Stat 643 Exam 2
Spring 2010

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

KEY

Name Signed **Date**

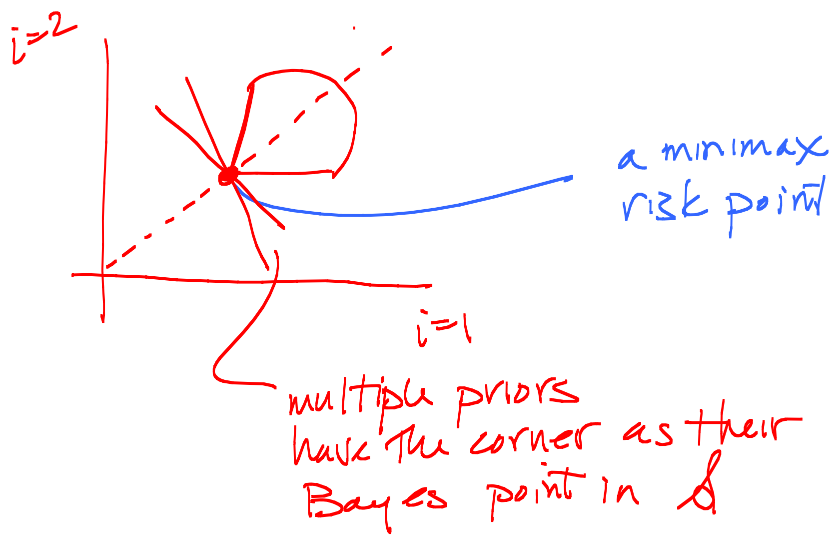
Name Printed

This exam is comprised of 10 equally weighted parts. Write answers for as many of those parts as you can in the 2 hours available for this exam. (You will surely NOT come close to finishing. Find things that you can do easily and completely.) Put your answers on the pages that follow, not on separate sheets.

1. Prove or give a counter-example to the following:

For a decision problem with finite parameter space $\Theta = \{1, 2, \ldots, m\}$, if a prior distribution G is least favorable, it is unique. (That is, if $G' \neq G$ then G' is not least favorable.)

This is false. It's easy to make an $m=2$ picture of a risk set where an equalizer rule produces a "corner" of $\mathcal{S}$ and thus there are many priors against which it is Bayes



2. Consider a decision problem with finite parameter space $\Theta = \{1, 2, \ldots, 9\}$, and action space consisting of two-element subsets of Θ, D @ $\{\{\theta, \theta'\} \mid \theta, \theta' \in \Theta, \theta \neq \theta'\}$. Then let

$$L(\theta, a) = I[\theta \notin a]$$

Suppose that a prior G on Θ and a likelihood $f_\theta(X)$ produce for each $x \in$ [a posterior over Θ that we will symbolize as $G(\theta \mid x)$. Give a prescription (in as explicit a form as possible) for a Bayes decision rule versus the prior G.

It suffices to for each x minimize posterior expected loss. This is

$$\sum_{\theta \notin a} G(\theta \mid x)$$

I do this by choosing a to contain two θ's with the largest possible corresponding posterior probabilities, i.e. θ and θ' maximizing

$$G(\theta \mid x) + G(\theta' \mid x)$$

3. Consider a decision-theoretic treatment of simple versus simple testing of $\mathrm{H}_0 : X \sim f_0$ versus $\mathrm{H}_1 : X \sim f_1$ with 0-1 loss, where

$$f_0(x) = I[0 < x < 1] \text{ and } f_1(x) = 2xI[0 < x < 1]$$

are densities with respect to Lebesgue measure on $\mathbb{R}$. Find a minimax test and identify a least favorable prior for this problem.

N-P tests produce risk points on the lower boundary of $\mathcal{S}$. Since

$$\frac{f_1}{f_0}(x) = 2x \quad \text{on } [0,1]$$

These are tests that decide in favor of f_1 for $x > c$. So risk points on the lower boundary of $\mathcal{S}$ are

$$\left(\int_c^1 dx, \int_0^c 2x\,dx \right) = (1-c, c^2)$$

An equalizer rule will then have $1-c = c^2$ i.e.

$$c^2 + c - 1 = 0 \quad \text{i.e.} \quad c = \frac{-1 \pm \sqrt{1+4}}{2} = -\frac{1}{2} \pm \frac{\sqrt{5}}{2}$$

Since $c > 0$ an equalizer rule has $c = -\frac{1}{2} + \frac{\sqrt{5}}{2}$.

Note that $x > c$ is $\frac{f_1}{f_0}(x) > 2c$. Neyman-Pearson tester with k is a Bayesian with prior $\frac{k}{k+1}$ on f_0 (and $\frac{1}{k+1}$ on f_1). So a least favorable prior here puts probability

$$\frac{\sqrt{5}-1}{\sqrt{5}-1+1} = \frac{\sqrt{5}-1}{\sqrt{5}} \quad \text{on } f_0$$

and $\frac{1}{\sqrt{5}}$ on f_1.

4. In the testing context of problem 3, suppose that one observes only $Y \equiv X \cdot I[X < .8]$ (and not X itself). Find most powerful tests of every size $\alpha \in [0,1]$ based on Y.

Note that $P_0[Y=0] = P_0[X \geq .8] = .2$ while

$P_1[Y=0] = P_1[X \geq .8] = x^2 \Big|_{.8}^{1} = 1 - .64 = .36$

So the likelihood ratio corresponding to $Y=0$ is $\frac{.36}{.2} = 1.8$.
For $x \in (0,.8)$ the likelihood ratio is $2x$. So (by N-P)

For $\alpha \in [0,.2]$ a most powerful size α test is

$$\phi_\alpha(y) = \begin{cases} \frac{\alpha}{.2} & \text{if } y=0 \\ 0 & \text{otherwise} \end{cases}$$

For $\alpha > .2$ a most powerful size α test is

$$\phi_\alpha(y) = \begin{cases} 1 & \text{if } y=0 \text{ or } y > 1-\alpha \\ 0 & \text{otherwise} \end{cases}$$

5. Consider squared error loss estimation of $\theta > 0$ based on X_1 and X_2 that are iid $\text{U}(0,\theta)$. The method of moments estimator of θ is $\hat{\theta} = X_1 + X_2$. $M \equiv \max(X_1, X_2)$ is well-known to be sufficient for θ, and conditional distributions for (X_1, X_2) given M are uniform on the union of two line segments in $\mathbb{R}^2$

$$\mathsf{S}_M = \{(x_1,x_2) \mid x_1 = M \text{ and } x_2 \in (0,M]\} \cup \{(x_1,x_2) \mid x_2 = M \text{ and } x_1 \in (0,M]\}$$

(Note that this means that marginally conditionally, X_1 is M with probability $\frac{1}{2}$ and with probability $\frac{1}{2}$ is $\text{U}(0,M)$.) Rao-Blackwellize $\hat{\theta}$ to produce an explicit formula for an estimator. What properties does this estimator have?

We want $E(X_1+X_2 \mid M) \overset{\text{symmetry}}{=} 2E(X_1 \mid M)$

$$= 2\left(\frac{1}{2}M + \frac{1}{2}\left(\frac{M}{2}\right)\right)$$

$$= \frac{3}{2}M$$

This is unbiased and better than $\hat{\theta}$. Provided M is complete (it is) Lehmann-Scheffé says that $\hat{\theta}$ is the UMVUE of θ.

6. A certain realized log-likelihood $L_n(\theta_1, \theta_2)$ for the two real parameters θ_1 and θ_2 in a regular problem is maximized at $\theta_1 = 2.0$ and $\theta_2 = 3.0$, where the Hessian of the log-likelihood (the matrix of second partials) is

$$\begin{pmatrix} -225 & -25 \\ -25 & -100 \end{pmatrix}$$

What are approximate 95% confidence limits for θ_1? What is an approximate significance level for testing $\mathrm{H}_0 : \theta_1 = 0$ and $\theta_2 = 0$? (You don't need to report a number, but tell me what tail probability of what distribution provides this.)

The negative inverse Hessian functions as an estimated covariance matrix for the (vector) MLE. This is (approximately)

$$-\begin{pmatrix} -225 & -25 \\ -25 & -100 \end{pmatrix}^{-1} \approx \begin{pmatrix} .004571 & -.001143 \\ -.001143 & .010286 \end{pmatrix}$$

Hence approximate 95% limits for θ_1 are

$$\hat{\theta}_1 \pm 1.96\, SE_{\hat{\theta}_1} \quad \text{i.e.} \quad 2.0 \pm 1.96\sqrt{.004571}$$

A Wald test statistic for $H_0: \theta_1 = \theta_2 = 0$ is

$$(\hat{\theta}_1 - 0, \hat{\theta}_2 - 0)(-\text{Hessian})\begin{pmatrix} \hat{\theta}_1 - 0 \\ \hat{\theta}_2 - 0 \end{pmatrix} = (2,3)\begin{pmatrix} 225 & 25 \\ 25 & 100 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

$$= 2175$$

An approximate p-value is then

$$P(\text{a } \chi^2_2 \text{ r.v.} > 2175) \approx 0$$

7. The interchange of orders of differentiation and integration is important to many of the arguments made in likelihood-based theory. For μ a sigma-finite measure on some space [, suppose $g(x,\theta): [\times \mathbb{R} \to \mathbb{R}$ is measureable in x for every fixed θ and differentiable in θ for every fixed x, ~~$g(x,\theta) \geq 0$~~, and there exists some integrable $M(x) \geq 0$ such

(handwritten: "and integrable" inserted after "measureable in x for every fixed θ"; "continuously" inserted before "differentiable")

$$\left| \frac{d}{d\theta} g(x,\theta) \right| \leq M(x) \quad \forall \ x \text{ and } \theta$$

(handwritten: (the bound need to be on the absolute derivative!!))

Prove then that

$$\frac{d}{d\theta} \int g(x,\theta)\, d\mu(x) = \int \frac{d}{d\theta} g(x,\theta)\, d\mu(x)$$

(Hint: Begin by writing the left hand side as a limit of a difference quotient. A first order Taylor approximation becomes relevant.)

* *

$$\frac{d}{d\theta} \int g(x,\theta)\, d\mu(x) = \lim_{\Delta \to 0} \frac{\int g(x,\theta+\Delta)\, d\mu(x) - \int g(x,\theta)\, d\mu(x)}{\Delta}$$

$$= \lim_{\Delta \to 0} \int \frac{g(x,\theta+\Delta) - g(x,\theta)}{\Delta}\, d\mu(x)$$

(need too that every $g(\cdot,\theta)$ is integrable)

$$= \lim_{\Delta \to 0} \int g'(x,\theta^*)\, d\mu(x)$$

(for some θ^* between θ and $\theta+\Delta$)

$$= \int \lim_{\Delta \to 0} g'(x,\theta^*)\, d\mu(x)$$

(dominated convergence Thm)

$$= \int \frac{d}{d\theta} g(x,\theta)\, d\mu(x)$$

($\theta^* \to \theta$ as $\Delta \to 0$)

(continuity of the derivative also needed)

8. Suppose that $X_1, X_2, \ldots, X_n$ are iid on $\mathbb{R}$ with marginal pdf $f_{\theta_1,\theta_2}(x)$ for real parameters θ_1 and θ_2. Write

$$m_1(\theta_1,\theta_2) = \mathrm{E}_{\theta_1,\theta_2} X \quad \text{and} \quad m_2(\theta_1,\theta_2) = \mathrm{E}_{\theta_1,\theta_2} X^2$$

and suppose that the mapping from $\mathbf{m} : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$\mathbf{m}(\theta_1,\theta_2) \equiv \left(m_1(\theta_1,\theta_2), m_2(\theta_1,\theta_2)\right)$$

is one-to-one and differentiable, with (differentiable) inverse $\mathbf{h}$. Identify a consistent estimator of (θ_1,θ_2) and identify a condition sufficient to guarantee that your estimator is root-n consistent. (Argue that your condition really is sufficient.)

The method of moments estimator

$$h\begin{pmatrix} \frac{1}{n}\sum X_i \\ \frac{1}{n}\sum X_i^2 \end{pmatrix}$$

is consistent for $\underset{\sim}{\theta}$ since by the LLN $\frac{1}{n}\sum X_i \xrightarrow{P_{\underset{\sim}{\theta}}} m_1(\underset{\sim}{\theta})$, $\frac{1}{n}\sum X_i^2 \xrightarrow{P_{\underset{\sim}{\theta}}} m_2(\underset{\sim}{\theta})$

(and the continuity of h).

Now if $E_{\underset{\sim}{\theta}} X_1^4 < \infty$ one has a bivariate CLT for

$$\sqrt{n}\begin{pmatrix} \frac{1}{n}\sum X_i - m_1(\underset{\sim}{\theta}) \\ \frac{1}{n}\sum X_i^2 - m_2(\underset{\sim}{\theta}) \end{pmatrix}$$

and so the Δ-method shows that

$$\sqrt{n}\left(h\begin{pmatrix} \frac{1}{n}\sum X_i \\ \frac{1}{n}\sum X_i^2 \end{pmatrix} - \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}\right)$$

converges in dsn under $\underset{\sim}{\theta}$.

9. Suppose that the density of X with respect to some sigma-finite measure μ is of 1-parameter exponential family form

$$f_\eta(x) = \mathrm{N}(\eta)\exp(\eta T(x))$$

for some real-valued function T. Under what conditions is the Cauchy-Schwarz inequality an equality? How then is it obvious that the Cramér-Rao inequality is achieved for the statistic $T(X)$? (Argue this not from a knowledge of what the C-R bound turns out to be, but from the condition that produces equality in Cauchy-Schwarz.)

C-S says that $EU^2 \cdot EV^2 \geq (EUV)^2$ with equality only when $P(U = aV + b) = 1$ for some a, b. The Cramér-Rao inequality is C-S applied to $\delta(X)$ and

$$g_\theta(X) = \frac{d}{d\theta}\ln f_\theta(X)$$

For a 1-parameter exponential family

$$\frac{d}{d\eta}\ln f_\eta(X) = \frac{d}{d\eta}\ln K(\eta) + T(X)$$

Thus this ↑ is clearly a linear function of $T(X)$ and we get equality in C-S and the C-R lower bound is achieved for $T(X)$.

10. Consider a "compound decision problem" as follows. $\Theta = \mathcal{D} = \{0,1\}^N$ (so both parameters and actions are N-vectors of 0's and 1's). Let

$$L(\theta, a) = \frac{1}{N}\sum_{i=1}^{N} I[a_i \neq \theta_i]$$

Suppose that for $\theta \in \Theta$, $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ has independent components, $X_i \sim \mathrm{N}(\theta_i, 1)$. (This is N discriminations between $\mathrm{N}(0,1)$ and $\mathrm{N}(1,1)$ where loss is an overall error rate.)

Let $\pi : \{1,2,\ldots,N\} \to \{1,2,\ldots,N\}$ be a permutation of the integers 1 to N, and define for $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ the transformation $\mathbb{R}^N \to \mathbb{R}^N$ $g_\pi(\mathbf{x}) = (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(N)})$. Then

$$\mathcal{G} = \{g_\pi \mid \text{such that } \pi \text{ is a permutation}\}$$

is a group. Show that the problem is invariant and say explicitly as possible what it means for a non-randomized decision rule to be equivariant in this context.

① Since for $\underset{\sim}{X} \sim P_{\underset{\sim}{\theta}}$, $\underset{\sim}{X}$ has independent $N(\theta_i, 1)$ components, $g_\pi(\underset{\sim}{X})$ has independent $N(\eta_i, 1)$ components where $\eta_i = \theta_{\pi(i)}$. So $\mathcal{G}$ leaves the model invariant and $\bar{g}_\pi : \Theta \to \Theta$ defined by

$$\bar{g}_\pi(\underset{\sim}{\theta}) = (\theta_{\pi(1)}, \theta_{\pi(2)}, \ldots, \theta_{\pi(N)})$$

satisfies $X \sim P_{\underset{\sim}{\theta}} \Rightarrow g_\pi(X) \sim P_{\bar{g}_\pi(\underset{\sim}{\theta})}$

② The loss is invariant because

$$L(\underset{\sim}{\theta}, \underset{\sim}{a}) = \frac{1}{N}\sum I[\theta_i \neq a_i] = \frac{1}{N}\sum I[\theta_{\pi(i)} \neq a_i'] = L(\bar{g}_\pi(\underset{\sim}{\theta}), \underset{\sim}{a}')$$

$\forall \underset{\sim}{\theta} \Rightarrow \underset{\sim}{a}' = (a_{\pi(1)}, a_{\pi(2)}, \ldots, a_{\pi(N)})$ (This probably deserves some careful argument...) So define $\tilde{g}_\pi : \mathcal{A} \to \mathcal{A}$ by $\tilde{g}_\pi(\underset{\sim}{a}) = (a_{\pi(1)}, a_{\pi(2)}, \ldots, a_{\pi(N)})$

And the problem is invariant.

In this context, an equivariant non randomized rule is one for which

$$\delta(g_\pi(\underset{\sim}{x})) = \tilde{g}_\pi(\delta(\underset{\sim}{x})) = (\delta(\underset{\sim}{x})_{\pi(1)}, \delta(\underset{\sim}{x})_{\pi(2)}, \ldots, \delta(\underset{\sim}{x})_{\pi(N)})$$

That is, permuting a data vector and making the N component decisions is the same as permuting N component decisions made with the original data vector.