A1/12 B1/15 **Principles of Statistics**

(i) A public health official is seeking a rational policy of vaccination against a relatively mild ailment which causes absence from work. Surveys suggest that 60% of the population are already immune, but accurate tests to detect vulnerability in any individual are too costly for mass screening. A simple skin test has been developed, but is not completely reliable. A person who is immune to the ailment will have a negligible reaction to the skin test with probability 0.4, a moderate reaction with probability 0.5 and a strong reaction with probability 0.1. For a person who is vulnerable to the ailment the corresponding probabilities are 0.1, 0.4 and 0.5. It is estimated that the money-equivalent of work-hours lost from failing to vaccinate a vulnerable person is 20, that the unnecessary cost of vaccinating an immune person is 8, and that there is no cost associated with vaccinating a vulnerable person or failing to vaccinate an immune person. On the basis of the skin test, it must be decided whether to vaccinate or not. What is the Bayes decision rule that the health official should adopt?

(ii) A collection of I students each sit J exams. The ability of the ith student is represented by θ_i and the performance of the ith student on the jth exam is measured by X_{ij}. Assume that, given $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)$, an appropriate model is that the variables $\{X_{ij}\,,\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J\}$ are independent, and

$$X_{ij} \sim N(\theta_i, \tau^{-1}),$$

for a known positive constant τ. It is reasonable to assume, *a priori*, that the θ_i are independent with

$$\theta_i \sim N(\mu, \zeta^{-1}),$$

where μ and ζ are population parameters, known from experience with previous cohorts of students.

Compute the posterior distribution of $\boldsymbol{\theta}$ given the observed exam marks vector $\mathbf{X} = \{X_{ij}\,,\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J\}$.

Suppose now that τ is also unknown, but assumed to have a Gamma(α_0, β_0) distribution, for known α_0, β_0. Compute the posterior distribution of τ given $\boldsymbol{\theta}$ and $\mathbf{X}$. Find, up to a normalisation constant, the form of the marginal density of $\boldsymbol{\theta}$ given $\mathbf{X}$.

A2/11 B2/16 **Principles of Statistics**

(i) Outline briefly the Bayesian approach to hypothesis testing based on Bayes factors.

(ii) Let Y_1, Y_2 be independent random variables, both uniformly distributed on $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Find a minimal sufficient statistic for θ. Let $Y_{(1)} = \min\{Y_1, Y_2\}$, $Y_{(2)} = \max\{Y_1, Y_2\}$. Show that $R = Y_{(2)} - Y_{(1)}$ is ancilliary and explain why the Conditionality Principle would lead to inference about θ being drawn from the conditional distribution of $\frac{1}{2}\{Y_{(1)} + Y_{(2)}\}$ given R. Find the form of this conditional distribution.

A3/12 B3/15 **Principles of Statistics**

(i) Let $X_1, \ldots, X_n$ be independent, identically distributed random variables, with the exponential density $f(x;\theta) = \theta e^{-\theta x}$, $x > 0$.

Obtain the maximum likelihood estimator $\hat{\theta}$ of θ. What is the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$?

What is the minimum variance unbiased estimator of θ? Justify your answer carefully.

(ii) Explain briefly what is meant by the *profile log-likelihood* for a scalar parameter of interest γ, in the presence of a nuisance parameter ξ. Describe how you would test a null hypothesis of the form $H_0 : \gamma = \gamma_0$ using the profile log-likelihood ratio statistic.

In a reliability study, lifetimes $T_1, \ldots, T_n$ are independent and exponentially distributed, with means of the form $E(T_i) = \exp(\beta + \xi z_i)$ where β, ξ are unknown and $z_1, \ldots, z_n$ are known constants. Inference is required for the mean lifetime, $\exp(\beta + \xi z_0)$, for covariate value z_0.

Find, as explicitly as possible, the profile log-likelihood for $\gamma \equiv \beta + \xi z_0$, with nuisance parameter ξ.

Show that, under $H_0 : \gamma = \gamma_0$, the profile log-likelihood ratio statistic has a distribution which does not depend on the value of ξ. How might the parametric bootstrap be used to obtain a test of H_0 of exact size α?

[*Hint: if* Y *is exponentially distributed with mean* 1, *then* μY *is exponentially distributed with mean* μ.]

A4/13 B4/15 **Principles of Statistics**

Write an account, with appropriate examples, of inference in multiparameter exponential families. Your account should include a discussion of natural statistics and their properties and of various conditional tests on natural parameters.