Contributions to the statistical analysis of DNA microarray data

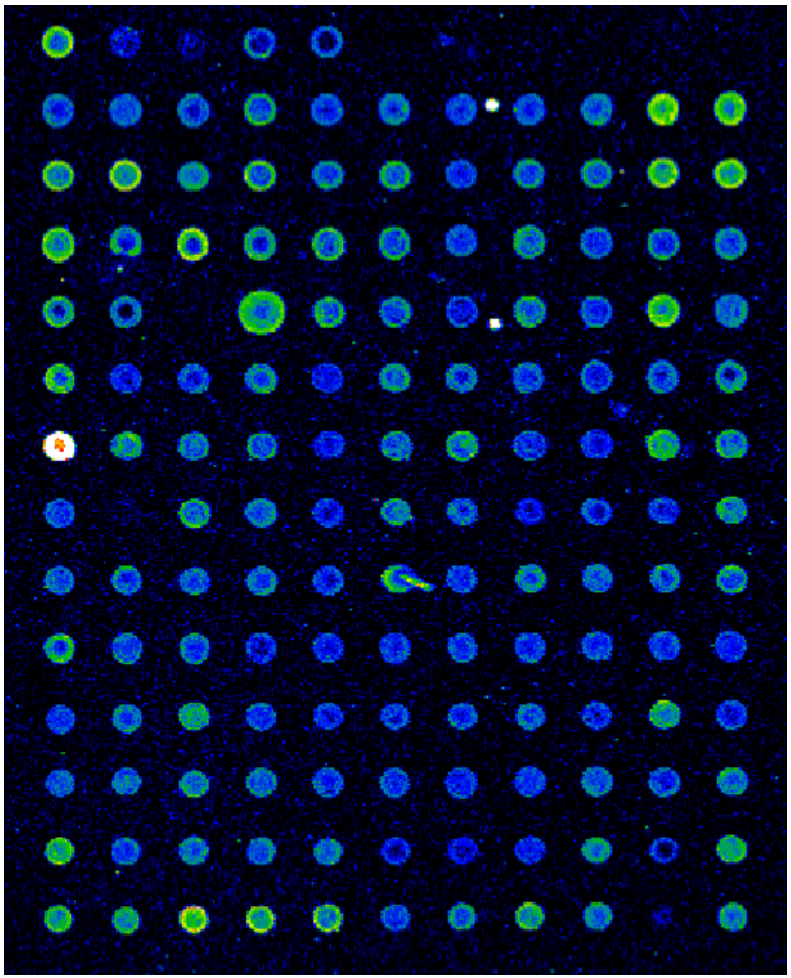
Pierre Neuvial[1,2]

[1]Université Paris VII, Laboratoire de Probabilités et Modèles Aléatoires

[2]Institut Curie / INSERM U900 / Mines ParisTech

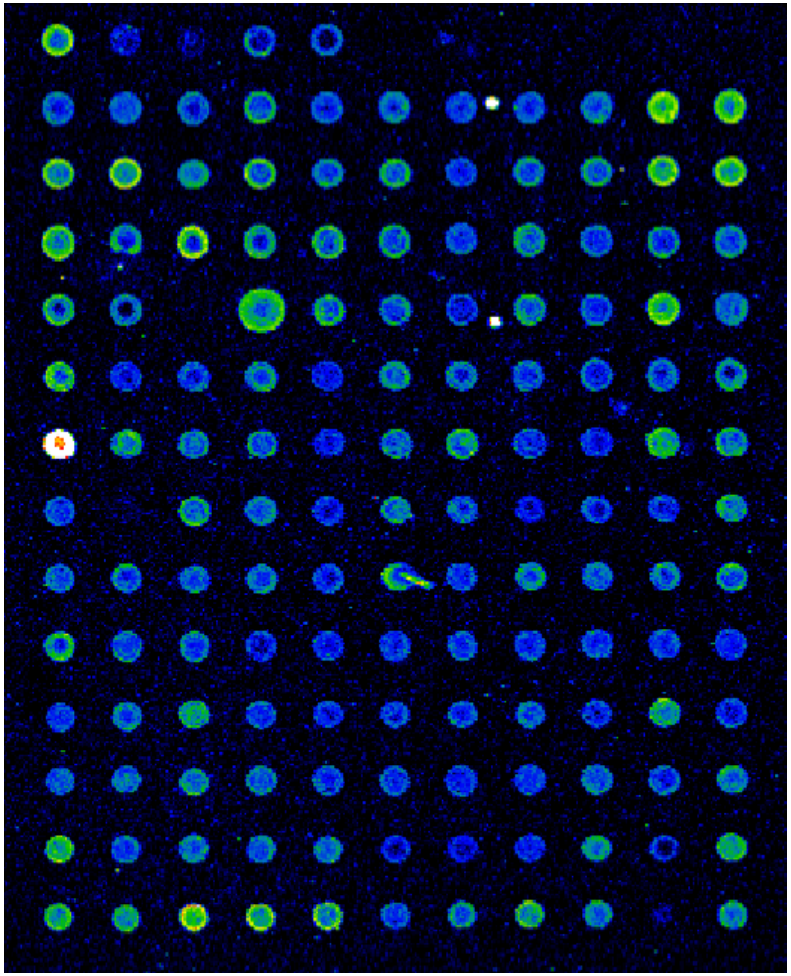
PhD thesis defence
September 30th, 2008

DNA microarray experiments



Small part of a scanned microarray

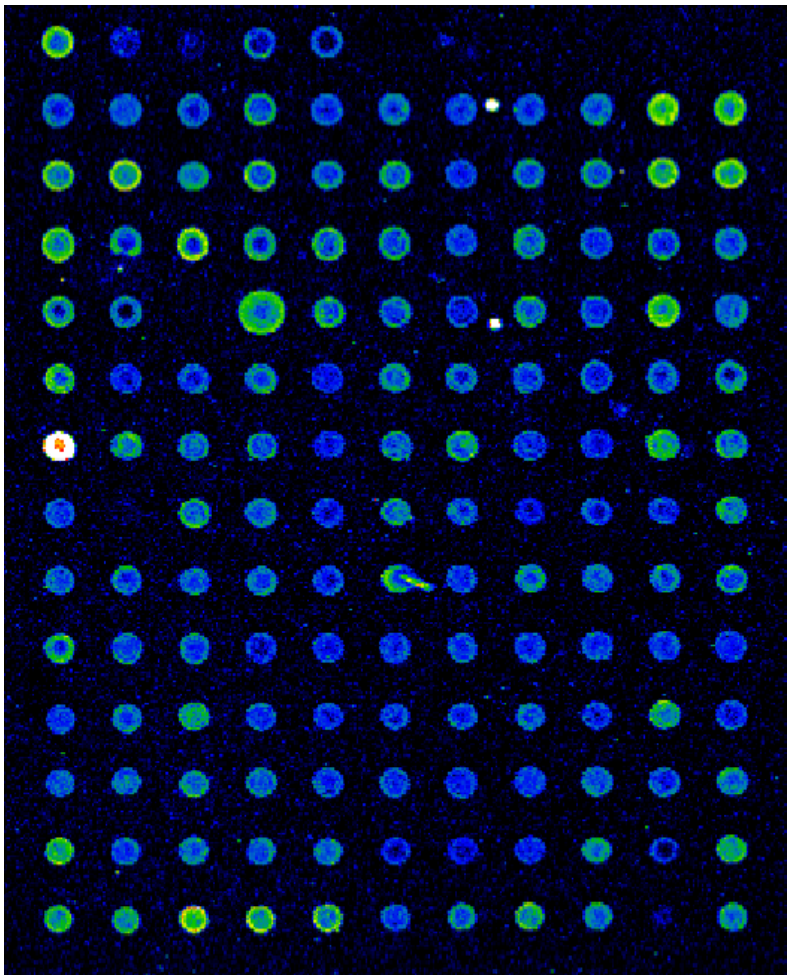
DNA microarray experiments



Small part of a scanned microarray :

- 1 spot $\leftrightarrow$ 1 variable
  e.g. one *gene*

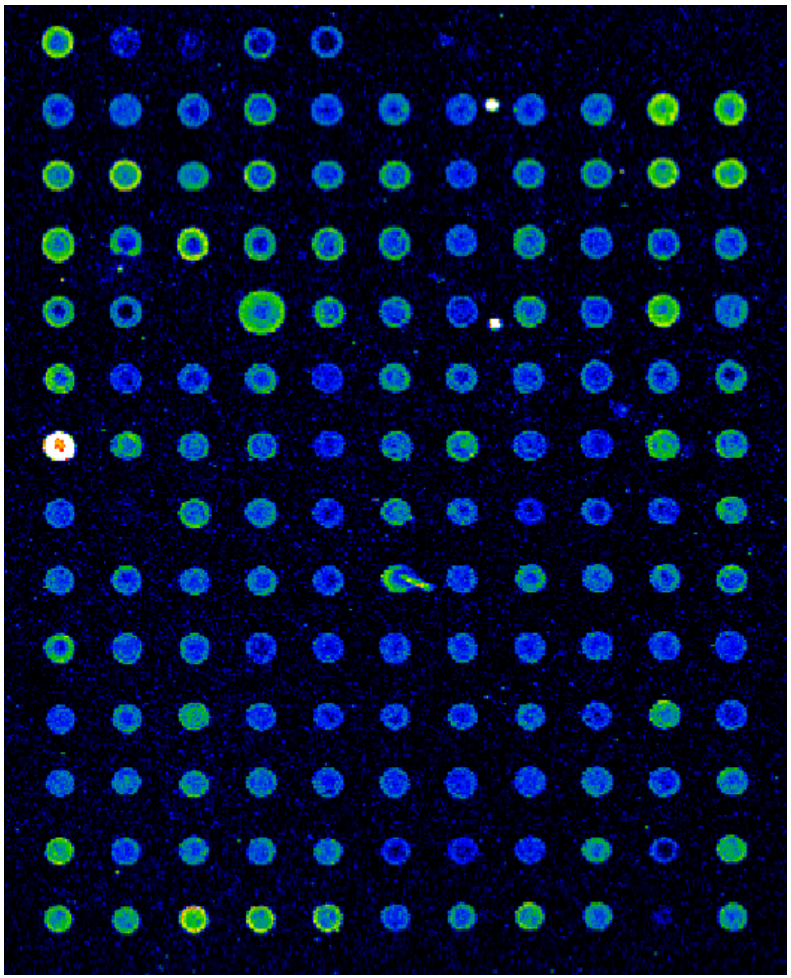
DNA microarray experiments



Small part of a scanned microarray :

- 1 spot $\leftrightarrow$ 1 variable
  e.g. one *gene*
- color $\leftrightarrow$ quantitative measurement
  e.g. that gene's *expression level*

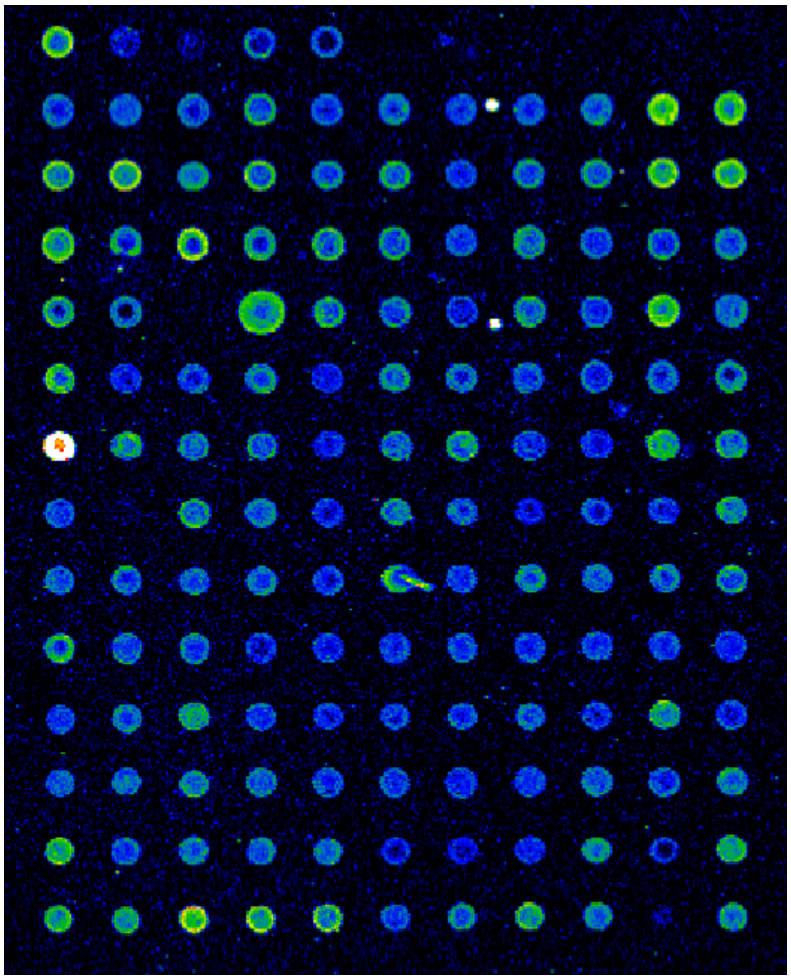
DNA microarray experiments



Small part of a scanned microarray :

- 1 spot $\leftrightarrow$ 1 variable
  e.g. one *gene*
- color $\leftrightarrow$ quantitative measurement
  e.g. that gene's *expression level*
- 1 experiment $\leftrightarrow$ 1 sample

DNA microarray experiments

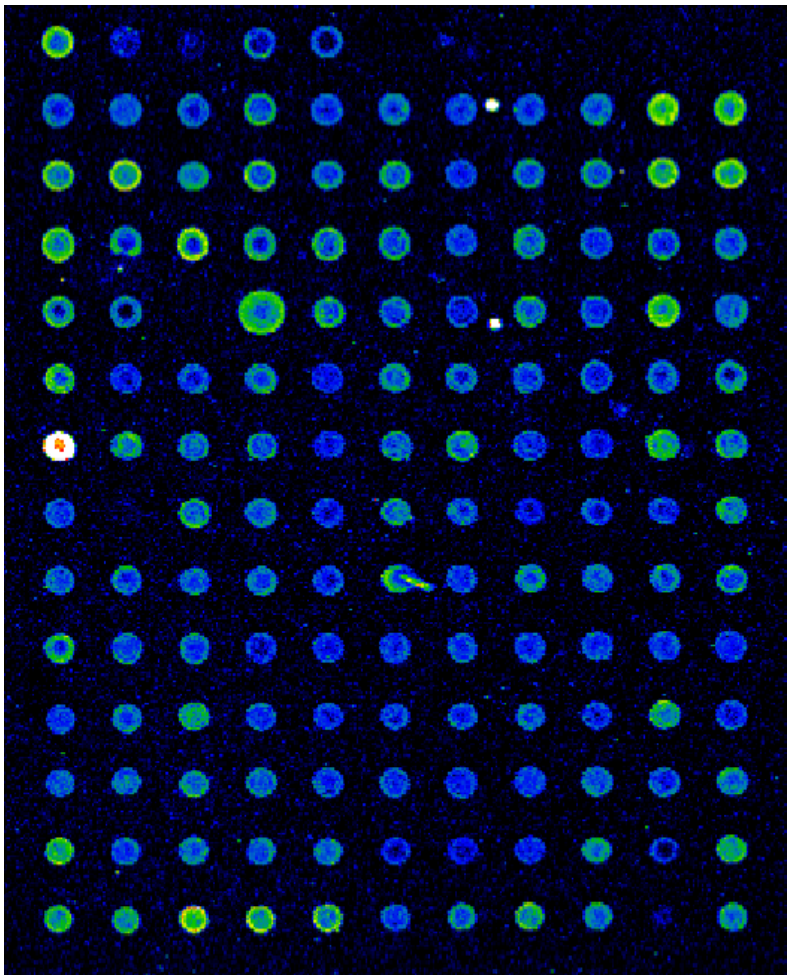


Small part of a scanned microarray :

- 1 spot $\leftrightarrow$ 1 variable
  e.g. one *gene*
- color $\leftrightarrow$ quantitative measurement
  e.g. that gene's *expression level*

- 1 experiment $\leftrightarrow$ 1 sample
- 1 experiment $\leftrightarrow$ many variables
  typically $10^5 - 10^6$

DNA microarray experiments



Small part of a scanned microarray :

- 1 spot $\leftrightarrow$ 1 variable
  e.g. one *gene*
- color $\leftrightarrow$ quantitative measurement
  e.g. that gene's *expression level*

- 1 experiment $\leftrightarrow$ 1 sample
- 1 experiment $\leftrightarrow$ many variables
  typically $10^5 - 10^6$

Why are microarrays relevant to cancer research ?

Cancers and genes

Cancer cells

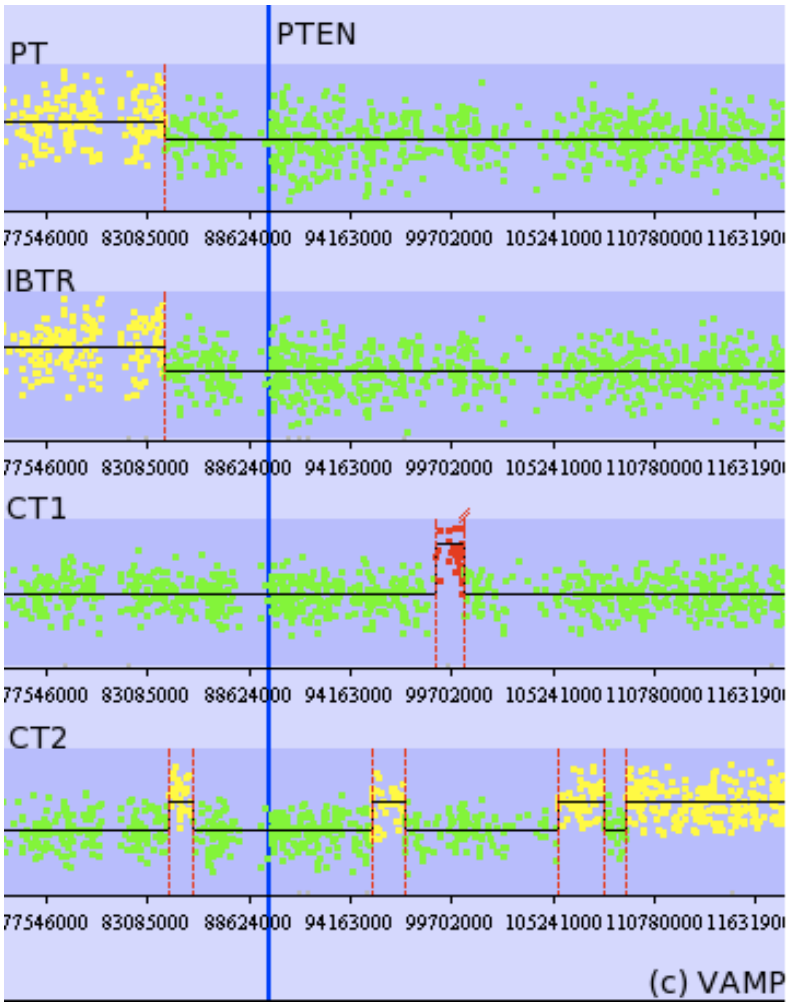
- grow without control
- avoid programmed cell death
- may invade adjacent tissues

Cancer involves dynamic changes in the genome

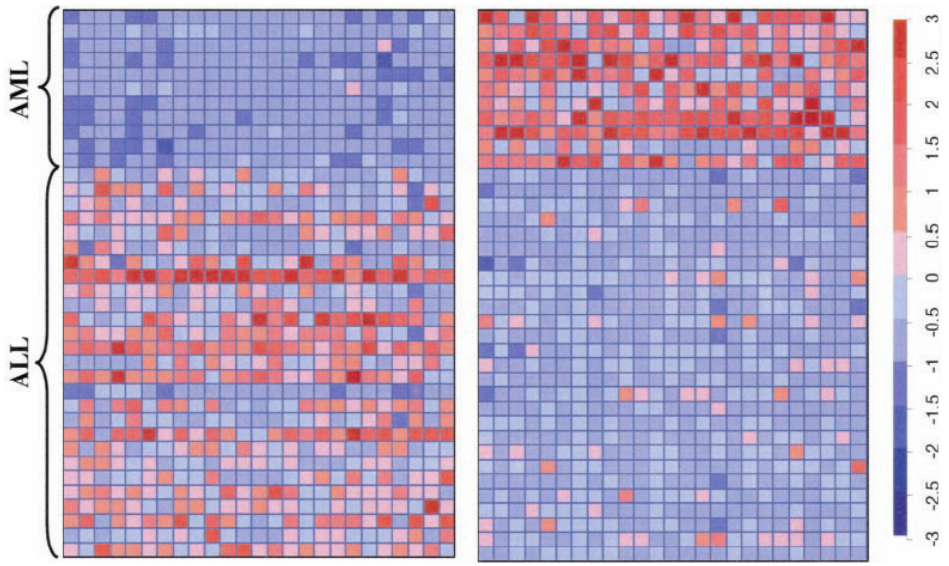
- DNA copy number alterations (e.g. mutations and aneuploidies)
- under- or over-expressed genes

Microarrays help identify genomic aberrations

DNA copy number changes



under- or over-expressed genes



Questions

Biological and clinical questions

1. understand tumor progression
2. find new therapeutic targets
3. identify prognosis and predictive factors

$\Rightarrow$ provide treatments adapted to each cancer subtype

Statistical questions raised

- unsupervised classification
- testing theory
- supervised classification, regression

Microarray data analysis

Characteristics of microarray data

- $10^5 - 10^6$ variables and only $10 - 100$ observations
- high experimental variability
- variables are not independent

Role of bioinformaticians and statisticians

1. understand biological or clinical questions
2. use or design adapted methods and software
3. analyze statistical properties of the methods used

Contributions of the thesis

- Normalization of DNA copy number data

  P. Neuvial, P. Hupé *et al*, *BMC Bioinformatics*, 2006

- Correlation between DNA copy number and expression

  P. Neuvial, P. Gestraud *et al*, poster at ISMB 2007

Contributions of the thesis

- Normalization of DNA copy number data

  P. Neuvial, P. Hupé *et al*, *BMC Bioinformatics*, 2006

- Correlation between DNA copy number and expression

  P. Neuvial, P. Gestraud *et al*, poster at ISMB 2007

- Unsupervised reconstruction of transcriptional regulatory networks

  M. Elati *et al*, *Bioinformatics*, 2007

- Definition of true recurrences among ipsilateral breast cancers

  M. Bollet, N. Servant *et al*, *JNCI*, 2008

Contributions of the thesis

- Normalization of DNA copy number data
  P. Neuvial, P. Hupé *et al*, *BMC Bioinformatics*, 2006
- Correlation between DNA copy number and expression
  P. Neuvial, P. Gestraud *et al*, poster at ISMB 2007
- Unsupervised reconstruction of transcriptional regulatory networks
  M. Elati *et al*, *Bioinformatics*, 2007
- Definition of true recurrences among ipsilateral breast cancers
  M. Bollet, N. Servant *et al*, *JNCI*, 2008
- Asymptotic properties of multiple testing procedures
  P. Neuvial, in revision for *EJS*
- Intrinsic bounds and FDR control in multiple testing problems
  P. Neuvial, submitted to *JMLR*

Contributions of the thesis

- Normalization of DNA copy number data

  P. Neuvial, P. Hupé *et al*, *BMC Bioinformatics*, 2006

- Correlation between DNA copy number and expression

  P. Neuvial, P. Gestraud *et al*, poster at ISMB 2007

- Unsupervised reconstruction of transcriptional regulatory networks

  M. Elati *et al*, *Bioinformatics*, 2007

- Definition of true recurrences among ipsilateral breast cancers

  M. Bollet, N. Servant *et al*, *JNCI*, 2008

- Asymptotic properties of multiple testing procedures

  P. Neuvial, in revision for *EJS*

- Intrinsic bounds and FDR control in multiple testing problems

  P. Neuvial, submitted to *JMLR*

Outline

1. Defining True Recurrences Among Ipsilateral Breast Cancers
   - Background: breast tumor recurrences
   - Method: a partial identity score
   - Result: improved definition of true recurrence

Outline

1. Defining True Recurrences Among Ipsilateral Breast Cancers
   - Background: breast tumor recurrences
   - Method: a partial identity score
   - Result: improved definition of true recurrence
2. Multiple testing procedures
   - Multiple testing
   - False Discoveries
   - Multiple testing procedures studied

Outline

1. Defining True Recurrences Among Ipsilateral Breast Cancers
   - Background: breast tumor recurrences
   - Method: a partial identity score
   - Result: improved definition of true recurrence
2. Multiple testing procedures
   - Multiple testing
   - False Discoveries
   - Multiple testing procedures studied
3. Asymptotic properties of FDR controlling procedures
   - Connections between Multiple Testing Procedures
   - Asymptotic false discovery proportion

Outline

Outline

Breast-conservative cancer treatment

Breast-conservative as compared to mastectomy

= equal survival

\+ superior psychosocial outcomes

\- risk of ipsilateral breast tumor recurrence (IBTR)

IBTR: New Primaries (NP) vs True Recurrences (TR)

- NP may be treated as the first tumor
- TR should get a more aggressive treatment
- Problem: no perfect definition

Our goal: improve current definition of NP/TR

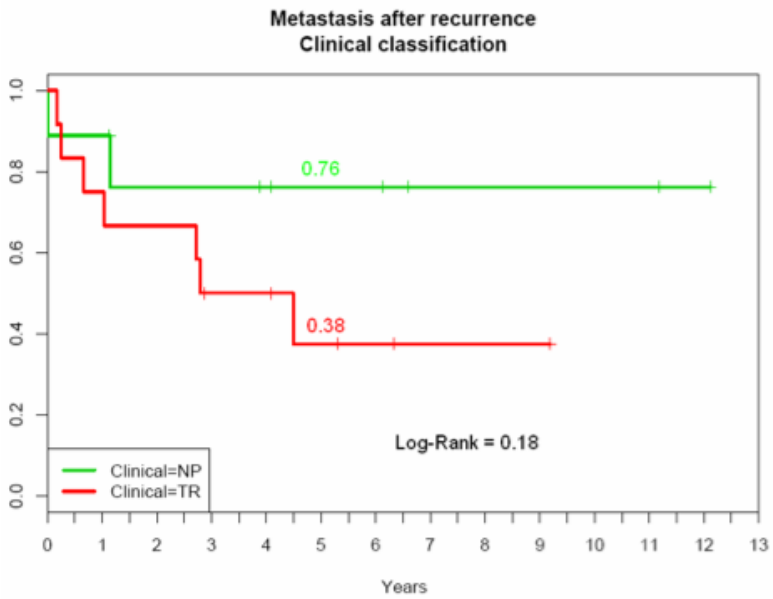
Current classification of breast tumor recurrences

Clinical definition: related tumors should share location, histological type, grade

Genomic definition: related tumors should share DNA copy number alterations (CNA)

Validation

- difficulty: no ground truth
- a good (posterior) indicator: metastasis-free survival



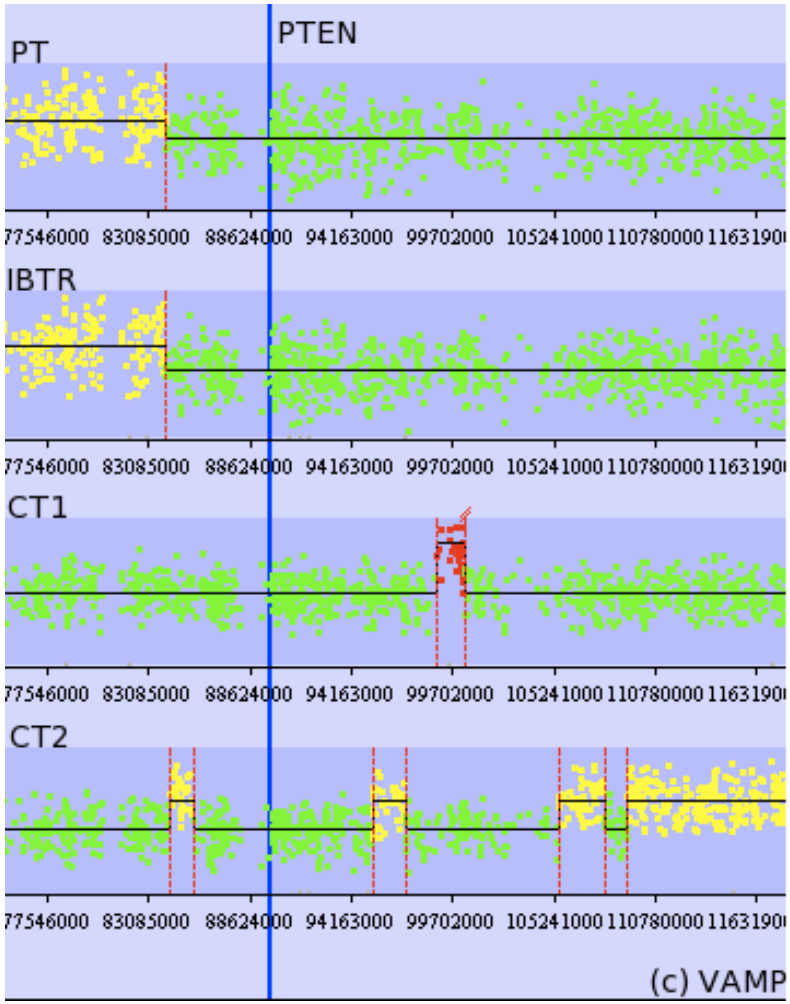
Data

Samples

- primary tumor (PT) and ipsilateral breast tumor recurrence (IBTR) for 22 patients
- 44 control breast tumors

Microarray data

- Affymetrix SNP 50k (Xba)
- copy number estimated using `ITALICS` (Rigaill *et al*, 2008)



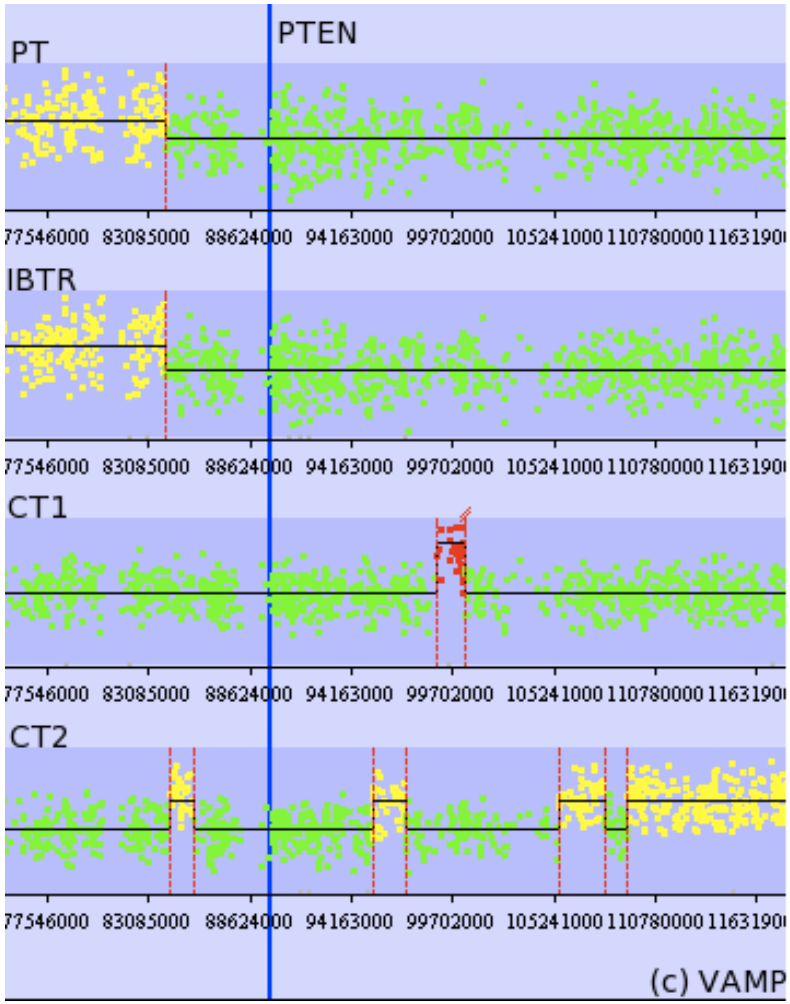
Outline

Biological idea

CNA vs breakpoints

- PTEN loss can be found in many breast cancers
- the breakpoint location is identical in the PT and IBTR of pair 5
- it is different for all other tumors in the study

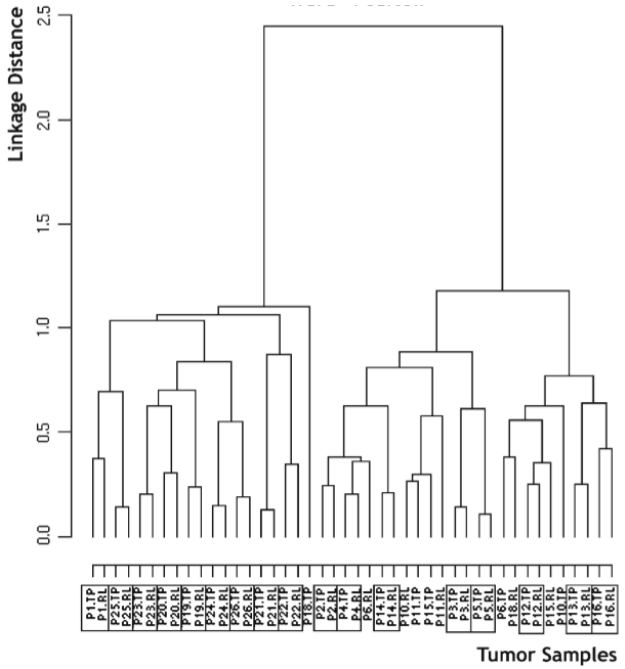
$\Rightarrow$ Use breakpoint locations as informative markers



Statistical idea

Current method to classify NP/TR

- hierarchical clustering of samples based on copy number alterations
- TR $\Longleftrightarrow$ IBTR and PT are neighbors on the dendrogram



Problems

- no significance estimation
- not robust to the addition/removal of a sample

$\Rightarrow$ using a score should be more appropriate

A partial identity score between tumors

Starting point: Dice's formula (*Ecology*, 1945)

$$S_D(i,j) = \frac{\text{number of common breakpoints between tumors } i \text{ and } j}{\text{mean number of breakpoints of } i \text{ and } j}$$

Proposed score

Taking breakpoint frequencies among controls into account:

$$PS(i,j) = \frac{\sum_{s \in S_i \cap S_j} (1 - f_s)^2}{\frac{1}{2}\left(\sum_{s \in S_i} (1 - f_s) + \sum_{s \in S_j} (1 - f_s)\right)}$$

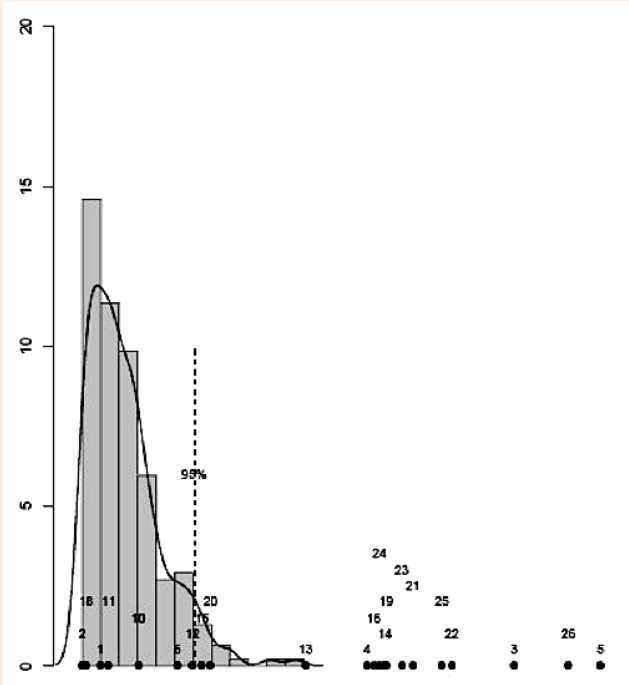
- S_k: set of breakpoints of tumor k
- f_s: frequency of breakpoint s among 44 control tumors

Statistical significance

Using “artificial pairs” to estimate a null distribution

Null hypothesis : no partial identity between the two tumors

1. match each of the 22 primary tumors with the IBTR of the 21 other patients
2. calculate the scores of all $22 \times 21 = 462$ such artificial pairs
3. true recurrence = IBTR with score higher 95% percentile



Outline

Results

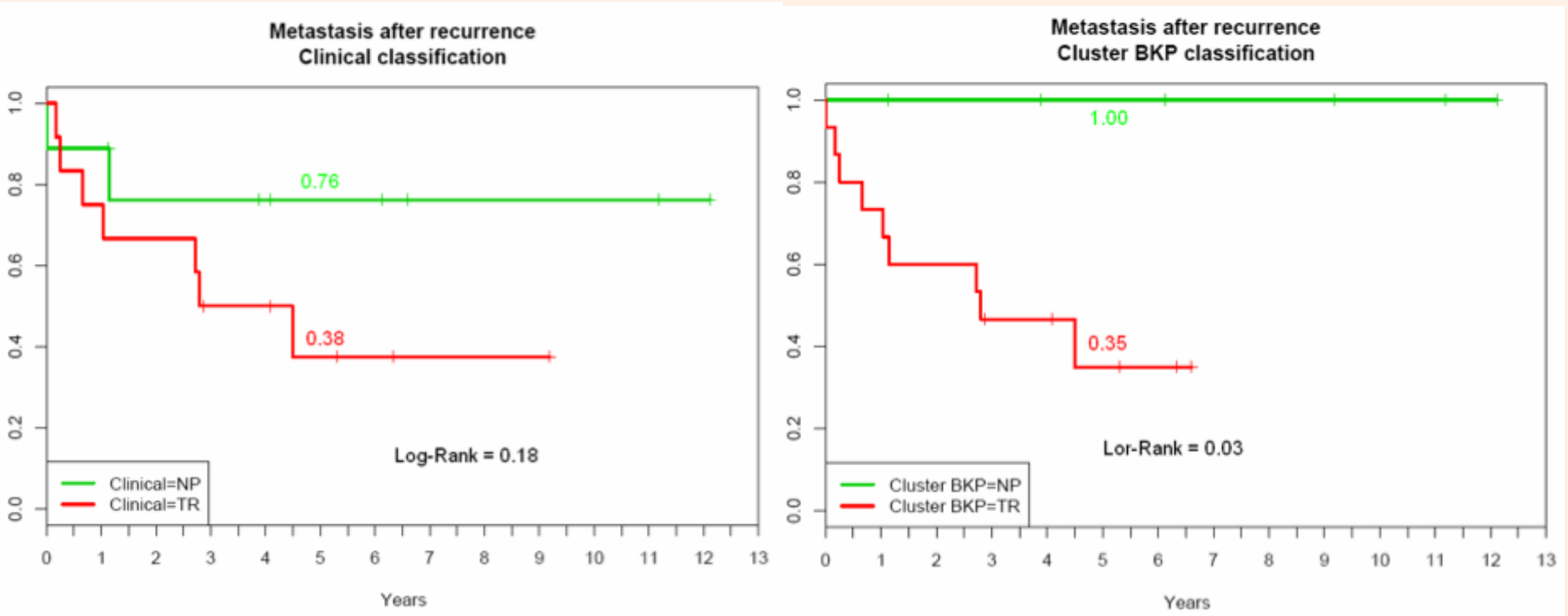
Assets of breakpoint information

- better concordance with clinical information than CNAs

Results

Assets of breakpoint information

- better concordance with clinical information than CNAs
- outperforms clinical information in terms of prognosis:



Further works

Differential analyses

Finding genes whose copy number differ between

- primary tumors whose IBTR is a true recurrence
- primary tumors from those whose IBTR is a new primary

Distant metastases vs primary tumors

- breast tumors often have ovarian metastases
- distinguish such metastases from primary ovarian cancers

Outline

Outline

Outline

Example: Differential analysis of gene expression data

Expression matrix from Golub data

expression levels of $m = 3051$ genes among $n = 38$ samples:

AML	Acute Myeloblastic Leukemia	$n_1 = 11$
ALL	Acute Lymphoblastic Leukemia	$n_2 = 27$

Goal

Find differentially expressed genes between AML and ALL

Example: Differential analysis of gene expression data

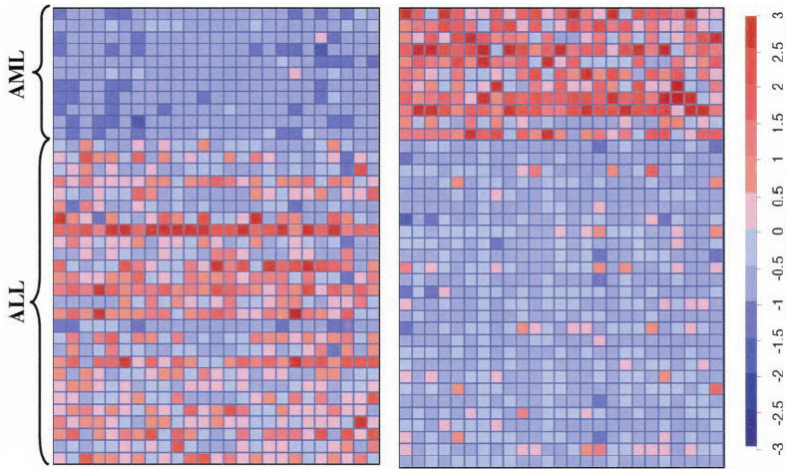
Expression matrix from Golub data

expression levels of $m = 3051$ genes among $n = 38$ samples:

AML	Acute Myeloblastic Leukemia	$n_1 = 11$
ALL	Acute Lymphoblastic Leukemia	$n_2 = 27$

Goal

Find differentially expressed genes between AML and ALL



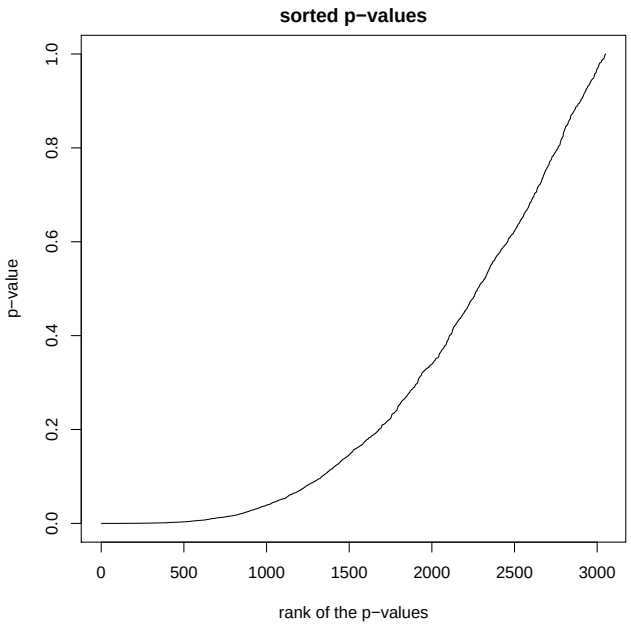
Mixture model

Notation and settings

- $\mathcal{H}_0, \mathcal{H}_1$: null and alternative hypotheses
- m: number of tested hypotheses
- π_0: (fixed) proportion of true null hypotheses

p-value distribution

- $P_i|\mathcal{H}_0 \sim \mathcal{U}(0,1)$
- $P_i|\mathcal{H}_1 \sim G_1$
- $(P_i)_{1\leq i\leq m} \overset{iid}{\sim} G$
- c.d.f. $G(x) = \pi_0 x + (1-\pi_0)G_1(x)$
- p.d.f. $g = \pi_0 + (1-\pi_0)g_1$



Multiple testing procedures

Multiple Testing Procedure (MTP)

$\mathcal{M} = (\mathcal{M}_m)_{m\in\mathbb{N}}$ such that $\mathcal{M}_m : [0,1]^m \to [0,1]$ rejects all hypotheses i verifying

$$P_i \leq \mathcal{M}_m(P_1, \dots P_m)$$

for any m-dimensional vector of p-values $(P_1, \dots P_m)$

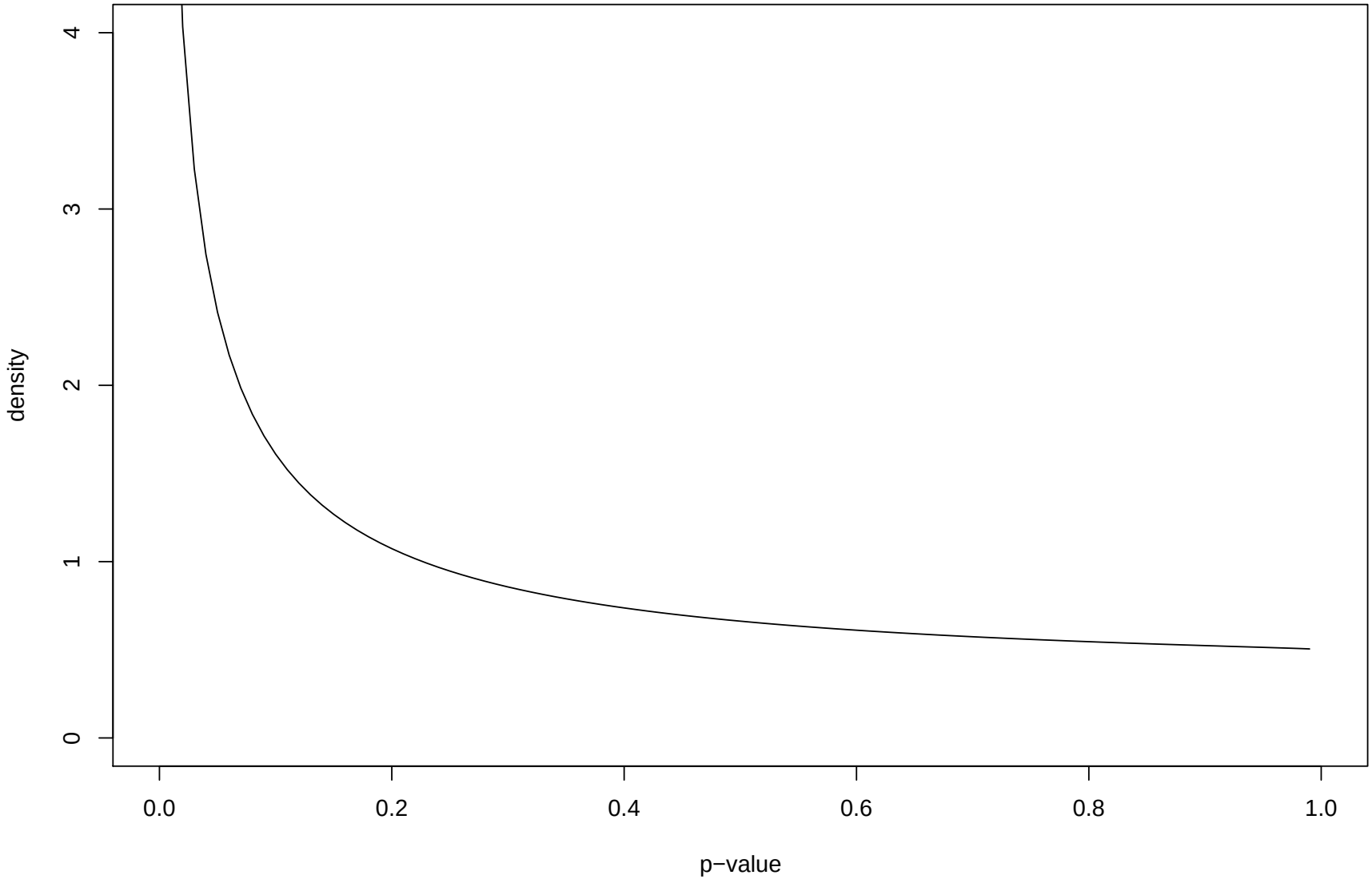
Threshold function

A multiple comparison procedure $\mathcal{M}$ has threshold function $\mathcal{T} : D[0,1] \to [0,1]$ iff

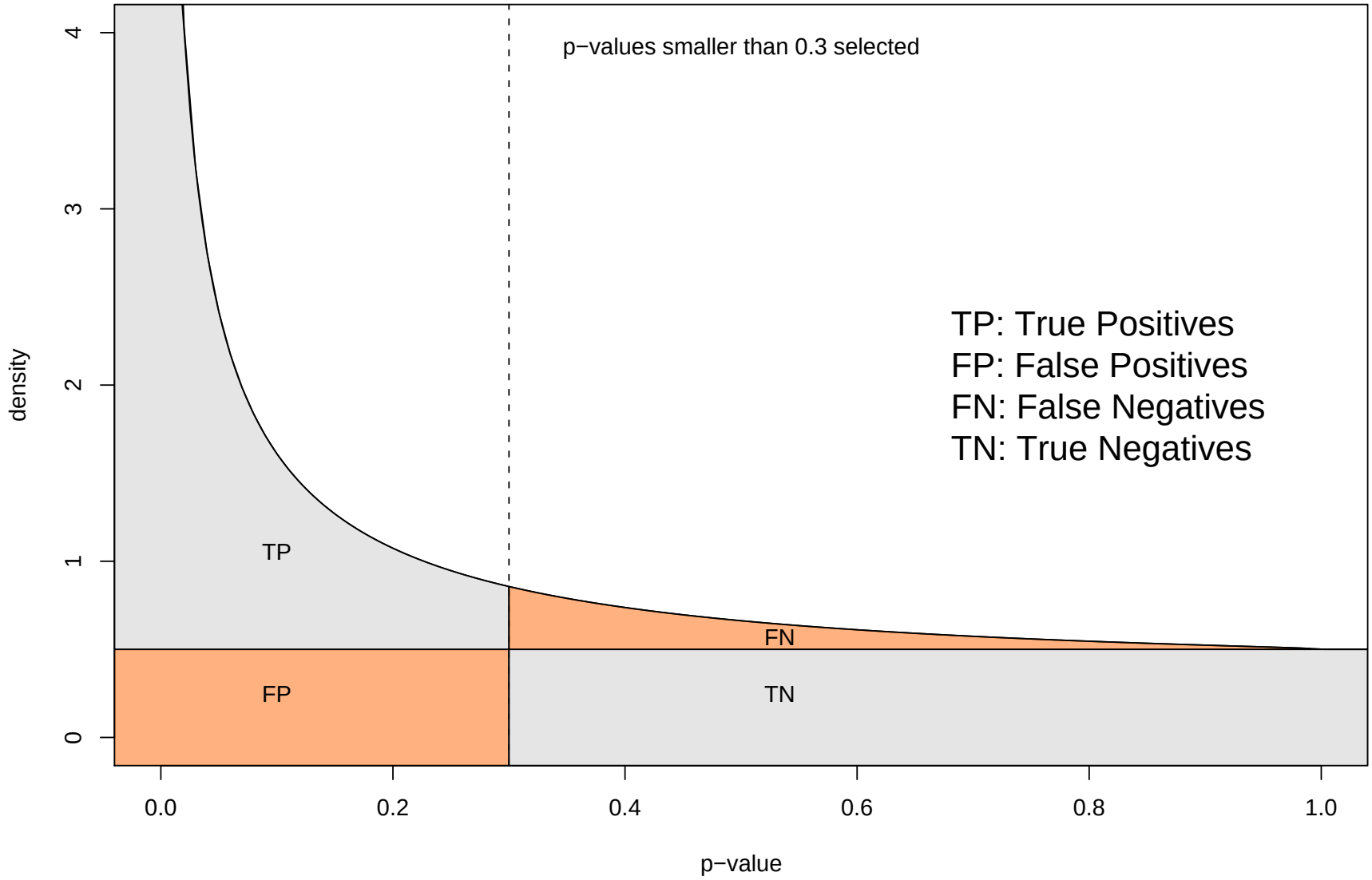
$$\forall m \in \mathbb{N}, \mathcal{M}_m(P_1, \dots P_m) = \mathcal{T}(\hat{\mathbb{G}}_m)$$

Outline

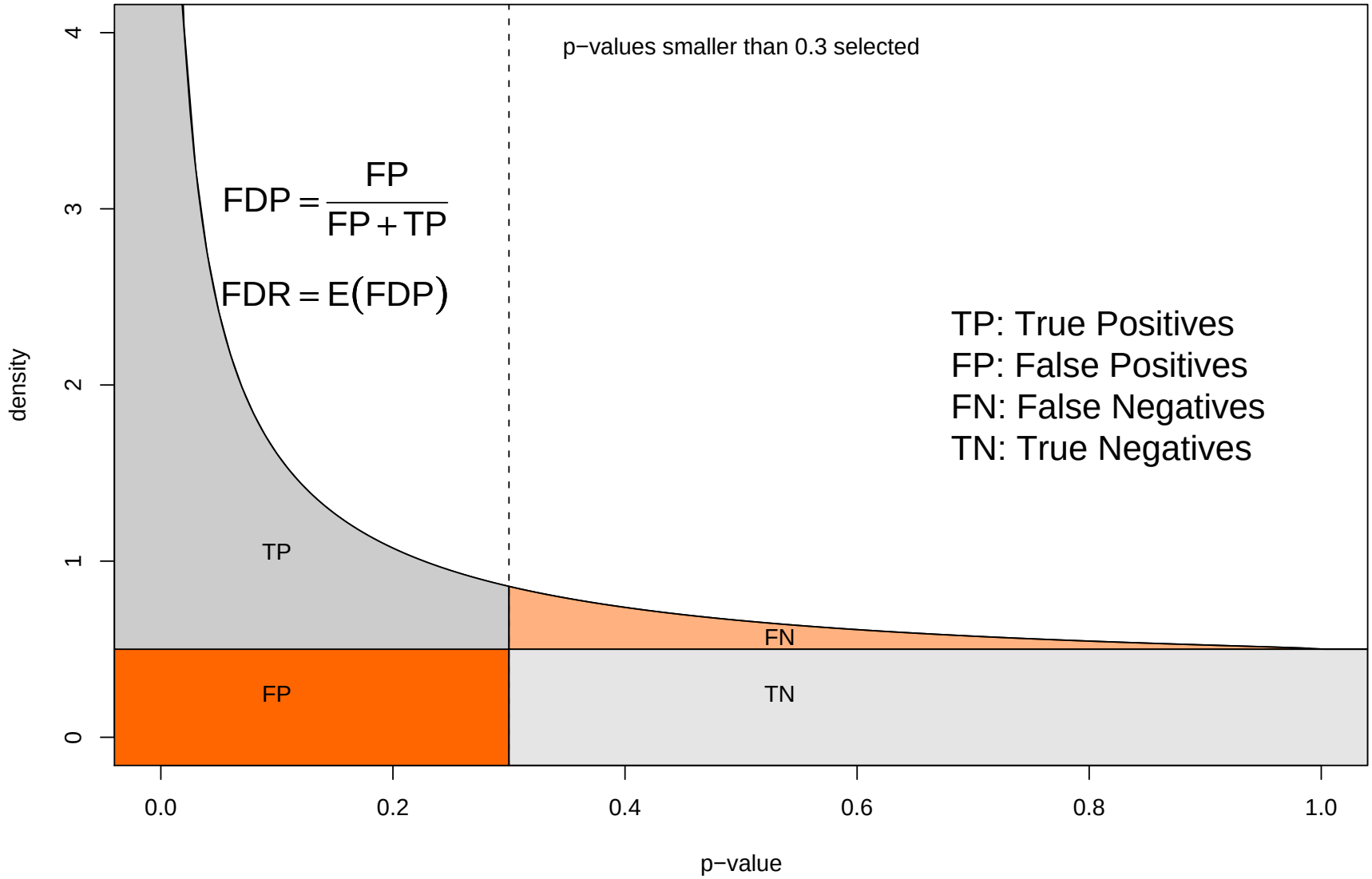
False Discovery Proportion and False Discovery Rate



False Discovery Proportion and False Discovery Rate



False Discovery Proportion and False Discovery Rate



FDP as a stochastic process of a random threshold

- we are interested in fluctuations of FDP around FDR
- $(FDP(t))_{0<t<1}$ as a stochastic process: (Genovese & Wasserman (Ann. Stat., 2004), Storey, Taylor & Siegmund (JRSS B, 2004))
- what about the FDP actually attained by a given MTP ?

$$FDP(\mathcal{T}(\hat{\mathbb{G}}_m))$$

Main idea

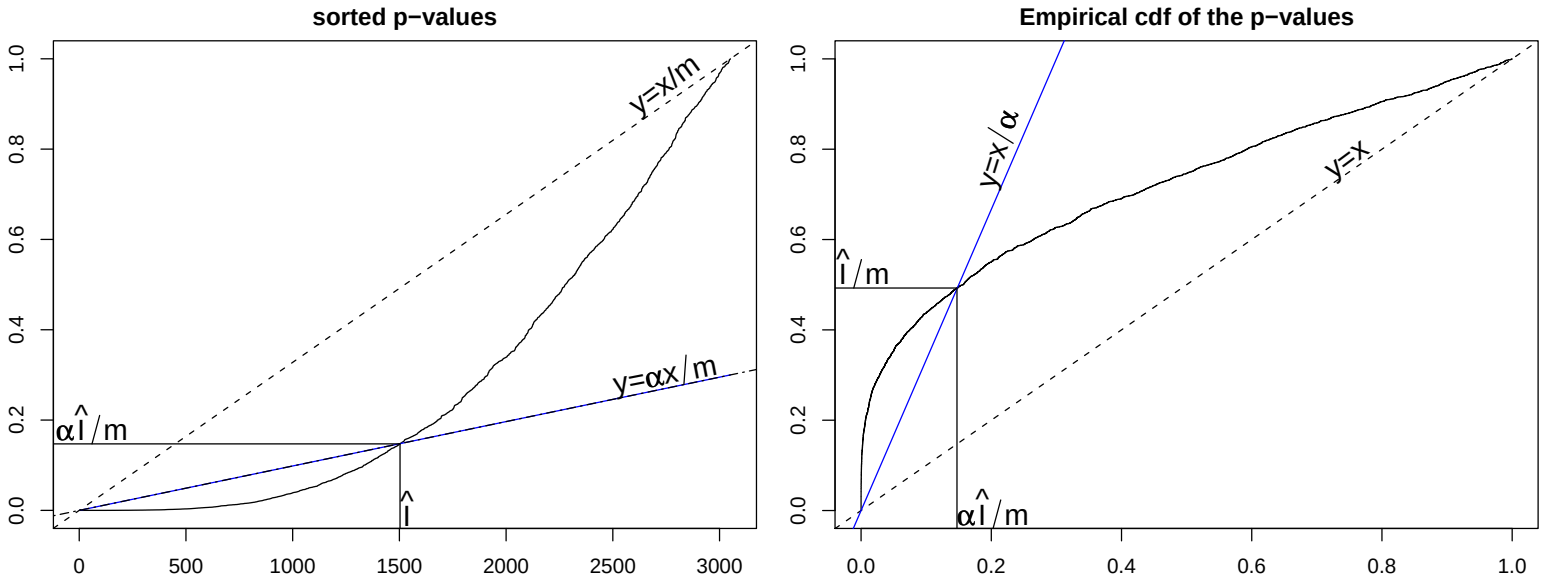
- derive asymptotic properties of the attained FDP from the functional Delta method
- understand how these properties rely on the procedure itself, and the *p*-value distribution

Outline

BH95 procedure (Benjamini & Hochberg, JRSS B, 1995)

The BH95 procedure at level α

1. Sort the m p-values : $P_{(1)} \leq \ldots \leq P_{(m)}$
2. Calculate $\widehat{l} = Max \left\{ k | P_{(k)} \leq \alpha \frac{k}{m} \right\}$
3. Reject all p-values smaller than $= \alpha \widehat{l}/m$



BH95 procedure: conservative FDR control

Threshold function of the BH95 procedure

$$\mathcal{T}(F) = \sup\{u \in [0,1], F(u) \geq u/\alpha\}$$

Conservativeness of the BH95 procedure

If the *p*-values are independent, procedure BH95 yields

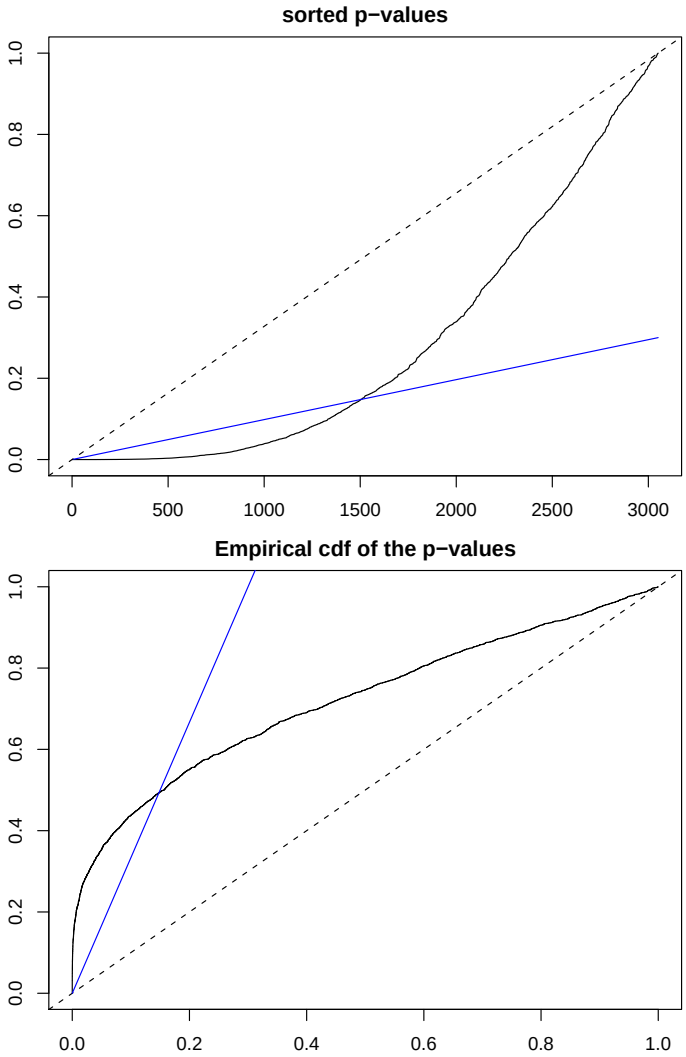
$$FDR \leq \pi_0 \alpha \leq \alpha$$

Two main types of refinements

plug-in procedures apply the BH95 procedure at level $\alpha/\widehat{\pi_0}$, where $\widehat{\pi_0}$ is an estimator of π_0

adaptive procedures use non-linear rejection curves, larger than Simes' line

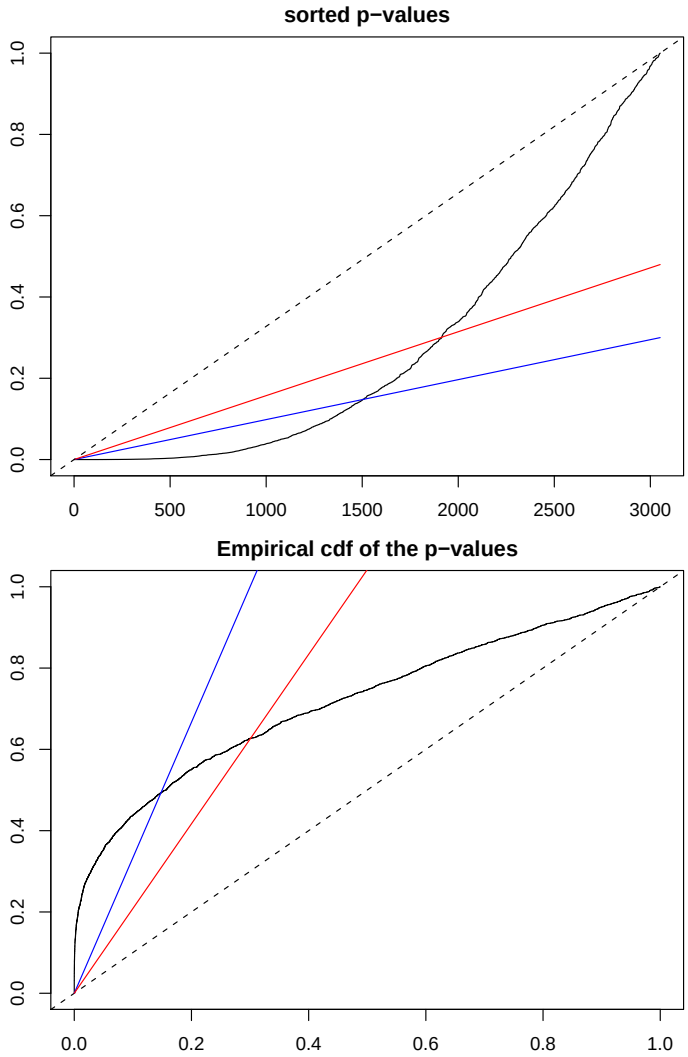
One-stage vs two-stage adaptive procedures



BH95 procedure

$$\mathcal{T}(F) = \sup\{u \in [0,1], F(u) \geq u/\alpha\}$$

One-stage vs two-stage adaptive procedures



BH95 procedure

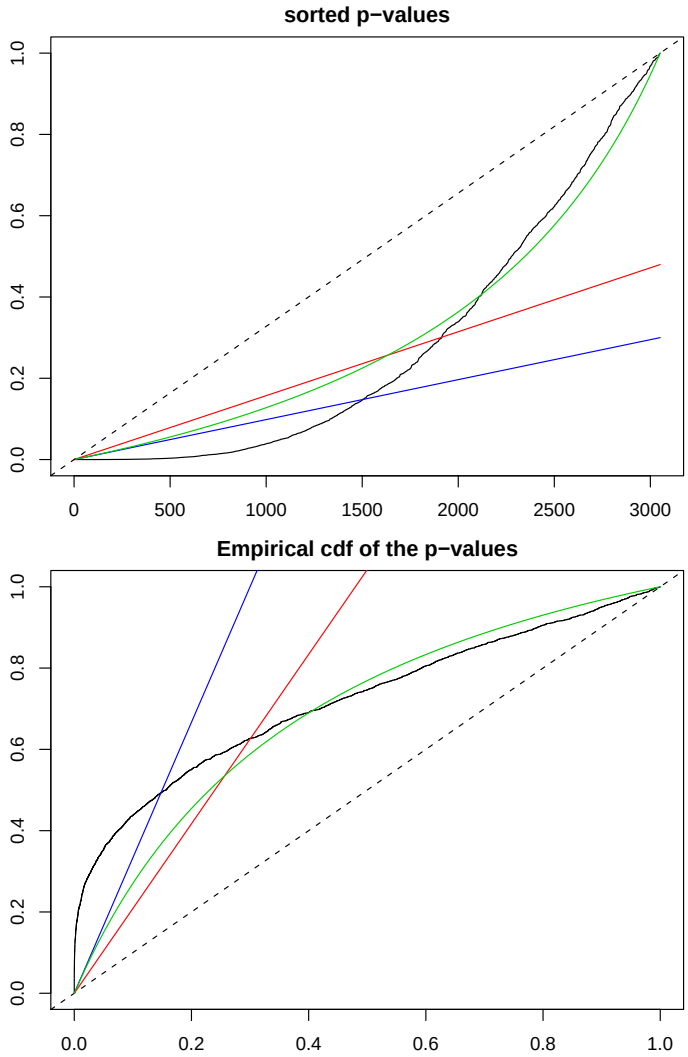
$$\mathcal{T}(F) = \sup\{u \in [0,1], F(u) \geq u/\alpha\}$$

Two-stage procedures: $\alpha \longrightarrow \alpha/\widehat{\pi_0}$

Sto02(λ): $\widehat{\pi_0}(\lambda) = \frac{1-\hat{\mathbb{G}}_m(\lambda)}{1-\lambda}$

BKY06 : $\widehat{\pi_0} = 1 - \hat{\mathbb{G}}_m(\widehat{\tau}^{\mathrm{BH95}}(\alpha))$

One-stage vs two-stage adaptive procedures



BH95 procedure

$$\mathcal{T}(F) = \sup\{u \in [0,1], F(u) \geq u/\alpha\}$$

Two-stage procedures: $\alpha \to \alpha/\widehat{\pi_0}$

Sto02(λ): $\widehat{\pi_0}(\lambda) = \frac{1-\hat{\mathbb{G}}_m(\lambda)}{1-\lambda}$

BKY06 : $\widehat{\pi_0} = 1 - \hat{\mathbb{G}}_m(\hat{\tau}^{\mathrm{BH95}}(\alpha))$

One-stage procedures: $u/\alpha \to r_\alpha$

FDR08: $r_\alpha(u) = u/(\alpha + (1-\alpha)u)$

BR08: $r_\alpha(u) = u/(\alpha + u)$

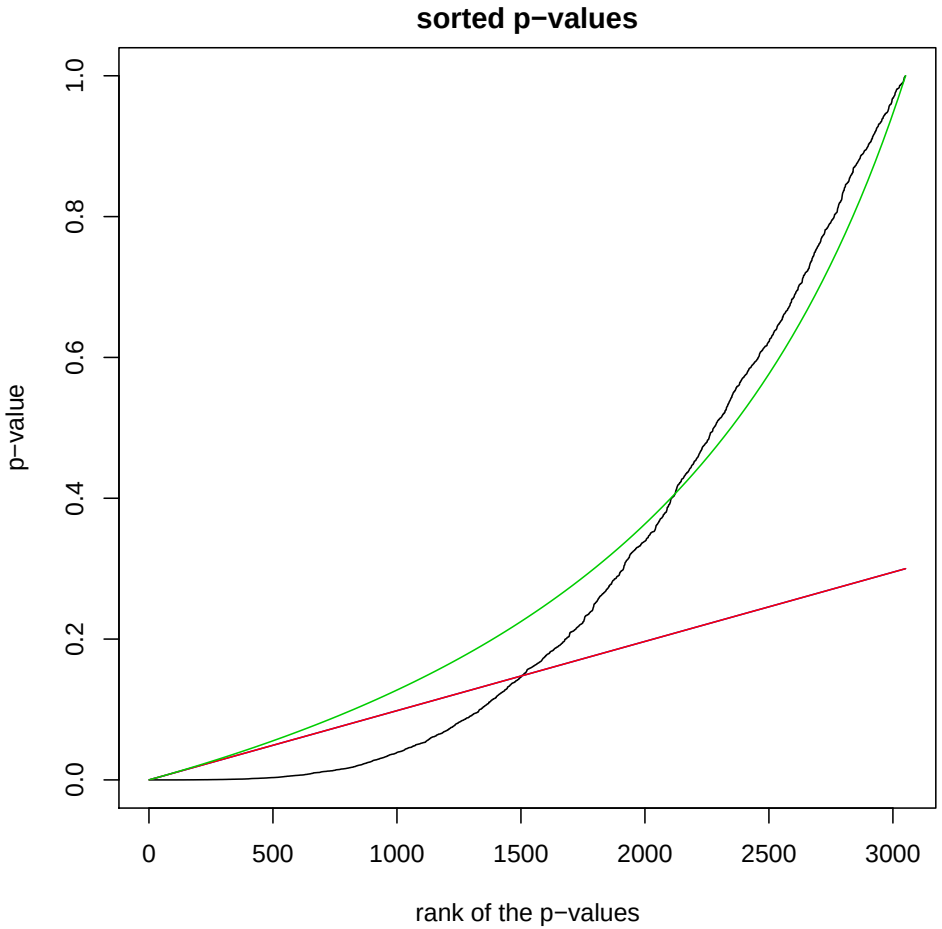
Outline

Outline

FDR08 as a fixed point of Sto02

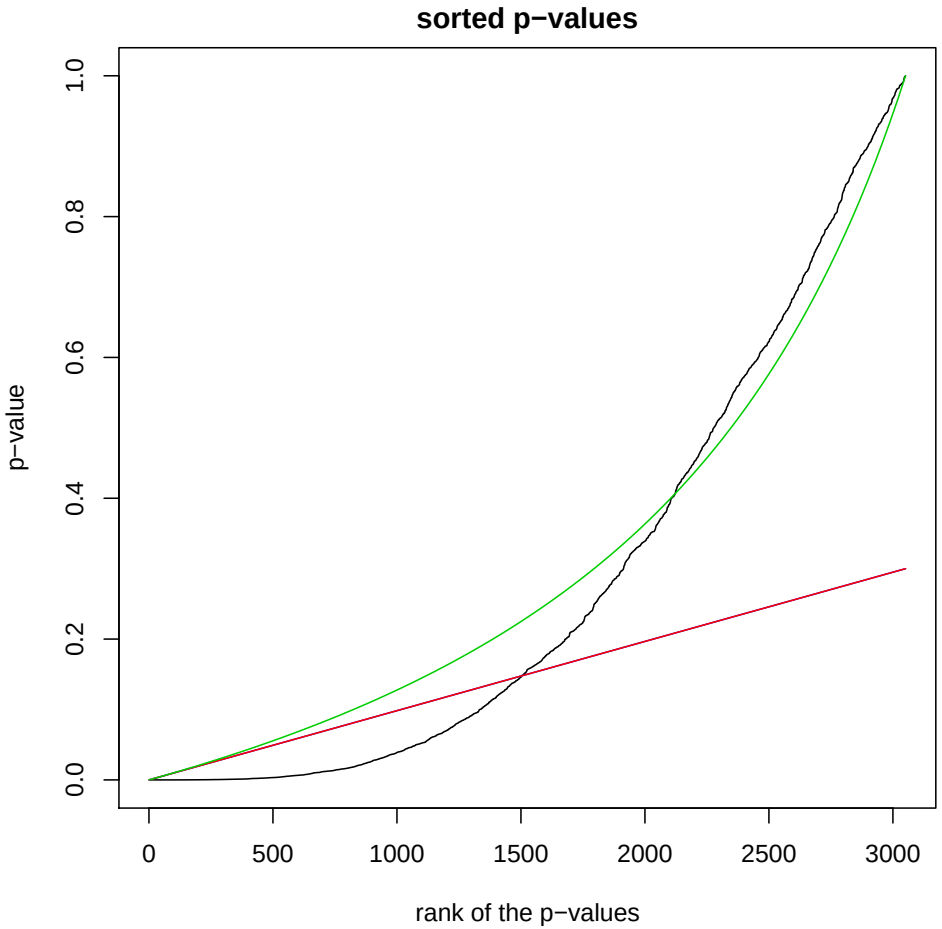
Procedure Sto02: apply BH95 at level $\alpha/\widehat{\pi_0}(\lambda)$

Procedure FDR08: a non-linear rejection curve



FDR08 as a fixed point of Sto02

Procedure Sto02: apply BH95 at level $\alpha/\widehat{\pi_0}(\lambda)$
Procedure FDR08: a non-linear rejection curve

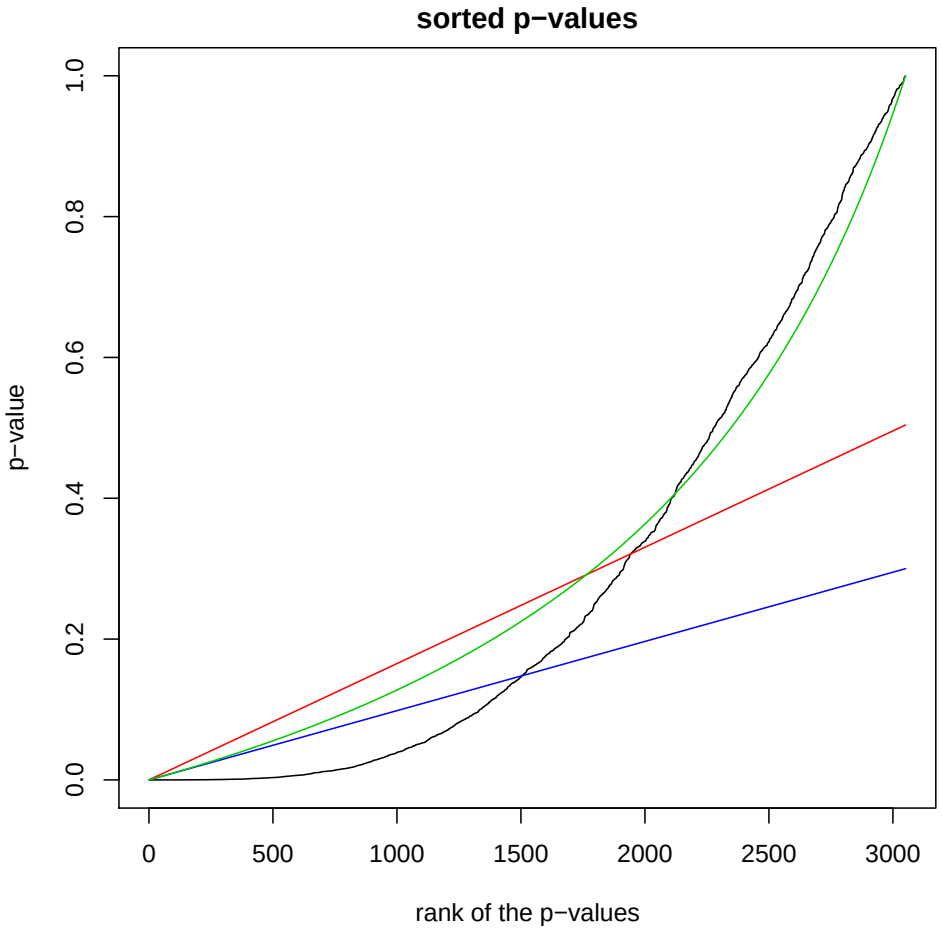


Iterate Procedure Sto02:

- choose $\lambda_0 \in (0, 1)$

FDR08 as a fixed point of Sto02

Procedure Sto02: apply BH95 at level $\alpha/\widehat{\pi_0}(\lambda)$
Procedure FDR08: a non-linear rejection curve

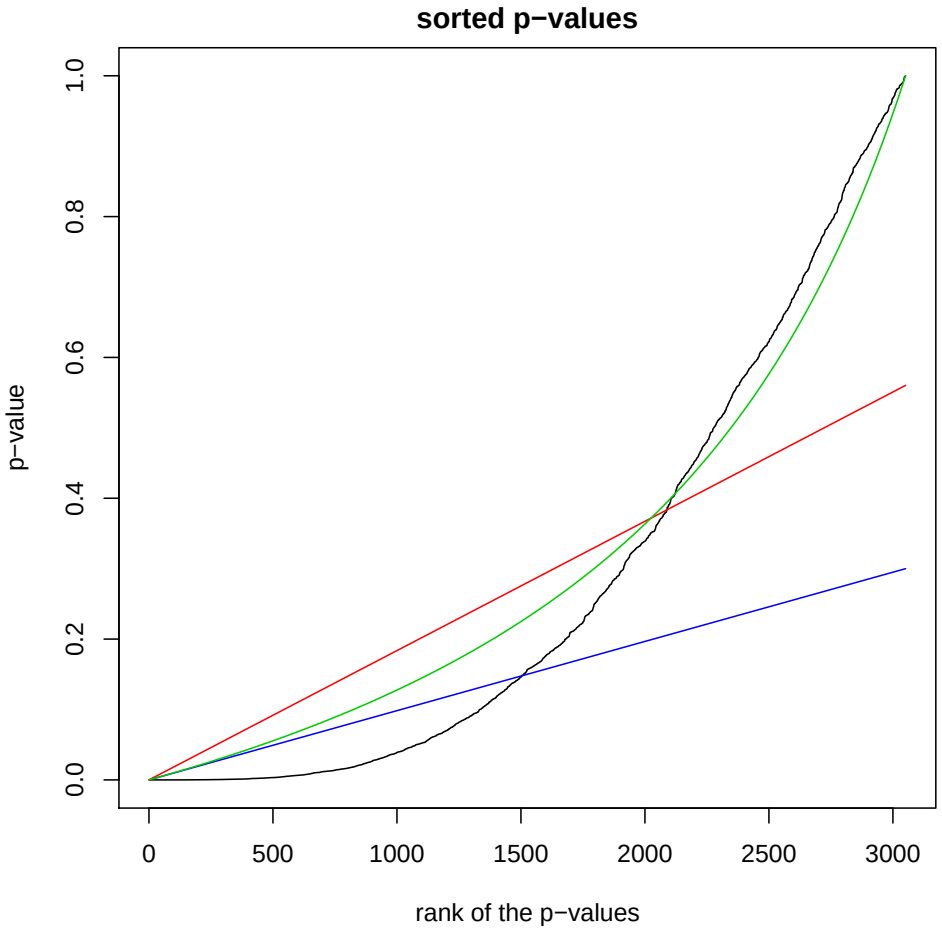


Iterate Procedure Sto02:

- choose $\lambda_0 \in (0, 1)$
- $\lambda_1 =$ rejection threshold of Sto02(λ_0)

FDR08 as a fixed point of Sto02

Procedure Sto02: apply BH95 at level $\alpha/\widehat{\pi_0}(\lambda)$
Procedure FDR08: a non-linear rejection curve

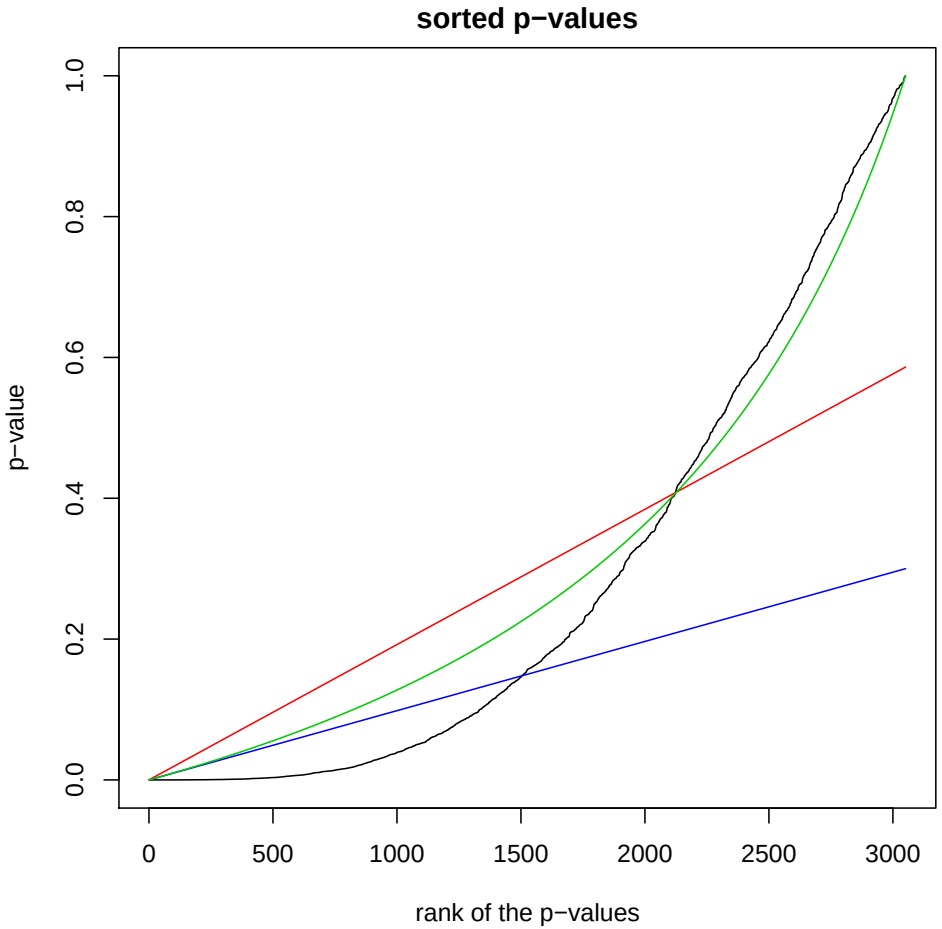


Iterate Procedure Sto02:

- choose $\lambda_0 \in (0, 1)$
- $\lambda_1 =$ rejection threshold of Sto02(λ_0)
- $\lambda_2 =$ rejection threshold of Sto02(λ_1)

FDR08 as a fixed point of Sto02

Procedure Sto02: apply BH95 at level $\alpha/\widehat{\pi_0}(\lambda)$
Procedure FDR08: a non-linear rejection curve

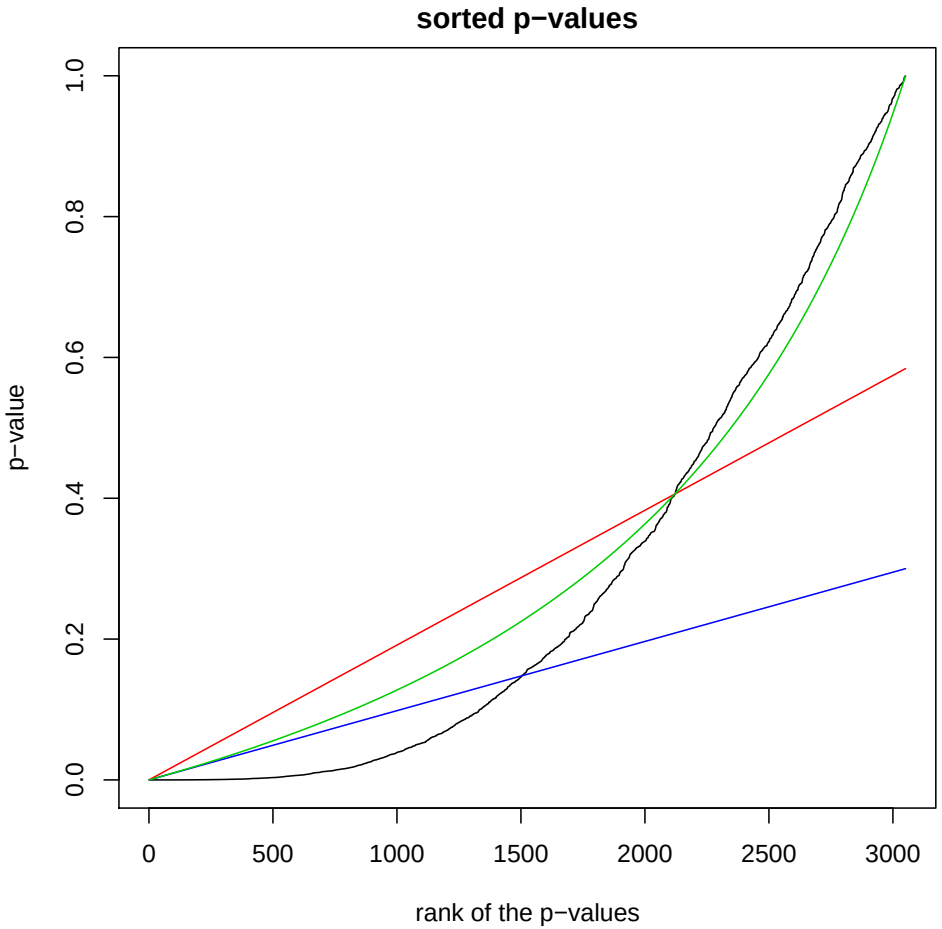


Iterate Procedure Sto02:

- choose $\lambda_0 \in (0, 1)$
- $\lambda_1 =$ rejection threshold of Sto02(λ_0)
- $\lambda_2 =$ rejection threshold of Sto02(λ_1)
- ...

FDR08 as a fixed point of Sto02

Procedure Sto02: apply BH95 at level $\alpha/\widehat{\pi_0}(\lambda)$
Procedure FDR08: a non-linear rejection curve



Iterate Procedure Sto02:

- choose $\lambda_0 \in (0, 1)$
- $\lambda_1 =$ rejection threshold of Sto02(λ_0)
- $\lambda_2 =$ rejection threshold of Sto02(λ_1)
- ...

(λ_n) converges to the rejection threshold of Procedure FDR08

Sto02 & FDR08: Comparison of threshold functions

$$\mathcal{T}^{\mathrm{Sto02}(\lambda)}(G) = \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha}\frac{1-G(\lambda)}{1-\lambda}\right\}$$

and

$$\begin{aligned}\mathcal{T}^{\mathrm{FDR08}}(G) &= \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha+(1-\alpha)u}\right\} \\ &= \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha}\frac{1-G(u)}{1-u}\right\}\end{aligned}$$

Comments

- self-consistency of the FDR08 procedure
- FDR08 is less conservative than Sto02 iff $\mathcal{T}^{\mathrm{FDR08}}(G) > \lambda$

Sto02 & FDR08: Comparison of threshold functions

$$\mathcal{T}^{\mathrm{Sto02}(\lambda)}(G) = \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha}\,\frac{1-G(\lambda)}{1-\lambda}\right\}$$

and

$$\begin{aligned}\mathcal{T}^{\mathrm{FDR08}}(G) &= \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha+(1-\alpha)u}\right\}\\ &= \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha}\,\frac{1-G(u)}{1-u}\right\}\end{aligned}$$

Comments

- self-consistency of the FDR08 procedure
- FDR08 is less conservative than Sto02 iff $\mathcal{T}^{\mathrm{FDR08}}(G) > \lambda$

BKY06 & BR08: Comparison of threshold functions

Letting $u_0 = \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha}\right\}$,

$$\mathcal{T}^{\text{BKY06}}(G) = \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha}(1 - G(u_0))\right\}$$

and

$$\begin{aligned}\mathcal{T}^{\text{BR08}}(G) &= \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha + u}\right\} \\ &= \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha}\ (1 - G(u))\right\}\end{aligned}$$

Comments

- self-consistency of the BR08 procedure
- BR08 is always more powerful than BKY06

BKY06 & BR08: Comparison of threshold functions

Letting $u_0 = \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha}\right\}$,

$$\mathcal{T}^{\text{BKY06}}(G) = \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha}\,(1 - G(u_0))\right\}$$

and

$$\begin{aligned}\mathcal{T}^{\text{BR08}}(G) &= \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha + u}\right\} \\ &= \sup\left\{u \in [0,1], G(u) \geq \frac{u}{\alpha}\ (1 - G(u))\right\}\end{aligned}$$

Comments

- self-consistency of the BR08 procedure
- BR08 is always more powerful than BKY06

Outline

Asymptotic distribution of FDP_m for procedure $\mathcal{T}$

Theorem

Let $\mathcal{T}$ be a threshold function, and $\tau^\star = \mathcal{T}(G)$. If $\mathcal{T}$ is Hadamard-differentiable at G, then

$$\sqrt{m}\left(FDP_m(\mathcal{T}(\hat{\mathbb{G}}_m)) - \frac{\pi_0 \tau^\star}{G(\tau^\star)}\right) \rightsquigarrow X,$$

where X is a centered Gaussian random variable whose variance depends on α, π_0, $\tau^\star$, and G.

- holds regardless of the form of the threshold function
- FDR is asymptotically controlled as soon as $\frac{\pi_0 \tau^\star}{G(\tau^\star)} \leq \alpha$

Asymptotic properties of the BH95 procedure

Theorem (BH95 procedure)

Let $\alpha^\star = 1/g(0)$, *and* $\tau^\star = \mathcal{T}(G)$. *If* $\alpha > \alpha^\star$, *then*

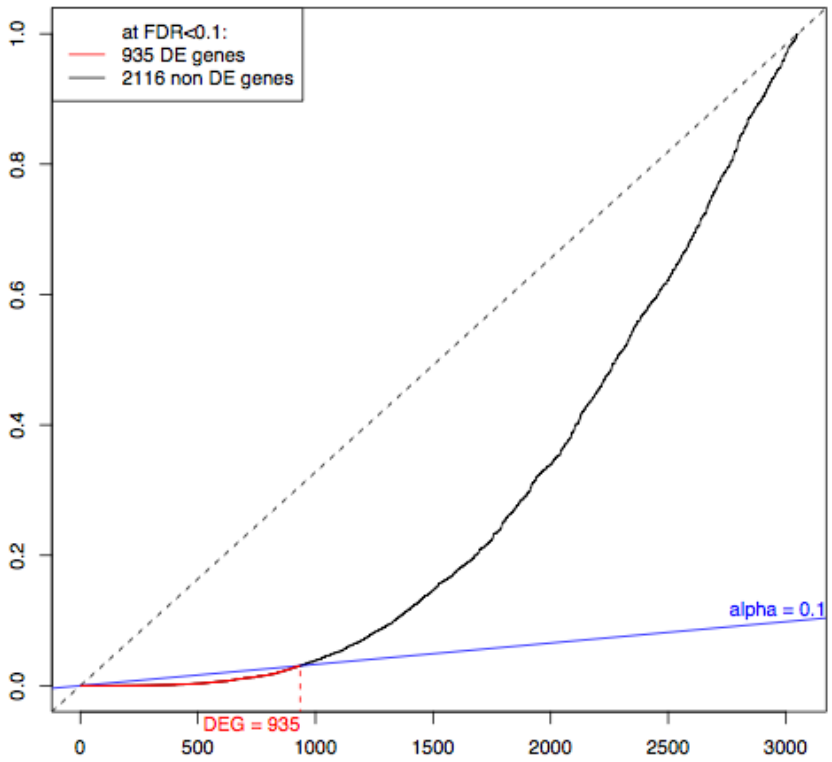
$$\sqrt{m}\left(FDP_m^{\mathrm{BH95}} - \pi_0\alpha\right) \rightsquigarrow \mathcal{N}\left(0, (\pi_0\alpha)^2 \frac{1-\tau^\star}{\tau^\star}\right)$$

Connection to criticality

$\alpha^\star$ is the critical value recently identified by Chi (Ann. Stat., 2007):

- if $\alpha < \alpha^\star$, the number of (true) discoveries is asymptotically bounded as the number of tested hypotheses increases;
- if $\alpha > \alpha^\star$, the proportion of discoveries converges in probability to a positive value $\tau^\star = \mathcal{T}(G)$.

Illustration of criticality



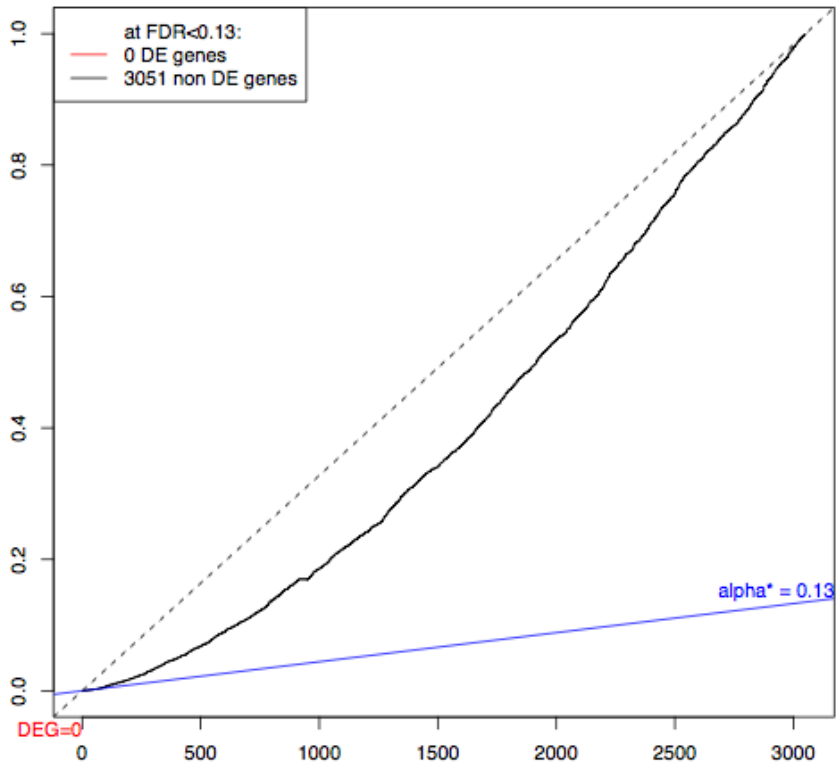
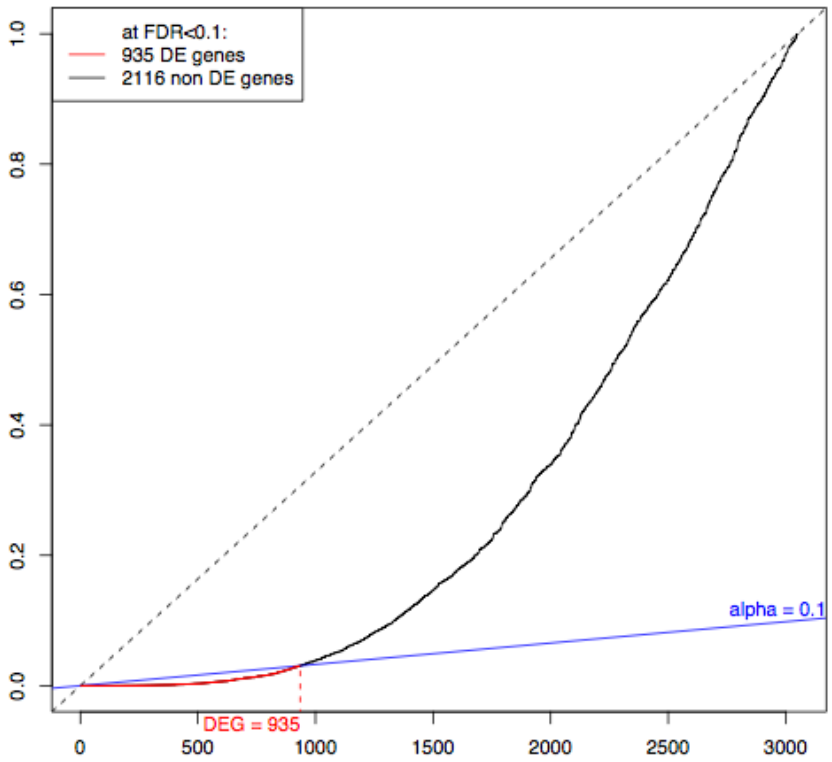


Illustration of criticality

27 vs 11 samples



8 vs 3 samples

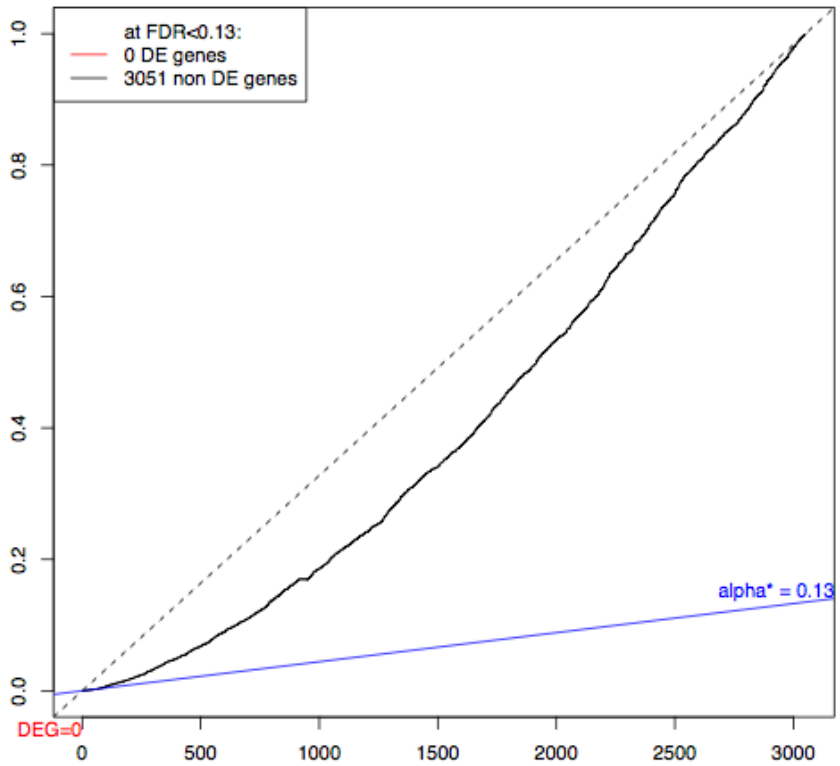
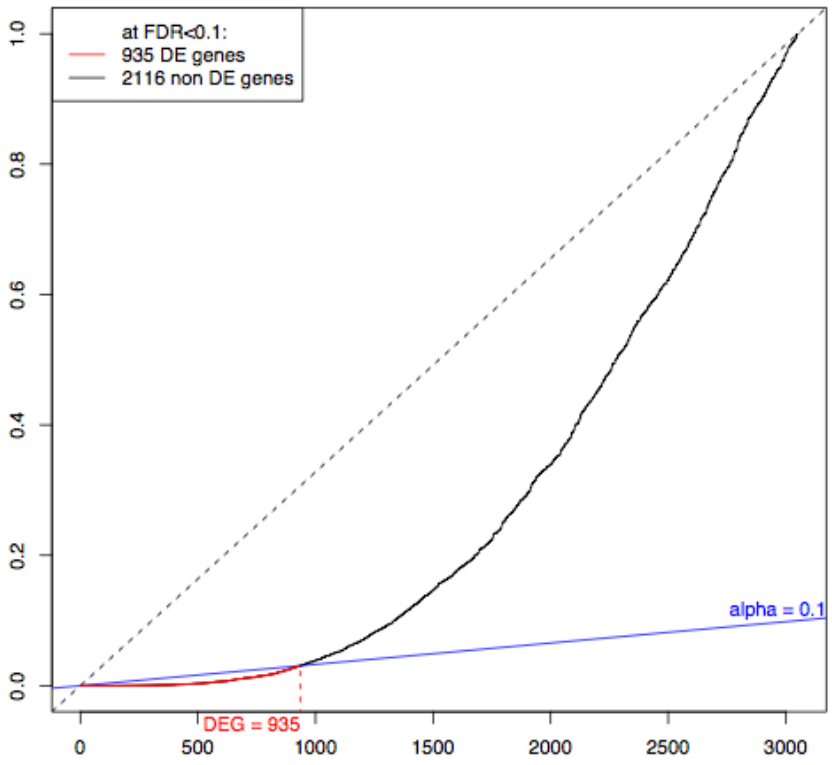
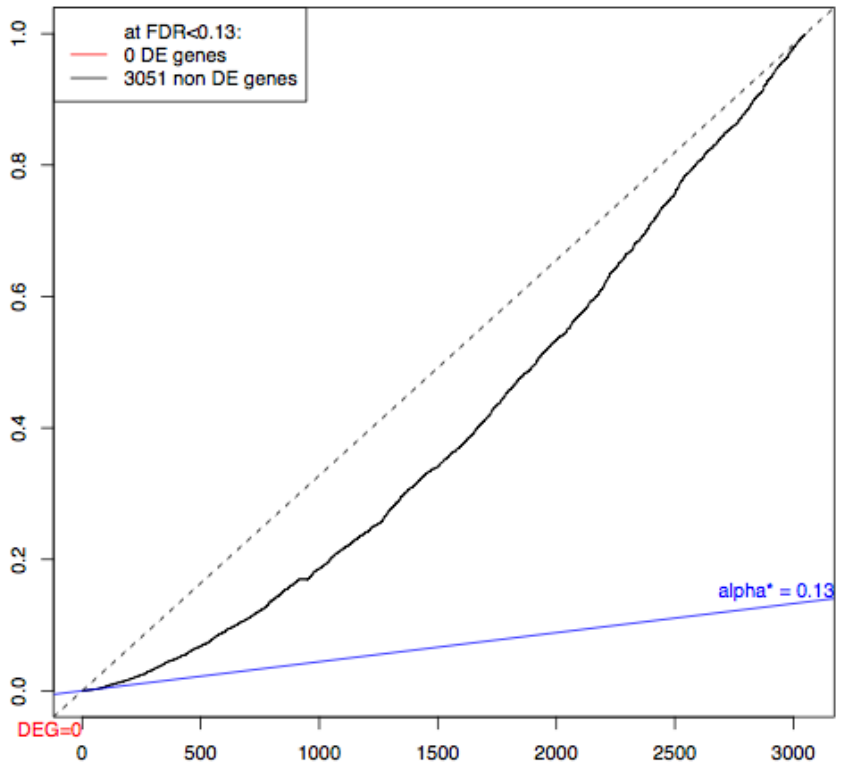


Illustration of criticality

27 vs 11 samples



8 vs 3 samples



$\Rightarrow$ Criticality vanishes as sample size increases

Conclusion

- Normalization of DNA copy number data

  P. Neuvial, P. Hupé *et al*, *BMC Bioinformatics*, 2006

- Joint analysis of DNA copy number and expression

  P. Neuvial, P. Gestraud *et al*, poster at ISMB 2007

- Unsupervised reconstruction of transcriptional regulatory networks

  M. Elati *et al*, *Bioinformatics*, 2007

- Definition of true recurrences among ipsilateral breast cancers

  M. Bollet, N. Servant *et al*, *JNCI*, 2008

- Asymptotic properties of multiple testing procedures

  P. Neuvial, in revision for *EJS*

- Intrinsic bounds and FDR control in multiple testing problems

  P. Neuvial, submitted to *JMLR*

Bonus

4 Appendix

- Asymptotic properties of Sto02 procedure
- Connections between one- and two-stage adaptive procedures

Bonus

4 Appendix
- Asymptotic properties of Sto02 procedure
- Connections between one- and two-stage adaptive procedures

Asymptotic properties of the Sto02(λ) procedure

$$\mathcal{T}(F) = \sup\left\{u \in [0,1], F(u) \geq \frac{u}{\alpha}\frac{1-F(\lambda)}{1-\lambda}\right\}$$

Theorem (Sto02(λ) procedure)

Let $\overline{\pi_0}(\lambda) = \frac{1-G(\lambda)}{1-\lambda}$, *and* $\tau^\star = \mathcal{T}(G)$. *If* $\alpha > \overline{\pi_0}(\lambda)\,\alpha^\star$, *then*

$$\sqrt{m}\left(FDP_m^{\text{Sto02}(\lambda)} - \frac{\pi_0}{\overline{\pi_0}(\lambda)}\alpha\right) \rightsquigarrow X^{\text{Sto02}(\lambda)},$$

where $X^{\text{Sto02}(\lambda)}$ *is a centered Gaussian random variable whose variance depends on* α, $\tau^\star$ *and* λ.

Optimal bandwidth — Storey's estimator

Theorem

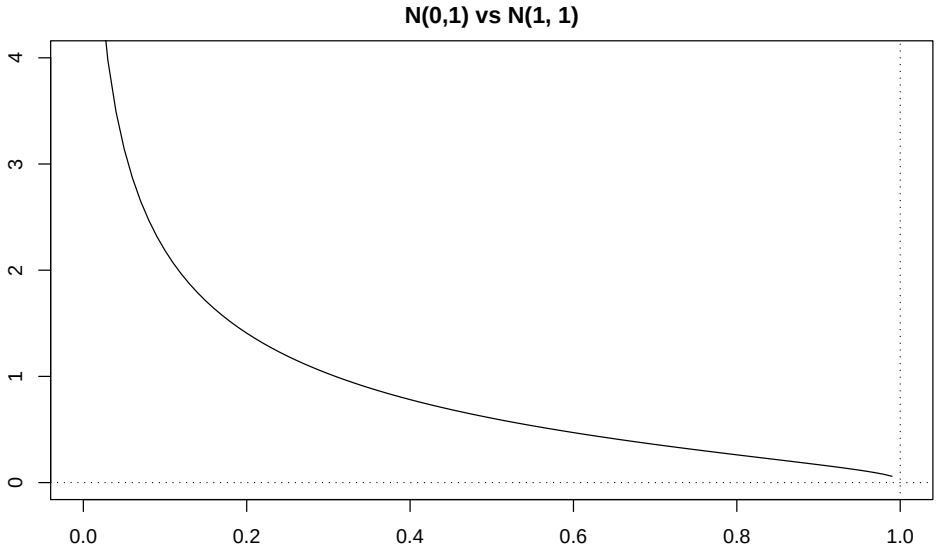
Assume that g is k times differentiable at 1, with $g^{(l)}(1) = 0$ for $0 \leq l < k$, and $g^{(k)}(1) \neq 0$.

1. *The optimal bandwidth in terms of MSE is given by $h_m(k) = C_k m^{-\frac{k}{2k+1}}$, where C_k is an explicit constant that depends on k, π_0, and $g^{(k)}(1)$;*
2. $$m^{\frac{k}{2k+1}} \left(FDP_m - \alpha \right) \rightsquigarrow \mathcal{N}\left(0, \frac{\alpha^2 C_k}{\pi_0} \right) .$$

Example: one-sided Gaussian location model

Proposition

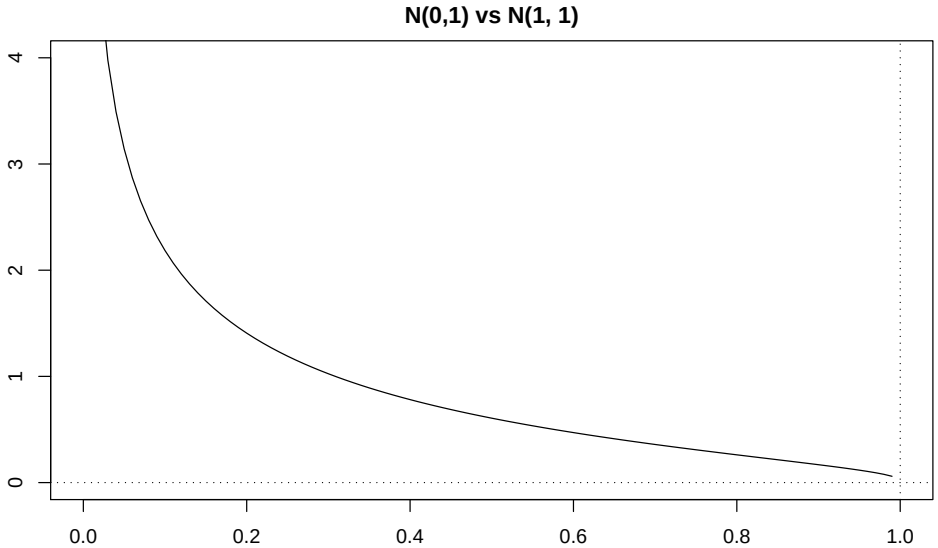
Assume that test statistics are distributed as $\mathcal{N}(0,1)$ *under* $\mathcal{H}_0$ *and as* $\mathcal{N}(\mu,1)$ *under* $\mathcal{H}_1$*. Then the p-value density under the alternative is not differentiable at* 1*.*

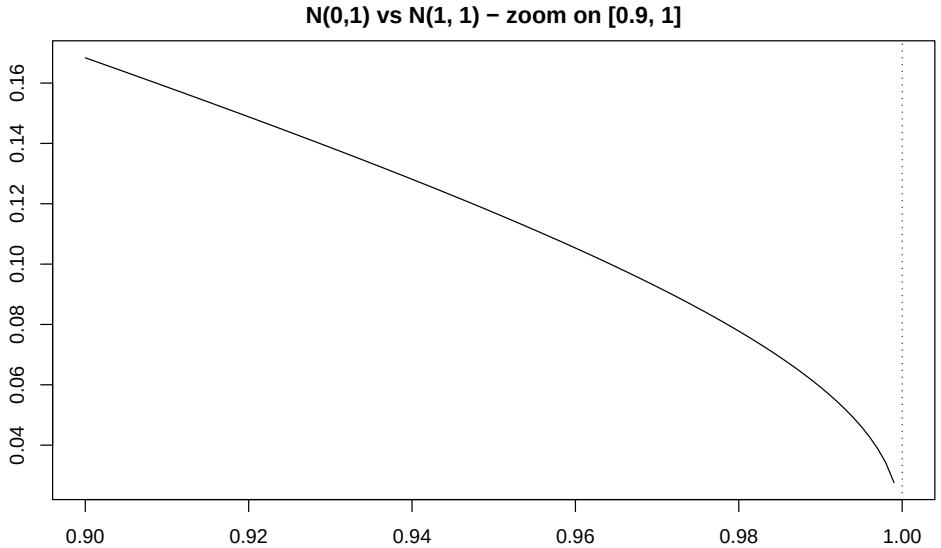


Example: one-sided Gaussian location model

Proposition

Assume that test statistics are distributed as $\mathcal{N}(0,1)$ *under* $\mathcal{H}_0$ *and as* $\mathcal{N}(\mu,1)$ *under* $\mathcal{H}_1$*. Then the p-value density under the alternative is not differentiable at* 1*.*





Bonus

4 Appendix
- Asymptotic properties of Sto02 procedure
- Connections between one- and two-stage adaptive procedures

FDR08 as a fixed point of Sto02

Theorem

For $\alpha \in [0,1]$, *and* $t_0 \in (0,1)$, *let*

- $\tau^\star = \mathcal{T}^{\mathrm{FDR08}}(G)$
- $\tau(u) = \mathcal{T}^{\mathrm{Sto02}(u)}(G)$ *for* $u \in [0,1]$
- $t_{i+1} = \tau(t_i)$ *for* $n \in \mathbb{N}$

If $\alpha^\star < \alpha < \pi_0$, *and if* G *and* f_α *have at most one interior crossing point, then*

$$\lim_{n\to\infty} t_n = \tau^\star$$

BR08 as a fixed point of BKY06

Theorem

For $\alpha \in [0,1]$, *and* $t_0 = 0$, *let*

- $\tau^\star = \mathcal{T}^{\mathrm{FDR08}}(G)$
- $\tau(u) = \mathcal{U}\left(G, \frac{\alpha/(1+\alpha)}{1-G(u)}\right)$
- $t_{i+1} = \tau(t_i)$ *for* $n \in \mathbb{N}$

With this notation, we have $u_0 = \tau(0)$, *and* $\mathcal{T}(G) = \tau(u_0)$ *is the asymptotic threshold of the* BKY06 *procedure.*

Assume that $\frac{\alpha}{1+\alpha} > \alpha^\star$ *and* $G\left(\frac{\alpha}{1+\alpha}\right) \leq \frac{1}{2}$. *If* G *and* b_α *have at most one interior crossing point, then*

$$\lim_{n\to\infty} t_n = \tau^\star$$

Adaptive multiple testing procedures

Y. Benjamini, A. M. Krieger, and D. Yekutieli.
Adaptive linear step-up procedures that control the false discovery rate.
Biometrika, 93(3):491, 2006.

G. Blanchard and E. Roquain.
Adaptive FDR control under independence and dependence.
Arxiv preprint math.ST/0707.0536v2, 2008.

H. Finner, T. Dickhaus, and M. Roters.
On the False Discovery Rate and an Asymptotically Optimal Rejection Curve.
Ann. Statist. (to appear).

J. D. Storey.
A direct approach to false discovery rates.
J. R. Stat. Soc. Ser. B Stat. Methodol., 64(3):479–498, 2002.