Comparing two-sample tests

In order to compare some of the nonparametric tests that we have been studying to the parametric t-test, we will assume that the distributions in the two groups are continuous and differ only by a shift parameter Δ, ($F_1(x) = F_2(x - \Delta)$) so that the null hypothesis is $H_0 : \Delta = 0$, with either one or two-sided alternative hypotheses. It is known that for normal populations, the t-test holds its' significance level and is the Uniformly Most Powerful Unbiased (UMPU) test.

Simulation results

See the text and Table 2.9.1. One result is that the Wilcoxon test is generally more powerful for heavy-tailed distributions and moderate-to-large sample sizes. The one area where the t-test is often more powerful is when sample sizes are small, which is when there may be less certainty about the t-test holding its' Type I error.

The concept of relative efficiency

Suppose that we are interested in comparing two tests of the hypothesis $H_0 : \Delta = 0$ versus $H_a : \Delta > 0$, at the same level of significance. Let the sample sizes for the two tests be $m_1 + n_1 = N_1$ and $m_2 + n_2 = N_2$, with $m_1/n_1 = m_2/n_2$. If N_1 and N_2 are chosen so that the tests have the same power, then the relative efficiency of test 1 to test 2 is defined to be:

$$\text{eff}(1 \text{ vs } 2) = \frac{N_2}{N_1}.$$

A relative efficiency eff(1 vs 2) greater than 1 thus indicates that test 1 requires less data and hence is more efficient than test 2. Under general conditions we can take the limit of this expression as $N_1 \to \infty, N_2 \to \infty$, and $\Delta \to 0$. This limit is called the asymptotic relative efficiency (a.r.e.) of test 1 to test 2, and is independent of the Type I error and power. The a.r.e. of two tests gives us an idea of how they compare in large samples. Some a.r.e. values for the Wilcoxon test compared to the t-test are shown in Table 2.9.2.

Permutation tests

The text presents some results on how many samples are needed from the permutation distribution, and on comparisons between permutation tests based on different statistics.