STATISTICS QUALIFYING EXAMINATION OCTOBER 3, 1980

<u>INSTRUCTIONS</u>: Answer <u>ANY FOUR</u> problems, each in a separate blue book.

1. Let $U_1, U_2, \ldots$ be i.i.d. uniform on $(0,1)$ and let Y_n denote the ν th order statistic of $U_1, \ldots, U_n$, where

$$\nu = n/2 + o(\sqrt{n}) \quad \text{as} \quad n \to \infty.$$

   (a) Show that $2\sqrt{n}\,(Y_n - \tfrac{1}{2}) \to N(0,1)$

$$\sqrt{2n}\,(Y_n - \tfrac{1}{2}) \to N(0,1) \quad \text{in distribution as} \quad n \to \infty.$$

   (b) Let $X_1, X_2, \ldots$ be i.i.d. with any continuous distribution function F and let M_n denote the median value of $X_1, \ldots, X_n$. Using the result of (a), find the asymptotic distribution of M_n as $n \to \infty$.

2. Let $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ be independent observations such that X_i has a continuous distribution function F and Y_i has a continuous distribution function G.

Let $k = \min(m,n)$

   (i) Give an unbiased estimate of $P[X < Y]$, where X and Y are independent observations from F and G respectively. $\frac{\sum_{i=1}^{k} I_{[x_i < y_i]}}{k}$ is unb

   (ii) Find the variance of the estimator in (i). $= P(x<y)[1-P(x<y)] \cdot \frac{1}{k}$.

   (iii) How would you test the hypothesis $F=G$? vs what? $F \neq G$
use rank statistic
or, using Smirnov test.

3. Consider the multinomial distribution

$$P(X_1=x_1, X_2=x_2, X_3=x_3 \mid X_1+X_2+X_3=N) = \frac{N!}{x_1!\,x_2!\,x_3!}\, p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

with $p_1+p_2+p_3 = 1$ and $p_i \geq 0$

(a) What is the maximum likelihood (ML) estimator of $p = (p_1, p_2, p_3)$?

(b) Let $p_1 = \theta^2$, $p_2 = 2\theta(1-\theta)$, $p_3 = (1-\theta)^2$, $0 \le \theta \le 1$. Find the ML estimate of θ.

(c) Find the ML estimate of θ if only $y_1 = x_1$ and $y_2 = x_2 + x_3$ are observable. Call this estimate $\overline{\theta}$.

(d) Find the exact or approximate mean and variance of $\hat{\theta}$ and $\overline{\theta}$, and compare the "goodness" in a suitable sense of $\hat{\theta}$ and $\overline{\theta}$ as estimates when x_1, x_2 and x_2 are observed.

4. Two laboratories each take n measurements on the same standard μ. Consider the model

$$y_{ij} = \mu + \epsilon_{ij} \quad i=1,2,\ldots;\ j=1,\ldots,n,$$

where the ϵ_{ij} are independent random variables with mean 0 such that $\operatorname{Var} \epsilon_{1j} = \sigma_1^2$, $\operatorname{Var} \epsilon_{2j} = \sigma_2^2$.

(i) Suppose it is known that $\sigma_2^2/\sigma_1^2 = 4$. Show that the minimum variance linear unbiased estimator of μ is given by

$$\hat{\mu} = \frac{4\,\overline{y}_1 + \overline{y}_2}{5}, \text{ where } \overline{y}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij}.$$

(ii) Generalize the above result to the case where there are k laboratories and the errors for the i^{th} laboratory are all independent with mean 0 and variance $a_i\sigma^2$ $(i=1,\ldots,k)$, where the a_i are known constants such that $\sum_{i=1}^{k} a_i = 1$. Discuss how you would estimate σ^2.

$$\hat{u} = \frac{\dfrac{\sum_{j=1}^{n_1} y_{1j}}{a_1} + \dfrac{\sum_{j=1}^{n_2} y_{2j}}{a_2} + \cdots + \dfrac{\sum_{j=1}^{n_k} y_{kj}}{a_k}}{\dfrac{n_1}{a_1} + \dfrac{n_2}{a_2} + \cdots + \dfrac{n_k}{a_k}}$$

if all n_k all equal $= n$

$$\frac{\overline{y}_1}{a_1} + \frac{\overline{y}_2}{a_2} + \cdots + \frac{\overline{y}_n}{a_k}$$

OK,

σ^2 by standard linear model

5. Let $\underset{\sim}{X} = (X_1, \ldots, X_n)$ be a random sample of size n from the density

$$\frac{x+1}{\theta(\theta+1)} e^{-x/\theta}, \quad x \geq 0,$$

$\log p(x) = \log(x+1) - \log\theta - \log(\theta+1) - \frac{x}{\theta}$

$\left[\frac{\partial}{\partial\theta}\log\right] = -\frac{1}{\theta} - \frac{1}{\theta+1} + \frac{x}{\theta^2} = \frac{x}{\theta^2} - \frac{2\theta+1}{\theta(\theta+1)} = \frac{1}{\theta^2}\left[x - \ldots\right.$

where θ is an unknown positive parameter.

(a) What is the Fisher information $I(\theta;n)$ $= n \cdot \frac{1}{\theta^4} Var(x)$ about θ contained in $\underset{\sim}{X}$?

under Linear transformation $a \sum x_i + b$ a, b constant

sample mean $\frac{\sum_1^n x_i}{n}$ est. of $\frac{\theta(2\theta+1)}{\theta+1} = EX$

(b) There is (essentially) one function $\tau(\theta)$ of the parameter, for which there exists an unbiased estimator $T(\underset{\sim}{X})$ with variance achieving the Cramér-Rao lower bound. Find (a version of) this function $\tau(\theta)$ and construct the appropriate estimator.

(c) Discuss the asymptotic properties (as $n \to \infty$) of your estimator $T(\underset{\sim}{X})$; in particular, see if you can exhibit a function $\sigma^2(\theta)$ of the parameter such that

$$\sqrt{n}\,[T(\underset{\sim}{X}) - \tau(\theta)] \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N[0, \sigma^2(\theta)].$$

$Var(x) = \frac{\theta^2(2\theta^2+4\theta+1)}{(\theta+1)^2}$

(d) Based on your answer in question (b), see whether you can construct a minimum variance unbiased estimate for

$$g(\theta) = \frac{(3+2\theta)(2+\theta)}{1+\theta}.$$

$EX + 6 = \frac{2\theta^2+\theta}{\theta+1} + \frac{6\theta+6}{\theta+1} = \frac{2\theta^2+7\theta+6}{\theta+1} = \frac{(3+2\theta)(2+\theta)}{\theta+1} = g(\theta)$

$\therefore \frac{\sum_1^n x_i}{n} + 6$ is a MVUE for $g(\theta)$ and achieves Cramér-Rao lower bound.

6. Consider the testing problem

H_0: $X_1, \ldots, X_n$ i.i.d. with the normal density

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\,\theta} e^{-x^2/2\theta^2}, \quad -\infty < x < \infty;$$

versus

H_1: $X_1, \ldots, X_n$ i.i.d. with the double exponential density

$$g_\theta(x) = \frac{1}{2\theta} e^{-|x|/\theta}, \quad -\infty < x < \infty,$$

where θ is an unknown parameter.

(i) Find the maximum likelihood estimator of θ under H_0 and under H_1 respectively.

(ii) Based on $X_1, \ldots, X_n$, find a "reasonable" test statistic for discriminating between H_0 and H_1, and describe how you would perform the test. Discuss any optimality properties your test might have. If you can think of more than one test, discuss their relative advantages and disadvantages.

using Kolmogorov test but θ unknown. 先估计.

$$\inf_\theta \sup_x \left(|F_n(x) - F_\theta(x)| \right) = \inf_\theta D_n(x; \theta)$$ for both

distributions, see $P(D_{normal}(x) < t)$

$P(D_{EXP}(x) < t)$ 那个大 就 assume 那个 dist.

如果能算到 t.