Journal Club: Higher Criticism

David Donoho (2002):

Higher Criticism for Heterogeneous Mixtures,

Technical Report No. 2002-12, Dept. of Statistics, Stanford University.

Introduction

- John Tukey (1976):
  Observe 11 significant tests (level 5%) out of 250.
  Cool???

Introduction

- John Tukey (1976):
  Observe 11 significant tests (level 5%) out of 250.
  Cool???

- Not really:
  You can expect 12.5 significant tests by chance.
  11 are rather disappointing.

- Tukey introduced the expression Higher Criticism.

What is a Higher Criticism?

- The expression Higher Criticism originates from bible exegesis.

- `www.familybiblestudy.net`:

  "The ordinary study or criticism is directed to finding out the meaning of the passages, their correct translation and their significance and bearing on doctrines.

  The higher critics go above and back of all that, applying to the books of the Bible the same tests and methods of examination as are applied to other ancient books. [...]

  The best scholars of the present day believe that many of the conclusions reached by the higher critics are erroneous, and that others are mere guesses for which there is not sufficient evidence."

Higher Criticism or Second-level significance testing

$$HC_{.05,n} = \sqrt{n} \quad \frac{(\text{Fraction Significant at .05}) - .05}{\sqrt{.05 \times .95}}$$

Higher Criticism or Second-level significance testing

$$HC_{.05,n} = \sqrt{n} \quad \frac{(\text{Fraction Significant at .05}) - .05}{\sqrt{.05 \times .95}}$$

$HC_{.05,n} > 2$: significance of the overall body of tests

Higher Criticism or Second-level significance testing

$$HC_{.05,n} = \sqrt{n} \quad \frac{(\text{Fraction Significant at .05}) - .05}{\sqrt{.05 \times .95}}$$

$HC_{.05,n} > 2$: significance of the overall body of tests

$$HC_n^\star = \max_{0<\alpha\leq\alpha_0} \sqrt{n} \quad \frac{(\text{Fraction Significant at } \alpha) - \alpha}{\sqrt{\alpha \times (1-\alpha)}}$$

"Sparse normal means" testing problem

H_0 : n normal means are all 0 vs. H_1 : a small fraction is $\neq 0$

"Sparse normal means" testing problem

H_0 : n normal means are all 0 vs. H_1 : a small fraction is $\neq 0$

- Early detection of bioweapons use
- Detection of Covert Communications
- Meta-analysis with heterogeneity
- (Microarray data)

The Model

H_0 : n normal means are all 0 vs. H_1 : a small fraction is $\neq 0$

$$H_0 : X_i \sim N(0,1) \quad \text{vs.} \quad H_1 : X_i \sim (1-\varepsilon)N(0,1) + \varepsilon N(\mu, 1)$$

The Model

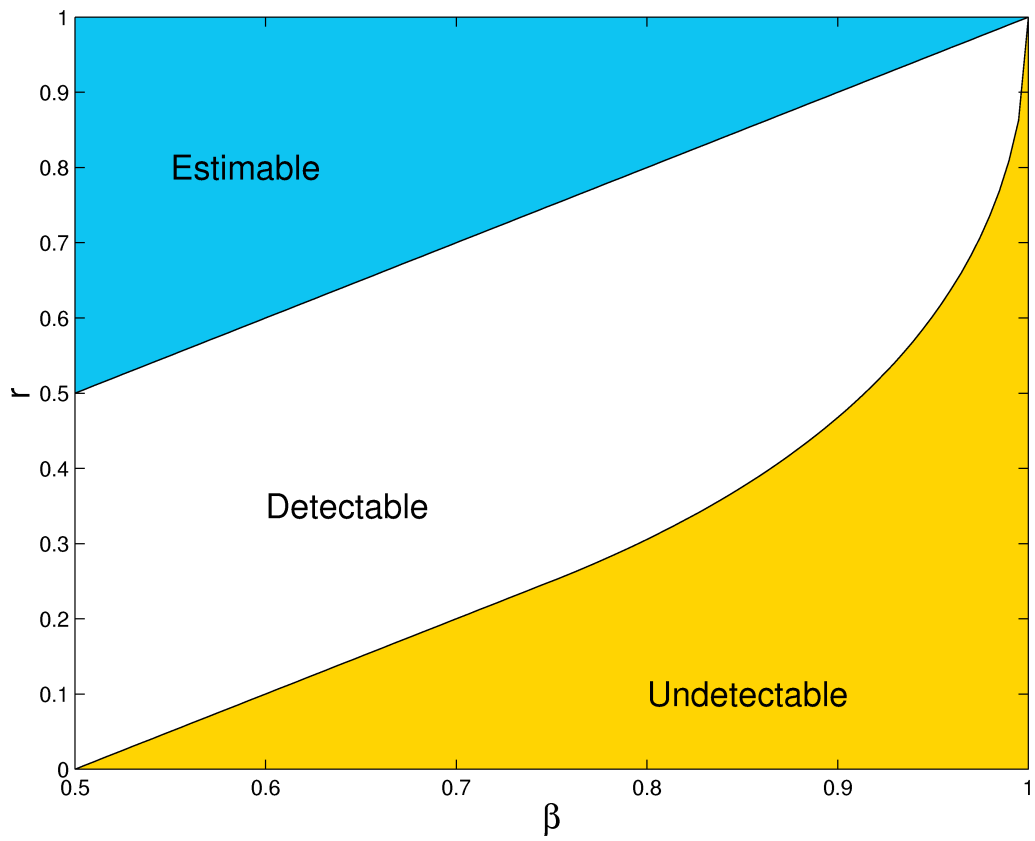
H_0 : n normal means are all 0 vs. H_1 : a small fraction is $\neq 0$

$$H_0 : X_i \sim N(0,1) \quad \text{vs.} \quad H_1 : X_i \sim (1-\varepsilon)N(0,1) + \varepsilon N(\mu, 1)$$

Choose the sparsity:

- Fraction $\varepsilon = n^{-\beta}$ with $\beta \in (1/2, 1)$, that means $\varepsilon \in \left(\frac{1}{n}, \frac{1}{\sqrt{n}}\right)$
- Normal mean $\mu = \sqrt{2r\log(n)}$ with $r \in (0,1)$

Detection boundary



Detection boundary

- For the model above:

$$\rho^\star(\beta) = \begin{cases} \beta - \frac{1}{2} & 0.5 < \beta < 0.75 \\ (1 - \sqrt{1-\beta})^2 & 0.75 \leq \beta < 1 \end{cases}$$

- In the region $r > \rho^\star(\beta)$ the Likelihood Ratio Test is optimal and detects the alternative reliably if r and β are known.

- If r and β are unknown, the Higher Criticism performs well.

Performance of Higher Criticism

- Main theorems:
  For $r > \rho^\star(\beta)$ the Higher Criticism has full power and completely separates the two hypotheses asymptotically.

- This holds also in other settings like Exponential, χ^2 and General Normal Distribution.

- Simulation study: Detection near the border needs large samples.

Comparison to Multiple Testing Procedures

- Maximum statistic $M_n = \max_i X_i$ is equivalent to taking the minimum of p-values (Bonferroni).

- Detection boundary:

$$\rho_M(\beta) = (1 - \sqrt{1-\beta})^2, \quad \beta \in (0.5, 1).$$

- The detection boundary for M_n is equal to the boundary for $HC^\star$ for very sparse alternatives, $\beta \in [0.75, 1)$.

- For $\beta \in (0.5, 0.75)$ it exceeds the boundary for $HC^\star$
  $\Rightarrow$ In that range M_n can be "dramatically outperformed" by $HC^\star$.

Comparison to Multiple Testing Procedures

- False Discovery Rate (Benjamini & Hochberg, 1995):
  Denote by $H_0^{(i)}$ the null hypothesis corresponding to the ith ordered p-value $p_{(i)}$.
  The FDR is controlled at level α by the procedure:

  Let k be the largest i for which $p_{(i)} \leq \frac{i}{n}\alpha$

  and reject all $H_0^{(i)}$ with $i = 1, \ldots, k$.

- The detection boundary is the same as for the maximum statistic and is therefore inefficient for $\beta < 0.75$.

Fisher's method

- Classical approach to combine p-values for overall test of significance:

$$F_n = -2 \sum_i \log(p_i) \overset{asy.}{\sim} \chi^2_n.$$

- Works well in case of homogeneity.

- In case of extreme heterogeneity:
  If $\varepsilon = n^{-\beta}, \beta > \frac{1}{2}$ and $\mu \leq \sqrt{2\log(n)}$, asymptotically F_n is unable to separate H_1 and H_0.

Summary

- The Higher Criticism is able to separate H_0 and H_1 throughout the same detection region as the Likelihood Ratio Test but needs no specification of the sparsity.
- Reaches "deeper" detection regions than other multiple testing procedures.
- Might also be used in microarray analysis.