Abstract

Parametric statistical methods are based on a number of assumptions not often satisfied by the data. Non-parametric methods require less assumptions but are less efficient. To overcome these limitations, it is possible to use distribution-free and assumption-free methods which depend only on the data on hand. But these procedures depend on resampling the data in a suitable manner. This involves a huge amount of computations but can be accomplished with modern computing resources.

Permutation Tests

- By T. Krishnan

Introduction

Parametric and nonparametric statistical methods require the data to satisfy certain assumptions and cannot be used on data that are small and ill-behaved. Analysis of such data can be tackled by "exact" statistical methods which require only simple and minimal assumptions about the source of data and their behavior.

Exact methods belong to the class of methods which depend on "resampling" from the data on hand and generating reference distributions thereof. These techniques include methods known as Resampling Methods, Permutation Methods, Monte Carlo Methods, Bootstrap Methods, Exact Methods, etc. This article is an introduction to these methods.

The implementation of these methods require a great deal of computing resources but can be achieved easily by modern computers.

Example: Paired t Test

Let us consider an example of data on the number of driving hours in a month of the husband and wife of 20 couples and examine if the mean driving hours of the husband and wife are the same. The observations are as follows:

Driving Hours

Husband	Wife
149	0
0	51
0	0
259	385
106	0
255	235
0	0
52	0
340	48
0	65
180	77
0	0
84	0
89	0
212	53
554	150
500	0
424	165
112	98
2600	0

Notice that this is paired (correlated) data. If you were to use parametric methods, you would opt for a paired t test with results:

Mean Difference = 229.45 SD Difference = 579.60 95.00t = 1.770395 df = 19 Prob = 0.092706

But the assumption required for it is that, the difference is normally distributed. But one look at the data will be enough to convince you that this is going to be no good, what with lack of symmetry, extreme values and outliers, etc.

A nonparametric test for this is the Wilcoxon Signed Rank Test. This test uses the differences, their signs, the mean values of positive and negative differences and constructs a statistic denoted by V and uses the asymptotic distribution of V under the null hypothesis with the following results:

Wilcoxon signed rank test with continuity correction

V = 135, p-value = 0.00604
alternative hypothesis: true location shift is not equal to 0

Warning message: In wilcox.test.default(husband, wife, paired = TRUE): cannot compute exact p-value with zeroes. The test is asymptotic and it has problems with 0's. We have only 20 observations. The p-values are vastly different.

Permutation Test for Paired Observations

An exact test approach to this problem is as follows: Under the hypothesis that there is no difference in the husband-wife distributions, in each pair either of the two numbers could have been the husband's and, the other the wife's. So under the hypothesis, we could have obtained any of the 2^{20} possible data sets and what we have observed is one of them.

The steps involved in developing a Permutation Test are as follows:

1. Analyze the problem: Set up a hypothesis; an alternative; understand data assumptions
2. Choose a statistic S to distinguish the hypothesis from the alternative
3. Compute statistic S from the given data
4. Find all re-arrangements of data as per the same data design consistent with the hypothesis
5. For each re-arrangement, compute statistic S
6. Form histogram of S values
7. Compare with S obtained in 3 above to determine the proportion of cases as more extreme than observed S, just like you do in conventional p-value computation
8. This proportion is the p-value.

Permutation Test for Paired Values

- Hypothesis to be tested is mean of husband = mean of wife; alternative: mean of husband $\neq$ mean of wife. The data appear in (husband, wife) pairs; each data point is a pair, which should not be separated during permutation.
- Let us choose the statistic as S = difference between husband mean and wife mean.
- For given data S = 229.45

- Re-arrangements will consist of interchanging the husband and wife values in a subset of data points and leaving the rest alone. This would mean that each data point has one of the two options: interchange or keep the same as observed. This means that the total number of possible permutations = $2^{20} = 1048576$. Note that keeping the pair together is the essence of the choice of permutations.

  Also, keep in mind that if the data is two-sample data, which means that there are 40 subjects with 20 husbands and 20 (independent) wives chosen, then the permutation set will consist of distributing the 40 numbers into subsets of 20 each, one for husbands and one for wives, amounting to $\binom{40}{20} = 137846528820$ permutations . This is not correct permuting for the paired case. Keeping the pairs intact naturally results in a much smaller number of permutations.

- The histogram of the S values are given in the graph below. Notice that for each permutation there is an exact opposite permutation for which the S value is the same with opposite sign. Thus it is enough to compute S for only half of the permutations.

- This distribution of S can be regarded as the reference or null distribution for computing p values and confidence intervals. In parametric inference, this reference distribution is computed theoretically from the assumptions. In non-parametric inference, the asymptotic distribution is computed theoretically.

Permutation Distribution

Histogram of the Mean Difference in the 2^{20} Permuted Samples
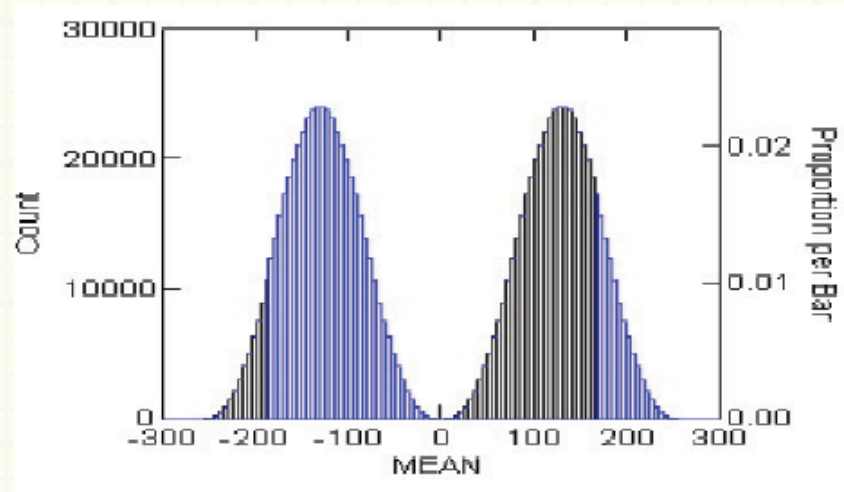
Hypothesis Testing with Permutation Distribution

- There are 1800 $\times 2$ observations in 'mean' which are outside of (-229.45,229.45)
- The proportion of observations in 'mean' outside (-229.45,229.45) is 0.003433

- Thus the two-sided permutation (exact) p-value = 0.003433
- Notice the enormous difference in the p-values of the three tests. The Permutation Test is the most valid of these as it is not based on any assumptions.

Properties of Permutation Tests

- This approach is applicable to a wide variety of problems
- The method is based only on observed data
- Data peculiarities are tolerated
- No models are used
- No sampling distribution assumptions or asymptotics are invoked
- The power of this test approximates to a corresponding parametric test when applicable
- The p-value computed is exact
- A permutation test is generally more powerful than a "traditional" nonparametric test

Randomization Test

Note that even for such a small sample in this example, the number of permutations is huge. In a larger data set, it will be quite impossible to generate and compute for all the permutations. In such cases, a random sample of permutations carefully chosen will be adequate. Tests based on such samples are called **Randomization Tests**. They have properties similar to Permutation Tests.

There are other variations of these tests such as Monte Carlo Tests, Bootstrap Tests, etc.

References

Manly, B.F.J. (2007). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Third Edition. Boca Raton, FL: Chapman & Hall/CRC.