University of California, Los Angeles
Department of Statistics

Statistics 100B **Instructor: Nicolas Christou**

Final exam
20 March 2020

Problem 1 (16 points)
Answer the following questions:

a. Let X and Y be independent exponential random variables with parameters θ_x and θ_y, respectively. Find the likelihood ratio test statistic based on X and Y, to test H_0: $\theta_x{=}2\theta_y$ against H_a: $\theta_x{\neq}2\theta_y$.

b. Suppose that independent random samples of size n from two normal populations with known variances σ_1^2 and σ_2^2 are to be used to test the null hypothesis $H_0 : \mu_1 - \mu_2 = \delta$ against the alternative hypothesis $H_a : \mu_1 - \mu_2 > \delta$, and that the probabilities of Type I and Type II errors are to have preassigned values α and β. Find an expression of the required sample size n in order to detect a shift from δ to δ', where $\delta' > \delta$. Please show all your work and draw any necessary graphs to support your answer.

c. Consider the regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, for $i = 1, \ldots, n$, with $E(\epsilon_i) = 0$, $var(\epsilon_i) = \sigma^2$, $cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$ because ϵ_i, ϵ_j are independent, and $\epsilon_i \sim N(0, \sigma)$. In class we developed the test statistic for testing $H_0 : \beta_1 = 0$ against $H_a : \beta_1 \neq 0$. What does this test mean? The test statistic we found was distributed at central t_{n-2} under H_0. Show that it is also distributed as central $F_{1,n-2}$ under H_0. Now suppose H_0 is not true. Find the mean and variance of this test statistic.

d. Suppose we have a single observation X from an exponential distribution with parameter $\frac{1}{\lambda}$. Use X to construct a confidence interval for λ using 90% confidence level.

Problem 2 (16 points)
Answer the following questions:

a. Answer TRUE/FALSE:

1. If the p-value for a test is 0.036, the null hypothesis can be rejected at the $\alpha = 0.05$ level of significance.

2. In a formal test of hypothesis, α is the probability that the null hypothesis is incorrect.

3. If the p-value is very small for a test to compare two population means, the difference between the means must be large.

4. Power is always computed by assuming that the null hypothesis is true.

5. If $0.01 < p-\text{value} < 0.025$, the null hypothesis can always be rejected at the $\alpha = 0.02$ level of significance.

6. When developing a likelihood ratio test, it is possible that $L(\hat{\omega}) > L(\hat{\Omega})$.

7. The non centrality parameter for the non central t distribution can be negative.

8. The unbiased estimator of σ^2 in simple regression with an intercept follows $\Gamma(\frac{n-2}{2}, \frac{2\sigma^2}{n-2})$. Assume the assumptions hold and normally distributed error terms.

b. For each of the following questions please answer: *could happen*, *impossible*, or *certainly.*

1. Based on a sample of $n = 131$ yes-or-no values, the 95% confidence interval for the population proportion p was found to be 0.711 ± 0.078. Also, with the same data the null hypothesis $H_0 : p = 0.785$ (tested against the alternative $H_a : p \neq 0.785$) was rejected at the 5% level of significance.

2. The hypothesis $H_0 : \mu = 1.8$ was tested against the alternative $H_a : \mu \neq 1.8$ at the 0.05 level of significance, using a sample size of $n = 85$. Unknown to the statistical analyst, the true value of μ was 1.74 and yet H_0 was accepted.

3. A hypothesis test was done at the 5% level of significance, and H_0 was not rejected. With the same data using the 1% level of significance, H_0 was not rejected.

4. A hypothesis test was done at the 5% level of significance, and H_0 was rejected. With the same data, but using the 1% level of significance, H_0 was not rejected.

5. A hypothesis test was done at the 5% level of significance, and H_0 was not rejected. With the same data, but using the 1% level of significance, H_0 was rejected.

6. A sample of size $n = 36$ was taken from a population that has a known population mean $\mu = 3.5$ and standard deviation $\sigma = 1.71$ yet the 95% confidence interval that was constructed failed to cover 3.5.

c. Using a sample size of $n = 250$ healthy adults a 95% confidence interval for the mean systolic blood pressure was found to be between 110 `mmHg` and 125 `mmHg`. Which one(s) of the following statements are false? Explain your answer!

1. If we select one hundred samples each one of size $n = 250$ from this population and compute a 95% confidence interval, in about 95 out of the 100 cases the confidence interval will be between 110 `mmHg` and 125 `mmHg`.

2. There is 95% probability that the mean systolic blood pressure will be between 110 `mmHg` and 125 `mmHg`.

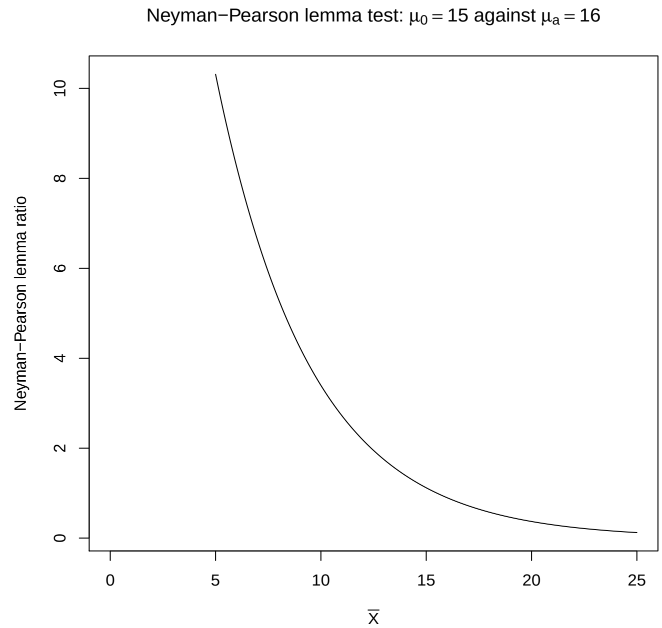
3. About 95% of healthy adults have systolic blood pressure between 110 `mmHg` and 125 `mmHg`.

d. The effectiveness of a new drug that lowers the blood cholesterol level will be tested against a placebo. The placebo will be administered to $n_1 = 25$ patients while the drug will be administered to $n_2 = 20$ patients. The hypothesis we want to test is
$H_0 : \mu_1 - \mu_2 = 13$
$H_a : \mu_1 - \mu_2 \neq 13$
Assume that $\sigma_1^2 = 50, \sigma_2^2 = 60$. Find the probability of detecting a shift from $\mu_1 - \mu_2 = 13$ to $\mu_1 - \mu_2 = 15$ if we are willing to accept a Type I error $\alpha = 0.05$. Draw the supporting graphs to show Type I and Type II errors.

e. Let $X_1, \ldots, X_n$ be i.i.d. random variables from $N(\mu_1, \sigma)$ and $Y_1, \ldots, X_m$ be i.i.d. random variables from $N(\mu_2, \sigma)$ with σ unknown. In testing the hypothesis $H_0 : \mu_1 = \mu_2$ we use the t statistic with $n + m - 2$ degrees of freedom. If H_0 is true the t statistic follows the central t distribution, while if H_0 is not true the t statistic follows the non-central t distribution with non-centrality parameter δ. Determine δ here and find the mean and variance of the non central distribution for this hypothesis test. Note: Assume $H_a : \mu_1 > \mu_2$.

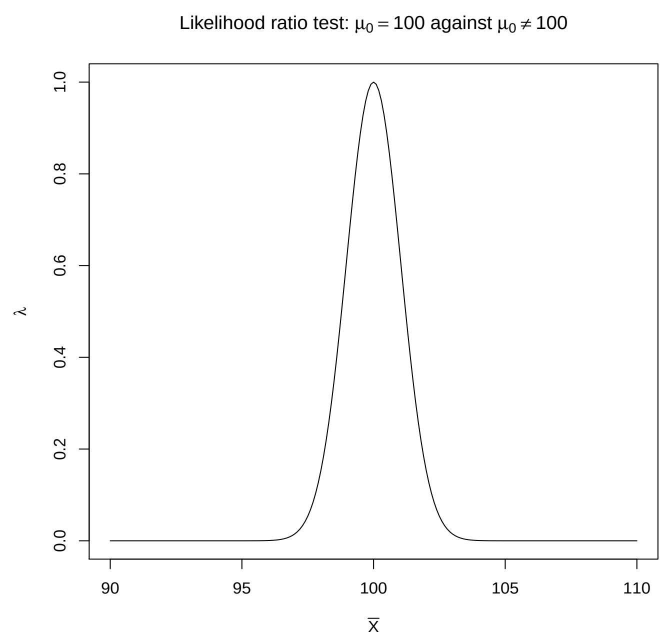
Problem 3 (16 points)
Answer the following questions:

a. Suppose X_1, X_2, X_3 are i.i.d. $N(\mu, \sqrt{13.5})$. We are interested in the following test: $H_0: \mu = 15$ against $\mu = 16$. The following plot is the graph of the Neyman-Pearson lemma ratio against values of $\bar{x}$.



Suppose $k = 0.7$. What is the exact significance level α? Note: You must find the exact value of $\bar{x}$ first that corresponds to the critical region for the given k.

b. Suppose $X_1, \ldots, X_{15}$ are i.i.d. $N(\mu, 4)$. We are interested in the following test: $H_0: \mu = 100$ against $\mu \neq 100$. The following plot is the graph of the likelihood ratio test against values of $\bar{x}$.



Suppose $k = 0.2$. What is the rejection region and the significance level α? Note: You must find the exact values of $\bar{x}$ first that correspond to the critical region for the given k.

c. Let $X_1, \ldots, X_n$ be i.i.d. random variables from normal distribution with mean μ_1 and variance σ_1^2 and $Y_1, \ldots, Y_m$ be i.i.d. random variables from normal distribution with mean μ_2 and variance σ_2^2. Find the likelihood ratio test for testing $H_0 : \mu_1 = \mu_2, \sigma_1^2 = \sigma_2^2$ against all alternatives. The two samples are independent.

d. Let $X_1, \ldots, X_{10}$ be i.i.d. Bernoulli(p) random variables. Find the most powerful test of size $\alpha = 0.0547$ for testing $H_0 : p = \frac{1}{2}$ against $H_a : p = \frac{1}{4}$. Find the power of this test.

e. Suppose that $X_1, \ldots, X_n$ are i.i.d. exponential random variables with $EX_i = \beta t_i$ where $t_1, \ldots, t_n$ are known constants, and $\beta > 0$ is an unknown parameter. It can be shown the MLE of β is given by $\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{t_i}$. (See also exam 2, problem 2c.) Suppose we want to test $H_0 : \beta = 1$ against $H_a : \beta \neq 1$. Find a test statistic given by the likelihood ratio test.

Problem 4 (16 points)
Answer the following questions:

a. Suppose $X_1, \ldots, X_n$ is a random sample from $N(\mu_1, \sigma)$ and $Y_1, \ldots, Y_m$ is a random sample from $N(\mu_2, \sigma)$. Assume σ^2 is unknown. Construct the likelihood ratio test for testing $H_0 : \mu_1 - \mu_2 = 0$ against $H_0 : \mu_1 - \mu_2 \neq 0$ and show that the test statistic follows t_{n+m-2}.

b. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from $U(\theta, \theta + 1)$. In testing the hypothesis $H_0 : \theta = 0$ against the alternative $H_a : \theta > 0$ we use the test: reject H_0 if $X_{(n)} > 1$ or $X_{(1)} > c$. Find c so that the test will have significance level α. Find an expression for the power function for this test.

c. Let $Y = ln(Z)$ with $Y \sim N(\mu, \sigma)$. Suppose the distribution of a predictor of a new value of Y follows a normal distribution, i.e. $\hat{Y}_0 \sim N(\mu, \sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}})$ where $\mathbf{a}$ is a vector of constants and $\mathbf{\Sigma}$ is the variance covariance matrix of $\mathbf{Y}$. (We assume that $\hat{Y}_0 = \mathbf{a}'\mathbf{Y}$, a linear combination of $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$.) In addition suppose $\hat{Y}_0 = ln(\hat{Z}_0)$. Find the unbiased predictor of Z_0.

d. Suppose $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))'$ represent ozone observations in ppm at spatial locations $s_1, \ldots, s_n$. Assume that $\mathbf{Y} \sim N_n(\mu\mathbf{1}, \sigma^2\mathbf{V})$, where $\mathbf{V}$ is matrix of known constants that can be constructed using a covariance function, for example, $c(h) = c_1 e^{-\frac{h}{\alpha}}$, where c_1, α are known parameters, and h is the Euclidean distance for the spatial locations $s_1, \ldots, s_n$. Find the Fisher information matrix for the parameters μ and σ^2 for this model.

e. In each of the following situations calculate the p-value of the observed data:

1. For testing $H_0 : \theta = \frac{1}{2}$ against $H_a : \theta > \frac{1}{2}$ when 7 successes are observed out of 10 Bernoulli trials.

2. For testing $H_0 : \lambda = 1$ against $H_a : \lambda > 1$, when $X = 3$ are observed, where $X \sim$ Poisson(λ).

3. For testing $H_0 : \lambda = 1$ against $H_a : \lambda > 1$, when $X_1 = 3, X_2 = 5$, and $X_3 = 1$ are observed, where $X \sim$ Poisson(λ).

Problem 5 (16 points)
Answer the following questions:

a. A method for finding confidence intervals for a discrete probability distribution is called "pivoting a discrete cdf," and it begins with a discrete test statistic T such that $P(L \leq T \leq U) = 1 - \alpha$, where L and U are given by $P(T \leq L) = \frac{\alpha}{2}$ and $P(T \geq U) = \frac{\alpha}{2}$. Let $X_1, \ldots, X_n$ be i.i.d. Poisson(λ) and consider the test statistic $Y = \sum_{i=1}^{n} X_i$. Suppose $Y = y_0$ is observed. Use the fact that if $Q \sim \Gamma(\alpha, \beta)$ (α is an integer) then $P(Q \leq q) = P(Y \geq \alpha)$, where $Y \sim$ Poisson($\frac{q}{\beta}$), to find a $1 - \alpha$ confidence interval for λ.

b. A confidence interval is unbiased if the expected value of the interval midpoint is equal to the estimated parameter. For example the midpoint of the interval $\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ is $\bar{x}$, and $E(\bar{x}) = \mu$. Now consider the confidence interval for σ^2. Show that the expected value of the midpoint of this confidence interval is not equal to σ^2. Assume that we are working with a random sample $X_1, \ldots X_n$ from $N(\mu, \sigma)$.

c. Suppose $X_1, \ldots, X_{15}$ are i.i.d. $N(\mu_1, \sigma)$ and $Y_1, \ldots, X_{16}$ are i.i.d. $N(\mu_2, \sigma)$. We want to test the hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against the alternative $H_a : \mu_1 - \mu_2 > 0$ using $\alpha = 0.05$. Compute the power of the test if $\mu_1 = 8, \mu_2 = 5$ and $\sigma = 3$. Use your t table to find the relevant critical values and also one of the following probabilities of the non-central t distribution. Please show the supporting graphs and explain all the details. $P[t_{29}(ncp = 2.782) > 1.699] = 0.8581$, $P[t_{29}(ncp = 2.782) < 1.699] = 0.1419$, $P[t_{29}(ncp = 7.7419) > 1.699] \approx 1$, $P[t_{29}(ncp = 0.4997) > 1.699] = 0.1237$, $P[t_{31}(ncp = 2.782) > 1.699] = 0.8582$.

d. Let $X_1, X_2, \ldots, X_n$ be a random sample from normal distribution with $E(X_i) = \mu, var(X_i) = \sigma^2$, and $cov(X_i, X_j) = \rho\sigma^2$, for $i \neq j$. Then $\bar{X} \sim N\left(\mu, \sigma\sqrt{\frac{1+(n-1)\rho}{n}}\right)$, and a 95% confidence interval for μ is

$$\bar{x} \pm 1.96\sigma\sqrt{\frac{1+(n-1)\rho}{n}}. \tag{1}$$

Note: If $\rho = 0$ then we get the usual confidence interval for μ when we are dealing with i.i.d. random variables:

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}. \tag{2}$$

Now, suppose we fail to see the dependence in our random variables, and instead of using (1) we decided to use (2). What is the actual coverage of our confidence interval if $n = 25, \sigma = 3, \rho = 0.2$ and $\bar{x} = 20$?

Problem 6 (20 points)
Answer the following questions:

a. Suppose $X_1, \ldots, X_n$ are i.i.d. with beta$(\theta_1, 1)$ pdf and $Y_1, \ldots, Y_m$ are i.i.d. with beta$(\theta_2, 1)$ pdf. Assume the two samples are independent. Find the likelihood ratio test for testing $H_0 : \theta_1 = \theta_2$ against $\theta_1 \neq \theta_2$.

b. Suppose $X_1, \ldots, X_n$ are i.i.d. random variables with $X \sim \Gamma(\alpha.\beta)$. Suppose α is a known integer. Use the pivot method to find a confidence interval for β using confidence level $1 - \alpha$.

c. Consider the simple regression model without intercept $y_i = \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$. Assume that $\epsilon_1, \ldots, \epsilon_n$ are independent and $\epsilon_i \sim N(0, \sigma)$ and that $x_1, \ldots, x_n$ are not random. Construct a prediction interval for Y_0 for a given new x_0.

d. Refer to question (c). Find a confidence interval for the expected value of Y_0 for a given new x_0.

e. Refer to question (c). Develop the likelihood ratio test for testing $H_0 : \beta_1 = 0$ against $H_a : \beta_1 \neq 0$ and show that it follows the F distribution.