

STAT 240 Homework 3

Rebecca Barter, Andrew Do, and Kellie Ottoboni

March 13, 2015

1) Show that a permutation test based on \bar{X} and a permutation test based on t are equivalent when $m = n$

Note that the t statistic is defined by

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{Var(X)}{n} + \frac{Var(Y)}{m}}}$$

and that a permutation test based on t involves

1. Fill a box with the observed data.
2. Draw a simple random sample of size n and call it X , call the remaining elements in the box Y .
3. Compute t as you would if X and Y were your original data. Call this $t^{*(1)}$.
4. Repeat steps 1-3 L times to get $t^{*(2)}, t^{*(3)}, \dots, t^{*(L)}$.
5. The distribution of the $t^{*(\ell)}$ approximates the true probability distribution of t under the strong null. In particular, a (left-tail) p -value can be computed as

$$\frac{1}{L} \# \{t^{*(\ell)} \leq t\}$$

We thus need to show that we can write

$$t^* = \frac{\bar{X}^* - \bar{Y}^*}{\sqrt{\frac{Var(X^*)}{n} + \frac{Var(Y^*)}{n}}}$$

in terms of \bar{X}^* only.

We note, however, that if we simply write $A = \sum_{i=1}^n X_i + \sum_{i=1}^n Y_i = \sum_{i=1}^n X_i^* + \sum_{i=1}^n Y_i^*$ to be the sum of all observations (which can be considered our new population from which we are drawing), we have that

$$\bar{X}^* - \bar{Y}^* = \left(1 + \frac{n}{n}\right) \bar{X}^* - \frac{1}{n} A = 2\bar{X}^* - \frac{1}{n} A$$

Thus the RHS depends only on \bar{X}^* . Next, if we write $B = \sum_{i=1}^n X_i^2 + \sum_{i=1}^n Y_i^2 = \sum_{i=1}^n X_i^{*2} + \sum_{i=1}^n Y_i^{*2}$, then

$$\begin{aligned}
\frac{Var(X^*)}{n} + \frac{Var(Y^*)}{n} &= \frac{1}{n} \left[\frac{1}{n} \sum_{i=1}^n X_i^{*2} - \frac{1}{n^2} \left(\sum_{i=1}^n X_i^* \right)^2 \right] + \frac{1}{n} \left[\frac{1}{n} \sum_{i=1}^n Y_i^{*2} - \frac{1}{n^2} \left(\sum_{i=1}^n Y_i^* \right)^2 \right] \\
&= \left[\frac{1}{n^2} \sum_{i=1}^n X_i^{*2} + \frac{1}{n^2} \sum_{i=1}^n Y_i^{*2} \right] - \frac{1}{n} \bar{X}^{*2} - \frac{1}{n^3} [A - n\bar{X}^*]^2 \\
&= \frac{1}{n^4} \left[n^2 \sum_{i=1}^n X_i^{*2} + n^2 \left(B - \sum_{i=1}^n X_i^{*2} \right) \right] - \frac{1}{n} \bar{X}^{*2} - \frac{1}{n^3} [A - n\bar{X}^*]^2 \\
&= \frac{B}{n^2} - \frac{1}{n} \bar{X}^{*2} - \frac{1}{n^3} [A - n\bar{X}^*]^2
\end{aligned}$$

which depends only on \bar{X}^* . Thus we have shown that we can write t^* in terms of \bar{X}^* only, implying that t and \bar{X} are equivalent test statistics for a permutation test when $m = n$.

2) Construct a hypothetical dataset (with at least 3 data points in treatment and at least 3 in control) for which the p -value of a permutation test based on \bar{X} is smaller than the p -value of a permutation test based on t . Try to make the difference substantial.

3) Construct a hypothetical dataset (with at least 3 data points in treatment and at least 3 in control) for which the p -value of a permutation test based on t is smaller than the p -value of a permutation test based on \bar{X} . Try to make the difference substantial.

4) Construct a hypothetical dataset (with at least 3 data points in treatment and at least 3 in control) for which the p -value of a permutation test based on \bar{X} is smaller than the p -value of a standard t test. Try to make the difference substantial.