

## Bootstrapping and Resampling

### 1 Bootstrapping Correction for False-Positive error

The goal of this method is to reduce the False-Positive (Type I) error, which is commonly due to finding more significant activation where there is not any. In other words, the test statistics rejects the NULL hypothesis (no activation) and effectively accepts the alternative hypothesis (activation is present). However, if a large number of tests are carried out there is data dragging, which will result in finding many activated regions, just by chance, which in fact are not significant in reality. The question is how to sift out the really significant statistics and avoid (or reduce) the False-Positive errors?

Suppose we have the functional volumetric (PET/SPECT) brain data for two groups of  $N$  subjects each; **group 1** –  $\{X_i\}_{i=1}^N$ , and **group 2** –  $\{Y_i\}_{i=1}^N$  (group sizes need not be equal, as independent tests are used). We first do regular SVT 4.2 (with no correction for multiple testing, like the Bonferroni-type correction implemented in SVT 5.2). Presumably, there will be some type-I error (regions declared significant while in fact they are not). Then for each of the significantly activated voxels, labeled by  $v_{a_o} \in \{v_a\}_{a \in I}$ , do the following:

Let  $Z_{v_{a_o}}^{obs}$  be the observed statistically significant Z-score at voxel location  $v_{a_o}$  according to the

SVT 4.2  $\left( Z_{v_{a_o}}^{obs} \geq 1.96 \right)$ . And similarly let

$$Z_{v_{a_o}}^b = \frac{\frac{1}{N} \sum_{l=1}^N D_l^b(v_{a_o})}{\sigma_{D_m}^b}$$

be the bootstrap statistic of interest (Z-score), where  $D_l^b(v_{a_o})$  is the intensity at  $v_{a_o}$  of the  $b^{th}$  randomized difference signal  $\left[ D_l^b(v_{a_o}) = X_{b_l}(v_{a_o}) - Y_{b_l}(v_{a_o}) \right]$  and  $\sigma_{D_m}^b$  is the corresponding (bootstrap sample) estimate of the standard deviation.

We select  $B$ , a number of bootstrap sub-samples ( $> 1,000$ , guaranteed to exist if  $N \geq 3$ , since  $\ln(2N) \geq 3/N$ ). Generate  $B$  bootstrap samples by randomly (uniformly) paring data intensities at voxel locations  $v_{a_o}$ , across all PET data, with replacement (and hence possible repetitions) from the mixed pool of all subjects. Basically mix all subjects and randomly cut in half allowing

repetition. Note that configurations in which the same subject is represented several times, or not at all, are allowed and certainly bound to appear.

Compute the value of the bootstrap statistic of interest at each of the  $\mathbf{B}$  bootstrapping samples for the fixed voxel location  $v_{a_o}$ . After obtaining these Z-scores,  $\left\{ Z_{v_{a_o}}^b \right\}_{b \in \mathbf{B}}$ , we determine the (bootstrap-sample) average Z-score

$$\bar{Z}_{v_{a_o}} = \frac{\sum_{b=1}^{\mathbf{B}} Z_{v_{a_o}}^b}{\mathbf{B}}$$

and its (bootstrap-sample) variance

$$\hat{\sigma}_{v_{a_o}}^2 = \frac{\sum_{b=1}^{\mathbf{B}} \left( Z_{v_{a_o}}^b - \bar{Z}_{v_{a_o}} \right)^2}{\mathbf{B} - 1}$$

Then, we use the usual bootstrap normal approximation to find  $CI(95\%, v_{a_o})$ , the 95%

confidence interval for  $Z_{v_{a_o}}^{obs}$ ,  $CI(95\%, v_{a_o}) = \bar{Z}_{v_{a_o}} \pm 1.96 \times \hat{\sigma}_{v_{a_o}}$ .

Finally,  $Z_{v_{a_o}}^{obs}$  is declared significant if it lies outside of  $CI(95\%, v_{a_o})$ , and not-significant if it is within this confidence interval. The same procedure is repeated across all SVT 4.2 significantly activated voxels  $\{v_a\}_{a \in I}$ .

**Another way** to apply bootstrapping is the following: As before, determine the bootstrap

statistics of interest  $\left\{ Z_{v_{a_o}}^b \right\}_{b \in \mathbf{B}}$  and report only the *p-value* at each voxel-location  $v_{a_o}$ .

$$p = \frac{C \left( \left\{ \left| \bar{Z}_{v_{a_o}}^b \right| \geq \left| \bar{Z}_{v_{a_o}}^{obs} \right| \right\}_{b=1}^{\mathbf{B}} \right)}{\mathbf{B}},$$

where  $C(\text{Set})$  is the cardinality of the set  $\text{Set}$ . This assigns a probability value to each voxel – indicating the ratio of the  $Z$ -scores in the bootstrap samples greater than the observed  $Z_{v_{a_0}}^{obs}$   $Z$ -score.

## 2. Traditional confidence limits

Let's leave bootstrapping for a minute, and just concentrate on standard confidence limits in parametric statistics. Again, we will focus on the population mean as a parameter of interest. We

know that the usual confidence limits can be found as  $CI\left(\frac{\alpha}{2}, \mu\right) = \bar{X} \pm t_{\alpha/2} S_{\bar{X}}$ . We can

solve for these limits just by taking the sample mean,  $\bar{X}$ , finding the critical value of  $t$  from Student's tables (for reasonable sample sizes it will be a bit more than 2.0), and multiplying by the standard error of the sample average, which is just the standard deviation of the sample divided by the square root of  $n$ . Notice that these limits will be symmetric because  $+t$  and  $-t$  will be equal except for the sign. How is this related to constructing bootstrapped  $t$  intervals?

## 3. Bootstrapped $T$ intervals

Efron (1979+) called the (bootstrap) extension intervals of the parametric statistics *bootstrapped- $t$  intervals*. These are actually surprisingly like the traditional intervals, but with a twist. To calculate the traditional intervals, we had to use the tables of the Student  $t$  distribution. But Gosset (Student) originally derived that distribution on the assumption that we were sampling from a normal population. And the whole purpose behind bootstrapping is to *get away from making that (parametric) kind of assumption*.

Suppose that we took our original sample, treated it as a pseudo-population, drew  $B$  bootstrapped samples, and calculated  $\bar{X}$  and  $S$  from each. From these statistics we could solve for  $t^*$ , where the asterisk (\*) is used to indicate that each of these is a  $t$  calculated on a bootstrapped sample. Now all we need are the 2.5% and 97.5% cutoffs of the  $t$  distribution we would have without assuming normality. And we can get those cutoffs just by drawing many bootstrapped samples,

and calculating  $t^*$  for each sample [i.e. for each sample ( $i$ ) we calculate  $\frac{(\bar{X}_i^* - \bar{X}_i)}{S_i^* / \sqrt{n}}$ , where

$\bar{X}_i^*$  is the mean of the  $i^{th}$  bootstrapped sample,  $\bar{X}_i$  is the mean from our original (observed) sample, and  $S_i^*$  is the (bootstrap-sample) standard deviation.]

After drawing  $B$  bootstrapped samples, we take the resulting sampling distribution of  $t^*$  and find its 2.5% ( $t^*_{0.025}$ ) and 97.5% ( $t^*_{0.975}$ ) cutoffs, and substitute those, instead of tabled values, in the traditional formula. This gives us  $CI\left(\frac{\alpha}{2} = 0.025, \mu\right) = \left[\bar{X} - t^*_{0.975}S_{\bar{X}} ; \bar{X} + t^*_{0.025}S_{\bar{X}}\right]$ .