UNIVERSITY OF MASSACHUSETTS
Department of Mathematics and Statistics
ADVANCED EXAM - Mathematical Statistics and Probability
January 24, 2008

70 points are required to pass. Of these a minimum of 25 point is needed from each section; Probability (problems 1-3) and Statistics (problems 4 and 5). Problem 1 is worth 15 points, problems 2-4 are each worth 20 points and problem 5 is worth 25 points.

1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and X a random variable mapping X into $\mathbf{R}$.

   (a) Prove that

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n).$$

   (b) Let $c > 0$ be a fixed positive number. Prove that

$$E(|X|) < \infty \text{ if and only if } \sum_{n=1}^{\infty} P(|X| \geq cn) < \infty.$$

   Conclude that if $\sum_{n=1}^{\infty} P(|X| \geq cn) < \infty$ for one value of $c > 0$, then this series converges for all values of $c > 0$.

2. Let $(\Omega, \mathcal{F}, P)$ be a probability space, X an integrable random mapping X into $\mathbf{R}$, and $\mathcal{G}$ a σ-algebra of subsets of Ω satisfying $\mathcal{G} \subset \mathcal{F}$ (i.e., $\mathcal{G}$ is a σ-subalgebra of $\mathcal{F}$).

   (a) State the properties that characterize $E(X|\mathcal{G})$, the conditional expectation of X with respect to $\mathcal{G}$.

   (b) Using a well known theorem in measure theory, prove that $E(X|\mathcal{G})$ exists.

   (c) Let Y be another integrable random variable mapping X into $\mathbf{R}$. Prove that if $X \leq Y$ a.s., then $E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$ a.s.

   (d) Let $\mathcal{A} = \{A_j, j \in \mathbf{N}\}$ be a countable collection of disjoint subsets of $\mathcal{F}$ satisfying $P(A_j) > 0$ for all $j \in \mathbf{N}$. Define $\mathcal{G}$ to be the σ-subalgebra of $\mathcal{F}$ generated by $\mathcal{A}$.

      (i) Describe the form of the sets in $\mathcal{G}$.

      (ii) Give a formula for $E(X|\mathcal{G})$ involving the sets A_j.

3. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{X_i, i \in \mathbf{N}\}$ a sequence of independent, identically distributed random variables mapping Ω into $\mathbf{R}$ and satisfying $E(X_i) = 0$ and $M = E(|X_i|^4) < \infty$ for all $i \in \mathbf{N}$. Define $S_n = \sum_{i=1}^n X_i$.

   (a) Prove that $\sigma^2 = E(|X_i|^2) < \infty$ for all $i \in \mathbf{N}$. Then use this fact to prove the weak law of large numbers for S_n/n. Prove the weak law of large numbers directly; do not deduce it from part (c) of this problem.

   (b) Prove that for any $\delta > 0$ there exists $C < \infty$ such that for all $n \in \mathbf{N}$

   $$P(|S_n/n| \geq \delta) \leq C/n^2.$$

   (c) Prove the strong law of large numbers for S_n/n. (**Hint.** Using part (b), derive a suitable upper bound for $P(A^c)$, where $A = \{S_n/n \to 0\}$).

4. Let $X_1, \ldots, X_n$ be i.i.d. from a gamma distribution with parameters α and γ, both positive; i.e.,

   $$f(x_i) = \frac{x_i^{\alpha-1} e^{-x_i/\gamma}}{\Gamma(\alpha)\gamma^\alpha} I_{(0,\infty)}(x_i)$$

   (a) Write out the log-likelihood function and get the likelihood/score equations for obtaining the maximum likelihood estimates of α and γ. You can just leave the derivative of $\Gamma(\alpha)$ denoted as $\Gamma'(\alpha)$ in your solution.

   (b) Find the Fisher Information matrix.

   (c) The likelihood equations do not have a closed form solution. Describe how you would proceed using either Fisher's scoring method or Newton-Raphson to proceed iteratively to obtain a solution to the likelihood equation (assuming for now that it exists).

   (d) Assuming that $g(\alpha) = log(\alpha) - (\Gamma'(\alpha)/\Gamma(\alpha))$ is monotonic in α, argue that a solution to the likelihood equations exists almost surely.

5. Let $X_1, \ldots X_n$ be i.i.d. $N(\mu, \sigma^2)$.

   (a) Find a set of two complete sufficient statistics. Note: You can apply a result for exponential families, but state it carefully and justify any conditions that need to hold to use the result.

(b) Find the UMVUE for each of μ, σ^2 and σ. State (without proof) what general result you are applying to construct UMVUE's but be sure to prove the unbiasedness of your answers. HINT: Recall that a chi-square with d degrees of freedom, is equivalent to a gamma distribution from the previous problem with $\alpha = d/2$ and $\gamma = 2$.

(c) Consider testing $H_0 : \mu \geq \mu_0$ versus $H_A : \mu < 0$. and define $T = (\bar{X} - \mu_0)/(S/n^{1/2})$ where $\bar{X} = \sum_i X_i/n$ and $S^2 = \sum_i (X_i - \bar{X})^2/(n-1)$.

Note: In answering this question you can use without proof what you know about distribution of the sample mean $\bar{X}$ and sample variance S^2 and the fact that they are independent.

i. First argue that when $\mu = \mu_0$, T is distributed t with $n-1$ degrees of freedom.

ii. Show that the test which rejects H_0 if $T < -t_{n-1,\alpha}$ is a likelihood ratio test, where $t_{n-1,\alpha}$ is the 100α the percentile of the t with $n-1$ degrees of freedom.

iii. Define the size of a test and argue that this test has size α. Note that you need to define the power function in doing this and address the monotonicity of the power function. Hint: Think conditioning.

(d) Now, drop the normality assumption.

- First prove that S^2 is unbiased for σ^2. Assume the second moment of X_i is finite.

- Now, assuming that $E(X_i^4) < \infty$, prove that S^2 converges in probability to (i.e., is consistent for) σ^2. (Aside: This is a key piece in arguing that the limiting distribution of T as n increases is a standard normal, which provides a robust large sample test for the problem in part c)).