MATH 562-01 MATHEMATICAL STATISTICS

Final Exam (Due by 12:00 pm, 12/05/2016, Monday) Name: ______________________

1. (10 points each) Let Z_1, Z_2 and Z_3 be independent normal random variables, each with mean zero and variance one. Find the distribution of the following variables, and justify your answers.

   (1) $\dfrac{Z_1 + Z_2 - Z_3}{2}$ (2). $\dfrac{(Z_1 + Z_2)^2}{2Z_3{}^2}$

2. (10 points each) Let Z_1, Z_2, Z_3, Z_4 be a random sample from $N(0,1)$, and X_1, X_2, X_3, X_4 a random sample from $N(2,1)$. Determine the sampling distributions of the following statistics. Explain why.

   (1). $\dfrac{(X_1 - X_2)^2 + (Z_1 + Z_2)^2 + (X_3 - X_4)^2}{2}$ (2). $\dfrac{\sum_{k=1}^{4}(X_k - \overline{X})^2}{\sum_{k=1}^{4}(Z_k - \overline{Z})^2}$

3. (10 points) Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be independent random samples from a normal distribution with unknown mean μ and variance σ^2. Let

   $$\overline{X} = \frac{1}{m}\sum_{k=1}^{m} X_k,\ \ \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \text{ and } S_x^2 = \frac{1}{m-1}\sum_{k=1}^{m}(X_k - \overline{X})^2.$$

   Find the constant c so that the statistic

   $$T = c\frac{\overline{Y} - \overline{X}}{S_x}$$

   has a Student's t-distribution with $m-1$ degree of freedom.

4. (10 points) Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, 10)$, If $P\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 \le 52.3\right] = 0.05$, find the sample size *n*.

5. (10 points each) Let $X_1, X_2, \cdots, X_{16}$ be a random sample from $N(1,4)$. Find

   (1). $P\left[1.753 < \dfrac{4(\overline{X} - 1)}{S}\right]$ (2). $P\left[S^2 \le 5.95\right]$

6. (10 points) Let $X_1, \cdots, X_6$ and $Y_1, \cdots, Y_8$ be independent random samples from a standard normal distribution, and $V = \dfrac{4}{3}\left[\sum_{i=1}^{6} X_i^2 \Big/ \sum_{j=1}^{8} Y_j^2\right]$. What is the 99th percentile of the distribution of *V*?

7. (10 points) Let $X_1, X_2, \cdots, X_n$ be a random sample from a population with probability density function

$$f(x;p) = \binom{10}{x} p^x (1-p)^{10-x},\ x = 0,1,\cdots,10.$$

where $0 < p < 1$ is an unknown parameter. Find the Fisher Information in the random sample about the parameter p.

8. (10 points) Let $X_1, X_2, \cdots, X_n$ be a random sample from a population with probability density function

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!},\quad x = 0,1,2,\cdots$$

where $\lambda > 0$ is an unknown parameter. Find the maximum likelihood estimator $\widetilde{\lambda}$, and determine if $\widetilde{\lambda}$ is efficient (i.e., a UMVUE).

9. (10 points) Let $X_1, X_2, \cdots, X_n$ be a random sample from a population with probability density function

$$f(x;\theta) = \frac{\theta \alpha^\theta}{x^{(\theta+1)}},\quad x > \alpha,$$

where $\alpha > 0$ is given and $\theta > 0$ is an unknown parameter. Find a sufficient statistic for θ.

10. (10 points) A random sample of size n is drawn from a normal population $N(\mu_1, \sigma_1^2)$, and another random sample of the same size is drawn independently from another normal population $N(\mu_2, \sigma_2^2)$. Find the MLE $\widetilde{\theta}$ for $\theta = \mu_1 - \mu_2$. If the variances σ_1^2 and σ_2^2 are assumed to be known, is $\widetilde{\theta}$ an efficient estimator?

11. (10 points) A sequence of independent Bernoulli trials with probability of success p is performed. Let X be the number of trials until the first success occurs. Four independent realizations of X are $x_1 = 1$, $x_2 = 1$, $x_3 = 2$ and $x_4 = 1$. If, a priori, p is uniformly distributed on $(0,1)$ and squared error loss is used, find the Bayesian estimate for p, i.e., $E[p|x_1, x_2, x_3, x_4]$.

12. (10 points) Let $X_1, \cdots, X_9$ be a random sample from a normal distribution with mean μ and variance σ^2, both unknown. Find a 90% confidence interval for μ.

13. (10 points) Let $X_1, X_2, \cdots, X_{21}$ be a random sample from a normal distribution with unknown mean μ and variance 1. Then $(L,U) = \left(\overline{X} - \frac{1.96}{\sqrt{21}}, \overline{X} + \frac{1.96}{\sqrt{21}}\right)$ is a 95% confidence interval for μ. From a particular random sample, we observed $\bar{x}$, and found that $L = 10.54$ and $U = 11.40$. Which of the following interpretations of our finding are true? Explain why. (No credit without explanation.)

    (1). The probability that μ will assume a value between 10.54 and 11.40 is 0.95.

(2). If we were to repeat this entire sampling and interval computation process 100,000 times independently, we would expect 95,000 of the resulting intervals containing the true value of μ.

(3). If we were to collect one additional independent observation from this normal population, the probability that this new observation would fall between 10.54 and 11.40 would be 0.95.

14. (10 points) Let X have density function $f(x) = (\theta+1)x^{\theta}$, $0 < x < 1$, and zero otherwise. The hypothesis $H_0 : \theta = 1$ is to be rejected in favor of $H_a : \theta = 2$ if $X > 0.90$. What is the probability of Type I error? What is the power of the test.

15. (10 points) Let $\overline{X}$ be the sample mean of a random sample from a normal distribution with variance 9. The hypothesis $H_0 : \mu = 100$ is rejected in favor of $H_a : \mu = 101$ if $\overline{X} > c$, where c is a constant. If the size of the test is required to be 0.05, find the minimum sample size necessary to achieve 0.5 as the power of the test.

16. (10 points) It is hypothesized that of all marathon runners, 70% are adult men, 25% are adult women, and 5% are youths. To test this hypothesis, the following data from a recent marathon are used:

Adult men	Adult women	Youths	Total
630	300	70	1,000

A chi-square goodness-of-fit test is used at $\alpha = 0.05$. What is the value of the test statistic? What would be the conclusion?

17. (10 points) Let $Y_1, Y_2, \cdots, Y_n$ be random variables such that $E[Y_i] = 2 + \beta x_i$, $i = 1,2,\cdots,n$, where $x_1, x_2, \cdots, x_n$ can be observed and β is an unknown parameter. If a random sample gives $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, the **least squares estimate of $\boldsymbol{\beta}$** is defined as $\hat{\beta}$ such that

$$S(\beta) = \sum_{i=1}^{n} (y_i - 2 - \beta x_i)^2$$

is minimized. Find the $\hat{\beta}$.

18. (10 points) A single observation is taken from a Cauchy distribution with density function $f(x) = \dfrac{1}{\pi\left[1 + (x-\theta)^2\right]}$. For testing $H_0 : \theta = 0$ versus $H_a : \theta \neq 0$ at the 0.05 significant level using the generalized likelihood ratio test, find the critical region