Stat 643 Exam 1
Spring 2010

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

____

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 2^{nd} 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,

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

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.

3. Suppose that $\{X_i\}$ are independent $\text{Bernoulli}(p_i)$ 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

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

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,

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

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

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

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

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

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

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

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

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

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.

7. Consider the set of distributions $\mathscr{P}$ on a finite set $\mathscr{X} = \{x_1, x_2, \ldots, x_k\}$ 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

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

Relate entropy to K-L information for distributions on $\mathscr{X}$ and identify the maximum-entropy element of $\mathscr{P}$. (Explain your choice.)

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 $I(\eta)$.

9. Consider the small family of four distributions $\mathscr{P} = \{P_1, P_2, P_3, P_4\}$ on the interval $\mathscr{X} = (1, 2)$ 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 $\mathscr{X}^n = (1, 2)^n$ and the family of possible distributions is $\{P_1^n, P_2^n, P_3^n, P_4^n\}$.)

10. Consider an FI regular model for a continuous real-valued random variable X, say $\mathcal{P} = \{P_\theta\}$, 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 $I_{[X]}(\theta) = I_X(\theta)$).