PSETs Landing Page*

Anish Krishna Lakkapragada

This is the documentation for using my PSET PDFs responsibly. I post these LaTEX'd PSETs (1) as an education resource for friends at other universities, fellow Yalies, and all those interested and (2) for quick reference. These PSETs are not to be used irresponsibly; only look at the solution after giving each problem an honest attempt. **If YOU USE THESE PSETS TO CHEAT, YOU ARE NOT ONLY STUPID BUT YOU ARE CHEATING YOURSELF OUT OF THE ABILITY TO FALL IN LOVE WITH MATH.** Furthermore, I am not smarter than you and my solutions did not always get a perfect score.

The general format for accessing the (one-indexed) `N`th assigned PSET PDF of a Yale course with course number `CODE` is:

`https://anish.lakkapragada.com/notes/TYPE-CODE/psets/N.pdf`

where `TYPE` is `stats` or `math`. Similarily, to access my solution for this PSET you can go to:

`https://anish.lakkapragada.com/notes/TYPE-CODE/sols/N.pdf`

These PSETs and associated solution PDFs are synchronized daily at 4:20AM with my computer files through a Cronjob Shell Script. If you want to contribute any corrections, please email `anish.lakkapragada@yale.edu`.

*Note that PDF here is referring to Portable Document Format, not to be confused with the veritable Probability Density Function.

S&DS 242/542: Homework 4

Due Wednesday, February 12, at 1PM

1. **Signed rank test.** Suppose $X_1, \ldots, X_n$ are IID continuous random variables with an unknown PDF f. Consider testing the null hypothesis

$$H_0 : f \text{ is symmetric around } 0$$

(This means that $f(x) = f(-x)$ for every $x \in \mathbb{R}$.) The **Wilcoxon signed-rank statistic** is

$$W = \sum_{i=1}^{n} S_i R_i$$

where

$$S_i = \begin{cases} 1 & \text{if } X_i > 0 \\ 0 & \text{if } X_i \leq 0 \end{cases}$$

and R_i is the rank of $|X_i|$ among the values $\{|X_1|, \ldots, |X_n|\}$ sorted in increasing order (so $R_i = 1$ for the smallest $|X_i|$, $R_i = 2$ for the second smallest $|X_i|$, etc.). Thus, W sums these ranks corresponding to only the positive values of X_i.

(a) Explain briefly why W is pivotal under H_0. To test against a one-sided alternative H_1 that the X_i's tend to take positive values, would you reject H_0 for large or small values of W?

(b) Under H_0, show that

$$\mathbb{E}[W] = \frac{n(n+1)}{4}$$
$$\mathrm{Var}[W] = \frac{n(n+1)(2n+1)}{24}$$

Assuming that W has an approximate normal distribution under H_0 for large n, explain how you would use this approximation to perform your test in part (a) at significance level α.

(Hint: To compute the mean and variance of W, write $W = \sum_{k=1}^{n} k I_k$, where $I_k = 1$ if the observation i with rank $R_i = k$ has $S_i = 1$, and $I_k = 0$ if this observation has $S_i = 0$.)

2. **Permutation tests for paired samples.** Suppose $X_1, \ldots, X_n$ are IID continuous random variables with an unknown PDF f. Consider testing the same null hypothesis as in Problem 1,

$$H_0 : f \text{ is symmetric around } 0$$

Let $T(X_1, \ldots, X_n)$ be any test statistic.

(a) Describe the distribution of T conditional on $|X_1|, \ldots, |X_n|$, under H_0. (What values can T take conditional on $|X_1|, \ldots, |X_n|$, and with what probabilities? You may assume that no X_i is exactly equal to 0.)

(b) Using part (a), explain how computer simulation can be used to perform a level-α test that rejects H_0 for large values of T.

If each X_i is the difference $X_i = Y_i - Z_i$ where $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ are n IID data pairs (e.g. $X_1, \ldots, X_n$ are the differences between two test scores for n students), explain why your procedure may be interpreted as a permutation test for testing the null hypothesis

$$H_0 : (Y_i, Z_i) \text{ has the same bivariate distribution as } (Z_i, Y_i)$$

3. **Testing a uniform null (Rice 9.20).** Consider two PDFs over $x \in [0,1]$: $f_0(x) = 1$ and $f_1(x) = 2x$. Consider a single observation $X \in [0,1]$ generated from one of these two distributions. Among all tests of the null hypothesis $H_0 : X \sim f_0(x)$ versus the alternative $H_1 : X \sim f_1(x)$ with significance level $\alpha = 0.10$, how large can the power possibly be?

4. **Most-powerful test for the normal variance.**
(a) For data $X_1, \ldots, X_n$ and two known and pre-specified values $\sigma_0^2 < \sigma_1^2$, consider testing

$$H_0 : X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(0, \sigma_0^2)$$
$$H_1 : X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(0, \sigma_1^2)$$

What is the most powerful test for testing H_0 versus H_1 at significance level α? Letting $\chi_n^2(\alpha)$ denote the upper-α point of the χ_n^2 distribution, describe explicitly both a test statistic T for your test and its associated rejection region.

(b) What is the distribution of your test statistic T under the alternative hypothesis H_1? Letting F denote the CDF of the χ_n^2 distribution, provide a formula for the power of this test against H_1, in terms of $\chi_n^2(\alpha)$, σ_0^2, σ_1^2, and F. Keeping σ_0^2 and α fixed, what happens to the power of the test as σ_1^2 increases to ∞?