Statistics for Financial Engineering - Refresher Seminar, Homework 2

Radoš Radoičić

July 27, 2017

Homework has two types of problems: theoretical (T) and practical (P).

You need to show all your work. It is due on 08/01/17.

Problem T1: "Exponential" Suppose that $X_1, \ldots, X_n$ are i.i.d. observations from the exponential distribution $E(\alpha)$. Consider the following two statistics: $T = \sum_{i=1}^n X_i$, and $S = 2\alpha T$. Show that $T \sim \Gamma(n, \alpha)$ and that $S \sim \Gamma(n, \frac{1}{2}) \sim \chi^2_{2n}$. Construct a 95% confidence interval for the mean $1/\alpha$.

Problem T2: "Normal Hypothesis"

1. Suppose that a sample $X_1, X_2, \ldots, X_{15} \sim N(\mu_X, \sigma_X^2)$, where μ_X and σ_X^2 are unknown, has sample mean and sample variance
$$\hat{\mu}_X = \overline{X} = 2.4, \hat{\sigma}_X^2 = \overline{X^2} - (\overline{X})^2 = 0.55.$$
Find 95% confidence intervals for μ_X and σ_X^2.

2. In addition to the sample $X_1, \ldots, X_{15}$ above, suppose that we are given a sample $Y_1, \ldots, Y_{10} \sim N(\mu_Y, \sigma_Y^2)$ from a distribution with the same variance $\sigma_Y^2 = \sigma_X^2$, but possibly different mean μ_Y. Suppose that
$$\hat{\mu}_Y = \overline{Y} = 2.8, \hat{\sigma}_Y^2 = \overline{Y^2} - (\overline{Y})^2 = 0.37.$$
Find a 90% confidence interval for $\mu_X - \mu_Y$.

3. For samples above, perform the t-test of the hypothesis $H_0 : \mu_X \leq \mu_Y$ (under the assumption of equal variances). Test whether the variances are equal using the F-test. In both tests, find the p-value and use level of significance $\alpha = 0.05$.

4. Suppose we have two normal samples $X = (X_1, X_2, \ldots, X_n)$ and $Y = (Y_1, Y_2, \ldots, Y_n)$. You want to test $H_0 : \mu_X = \mu_Y$ for the means of the two samples. As usual, let $\hat{\sigma}_{X,Y}$ be the sample covariance between X and Y, i.e.
$$\hat{\sigma}_{X,Y} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}).$$
Also, form $Z = (Z_1, Z_2, \ldots, Z_n)$ by $Z_i = X_i - Y_i$.

(A) Show that $\hat{\sigma^2}_Z = \hat{\sigma^2}_X + \hat{\sigma^2}_Y - 2\hat{\sigma}_{X,Y}$.

(B) If X and Y were paired, you would use the t-statistics for Z from Lecture Notes Part 1 (i.e. $T_Z = \sqrt{n-1} \cdot \frac{\hat{\mu}_Z}{\hat{\sigma}_Z}$) to test H_0 (here $\mu_0 = 0$). If X and Y were not paired (i.e. independent), but you assumed the equal variances, you would use the t-statistic $T_{X,Y}$ from Lecture Notes Part 1. Compare the two statistics and discuss how this affects the power of the two tests (one based on paired samples, one based on independent samples).

Problem T3: "Poisson" Suppose that $X_1, \ldots, X_n$ form a random sample from a Poisson distribution for which the value of the mean λ is unknown. Let λ_1 and λ_2 be specified values such that $\lambda_2 > \lambda_1 > 0$ and suppose that it is desired to test the following simple hypotheses:

$$H_1 : \lambda = \lambda_1 \; H_2 : \lambda = \lambda_2.$$

Show that the value of $\alpha_1(\delta) + \alpha_2(\delta)$ is minimized by a test which rejects H_1 when $\overline{X} > c$. Find the value of c. For $\lambda_1 = 1/4$, $\lambda_2 = 1/2$, and $n = 20$, determine the minimum value of $\alpha_1(\delta) + \alpha_2(\delta)$ that can be attained.

Problem T4: "Goodness of height" Suppose that the distribution of the heights of men who reside in a certain large city is a normal distribution for which the mean is 68 inches and the standard deviation is 1 inch. Suppose also that when the heights of 500 men who reside in a certain neighborhood of the city were measured, the following distribution was obtained: number of men of height less than 66 in. is 18, between 66 and 67.5 in. is 177, between 67.5 and 68.5 in. is 198, between 68.5 and 70 in. is 102, and greater than 70 in. is 5. Test the hypothesis that with regard to height these 500 men form a random sample from all the men who reside in the city. (hint: χ^2 goodness-of-fit)

Problem T5: "NBA" At the fifth NBA game of the season at a certain arena, 200 people were selected at random and asked how many of the previous four games they had attended. 33 said 0, 67 said 1, 66 said 2, 15 said 3, and 19 said 4. Test the hypothesis that these 200 observed values can be regarded as a random sample from a binomial distribution; that is, there exists a number θ, $0 < \theta < 1$, such that the probabilities are as follows: $p_0 = (1-\theta)^4, p_1 = 4\theta(1-\theta)^3, \ldots, p_4 = \theta^4$. ($\chi^2$ goodness-of-fit for composite hypotheses).

Problem P1: Chapter 7 R-lab.