Alpha-Investing

Sequential Control of Expected False Discoveries

Dean Foster   Bob Stine

Department of Statistics
Wharton School of the University of Pennsylvania
www-stat.wharton.upenn.edu/∼stine

Joint Statistics Meeting, Salt Lake City
August 2007

Overview

Opportunities for Using Domain Knowledge in Testing

Situations in applications

- Clinical trial
  - Choice of secondary hypotheses to test in a clinical trial depends on the outcome of the primary test.
- Variable selection
  - Pick interactions to add to a regression model after detect interesting main effects (select from p rather than p^2).
- Data preparation
  - Construct retrieval instructions for extraction from database.
  - Geographic search over region based on neighbors.

Sequential decisions

- Choice of next action depends on what has happened so far.
- Maintain control chance for false positive error.

Keeping Track of a Sequence of Tests and Errors

- Collection of m **null** hypotheses

$$H_1, H_2, \ldots, H_m, \ldots$$

specify values of parameters θ_j ($H_j : \theta_j = 0$).

- Tests produce p-values $p_1,\ p_2, \ldots, p_m, \ldots$
- Reject H_j if p_j is smaller than α_j

$$R(m) = \sum_j R_j, \quad R_j = \begin{cases} 1 & \text{if } p_j < \alpha_j \\ 0 & \text{otherwise} \end{cases}$$

- How to control the **un**observed number of incorrect rejections?

$$V^{\theta}(m) = \sum V_j^{\theta}, \quad V_j^{\theta} = \begin{cases} 1 & \text{if } p_j < \alpha_j \text{ but } \theta_j = 0 \\ 0 & \text{otherwise} \end{cases}$$

Several Criteria Are Used to Control Error Rates

- Family-wise error rate, the probability for **any** incorrect rejection

$$FWER(m) = P(V^\theta(m) > 0)$$

Conservative when testing 1,000s of tests.

- False discovery rate, the expected proportion of false rejections among the rejected hypotheses

$$FDR(m) = E\left(\frac{V^\theta(m)}{R(m)} | R(m) > 0\right) P(R(m) > 0)$$

Less conservative with larger power.

- Marginal false discovery rate, the ratio of expected counts

$$mFDR_\eta(m) = \frac{E\ V^\theta(m)}{E\ R(m) + \eta}$$

Typically set $\eta = 1 - \alpha \approx 1$. (Convexity: $FDR \geq mFDR$)

Batch Procedures Vary the Level α_j

- "Batch" procedures have all m p-values at the start.
- Bonferroni (alpha-spending) controls $FWER(m) < \alpha$.

$$\text{Reject } H_j \text{ if } p_j < \alpha/m$$

- Benjamini-Hochberg "step-down" procedure (BH) controls $FDR(m) < \alpha$ for independent tests (and some dependent tests). For the ordered p-values $p_{(1)} < p_{(2)} < \cdots < p_{(m)}$

$$\text{Reject } H_{(j)} \text{ if } p_{(j)} < j\alpha/m$$

- Weighted BH procedure (wBH, Genovese et al, 2006) controls $FDR(m) < \alpha$ using *a priori* information to weight tests.

$$\text{Reject } H_{(j)} \text{ if } p_{(j)} < W_{(j)}\, j\, \alpha/m$$

More power: $W_j > 1$ for false nulls, else $W_j < 1$.

Alpha-Investing Resembles Alpha-Spending

- Initial alpha-wealth to "invest" in testing $\{H_j\}$

$$W(0) = \alpha$$

- Alpha-investing rule determines level for test of H_j, possibly using outcomes of prior tests

$$\alpha_j = \mathcal{I}_{W(0)}(\{R_1, R_2, \ldots, R_{j-1}\})$$

- Difference from alpha-spending:
  Rule earns more alpha-wealth when it rejects a null hypothesis

$$W(j) - W(j-1) = \begin{cases} \omega & \text{if } p_j \le \alpha_j \,, \\ -\alpha_j/(1-\alpha_j) & \text{if } p_j > \alpha_j \,. \end{cases}$$

  Earns payout ω if rejects H_j; pays $\alpha_j/(1-\alpha_j)$ if not.

Examples of Policies for Alpha-Investing Rules

- Aggressive policy anticipates clusters of $\theta_j \neq 0$
  - Investing rule: If last rejected hypothesis is H_{k^*}, then

$$\mathcal{I}_{W(0)}(\{R_1, R_2, \ldots, R_{j-1}\}) = \frac{W(j-1)}{1+j-k^*}, \qquad j > k^*$$

  - Invest most immediately after reject H_{k^*}:
    Invest $\frac{1}{2}$ of current wealth to test H_{k^*+1}
    Invest $\frac{1}{3}$ of current wealth to test H_{k^*+2}
    ...
- Revisiting policy mimics BH step-down procedure
  - Test every hypothesis first at level α/m.
  - If reject at least one, alpha-wealth remains $\geq W(0)$.
  - Test remaining hypotheses conditional on $p_j > \alpha/m$.
  - Rejects H_j if $p_j \leq 2\,\alpha/m$ (like BH).
  - Continue while at least one is rejected until wealth is spent.

Theory: Alpha-Investing Uniformly Controls mFDR

- Stop early: Do you care about every hypothesis that's rejected, or are you most interested in the first few?
  - Scientist studies first 10 genes identified from micro-array.
  - What is FDR when stop early?
- Uniform control of mFDR
  A test procedure *uniformly controls mFDR$_\eta$* at level α if for any finite stopping time T,

$$\sup_\theta \frac{E_\theta\left(V^\theta(T)\right)}{E_\theta\left(R(T)\right)+\eta} < \alpha$$

Theorem

An alpha-investing rule $\mathcal{I}_{W(0)}$ with initial alpha-wealth $W(0) \le \alpha\,\eta$ and pay-out $\omega \le \alpha$ uniformly controls mFDR$_\eta$ at level α.

Why control mFDR rather than FDR?

$$FDR(m) \approx E\left(\frac{V^\theta(m)}{R(m)}\right) \qquad mFDR_\eta(m) = \frac{E\ V^\theta(m)}{E\ R(m) + \eta}$$

- They produce similar control in the type of problems we consider, as shown in simulation. See simulation results
- By controlling a ratio of means, we are able to identify a martingale:

Lemma

The process

$$A(j) = \alpha R(j) - V^\theta(j) + \eta\,\alpha - W(j)$$

is a sub-martingale

$$E(A(j) \mid A(j-1), \ldots, A(1)) \geq A(j-1)\ .$$

Two Simulations of Alpha-Investing

Comparison to batch

- Fixed collection of hypotheses $H_1, \ldots, H_{200}$
- $H_j : \mu_j = 0$
- Spike and slab mixture, iid sequence
  $$\mu_j = \begin{cases} N(0, 2\log m) \\ 0 \end{cases}$$
- 10,000 replications

Testing an infinite stream

- Infinite sequence of hypotheses $H_1, \ldots, H_{4000}, \ldots$
- $H_j : \mu_j = 0$
- Hidden Markov chain
  - 10% or 20% $\mu_j = 3$
  - Average length of cluster varies
- 1,000 replications, halted at 4,000 tests

Procedures That Use Domain Knowledge

Oracle-based Weighted BH

- Oracle reveals which hypotheses to test
- Only test $m - m_0$ that are false
- Threshold for p-values

$$j\,\alpha/m \quad \Rightarrow \quad j\alpha/(m - m_0)$$

- Spread available alpha-level over fewer hypotheses

Alpha-investing

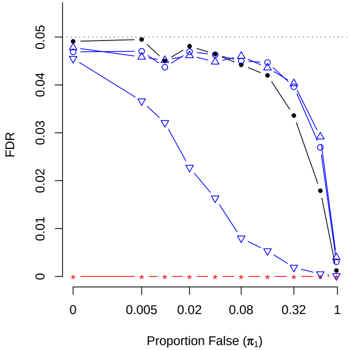
- Scientist able to order hypotheses by μ_j
- Test them all, but start with false
- Aggressive investing

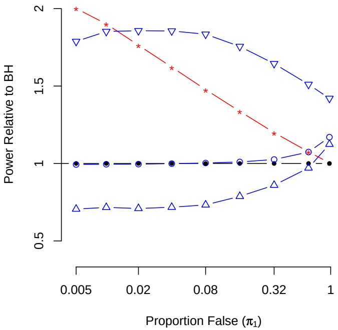
$$\alpha/2 \quad \Rightarrow \quad (\alpha + \omega)/2$$

- Initial rejections produce alpha-wealth for subsequent tests.

Alpha-Investing+Order Outperforms wBH+Oracle

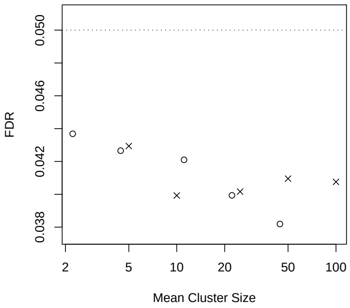
- Test $m = 200$ hypotheses, $\mu_j \sim$ spike–and–slab mixture
- Step-down: BH, wBH with oracle,
- Alpha-investing: aggressive($\bigtriangledown$, $\triangle$), mimic BH ($\circ$)

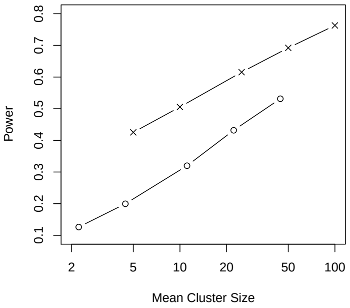




Testing an Infinite Stream of Hypotheses

- Generate μ_j from Markov chain
  - 10% ($\circ$) or 20% ($\times$) non-zero means
  - Fixed alternative: $\mu_j = 0$ or 3
- 1,000 sequences of hypotheses, snapshot at 4,000 tests
- Investing rule: Aggressive alpha-investing





Summary

Alpha-investing ...

- Allows testing of a dynamically chosen, infinite stream of hypotheses
- Underlying martingale proves alpha-investing obtains uniform control of mFDR ($\approx$ FDR)
- Exploits domain knowledge to improve power of tests
- Further details in paper at

  `stat.wharton.upenn.edu/~stine`

- What's next?
  - Applications in variable selection
  - Universal policies for alpha-spending

FDR and mFDR Produce Similar Types of Control

Simulation of tests

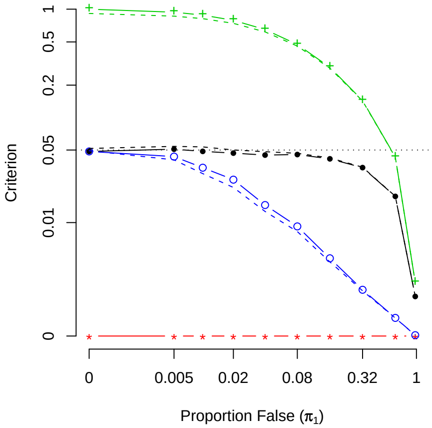
- $m = 200$ hypotheses
- Proportion π_1 false
- Spike–and–slab mixture

$$\mu_j = \begin{cases} N(0, 2\log m) \\ 0 \end{cases}$$

- 10,000 replications

Procedures

- Naive, Bonferroni, BH step-down, wBH with oracle
- Solid: FDR
  Dashed: mFDR



Return