we need the Θ to be an open subset of $\mathbb{R}$, existence and boundedness of second partial derivatives (third derivatives for weakly consistent solutions), the positive-definiteness of the expectation of the matrix Ψ of second partial derivatives, and identifiability.

4 MARKS

(c) (i) We have

$$L(\theta,\eta) = \eta^2\theta\exp\{-[\eta x + \theta\eta y]\} \qquad x, y > 0$$

so that

$$l(\theta,\eta) = \log L(\theta,\eta) = 2\log\eta + \log\theta - (\eta x + \theta\eta y)$$

and

$$\frac{\partial l}{\partial\eta} = \frac{2}{\eta} - x - \theta y \qquad\qquad \frac{\partial l}{\partial\theta} = \frac{1}{\theta} - \eta y$$

$$\frac{\partial^2 l}{\partial\eta^2} = -\frac{2}{\eta^2} \qquad\qquad \frac{\partial^2 l}{\partial\theta^2} = -\frac{1}{\theta^2}$$

$$\frac{\partial^2 l}{\partial\eta\partial\theta} = -y$$

yielding the observed and Fisher information

$$\mathcal{I}(\theta,\eta) = \begin{bmatrix} \dfrac{2}{\eta^2} & y \\ y & \dfrac{1}{\theta^2} \end{bmatrix} \qquad I(\theta,\eta) = \begin{bmatrix} \dfrac{2}{\eta^2} & \dfrac{1}{\eta\theta} \\ \dfrac{1}{\eta\theta} & \dfrac{1}{\theta^2} \end{bmatrix}$$

as $E[Y] = 1/(\eta\theta)$.

4 MARKS

(ii) The parameters are not orthogonal as the off-diagonal element of $I(\theta,\eta)$ is non-zero.

2 MARKS

SEEN TECHNIQUE

2. **(a)** **(i)** $\{X_n\}$ converges almost surely to a limiting random variable X if

$$P\left[\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right] = 1$$

that is, the set of ω for which $X_n(\omega) \to X(\omega)$ has P-measure one. Equivalently,

$$X_n \overset{a.s.}{\to} X \qquad \Longleftrightarrow \qquad P\left[\lim_{n\to\infty} |X_n - X| < \varepsilon\right] = 1$$

for all $\varepsilon > 0$.

2 MARKS

SEEN

(ii) **THEOREM**

Let $\{A_k\}$ be a sequence of events in sample space Ω. If

$$A^{(S)} = \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} A_j$$

is the limsup event of the infinite sequence; $A^{(S)}$ occurs if any only if infinitely many of the $A_j s$ occur, or the $A_j s$ occur infinitely often (i.o.)

(I) If $\sum_{k=1}^{\infty} P(A_k) < \infty$, then $P\left(A^{(S)}\right) = P\left(A_j \text{ occurs i.o.}\right) = 0$,

(II) If the events $\{A_k\}$ are independent, and $\sum_{k=1}^{\infty} P(A_k) = \infty$, then $P\left(A^{(S)}\right) = 1$.

PROOF (I) Note first that

$$\sum_{k=1}^{\infty} P(A_k) < \infty \Longrightarrow \lim_{k\to\infty} \sum_{j=k}^{\infty} P(A_j) = 0.$$

because if the sum on the left-hand side is bounded above, then the sum on the right-hand side tends to zero as $k \to \infty$. Now, for every $k \geq 1$,

$$A^{(S)} = \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} A_j \subseteq \bigcup_{j=k}^{\infty} A_j$$

and therefore, as $k \longrightarrow \infty$

$$P\left(A^{(S)}\right) \leq P\left(\bigcup_{j=k}^{\infty} A_j\right) \leq \sum_{j=k}^{\infty} P(A_j) \to 0$$

(II) Consider $K \geq k$, and the union of events

$$A = \bigcup_{j=k}^{K} A_j.$$

Then

$$A' = \bigcap_{j=k}^{K} A_j' \subseteq \bigcup_{j=k}^{\infty} A_j'$$

Now

$$P(A) = P\left(\bigcup_{j=k}^{K} A_j\right) \leq P\left(\bigcup_{j=k}^{\infty} A_j\right).$$

Therefore

$$
\begin{aligned}
1-P\left(\bigcup_{j=k}^{\infty}A_j\right) &\leq 1-P\left(\bigcup_{j=k}^{K}A_j\right)=1-P(A)=P\left(A'\right)=P\left(\bigcap_{j=k}^{K}A'_j\right)\\
&= \prod_{j=k}^{K}P\left(A'_j\right) \qquad \text{by independence}\\
&= \prod_{j=k}^{K}\left(1-P\left(A_j\right)\right)\leq \exp\left\{-\sum_{j=k}^{K}P\left(A_j\right)\right\}
\end{aligned}
$$

as $1-x\leq\exp\{-x\}$ for $0<x<1$. Now, taking the limit of both sides as $K\to\infty$, for fixed k,

$$
1-P\left(\bigcup_{j=k}^{\infty}A_j\right)\leq\lim_{K\to\infty}\exp\left\{-\sum_{j=k}^{K}P\left(A_j\right)\right\}=\exp\left\{-\sum_{j=k}^{\infty}P\left(A_j\right)\right\}=0
$$

as, by assumption $\sum\limits_{k=1}^{\infty}P(A_k)=\infty$. Thus, for each k, we have that

$$
P\left(\bigcup_{j=k}^{\infty}A_j\right)=1 \qquad \therefore \qquad \lim_{k\to\infty}P\left(\bigcup_{j=k}^{\infty}A_j\right)=1.
$$

By continuity of probability measure

$$
\lim_{k\to\infty}P\left(A_k\right)=P\left(\lim_{k\to\infty}A_k\right)=P\left(\bigcap_{k=1}^{\infty}A_k\right)=P\left(\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty}A_j\right)=P\left(A^{(S)}\right)
$$

Hence$P\left(A^{(S)}\right)=1$.

This result is related to almost sure convergence; if we let

$$
A_j\left(\varepsilon\right)\equiv\left\{\omega:\left|X_j\left(\omega\right)-X\left(\omega\right)\right|<\varepsilon\right\}
$$

then if A_j occurs i.o. we have a.s. convergence of $\{X_n\}$ to X.

12 MARKS

SEEN

(b) (i) Let A_n be the event $(X_n\neq 0)$. Then $P(A_n)=1/n$, and hence

$$
\sum_{n=1}^{\infty}P(A_n)=\infty.
$$

The events $A_1,A_2,\ldots$ are independent, so by the BC Lemma part (II),

$$
P(A_n \text{ occurs i.o})=1,
$$

so X_n does not converge a.s. to 0. X_n only takes values in $\{0,1\}$, and $P[X_n=0]>0$ for any finite n, so X_n does not converge to 1 a.s. either. Hence X_n does not converge a.s. to any real value.

3 MARKS

(ii) We have

$$
E\left[\left|X_n\right|\right]=E\left[I_{[0,n^{-1})}\left(U_n\right)\right]=P\left[U_n\leq n^{-1}\right]=\frac{1}{n}
$$

so

$$
X_n\overset{r=1}{\to}X_B
$$

where $P\left[X_B=0\right]=1$, and we have convergence in r^{th} mean to zero for $r=1$.

3 MARKS

UNSEEN

3. (a) **THEOREM** Let $F_n(x)$ denote the empirical distribution function (edf) derived from an i.i.d. sample $X_1, ..., X_n$ from a distribution with cdf F_X, that is,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[X_i,\infty)}(x) \qquad x \in \mathbb{R}.$$

Then the edf converges almost surely to the true cdf, uniformly in x, that is

$$P\left[\sup_x |F_n(x) - F_X(x)| \to 0\right] = 1.$$

PROOF. First note that

$$F_n(x) \stackrel{a.s.}{\to} F_X(x)$$

pointwise for $x \in \mathbb{R}$, by the Strong Law of Large numbers, by definition of F_n as the sample mean of a collection of iid (indicator) random variables. Now let $\varepsilon > 0$ be specified, and choose $k > 1/\varepsilon$, and numbers

$$-\infty = x_0 < x_1 \leq x_2 \leq ... \leq x_{k-1} < x_k = \infty$$

such that

$$P[X < x_j] = F_X\left(x_j^-\right) \leq \frac{j}{k} \leq F_X(x_j) = P[X \leq x_j]$$

for $j = 1, 2, ..., k-1$. Note that if $x_{j-1} < x_j$ then $F_X\left(x_j^-\right) - F_X(x_{j-1}) \leq \frac{1}{k} < \varepsilon$. By the Strong Law, as $n \to \infty$,

$$F_n(x_j) \stackrel{a.s.}{\to} F_X(x_j) \qquad \text{and} \qquad F_n\left(x_j^-\right) \stackrel{a.s.}{\to} F_X\left(x_j^-\right)$$

for each j. Thus, also by the Strong Law, as $n \to \infty$,

$$\triangle_n = \max_j \left\{|F_n(x_j) - F_X(x_j)|, \left|F_n(x_j^-) - F_X(x_j^-)\right|\right\} \stackrel{a.s.}{\to} 0. \qquad \text{(A3.1)}$$

Let $x \in \mathbb{R}$, and find j such that $x_{j-1} \leq x < x_j$. Then, as

$$x < x_j \Longrightarrow F_n(x) \leq F_n\left(x_j^-\right) \qquad \text{and} \qquad F_X(x) \leq F_X\left(x_j^-\right),$$

by definition of the regular grid defined by the x_js,

$$\begin{aligned} F_n(x) - F_X(x) &\leq F_n(x_j^-) - F_X(x_{j-1}) \\ &\leq F_n(x_j^-) - F_X\left(x_j^-\right) + \varepsilon \end{aligned}$$

and also

$$\begin{aligned} F_n(x) - F_X(x) &\geq F_n(x_{j-1}) - F_X\left(x_j^-\right) \\ &\geq F_n(x_{j-1}) - F_X(x_{j-1}) - \varepsilon. \end{aligned}$$

Hence, for any such x,

$$|F_n(x) - F_X(x)| \leq \triangle_n + \varepsilon$$

and the RHS converges almost surely to ε, by (A3.1). This result holds uniformly in x, so we have

$$\sup_x |F_n(x) - F_X(x)| \stackrel{a.s.}{\to} \varepsilon$$

and hence the result follows, as the choice of $\varepsilon > 0$ is arbitrary.

12 MARKS

SEEN

(b) For any p

$$p = \frac{e^{x_p}}{1+e^{x_p}} \Longrightarrow x_p = \log\left(\frac{p}{1-p}\right)$$

and, here,

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{e^x}{(1+e^x)^2}$$

so

$$f_X(x_p) = \frac{p/(1-p)}{(1+p/(1-p))^2} = p(1-p)$$

Now, from the Central Limit Theorem result for the sample quantiles, as $n \to \infty$,

$$\sqrt{n}\left(\begin{pmatrix} X_{(k_1)} \\ X_{(k_2)} \end{pmatrix} - \begin{pmatrix} x_{p_1} \\ x_{p_2} \end{pmatrix}\right) \to Z \sim N(0, \Sigma)$$

where

$$\begin{aligned}
\Sigma &= \begin{bmatrix} \dfrac{p_1(1-p_1)}{f_X(x_{p_1})^2} & \dfrac{p_1(1-p_2)}{f_X(x_{p_1})f_X(x_{p_2})} \\[2ex] \dfrac{p_1(1-p_2)}{f_X(x_{p_1})f_X(x_{p_2})} & \dfrac{p_2(1-p_2)}{f_X(x_{p_2})^2} \end{bmatrix} \\[3ex]
&= \begin{bmatrix} \dfrac{p_1(1-p_1)}{(p_1(1-p_1))^2} & \dfrac{p_1(1-p_2)}{p_1(1-p_1)p_2(1-p_2)} \\[2ex] \dfrac{p_1(1-p_2)}{p_1(1-p_1)p_2(1-p_2)} & \dfrac{p_2(1-p_2)}{(p_2(1-p_2))^2} \end{bmatrix} \\[3ex]
&= \begin{bmatrix} \dfrac{1}{p_1(1-p_1)} & \dfrac{1}{(1-p_1)p_2} \\[2ex] \dfrac{1}{(1-p_1)p_2} & \dfrac{1}{p_2(1-p_2)} \end{bmatrix}
\end{aligned}$$

Hence

$$\begin{pmatrix} X_{(k_1)} \\ X_{(k_2)} \end{pmatrix} \sim AN\left(\begin{pmatrix} x_{p_1} \\ x_{p_2} \end{pmatrix}, \frac{1}{n}\Sigma\right).$$

8 MARKS

SEEN TECHNIQUE

4. **(a)** The Kullback-Liebler (KL) divergence between two probability measures that have densities f_0 and f_1 with respect to measure υ is defined as

$$K(f_0, f_1) = \int f_0(x) \log \frac{f_0(x)}{f_1(x)} d\upsilon(x) = E_{f_0}\left[\log \frac{f_0(x)}{f_1(x)}\right]$$

2 MARKS

(b) Using Jensen's Inequality on the convex function $-\log x$

$$\begin{aligned} -K(f_0, f_1) &= E_{f_0}\left[-\log \frac{f_0(x)}{f_1(x)}\right] = E_{f_0}\left[\log \frac{f_1(x)}{f_0(x)}\right] \\ &\leq \log E_{f_0}\left[\frac{f_1(x)}{f_0(x)}\right] = \log\left\{\int \frac{f_1(x)}{f_0(x)} f_0(x) d\upsilon(x)\right\}. \\ &\leq \log\left\{\int_{S_0} f_1(x) d\upsilon(x)\right\} \leq \log 1 = 0 \end{aligned}$$

where S_0 is the support of f_0, with equality if $\int_{S_0} f_1(x)d\upsilon(x) = 1$. Hence $K(f_0, f_1) \geq 0$.

6 MARKS

SEEN

(c) We have, for $\theta \in \Theta$

$$T_n = \frac{1}{n} \log \frac{L_n(\theta_0)}{L_n(\theta)} = \frac{1}{n} \sum_{i=1}^{n} \log \frac{f_X(X_i; \theta_0)}{f_X(X_i; \theta)}$$

and thus by the Strong Law of Large numbers

$$T_n \overset{a.s}{\to} E_{f_0}\left[\log \frac{f_X(X_i; \theta_0)}{f_X(X_i; \theta)}\right] = K(f_{\theta_0}, f_\theta)$$

and by the previous result $K(f_{\theta_0}, f_\theta) = 0 \Longleftrightarrow \theta = \theta_0$

6 MARKS

SEEN TECHNIQUE

(d) **(i)**

$$\begin{aligned} K(f_0, f_1) &= \int_0^\infty f_0(x) \log \frac{f_0(x)}{f_1(x)} dx = \int_0^\infty \left\{\lambda_0 e^{-\lambda_0 x} \times \left[\log \frac{\lambda_0}{\lambda_1} + (\lambda_1 - \lambda_0) x\right]\right\} dx \\ &= \log \frac{\lambda_0}{\lambda_1} + (\lambda_1 - \lambda_0) E_{f_0}[X] = \log \frac{\lambda_0}{\lambda_1} + \frac{(\lambda_1 - \lambda_0)}{\lambda_0} \end{aligned}$$

2 MARKS

(ii)

$$\begin{aligned} K(f_0, f_1) &= \int_0^\infty f_0(x) \log \frac{f_0(x)}{f_1(x)} dx \\ &= \int_0^\infty \left\{\frac{1}{\Gamma(\alpha_0)} x^{\alpha_0 - 1} e^{-x} \times \left[\log \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_0)} + (\alpha_1 - \alpha_0) \log x\right]\right\} dx \\ &= \log \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_0)} + (\alpha_1 - \alpha_0) E_{f_0}[\log X] \\ &= \log \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_0)} + (\alpha_1 - \alpha_0) \, \mathrm{Di}\Gamma(\alpha_0) \end{aligned}$$

6 MARKS

UNSEEN

5. (a) **THEOREM (**Following the notation and proof of Bernardo and Smith (1994))

If $X_1, X_2, ...$ is an infinitely exchangeable sequence of 0-1 variables with probability measure P, then there exists a distribution function Q such that the joint mass function of $(X_1, X_2, ..., X_n)$ has the form

$$p(X_1, X_2, ..., X_n) = \int_0^1 \left\{ \prod_{i=1}^n \theta^{X_i} (1-\theta)^{1-X_i} \right\} dQ(\theta)$$

where

$$Q(\theta) = \lim_{n\to\infty} P\left[\frac{Y_n}{n} \leq \theta\right]$$

and $Y_n = \sum_{i=1}^n X_i$, and $\theta = \lim_{n\to\infty} Y_n/n$ is the (strong-law) limiting relative frequency of 1s.

PROOF By exchangeability, for $0 \leq y_n \leq n$

$$P[Y_n = y_n] = \binom{n}{y_n} p(x_1, x_2, ..., x_n) = \binom{n}{y_n} p\left(x_{\pi(1)}, x_{\pi(2)}, ..., x_{\pi(n)}\right) \tag{A5.0}$$

where $X_i = x_i$ and $y_n = \sum_{i=1}^n x_i$, and $\pi()$ is any permutation of the indices. For finite N, let $N \geq n \geq y_n \geq 0$. Then, by exchangeability

$$P[Y_n = y_n] = \sum P[Y_n = y_n | Y_N = y_N] P[Y_N = y_N] \tag{A5.1}$$

where the summation extends over $(y_n, ..., N-(n-y_n))$. Now the conditional probability $P[Y_n = y_n|Y_N = y_N]$ is a hypergeometric mass function

$$P[Y_n = y_n|Y_N = y_N] = \frac{\binom{y_N}{y_n}\binom{N-y_N}{n-y_n}}{\binom{N}{n}} \qquad 0 \leq y_n \leq n.$$

Rewriting the binomial coefficients, we have

$$P[Y_n = y_n] = \binom{n}{y_n} \sum \frac{(y_N)_{y_n} (N-y_N)_{n-y_n}}{(N)_n} P[Y_N = y_N] \tag{A5.2}$$

where $(x)_r = x(x-1)(x-2)...(x-r+1)$.

Define function $Q_N(\theta)$ on $\mathbb{R}$ as the step function which is zero for $\theta < 0$, and has steps of size $P[Y_N = y_N]$ at $\theta = y_N/N$ for $y_N = 0, 1, 2, ..., N$. Hence, utilizing the Lebesgue-Stieltjes notation, we can re-write

$$P[Y_n = y_n] = \binom{n}{y_n} \int_0^1 \frac{(\theta N)_{y_n} ((1-\theta)N)_{n-y_n}}{(N)_n} dQ_N(\theta). \tag{A5.3}$$

This result holds for any finite N, but in (A5.1) we need to consider $N \to \infty$. In the limit,

$$\frac{(\theta N)_{y_n} ((1-\theta)N)_{n-y_n}}{(N)_n} \to \theta^{y_n}(1-\theta)^{n-y_n} = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}$$

as $(x)_r \to x^r$ if $x \to \infty$ with r fixed. Also, by the Helly Theorem $\{Q_N(\theta)\}$ has a convergent subsequence $\{Q_{N_j}(\theta)\}$ such that, for a distribution function Q,

$$\lim_{j\to\infty} Q_{N_j}(\theta) = Q(\theta)$$

Thus the result follows comparing (A5.0) and the limiting form of (A5.3) the result follows.

12 MARKS

SEEN

(b) For $1 \leq m \leq n$

$$
\begin{aligned}
p\left(X_{m+1}, X_2, ..., X_n | X_1, X_2, ..., X_m\right) &= \frac{p\left(X_1, X_2, ..., X_n\right)}{p\left(X_1, X_2, ..., X_m\right)} \qquad \text{(A5.4)} \\
&= \int_0^1 \left\{ \prod_{i=m+1}^{n} \theta^{X_i} \left(1-\theta\right)^{1-X_i} \right\} dQ\left(\theta | X_1, ..., X_m\right)
\end{aligned}
$$

where, if

$$Q\left(\theta\right) = \int_0^\theta dQ\left(t\right)$$

we have

$$dQ\left(\theta | X_1, ..., X_m\right) = \frac{\prod_{i=1}^{m} \theta^{X_i} \left(1-\theta\right)^{1-X_i} dQ\left(\theta\right)}{\int_0^1 \prod_{i=1}^{m} \theta^{X_i} \left(1-\theta\right)^{1-X_i} dQ\left(\theta\right)}$$

as the updated "prior" measure. Hence, if $Y_{n-m} = \sum_{i=m+1}^{n} X_i$, we have from (A5.4)

$$p\left(Y_{n-m} | X_1, ..., X_m\right) = \int_0^1 \binom{n-m}{y_{n-m}} \theta^{Y_{n-m}} \left(1-\theta\right)^{(n-m)-Y_{n-m}} dQ\left(\theta | X_1, ..., X_m\right)$$

which identifies $Q\left(\theta | X_1, ..., X_m\right)$ as the *limiting posterior predictive distribution*, as from (A5.4) and the representation theorem itself

$$\lim_{n \to \infty} \left[\frac{Y_{n-m}}{n-m}\right] = Q\left(\theta | X_1, ..., X_m\right)$$

8 MARKS

SEEN