STATISTICS QUALIFYING EXAMINATION SEPTEMBER 25, 1981

**INSTRUCTIONS:** Answer ANY FOUR problems, each in a separate blue book.

1. "Directional Data"

Suppose $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \ldots \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$

are independent identically distributed random vectors with

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} R\cos\theta \\ R\sin\theta \end{pmatrix}, I\right)$$

where $R$ is a known constant and $\theta$, $-\pi \le \theta < \pi$, is unknown.

(NOTE: the density of $\begin{pmatrix} X \\ Y \end{pmatrix}$ is given by

$$p_\theta(x,y) = (2\pi)^{-1}\exp\{-[(x-R\cos\theta)^2 + (y-R\sin\theta)^2]/2\}) .$$

(A) Calculate the MLE $\hat\theta_n$ of $\theta$ based on the observations $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$.

$\hat\theta = tg\left(\frac{\Sigma x_i}{\Sigma y_i}\right)$

(B) Show from first principles that

$$\sqrt{n}\,(\hat\theta_n - \theta) \xrightarrow{\mathcal{L}} N(0, v_\theta^2)$$

$= \frac{1}{R^2}$

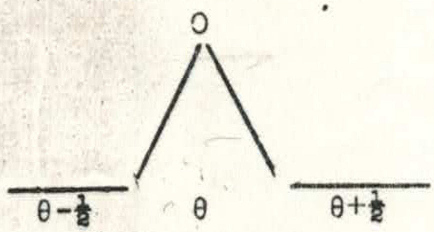
when $\theta$ is the true parameter. What is $v_\theta^2$ ?

NOTE: Assume $\theta \neq -\pi$. $v_\theta^2 = 1$ $\frac{1}{R^2}$

(C) Can you think of a better estimate of $\theta$ ?

2. Let $X_1, \ldots, X_n$ be iid from the density

$$f_\theta(x) = 2 - 4|x-\theta| \quad \text{if } |x-\theta| < \tfrac{1}{2}$$
$$= 0 \quad \text{if } |x-\theta| \ge \tfrac{1}{2} .$$

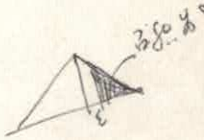



where $\theta$ is unknown

Let $\hat{\theta}_n$ be the midrange, ie,

$$\hat{\theta}_n = \tfrac{1}{2}\,[\max(X_1,\ldots,X_n) + \min(X_1,\ldots,X_n)]$$

and $\tilde{\theta}_n$ be the median of $X_1,\ldots,X_n$ .



(A) Show that $\{\hat{\theta}_n\}$ and $\{\tilde{\theta}_n\}$ are both weakly consistent sequences of estimators for $\theta$ , ie, $\forall \epsilon>0,\ \theta\in R$,

$$P_\theta\,\{|\hat{\theta}_n - \theta| > \epsilon\} \to 0$$

$\max X_i \xrightarrow{P} \theta+\frac{1}{2}$

$\min X_i \xrightarrow{P} \theta-\frac{1}{2}$.

$\sum_{i=1}^{\frac{n}{2}-1} \binom{n}{i} q^i p^{n-i} \quad (q<\tfrac{1}{2}) \to 1$ #

$\geq \sum_{|\frac{i}{n}-q|<\frac{1}{2}-q=\varepsilon}$

$$P_\theta\,\{|\tilde{\theta}_n - \theta| > \epsilon\} \to 0 .$$

$P(|m-\theta|<\varepsilon) \geq P(\sum I_{[X_i>\theta+\varepsilon)} < \frac{n}{2},\ \sum I_{[X_i<\theta-\varepsilon]} < \frac{n}{2})$

and $P(\sum I_{[X_i \gtrless \theta\pm\varepsilon]} < \frac{n}{2}) \to 1$

(B) Determine the asymptotic distribution theory for $\{\hat{\theta}_n\}$ and $\{\tilde{\theta}_n\}$: ie, find constants $\alpha_1$, $\alpha_2$ and nondegenerate distributions.

$F_1$ and $F_2$ such that

$$n^{\alpha_1}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} F_1$$
$$n^{\alpha_2}(\tilde{\theta}_n - \theta) \xrightarrow{\mathcal{D}} F_2 .$$

Which of the two estimators is better?

3. Let $\underset{\sim}{X} = (X_1,\ldots,X_n)$ and $\underset{\sim}{Y} = (Y_1,\ldots,Y_n)$ be independent random samples from two exponential distributions:

$$f_x(x;\theta_1) = \frac{1}{\theta_1}\exp\left(-\frac{x}{\theta_1}\right) \; ; \quad x \geq 0$$

and

$$f_y(y;\theta_2) = \frac{1}{\theta_2}\exp\left(-\frac{y}{\theta_2}\right) \; ; \quad y \geq 0$$

with unknown parameters $\theta_1$, $\theta_2$, respectively.

(i) Show that the critical region of a generalized likelihood ratio test of the hypothesis

H: $\theta_1 = \theta_2$ versus the alternative A: $\theta_1 \neq \theta_2$ depends only on the ratio ,

$$\frac{\sum_{i=1}^{m} x_i}{\sum_{j=1}^{n} Y_j}$$

(ii) Show that when $\theta_1 = \theta_2$ , the random variables

$$\frac{\sum_{i=1}^{m} X_i}{\sum_{j=1}^{n} Y_j} \text{ and } \left(\sum_{i=1}^{m} X_i + \sum_{j=1}^{n} Y_j\right)$$

$$\frac{\theta_1}{\theta_2 B(m,n)} \frac{\left(\frac{\theta_1}{\theta_2} x\right)^{m-1}}{\left(1+\frac{\theta_1}{\theta_2}x\right)^{n+m}}$$

are independent.

(iii) What is the distribution of the test statistic?

$$\frac{\sum_{i=1}^{m} X_i}{\sum_{j=1}^{n} Y_j}$$

$$\frac{1}{B(p,q)} \frac{V^{p-1}}{(V+1)^{p+q}}$$

(A) Under the null hypothesis $\theta_1 = \theta_2$ ?

$p = m$ , $q = n$

(B) Under the alternative hypothesis $\{\theta_1 \neq \theta_2\}$

$\frac{1}{\theta_1} X \sim \Gamma(m,1)$

$\frac{1}{\theta_2} Y \sim \Gamma(n,1)$

$\Rightarrow \dfrac{\frac{1}{\theta_1}\sum_1^m x_i}{\frac{1}{\theta_2}\sum_1^n y_i} \sim \frac{1}{B(p,q)}\frac{V^{p-1}}{(V+1)^{p+q}}$ ; $\dfrac{\sum^m x_i}{\sum^n y_i} \sim \dfrac{\theta_1}{\theta_2 B(p,q)} \left(\dfrac{V\theta_1}{\theta_2}\right)^{p-1}$

4. (A) Consider the general linear model

(1) $Y = X\beta + \epsilon$

where $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, $X$ is a full rank $n \times k$ matrix and

$\epsilon_1, \ldots, \epsilon_n$ are uncorrelated mean zero random variables, each with variance $\sigma^2$. The parameters $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$ are unknown.

Let $\hat{\beta}_1$ be the Least Squares estimate of $\beta_1$.

Now consider the restricted linear model

(2) $Y = X\beta + \epsilon$

where everything is the same as in (1) except that now $\beta_k$ is <u>known</u>: $\beta_k = b_k$ for some real number $b_k$.

Let $\tilde{\beta}_1$ be the Least Squares estimate of $\beta_1$ subject to the constraint $\beta_k = b_k$

Assuming that $\beta_k = b_k$, which of the two estimators $\hat{\beta}_1$, $\tilde{\beta}_1$ has smaller variance? Give a proof.

HINT: Use a famous theorem of Linear Models.

(B) Consider the model

$$Y = X\beta + \epsilon$$

where $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}$, $X = (x_{ij})$ is a four x four matrix,

$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_4 \end{pmatrix}$ are unknown parameters, and $\epsilon_1, \ldots, \epsilon_4$ are iid mean zero, variance $\sigma^2$.

Show that among the class of all 4x4 matrices $X = (x_{ij})_{i,j=1,\ldots,}$ with the property that

$$x_{ij} = 0, 1, \text{ or } -1 \quad \forall (i,j)$$

<u>none</u> gives rise to least-squares estimates of $\beta_1, \beta_2, \beta_3, \beta_4$ with smaller variances than those of the LS estimators determined by

张光庭

$$X^* = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

HINT:

(i) The LS estimates determined by $X^*$ all have variances 1/4.

(ii) Use the result of (A) to show that you can never do better than 1/4.

5. Suppose $\{\epsilon_n\}_{n=1}^{\infty}$, $\{X_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$ are independent sequences of iid random variables with

$$P\{\epsilon_n = 1\} = 1 - P\{\epsilon_n = 0\} = 1/2$$

$$P\{X_n = k\} = r^k(1-r) \ ; \quad k=0,1,2,\ldots$$

$$P\{Y_n = k\} = \lambda^k e^{-\lambda}/k! \ ; \quad k=0,1,2,\ldots$$

The parameters $r$ and $\lambda$ are unknown to the statistician, and he would like to estimate them. Unfortunately, he cannot observe the entire vector $\begin{pmatrix} \epsilon_n \\ X_n \\ Y_n \end{pmatrix}$ ; in fact, he only observes

$$W_n = \epsilon_n X_n + (1-\epsilon_n)Y_n$$

(thus at each stage he sees $X_n$ or $Y_n$ but he doesn't know which he is seeing).

(A) Find consistent estimates of $r$ and $\lambda$ based on $\{W_n\}_{n=1}^{\infty}$: that is, find functions

$$\hat{r}_n(w_1,\ldots,w_n) \quad \text{and} \quad \hat{\lambda}_n(w_1,\ldots,w_n)$$

such that

$$\begin{cases} 2EW_n = \frac{r}{1-r} + \lambda \\ 2\,EW_n^2 = \frac{r(1+r)}{(1-r)^2} + \lambda + \lambda^2 \end{cases}$$

$$\hat{r}_n(w_1,\ldots,w_n) \xrightarrow{\text{A.S.}} r$$

and $$\hat{\lambda}_n(w_1,\ldots,w_n) \xrightarrow{\text{A.S.}} \lambda$$

(B) Discuss the asymptotic distribution theory of your estimators. How does the asymptotic variance of $\hat{\lambda}_n$ compare with the asymptotic variance of the MLE of $\lambda$ based on $Y_1, Y_2, \ldots, Y_n$ ? (That is, what is the asymptotic "cost" of not knowing which population you're observing?).

6. Let $X_{ij}$ $(i=1,\ldots,k;\ j=1,\ldots,n)$ be independent random variables with

$$X_{ij} \sim N(\mu_i, \sigma^2) .$$

Suppose that the parameters $\mu_i$ $(i=1,\ldots,k)$ and $\sigma^2$ are unknown. How would you test the hypothesis that <u>at least one</u> of the means $\mu_i$ $(i=2,\ldots,k)$ satisfies

$$\mu_i > \mu_1 \ ?$$

7. State and prove the Gauss-Markov theorem in regression theory. Be sure to define all variables involved.