Paper 1, Section I

7H Statistics

Consider an estimator $\hat{\theta}$ of an unknown parameter θ, and assume that $\mathbb{E}_\theta\big(\hat{\theta}^2\big) < \infty$ for all θ. Define the *bias* and *mean squared error* of $\hat{\theta}$.

Show that the mean squared error of $\hat{\theta}$ is the sum of its variance and the square of its bias.

Suppose that $X_1, \ldots, X_n$ are independent identically distributed random variables with mean θ and variance θ^2, and consider estimators of θ of the form $k\bar{X}$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

(i) Find the value of k that gives an unbiased estimator, and show that the mean squared error of this unbiased estimator is θ^2/n.

(ii) Find the range of values of k for which the mean squared error of $k\bar{X}$ is smaller than θ^2/n.

Paper 2, Section I

8H Statistics

There are 100 patients taking part in a trial of a new surgical procedure for a particular medical condition. Of these, 50 patients are randomly selected to receive the new procedure and the remaining 50 receive the old procedure. Six months later, a doctor assesses whether or not each patient has fully recovered. The results are shown below:

	Fully recovered	Not fully recovered
Old procedure	25	25
New procedure	31	19

The doctor is interested in whether there is a difference in full recovery rates for patients receiving the two procedures. Carry out an appropriate 5% significance level test, stating your hypotheses carefully. [You do not need to derive the test.] What conclusion should be reported to the doctor?

[*Hint: Let* $\chi_k^2(\alpha)$ *denote the upper* 100α *percentage point of a* χ_k^2 *distribution. Then*

$$\chi_1^2(0.05) = 3.84,\ \chi_2^2(0.05) = 5.99,\ \chi_3^2(0.05) = 7.82,\ \chi_4^2(0.05) = 9.49.]$$

Paper 4, Section II

19H Statistics

Consider a linear model

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (\dagger)$$

where X is a known $n \times p$ matrix, $\boldsymbol{\beta}$ is a $p \times 1$ $(p < n)$ vector of unknown parameters and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of independent $N(0, \sigma^2)$ random variables with σ^2 unknown. Assume that X has full rank p. Find the least squares estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and derive its distribution. Define the residual sum of squares RSS and write down an unbiased estimator $\hat{\sigma}^2$ of σ^2.

Suppose that $V_i = a + bu_i + \delta_i$ and $Z_i = c + dw_i + \eta_i$, for $i = 1, \ldots, m$, where u_i and w_i are known with $\sum_{i=1}^m u_i = \sum_{i=1}^m w_i = 0$, and $\delta_1, \ldots, \delta_m, \eta_1, \ldots, \eta_m$ are independent $N(0, \sigma^2)$ random variables. Assume that at least two of the u_i are distinct and at least two of the w_i are distinct. Show that $\mathbf{Y} = (V_1, \ldots, V_m, Z_1, \ldots, Z_m)^T$ (where T denotes transpose) may be written as in ($\dagger$) and identify X and $\boldsymbol{\beta}$. Find $\hat{\boldsymbol{\beta}}$ in terms of the V_i, Z_i, u_i and w_i. Find the distribution of $\hat{b} - \hat{d}$ and derive a 95% confidence interval for $b - d$.

[*Hint: You may assume that* $\frac{RSS}{\sigma^2}$ *has a* χ^2_{n-p} *distribution, and that* $\hat{\beta}$ *and the residual sum of squares are independent. Properties of* χ^2 *distributions may be used without proof.*]

Paper 1, Section II

19H Statistics

Suppose that X_1, X_2, and X_3 are independent identically distributed Poisson random variables with expectation θ, so that

$$\mathbb{P}(X_i = x) = \frac{e^{-\theta}\theta^x}{x!} \quad x = 0, 1, \ldots,$$

and consider testing $H_0 : \theta = 1$ against $H_1 : \theta = \theta_1$, where θ_1 is a known value greater than 1. Show that the test with critical region $\{(x_1, x_2, x_3) : \sum_{i=1}^{3} x_i > 5\}$ is a likelihood ratio test of H_0 against H_1. What is the size of this test? Write down an expression for its power.

A scientist counts the number of bird territories in n randomly selected sections of a large park. Let Y_i be the number of bird territories in the ith section, and suppose that $Y_1, \ldots, Y_n$ are independent Poisson random variables with expectations $\theta_1, \ldots, \theta_n$ respectively. Let a_i be the area of the ith section. Suppose that $n = 2m$, $a_1 = \cdots = a_m = a (> 0)$ and $a_{m+1} = \cdots = a_{2m} = 2a$. Derive the generalised likelihood ratio Λ for testing

$$H_0 : \theta_i = \lambda a_i \text{ against } H_1 : \theta_i = \begin{cases} \lambda_1 & i = 1, \ldots, m \\ \lambda_2 & i = m+1, \ldots, 2m. \end{cases}$$

What should the scientist conclude about the number of bird territories if $2\log_e(\Lambda)$ is 15.67?

[*Hint: Let $F_\theta(x)$ be $\mathbb{P}(W \leqslant x)$ where W has a Poisson distribution with expectation θ. Then*

$$F_1(3) = 0.998, \quad F_3(5) = 0.916, \quad F_3(6) = 0.966, \quad F_5(3) = 0.433\,.]$$

Paper 3, Section II

20H Statistics

Suppose that $X_1, \ldots, X_n$ are independent identically distributed random variables with

$$\mathbb{P}(X_i = x) = \binom{k}{x} \theta^x (1-\theta)^{k-x}, \quad x = 0, \ldots, k,$$

where k is known and θ $(0 < \theta < 1)$ is an unknown parameter. Find the maximum likelihood estimator $\hat{\theta}$ of θ.

Statistician 1 has prior density for θ given by $\pi_1(\theta) = \alpha\theta^{\alpha-1}$, $0 < \theta < 1$, where $\alpha > 1$. Find the posterior distribution for θ after observing data $X_1 = x_1, \ldots, X_n = x_n$. Write down the posterior mean $\hat{\theta}_1^{(B)}$, and show that

$$\hat{\theta}_1^{(B)} = c\hat{\theta} + (1-c)\tilde{\theta}_1,$$

where $\tilde{\theta}_1$ depends only on the prior distribution and c is a constant in $(0,1)$ that is to be specified.

Statistician 2 has prior density for θ given by $\pi_2(\theta) = \alpha(1-\theta)^{\alpha-1}$, $0 < \theta < 1$. Briefly describe the prior beliefs that the two statisticians hold about θ. Find the posterior mean $\hat{\theta}_2^{(B)}$ and show that $\hat{\theta}_2^{(B)} < \hat{\theta}_1^{(B)}$.

Suppose that α increases (but n, k and the x_i remain unchanged). How do the prior beliefs of the two statisticians change? How does c vary? Explain briefly what happens to $\hat{\theta}_1^{(B)}$ and $\hat{\theta}_2^{(B)}$.

[*Hint: The Beta*(α, β) $(\alpha > 0,\ \beta > 0)$ *distribution has density*

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1,$$

with expectation $\frac{\alpha}{\alpha+\beta}$ *and variance* $\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$. *Here,* $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$, $\alpha > 0$, *is the Gamma function.*]