Stat 643 Exam 1
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. A first order version of Taylor's Theorem for real-valued functions of a single real variable says that if a function $f(x)$ has a continuous 2nd derivative $f''(x)$ on an interval I containing a point x_0, then for any other $x \in I$

$$f(x) = f(x_0) + (x - x_0) f'(x_0) + \int_{x_0}^{x} f''(t)(x-t)\,dt$$

($f'(x)$ here is, of course, the first derivative of $f(x)$).

Use this fact about real-valued functions of a single real variable to prove the (complex number) bound used in class that says for any real number α,

$$\left|e^{i\alpha} - 1 - i\alpha\right| \le \frac{\alpha^2}{2}$$

$e^{i\alpha} = \cos\alpha + i\sin\alpha$ (expand both $\cos\alpha$ and $\sin\alpha$ at $\alpha = 0$)

$= 1 + \alpha(-\sin 0) + \int_0^\alpha (-\cos t)(\alpha - t)\,dt$

$+ i\left(0 + \alpha\cos 0 + \int_0^\alpha (-\sin t)(\alpha - t)\,dt\right)$

So $e^{i\alpha} - 1 - \alpha = -\int_0^\alpha (\cos t + i\sin t)(\alpha - t)\,dt$

and $\left|e^{i\alpha} - 1 - \alpha\right| = \left|\int_0^\alpha (\cos t + i\sin t)(\alpha - t)\,dt\right|$

$\alpha > 0$ case $\le \int_0^\alpha |\cos t + i\sin t|\,(\alpha - t)\,dt$

$\le \frac{\alpha^2}{2}$

2. Argue carefully that if $Y_1, Y_2, \ldots$ are iid random variables with $\mathrm{E}Y_1 = 0$ and $\mathrm{E}Y_1^2 = \sigma^2 < \infty$, then the double array $\{X_{ij}\}$ defined by $n_i = i$ and $X_{ij} = Y_j$ satisfies the Lindeberg condition.

Note than $B_i^2 = \mathrm{Var}\, S_i = i\sigma^2$ and

$$\frac{1}{B_i^2} \sum_{j=1}^{n_i} E X_{ij}^2 I[|X_{ij}| > \epsilon B_i]$$

$$= \frac{1}{i\sigma^2}\, i\, E Y_1^2 I[|Y_1| > \epsilon \sqrt{i}\,\sigma]$$

$$= \frac{1}{\sigma^2} E\, Y_1^2 I[|Y_1| > \epsilon\sqrt{i}\,\sigma]$$

$\leq Y_i^2$ and $EY_i^2 < \infty$

So the dominated convergence theorem says that since $Y_1^2 I[|Y_1| > \epsilon\sqrt{i}\,\sigma] \xrightarrow{a.s.} 0$ the above converges to 0 and Lindeberg is satisfied.

3. Suppose that $\{X_i\}$ are independent Bernoulli(p_i) random variables and let $S_n \equiv \sum_{i=1}^{n} X_i$. Let $0 < c < \frac{1}{2}$ and suppose further that each $p_i \in [c, 1-c]$. Argue carefully that

$$\frac{S_n - \mathrm{E}S_n}{\sqrt{\mathrm{Var}S_n}} \xrightarrow{d} \mathrm{N}(0,1)$$

Liapounov with $\delta = 2$ works fine here. For $X_i \sim \mathrm{Ber}(p_i)$

$$B_n^2 = \mathrm{Var}\, S_n = \sum_{i=1}^{n} p_i(1-p_i) \geq nc(1-c)$$

$E|X_i - p_i|^4 \leq 1$ so

$$\frac{1}{B_n^4} \sum_{i=1}^{n} E|X_i - p_i|^4 \leq \frac{n}{n^2 c^2 (1-c)^2} = \frac{1}{n(c(1-c))^2} \to 0,$$

So the Liapounov condition is met and Theorem 21 and Theorem 19 imply that

$$\frac{S_n - ES_n}{\sqrt{\mathrm{Var}\, S_n}} \xrightarrow{d} N(0,1)$$

4. Use a probabilistic argument to evaluate

$$\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T}\left(\frac{\sin(t/2)}{t/2}\right)^2 \exp(\lambda\cos t-\lambda)\cos(\lambda\sin t)\,dt$$

Hints: You may use the fact that for real $a>0$

$$e^{iat}-e^{-iat}=2i\sin(at)$$

Further,

$$\operatorname{Re} e^{\lambda\left(e^{it}-1\right)}=\exp(\lambda\cos t-\lambda)\cos(\lambda\sin t)$$

$\phi(t)=\frac{\sin\left(\frac{t}{2}\right)}{\frac{t}{2}}\exp(\lambda\cos t-\lambda)\cos(\lambda\sin t)$ is the chf of the sum of two independent variables, the first of which has (as in 8 on page 4 of the notes) a dsn on the integers which is an average of a Poisson λ and that of such a dsn reflected on the non-positive integers. The second is uniform on $\left(-\frac{1}{2},\frac{1}{2}\right)$.

Then the limit is exactly that of Corollary 6 for the interval $\left(-\frac{1}{2},\frac{1}{2}\right)$ and so is

$$e^{-\lambda}$$

(the Poisson(λ) probability at 0).

5. Consider the space $[= [-1,1]$ with Borel sigma-algebra, and probability measure P uniform on $[$. Further, consider the sets of subsets of $[$

$$A = \left\{ (a,b) \mid \ -1 < a < b < -\frac{3}{4} \right\}$$

$$B = \left\{ (c,d) \mid \ \frac{3}{4} < c < d < 1 \right\}$$

$$C = \left\{ (-c,-d) \cup (c,d) \mid \ \frac{1}{2} < c < d < \frac{3}{4} \right\}$$

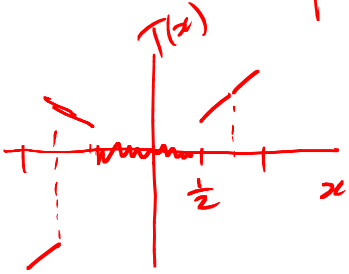
$$D = \left\{ \left(-\frac{1}{2}, \frac{1}{2} \right) \right\}$$

and let J be the σ-algebra generated by $A \cup B \cup C \cup D$. Identify a statistic $T : [\to \mathbb{R}$ that generates J. Then for $X(x) = x$, identify a function of $x \in [$ that is a version of $\mathrm{E}(X \mid J)$.

We need something for T that 1) is constant on $(-\frac{1}{2}, \frac{1}{2})$
2) maps $-x$ and x to the same value for $x \in (\frac{1}{2}, \frac{3}{4})$ and
3) is 1-1 on $(-1,-\frac{3}{4}) \cup (\frac{3}{4}, 1)$

One such statistic is

$$T(x) = \begin{cases} x & \text{for } x \in (-1,-\frac{3}{4}) \cup (\frac{3}{4},1) \\ |x| & \text{for } x \in (-\frac{3}{4},-\frac{1}{2}) \cup (\frac{1}{2},\frac{3}{4}) \\ 0 & \text{for } x \in (-\frac{1}{2},\frac{1}{2}) \end{cases}$$



$$E(X \mid \mathcal{G})(x) = \begin{cases} x & x \in (-1,-\frac{3}{4}) \cup (\frac{3}{4},1) \\ 0 = (\frac{1}{2})x + (\frac{1}{2})(-x) & \text{for } x \in (-\frac{3}{4},-\frac{1}{2}) \cup (\frac{1}{2},\frac{3}{4}) \\ 0 = \int_{-\frac{1}{2}}^{\frac{1}{2}} x\,dx & \text{for } x \in (-\frac{1}{2},\frac{1}{2}) \end{cases}$$

(Presumably, the conditional means of X given T are as above.)

6. Suppose that X_1, X_2, X_3 are iid with marginal probability density on $(0,1)$

$$f(x\,|\,\theta) = \frac{2}{\theta+1} I[0 < x < \theta] + \frac{1}{\theta+1} I[\theta \le x < 1]$$

for $\theta \in (0,1)$. Argue carefully that the order statistics $X_{(1)}, X_{(2)}, X_{(3)}$ are minimal sufficient for θ. Hint: The likelihood is not continuous.

The likelihood is discontinuous exactly at the order statistics. For two samples to have proportional likelihood, they must have exactly the same set of discontinuities and therefore must have the same set of order statistics. (Ties have 0 probability.) Therefore Theorem 55 says that the order statistics are minimal sufficient.

7. Consider the set of distributions S on a finite set [$=\{x_1, x_2, \ldots, x_k\}$ with k elements specified/parameterized by probability vectors $\mathbf{p} = (p_1, p_2, \ldots, p_k)$ with $P_{\mathbf{p}}(x_i) \equiv p_i \geq 0\ \forall i$ and $\sum_{i=1}^{k} p_i = 1$. The "entropy" associated with distribution $P_{\mathbf{p}}$ is

$$\mathrm{H}(\mathbf{p}) \equiv -\sum_{i=1}^{k} p_i \ln p_i$$

Relate entropy to K-L information for distributions on [and identify the maximum-entropy element of S . (Explain your choice.)

Consider for $\underset{\sim}{u} = \left(\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k}\right)$

$$I(P_{\underset{\sim}{p}}, P_{\underset{\sim}{u}}) = E_{\underset{\sim}{p}} \ln\left(\frac{p_x}{u_x}\right) = E_{\underset{\sim}{p}}\left(\ln p_x - \ln\left(\tfrac{1}{k}\right)\right)$$

$$= -\mathcal{E}(\underset{\sim}{p}) + \ln k$$

Now $I(P_{\underset{\sim}{p}}, P_{\underset{\sim}{u}}) \geq 0$ with equality only when $P_{\underset{\sim}{p}} = P_{\underset{\sim}{u}}$.

Since

$$\mathcal{E}(\underset{\sim}{p}) = \ln k - I(P_{\underset{\sim}{p}}, P_{\underset{\sim}{u}})$$

maximum-entropy $\underset{\sim}{p}$ is $\underset{\sim}{p} = \underset{\sim}{u}$.

8. Consider the measure $\mu = \lambda + \gamma$ on $\mathbb{R}$ where λ is Lebesgue measure on the interval $(0,1)$ and γ is counting measure on $\mathbb{Z}^+ = \{0,1,2,\ldots\}$. For $\eta \in \mathbb{R}$ such that

$$\int \exp(\eta x)\, d\mu(x) < \infty \quad .$$

let

$$f_\eta(x) = \frac{1}{\int \exp(\eta x)\, d\mu(x)} \exp(\eta x)$$

be an RN derivative with respect to μ of a probability distribution P_η on $\mathbb{R}$. Identify the natural parameter space for this family of distributions, and find the Fisher Information $I(\eta)$.

$$\int \exp(\eta x)\, d\mu(x) = \int \exp(\eta x)\, d\lambda(x) + \int \exp(\eta x)\, d\gamma(x)$$

$$= \underbrace{\int_0^1 \exp(\eta x)\, dx}_{\text{always finite}} + \underbrace{\sum_{x=0}^{\infty} e^{\eta x}}_{\text{finite only if } e^\eta < 1}$$

So $\Gamma = (-\infty, 0)$. For an exponential family, the FI is the variance of the natural sufficient statistic, the negative 2nd derivative of $\ln K(\eta)$, for $K(\eta)$ the normalizer. For $\eta < 0$

$$\int \exp(\eta x)\, d\mu = \frac{1}{\eta} \exp(\eta x)\Big|_0^1 + \frac{1}{1-e^\eta}$$

$$= \frac{\exp\eta - 1}{\eta} + \frac{1}{1-e^\eta} = \frac{1-(e^\eta-1)^2}{\eta(1-e^\eta)}$$

$K(\eta) = \frac{1}{\text{above}}$ $\quad -\ln K(\eta) = \ln(\text{above})$

$$= \ln\left(1-(e^\eta-1)^2\right) - \ln\eta - \ln(1-e^\eta)$$

and the 2nd derivative of this is $I(\eta)$

9. Consider the small family of four distributions $\mathtt{S} = \{P_1, P_2, P_3, P_4\}$ on the interval $[\ = (1,2)$ with densities with respect to Lebesgue measure

$$f_1(x) = I[1 < x < 2]$$

$$f_2(x) = \frac{2}{3} x \cdot I[1 < x < 2]$$

$$f_3(x) = \frac{3}{7} x^2 \cdot I[1 < x < 2]$$

$$f_4(x) = \frac{1}{\ln 2}\left(\frac{1}{x}\right) I[1 < x < 2]$$

Identify a one-dimensional sufficient statistic in a statistical problem involving n iid observations, each P_θ distributed. (That is, the observation space is $[\,^n = (1,2)^n$ and the family of possible distributions is $\{P_1^n, P_2^n, P_3^n, P_4^n\}$.)

Theorem 56 promises that a minimal sufficient statistic here will be

$$\left(\frac{f_2(\underset{\sim}{x})}{f_1(\underset{\sim}{x})}, \frac{f_3(\underset{\sim}{x})}{f_1(\underset{\sim}{x})}, \frac{f_4(\underset{\sim}{x})}{f_1(\underset{\sim}{x})} \right)$$

This is essentially

$$\left(\prod_{i=1}^n X_i, \left(\prod_{i=1}^n X_i\right)^2, \prod_{i=1}^n \left(\frac{1}{X_i}\right) \right)$$

which in turn is clearly a function of the 1-dimensional statistic

$$T(X) = \prod_{i=1}^n X_i$$

10. Consider an FI regular model for a continuous real-valued random variable X, say $\mathcal{S} = \{P_\theta\}$, with probability densities $f_\theta(x)$ (for, say, $\theta > 0$). Let

$$[X] = \text{the integer closest to } X$$

You may assume that the family of distributions for $[X]$ inherits FI regularity from the model for X. Let $I_X(\theta)$ and $I_{[X]}(\theta)$ be Fisher Information about θ in respectively X and $[X]$. Give a necessary and sufficient condition on the densities $f_\theta(x)$ under which the integer rounding of X to $[X]$ does not reduce the Fisher Information (that is $I_{[X]}(\theta) = I_X(\theta)$).

It is necessary and sufficient that $[X]$ be sufficient for $\mathcal{P}$. That requires that the conditional densities for $X \mid [X]$ are free of θ. This is

$$\frac{f_\theta(x)}{\int_{[X]-\frac{1}{2}}^{[X]+\frac{1}{2}} f_\theta(x)\,dx} I\left[[X]-\frac{1}{2} < x < [X]+\frac{1}{2} \right]$$

is independent of θ. That is, there must be functions h_j $(j = 0, \pm 1, \pm 2, \ldots)$ with $h_j(x) \geq 0$ and $\int_{j-1/2}^{j+1/2} h_j(x)\,dx = 1$ such that

for $p_j(\theta) = P_\theta\left[[X] = j \right]$

$$f_\theta(x) = \sum_{j=-\infty}^{\infty} p_j(\theta)\, h_j(x)$$