CS5960: Nonparametric Methods **Due Wed. 4/8**

Homework 6: Hypothesis Testing

Instructions: Be sure to electronically submit your answers in pdf format for the written part and as an R file for the coding part. You may work together and discuss the problems with your classmates, but write up your final answers entirely on your own. **Parameters:** In all simulations below use 1000 repititions, 1000 permutations, and a sample size of $n = 10$ for all variables.

Permutation tests of two means. Let X and Y be two Gaussian random variables with means μ_X, μ_Y and both with variance $\sigma^2 = 1$. Consider the hypothesis test

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

1. Simulate the null hypothesis H_0. Compare the Type I error rate for the parametric t-test and a permutation test of the t statistic.

2. Simulate the alternative hypothesis with $\mu_X = 0.5$ and $\mu_Y = -0.5$. Compare the Type II error rate of the parametric and permutation tests.

3. Now let X be distributed according to a non-central t-distribution with 2 degrees-of-freedom and non-centrality parameter of 2. Let $Y \sim -X$, i.e., Y is distributed as the negation of the distribution of X. The distribution X has mean $\mu_X = 2\sqrt{\pi}$. Simulate the following:

   - $H_0 : \mu_X = \mu_Y$, simulate $X - 2\sqrt{\pi}$ and $Y + 2\sqrt{\pi}$
   - $H_1 : \mu_X \neq \mu_Y$, simulate X and Y

   Repeat the same experiments to test the Type I and II error rates for the parametric t and permutation tests in this case. (**Important:** X and Y are still independent! Don't simulate X and then negate those numbers for Y.)

4. Discuss the results you got in Problems 1-3. Which test was better for each scenario? What two properties do the distributions in Problem 3 have that make them different from Gaussians? How does this affect the performance of the parametric and permutation tests?

Permutation tests of correlation. Let X and Y be two Gaussian random variables with zero mean and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}.$$

Let $\rho = \sigma_{XY}/(\sigma_X \sigma_Y)$ be the correlation between X and Y, and consider the hypothesis test

$$H_0 : \rho = 0$$
$$H_1 : \rho > 0$$

5. Look up "Fisher's Transformation" and describe how it can be used to test this hypothesis. What is the distribution of ρ under the null hypothesis after applying this transformation?

6. Describe how you could use a permutation test to test this hypothesis. Explain why your test is valid.

7. Simulate the null distribution with Σ equal to the identity matrix. Compare the Type I error rates for the parametric and permutation tests.

8. Simulate the alternative hypothesis with $\sigma_X^2 = \sigma_Y^2 = 1$ and $\sigma_{XY} = 0.8$. Compare the Type II error rate for the parametric and permutation tests.