False Discovery Control: Exact and Large-sample Approaches

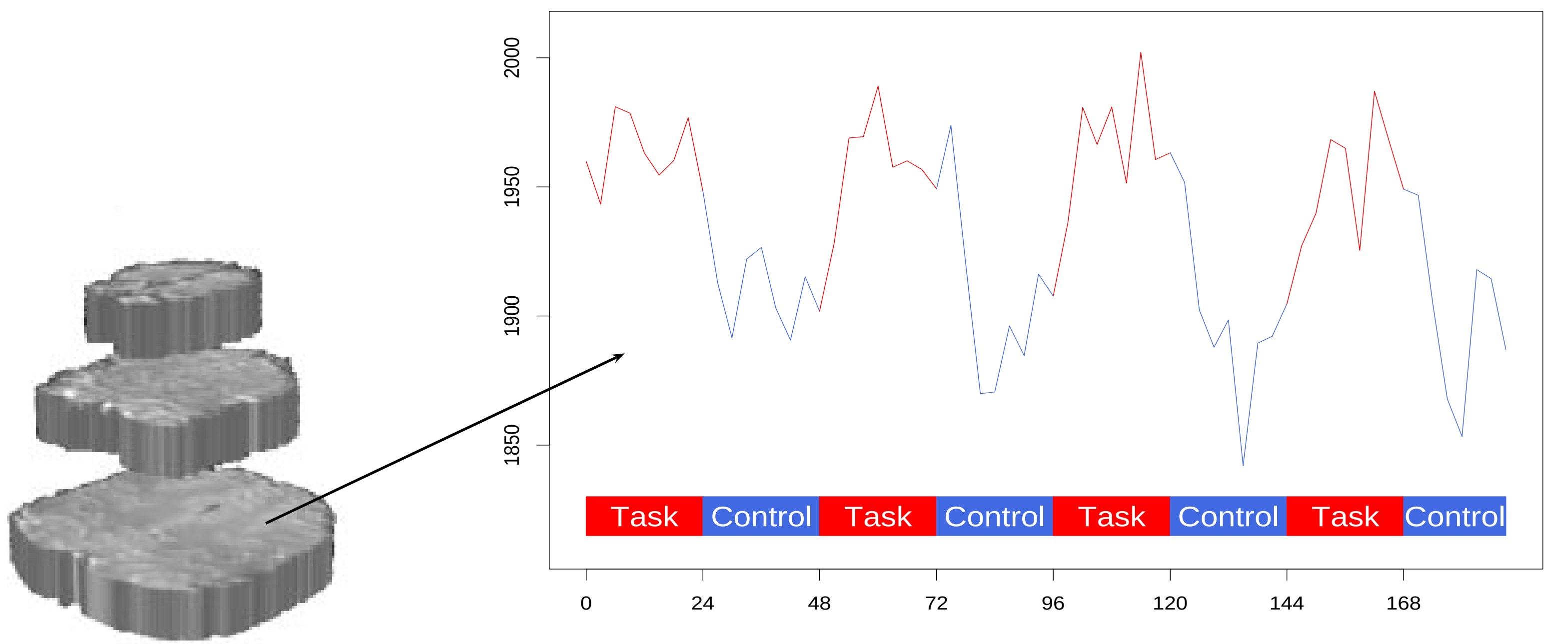
Christopher R. Genovese
Department of Statistics
Carnegie Mellon University
http://www.stat.cmu.edu/~genovese/

Larry Wasserman
Department of Statistics
Carnegie Mellon University

This work partially supported by NSF Grant SES 9866147.

Motivating Example #1: fMRI

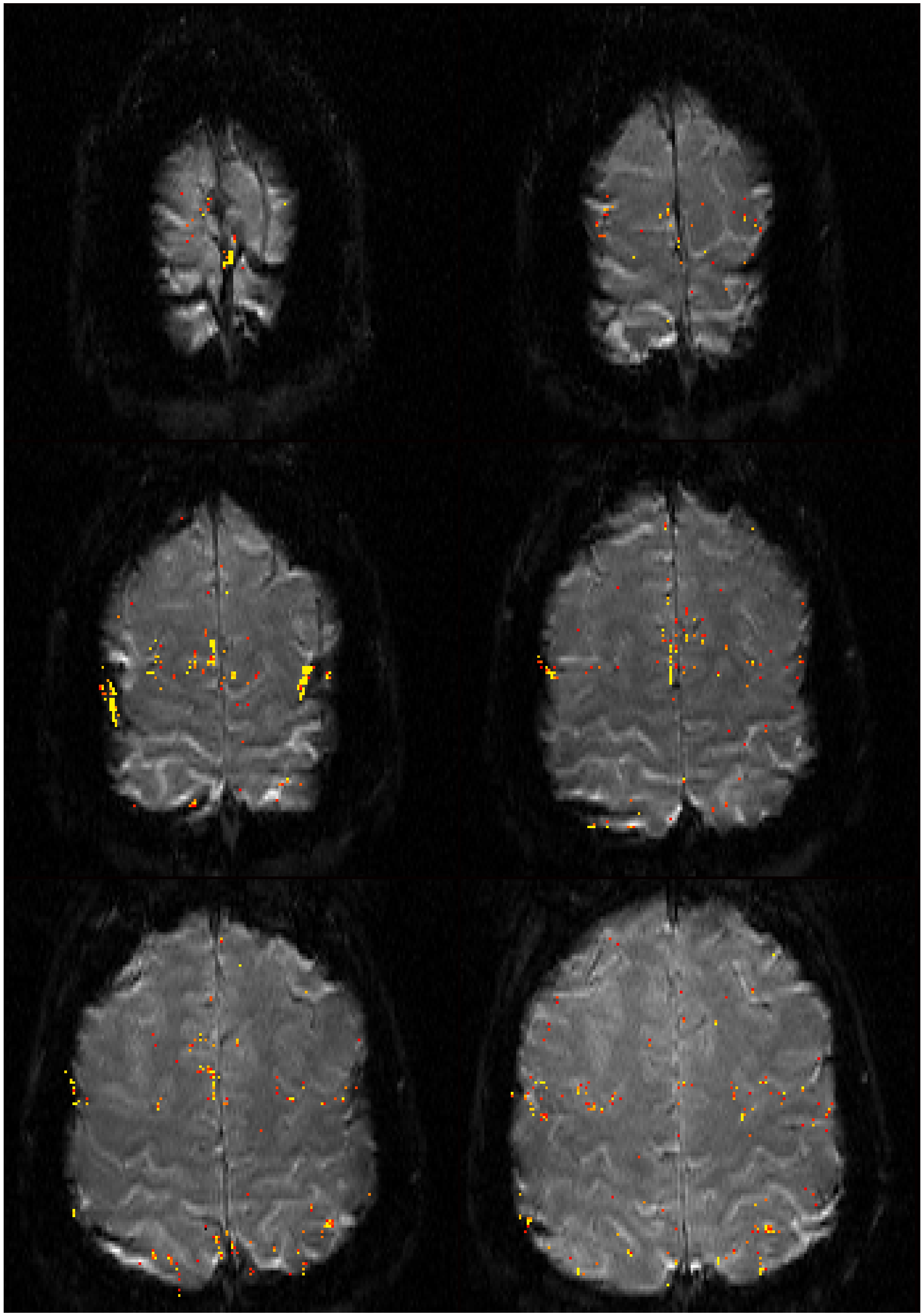
- fMRI Data: Time series of 3-d images acquired while subject performs specified tasks.



- Goal: Characterize task-related signal changes caused (indirectly) by neural activity. [See, for example, Genovese (2000), *JASA* 95, 691.]

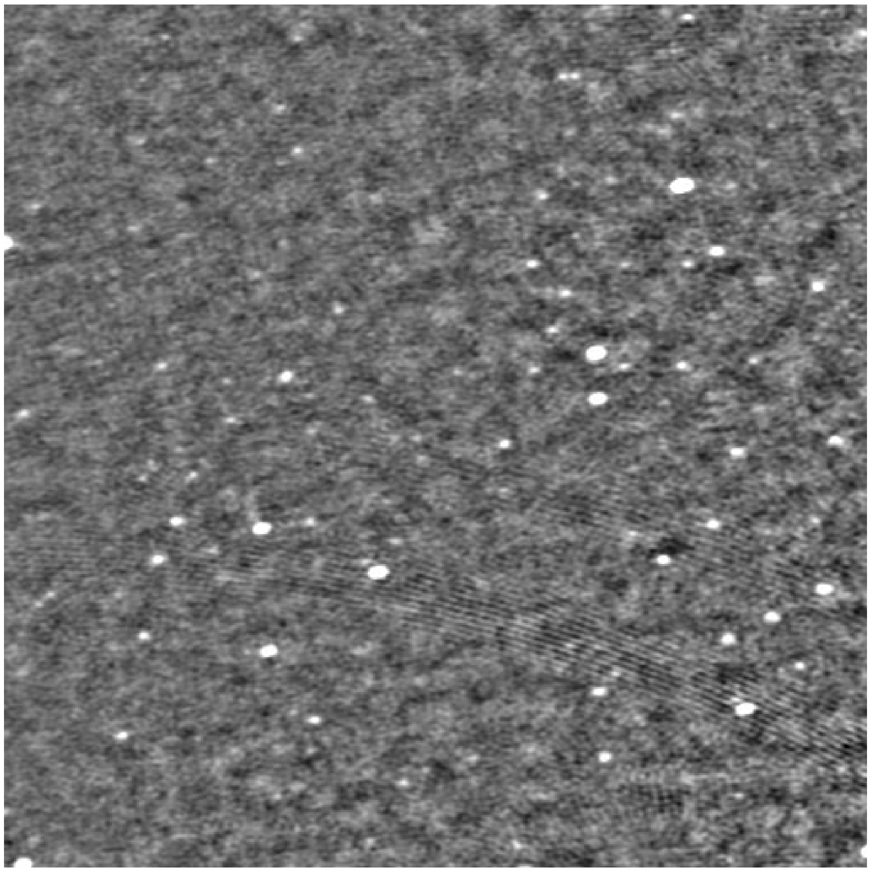
fMRI (cont'd)

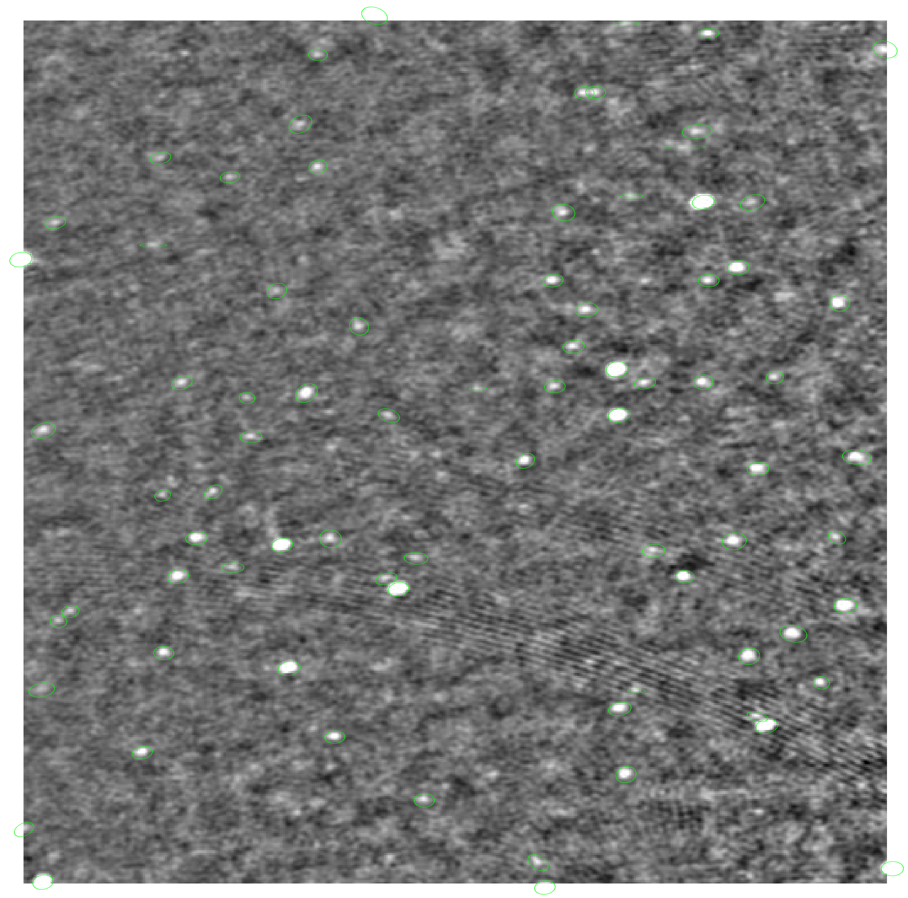
Perform hypothesis tests at many thousands of volume elements to identify loci of activation.



Motivating Example #2: Source Detection

- Interferometric radio telescope observations processed into digital image of the sky in radio frequencies.
- Signal at each pixel is a mixture of source and background signals.





Motivating Example #3: DNA Microarrays

- New technologies allow measurement of gene expression for thousands of genes simultaneously.

Condition 1

Gene	Subject 1	Subject 2	Subject 3	...
1	X_{111}	X_{121}	X_{131}	...
2	X_{211}	X_{221}	X_{231}	...
3	$\vdots$	$\vdots$	$\vdots$	...
4				
5				
6				
$\vdots$				

Condition 2

Gene	Subject 1	Subject 2	Subject 3	...
1	X_{112}	X_{122}	X_{132}	...
2	X_{212}	X_{222}	X_{232}	...
3	$\vdots$	$\vdots$	$\vdots$	...
4				
5				
6				
$\vdots$				

- Goal: Identify genes associated with differences among conditions.
- Typical analysis: hypothesis test at each gene.

The Multiple Testing Problem

- Perform m simultaneous hypothesis tests.
- Classify results as follows:

	H_0 Retained	H_0 Rejected	Total
H_0 True	$M_{0\|0}$	$M_{1\|0}$	M_0
H_0 False	$M_{0\|1}$	$M_{1\|1}$	M_1
Total	$m - R$	R	m

  Here, $M_{i|j}$ is the number of H_i chosen when H_j true.

- Only R and m are observed.
- Traditional methods seek strong control of type I error.

  Typical guarantee: $\mathsf{P}\left\{M_{1|0} > 0\right\} \leq \alpha$.

False Discovery and Nondiscovery Proportions

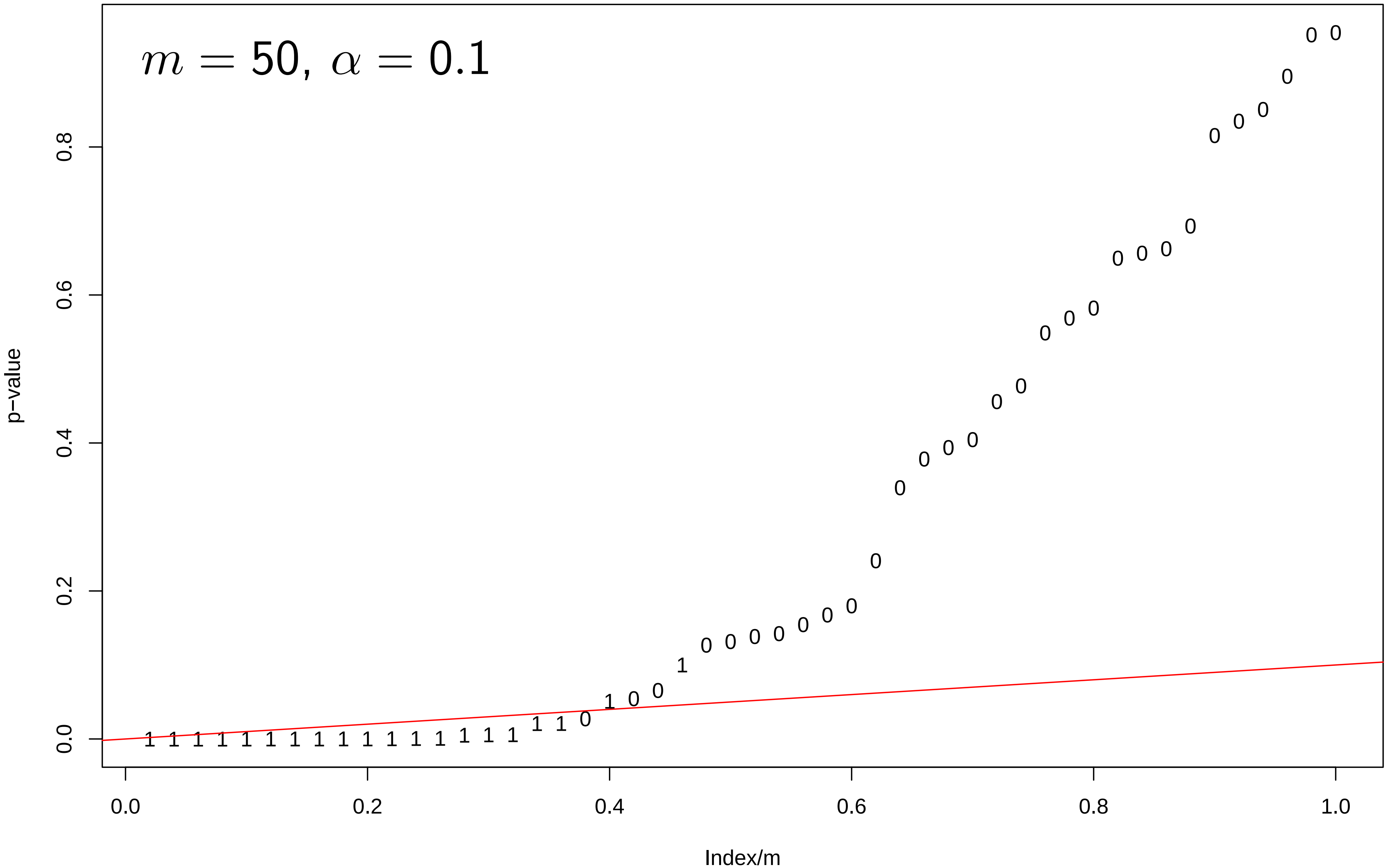
- Define the False Discovery Proportion (FDP) and the False Nondiscovery Proportion (FNP) as follows:

$$\text{FDP} = \begin{cases} \dfrac{M_{1|0}}{R} & \text{if } R > 0, \\ 0, & \text{if } R = 0. \end{cases} \qquad \text{FNP} = \begin{cases} \dfrac{M_{0|1}}{m - R} & \text{if } R < m, \\ 0, & \text{if } R = m. \end{cases}$$

- Then, the False Discovery Rate (FDR) and the False Nondiscovery Rate (FNR) are given by

$$\text{FDR} = \text{E(FDP)} \qquad \text{FNR} = \text{E(FNP)}.$$

- Benjamini and Hochberg (1995) introduced FDR and produced a procedure to guarantee that FDR $\leq \alpha$.



$m = 50$, $\alpha = 0.1$
p−value
Index/m
0.0
0.2
0.4
0.6
0.8
1.0

Selected Recent Work on FDR

Abromovich, Benjamini, Donoho, and Johnstone. (2000)

Benjamini & Hochberg (1995, 2000)

Benjamini & Liu (1999)

Benjamini & Hochberg (2000)

Benjamini & Yekutieli (2001)

Efron, et al. (2001)

Finner and Roters (2001, 2002)

Hochberg & Benjamini (1999)

Genovese & Wasserman (2001,2002,2003)

Pacifico, Genovese, Verdinelli & Wasserman (2003)

Sarkar (2002)

Storey (2001,2002)

Storey & Tibshirani (2001)

Seigmund, Taylor, and Storey (2003)

Tusher, Tibshirani, Chu (2001)

Road Map

1. Preliminaries
   - Models for FDP and FNP
   - FDP and FNP as stochastic processes
2. Plug-in Procedures
   - Asymptotic behavior of BH procedure
   - Optimal Thresholds
3. Confidence Envelopes and Thresholds
   - Exact Confidence Envelopes for FDP
   - Controlling exceedance probabilities for FDP
4. False Discovery Control for Random Fields
   - Confidence Supersets and Thresholds
   - Fast Algorithm
5. Estimating the p-value distribution

Road Map

1. Preliminaries
   - Models for FDP and FNP
   - FDP and FNP as stochastic processes
2. Plug-in Procedures
   - Asymptotic behavior of BH procedure
   - Optimal Thresholds
3. Confidence Envelopes and Thresholds
   - Exact Confidence Envelopes for FDP
   - Controlling exceedance probabilities for FDP
4. False Discovery Control for Random Fields
   - Confidence Supersets and Thresholds
   - Fast Algorithm
5. Estimating the p-value distribution

Basic Models

- Let $P^m = (P_1, \ldots, P_m)$ be the p-values for the m tests.
- Let $H^m = (H_1, \ldots, H_m)$ where $H_i = 0$ (or 1) if the i^{th} null hypothesis is true (or false).
- We assume the following model:

$$
\begin{aligned}
H_1, \ldots, H_m &\text{ iid Bernoulli}\langle a \rangle \\
\Xi_1, \ldots, \Xi_m &\text{ iid } \mathcal{L}_{\mathcal{F}} \\
P_i \mid H_i = 0, \Xi_i = \xi_i &\sim \text{Uniform}\langle 0, 1 \rangle \\
P_i \mid H_i = 1, \Xi_i = \xi_i &\sim \xi_i.
\end{aligned}
$$

where $\mathcal{L}_{\mathcal{F}}$ denotes a probability distribution on a class $\mathcal{F}$ of distributions on $[0, 1]$.

Basic Models (cont'd)

- Marginally, $P_1, \ldots, P_m$ are drawn iid from

$$G = (1 - a)U + aF,$$

where U is the Uniform$\langle 0, 1\rangle$ cdf and

$$F = \int \xi \, d\mathcal{L}_{\mathcal{F}}(\xi).$$

- Typical examples:
  - Parametric family: $\mathcal{F}_\Theta = \{F_\theta \colon \theta \in \Theta\}$
  - Concave, continuous distributions

$$\mathcal{F}_C = \{F \colon F \text{ concave, continuous cdf with } F \geq U\}.$$

- Can also work under what we call the *conditional model* where $H_1, \ldots, H_m$ are fixed, unknown.

Multiple Testing Procedures

- A multiple testing procedure T is a map $[0,1]^m \longrightarrow [0,1]$, where the null hypotheses are rejected in all those tests for which $P_i \leq T(P^m)$. We call T a *threshold*.
- Examples:

Uncorrected testing	$T_{\mathrm{U}}(P^m) = \alpha$
Bonferroni	$T_{\mathrm{B}}(P^m) = \alpha/m$
Fixed threshold at t	$T_t(P^m) = t$
First r	$T_{(r)}(P^m) = P_{(r)}$
Benjamini-Hochberg	$T_{\mathrm{BH}}(P^m) = \sup\{t \colon \widehat{G}(t) = t/\alpha\}$
Oracle	$T_{\mathrm{O}}(P^m) = \sup\{t \colon G(t) = (1-a)t/\alpha\}$
Plug In	$T_{\mathrm{PI}}(P^m) = \sup\{t \colon \widehat{G}(t) = (1-\widehat{a})t/\alpha\}$
Regression Classifier	$T_{\mathrm{Reg}}(P^m) = \sup\{t \colon \widehat{\mathsf{P}}\{H_1{=}1 \mid P_1{=}t\} > 1/2\}$

FDP and FNP as Stochastic Processes

- Inherent difficulty: FDP, FNP, and a general threshold all depend on the same data.
- Define the FDP and FNP processes, respectively, by

$$\mathrm{FDP}(t) \equiv \mathrm{FDP}(t; P^m, H^m) = \frac{\sum_i 1\left\{P_i \leq t\right\}(1 - H_i)}{\sum_i 1\left\{P_i \leq t\right\} + 1\left\{\text{all } P_i > t\right\}}$$

$$\mathrm{FNP}(t) \equiv \mathrm{FNP}(t; P^m, H^m) = \frac{\sum_i 1\left\{P_i > t\right\} H_i}{\sum_i 1\left\{P_i > t\right\} + 1\left\{\text{all } P_i \leq t\right\}}.$$

- For procedure T, the FDP and FNP are obtained by evaluating these processes at $T(P^m)$.

FDP and FNP as Stochastic Processes (cont'd)

- Both these processes converge to Gaussian processes outside a neighborhood of 0 and 1 respectively.
- For example, define

$$Z_m(t) = \sqrt{m}\,(\mathsf{FDP}(t) - Q(t)), \quad \delta \le t \le 1,$$

where $0 < \delta < 1$ and $Q(t) = (1-a)U/G$.

- Let Z be a mean 0 Gaussian process on $[\delta, 1]$ with covariance kernel

$$K(s,t) = a(1-a)\,\frac{(1-a)stF(s \wedge t) \,+\, aF(s)F(t)(s \wedge t)}{G^2(s)\,G^2(t)}.$$

- Then, $Z_m \rightsquigarrow Z$.

Road Map

1. Preliminaries
   - Models for FDP and FNP
   - FDP and FNP as stochastic processes
2. Plug-in Procedures
   - Asymptotic behavior of BH procedure
   - Optimal Thresholds
3. Confidence Envelopes and Thresholds
   - Exact Confidence Envelopes for FDP
   - Controlling exceedance probabilities for FDP
4. False Discovery Control for Random Fields
   - Confidence Supersets and Thresholds
   - Fast Algorithm
5. Estimating the p-value distribution

Plug-in Procedures

- Let $\widehat{G}_m$ be the empirical cdf of P^m under the mixture model. Ignoring ties, $\widehat{G}_m(P_{(i)}) = i/m$, so BH equivalent to

$$T_{\mathrm{BH}}(P^m) = \max\left\{t\colon\ \widehat{G}_m(t) = \frac{t}{\alpha}\right\}.$$

  as Storey (2002) first noted.

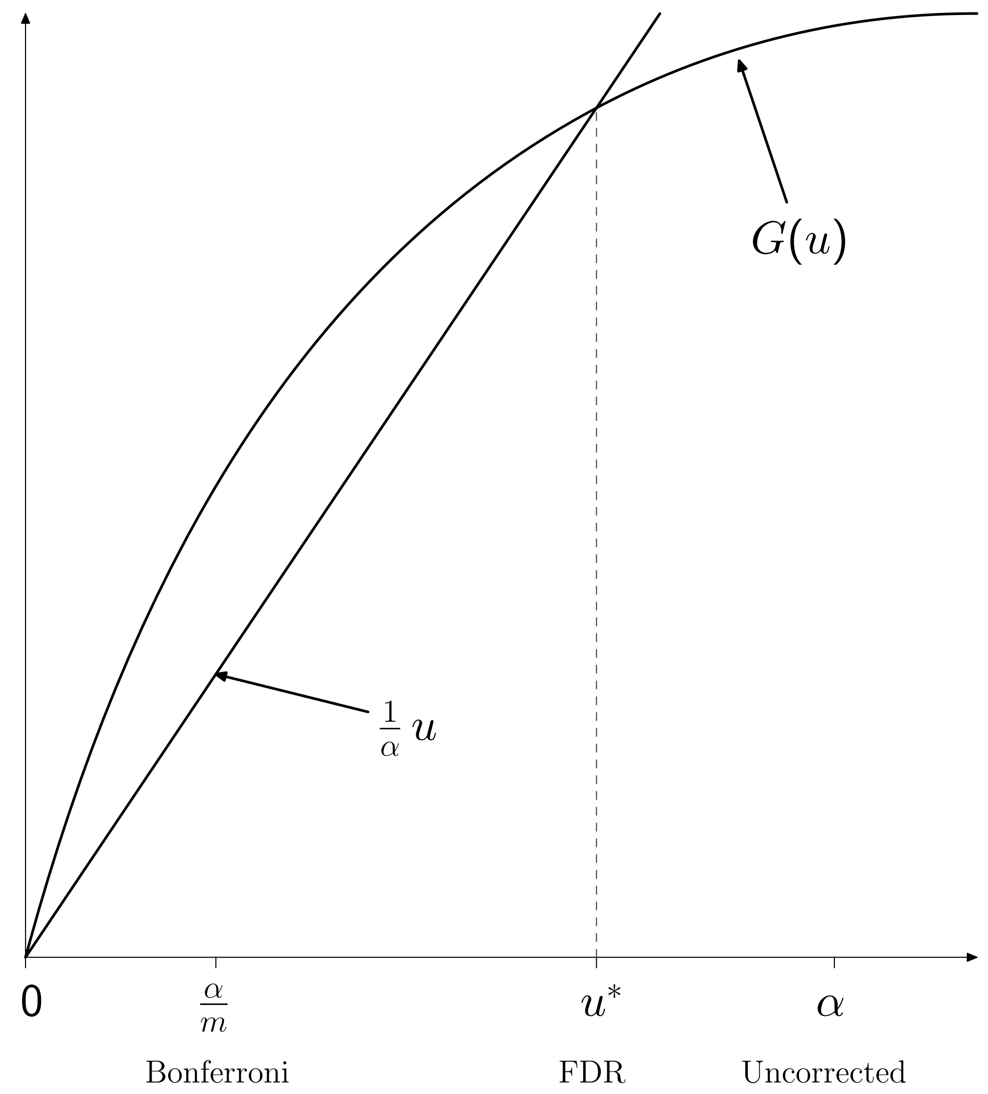
- One can think of this as a plug-in procedure for estimating

$$u^*(a, G) = \max\left\{t\colon\ G(t) = \frac{t}{\alpha}\right\}.$$

- Genovese and Wasserman (2002) showed that BH converges to a fixed-threshold at u^*.

Asymptotic Behavior of BH Procedure

This yields the following picture:



Optimal Thresholds

- In the continuous case, Benjamini and Hochberg's argument shows that

$$\mathsf{E}\Big[\mathsf{FDP}(T_{\mathrm{BH}}(P^m))\Big] = (1-a)\alpha.$$

- The BH procedure overcontrols FDR and thus will not in general minimize FNR.

- This suggests using T_{PI}, the plug-in estimator for

$$t^*(a, G) = \max\left\{t\colon\ G(t) = \frac{(1-a)t}{\alpha}\right\}.$$

- Note that $t^* \geq u^*$. If we knew a, this would correspond to using the BH procedure with $\alpha/(1-a)$ in place of α.

Optimal Thresholds (cont'd)

- For each $0 \leq t \leq 1$,

$$\mathrm{E}(\mathrm{FDP}(t)) = \frac{(1-a)\,t}{G(t)} + O\left((1-t)^m\right)$$

$$\mathrm{E}(\mathrm{FNP}(t)) = a\,\frac{1-F(t)}{1-G(t)} + O\left((a+(1-a)t)^m\right).$$

- Ignoring $O()$ terms and choosing t to minimize $\mathrm{E}(\mathrm{FNP}(t))$ subject to $\mathrm{E}(\mathrm{FDP}(t)) \leq \alpha$, yields $t^*(a, G)$ as the optimal threshold.
- GW (2002) show that

$$\mathrm{E}(\mathrm{FDP}(t^*(\widehat{a}, \widehat{G}))) \leq \alpha + O(m^{-1/2}).$$

Road Map

1. Preliminaries
   - Models for FDP and FNP
   - FDP and FNP as stochastic processes
2. Plug-in Procedures
   - Asymptotic behavior of BH procedure
   - Optimal Thresholds
3. Confidence Envelopes and Thresholds
   - Exact Confidence Envelopes for FDP
   - Controlling exceedance probabilities for FDP
4. False Discovery Control for Random Fields
   - Confidence Supersets and Thresholds
   - Fast Algorithm
5. Estimating the p-value distribution

Confidence Envelopes and Thresholds

- In practice, it would be useful to be able to control quantiles of the FDP process.
- We want a procedure T_C that, for some specified C and α, guarantees

$$\mathsf{P}_G\Big\{\mathsf{FDP}(T_C) > C\Big\} \leq \alpha.$$

  We call this a $(1-\alpha, C)$ *confidence-threshold procedure*.

- Three methods: (i) asymptotic closed-form threshold, (ii) asymptotic confidence envelope, and (iii) exact small-sample confidence envelope.

  I'll focus here on the latter.

Confidence Envelopes and Thresholds (cont'd)

- A $1 - \alpha$ confidence envelope for FDP is a random function $\overline{\mathsf{FDP}}(t)$ on $[0, 1]$ such that

$$\mathsf{P}\Big\{\mathsf{FDP}(t) \leq \overline{\mathsf{FDP}}(t) \text{ for all } t\Big\} \geq 1 - \alpha.$$

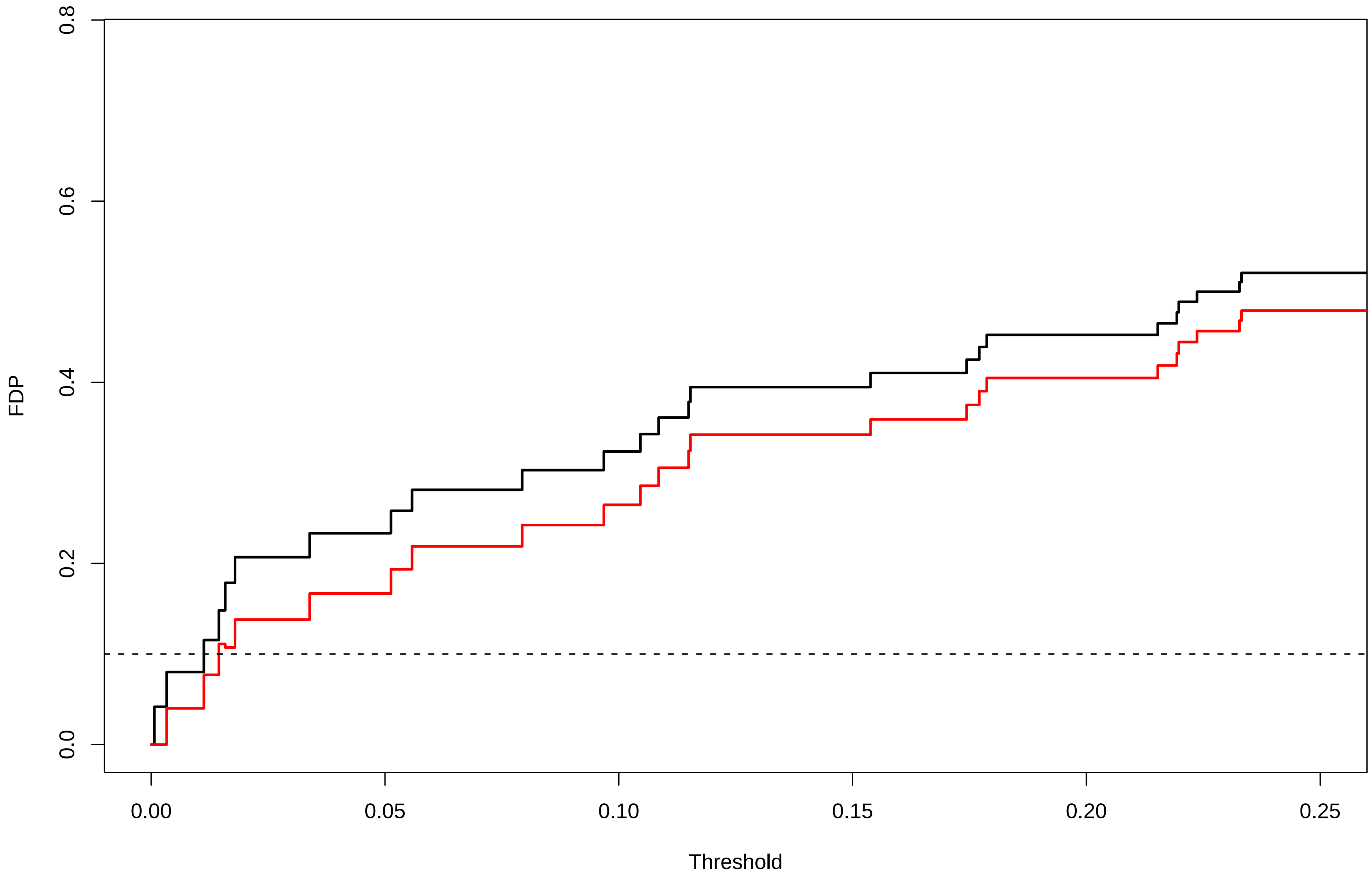
- Given such an envelope, we can construct confidence thresholds.

  Two special cases have proved useful:

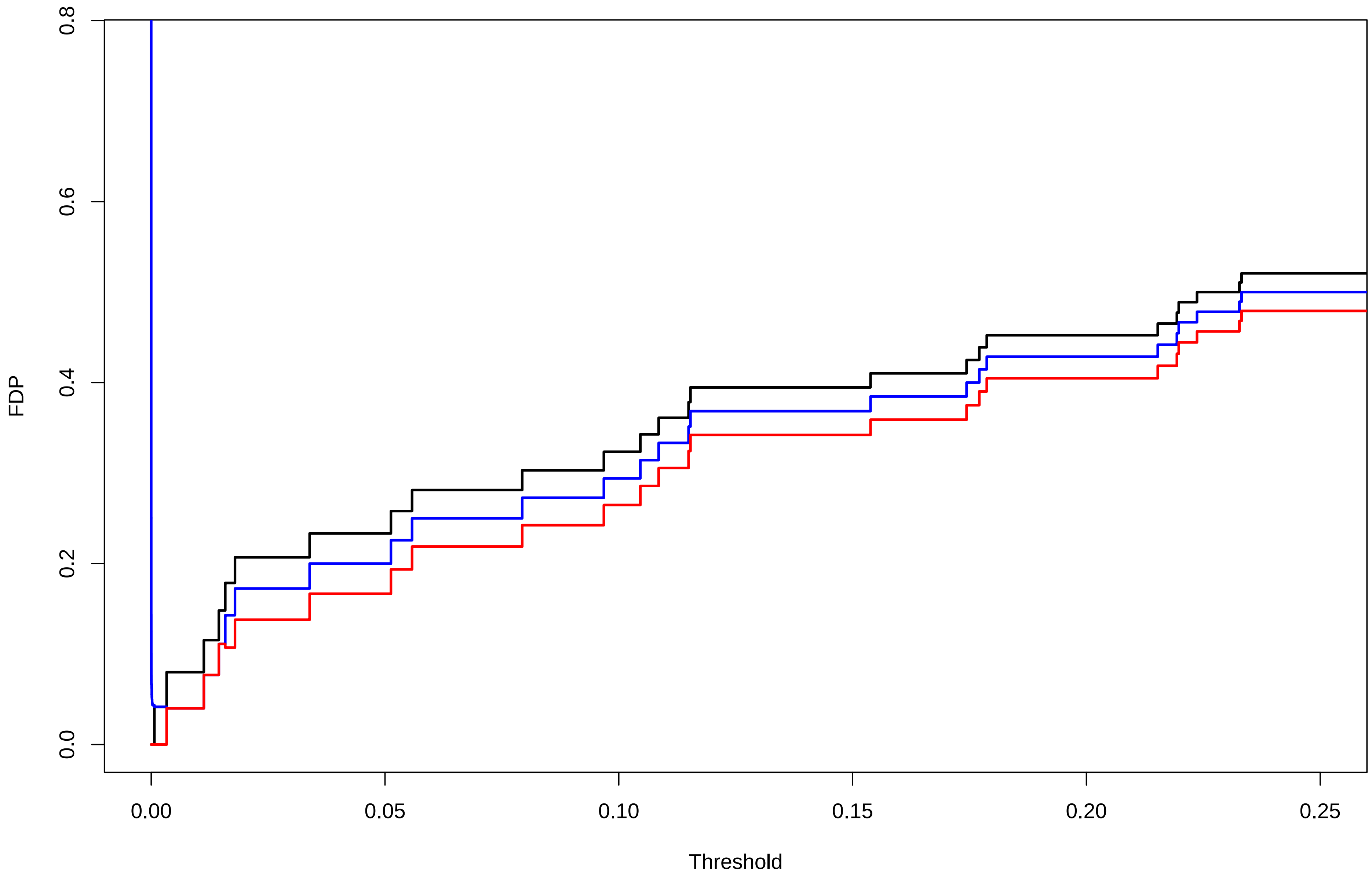
  – *Fixed-ceiling thresholds* define C to be a pre-determined constant (the ceiling) and take T_C to be the maximum t for which $\overline{\mathsf{FDP}}(t) \leq C$.

  – *Minimum-envelope thresholds* define C to be the $\min_t \overline{\mathsf{FDP}}(t)$ and take T_C to be the maximum t for which this minimum is achieved.

Exact Confidence Envelopes



Exact Confidence Envelopes (cont'd)



Exact Confidence Envelopes (cont'd)

- Given $V_1, \ldots, V_k$, let $\varphi_k(v_1, \ldots, v_k)$ be a level α test of the null that $V_1, \ldots, V_k$ are IID Uniform(0, 1).

- Define
$$p_0^m(h^m) = (p_i \colon h_i = 0,\ 1 \le i \le m)$$
$$m_0(h^m) = \sum_{i=1}^{m} (1 - h_i)$$
and
$$\mathcal{U}_\alpha(p^m) = \left\{ h^m \in \{0,1\}^m \colon \varphi_{m_0(h^m)}\left(p_0^m(h^m)\right) = 0 \right\}.$$
Note that as defined, $\mathcal{U}_\alpha$ always contains the vector $(1, 1, \ldots, 1)$.

- Let
$$\mathcal{G}_\alpha(p^m) = \left\{ \mathsf{FDP}(\cdot, h^m, p^m) \colon\ h^m \in \mathcal{U}_\alpha(p^m) \right\}$$
$$\mathcal{M}_\alpha(p^m) = \left\{ m_0(h^m) \colon\ h^m \in \mathcal{U}_\alpha(p^m) \right\}.$$

Exact Confidence Envelopes (cont'd)

- THEOREM. For all $0 < a < 1$, F, and positive integers m,

$$\mathsf{P}\Big\{H^m \in \mathcal{U}_\alpha(P^m)\Big\} \geq 1 - \alpha$$
$$\mathsf{P}\Big\{M_0 \in \mathcal{M}_\alpha(P^m)\Big\} \geq 1 - \alpha$$
$$\mathsf{P}\Big\{\mathsf{FDP}(\cdot, H^m, P^m) \in \mathcal{G}_\alpha\Big\} \geq 1 - \alpha.$$

- Define $\overline{\mathsf{FDP}}$ to be pointwise supremum over $\mathcal{G}_\alpha$. Then, $\overline{\mathsf{FDP}}$ is a $1 - \alpha$ confidence envelope for FDP.

- Confidence thresholds are then easy to construct. For example

$$T_c = \sup\{t : \Gamma(t) \leq c \text{ and } \Gamma \in \mathcal{G}_\alpha(P^m)\}$$

is a $1 - \alpha$ fixed-ceiling confidence threshold with ceiling c.

Choice of Tests

- The choice of uniformity tests has a big impact on performance of the confidence envelopes.
- There are two desiderata:

  A. "Power": $\overline{\text{FDP}}$ should be close to FDP, and

  B. Computability: Need to to carry out all 2^m tests quickly.

- Both are met by using the kth order statistic of any subset as a test statistic, for some k. We call these the $P_{(k)}$ tests.

  For small k, these are sensitive to departures that have a large impact on FDP. They can also be computed in m or few steps.

- In contrast, traditional uniformity tests, such as the (one-sided) Kolmogorov-Smirnov test do not fare as well.

  The Kolmogorov-Smnirov test looks for deviations from uniformity equally though all the p-values.

Computing $P_{(k)}$ Envelopes

- Let q_{mkj} denote the α quantile of the Beta$(k, m - j + 1)$ for $k \le j \le m$.

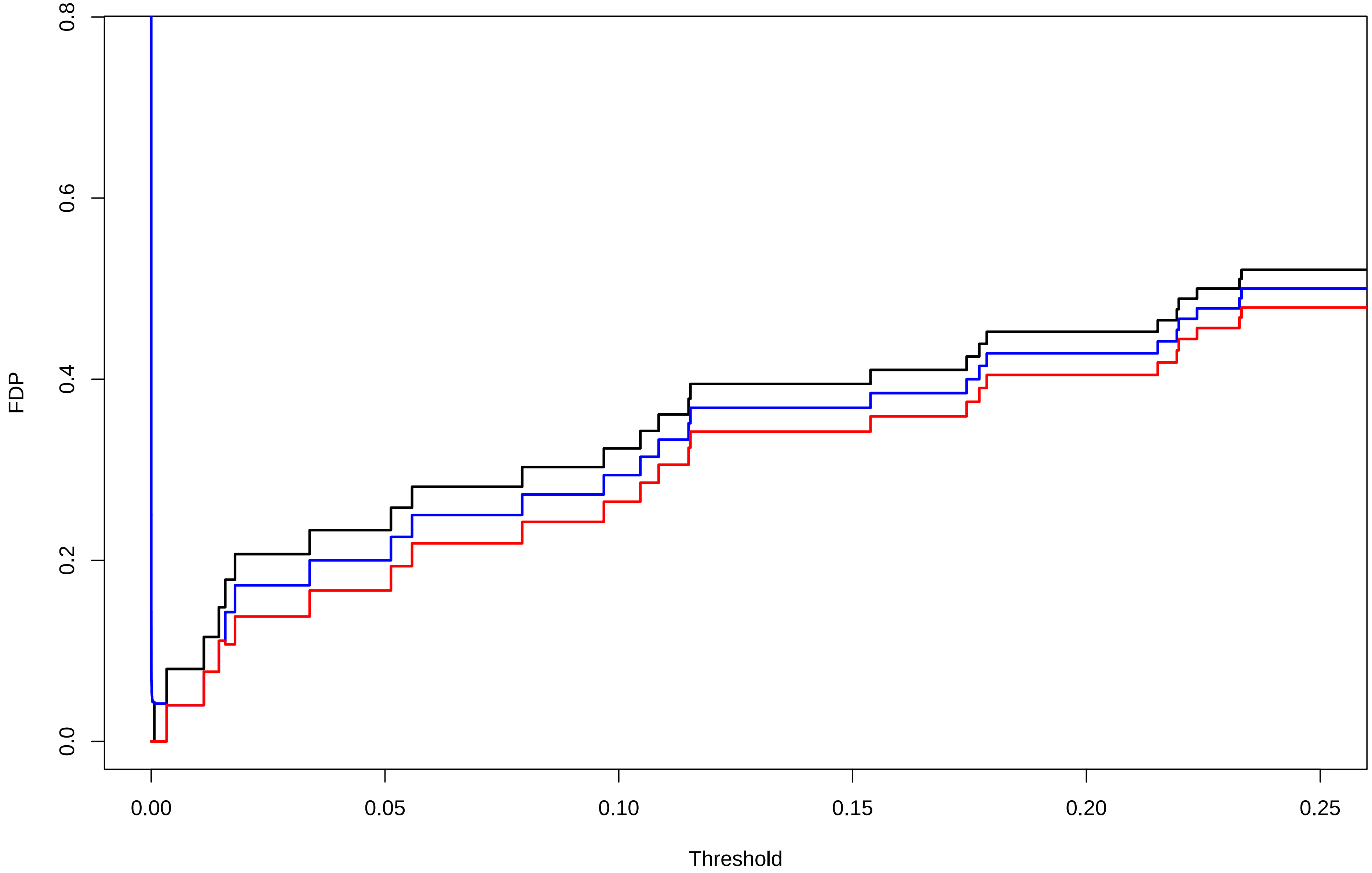
- Let J_k be the index of the smallest $P_{(j)}$ which is $\ge q_{mkj}$.

- The confidence envelope for the $P_{(k)}$-test is achieved by the configuration of nulls (0) and alternatives (1) in the ordered p-values.

$$\underbrace{0 \cdots 0}_{k-1} \overbrace{1 \cdots 1}^{J_k - k} 0 \cdots 0$$

$$\overline{\mathrm{FDP}}_k(t) = \begin{cases} 1 & \text{if } t \le \frac{k-1}{m} \\ \frac{k-1}{m\widehat{G}(t)} & \text{if } \frac{k-1}{m} < t \le \frac{J_k}{m} \\ 1 - \frac{J_k - k + 1}{m\widehat{G}(t)} & \text{if } t > \frac{J_k}{m} \end{cases}$$

Computing $P_{(k)}$ Envelopes (cont'd)



Choice Among $P_{(k)}$ Tests

- For any k, let $V_k = J_k - k$.
- In any pairwise comparison of $P_{(k)}$ and $P_{(k')}$ tests with $k < k'$, there are only three possible orderings:

A. $P_{(k)}$ dominates everwhere if $V_k \geq V_{k'}$,

B. $P_{(k')}$ dominates everywhere if $V_{k'} > V_k \left[1 + \frac{k' - k}{k - 1}\right] + \frac{k' - k}{k - 1}$,

C. Otherwise, the two profiles cross at $J_{k'}$ with value $(k' - 1)/J_{k'}$.

- The result for any k can be put in terms of Uniform hitting times for a boundary of the form $G(q_{mkj}) \approx G(\tilde{q}_{mk}/(m - j + 1))$.

  The distribution of these hitting times can be computed exactly (with difficulty) via Steck's equality.

Choice Among $P_{(k)}$ Tests (cont'd)

- Alternatively, using a family of alternative distributions, such as Uniform$(0, 1/\theta)$ or Normal$(\theta, 1)$, we can compute $k^*(\theta)$, the optimal k for each θ.

- So far, this is consistent with our simulation results across a wide variety of families.

- The $P_{(1)}$ and $P_{(2)}$ tests appear to perform well under a wide range of alternatives.

- Next steps: data dependent choice of k, adjusted test procedures.

  Plug-in estimation into $k^*(\theta)$ for approximate family is a simple but effective data-dependent choice.

Road Map

1. Preliminaries
   - Models for FDP and FNP
   - FDP and FNP as stochastic processes
2. Plug-in Procedures
   - Asymptotic behavior of BH procedure
   - Optimal Thresholds
3. Confidence Envelopes and Thresholds
   - Exact Confidence Envelopes for FDP
   - Controlling exceedance probabilities for FDP
4. False Discovery Control for Random Fields
   - Confidence Supersets and Thresholds
   - Fast Algorithm
5. Estimating the p-value distribution

False Discovery Control for Random Fields

- Multiple testing methods based on the excursions of random fields are widely used, especially in functional neuroimaging (e.g., Cao and Worsley, 1998) and scan clustering (Glaz, Naus, and Wallenstein, 2001).

- False Discovery Control extends to this setting as well.

- For a set S and a random field $X = \{X(s): s \in S\}$ with mean function $\mu(s)$, use the realized value of X to test the collection of one-sided hypotheses

$$H_{0,s} : \mu(s) = 0 \text{ versus } H_{1,s} : \mu(s) > 0.$$

Let $S_0 = \{s \in S : \ \mu(s) = 0\}$.

False Discovery Control for Random Fields

- Define a spatial version of FDP by

$$\mathsf{FDP}(t) = \frac{\lambda(S_0 \cap \{s \in S : X(s) \geq t\})}{\lambda(\{s \in S : X(s) \geq t\})},$$

  where λ is usually Lebesgue measure.

- As in the cases discussed earlier, we can control FDR or quantiles of FDP.

- Our approach is again based on finding a confidence envelope for FDP by finding a confidence superset U of S_0.

Confidence Supersets and Envelopes

1. For every $A \subset S$, test $H_0 : A \subset S_0$ versus $H_1 : A \not\subset S_0$ at level α using the test statistic $X(A) = \sup_{s \in A} X(s)$.

   The tail area for this statistic is $p(z, A) = \mathsf{P}\Big\{X(A) \geq z\Big\}$.

2. Let $\mathcal{C} = \{A \subset S\colon\ p(x(A), A) \geq \alpha\}$.

3. Then, $U = \bigcup_{A \in \mathcal{C}} A$ satisfies $\mathsf{P}\Big\{U \supset S_0\Big\} \geq 1 - \alpha$.

4. And,

$$\overline{\mathsf{FDP}}(t) = \frac{\lambda(U \cap \{s \in S : X(s) > t\})}{\lambda(\{s \in S : X(s) > t\})},$$

   is a confidence envelope for FDP.

Confidence Supersets and Envelopes (cont'd)

- The challenge of this strategy is to find U without computing the tests for every subset.

- In general, define a sequence of nested partitions that separates points
$$\mathcal{S}_n = \{S_{n1}, \ldots, S_{nN_n}\}.$$
Example: unions of cubes.
Our algorithm (below) applied to $\mathcal{S}_n$ produces a set U_n.
The set $U = \overline{\lim_n U_n}$ is a confidence superset for S_0.

- For a given partition $S_1, \ldots, S_N$ of S, our algorithm requires at most N steps though in effect computing 2^N tests.
We assume the null distribution of $\sup_{j \in \mathcal{I}} X(S_j)$ can be computed for any $\mathcal{I} \subset \{1, \ldots, N\}$

Confidence Supersets and Envelopes (cont'd)

Algorithm

1. Compute all realized values of the test statistics $x(S_j)$
2. Sort these in decreasing order $x_{(1)} \geq \cdots \geq x_{(N)}$.

   Let $S_{(j)}$ be the partition element corresponding to $x_{(j)}$.
3. For $k = 1, \ldots, N$ do the following:

   a. Set $V_k = \bigcup_{j=k}^{N} S_{(j)}$.

   b. Compute $p(x_{(k)}, V_k)$.

   c. If $p(x_{(k)}, V_k) \geq \alpha$: STOP and set $V^* = V_k$.

   d. If $p(x_{(k)}, V_k) < \alpha$: increase k by 1 and GOTO 3a.

Extracting Thresholds

- Using U, we can define FDR-controlling thresholds, confidence thresholds, and thresholds that control the number of false clusters to some tolerance.
- For the latter, decompse the t-level set of X into its connected components $C_{t1}, \ldots, C_{tk_t}$.
- Say a cluster C is false at tolerance ϵ if
$$\frac{\lambda(C \cap S_0)}{\lambda(C)} \geq \epsilon.$$
- For level t, let $\xi(t)$ denote the proportion of false clusters (at tol ϵ) out of k_t clusters.
- Then,
$$\overline{\xi}(t) = \frac{\#\left\{1 \leq i \leq k_t : \frac{\lambda(C_{ti} \cap U)}{\lambda(C_{ti})} \geq \epsilon\right\}}{k_t}$$
gives a $1 - \alpha$ confidence envelope for ξ.

Gaussian Fields

- Assume $S = [0,1]^d$ and that X is a zero-mean, homogeneous Gaussian field with covariance

  $$\text{Cov}(X(r), X(s)) = \rho(r-s),$$

  where we assume that ρ gives X almost surely continuous sample paths.

  Example: $\rho(u) = 1 - u^T C^{-1} u + o(\|u\|^2)$ for some matrix C.

- The key challenge here is to approximate $p(z, A)$.

  A common method uses the expected Euler characteristic of the level sets.

Gaussian Fields (cont'd)

- For our purposes, this will not work because the Euler characteristic approximation is monotone for non-convex sets.

  Note also that for non-convex sets, not all terms in the Euler approximation are accurate.

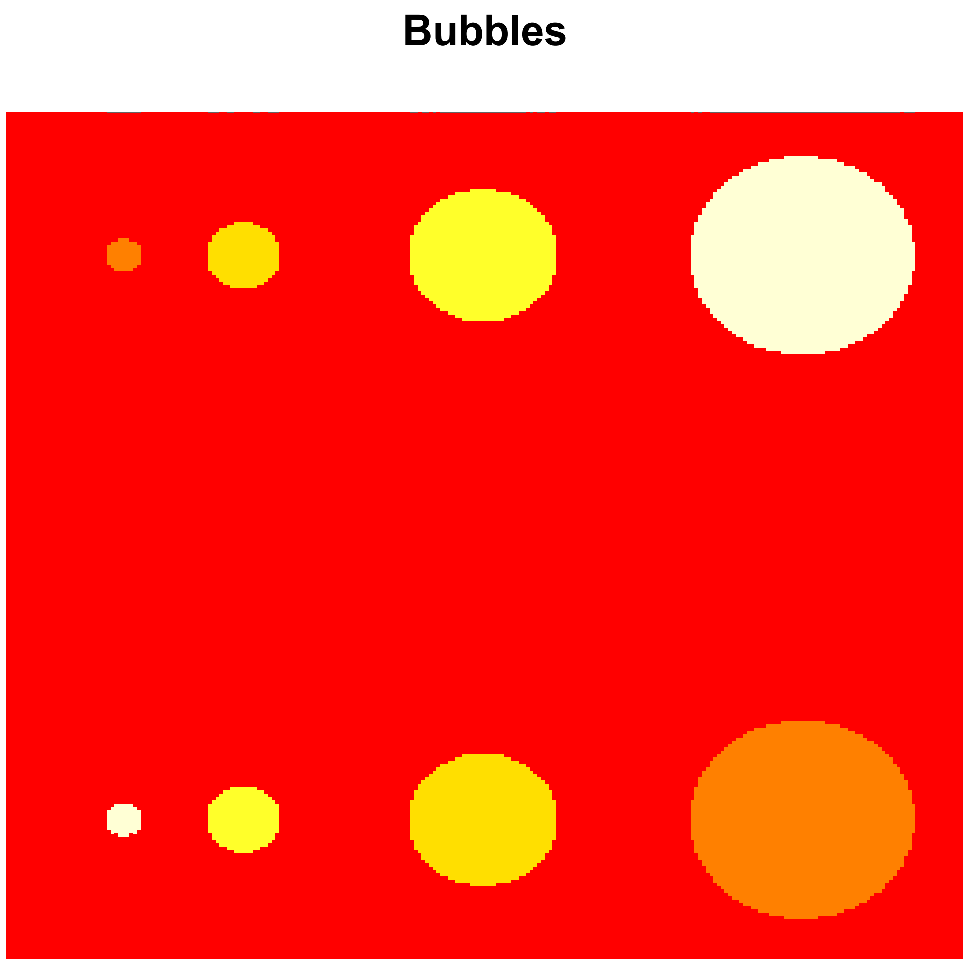
- Instead we use a result of Piterbarg (1996) to obtain

$$p(z, A) = \mathsf{P}\left\{ \sup_{s \in A} \frac{X(s)}{\sigma} \geq \frac{z}{\sigma} \right\} \simeq \frac{\pi^{-\frac{d}{2}}}{|\det C|} \lambda(A) \left(\frac{z}{\sigma}\right)^d \left[1 - \Phi\left(\frac{z}{\sigma}\right)\right],$$

  for C as in the quadratic form above.

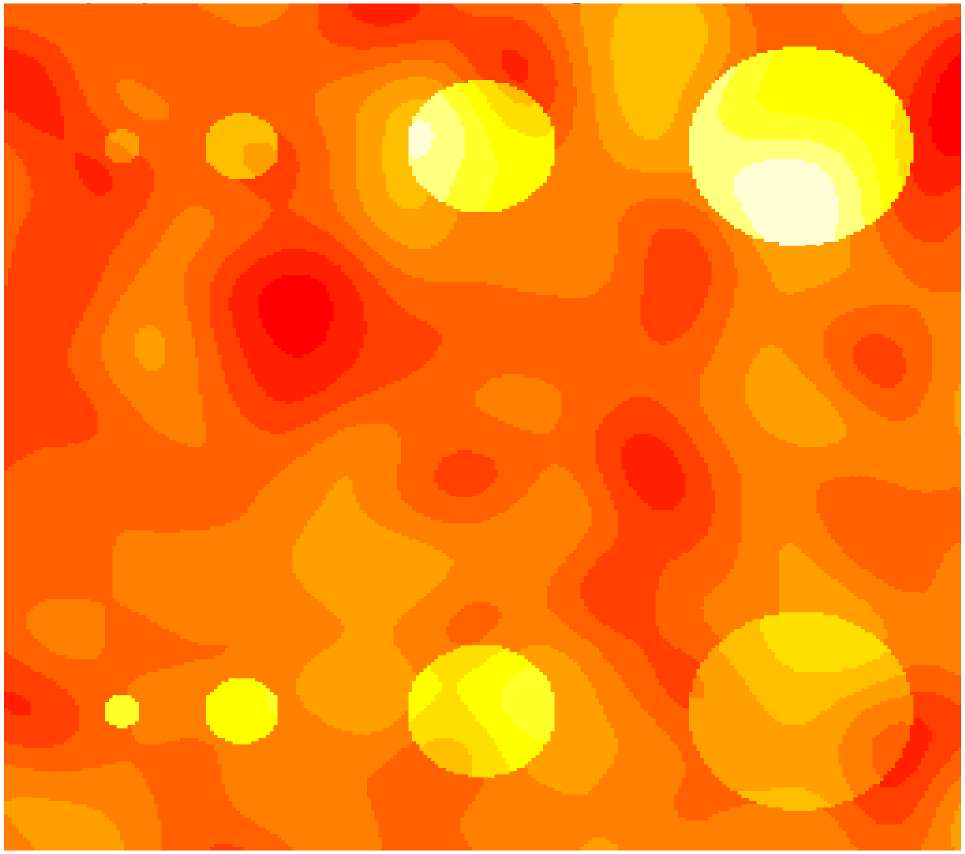
- Simulations over a wide variety of S_0s and covariance structures show that coverage of U rapidly converges to the target level.

Gaussian Fields: Example

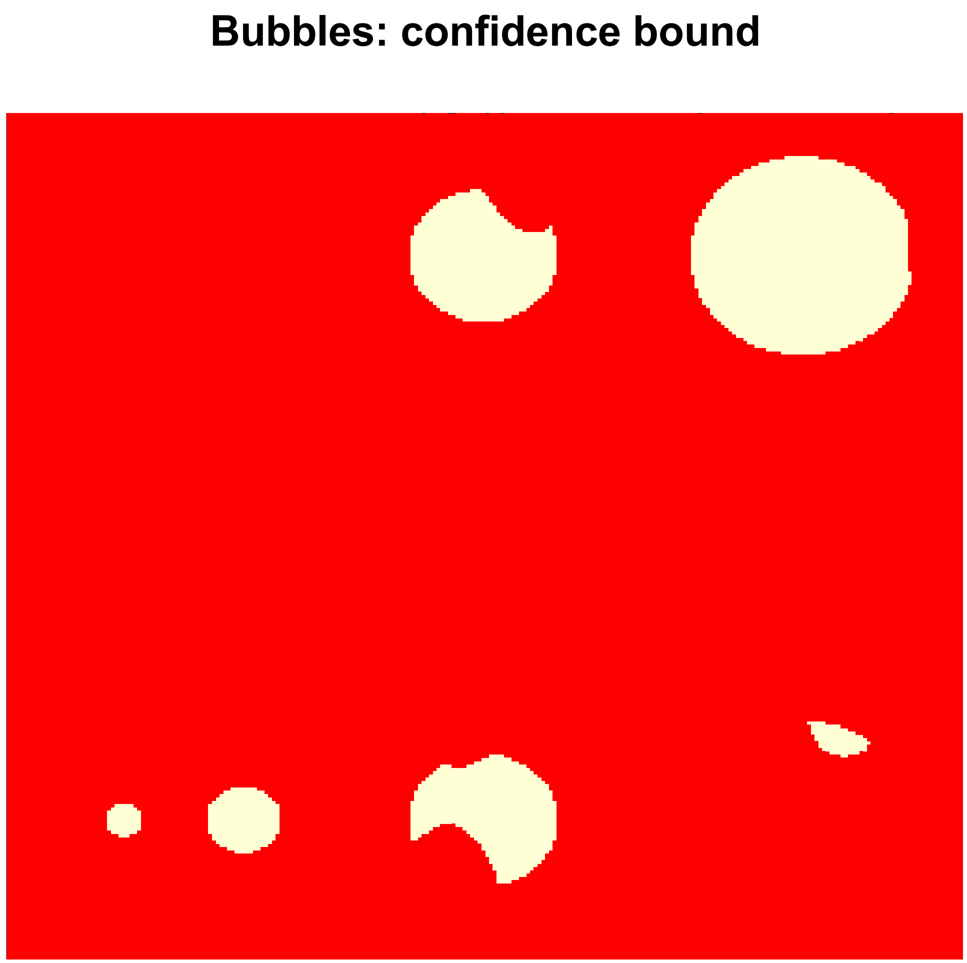


Gaussian Fields: Example (cont'd)

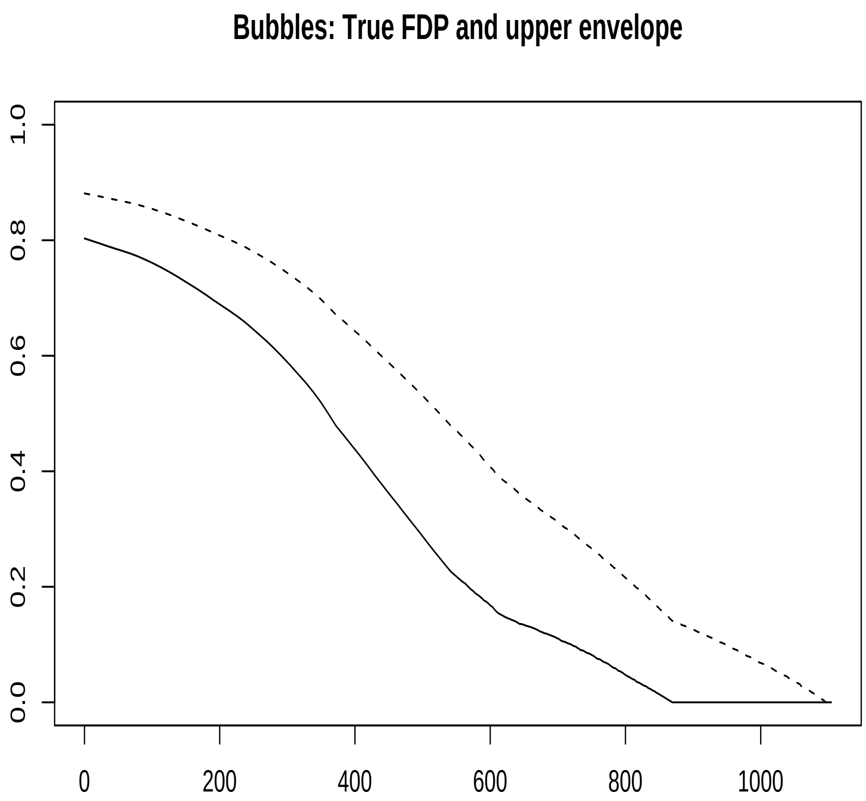
Bubbles + noise



Gaussian Fields: Example (cont'd)



Gaussian Fields: Example (cont'd)



Road Map

1. Preliminaries
   - Models for FDP and FNP
   - FDP and FNP as stochastic processes
2. Plug-in Procedures
   - Asymptotic behavior of BH procedure
   - Optimal Thresholds
3. Confidence Envelopes and Thresholds
   - Exact Confidence Envelopes for FDP
   - Controlling exceedance probabilities for FDP
4. False Discovery Control for Random Fields
   - Confidence Supersets and Thresholds
   - Fast Algorithm
5. Estimating the p-value distribution

Estimating a and F

- Recall that the p-value distribution $G = (1-a)U + aF$ where a and F are unknown.

- We need a good estimate of a for plug-in estimates,

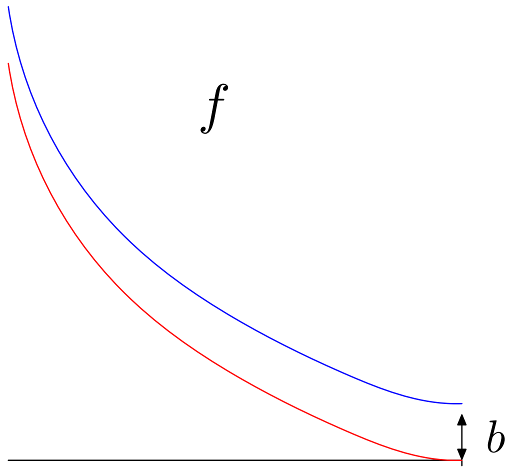
$$T_{\mathrm{PI}}(P^m) = \max\left\{t\colon\ \widehat{G}(t) = \frac{(1-\widehat{a})t}{\alpha}\right\},$$

  that approximate the optimal threshold.

- Good estimates of a and F can be useful for some types of confidence thresholds.

Estimating a and F (cont'd)

- Identifiability and Purity



If $\min f = b > 0$, can write $F = (1-b)U + bF_0$, $\mathcal{O}_G = \{(\widetilde{a}, \widetilde{F}) : \widetilde{F} \in \mathcal{F}, G = (1-\widetilde{a})U + \widetilde{a}\widetilde{F}\}$ may contain more than one element.

If $f = F'$ is decreasing with $f(1) = 0$, then (a, F) is identifiable.

- In general, let $\underline{a} \leq a$ be the smallest mixing weight in the orbit: $\underline{a} = 1 - \min_t g(t)$. This is identifiable.

  Storey (2002) notes that $0 \leq \sup_{0<t<1} \dfrac{G(t) - t}{1 - t} \leq \underline{a} \leq a \leq 1$.

- $a - \underline{a}$ is typically small: $a - \underline{a} = ae^{-n\theta^2/2}$ in the two-sided test of $\theta = 0$ versus $\theta \neq 0$ in the Normal$\langle \theta, 1 \rangle$ model.

Estimating a and F (cont'd)

- Parametric Case
  - Derived a $1-\beta$ one-sided conf. int. for $\underline{a}$ and thus a. (a, θ) typically identifiable even if $a > \underline{a}$; use MLE.

- Non-parametric case:
  - Derived a $1-\beta$ one-sided conf. int. for $\underline{a}$ and thus a.
  - When F concave, get $\widehat{a}_{\mathrm{HS}} = \underline{a} + O_P(m^{-1/3}(\log m)^{1/3})$.
  - When F smooth enough, get $\widehat{a}_{\mathrm{S}} = \underline{a} + O_P(m^{-2/5})$.
  - Consistent estimate for F_0 if $\widehat{a}$ consistent for $\underline{a}$:

$$\widehat{F}_m = \operatorname*{argmin}_{H \in \mathcal{F}} \|\widehat{G} - (1-\widehat{a})U - \widehat{a}H\|_\infty.$$

Estimating a and F (cont'd)

- $\widehat{a}_{\mathrm{S}}$ uses "spacings" estimator (Swanepoel, 1999) to estimate $\min g(t)$. This yields

$$\frac{m^{2/5}}{(\log m)^{\delta}} (\widehat{a} - \underline{a}) \rightsquigarrow \text{Normal}\langle 0, (1 - \underline{a})^2 \rangle$$

- $\widehat{a}_{\mathrm{HS}} = 1 - \min\{h(1)\colon \gamma_- \le h \le \gamma_+\}$, where $[\gamma_-, \gamma_+]$ is the $1 - \alpha$ finite-sample confidence envelope for g derived in Hentgartner and Stark (1995).

  A $1 - \alpha$ confidence interval for a is $[1 - \gamma_+(1), 1]$.

- Storey's estimator for fixed $0 \le t_0 \le 1$

$$\widehat{a}_0 = \left(\frac{\widehat{G}(t_0) - t_0}{1 - t_0} \right)_+ ,$$

  though asymptotically biased can also be useful.

Estimating a and F (cont'd)

- Confidence interval for a given by

$$\mathcal{A}_m = \left[\max_t \frac{\widehat{G}_m(t) - t - \epsilon_m(\alpha)}{1-t}, 1\right],$$

where $\widehat{G}_m$ is EDF and $\epsilon_m(\alpha) = \sqrt{\log(2/\alpha)/2m}$.

Then,

$$1 - \alpha \le \inf_{a,F} \mathsf{P}\left\{a \in A_m\right\} \le 1 - \alpha + R_m$$

where

$$R_m = \sum_j (-1)^j \frac{\alpha^{j^2}}{2^{j^2-1}} + O\left(\frac{(\log m)^2}{\sqrt{m}}\right)$$

Road Map

1. Preliminaries
   - Models for FDP and FNP
   - FDP and FNP as stochastic processes
2. Plug-in Procedures
   - Asymptotic behavior of BH procedure
   - Optimal Thresholds
3. Confidence Envelopes and Thresholds
   - Exact Confidence Envelopes for FDP
   - Controlling exceedance probabilities for FDP
4. False Discovery Control for Random Fields
   - Confidence Supersets and Thresholds
   - Fast Algorithm
5. Estimating the p-value distribution

Take-Home Points

- It's helpful to think of FDP (FNP, FDR, ...) as stochastic processes.

  Dependence between threshold and FDP can have a big effect.

- Asymptotic approach motivated by particular applications, but asymptotics appear to kick in rather quickly.

- Confidence thresholds have practical advantages over FDR control.

- Dependence complicates the analysis greatly; confidence envelopes appear to be valid under positive dependence.

- For spatial applications, adjacency can be highly informative but is ignored by standard multiple testing methods.

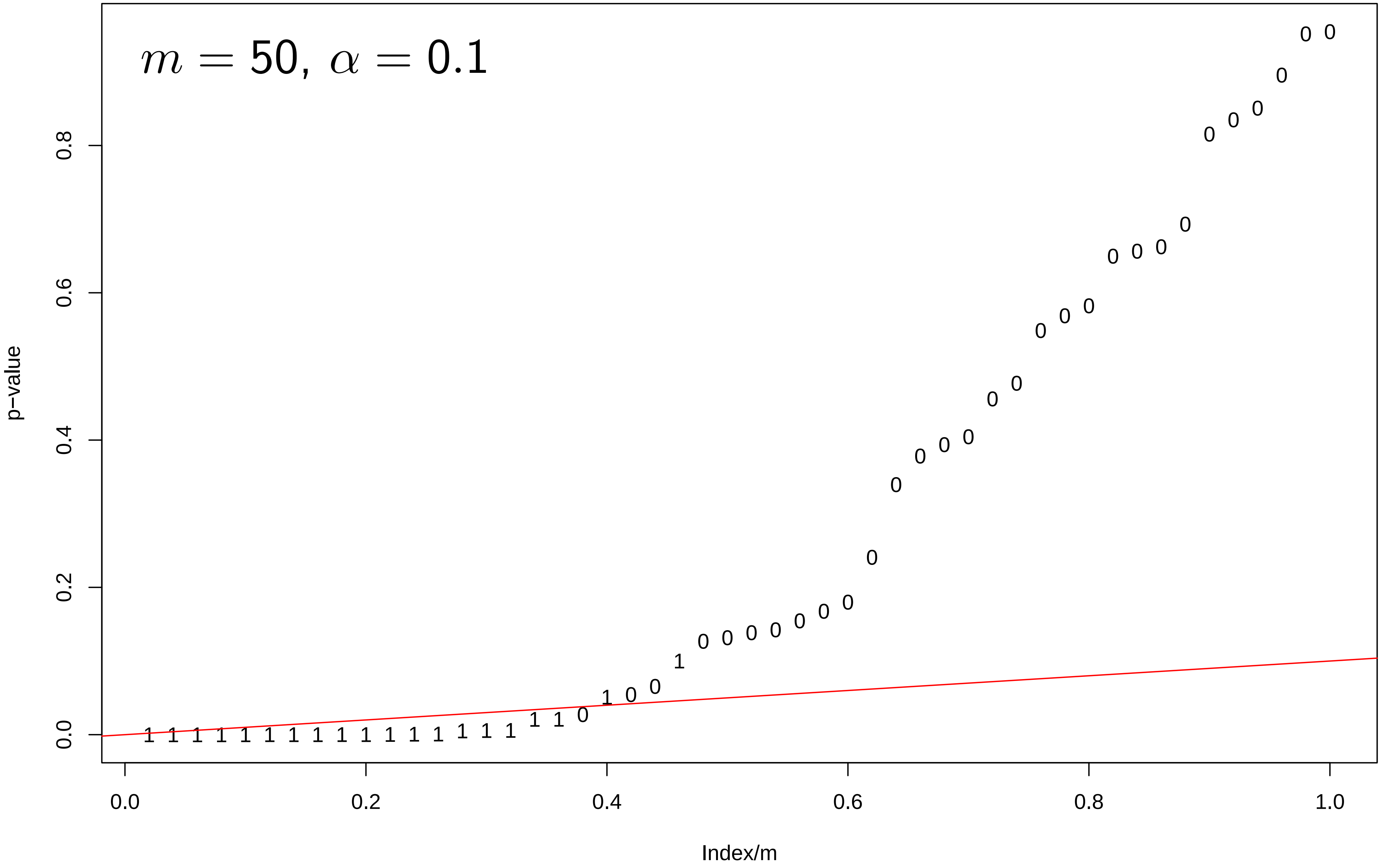
  Cluster-based false discovery control (work in progress) offers an advantage in these cases.

Appendix

1. Notation
2. BH Picture
3. Asymptotic Confidence Thresholds
4. Bayes and Empirical Bayes Thresholds

Recurring Notation

$m, M_0, M_{1\|0}$	# of tests, true nulls, false discoveries
a	Mixture weight on $a$lternative
$H^m = (H_1, \ldots, H_m)$	Unobserved true classifications
$P^m = (P_1, \ldots, P_m)$	Observed p-values
U	CDF of Uniform$\langle 0, 1\rangle$
F, f	Alternative CDF and density
$G = (1-a)U + aF$	Marginal CDF of P_i
$g = G'$	Marginal density of P_i
$\widehat{G}_m$	Estimate of G (e.g., empirical CDF of P^m)
$\epsilon_k(\beta) = \sqrt{\frac{1}{2k} \log\left(\frac{2}{\beta}\right)}$	DKW bound $1-\beta$ quantile of $\|\widehat{G}_k - G\|_\infty$



$m = 50$, $\alpha = 0.1$
p−value
Index/m
0.0
0.2
0.4
0.6
0.8
1.0

Closed-Form Asymptotic Confidence Thresholds

- Let

$$t_0 = Q^{-1}(c) \qquad \hat{t}_0 = \widehat{Q}^{-1}(c).$$

- Then define

$$T_C = \hat{t}_0 + \frac{\widehat{\Delta}_{m,\alpha}}{\sqrt{m}},$$

where $\widehat{\Delta}_{m,\alpha}$ is depends on a density estimate of $g = G'$.

- Then,

$$\mathsf{P}_G\Big\{ \mathsf{FDP}(T_C) \leq c \Big\} \geq 1 - \alpha + o(1).$$

Closed-Form Asymptotic Confidence Thresholds

- Details:

$$\widehat{\Delta}_{m,\alpha} = \frac{z_{\alpha/2}\left(\sqrt{\widehat{K}_{Q^{-1}}(\widehat{t_0},\widehat{t_0})} + \widehat{g}(\widehat{t_0})\right) + 2\sqrt{\log m}}{1 - \widehat{a} - c\widehat{g}(\widehat{t_0})}$$

$$\widehat{K}_{Q^{-1}}(s,t) = \frac{\widehat{K}_Q(\widehat{Q}^{-1}(s), \widehat{Q}^{-1}(t))}{\widehat{Q'}(\widehat{Q}^{-1}(s))\widehat{Q'}(\widehat{Q}^{-1}(t))}$$

$$\widehat{K}_Q(s,t) = \frac{(1-\widehat{a})^2 st}{\widehat{G}^2(s)\widehat{G}^2(t)}\left[\widehat{G}(s \wedge t) - \widehat{G}(s)\widehat{G}(t)\right].$$

- This requires no bootstrapping but does require density estimation.

  This is analogous to the situation faced when estimating the standard error of a median.

Bayesian Thresholds

- Bayesian Threshold bounds posterior FDR:

$$T_{\text{Bayes}} = \mathsf{sup}\{t : \ \mathsf{E}(\mathsf{FDP}(t) \mid P^m) \leq \alpha\}$$

- Similarly, can construct a posterior (c, α) confidence threshold $T_{\text{Bayes},c}$ by

$$T_{\text{Bayes},c} = \mathsf{sup}\{t : \ \mathsf{P}\Big\{\mathsf{FDP}(t) \leq c \mid P^m\Big\} \leq \alpha\}$$

EBT (Empirical Bayes Testing)

- Efron et al (2001) note that

$$\mathsf{P}\Big\{H_i = 0 \mid P^m\Big\} = \frac{(1-a)}{g(P_i)} \equiv q(P_i)$$

- Reject whenever $q(p) \le \alpha$?
- For a, f unknown, $f \ge 0$ implies that

$$a \ge 1 - \min_p g(p) \Longrightarrow \widehat{a} = 1 - \min_p \widehat{g}(p).$$

- Then,

$$\widehat{q}(p) = \frac{1-\widehat{a}}{\widehat{g}(p)} = \frac{\min_s \widehat{g}(s)}{\widehat{g}(p)}$$

EBT versus FDR

- If we reject when $\mathsf{P}\Big\{H_i = 0 \mid P^m\Big\} \leq \alpha$, how many errors are we making?

- Under weak conditions, can show that

$$q(t) \leq \alpha \;\; \text{implies} \;\; Q(t) < \alpha$$

So EBT is conservative.

Behavior of $\widehat{q}$

- THEOREM. Let $\widehat{q}(t) = \frac{(1-a)}{\widehat{g}(t)}$. Suppose that

$$m^{\alpha}(\widehat{g}(t) - g(t)) \rightsquigarrow W$$

for some $\alpha > 0$, where W is a mean 0 Gaussian process with covariance kernel $\tau(v, w)$. Then

$$m^{\alpha}\,(\widehat{q}(t) - q(t)) \rightsquigarrow Z$$

where Z is a Gaussian process with mean 0 and covariance kernel

$$K_q(v, w) = \frac{(1-a)^2 \tau(v, w)}{g(v)^4 g(w)^4}.$$

Behavior of $\widehat{q}$ (cont'd)

- Parametric Case: $g \equiv g_\theta = (1-a) + a f_\theta(v)$ Then,

$$\mathsf{rel}(v) = \frac{\widehat{\mathsf{se}}(\widehat{q}(v))}{q(v)} \approx O\left(\frac{1}{\sqrt{m}}\right)\left|\frac{\partial \log g_\theta}{\partial d\theta}\right| \;=\; O\left(\frac{1}{\sqrt{m}}\right)|v-\theta| \quad \text{Normal case}$$

- Nonparametric Case

$$\widehat{g}(t) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{h_m}K\left(\frac{t-P_i}{h_m}\right)$$

$h_m = cm^{-\beta}$ where $\beta > 1/5$ (undersmooth). Then

$$\mathsf{rel}_v = \frac{c}{m^{(1-\beta)/2}\sqrt{g(v)}}.$$