

Confidence intervals for an estimated mean

Some reflection will show that it is not possible to estimate confidence intervals for a mean without making some distributional assumption. Without assumptions we cannot address the possibility that the distribution takes very large positive values with very low probability: this could give it an arbitrarily high mean, which we would not expect to detect.

A symmetry-based test

If we are willing to assume our distribution is symmetric, however, we can get confidence intervals for the mean by permutation [1]. If we have n observations \hat{x}_i , and a symmetric distribution with hypothesized mean h , the null model is that we were equally likely to have observed \hat{x}_i or the value the same distance away from \hat{x}_i but on the other side: $2h - \hat{x}_i$. We can thus construct a null distribution of means under the null hypothesis: $\mu_{\text{hyp}} = \langle \tilde{x} \rangle$, where \tilde{x} takes the values \tilde{x}_i , each of which is \hat{x}_i with probability $1/2$ and $2h - \hat{x}_i$ with probability $1/2$. This distribution has 2^n values: depending on the size of n we can either enumerate it or sample from it.

The critical value of h

This method allows us to calculate a P value for a given value of h , but it is not yet a practical method for estimating confidence intervals. To estimate confidence intervals, we wish to do a single set of permutations that will tell us the answer for all values of h . The dividing line for whether a value of μ_{hyp} is more extreme than the observed mean μ_{obs} is of course the point where the two values are equal. If we write C for the set of indices over which $\tilde{x}_i \neq \hat{x}_i$, then this critical point is given by:

$$\sum_i \hat{x}_i - \tilde{x}_i = 0 \tag{1}$$

$$\sum_{i \in C} \hat{x}_i - (2h - \hat{x}_i) = 0 \tag{2}$$

$$\sum_{i \in C} \hat{x}_i = \sum_{i \in C} h \tag{3}$$

We can thus define the critical value of h_c that corresponds to a choice of indices to switch C as the average over x_i , for $i \in C$. We can then examine a single enumeration of or sample from the subsets of the indices, and take our confidence intervals to be the desired quantiles of the distribution of h_c .

Note that the mean corresponding to the empty set is undefined. This set (if it appears in our sample) should be counted on both sides of the distribution (i.e., it does not provide any confidence that the true mean is larger or smaller than any value). In particular, this means that we require at least six samples to obtain a two-sided 95% confidence interval for the mean using a

permutation test with the symmetry assumption. This corresponds intuitively to the observation that, with five samples, we cannot be confident at the 95% level that we did not randomly choose all of them to be on the same side of the mean.

The conservative bootstrap

I have shown above that the standard symmetry-based permutation test for confidence intervals of the mean of a sample is equivalent to a resampling test. Instead of resampling observed values with replacement, however, this test chooses at random whether to ignore or not ignore a given value. It seems to me that this approach addresses a possible concern with the conventional bootstrap: the conventional bootstrap is fundamentally non-conservative: in particular, as sample size gets smaller, our nominal confidence becomes greater.

I refer to a test that bootstraps over subsets, rather than equal-sized resamples, as a “conservative bootstrap”. It seems natural to wonder whether the conservative bootstrap could be a useful alternative to standard bootstrapping in a variety of applications, for example constructing bootstrap support estimates for phylogenetic trees.

R files

The calculations above are implemented in `permmean`, available in the file `permtests.R`. This test depends on functions that create subsets and evaluate lists of critical values from permutation tests, available in the file `permutils.R`.

References

- [1] Philip Good. *Permutation Tests*. Springer, New York, 2000.