BMI/STAT-768
Statistical Methods for Medical Image Analysis

Moo K. Chung
Waisman Laboratory for Brain Imaging and Behavior
University of Wisconsin-Madison

www.stat.wisc.edu/~mchung
Permutation & Jackknife resampling methods

Hayasaka, S. Nichols, T.E. Validating cluster size inference: random field and permutation methods
NeuroImage 20:2343-2356

Permutation Method

• R.A. Fisher invented the method in 1935 in The Design of Experiments. This is the only exact method for computing p-value.

• Requires permutation invariance: Find units exchangeable under the null hypothesis.

Ex. Under null, two distributions are identical.

• There may be a situation we may not able to permute.
Permutation Method

• Does not assume any statistical distribution - nonparametric.

• For multiple comparisons correction: numerically estimate the distribution of the sup $T(t)$ under null.
Two sample example

• Group A (n subjects) vs. Group B (m subjects)
  Null hypothesis: equal distribution between groups.

• Mix all subjects together.

• Separate the mixed subjects into n subjects and m subjects.
  There are \( \binom{n+m}{n} \) possible permutations.

• Need to assume the joint distribution of each permutation is identical.

Small sample example

Group A (AAA) Group B (BBB)

# permutation = \( \binom{3+3}{3} \) = 20 combinations

1. AAABBB
2. AABABB
3. AABBBAB
4. AABBBBA
5. ABAABBB
6. ABABAB
7. ABABBA
8. ABBAABB
9. ABBABA
10. ABBBBAA
11. BAAABB
12. BAABAB
13. BAABBA
14. BABAAB
15. BABABA
16. BABBAA
17. BBAABA
18. BBAABA
19. BBABAA
20. BBBAAA
• Compute the $\sup T(t)$ for each permutation.

• Construct histogram of $\sup T(t)$ and threshold at 5% level.
Permutation test - model free statistical inference
Not recommended for most studies.

More than 1500 permutations are needed to guarantee the convergence.
8 hours of running time in MATLAB.
MATLAB Demo
For a group with $n$ subjects, one subject is removed and the remaining $n-1$ subjects are used in computing statistics (leave-one-out scheme).

This process is repeated for each subject to produce $n$ statistics.
Jackknife estimation

Leave-one-out mean

\[ \mu_{(i)} = \frac{1}{n-1} \sum_{j \neq i} X_j = \frac{n \bar{X} - X_i}{n-1} \]

Jackknife estimate of mean

\[ \mu_{(\cdot)} = \frac{1}{n} \sum_{i=1}^{n} \mu_{(i)} = \frac{n}{n-1} \bar{X} - \frac{1}{n-1} \bar{X} = \bar{X} \]
Example: Correlation between gill weight and body weight in 12 crabs

Sample correlation

\[ r = 0.865 \]

Jackknife

\[ r = 0.878 \]

<table>
<thead>
<tr>
<th>Gill (mg)</th>
<th>Body (g)</th>
<th>( r_i )</th>
<th>( \varphi_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1590</td>
<td>14.40</td>
<td>0.888</td>
<td>0.607</td>
</tr>
<tr>
<td>1790</td>
<td>15.20</td>
<td>0.884</td>
<td>0.656</td>
</tr>
<tr>
<td>10</td>
<td>11.30</td>
<td>0.892</td>
<td>0.570</td>
</tr>
<tr>
<td>450</td>
<td>2.50</td>
<td>0.830</td>
<td>1.249</td>
</tr>
<tr>
<td>3840</td>
<td>22.70</td>
<td>0.811</td>
<td>1.452</td>
</tr>
<tr>
<td>23</td>
<td>14.90</td>
<td>0.863</td>
<td>0.879</td>
</tr>
<tr>
<td>10</td>
<td>1.41</td>
<td>0.875</td>
<td>0.751</td>
</tr>
<tr>
<td>32</td>
<td>15.81</td>
<td>0.872</td>
<td>0.779</td>
</tr>
<tr>
<td>8</td>
<td>4.19</td>
<td>0.845</td>
<td>1.078</td>
</tr>
<tr>
<td>22</td>
<td>15.39</td>
<td>0.867</td>
<td>0.843</td>
</tr>
<tr>
<td>32</td>
<td>17.25</td>
<td>0.858</td>
<td>0.940</td>
</tr>
<tr>
<td>21</td>
<td>9.52</td>
<td>0.877</td>
<td>0.725</td>
</tr>
</tbody>
</table>
Bar code on the first Betti number

24 attention deficit hyperactivity disorder (ADHD) children
26 autism spectrum disorder (ASD) children
11 pediatric control subjects

1 correlation matrix per group
1 barcode per group

Jackknife is used to generate 1 barcode per subject

Lee et al. 2011 MICCAI
Gromov-Hausdorff distance between networks

\[ d_{GH} = \frac{1}{2} \max_{i,j} |\rho_{1ij} - \rho_{2ij}| \]

Jacknife on 24, 26, 11 subjects

(24+26+11) x (24+26+11)
Gromov-Hausdorff distance map
Example: Jacobian determinant network

One barcode per subject
→ resampling to obtain empirical distribution
Jackknife within each group

Inference done similar to KS-test procedure: 

$p$-value = 0.0001
Bootstrap resampling

- Introduced by Efron in 1979.
- Motivated by Jackknife.
- A “bootstrap” data set is one created by randomly selecting $n$ points from the training set $D$, with replacement.
- In bootstrap estimation, this selection process is independently repeated $B$ times to yield $B$ bootstrap data sets.
The Bootstrap

• Data $D = X_1, X_2, X_3, \ldots, X_n \rightarrow$ statistic $s$

• Bootstrap replicate:
  - $D^*1 = X^*_1, X^*_2, X^*_3, \ldots, X^*_n \rightarrow$ statistic $s^*1$
  - $D^*2 = X^*_1, X^*_2, X^*_3, \ldots, X^*_n \rightarrow$ statistic $s^*2$
  - ...

• $X^*_1, X^*_2, X^*_3, \ldots, X^*_n$ are randomly selected with replacement, from $X_1, X_2, X_3, \ldots, X_n$

• Usually use 100-10,000 bootstrap replicates and obtain the empirical distribution
Example: *Median Gill Weight in Crabs*

Gill weights (in mg):
159 179 100 45 384 230 100 320 80 220 320 210

Median = 195mg

<table>
<thead>
<tr>
<th>Real</th>
<th>159</th>
<th>179</th>
<th>100</th>
<th>45</th>
<th>384</th>
<th>230</th>
<th>100</th>
<th>320</th>
<th>80</th>
<th>220</th>
<th>320</th>
<th>210</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Median</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Bootstrap replicates:**

<table>
<thead>
<tr>
<th>B1</th>
<th>320</th>
<th>159</th>
<th>45</th>
<th>320</th>
<th>100</th>
<th>320</th>
<th>100</th>
<th>320</th>
<th>100</th>
<th>230</th>
<th>100</th>
<th>210</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2</td>
<td>384</td>
<td>384</td>
<td>45</td>
<td>384</td>
<td>100</td>
<td>80</td>
<td>45</td>
<td>179</td>
<td>230</td>
<td>230</td>
<td>205</td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>159</td>
<td>320</td>
<td>80</td>
<td>45</td>
<td>45</td>
<td>80</td>
<td>220</td>
<td>210</td>
<td>230</td>
<td>320</td>
<td>220</td>
<td>215</td>
</tr>
<tr>
<td>B4</td>
<td>220</td>
<td>179</td>
<td>384</td>
<td>100</td>
<td>80</td>
<td>100</td>
<td>230</td>
<td>230</td>
<td>179</td>
<td>230</td>
<td>384</td>
<td>45</td>
</tr>
<tr>
<td>B5</td>
<td>320</td>
<td>220</td>
<td>210</td>
<td>100</td>
<td>159</td>
<td>320</td>
<td>220</td>
<td>210</td>
<td>100</td>
<td>80</td>
<td>100</td>
<td>210</td>
</tr>
<tr>
<td>B6</td>
<td>80</td>
<td>100</td>
<td>230</td>
<td>100</td>
<td>210</td>
<td>384</td>
<td>159</td>
<td>220</td>
<td>320</td>
<td>45</td>
<td>45</td>
<td>210</td>
</tr>
<tr>
<td>B7</td>
<td>179</td>
<td>210</td>
<td>80</td>
<td>320</td>
<td>100</td>
<td>230</td>
<td>159</td>
<td>320</td>
<td>100</td>
<td>45</td>
<td>384</td>
<td>320</td>
</tr>
<tr>
<td>B8</td>
<td>384</td>
<td>159</td>
<td>100</td>
<td>159</td>
<td>100</td>
<td>179</td>
<td>100</td>
<td>179</td>
<td>220</td>
<td>384</td>
<td>220</td>
<td>159</td>
</tr>
<tr>
<td>B9</td>
<td>320</td>
<td>210</td>
<td>45</td>
<td>320</td>
<td>179</td>
<td>159</td>
<td>100</td>
<td>210</td>
<td>159</td>
<td>45</td>
<td>210</td>
<td>100</td>
</tr>
</tbody>
</table>

\[\ldots\]

Empirical distribution
Measure of complexity
Fractal dimension
Network complexity