Paper 1, Section I

7H Statistics

$X_1, X_2, \ldots, X_n$ form a random sample from a distribution whose probability density function is

$$f(x;\theta) = \begin{cases} \dfrac{2x}{\theta^2} & 0 \leqslant x \leqslant \theta \\ 0 & \text{otherwise,} \end{cases}$$

where the value of the positive parameter θ is unknown. Determine the maximum likelihood estimator of the median of this distribution.

Paper 2, Section I

8H Statistics

Define a *simple hypothesis*. Define the terms *size* and *power* for a test of one simple hypothesis against another. State the Neyman-Pearson lemma.

There is a single observation of a random variable X which has a probability density function $f(x)$. Construct a best test of size 0.05 for the null hypothesis

$$H_0 : \quad f(x) = \tfrac{1}{2}, \qquad\qquad -1 \leqslant x \leqslant 1,$$

against the alternative hypothesis

$$H_1 : \quad f(x) = \frac{3}{4}(1 - x^2), \quad -1 \leqslant x \leqslant 1.$$

Calculate the power of your test.

Paper 1, Section II
19H Statistics

(a) Consider the general linear model $Y = X\theta + \varepsilon$ where X is a known $n \times p$ matrix, θ is an unknown $p \times 1$ vector of parameters, and ε is an $n \times 1$ vector of independent $N(0, \sigma^2)$ random variables with unknown variances σ^2. Show that, provided the matrix X is of rank p, the least squares estimate of θ is

$$\hat{\theta} = (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} Y.$$

Let

$$\hat{\varepsilon} = Y - X\hat{\theta}.$$

What is the distribution of $\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}$? Write down, in terms of $\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}$, an unbiased estimator of σ^2.

(b) Four points on the ground form the vertices of a plane quadrilateral with interior angles $\theta_1, \theta_2, \theta_3, \theta_4$, so that $\theta_1 + \theta_2 + \theta_3 + \theta_4 = 2\pi$. Aerial observations Z_1, Z_2, Z_3, Z_4 are made of these angles, where the observations are subject to independent errors distributed as $N(0, \sigma^2)$ random variables.

(i) Represent the preceding model as a general linear model with observations $(Z_1, Z_2, Z_3, Z_4 - 2\pi)$ and unknown parameters $(\theta_1, \theta_2, \theta_3)$.

(ii) Find the least squares estimates $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$.

(iii) Determine an unbiased estimator of σ^2. What is its distribution?

Paper 4, Section II
19H Statistics

There is widespread agreement amongst the managers of the Reliable Motor Company that the number X of faulty cars produced in a month has a binomial distribution

$$P(X = s) = \binom{n}{s} p^s (1-p)^{n-s} \quad (s = 0, 1, \ldots, n; \quad 0 \leqslant p \leqslant 1),$$

where n is the total number of cars produced in a month. There is, however, some dispute about the parameter p. The general manager has a prior distribution for p which is uniform, while the more pessimistic production manager has a prior distribution with density $2p$, both on the interval $[0, 1]$.

In a particular month, s faulty cars are produced. Show that if the general manager's loss function is $(\hat{p} - p)^2$, where $\hat{p}$ is her estimate and p the true value, then her best estimate of p is

$$\hat{p} = \frac{s+1}{n+2}.$$

The production manager has responsibilities different from those of the general manager, and a different loss function given by $(1-p)(\hat{p} - p)^2$. Find his best estimate of p and show that it is greater than that of the general manager unless $s \geqslant \frac{1}{2} n$.

[*You may use the fact that for non-negative integers* α, β,

$$\int_0^1 p^\alpha (1-p)^\beta \, dp = \frac{\alpha! \beta!}{(\alpha + \beta + 1)!}. \quad]$$

Paper 3, Section II

20H Statistics

A treatment is suggested for a particular illness. The results of treating a number of patients chosen at random from those in a hospital suffering from the illness are shown in the following table, in which the entries a, b, c, d are numbers of patients.

	Recovery	Non-recovery
Untreated	a	b
Treated	c	d

Describe the use of Pearson's χ^2 statistic in testing whether the treatment affects recovery, and outline a justification derived from the generalised likelihood ratio statistic. Show that

$$\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}.$$

[*Hint: You may find it helpful to observe that* $a(a+b+c+d)-(a+b)(a+c) = ad-bc$.]

Comment on the use of this statistical technique when

$$a = 50, \quad b = 10, \quad c = 15, \quad d = 5.$$