Department of Mathematics
Ph.D. Qualifying Examination
Time: 180 minutes

Statistics

May 18, 2007

General Instructions:
- There are two parts in this examination. Part A (STA 584) has 9 questions and Part B (STA 684) has 5 questions.
- Begin each question on a new sheet with the question number clearly labeled. Write on one side only. When finished, please arrange all pages according to the question numbers and then number the pages accordingly.
- You must show all your work correctly to earn full credits. Partial credits will be given for partially correct solutions.

Part A [Answer all questions]

1. In a certain community, 8% of all adults over 50 have diabetes. If a health service in this community correctly diagnoses 95% of all people with diabetes as having the disease and incorrectly diagnoses 2% of all persons without diabetes as having the disease, find the probabilities that a person over 50 diagnosed by the health service as having diabetes actually has the disease.

2. A random variable $X$ has a gamma distribution if and only if its probability density is given by $f(x) = \begin{cases} kx^{\alpha-1}e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$ where $\alpha > 0$ and $\beta > 0$.
   a) Find $k$.
   b) Derive the moment generating function of $X$.
   c) Use your result of part (b) to find the mean and variance of $X$.
   d) Use your result of part (b) to find the moment generating function of an exponential random variable with a mean of $\lambda$.
   e) Use your result of part (c) to find the mean and variance of a chi-square random variable with $n$ degrees of freedom.

3. Suppose that the joint density function of $X$ and $Y$ is defined by
   \[ f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1 \\ 0 & \text{elsewhere} \end{cases} \]
   a) What is density function of $X$ given $Y = y$?
   b) Find the conditional mean of $X$ given $Y = y$.
   c) Find the conditional probability $P(X > 1/2 | Y = 1/4)$.
   d) Assume that $E(X) = 5/9$ and $E(Y) = 5/6$, find the covariance between $X$ and $Y$.
   e) Find $E(X^2 + Y^2)$. 
4. Suppose that \( X \) and \( Y \) are independent continuous random variables with the following density functions: \( f_X(x) = 1, \) for \( 0 < x < 1 \) and \( f_Y(y) = 2y, \) for \( 0 < y < 1. \) Find \( P(Y < X). \)

5. Employees of a large company all choose one of three levels of health insurance coverage, for which premiums, denoted by \( X, \) are 1, 2, and 3, respectively. Premiums are subject to a discount, denoted by \( Y, \) of 0 for smokers and 1 for non-smokers. The joint distribution of \( X \) and \( Y \) is given by:
\[
P(X = x, Y = y) = \frac{x^2 + y^2}{31}, \text{ for } x = 1, 2, 3 \text{ and } y = 0, 1.
\]

a) What is the probability distribution of \( X - Y, \) the total premium paid by a randomly chosen employee?

b) Calculate \( \text{Var}(Y|X = 1). \)

6. Three individuals are running a one kilometer race. The completion time for each individual is a random variable. Let \( X_i \) be the completion time, in minutes, for person \( i. \)

\[\begin{align*}
X_1 &: \text{uniform distribution on the interval } [2.5, 3.1] \\
X_2 &: \text{uniform distribution on the interval } [2.6, 3.2] \\
X_3 &: \text{uniform distribution on the interval } [2.7, 3.3]
\end{align*}\]

a) Find the probability that the earliest completion is less than 3 minutes.

b) Find the probability that the latest completion is less than 3 minutes.

7. Suppose that \( X \) and \( Y \) are independent exponential random variables, each with mean 1. Suppose that \( U = Y/X. \) Find the probability density function of \( U. \)

8. If the distribution of the weights of all men traveling by air between Dallas and El Paso has a mean of 153 pounds and a standard deviation of 18 pounds, what is the probability that the total weight of 36 men traveling on a plane between these two cities is more than 5,400 pounds?

9. Let \( X, Y, \) and \( Z \) be independent Poisson random variables with \( E(X) = 3, \) \( E(Y) = 1 \) and \( E(Z) = 1. \) Find \( P(X + Y + Z < 2). \)
Part B [Answer all questions]

Question #10
(a) Let $X$ be a random variable such that $P(X \leq 0) = 0$, $k \geq 1$ and let $\mu = \text{E}(X)$ exist. Prove that $P(X \geq \mu k) \leq k^{-1}$.

(b) Define each of the following concepts in respect of a sequence $\{X_n\}$, $(n \geq 1)$ of random variables:
(i) Convergence in probability.
(ii) Convergence in distribution.
(c) Let $Y_n$ denote the maximum of a random sample of size $n$ from a distribution that has the probability density function $f(x) = \theta^{-1}$, $0 < x < \theta$, zero elsewhere. Let $U_n = nY_n$. Does $U_n$ converge in distribution to some random variable $U$? If so, find the probability density function of $U$.
(d) For the following sequences of independent random variables, does the weak law of large numbers hold?
   (i) $P(X_n = \pm n) = 1/(2\sqrt{n})$, $P(X_n = 0) = 1-1/\sqrt{n}$
   (ii) $P(X_n = \pm 2^n) = 2^{-3n-1}$, $P(X_n = \pm 1) = (1-2^{-3n})/2$
   (iii) $P(X_n = \pm \sqrt{n}) = 2^{-1}$

Question #11
(a) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution that has a mean $\mu$ and a variance $\sigma^2$. Let $\bar{X}_n$ denote the sample mean. Prove that the random variable $Y_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting distribution that is normal with mean zero and variance 1.

(b) Let $\bar{X}_n$ denote the mean of a random sample of size $n$ from a distribution that has the probability density function $f(x) = e^x$ for $x < 0$.
   (i) Show that the moment generating function of $X$ is $M(t) = (1+t)^{-1}$.
   (ii) Find the moment generating function $M(t, n)$ of $Y_n = \sqrt{n}(\bar{X}_n + 1)$.
   (iii) By taking the limit of the moment generating function $M(t, n)$ in (ii), show that the limiting distribution of $Y_n$ as $n \to \infty$ is a standard normal distribution.

Question #12
(a) State the Neyman-Pearson theorem.

(b) Suppose a random variable $X$ has the gamma probability density function $f(x; \theta) = \theta^{-2}xe^{-x/\theta}$ for $x > 0$. Consider the simple null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta < \theta_0$. Let $X_1, X_2, \ldots, X_5$ denote a random sample of size 5 from the distribution.
   (i) Use the Neyman-Pearson theorem to find the most powerful critical region of size $\alpha$.
   (ii) Find the constant in (i) by taking $\theta_0 = 2$ and $\alpha = 0.05$. 


Question #13
(a) Explain each of the following and give an example to illustrate your explanation.
   (i) Sufficient statistic
   (ii) Ancillary statistic
   (iii) Exponential family of probability density functions
(b) Suppose a random sample of size \( n \) is taken from generalized negative binomial distribution (GNBD) with the probability mass function

\[
f(x; \theta) = \frac{m}{m+2x} \left( \frac{m+2x}{x} \right) \theta^x (1-\theta)^{m+x}, \quad \text{for } x = 0, 1, 2, 3, \ldots,
\]

where \( 0 < \theta < 0.5 \) and \( m > 0 \). The population mean for the distribution is \( \mu = m\theta (1-2\theta)^{-1} \) and the population variance is \( \sigma^2 = m\theta (1-\theta)(1-2\theta)^{-3} \).

   (i) If the parameter \( m \) is known, obtain a sufficient statistic for parameter \( \theta \).
   (ii) If the parameter \( m \) is known, find the maximum likelihood estimator of \( \theta \).
   (iii) If the parameter \( m \) is known, find the moment estimator of \( \theta \).
   (iv) Determine the range of values of \( \theta \) for which the variance \( \sigma^2 \) is greater than the mean \( \mu \). Based on your answer, is \( \sigma^2 \) always greater than \( \mu \) for GNBD? Explain.

Question #14
(a) Let \( X_1, X_2, \ldots, X_n \) be a random sample from a Poisson distribution \( f(x; \theta) = \frac{e^{-\theta} \theta^x}{x!} \), for \( x = 0, 1, 2, \ldots \).

   (i) Show that the likelihood ratio test of \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta \neq \theta_0 \) is based upon the statistic \( Y = \sum_{i=1}^{n} X_i \). Show that the rejection region is of the form \( Y \leq c_1 \) or \( Y \geq c_2 \).
   (ii) Obtain the null distribution of \( Y = \sum_{i=1}^{n} X_i \).
   (iii) For \( \theta_0 = 0.5 \) and \( n = 50 \), find the approximate significance level of the test that rejects \( H_0 \) if \( Y \leq 15 \) or \( Y \geq 35 \). Use the Central Limit Theorem and ignore continuity correction.
   (iv) Use the rejection region in (iii) to find the power of the test when \( \theta = 0.32 \) and \( n = 50 \).

(b) Let \( Y_1 < Y_2 < \ldots < Y_5 \) be the order statistics of a random sample of size \( n = 5 \) from a distribution with probability density function \( f(x; \theta) = 1/\theta, \quad 0 < x < \theta \), zero elsewhere, where \( \theta > 0 \). The hypothesis \( H_0 : \theta = 1 \) is rejected and \( H_1 : \theta > 1 \) is accepted if the observed \( Y_3 \geq c \).

   (i) Find the constant \( c \) so that the significance level \( \alpha = 0.05 \)
   (ii) Determine the power function of the test.
   (iii) Can you compute the power at \( \theta = 0.5 \)? Why or why not?