1. (a) (i) Define the terms *measurable space*, *measure space* and *probability space,* and any terms you use in the definitions.

(ii) Suppose ψ is a *simple function* defined on an arbitrary measure space. Give the mathematical form of ψ, give an expression for its *Lebesgue-Stieltjes integral*, and give the *supremum definition* for the Lebesgue-Stieltjes integral of a non-negative Borel function.

(b) Explain the relevance of the *Wald* and *Cramér* theorems to the asymptotic behaviour of maximum likelihood estimators. Give brief details of the regularity conditions under which these theorems operate.

(c) Suppose that X and Y are independent *Exponential* random variables with (rate) parameters η and $\theta\eta$ respectively, so that the likelihood function is

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

for parameters $\theta, \eta > 0$.

(i) Find the *Fisher Information* for (θ, η), $I(\theta, \eta)$, derived from this likelihood.

(ii) Are θ and η *orthogonal* parameters ? Justify your answer.

2. (a) (i) Give the definition for *almost sure* convergence of a sequence of random variables $\{X_n\}$ to a limiting random variable X.

(ii) State and prove the Borel-Cantelli Lemma. Explain the connection between this result and the concept of almost sure convergence.

(b) Consider the sequence of random variables defined for $n = 1, 2, 3, ...$ by

$$X_n = I_{[0,n^{-1})}(U_n)$$

where $U_1, U_2, ...$ are a sequence of independent *Uniform*$(0, 1)$ random variables, and I_A is the indicator function for set A

$$I_A(\omega) = \begin{cases} 1 & \omega \in A; \\ 0 & \omega \notin A. \end{cases}$$

Does the sequence $\{X_n\}$ converge

(i) almost surely ?

(ii) in r^{th} mean for $r = 1$?

Justify your answers.

[*Hint: Consider the events* $A_n \equiv (X_n \neq 0)$ *for* $n = 1, 2, \ldots$]

3. (a) State and prove the Glivenko-Cantelli Theorem on the uniform convergence of the empirical distribution function.

(b) Let $X_{(1)} < X_{(2)} < ... < X_{(n)}$ denote the order statistics derived from a random sample of size n from the log-logistic distribution which has distribution function

$$F_X(x) = \frac{e^x}{1+e^x} \qquad x \in \mathbb{R}.$$

Let $0 < p_1 < p_2 < 1$ be two probabilities with corresponding quantiles x_{p_1} and x_{p_2}. Let $k_1 = \lceil np_1 \rceil$ and $k_2 = \lceil np_2 \rceil$, so that $X_{(k_1)}$ and $X_{(k_2)}$ are the sample quantiles that act as natural estimators of x_{p_1} and x_{p_2}.

Find an asymptotic normal approximation (for large n) to the joint distribution of

$$\begin{pmatrix} X_{(k_1)} \\ X_{(k_2)} \end{pmatrix}.$$

4. (a) Define the Kullback-Liebler (KL) divergence between two probability measures that have densities f_0 and f_1 with respect to measure υ.

(b) Show that the KL divergence is a non-negative quantity.

(c) Let $L_n(\theta)$ denote the likelihood for independent and identically distributed random variables $X_1, ..., X_n$ having probability density function $f_X(\cdot\,;\theta)$, with common support $\mathbb{X}$ that does not depend on θ, for $\theta \in \Theta$. Let θ_0 denote the true value of θ, and suppose that θ is identifiable, that is,

$$f_X(x;\theta_1) = f_X(x;\theta_2)\,, \text{ for all } x \in \mathbb{X} \Longrightarrow \theta_1 = \theta_2.$$

Prove that the random variable

$$\frac{1}{n}\log\frac{L_n(\theta_0)}{L_n(\theta)}$$

converges almost surely to zero if and only if $\theta = \theta_0$.

(d) Evaluate the KL divergence $K(f_0, f_1)$ (with respect to Lebesgue measure) between two *Exponential* densities with rate parameters λ_0 and λ_1

$$f_0(x) = \lambda_0 \exp\{-\lambda_0 x\} \qquad f_1(x) = \lambda_1 \exp\{-\lambda_1 x\}$$

for $x > 0$, and zero otherwise.

5. (a) State and prove the Bayesian representation theorem (of De Finetti) for an exchangeable sequence of 0-1 random variables.

(You may quote without proof the Helly Theorem on the existence of a convergent sequence of distribution functions.)

(b) Suppose that $X_1, ..., X_m, X_{m+1}, ..., X_n$ are a (finitely) exchangeable collection of 0-1 random variables. Give an expression for the (posterior) predictive distribution

$$p(X_{m+1}, ..., X_n | X_1, ..., X_m)$$

explaining carefully any notation that you use.

Discuss the limiting behaviour of the posterior predictive as $n \to \infty$ with m fixed.