Paper 1, Section I

6H Statistics

State the Rao-Blackwell theorem.

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. Geometric(p) random variables; i.e., X_1 is distributed as the number of failures before the first success in a sequence of i.i.d. Bernoulli trials with probability of success p.

Let $\theta = p - p^2$, and consider the estimator $\hat{\theta} = 1_{\{X_1=1\}}$. Find an estimator for θ which is a function of the statistic $T = \sum_{i=1}^n X_i$ and which has variance strictly smaller than that of $\hat{\theta}$. [*Hint: Observe that T is a sufficient statistic for p.*]

Paper 2, Section I

6H Statistics

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with probability density function

$$f_\theta(x) = \frac{1}{2} + \frac{1_{\{x<\theta\}}}{2\theta} \quad \text{for } x \in [0,1],$$

with parameter $\theta \in (0,1)$.

(a) Write down the likelihood function, and show that the maximum likelihood estimator coincides with one of the samples.

(b) Consider the estimator $\tilde{\theta} = 4\overline{X} - 1$ where $\overline{X} = n^{-1}\sum_{i=1}^n X_i$. Is $\tilde{\theta}$ unbiased? Construct an asymptotic $(1-\alpha)$-confidence interval for θ around this estimator.

Paper 1, Section II
18H Statistics

A clinical study follows n patients being treated for a disease for T months. Suppose we observe $X_1, \ldots, X_n$, where $X_i = t$ if patient i recovers at month t, and $X_i = T+1$ if the patient does not recover at any point in the observation period. For $t = 1, \ldots, T$, the parameter $q_t \in [0,1]$ is the probability that a patient recovers at month t, given that they have not already recovered.

We select a prior distribution which makes the parameters $q_1, \ldots, q_T$ i.i.d. and distributed as $\text{Beta}(T, 1)$.

(a) Write down the likelihood function. Compute the posterior distribution of $(q_1, \ldots, q_T)$.

(b) The parameter γ is the probability that a patient recovers at or before month M. Write down γ in terms of $q_1, \ldots, q_T$. Compute the Bayes estimator for γ under the quadratic loss.

(c) Suppose we wish to estimate γ, but our loss function is asymmetric; i.e., we prefer to underestimate rather than overestimate the parameter. In particular, the loss function is given by

$$L(\delta, \gamma) = \begin{cases} 2|\gamma - \delta| & \text{if } \delta \geqslant \gamma \\ |\gamma - \delta| & \text{if } \delta < \gamma. \end{cases}$$

Find an expression for the Bayes estimator of γ under this loss function, in terms of the posterior distribution function F of γ. [You need not derive F.]

Paper 3, Section II

18H Statistics

Consider a linear model $Y = X\beta + \varepsilon$, where $X \in \mathbb{R}^{n\times p}$ is a fixed design matrix of rank $p < n/2$, $\beta \in \mathbb{R}^p$, and $\varepsilon \sim N(0, \sigma^2\Sigma_0)$, for some known positive definite matrix $\Sigma_0 \in \mathbb{R}^{n\times n}$ and an unknown scalar $\sigma^2 > 0$.

(a) Derive the maximum likelihood estimators $(\hat{\beta}, \hat{\sigma}^2)$ for the parameters (β, σ^2).

(b) Find the distribution of $\hat{\beta}$.

(c) Prove that $\hat{\beta}$ is the Best Linear Unbiased Estimator for β.

Now, suppose that $\varepsilon \sim N(0, \Sigma)$ where $\Sigma \in \mathbb{R}^{n\times n}$ is a diagonal matrix with

$$\Sigma_{ii} = \begin{cases} \sigma_1^2 & \text{if } i \leqslant n/2\,, \\ \sigma_2^2 & \text{if } i > n/2\,, \end{cases}$$

and where σ_1^2 and σ_2^2 are unknown parameters and n is even.

(d) Describe a test of size α for the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against the alternative $H_1 : \sigma_1^2 < \sigma_2^2$, using the test statistic

$$T = \frac{\|Y_1 - X_1(X_1^T X_1)^{-1} X_1^T Y_1\|^2}{\|Y_2 - X_2(X_2^T X_2)^{-1} X_2^T Y_2\|^2}$$

where,

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},$$

with $Y_1, Y_2 \in \mathbb{R}^{n/2}$ and $X_1, X_2 \in \mathbb{R}^{n/2\times p}$. [You must specify the null distribution of T and the critical region, and you may quote any result from the lectures that you need without proof.]

Paper 4, Section II
17H Statistics

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. observations from a zero-inflated Poisson distribution with parameters $\pi \in [0,1]$ and $\lambda > 0$, which has probability mass function

$$f_{\pi,\lambda}(x) = \begin{cases} \pi + (1-\pi)e^{-\lambda} & \text{if } x = 0, \\ (1-\pi)\frac{\lambda^x e^{-\lambda}}{x!} & \text{if } x = 1, 2, \ldots. \end{cases}$$

Let $n_0 = \sum_{i=1}^n 1_{\{X_i=0\}}$ and $S = \sum_{i=1}^n X_i$.

(a) What is meant by *sufficient statistic* and *minimal sufficient statistic*? Show that $T = (n_0, S)$ is a sufficient statistic. Is it minimal sufficient?

(b) Suppose the parameter λ is known to be equal to some value λ_0. We wish to test the null hypothesis $H_0 : \pi = 0$ against the alternative $H_1 : \pi = 1/2$. Suppose there exists a likelihood ratio test of size α for H_0 against H_1. Specify the test statistic and the critical region. Is this test uniformly most powerful for the alternative $H_1 : \pi > 0$?

(c) Now suppose that both π and λ are unknown. We wish to test the null hypothesis $H_0 : \pi = 1/2$ against the alternative $H_1 : \pi \in [0,1]$. State the asymptotic null distribution of the generalised likelihood ratio statistic:

$$W = 2 \log \frac{\max_{\lambda>0, \pi\in[0,1]} L(\lambda, \pi; X)}{\max_{\lambda>0} L(\lambda, 1/2; X)},$$

where $L(\lambda, \pi; X)$ is the likelihood function. Describe a test of size α using this statistic.

[*You may quote any result from the lectures that you need without proof.*]