Paper 1, Section I

7H Statistics

What does it mean to say that an estimator $\hat{\theta}$ of a parameter θ is *unbiased*?

An n-vector Y of observations is believed to be explained by the model

$$Y = X\beta + \varepsilon,$$

where X is a known $n \times p$ matrix, β is an unknown p-vector of parameters, $p < n$, and ε is an n-vector of independent $N(0, \sigma^2)$ random variables. Find the maximum-likelihood estimator $\hat{\beta}$ of β, and show that it is unbiased.

Paper 3, Section I

8H Statistics

In a demographic study, researchers gather data on the gender of children in families with more than two children. For each of the four possible outcomes GG, GB, BG, BB of the first two children in the family, they find 50 families which started with that pair, and record the gender of the third child of the family. This produces the following table of counts:

First two children	Third child B	Third child G
GG	16	34
GB	28	22
BG	25	25
BB	31	19

In view of this, is the hypothesis that the gender of the third child is independent of the genders of the first two children rejected at the 5% level?

[*Hint: the* 95% *point of a* χ_3^2 *distribution is* 7.8147*, and the* 95% *point of a* χ_4^2 *distribution is* 9.4877.]

Paper 1, Section II

18H Statistics

What is the *critical region* C of a test of the null hypothesis $H_0 : \theta \in \Theta_0$ against the alternative $H_1 : \theta \in \Theta_1$? What is the *size* of a test with critical region C? What is the *power function* of a test with critical region C?

State and prove the Neyman–Pearson Lemma.

If $X_1, \ldots, X_n$ are independent with common $\text{Exp}(\lambda)$ distribution, and $0 < \lambda_0 < \lambda_1$, find the form of the most powerful size-α test of $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda = \lambda_1$. Find the power function as explicitly as you can, and prove that it is increasing in λ. Deduce that the test you have constructed is a size-α test of $H_0 : \lambda \leqslant \lambda_0$ against $H_1 : \lambda = \lambda_1$.

Paper 4, Section II

19H Statistics

What is a *sufficient statistic*? State the factorization criterion for a statistic to be sufficient.

Suppose that $X_1, \ldots, X_n$ are independent random variables uniformly distributed over $[a, b]$, where the parameters $a < b$ are not known, and $n \geqslant 2$. Find a sufficient statistic for the parameter $\theta \equiv (a, b)$ based on the sample $X_1, \ldots, X_n$. Based on your sufficient statistic, derive an unbiased estimator of θ.

Paper 2, Section II

19H Statistics

What does it mean to say that the random d-vector X has a *multivariate normal* distribution with mean μ and covariance matrix Σ?

Suppose that $X \sim N_d(0, \sigma^2 I_d)$, and that for each $j = 1, \ldots, J$, A_j is a $d_j \times d$ matrix. Suppose further that

$$A_j A_i^T = 0$$

for $j \neq i$. Prove that the random vectors $Y_j \equiv A_j X$ are independent, and that $Y \equiv (Y_1^T, \ldots, Y_J^T)^T$ has a multivariate normal distribution.

[*Hint: Random vectors are independent if their joint MGF is the product of their individual MGFs.*]

If $Z_1, \ldots, Z_n$ is an independent sample from a univariate $N(\mu, \sigma^2)$ distribution, prove that the sample variance $S_{ZZ} \equiv (n-1)^{-1} \sum_{i=1}^n (Z_i - \bar{Z})^2$ and the sample mean $\bar{Z} \equiv n^{-1} \sum_{i=1}^n Z_i$ are independent.