STA 413F2011 Assignment 7

Please read Sections 5.1, 5.5 and 5.6. Then do this assignment in preparation for Quiz Seven in tutorial on Friday Oct. 28th. The problems are practice for the quiz, and are not to be handed in. Any necessary material from the formula sheet will be supplied with the quiz. Several questions refer to the "Modified Central Limit Theorem." This refers to $\frac{\sqrt{n}(\overline{X}_n-\mu)}{\widehat{\sigma}_n} \stackrel{d}{\rightarrow} Z \sim N(0,1)$.

1. $X_1,\ldots,X_n$ be a random sample. For each of the distributions below, give the parameter space Ω.
   (a) Bernoulli
   (b) Binomial(m,θ) with m known
   (c) Binomial(m,θ) with m unknown
   (d) Poisson
   (e) Geometric
   (f) Uniform(α,β)
   (g) Exponential
   (h) Gamma
   (i) Normal

2. Let $X_1,\ldots,X_n$ be a random sample from a distribution (not necessarily normal) with expected value μ and variance σ^2. We will test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ using the critical region
$$C = \{(x_1,\ldots,x_n) : \left|\frac{\overline{x}-\mu_0}{s/\sqrt{n}}\right| > z_{\alpha/2}\}$$
   (a) What is the approximate size of the test? Just give the answer; you don't have to prove anything.
   (b) Show that H_0 is rejected if and only if the $(1-\alpha)100\%$ confidence interval for μ does *not* include μ_0. It is easiest to start by writing the set of $(x_1,\ldots,x_n)$ such that μ_0 is in the confidence interval, and then work on it until it becomes C-complement.

3. Let $X_1,\ldots,X_n$ be a random sample from a normal distribution with expected value μ and variance σ^2. We wish to test $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$ using the critical region
$$C = \{(x_1,\ldots,x_n) : \frac{(n-1)S^2}{\sigma_0^2} > k\}$$
   (a) What is the parameter vector $\boldsymbol{\theta}$?
   (b) What is the parameter space Ω?
   (c) What is ω_0? Is it simple or composite?
   (d) What is ω_1? Is it simple or composite?
   (e) Find k so that the size of the test is *exactly* α. Recall that for random sampling from a normal distribution, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. Make up your own notation for the critical value, and then draw a picture to illustrate what it means.

(f) Show that this same test is also size α for testing $H_0 : \sigma^2 \leq \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$.

(g) Write the function $P_{\boldsymbol{\theta}}(\mathbf{X} \in C)$ for this test (a function of $\boldsymbol{\theta}$ only through σ^2) in terms of the cumulative distribution function of a chi-square random variable.

4. Let $X_1, \ldots, X_{n_1}$ be a random sample from an exponential distribution with parameter θ. The null hypothesis is $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$.

   (a) Give the critical region C for an approximate size α test based on the modified Central Limit Theorem. Your answer will use a critical value from the standard normal distribution.

   (b) Write $P_\theta(\mathbf{X} \in C)$ for this test explicitly as a function of the true parameter θ. You answer will involve Φ, the cumulative distribution function of a standard normal.

   (c) Suppose $\theta_0 = 2$, $\alpha = 0.05$ and $n = 30$. What is the approximate power for $\theta = 2.5$? The answer is a single number. I get a power of 0.4129.

   (d) What is the smallest sample size that will guarantee an approximate power of 0.80 for $\theta = 2.5$? The idea here is to set $P_\theta(\mathbf{X} \in C)$ to 0.80 an then solve for n. The answer is a single number (an integer, of course). I get $n = 117$.

   (e) Again for $\theta_0 = 2$, $\alpha = 0.05$ and $n = 30$, suppose we observe $\overline{X}_n = 3$.

      i. What is the value of the test statistic? The answer is a number.

      ii. Is $\mathbf{X} \in C$? Answer Yes or No.

      iii. Do you reject H_0? Answer Yes or No.

      iv. Calculate the p-value. The answer is a number. I get 0.0031.

5. Here is another test for the exponential model of Question 4. Let

$$C = \{\mathbf{x} : \sqrt{n}(\ln(\overline{X}_n) - \ln(\theta_0)) > k\}$$

   (a) Find the value of k so that this test has approximately size α. Your answer will involve a critical value of the standard normal distribution. Start by finding the limiting distribution of the test statistic under the null hypothesis.

   (b) Express the function $P_\theta\{\mathbf{X} \in C\}$ in terms of the cumulative distribution function of a standard normal random variable.

   (c) Is the power function strictly increasing in θ? Answer Yes or No and prove your answer.

   (d) Suppose the null hypothesis were $H_0 : \theta \leq \theta_0$. Is this test still size α? Answer Yes or No and justify your answer.

   (e) Suppose $\theta_0 = 2$, $\alpha = 0.05$ and $n = 30$. What is the approximate power for $\theta = 2.5$? The answer is a single number. I get 0.336.

6. Here is yet another test for the exponential model of Question 4. Let

$$C = \{\mathbf{x} : \frac{2n\overline{X}_n}{\theta_0} > k\}$$

   (a) Find the value of k so that this test has *exactly* size α. Your answer will involve the chi-square distribution. Start by using moment-generating functions to find the distribution of the test statistic under the null hypothesis.

(b) Express the function $P_\theta\{\mathbf{X} \in C\}$ in terms of the cumulative distribution function of a chi-square random variable.

(c) Is the function $P_\theta\{\mathbf{X} \in C\}$ strictly increasing in θ? Answer Yes or No and prove your answer.

(d) For what values of θ is $P_\theta\{\mathbf{X} \in C\}$ the power function?

(e) Suppose the null hypothesis were $H_0 : \theta \leq \theta_0$. Is this test still size α? Answer Yes or No and justify your answer.

(f) This last item is just a comment. For $\theta_0 = 2$, $\alpha = 0.05$ and $n = 30$, the generic large-sample test had a power of approximately 0.4129 at $\theta = 2.5$, while the variance-stabilized test had a power of approximately 0.336. The exact test has a power of 0.362.

7. Let $X_1, \ldots, X_n$ be a random sample from a Gamma distribution with parameters α (unknown) and $\beta = 1$ (known). Using just the Modified Central Limit Theorem (no variance-stabilizing transformations, please!), give an approximate critical region of size 0.05 for the following null and alternative hypotheses. Please use specific numbers for your critical values, not symbols. You don't have to prove anything or justify your answers. But you do have to be aware of the justification in order to get the right answer.

(a) $H_0 : \alpha = \alpha_0$ versus $H_1 : \alpha \neq \alpha_0$.

(b) $H_0 : \alpha = \alpha_0$ versus $H_1 : \alpha > \alpha_0$.

(c) $H_0 : \alpha \leq \alpha_0$ versus $H_1 : \alpha > \alpha_0$.

(d) $H_0 : \alpha = \alpha_0$ versus $H_1 : \alpha < \alpha_0$.

(e) $H_0 : \alpha \geq \alpha_0$ versus $H_1 : \alpha < \alpha_0$.

8. For Question 7, suppose $n = 150$, $\alpha_0 = 7.5$ and you observe a sample mean of $\overline{X}_n = 8.2$. Test $H_0 : \alpha \leq \alpha_0$ versus $H_1 : \alpha > \alpha_0$.

(a) What is the value of the test statistic? The answer is a single number.

(b) What is the p-value? The answer is a single number.

(c) Do you reject H_0? Answer Yes or No.

(d) Is $p < \alpha$? Answer Yes or No.

9. Let $Y_i = x_i + \epsilon_i$, for $i = 1, \ldots, n$, where

- $x_1, \ldots, x_n$ are fixed, known constants
- $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed Normal(0,σ^2) random variables; the parameter σ^2 is unknown.
- The data consist of n pairs (x_i, Y_i). Of course the error terms ϵ_i are not observable.

(a) Find the distribution of $\frac{1}{\sigma^2}\sum_{i=1}^n (Y_i - x_i)^2$. Show your work.

(b) We wish to test $H_0 : \sigma^2 \leq \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$.

i. What is Ω?

ii. What is ω_0? Is it simple or composite?

iii. What is ω_1? Is it simple or composite?

(c) Find the constant k so that the following test will be size α for testing the *simple* null hypothesis $H_0 : \sigma^2 = \sigma_0^2$.

$$C = \{\mathbf{y} : \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sigma_0^2} > k\}$$

(d) Find the power function $P_{\sigma^2}\{\mathbf{X} \in C\} = \gamma(\sigma^2)$.

(e) Prove that the test C is also size α for testing $H_0 : \sigma^2 \leq \sigma_0^2$.

10. Let $X_1, \ldots, X_{n_1}$ be a random sample from a distribution (not necessarily normal) with expected value μ_1 and variance σ_1^2, and let $Y_1, \ldots, Y_{n_2}$ be a random sample from a distribution (not necessarily normal) with expected value μ_2 and variance σ_2^2. The random samples are independent of each other. The Central Limit Theorem tells us that for large n_1, the distribution of $\overline{X}_{n_1}$ is approximately $N(\mu_1, \frac{\sigma_1^2}{n_1})$. Similarly, the distribution of $\overline{Y}_{n_2}$ is approximately $N(\mu_2, \frac{\sigma_2^2}{n_2})$. So, what should[1] the approximate distribution of $\overline{X}_{n_1} - \overline{Y}_{n_2}$ be?

11. Let $X_1, \ldots, X_{n_1}$ be a random sample from a (possibly) non-normal distribution with mean μ_1 and variance σ_1^2), and let $Y_1, \ldots, Y_{n_2}$ be a random sample from a (possibly) non-normal distribution with mean μ_2 and variance σ_2^2). These are *independent* random samples, meaning that the data are independent between samples as well as within samples. We are interested in testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. Find the constant k so that the following critical region will have, for large n_1 and large n_2, a size of approximately α.

$$C = \left\{ (\mathbf{x}, \mathbf{y}) : \left| \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_2^2}{n_2}}} \right| > k \right\},$$

where $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ are consistent estimators of σ_1^2 and σ_2^2, respectively. There is almost no work to show.

12. Of a random sample of 150 Special Needs students in the Toronto District School Board, 19 were in regular classes, and the rest were in Special Education classes. Of a random sample of 200 Special Needs students in the Toronto Separate School Board, 48 were in regular classes, and the rest were in Special Education classes. Test for difference between the proportions of Special Needs students in regular classes in the two school boards. Use $\alpha = 0.01$. What do you conclude? Is the proportion of Special Needs students greater in one of the school boards? Of course you should use the test from the last question.

[1] You are being asked for an educated guess, not a proof.