Paper 1, Section I

7H Statistics

Describe the generalised likelihood ratio test and the type of statistical question for which it is useful.

Suppose that $X_1, \ldots, X_n$ are independent and identically distributed random variables with the Gamma$(2, \lambda)$ distribution, having density function $\lambda^2 x \exp(-\lambda x)$, $x \geqslant 0$. Similarly, $Y_1, \ldots, Y_n$ are independent and identically distributed with the Gamma$(2, \mu)$ distribution. It is desired to test the hypothesis $H_0 : \lambda = \mu$ against $H_1 : \lambda \neq \mu$. Derive the generalised likelihood ratio test and express it in terms of $R = \sum_i X_i / \sum_i Y_i$.

Let $F^{(1-\alpha)}_{\nu_1,\nu_2}$ denote the value that a random variable having the F_{ν_1,ν_2} distribution exceeds with probability α. Explain how to decide the outcome of a size 0.05 test when $n = 5$ by knowing only the value of R and the value $F^{(1-\alpha)}_{\nu_1,\nu_2}$, for some ν_1, ν_2 and α, which you should specify.

[You may use the fact that the χ^2_k distribution is equivalent to the Gamma$(k/2, 1/2)$ distribution.]

Paper 2, Section I

8H Statistics

Let the sample $x = (x_1, \ldots, x_n)$ have likelihood function $f(x; \theta)$. What does it mean to say $T(x)$ is a sufficient statistic for θ?

Show that if a certain factorization criterion is satisfied then T is sufficient for θ.

Suppose that T is sufficient for θ and there exist two samples, x and y, for which $T(x) \neq T(y)$ and $f(x; \theta)/f(y; \theta)$ does not depend on θ. Let

$$T_1(z) = \begin{cases} T(z) & z \neq y \\ T(x) & z = y. \end{cases}$$

Show that T_1 is also sufficient for θ.

Explain why T is not minimally sufficient for θ.

Paper 4, Section II

19H Statistics

From each of 3 populations, n data points are sampled and these are believed to obey

$$y_{ij} = \alpha_i + \beta_i(x_{ij} - \bar{x}_i) + \epsilon_{ij}, \quad j \in \{1, \ldots, n\},\ i \in \{1, 2, 3\},$$

where $\bar{x}_i = (1/n)\sum_j x_{ij}$, the ϵ_{ij} are independent and identically distributed as $N(0, \sigma^2)$, and σ^2 is unknown. Let $\bar{y}_i = (1/n)\sum_j y_{ij}$.

(i) Find expressions for $\hat{\alpha}_i$ and $\hat{\beta}_i$, the least squares estimates of α_i and β_i.

(ii) What are the distributions of $\hat{\alpha}_i$ and $\hat{\beta}_i$?

(iii) Show that the residual sum of squares, R_1, is given by

$$R_1 = \sum_{i=1}^{3} \left[\sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 - \hat{\beta}_i^2 \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 \right].$$

Calculate R_1 when $n = 9$, $\{\hat{\alpha}_i\}_{i=1}^3 = \{1.6, 0.6, 0.8\}$, $\{\hat{\beta}_i\}_{i=1}^3 = \{2, 1, 1\}$,

$$\left\{ \sum_{j=1}^{9} (y_{ij} - \bar{y}_i)^2 \right\}_{i=1}^{3} = \{138, 82, 63\}, \quad \left\{ \sum_{j=1}^{9} (x_{ij} - \bar{x}_i)^2 \right\}_{i=1}^{3} = \{30, 60, 40\}.$$

(iv) H_0 is the hypothesis that $\alpha_1 = \alpha_2 = \alpha_3$. Find an expression for the maximum likelihood estimator of α_1 under the assumption that H_0 is true. Calculate its value for the above data.

(v) Explain (stating without proof any relevant theory) the rationale for a statistic which can be referred to an F distribution to test H_0 against the alternative that it is not true. What should be the degrees of freedom of this F distribution? What would be the outcome of a size 0.05 test of H_0 with the above data?

Paper 1, Section II

19H Statistics

State and prove the Neyman-Pearson lemma.

A sample of two independent observations, (x_1, x_2), is taken from a distribution with density $f(x;\theta) = \theta x^{\theta-1}$, $0 \leqslant x \leqslant 1$. It is desired to test $H_0 : \theta = 1$ against $H_1 : \theta = 2$. Show that the best test of size α can be expressed using the number c such that

$$1 - c + c\log c = \alpha\,.$$

Is this the uniformly most powerful test of size α for testing H_0 against $H_1 : \theta > 1$?

Suppose that the prior distribution of θ is $P(\theta = 1) = 4\gamma/(1+4\gamma)$, $P(\theta = 2) = 1/(1+4\gamma)$, where $1 > \gamma > 0$. Find the test of H_0 against H_1 that minimizes the probability of error.

Let $w(\theta)$ denote the power function of this test at θ ($\geqslant 1$). Show that

$$w(\theta) = 1 - \gamma^\theta + \gamma^\theta \log \gamma^\theta.$$

Paper 3, Section II

20H Statistics

Suppose that X is a single observation drawn from the uniform distribution on the interval $\left[\theta-10, \theta+10\right]$, where θ is unknown and might be any real number. Given $\theta_0 \neq 20$ we wish to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = 20$. Let $\phi(\theta_0)$ be the test which accepts H_0 if and only if $X \in A(\theta_0)$, where

$$A(\theta_0) = \begin{cases} \left[\theta_0 - 8, \infty\right), & \theta_0 > 20 \\ \left(-\infty, \theta_0 + 8\right], & \theta_0 < 20\,. \end{cases}$$

Show that this test has size $\alpha = 0.10$.

Now consider

$$C_1(X) = \{\theta : X \in A(\theta)\},$$
$$C_2(X) = \left\{\theta : X - 9 \leqslant \theta \leqslant X + 9\right\}.$$

Prove that both $C_1(X)$ and $C_2(X)$ specify 90% confidence intervals for θ. Find the confidence interval specified by $C_1(X)$ when $X = 0$.

Let $L_i(X)$ be the length of the confidence interval specified by $C_i(X)$. Let $\beta(\theta_0)$ be the probability of the Type II error of $\phi(\theta_0)$. Show that

$$E[L_1(X) \mid \theta = 20] = E\left[\int_{-\infty}^{\infty} 1_{\{\theta_0 \in C_1(X)\}} d\theta_0 \;\middle|\; \theta = 20\right] = \int_{-\infty}^{\infty} \beta(\theta_0)\, d\theta_0.$$

Here $1_{\{B\}}$ is an indicator variable for event B. The expectation is over X. [Orders of integration and expectation can be interchanged.]

Use what you know about constructing best tests to explain which of the two confidence intervals has the smaller expected length when $\theta = 20$.