Statistics Preliminary Exam

August 2023

There are 7 problems with a total of 200 points.

Show all of your work.

1. (25 points) Let a random variable X have the binomial distribution $b(p,n)$, and let the function $g(p)$ defined as $g(p) = p(1-p)$.

   (a) Show that the UMVUE of $g(p)$ is $\hat{\delta} = X(n-X)/n(n-1)$.
   (UMVUE: uniformly minimum variance unbiased estimator).

   (b) Determine the limiting distribution of $\sqrt{n}(\hat{\delta} - g(p))$ and $n(\hat{\delta} - g(p))$ when $g'(p) \neq 0$ and $g'(p) = 0$, respectively.

2. (25 points) Let the random variable X follow the *inverse Gaussian* distribution $I(\mu, \tau)$ with density

$$\sqrt{\frac{\tau}{2\pi x^3}} \exp\left(-\frac{\tau}{2x\mu^2}(x-\mu)^2\right), \quad x > 0, \ \ \tau, \mu > 0.$$

   (a) Find the moment generating function of X.

   (b) Show that $V = \frac{\tau}{X\mu^2}(X-\mu)^2 \sim \chi_1^2$

   Let $X_1, \ldots, X_n$ be a random sample from $I(\mu, \tau)$.

   (c) Show that $\overline{X} = \sum_{i=1}^n X_i/n \sim I(\mu, n\tau)$.

   (d) Show that there exists a UMP test for testing $H_0 : \mu \leq \mu_0$ versus. $H_1 : \mu > \mu_0$ when τ is known. (UMP: uniformly most powerful).

3. (25 points) Let $X_1, X_2, \ldots, X_m$ be a random sample from an exponential distribution with mean λ, and $Y_1, Y_2, \ldots, Y_n$ be a random sample form an exponential distribution with mean μ, and assume that the two samples are independent.

   (a) Find the LRT statistic, T, for testing the null hypothesis $H_0 : \lambda = \mu$ versus the alternative hypothesis $H_1 : \lambda \neq \mu$. (LRT: Likelihood Ratio Test).

   (b) Using a suitable one-to-one transformation of T, find the exact 5% critical region for the LRT in (a). Give the critical region in terms of the percentile(s) of a known distribution. Clearly identify the distribution and which percentiles(s), upper or lower.

4. (25 points) Let $X_1, \ldots, X_n$ be a random sample from $U(\theta, \theta+1)$, where $-\infty < \theta < \infty$ and it is unknown. Assume a prior distribution for θ given by the probability density function, for $-\infty < \theta < \infty$,

$$\pi(\theta) = \frac{1}{2} e^{-|\theta|} \ .$$

   (a) Find the posterior distribution of θ, given $(X_1 = x_1, \ldots, X_n = x_n)$, i.e., $\pi(\theta | x_1, \cdots, x_n)$.

   (b) Find the Bayes estimator of θ under the loss function, $L(\theta, \delta) = (\theta - \delta)^2$.

5. (30 points) Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be $n \times p_1$ and $n \times p_2$ matrices of predictors whose columns are linearly independent to each other. We consider the linear regression model below:

$$\boldsymbol{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon},$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are p_1- and p_2- dimensional vectors, respectively, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

(a) Express the ordinary least square estimator for $\boldsymbol{\beta} = \begin{pmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{pmatrix}$ using $\mathbf{X}_1$, $\mathbf{X}_2$, and $\boldsymbol{y}$.

(b) Let

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix}\mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{X}_1\boldsymbol{y} \\ \mathbf{X}_2\boldsymbol{y}\end{pmatrix}$$

be the ordinary least square estimator found above in (a). Find the explicit forms of $\mathbf{G}_{11}$, $\mathbf{G}_{12}$, $\mathbf{G}_{21}$, and $\mathbf{G}_{22}$.

(c) Based on the results in part (b), show that $\boldsymbol{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1\boldsymbol{y}$ when $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$.

6. (35 points) Suppose that data $\{(x_{ij}, y_{ij}) : i = 1, \ldots, n, j = 1, \ldots, p\}$ can be modeled as having a common slope γ and possibly different intercepts θ_i using the linear model,

$$Y_{ij} = \theta_i + \gamma x_{ij} + \epsilon_{ij},$$

where $\{\epsilon_{ij}\}$ are independently and identically distributed $N(0, \sigma^2)$ random variables. Assume that no vector $(x_{i1}, \ldots, x_{ip})$, for $i = 1, \ldots, n$, is proportional to the vector of 1s.

(a) Determine the ordinary least squares estimator of $(\theta_1, \ldots, \theta_n, \gamma)'$.

(b) Give an explicit expression for the size α likelihood-ratio test of the hypothesis,

$$H_0 : \theta_1 = \cdots = \theta_n = 0 \text{ versus } H_a : \text{ not } H_0$$

(c) Compute the power of the test that you derived in part (b). (There are several ways of defining the non-centrality parameter for the test. Pick any one of these, and use it consistently in this part.) Show that the power is independent of γ.

(d) State the power of the test when $\theta_1 = \cdots = \theta_n = 0$ and $\gamma = 2$.

7. (35 points) For a linear model given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i,$$

with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2 > 0$ for $i = 1, \ldots, n$, consider a centered model given by

$$y_i = \alpha + \beta_1(x_{i1} - \bar{x}_1) + \beta_2(x_{i2} - \bar{x}_2) + \cdots + \beta_p(x_{ip} - \bar{x}_p) + \epsilon_i$$

with $\bar{x}_j = \frac{1}{n}\sum_{i=1}^n x_{ij}$ for $j = 1, \ldots, p$.

(a) Let X be an $n \times p$ matrix of the predictors, i.e.,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

The centered model can be written as

$$\boldsymbol{y} = [\boldsymbol{j}\ \boldsymbol{X}_c] \begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{\epsilon}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)$, $\boldsymbol{j}$ is a p-dimensional column vector whose elements are all 1s, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$. Express $\boldsymbol{X}_c$ using $\boldsymbol{X}$, $\boldsymbol{I}_n$, and $\boldsymbol{J}_n$ where $\boldsymbol{I}_n$ is an n by n identity matrix and $\boldsymbol{J}_n$ is an n by n matrix of 1s.

(b) Show that the ordinary least squares estimators for α and $\boldsymbol{\beta}$ are given by $\bar{y}$ and $(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'\boldsymbol{y}$.

(c) Now assume that the covariance matrix of $\boldsymbol{\epsilon}$ is given as $\boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I} + \rho\boldsymbol{J}]$ with $0 < \rho < 1$. Show that the generalized least squares estimator for α and $\boldsymbol{\beta}$ are the same as the ordinary least squares estimator found in part (b).

Hint: Use the fact that the inverse of $\boldsymbol{V} = (1-\rho)\boldsymbol{I} + \rho\boldsymbol{J}$ can be written as $\boldsymbol{V}^{-1} = \frac{1}{(1-\rho)}\left(\boldsymbol{I} - \frac{\rho}{1+(n-1)\rho}\boldsymbol{J}\right)$.