Cluster Based Permutation Tests

Sabrina Sghirripa

Revisiting the Curious Case of Nicki Minaj's Cousins Friend in Trinidad

- NULL HYPOTHESIS → **vax = no vax** (i.e. there will be NO difference -Nicki's cousin's friend in Trinidad was talking smack)
- ALTERNATIVE HYPOTHESIS → **vax =/= no vax** (i.e. there *will* be a difference- Nicki's cousin's friend in Trinidad actually became impotent from the vaccine)

Errors

	ACCEPT Null	REJECT Null
Null is TRUE (no difference)	1-α Correct (probability of correct retention)	α (Type 1 error) FALSE POSITIVE
Null is FALSE (difference)	β (Type 2 error) FALSE NEGATIVE	1-β (power of the test)

- We need to keep the probability of a type 1 error low by setting a significance level: 0.05, 0.01, 0.001…
- The *p*-value is the smallest Type 1 error rate you are willing to tolerate if you want to reject the null hypothesis
- Power: probability with which we reject a null when it is actually false

Errors

	ACCEPT Null	REJECT Null
Null is TRUE (no difference)	Vaccine doesn't cause impotence	Vaccine doesn't cause impotence but we say it does (oh shit) FALSE POSITIVE
Null is FALSE (difference)	Vaccine does cause impotence, but we say it doesn't (oh shit) FALSE NEGATIVE	Vaccine causes impotence

The multiple comparisons problem

- If we run many tests, we increase our chance of encountering a type 1 error (false positive)
- EEG data is multi-dimensional
  - Channels
  - Times of interest
  - Frequencies of interest
- 62 channels across x number of times and y numbers of frequencies = WAY too many tests
  - How do we know the significant results aren’t just false positives?

The multiple comparisons problem

- Solution 1: p-value correction
  - Use a method like Bonferroni to adjust the p-value
  - Original p-value divided by number of tests performed
- Issue → makes it *really* hard to actually find a difference because the p-value becomes miniscule (i.e. reducing statistical power)
- Solution 2: take a region of interest, time of interest and/or frequency of interest and collapse over dimensions, then run conventional statistical analyses on the data
  - This is useful in some circumstances when your hypotheses are strong
- Issue → you might not capture the full story of your data this way

Cluster Based Permutation Tests

- Controls for the type 1 error rate when comparing across multiple channels etc.
- Two part process
  - Calculate cluster-based test statistic
  - Calculate the significance probability
- Null hypothesis: data observed in experimental conditions are drawn from the *same* probability distribution
- Alternative hypothesis: data observed in experimental conditions are drawn from *different* probability distributions

1. Calculate Cluster-Based Test Statistic

1. For every sample pair/triplet, run a t-test that quantifies the effect at this particular sample (e.g. channel/frequency/time)
2. Test whether the t-value is larger than a given threshold to determine whether the sample is a candidate for a cluster of samples
3. Selected samples are clustered in connected sets on the basis of temporal, spatial and spectral adjacency
4. Cluster-level statistics are calculated by taking the sum of t-values within every cluster
5. The maximum of the cluster-level statistic is taken and becomes the test statistic

2. Calculate Significance Probability

1. Put all the trials in each condition into a single set
2. Randomly draw trials from the combined data into two subsets (random partitioning)
3. Calculate the test-statistic (i.e. maximum of cluster level summed t-values) on the random partitions
4. Rinse and repeat many times and construct a histogram of the test statistics
5. Calculate the proportion of random partitions that resulted in a larger test-statistic than the one calculated in step 4
6. If the cluster-level statistic observed from the original data was larger in value than >95% of random partitions (i.e. p-value of 0.05), then the cluster is deemed significant

Interpretation

- If your cluster is significant, you can report that there's a difference between your experimental conditions
- You CANNOT say
  - *There was a significant cluster in area x during time y*
  - Why? Because the significant p-value only tells you the conditions differ. Not where or when they differ.
- If you actually do have an a-priori hypothesis about latency and spatial regions, you can say:
  - *After selecting the a-priori time and region of interest in our data, the cluster-based permutation tests revealed a difference between condition 1 and condition 2. In the (insert time range) differences were most pronounced over (insert electrode region)*