

Nonparametric Statistics in Neuroimaging

Martin Lindquist
Department of Biostatistics
Johns Hopkins University

Parametric Statistics



- Most statistical methods used in the analysis of neuroimaging data are **parametric** in nature.
 - The data is assumed to come from a population that follows a certain probability distribution with a fixed set of parameters.
 - Examples include the general linear model (GLM) and Gaussian random field theory.
- Parametric procedures rely on a number of assumptions whose validity impact the results.

Parametric Statistics



- 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



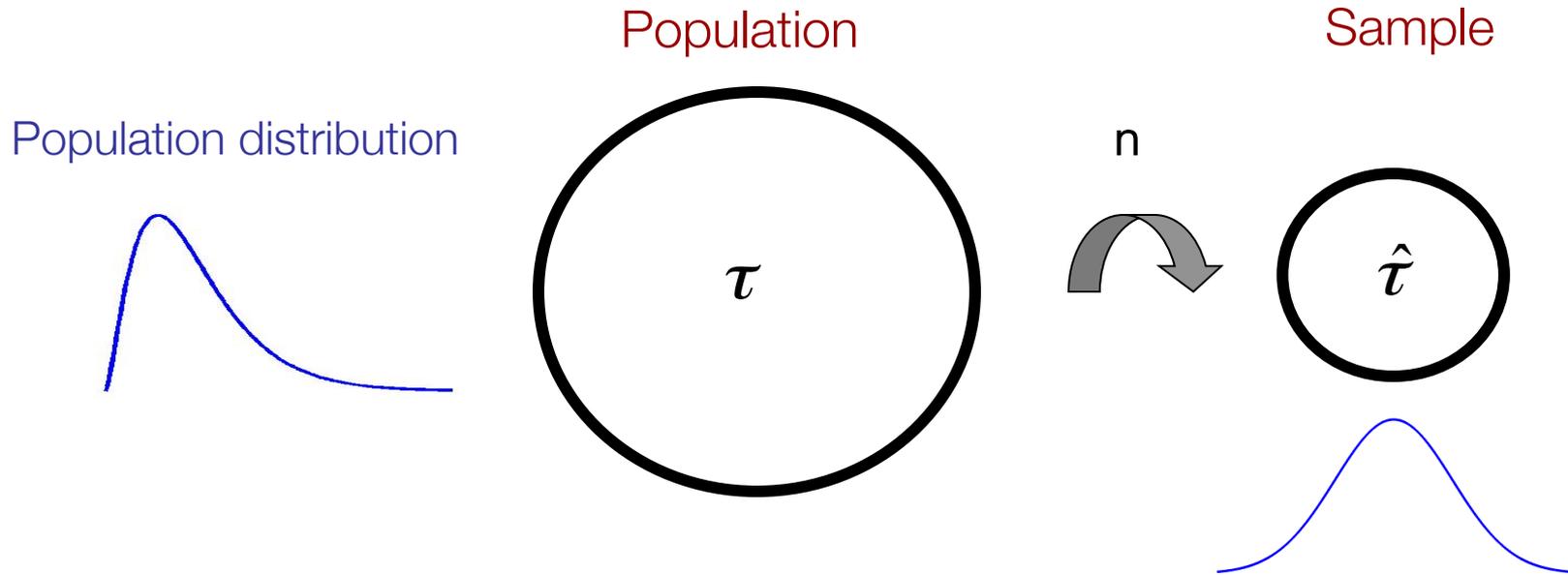
- The use of **computer intensive methods** to perform nonparametric inference has become increasingly popular in recent years.
- The most popular approaches include:
 - the bootstrap procedure, and
 - permutation tests
- Both techniques are based on randomly resampling the available data.

Statistical Inference Primer



- 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.

Statistical Inference Primer



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



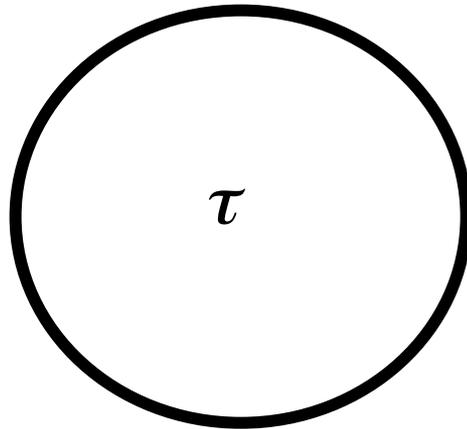
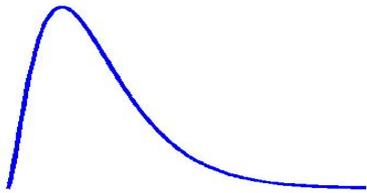
- The **bootstrap** is a computer-based method of inference that allows for estimation of the sampling distribution of almost any statistic.
- It can be used to construct confidence intervals for situations where traditional methods cannot (or should not) be used.
- By repeatedly **resampling with replacement** from the sample we approximate repeatedly sampling data from the population.

Illustration

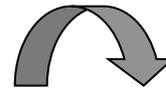


Population

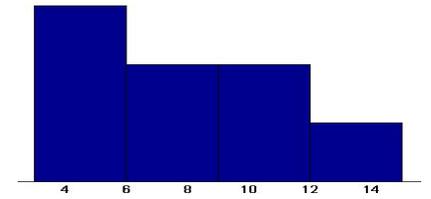
Population distribution



n



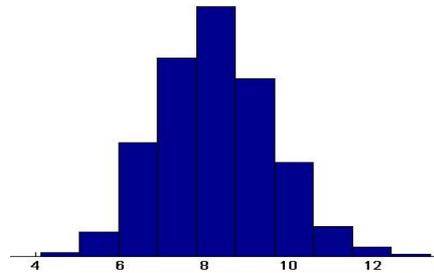
Data distribution



Take N bootstrap samples of size n



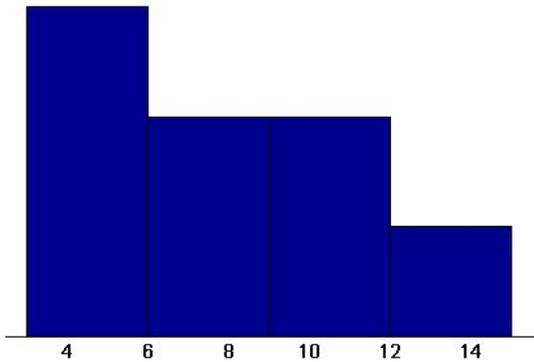
Bootstrap distribution



t_1, t_2, \dots, t_N

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

Data Distribution

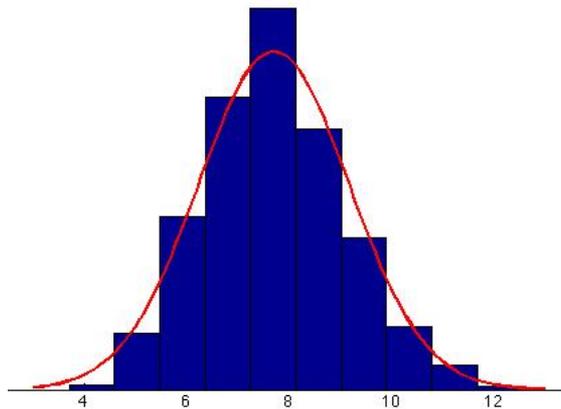


Take repeated resamples of size 7

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

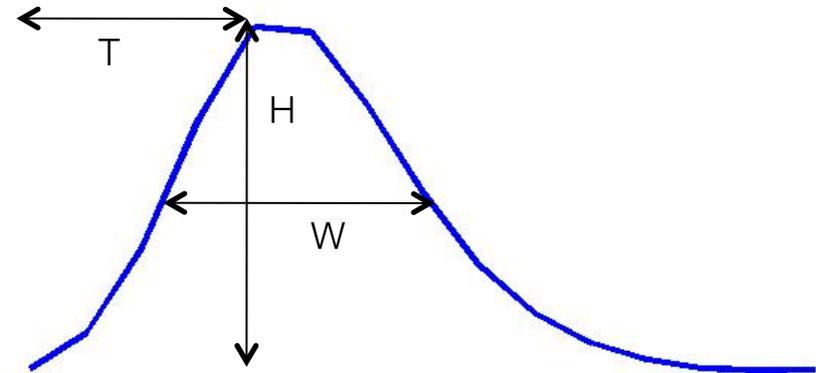


- The bootstrap has many uses in neuroimaging.
- Examples include:
 - Performing second-level inference when the first-level model uses basis functions to model the HRF.
 - Constructing confidence intervals for situations when the distribution of a statistic is unknown (e.g., mediation or dynamic correlation).

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.

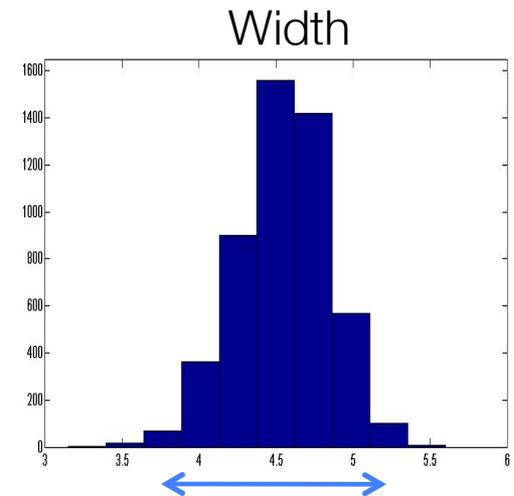
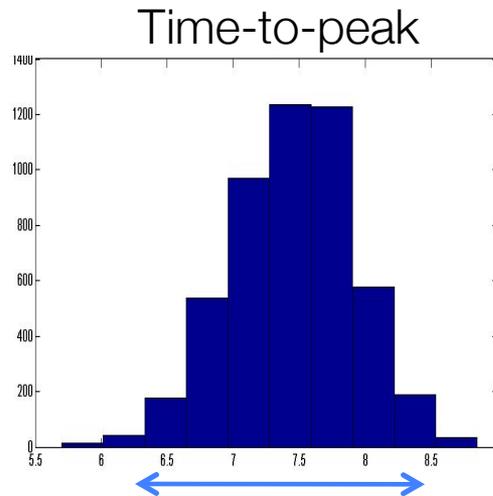
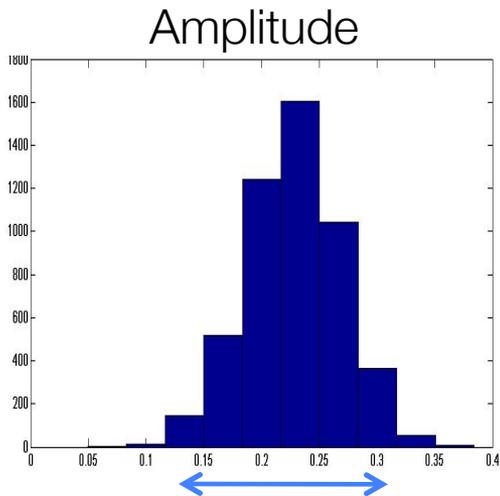


Bootstrap Procedure



Procedure:

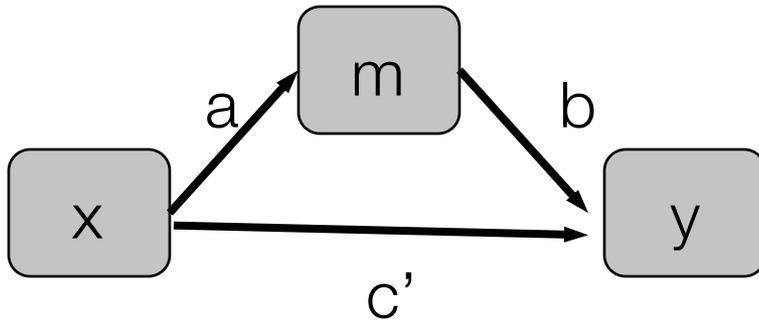
- 1) Sample among the n subjects with replacement.
- 2) Compute the statistics of interest (e.g., group mean of amplitude, time-to-peak, and width).
- 3) Repeat many times (e.g., 1,000-10,000 times).
- 4) Confidence intervals based on bootstrap distribution.



Mediation

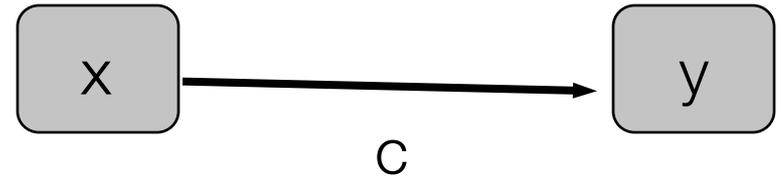


Full model, with mediator



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

Reduced model, without mediator



$$y = cx + e_y$$

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

$$c = c' + ab$$

Test the significance of the $a*b$ product

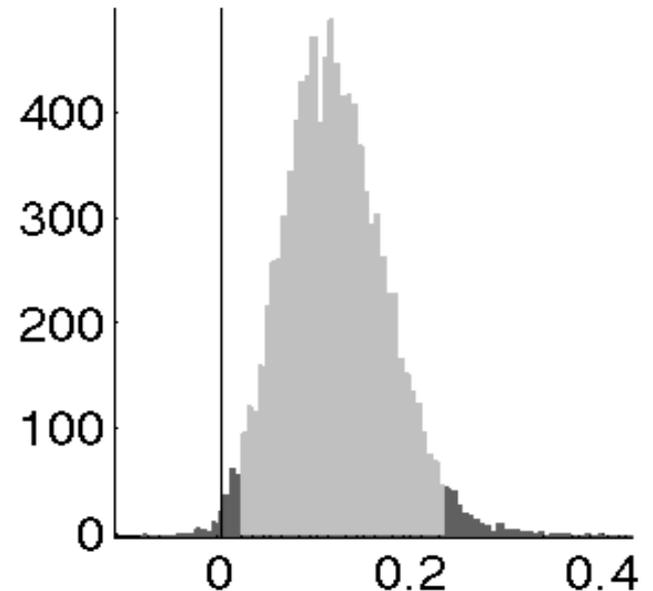
Bootstrap Procedure



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 based on bootstrap distribution.

Histogram of bootstrapped
Indirect ($a*b$) effects



Dependent Data



- In the two previous examples we performed the bootstrap across subjects.
- In certain cases (e.g., dynamic connectivity) we may want to bootstrap across time.
 - Difficult due to autocorrelations in the time course.
 - Need to use alternative methods that are more appropriate for dependent data.

Dependent Data

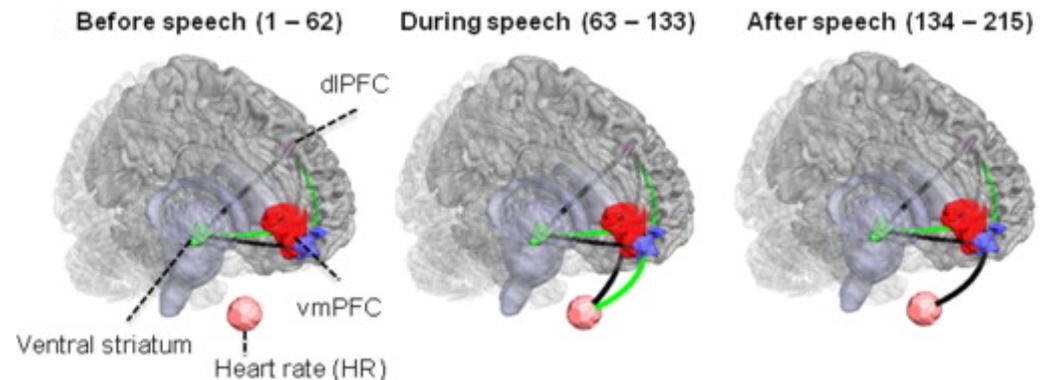
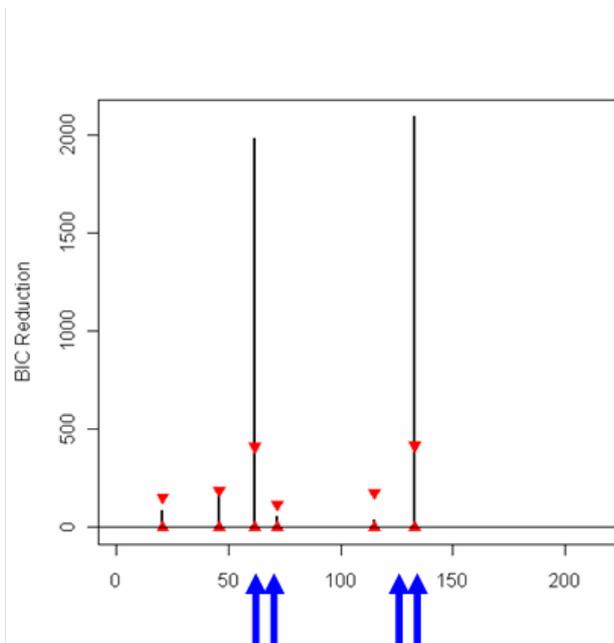


- The **block bootstrap** divides the time series into blocks and resamples with replacement these blocks to create a pseudo time series.
 - Successive time points are assumed to be correlated, but observations “far apart” uncorrelated.
 - By resampling sufficiently long blocks, the dependence structure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

Dynamic Connectivity



- 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

<u>Subject</u>	<u>Actual Treatment</u>
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 .

Example



- 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.

Example



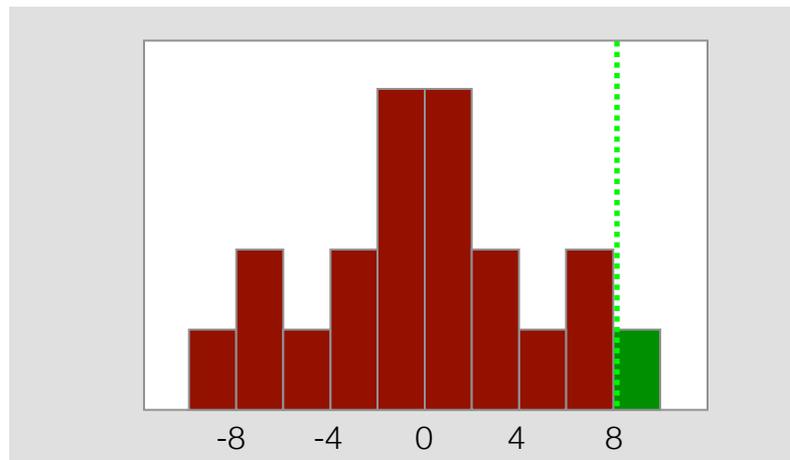
- Under H_0
 - Consider all equivalent re-labelings.
 - Assume exchangeability, i.e. the distribution of the statistic is the same whatever the relabeling.
 - Compute all possible statistic values.
 - Determine the permutation distribution.
 - Each relabeling is equally likely. Hence, each statistic 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

Example



- Permutation distribution



P-value = 0.05

Actual value = 9.45

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



- 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.
 - Special care needs to be taken when performing permutation tests on individual subjects.



- 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.



- Computational Intensity
 - Analysis needs to be repeated for each relabeling.
 - Computations can be parallelized.
 - Can use GPUs (graphics processing units).
- 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.

Applications in Neuroimaging



- Permutation tests have found wide usage in neuroimaging.
- They can be used in place of random field theory methods to correct for multiple comparisons.
 - The family-wise error rate (FWER) is often controlled by considering the distribution of the max statistic.
 - Use permutation methods to find this distribution.
- They are also often used to determine accuracy in predictive models.

Multiple Comparisons

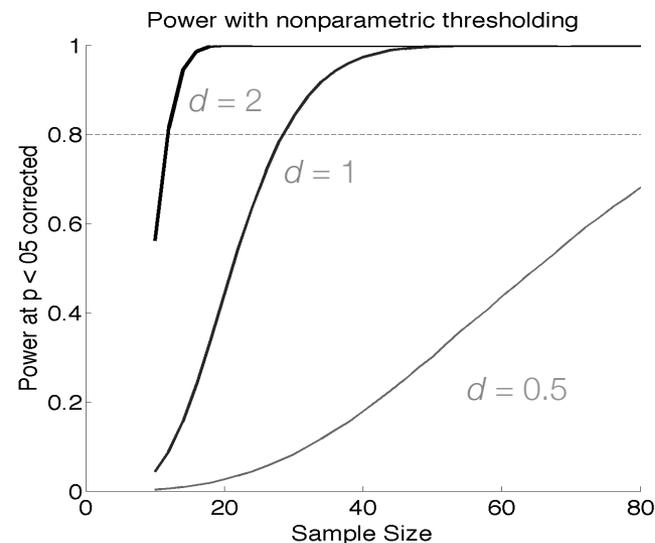
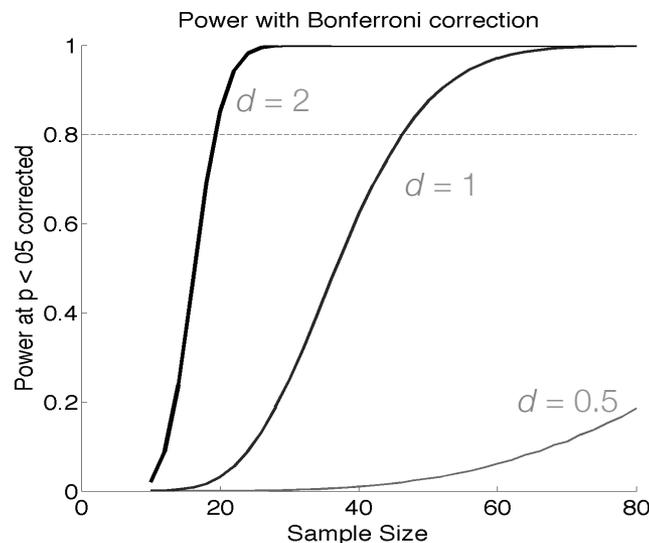


- One-sample test ($H_0: \beta=0$)
 - For each permutation randomly flip the sign of each subject's data and compute test statistics at each voxel.
 - Retain the max test statistic over the brain, creating the null maximum distribution needed for FWE inference.
 - Assumes the errors are symmetrically distributed.
- Two-sample test ($H_0: \beta_1=\beta_2$)
 - For each permutation randomly re-assign group labels (i.e., case/control) to subjects and compute test statistics at each voxel.
 - Retain the max test statistic over the brain, creating the null maximum distribution needed for FWE inference.

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.





- **Statistical nonparametric mapping** (SnPM; Nichols & Holmes 2002) is a nonparametric equivalent to SPM.
 - Uses permutation tests, rather than random field theory, to correct for multiple comparisons.
- **Randomise** (Winkler et al. 2014) is FSL's tool for nonparametric permutation inference.
 - Allows modeling and inference using standard GLM design setup.
 - Provides a mechanism for handling repeated measures and confound regressors.
- **Broccoli** (Eklund et al. 2014) also uses permutation tests at the second-level.

Classification



- Another area where non-parametric procedures have recently found wide usage is in multi-voxel pattern analysis (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.