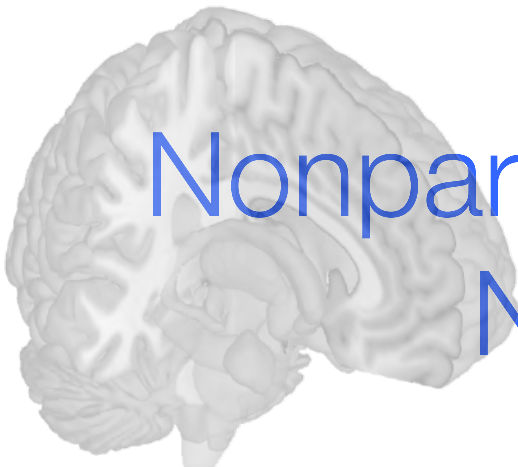


Nonparametric Statistics in Neuroimaging

Martin Lindquist
Department of Biostatistics
Johns Hopkins University

Statistical Parametric Mapping

- Statistical parametric mapping (SPM) refers to a data analytic framework commonly used in neuroimaging.

- It involves using the general linear model (GLM) to construct statistical maps and Gaussian random field theory to threshold them.

- Both are parametric models that make a number of assumptions whose validity impact the results.

Statistical Parametric Mapping

- Violations of any of these assumptions can produce incorrect p-values, an increased number of false positives and generally erroneous results.

- Arguably any number of the necessary assumptions will be violated in any given neuroimaging study.

Nonparametric Inference

- The goal of nonparametric inference is to use the data at hand to perform inference while making as few assumptions as possible.

- Because they only require minimal assumptions to be valid, they provide a flexible methodology for the statistical analysis of neuroimaging data.

Nonparametric Procedures

- Nonparametric equivalents exist to most statistical procedures used in neuroimaging.

- Classical nonparametric tests include:
  - Sign test (one-sample t-test)
  - Mann-Whitney test (two-sample t-test)
  - Wilcoxon signed rank test (paired t-test)
  - Kruskal-Wallis test (ANOVA)

- Classical nonparametric tests are often based on studying ranks.

Nonparametric Procedures

- The use of computer intensive methods to perform nonparametric inference has become increasingly popular in recent years.

- The most popular approaches include:
  - the bootstrap procedure, and
  - permutation tests

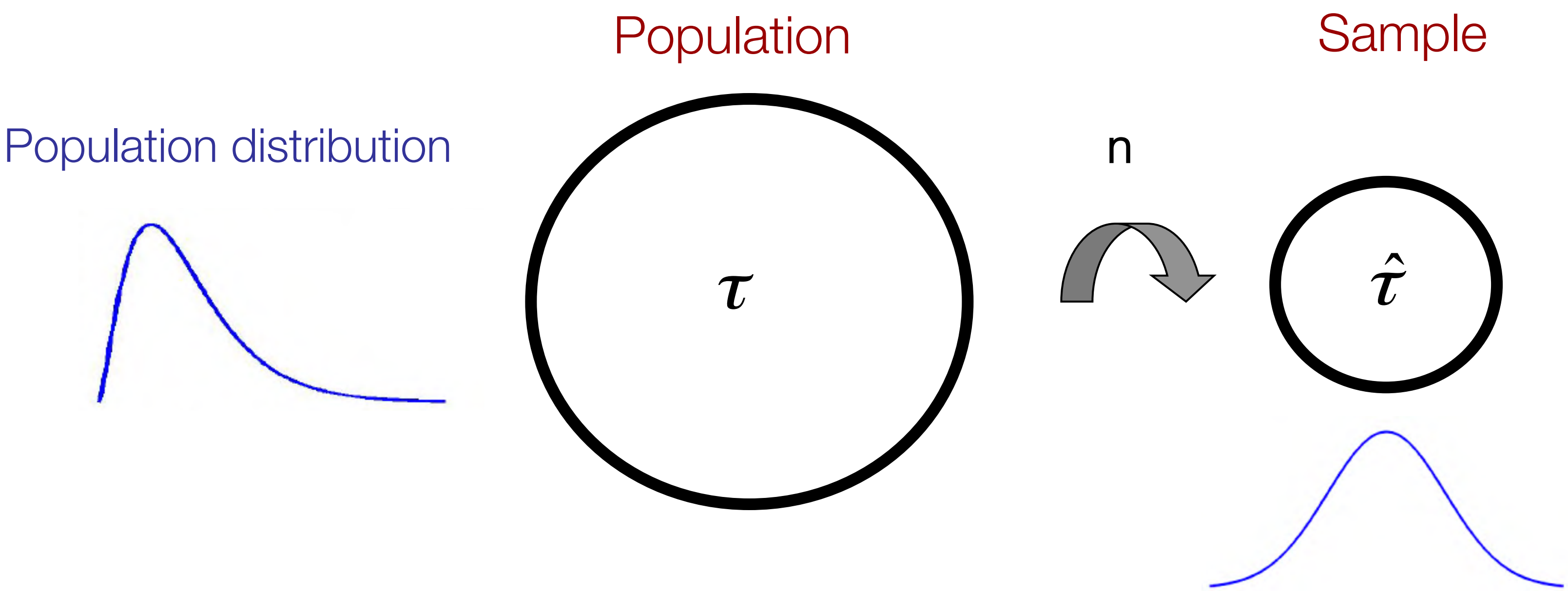
- Both techniques are based on randomly resampling the available data.

Statistical Inference

- A parameter is a number that describes the population, while a statistic describes a sample.

- In statistical inference we use a known statistic to estimate an unknown population parameter.
  - The statistic varies from sample to sample.
  - The sampling distribution is a mathematical model that provides information about this variation.
  - It allows us to construct confidence intervals and hypothesis tests.

Illustration



To perform inference we need the sampling distribution of $\hat{\tau}$.

- Describes what values it can take and how often it takes them.
- In principal can be obtained by repeatedly sampling from the population and studying how the estimate of τ varies.
  - Not feasible in practice.
- Can sometimes be derived theoretically (i.e., sample mean).

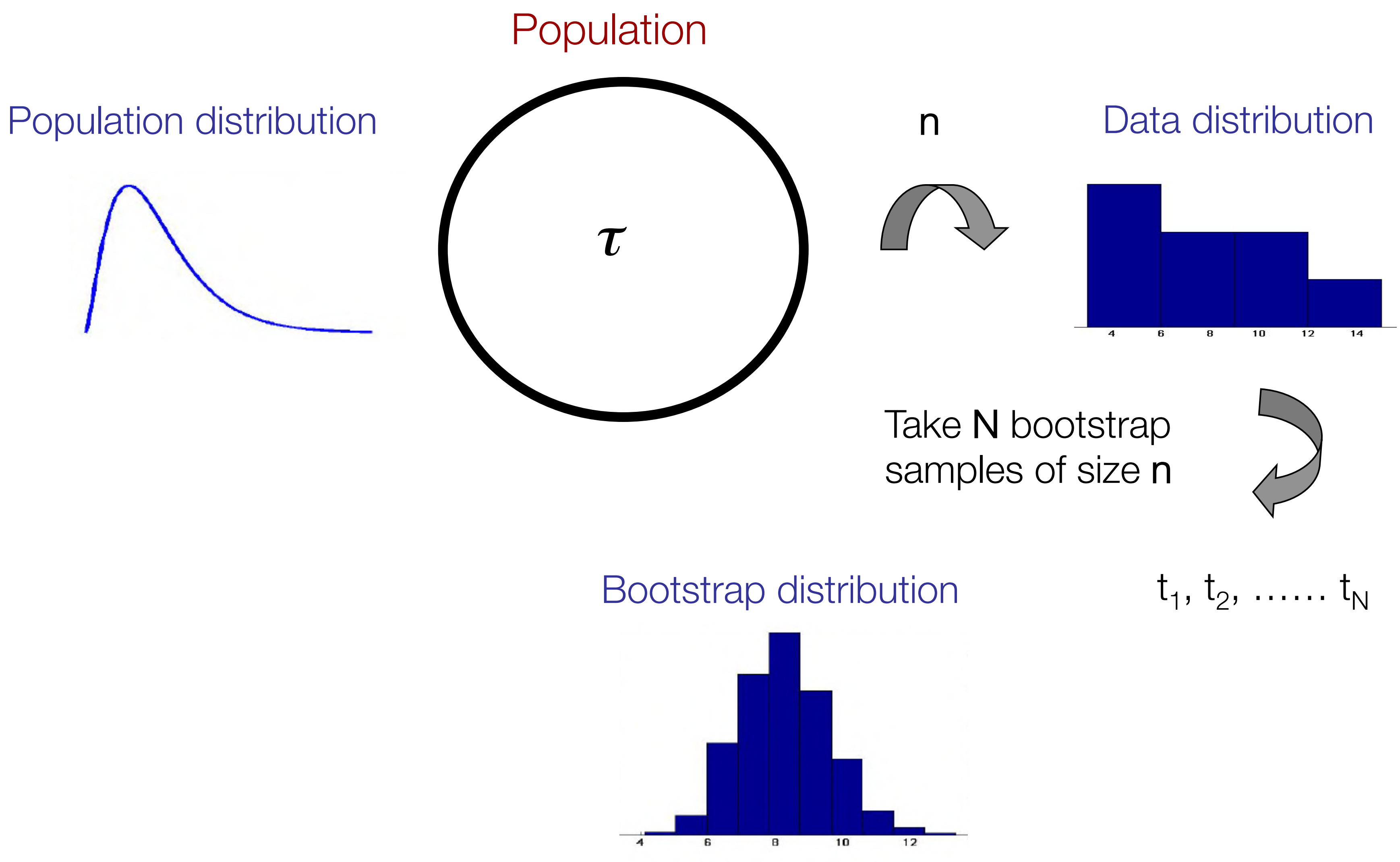
The Bootstrap

- The bootstrap is a computer-based method of inference that allows for estimation of the sampling distribution of almost any statistic.

- It can be used to construct confidence intervals for situations where traditional methods cannot (or should not) be used.

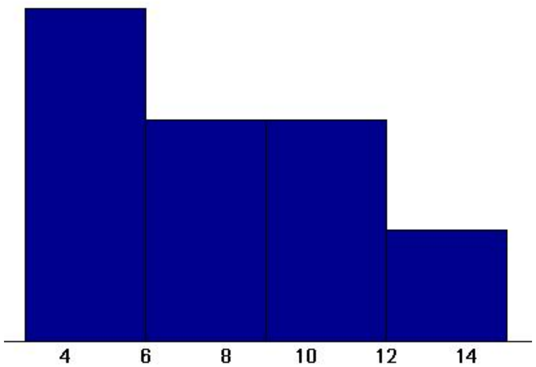
- By repeatedly resampling with replacement from the sample we approximate repeatedly sampling data from the population.

Illustration



Suppose we have the following data set: 3, 5, 6, 7, 8, 10 , 15

Data Distribution

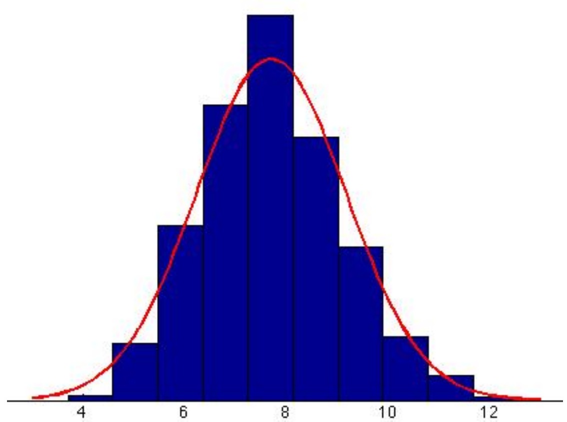


Take repeated resamples of size 7

3	3	3	5	7	7	15	$\bar{x} = 6.14$
5	5	7	8	10	10	10	$\bar{x} = 7.86$
3	3	7	7	15	15	15	$\bar{x} = 9.23$
6	6	7	7	7	10	15	$\bar{x} = 8.29$

......

$$N\left(\mu, \frac{\sigma^2}{n}\right)$$



Using the bootstrap distribution we can construct confidence intervals

Bootstrap distribution of the sample mean

Applications in Neuroimaging

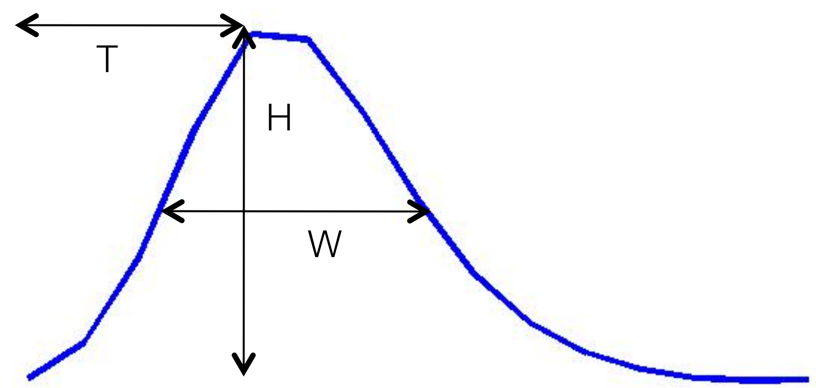
- The bootstrap has many potential uses in neuroimaging.

- In our work we have used it to:
  - Perform second-level inference when the first-level model uses a set of basis functions to model the HRF.
  - Construct confidence intervals for situations when the distribution of a statistic is unknown (e.g., mediation or dynamic correlation).

HRF Estimation

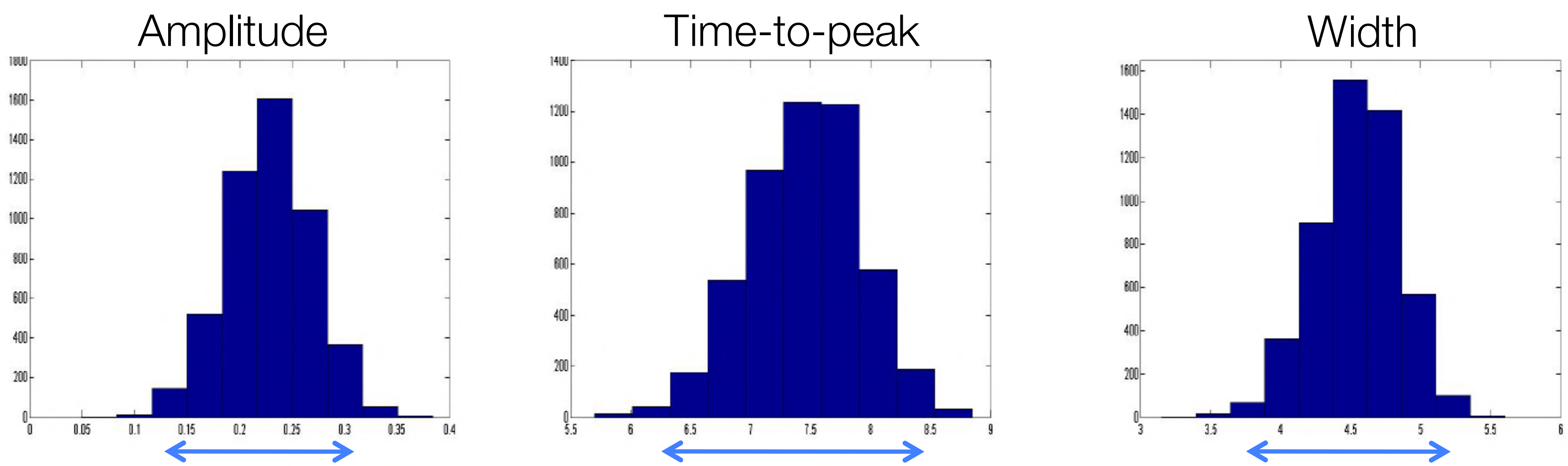
- Group analysis can be tricky when comparing conditions modeled with multiple basis functions.

- Procedure:
  - Use a flexible basis set (e.g., FIR, spline or IL model) to estimate the hemodynamic response.
  - Measure key features of the HRF (e.g., amplitude, time-to-peak and width).
  - Repeat for all subjects in the study.



Bootstrap Procedure

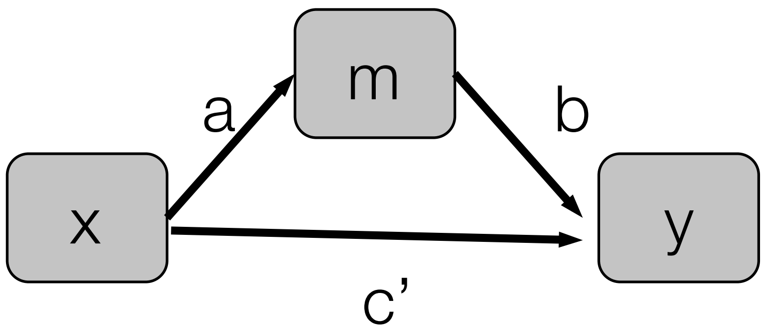
- Use the bootstrap procedure to get confidence intervals for parameters of interest.
  - Sample among the n subjects with replacement.
  - Compute the statistics of interest (e.g., group mean of amplitude, time-to-peak, and width).
  - Repeat many times.



Lindquist et al., 2008

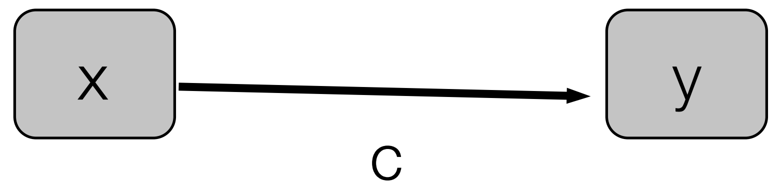
Mediation

<u>Full model, with mediator</u>



$$m = ax + e_m$$
$$y = bm + c'x + e'_y$$

<u>Reduced model, without mediator</u>



$$y = cx + e_y$$

Does *m* explain some or all of the x-y relationship?

Test the significance of the a*b product

Bootstrap Procedure

Procedure:

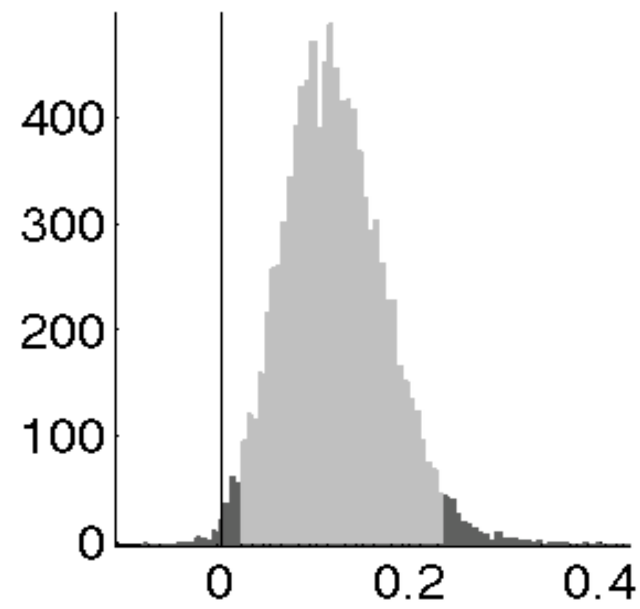
1) Sample rows of matrix [x m y] with replacement to fill sample of N subjects.

2) Calculate mediation effect a*b.

3) Repeat many times (e.g., 1,000-10,000 times).

4) Confidence interval/p-values based on bootstrapped distribution.

Histogram of bootstrapped Indirect (a*b) effects



Wager et al., 2008

Dependent Data

- In the two previous examples we performed the bootstrap across subjects.

- However, in many situations (e.g., dynamic connectivity) we may be more interested in bootstrapping across time.
  - This can be problematic due to autocorrelations in the time course.
  - Here we need to use alternative methods that are more appropriate for dependent data.

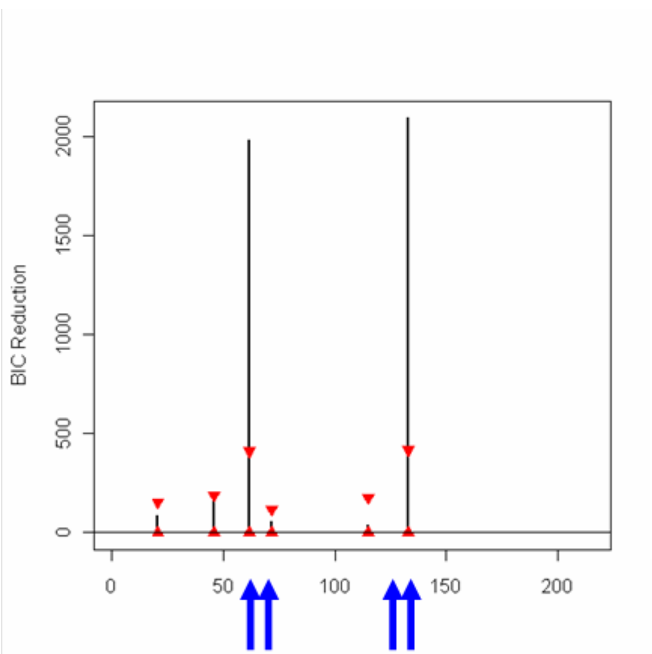
Dependent Data

- The block bootstrap divides the time series into blocks and resamples with replacement these blocks to create a pseudo time series.
  - Successive time points are assumed to be correlated, but observations "far apart" uncorrelated.
  - By resampling sufficiently long blocks, the dependence structure of the original time series is preserved.

- The stationary bootstrap allows for randomly varying block sizes.
  - Pseudo-time series generated by resampling blocks of random size, whose length has a geometric distribution.
  - The pseudo-time series guaranteed to be stationary.

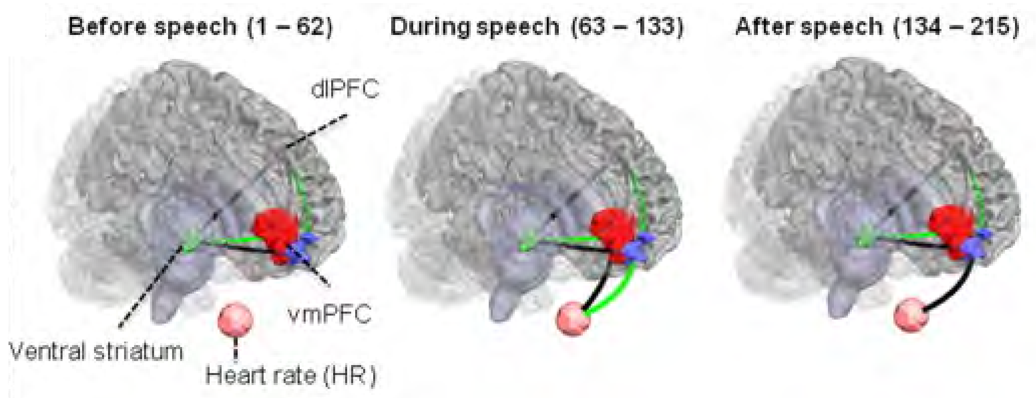
Bootstrap Procedure

- Dynamic connectivity between 4 ROIs and heart rate assessed using Dynamic Connectivity Regression (Cribben et al. 2012).
- Stationary bootstrap used to determine significant changes in connectivity.



Speech Instruction

"No Speech" Instruction



Permutation Tests

- Permutation tests are another example of a computer-intensive statistical technique.
- They are significance tests based on resamples drawn at random from the original data.
- In contrast to the bootstrap, the resamples are drawn without replacement in a manner consistent with the null hypothesis and the study design.

Illustration

H_0: Drug no more effective than placebo

Subject	Actual Treatment
1	Drug
2	Drug
3	Drug
4	Drug
5	Placebo
6	Placebo
7	Placebo
8	Placebo

If the drug has no effect the difference between groups should be the same regardless of whether we look at the actual or permuted treatment assignment.

Permutation Distribution

- The permutation distribution of the statistic of interest is formed using the values of the statistic from a large number of resamples.

- The permutation distribution estimates the sampling distribution under the condition that H_0 is true.

- The permutation distribution can be used to compute p-values to test H_0.

Illustration

- Data from V1 voxel in visual stimulus experiment

  A: Active, flashing checkerboard   B: Baseline, fixation

  6 blocks, ABABAB   Just consider block averages.

A	B	A	B	A	B
103.00	90.48	99.93	87.83	99.76	96.06

- Null hypothesis H_0
  - No experimental effect, i.e. the A and B labels are arbitrary.

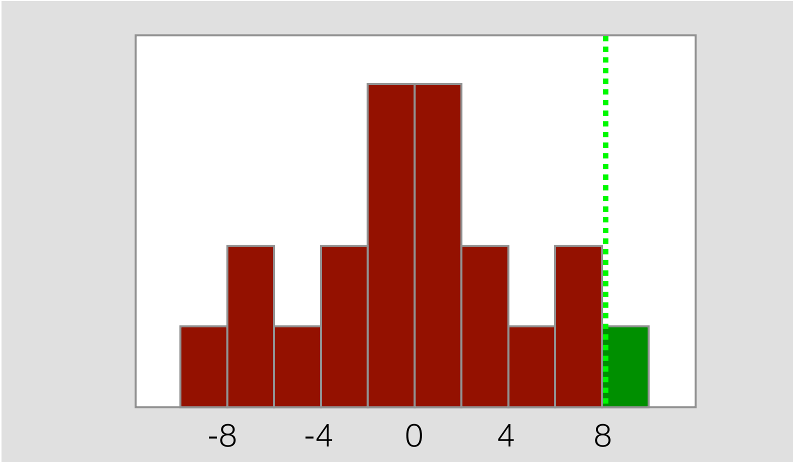
- Statistic
  - Mean difference between conditions.

Nichols & Holmes, 2002

- Under H_o
  - Consider all equivalent re-labelings.
    - Assume exchangeability, i.e. the distribution of the statistic is the same whatever the relabeling.
  - Compute all possible statistic values.
  - Determine the permutation distribution.
    - Each relabeling is equally likely. Hence, each statistic has equal probability.

AAABBB	4.82	ABABAB	9.45	BAAABB	-1.48	BABBAA	-6.86
AABABB	-3.25	ABABBA	6.97	BAABAB	1.10	BBAAAB	3.15
AABBAB	-0.67	ABBAAB	1.38	BAABBA	-1.38	BBAABA	0.67
AABBBA	-3.15	ABBABA	-1.10	BABAAB	-6.97	BBABAA	3.25
ABAABB	6.86	ABBBAA	1.48	BABABA	-9.45	BBBAAA	-4.82

- Permutation distribution



P-value = 0.05

Actual value = 9.45

Permutation tests don't work very well with small sample sizes, as they tend to be conservative.

Comments

- Requires only the assumption of exchangeability
  - Under H_0, the distribution is unchanged by permutation.
  - This allows us to build the permutation distribution.

- Subjects are exchangeable
  - Under H_0, each subject's A/B labels can be flipped.
  - Permutation tests useful for 2nd level analysis.

- fMRI scans not exchangeable under H_0
  - The problem is temporal autocorrelation.
  - Be careful performing permutation tests on individual subjects.

Issues

- Sample size
  - If there are N possible relabelings, the smallest attainable p-value is 1/N which can be problematic at small N.
  - If there are too many possible relabelings it may not be feasible to compute the statistic images for all of them.
    - Randomly sample from the population of relabelings.
    - Each relabeling should be equally likely to be chosen.

- Flexibility
  - The permutation approach is free to consider any statistic and is not bound to those that have a known distributional form.

Issues

- Computational Intensity
  - Analysis needs to be repeated for each relabeling.
  - Computations can be parallelized.
  - Can use GPUs (graphics processing units).
- Implementation
  - Each experimental design type needs unique code to generate permutations.
- Power and Validity
  - Nonparametric procedures are less powerful than their parametric counterparts when the assumptions of the latter hold. If not, nonparametric procedures more valid.

SnPM

- Statistical nonparametric mapping (Nichols & Holmes) is a nonparametric equivalent to SPM.
- It uses a permutation test, rather than random field theory, to correct for multiple comparisons.
- This allows one to avoid the assumptions needed for using random field theory.

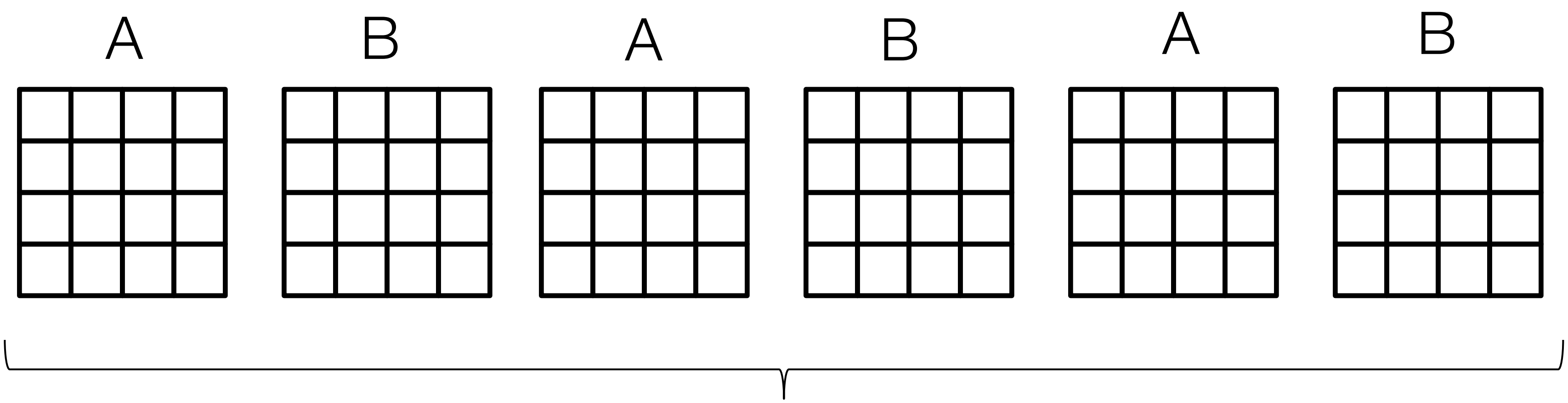
Multiple Comparisons

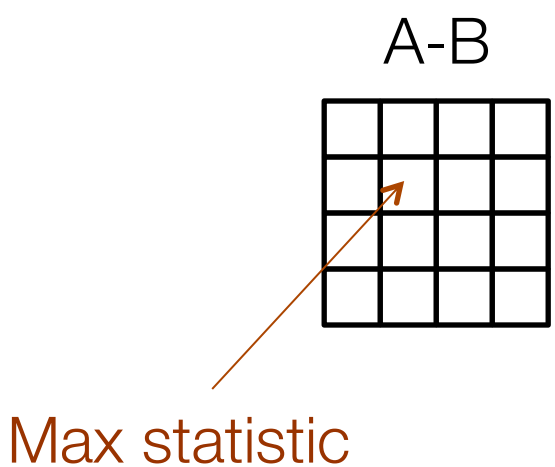
- The family-wise error rate (FWER) is often controlled by considering the distribution of the max statistic.

- If the max statistic is significant under the null we have a false positive.

- To study the max statistic all voxels need to be considered simultaneously.
  - Permutations carried out on the image level, i.e. entire images are relabeled.

Illustration





By repeatedly permuting the image labels and computing the max statistic we obtain the permutation distribution needed for correcting the FWER.

Comments

- For the test to be equally sensitive at all voxels, the sampling distribution should be roughly equal across voxels.

- Otherwise, areas where the statistic is highly variable tend to dominate the permutation distribution for the max statistic.

- The test will still be valid, but less sensitive at voxels with lower variability.

Pseudo t-statistic

- For small sample sizes (< 20) it makes sense to pool the variance estimate at each voxel with its neighbors to get a better variance estimate.

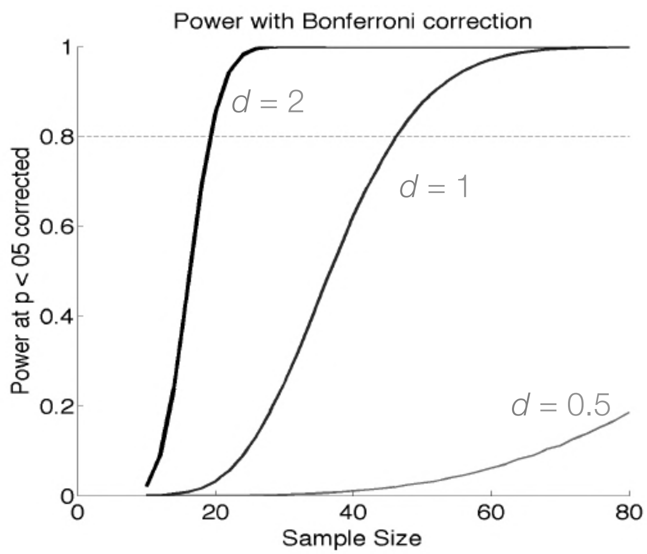
- The pseudo t-statistic images formed using the smoothed variance estimators are smoother than the standard t-statistic images.

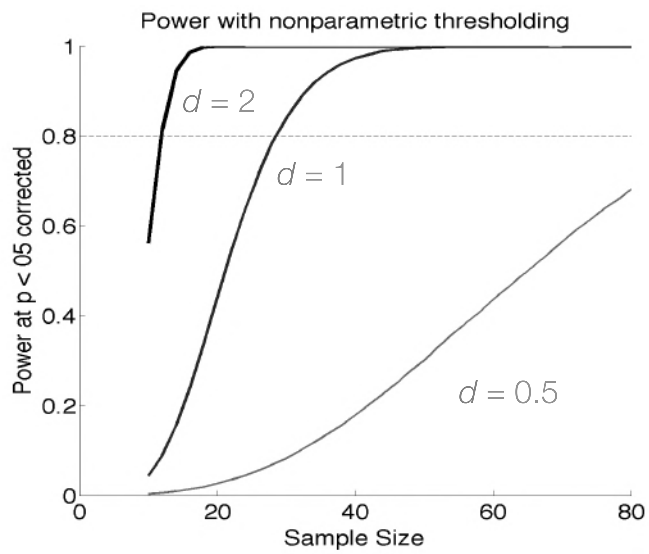
- It is difficult to obtain a parametric distribution for this statistic, but using non-parametric methods it is straightforward.

Power

- Bonferroni and SPM's GRF correction do not account accurately for spatial smoothness with the sample sizes and smoothness values typical in imaging experiments.

- Nonparametric tests more accurate and sensitive.





Wager et al., 2009

Classification

- Another area where non-parametric procedures have recently found wide usage is in MVPA.

- The bootstrap can be used to determine brain areas making reliable contributions to prediction.

- Permutation tests can be used to determine prediction accuracy.
  - Suppose we want to classify whether an observation belongs to either one of two classes.
  - Assuming no class information in the data, labels can be permuted without altering the expected accuracy.

Summary

- Nonparametric procedures are attractive tools as they require minimal assumptions for validity.
- They provide a flexible methodology for the statistical analysis of neuroimaging data.
- The Bootstrap and permutation tests have found wide usage in neuroimaging and should continue to gain in popularity.
  - Parallelization and GPUs allow for potential speed-up.