LECTURE 11A NOTES

1. Permutation tests. Permutation tests are tests of hypotheses that assert two sets of observations are exchangeable. The tests are based on the observation that if two sets of observations $\{\mathbf{x}_i\}_{i\in[n_x]}$ and $\{\mathbf{y}_i\}_{i\in[n_y]}$ are exchangeable, their joint distribution is invariant to swapping some $\mathbf{x}_i$'s with $\mathbf{y}_i$'s. The basic idea is best illustrated by an example.

EXAMPLE 1.1. *Let $\mathbf{x}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_x, 1)$ and $\mathbf{y}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_y, 1)$. Consider testing $H_0 : \mu_x = \mu_y = \mu$ versus $H_1 : \mu_x \neq \mu_y$. Since $\{\mathbf{x}_i\}_{i\in[n_x]}$ and $\{\mathbf{y}_i\}$ are* i.i.d. *under H_0, the observations are exchangeable.*

Let $n = n_x + n_y$,

$$\mathbf{z} := (\mathbf{x}_1, \dots, \mathbf{x}_{n_x}, \mathbf{y}_1, \dots, \mathbf{y}_{n_y}),$$

and

$$\mathbf{t} := \phi(\mathbf{z}) := \left| \tfrac{1}{n_x} \textstyle\sum_{i\in[n_x]} \mathbf{z}_i - \tfrac{1}{n_y} \textstyle\sum_{i\in[n_y]} \mathbf{z}_{n_x+i} \right|.$$

be the test statistic. Consider its values $\{\mathbf{t}_\pi\}$ for all $n!$ permutations of the components of $\mathbf{z}$. Under H_0, we expect all the permutations to be equally likely. Under H_1, we expect $\phi(\mathbf{z})$ to be large. Thus, a p-value is

$$p(\mathbf{t}) = \frac{1}{n!} \sum_\pi \mathbf{1}_{(\mathbf{t}_\pi, \infty)}(\mathbf{t}),$$

where the sum is over all $n!$ permutations.

When deriving permutation tests, one must take care to ensure that the two sets of observations are exchangeable under the null hypothesis.

EXAMPLE 1.2. *Let $\mathbf{x}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_x, \sigma_x^2)$ and $\mathbf{y}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_y, \sigma_y^2)$. Consider testing $H_0 : \mu_x = \mu_y = \mu$ versus $H_1 : \mu_x \neq \mu_y$. The preceding example is not a valid α-level test of H_0 because*

1. *$\{\mathbf{x}_i\}_{i\in[n_x]}$ and $\{\mathbf{y}_i\}_{i\in[n_y]}$ are not exchangeable (their variances differ).*
2. *some permutations are more likely than others: the permutations that label the largest and smallest observations as drawn from the distribution with larger variance are more likely.*

Thus $p(\mathbf{t})$ is not a p-value (it is not stochastically larger than $\mathrm{unif}(0,1)$ *under H_0), and rejecting H_0 when $p(\mathbf{t}) \leq \alpha$ does not control the Type I error rate at the nominal level.*

However, if σ_x^2, σ_y^2 are known, it is easy to normalize the observations by their standard deviations and conduct a permutation test on the normalized observations.

Formally, permutation tests are based on the observation that the order statistics are sufficient for any model of exchangeable random variables. Indeed, if $\mathbf{z}$ is a set of exchangeable random variables, then

$$\mathbf{P}\big(\mathbf{z} \mid \{\mathbf{z}_{(i)}\}_{i\in[n]}\big) = \frac{1}{n!}$$

no matter the particular distribution of $\mathbf{z}$, which in turn determines the conditional distribution of $\phi(\mathbf{z}) \mid \{\mathbf{z}_{(i)}\}_{i\in[n]}$. By permuting the $\mathbf{z}_i$'s and evaluating the test statistic on the permuted observations, we are simulating the conditional distribution of $\mathbf{t} \mid \{\mathbf{z}_{(i)}\}_{i\in[n]}$. Once the conditional distribution is known,

$$p(\mathbf{t}) = 1 - F(\mathbf{t}),$$

where F is the CDF of $\mathbf{t} \mid \{\mathbf{z}_{(i)}\}_{i\in[n]}$, is a p-value. When n is not too large, it is possible to exactly characterize the distribution of $\mathbf{t} \mid \{\mathbf{z}_{(i)}\}_{i\in[n]}$ by evaluating $\mathbf{t}_\pi$ for all $n!$ permutations. Otherwise, we resort to simulation.

We remark again the one must be careful when considering permutation tests. Ultimately, a permutation test is a test of the exchangeability of the observations. Thus, in order to justify a permutation test of a null hypothesis H_0, one must establish that the observations are exchangeable under H_0. Otherwise, the permutation test may not control the Type I error rate at the nominal level.

Yuekai Sun
Berkeley, California
November 28, 2015