Stat 643 Exam 2

May 3, 2007

I have neither given nor received unauthorized assistance on this examination.

KEY

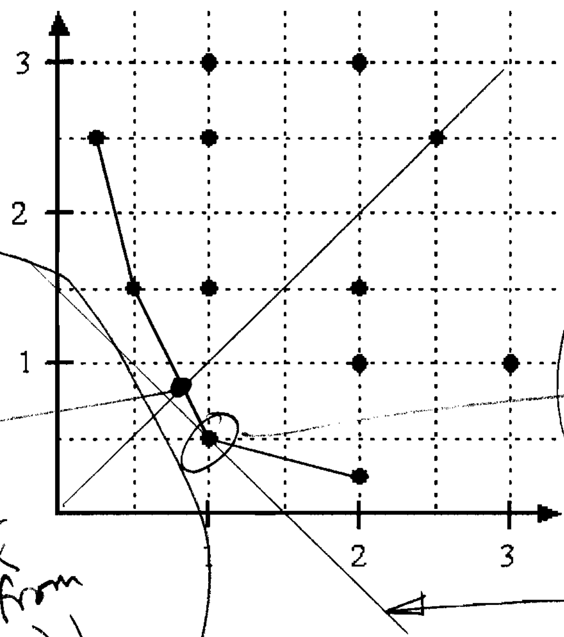
signature ______________________ date ______________

There are 10 "small" problems on this exam. Some are easier than others. All will be scored out of 10 points to make a total possible score of 100 pints.

1. Suppose that the figure below shows risk vectors for all non-randomized rules in a 2-state decision problem. Identify (give the coordinates of) the (possibly randomized) risk vectors that are

a) minimax

b) Bayes versus a prior that is uniform on $\Theta = \{\theta_1, \theta_2\}$



a)

$(1, .5)$

risk point Bayes vs the uniform prior

$\left(\frac{5}{6}, \frac{5}{6}\right)$

risk point that is minimax

(at the intersection of the line $y_1 = y_2$ and the line segment from $(1,.5)$ to $(.5, 1.5)$)

line $.5y_1 + .5y_2 = .75$

i.e. $y_1 + y_2 = 1.5$

line segment from $(1,.5)$ to $(.5, 1.5)$ is described by

$$\alpha(1,.5) + (1-\alpha)(.5, 1.5) \quad \text{for } \alpha \in [0,1]$$

When the 1st and second coordinates are equal

$$\alpha(1) + (1-\alpha)(.5) = \alpha(.5) + (1-\alpha)(1.5)$$

i.e. $\quad .5\alpha + .5 = -\alpha + 1.5$

or $\quad \alpha = \frac{2}{3}$

So the minimax risk vector has 1st coordinate

$$\frac{2}{3}(1) + \frac{1}{3}(.5) = \frac{5}{6}$$

2. Suppose that X and Y are independent, $X \sim \text{Bi}(n,p)$ and $Y \sim \text{Ber}(p)$. Consider squared error loss estimation of p and an estimator $\hat{p} = \delta(X)$. Let $W = X + Y$. Find an estimator $\hat{p}^* = \gamma(W)$ that improves upon $\hat{p}$. (Carefully say why your new estimator is better in terms of risk.)

W is sufficient for p. Our version of Rao-Blackwell promises an improvement on $\hat{p}$ since squared error is strictly convex in a. Consider the conditional dsn of $X \mid W = w$.

If $W=0$ Then $X=0$

If $W=n+1$ Then $X=n$

For $0<w<n+1$ if $W=w$ by symmetry

$X=w$ and $Y=0$ with conditional probability $\frac{n+1-w}{n+1}$

$X=w-1$ and $Y=1$ with conditional probability $\frac{w}{n+1}$

So

$$\gamma(W) = E\left[\delta(X) \mid W\right] = \begin{cases} \delta(0) & \text{if } W=0 \\ \delta(1) & \text{if } W=1 \\ \frac{W}{n+1}\delta(W-1) + \frac{n+1-W}{n+1}\delta(W) & \text{otherwise} \end{cases}$$

is an improvement on $\delta(X)$.

3. Consider a model for an observable X with real parameter $\alpha \in (0,1)$ and pdf on $(0,1)$

$$f(x \mid \alpha) = 1 + \alpha(2x - 1)$$

Write out explicitly as possible the Fisher information in X about α at α_0, $I_X(\alpha_0)$. (You'll probably have to leave this in terms of an integral that would need to be evaluated numerically.)

$$\frac{\partial}{\partial \alpha} \log f(x \mid \alpha) = \frac{2x-1}{1+\alpha(2x-1)}$$

So

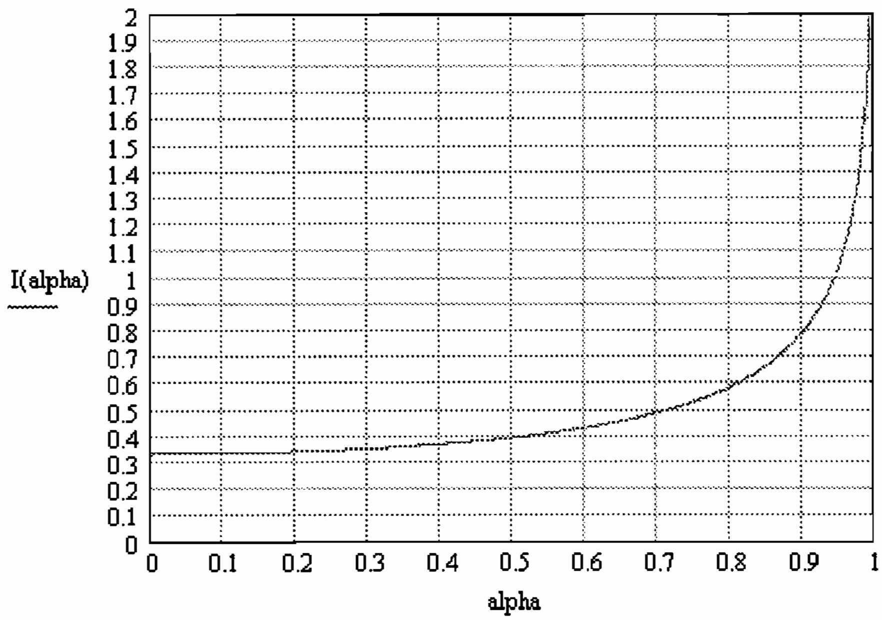
$$I_X(\alpha_0) = E_{\alpha_0}\left(\frac{2x-1}{1+\alpha_0(2x-1)}\right)^2$$

$$= \int_0^1 \frac{(2x-1)^2}{1+\alpha_0(2x-1)}\, dx$$

4. In the context of problem 3, $I_X(\alpha)$ can be computed numerically and looks as in the plot below. Based on the information in this plot, if $X_1, X_2, \ldots, X_{100}$ are iid with marginal density $f(x \mid \alpha)$,

a) what do you propose as an approximation to the distribution of an "MLE" of α if it is the case that $\alpha = .5$, and

b) what do you propose as (realizable/implementable) "large sample conservative" approximately 95% two-sided confidence limits for α, based on an "MLE" of α, say $\hat{\alpha}_{100}$?



a) $I_X(.5) = .354$ (about .4 read from the graph)

So an MLE will be approximately normal (with mean $\alpha = .5$ and variance $\dfrac{1}{100(.354)}$)

b) From the plot, $I_X(\alpha) \geq .33$ so using the approximately normal dsn of $\hat{\alpha}_{100}$ (with variance $\dfrac{1}{100(.33)}$) use

$$\hat{\alpha}_{100} \pm z \left(\frac{1}{\sqrt{100\, I_1(\alpha)}} \right) \leq \frac{1}{\sqrt{100(.33)}}$$

So use

$$\hat{\alpha}_{100} \pm 1.96 \frac{1}{\sqrt{100(.33)}}$$

5. Again in the context of problem 3, if $X_1, X_2, \ldots, X_n$ are iid with marginal density $f(x \mid \alpha)$,

a) what is the likelihood equation (that would need to be solved in order to find an MLE of α)?

b) give an *explicit* formula for an estimator of α that will be "asymptotically efficient" here.

a) $L_n(\alpha) = \sum_{i=1}^{n} \log f(X_i \mid \alpha)$ so the likelihood equation is $L_n'(\alpha) = 0$ i.e.

$$0 = \sum_{i=1}^{n} \frac{2X_i - 1}{1 + \alpha(2X_i - 1)}$$

b) Here we can make a "1-step Newton Improvement" on any consistent estimator $\tilde{\alpha}_n$. That is with

$$L_n'(\alpha) = \sum_{i=1}^{n} \frac{2X_i - 1}{1 + \alpha(2X_i - 1)} \quad \text{and} \quad L_n''(\alpha) = \sum_{i=1}^{n} \frac{-(2X_i - 1)^2}{(1 + \alpha(2X_i - 1))^2}$$

use

$$\hat{\alpha}_n = \tilde{\alpha}_n - \frac{L_n'(\tilde{\alpha}_n)}{L_n''(\tilde{\alpha}_n)}$$

$\tilde{\alpha}_n$ can, for example, be a method of moments estimator derived from the fact that

$$\begin{aligned} E_\alpha X &= \int_0^1 x(1 + \alpha(2x - 1))\,dx \\ &= \int_0^1 x\,dx + \alpha \int_0^1 2x^2 - x\,dx \\ &= \frac{1}{2} + \alpha\left(\frac{2}{3}x^3 \Big|_0^1 - \frac{x^2}{2}\Big|_0^1 \right) \\ &= \frac{1}{2} + \alpha\left(\frac{1}{6}\right) \end{aligned}$$

So $\alpha = 6\left(E_\alpha X - \frac{1}{2}\right)$ and a m.o.m. estimator is $\tilde{\alpha}_n = 6\left(\bar{X}_n - \frac{1}{2}\right)$.

6. Consider the problem of likelihood ratio testing in the *non-regular* family of Uniform $(0,\theta)$ distributions. That is, suppose that $X_1, X_2, \dots, X_n$ are iid Uniform $(0,\theta)$ and consider testing $\mathrm{H}_0: \theta = \theta_0$ versus $\mathrm{H}_a: \theta \neq \theta_0$. Let

$$\Lambda_n = 2\left(\sup_\theta L_n(\theta) - L_n(\theta_0)\right)$$

where (as usual) $L_n(\theta)$ is the n observation log-likelihood. What is the large sample distribution of Λ_n under the null hypothesis? (Hint: You may use without proof the facts that the MLE of θ is $\max_{i=1,2,\dots,n} X_i$, and that under the null hypothesis, $n\left(\theta_0 - \max_{i=1,2,\dots,n} X_i\right)$ is asymptotically Exponential with mean θ_0.)

The loglikelihood here is

$$\log\left(\left(\frac{1}{\theta}\right)^n I\left[\max X_i \le \theta\right]\right) = \begin{cases} -\infty & \text{if } \theta < \max X_i \\ -n \log \theta & \theta \geq \max X_i \end{cases}$$

So

$$\Lambda_n = 2\left(-n \log(\max X_i) + n \log \theta_0\right)$$

$$= 2n\left(\log(\theta_0) - \log(\max X_i)\right)$$

The Δ-method then shows that under H_0

$$n\left(\log(\theta_0) - \log(\max X_i)\right) \xrightarrow{d} \frac{1}{\theta_0} \operatorname{Exp}(\theta_0) \text{ i.e. } \operatorname{Exp}(1)$$

mean

So under H_0, $\Lambda_n \xrightarrow{d} \chi^2_2$

<u>not</u> 1 !!!

7. Suppose that as on the Exponential Families handout, for $\eta \in \Gamma \subset \Re$, distributions P_η have densities

$$f_\eta(x) = K(\eta)\exp(\eta T(x))h(x)$$

If $X_1, X_2, \ldots, X_n$ are iid P_{η_0} and $L_n(\eta)$ is the n observation log-likelihood, to what does $\frac{1}{n}L_n(\eta)$ converge in probability? (Give a formula in terms of the factors of f_η.)

$$\frac{1}{n}L_n(\eta) = \frac{1}{n}\sum_{i=1}^{n}\left(\log K(\eta) + \eta T(X_i) + \log h(X_i)\right)$$

$$= \log K(\eta) + \eta\left(\frac{1}{n}\sum_{i=1}^{n} T(X_i)\right) + \frac{1}{n}\sum_{i=1}^{n}\log h(X_i)$$

By the LLN, under P_{η_0} this converges in probability to

$$\log K(\eta) + \eta \underbrace{E_{\eta_0}T(X)}_{-\frac{K'(\eta_0)}{K(\eta_0)}} + E_{\eta_0}\log h(X)$$

8. Bayesians sometimes argue that their "posteriors are consistent." Consider the simplest possible version of this. Suppose that $\Theta = \{1,2\}$ and for some σ-finite measure μ, f_1 and f_2 are densities for two different probability distributions P_1 and P_2. For a prior distribution G on Θ, with $g_i = G(\{i\}) \in (0,1)$ for $i = 1,2$ and $X_1, X_2, \ldots, X_n$ iid with marginal one of P_1 and P_2,

a) what is the posterior probability (based on the n observations) that $\theta = 1$?
b) argue carefully that under the $\theta = 1$ model, the posterior probability that $\theta = 1$ from a) converges to 1 in probability.

a) The posterior probability that $\theta=1$ is the random variable

$$\frac{g_1 \prod_{i=1}^n f_1(X_i)}{g_1 \prod_{i=1}^n f_1(X_i) + g_2 \prod_{i=1}^n f_2(X_i)} = \frac{g_1 \prod_{i=1}^n \frac{f_1(X_i)}{f_2(X_i)}}{g_1 \prod_{i=1}^n \frac{f_1(X_i)}{f_2(X_i)} + g_2}$$

b) It will suffice to show that

$$\log\left(\prod_{i=1}^n \frac{f_1(X_i)}{f_2(X_i)}\right) \xrightarrow{P} \infty$$

for which it suffices to show that

$$\frac{1}{n} \log\left(\prod_{i=1}^n \frac{f_1(X_i)}{f_2(X_i)}\right) \xrightarrow{P} c > 0$$

But

$$\frac{1}{n} \log\left(\prod \frac{f_1(X_i)}{f_2(X_i)}\right) = \frac{1}{n} \sum_{i=1}^n \log \frac{f_1(X_i)}{f_2(X_i)}$$

$$\xrightarrow{P} I_X(P_1, P_2) > 0$$

under P_1

9. Suppose that $X \sim \text{Bi}(2, p)$ and that squared error loss of $p \in [0,1]$ is under consideration. Argue directly that for any $c \in (0,1)$ the estimator

$$\delta_c(X) = \begin{cases} 0 & \text{if } x = 0 \\ c & \text{if } x = 1 \\ 1 & \text{if } x = 2 \end{cases}$$

is admissible. (Hint: What must be the $p = 0$ risk and the $p = 1$ risk of any ϕ improving upon a δ_c. What does that say about the form of ϕ?)

If $p = 0$ then $P_0[X = 0] = 1$ and δ_c has 0 risk. Similarly $P_1[X = 2] = 1$ and δ_c has 0 risk.

If there is an estimator ϕ better than δ_c there is nonrandomized one better than δ_c, so assume ϕ is non-randomized. ϕ must have risk 0 at $p = 0$ and $p = 1$ and thus must have $\phi(0) = 0$ and $\phi(2) = 1$. (Otherwise it can not be at least as good as δ_c.) But then if $\phi(1) = a$ we somehow must have the p risk of ϕ no worse than that of δ_c $\forall$ $p \in (0,1)$ and better than that of δ_c for some $p \in (0,1)$. But

$$R(p, \delta_c) = (p)^2(1-p)^2 + (p-c)^2 2p(1-p) + (1-p)^2 p^2$$

and $$R(p, \phi) = p^2(1-p)^2 + (p-a)^2 2p(1-p) + (1-p)^2 p^2$$

So $$R(p, \delta_c) - R(p, \phi) = 2p(1-p)\left[(p-c)^2 - (p-a)^2\right]$$

But unless $a = c$ this is negative for $p = c$. On the other hand, if $a = c$ $\phi = \delta_c$ and is not better than δ_c. Thus there is no ϕ better than δ_c and δ_c is admissible.

10. An alternative to 0-1 loss for decision-theoretic treatments of some testing problems is the so-called "linear loss." That is, for an interval $\Theta \subset \Re$ and action space $\mathcal{A} = \{0,1\}$, if $\theta_0 \in \Theta$, testing $\mathrm{H}_0{:}\theta \le \theta_0$ vs $\mathrm{H}_a{:}\theta > \theta_0$ might be phrased in decision theoretic terms using a loss function

$$L(\theta,a) = \begin{cases} \max\{0, \theta - \theta_0\} & \text{for } a = 0 \\ \max\{0, \theta_0 - \theta\} & \text{for } a = 1 \end{cases}$$

For a prior G on Θ, and distributions P_θ of an observable X, what is the form of a Bayes rule for G? (Hint: Consider $L(\theta,0) - L(\theta,1)$.)

$$L(\theta,0) - L(\theta,1) = \theta - \theta_0$$

Action 0 is preferable when the posterior mean of this is negative and action 1 is preferable when the posterior mean is positive. But this is

Take action 0 when $\quad E[\theta|X] - \theta_0 < 0$

$$E[\theta|X] < \theta_0$$

Take action 1 when $\quad E[\theta|X] > \theta_0$