L'astuce bayésienne en test multiple

Etienne Roquain[1]

[1]Laboratory LPMA, Université Pierre et Marie Curie (Paris 6), France

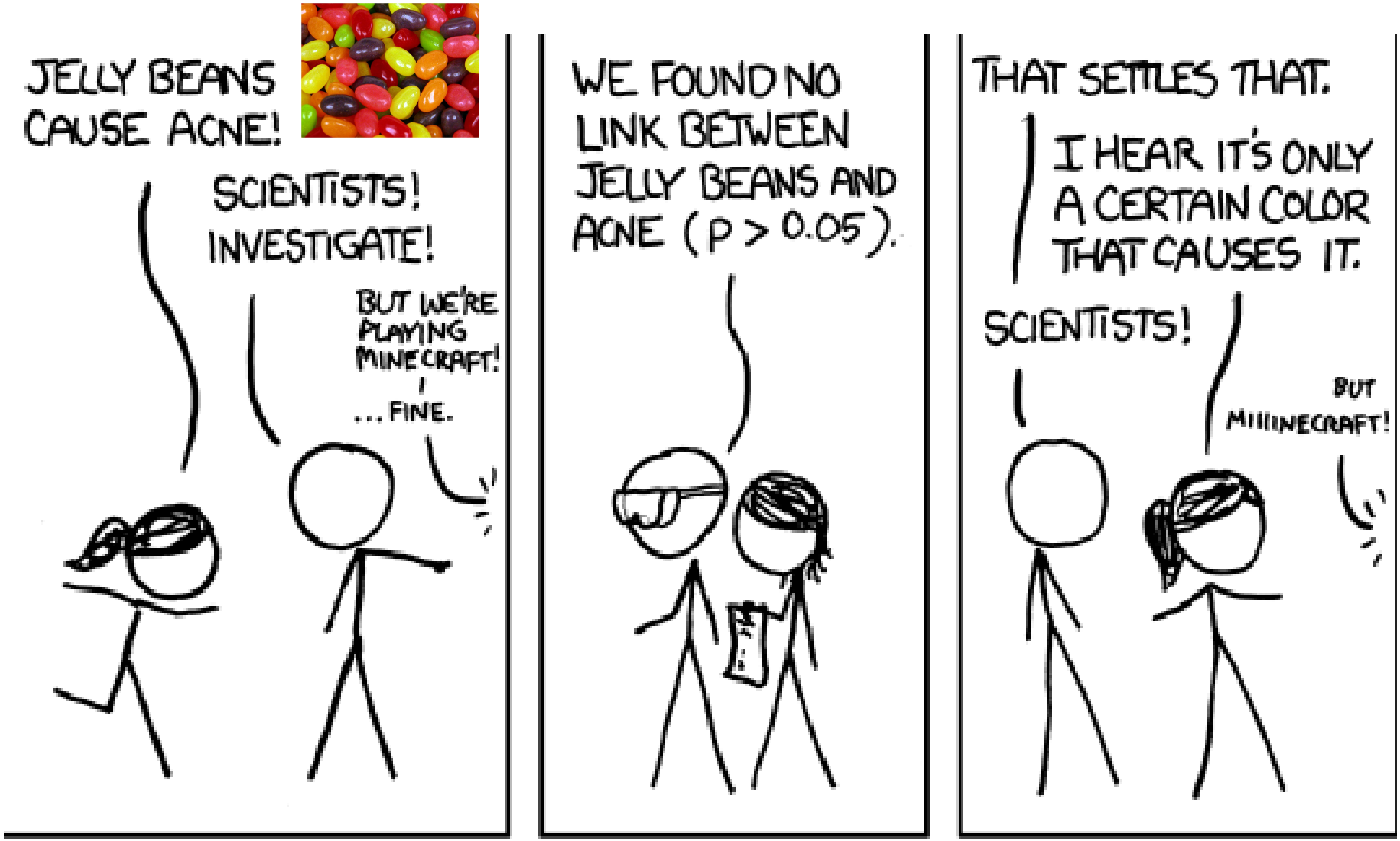
Journée AppliBUGS 2016

1 Introduction

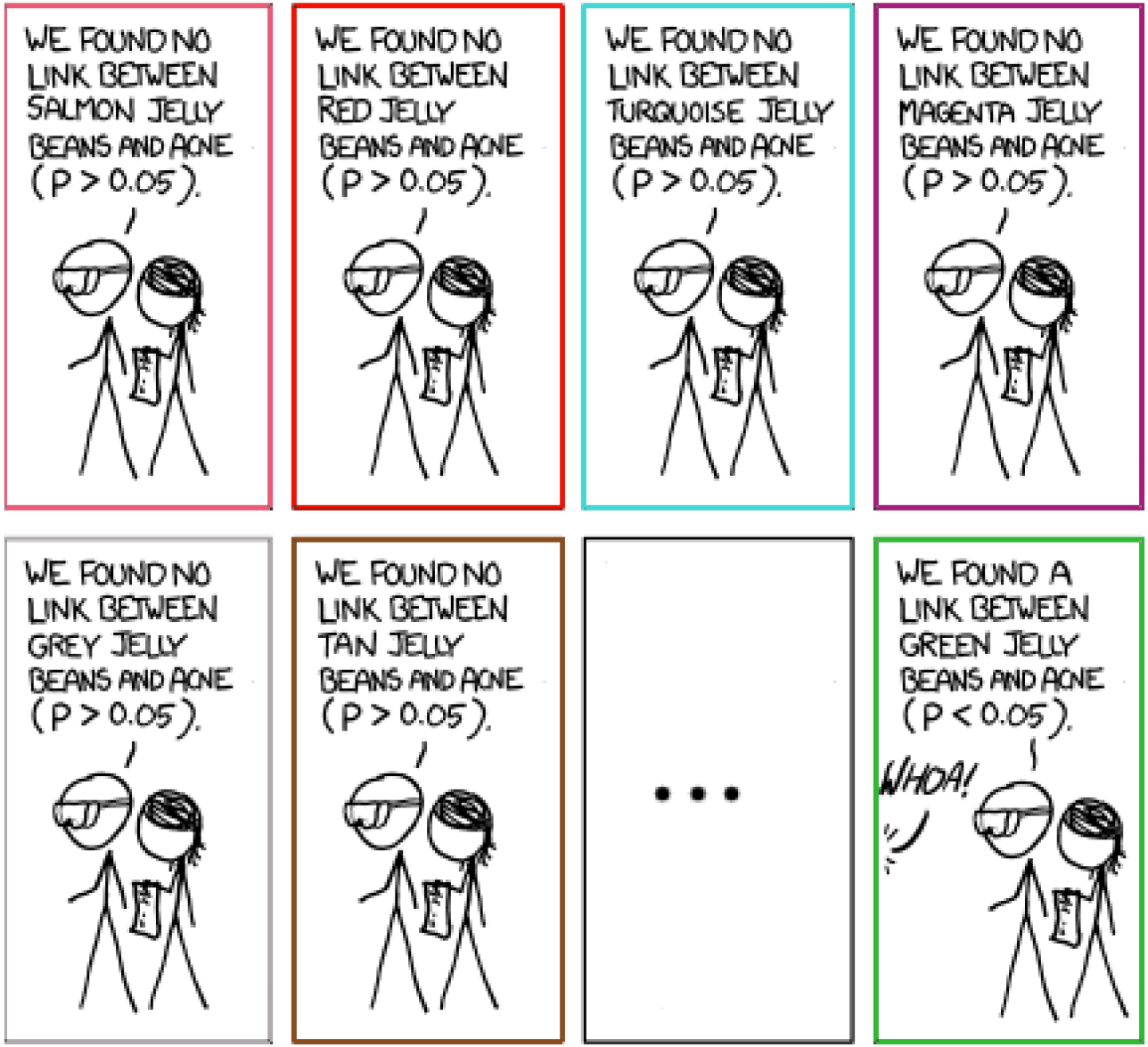
2 BH procedure

3 Bayesian tricks

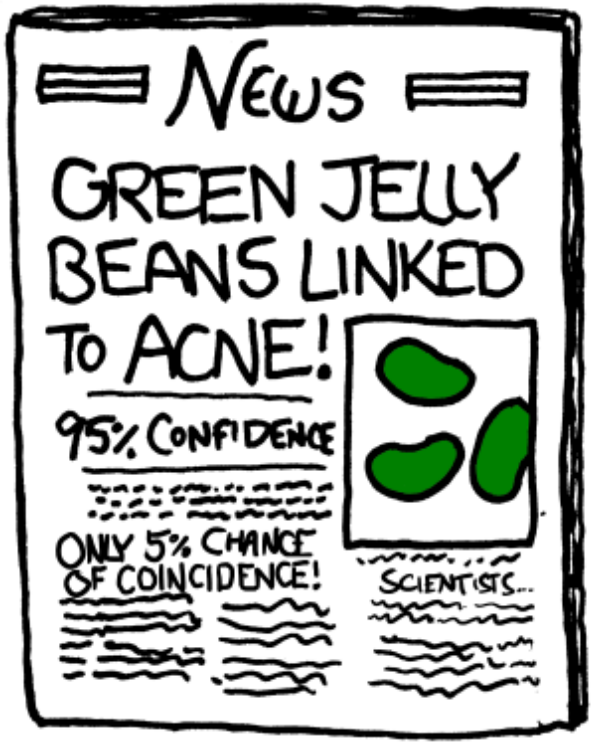
A “multiple testing story" (http://xkcd.com)



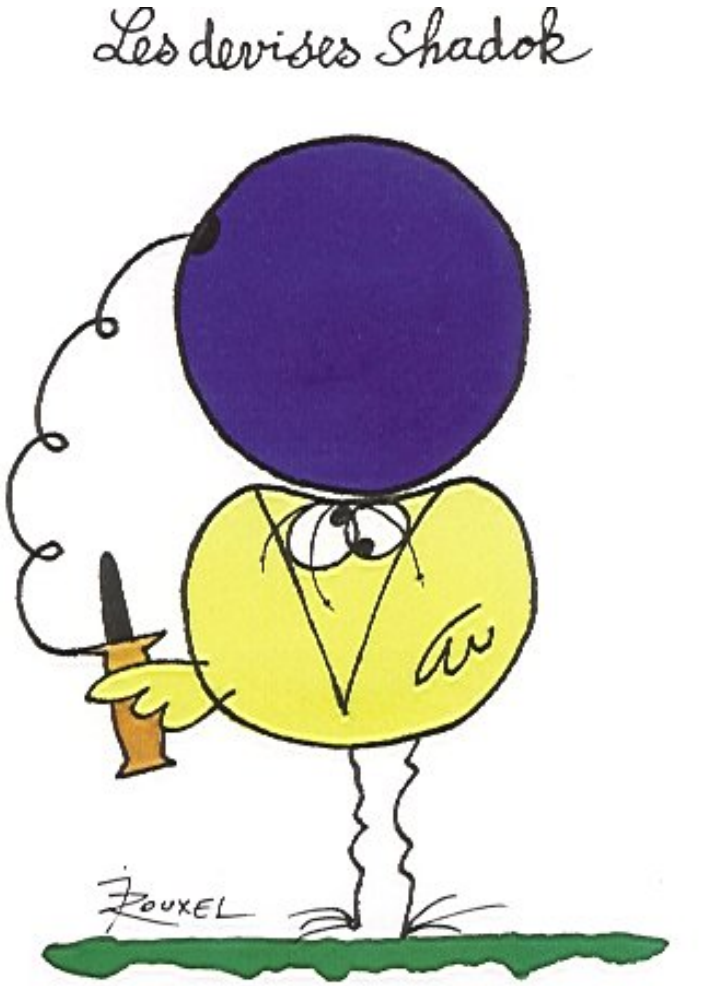
A “multiple testing story" (http://xkcd.com)



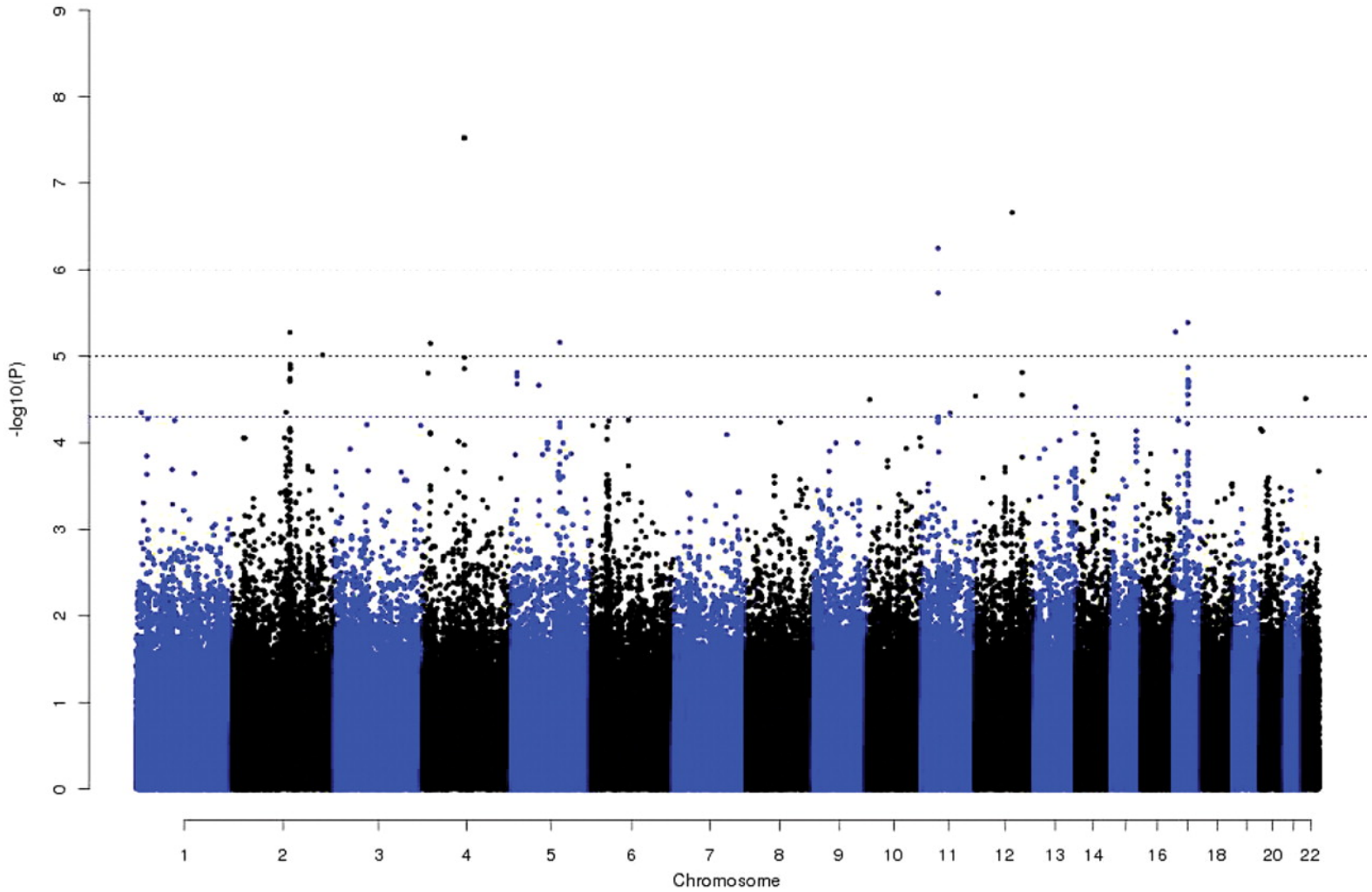
A “multiple testing story" (http://xkcd.com)



Shadoks motto



Manhattan plot (GWAS data)



- Find interesting SNPs
- Appropriate threshold ?

Simple multiple testing setting

Model:

- $\theta = (\theta_1, \ldots, \theta_m) \in \{0,1\}^m$ arbitrary
- $X = (X_1, \ldots, X_m) \mid \theta$ mutually independent,

$$X_i \sim \mathcal{N}(0,1) \text{ if } \theta_i = 0$$
$$X_i \sim \mathcal{N}(\mu,1) \text{ if } \theta_i = 1,$$

for some $\mu > 0$.

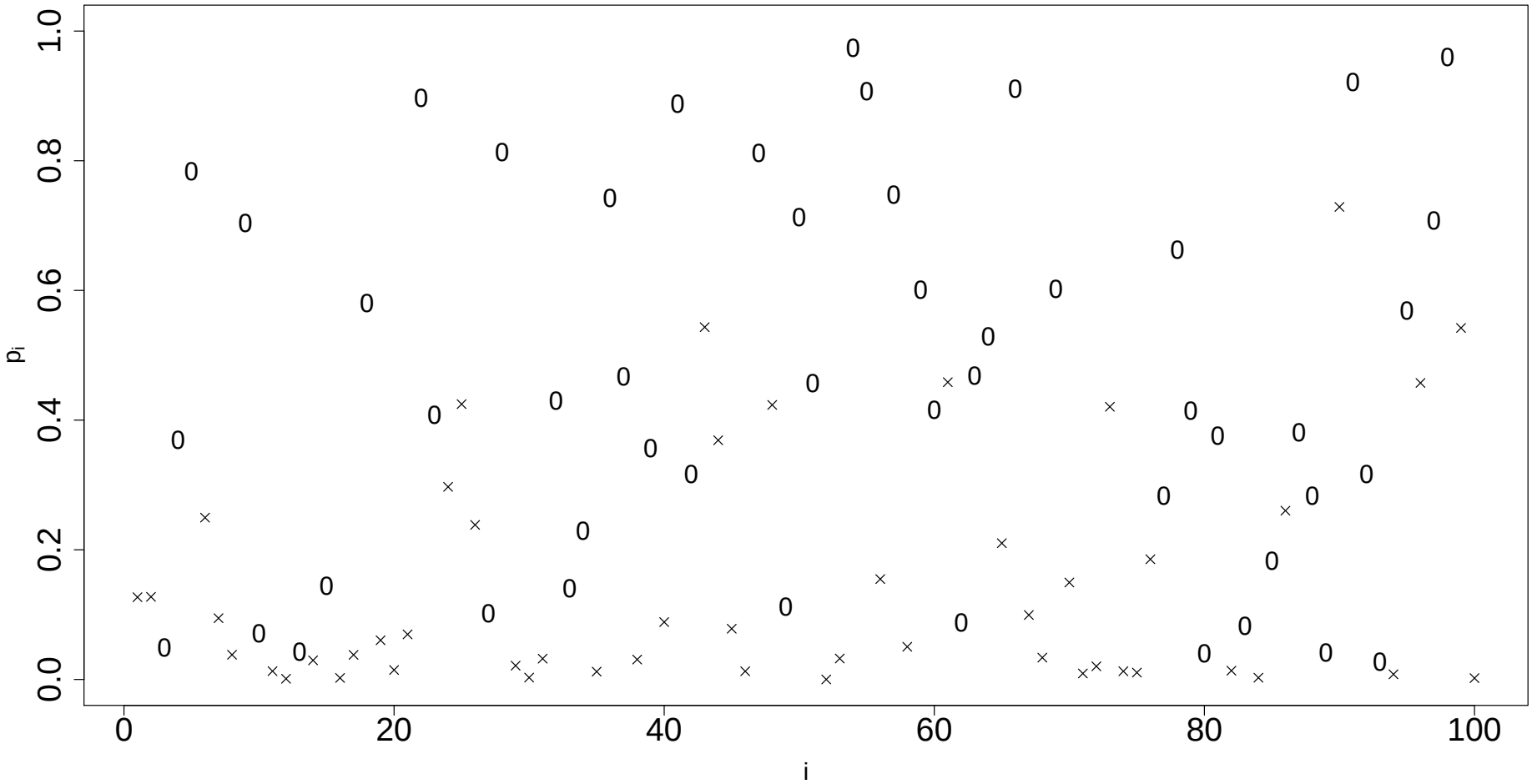
Aim: recover θ from $(X_i, 1 \leq i \leq m)$

Normalization:

- p-values

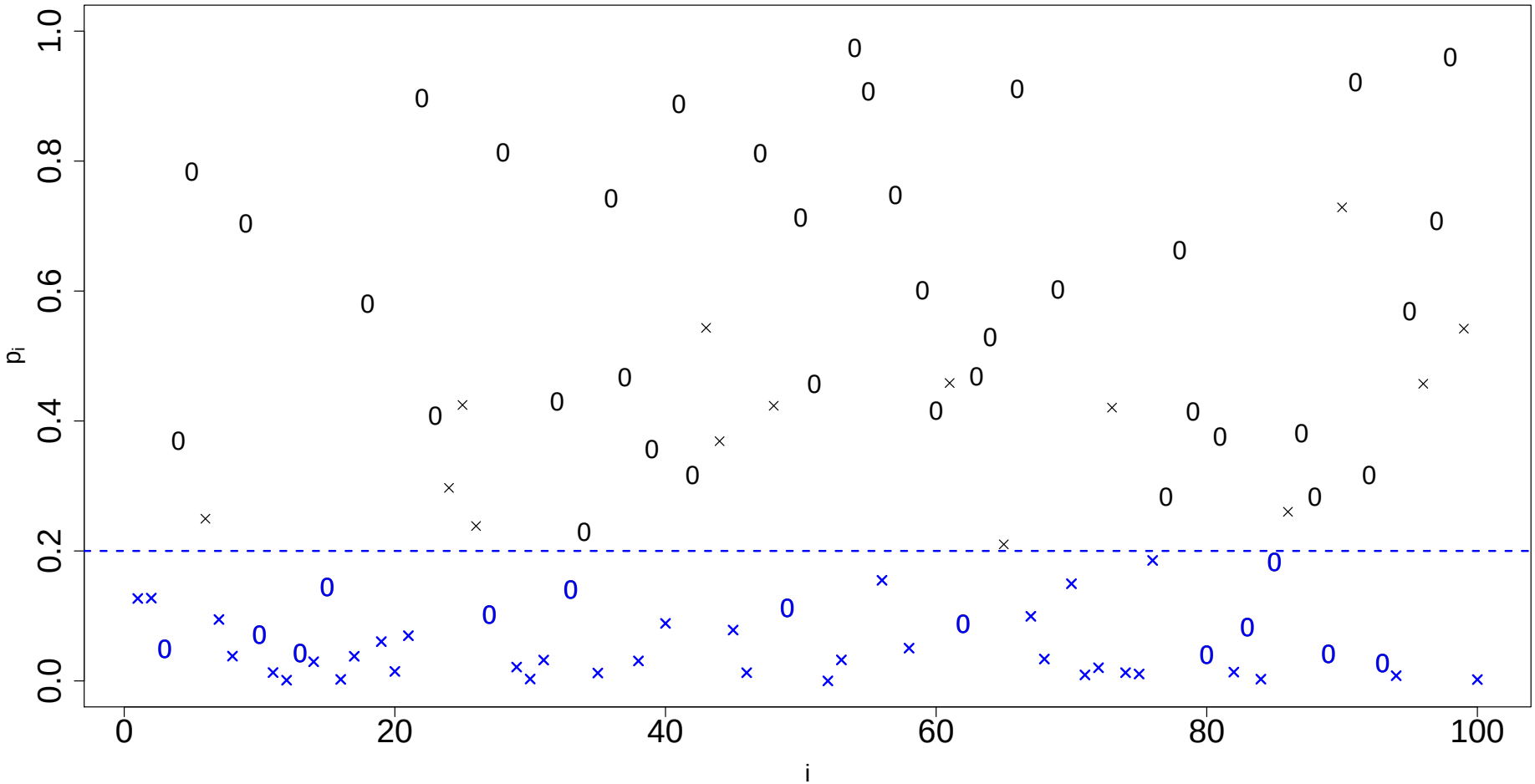
$$p_i(X) = \overline{\Phi}(X_i), \quad 1 \leq i \leq m$$

Thresholding



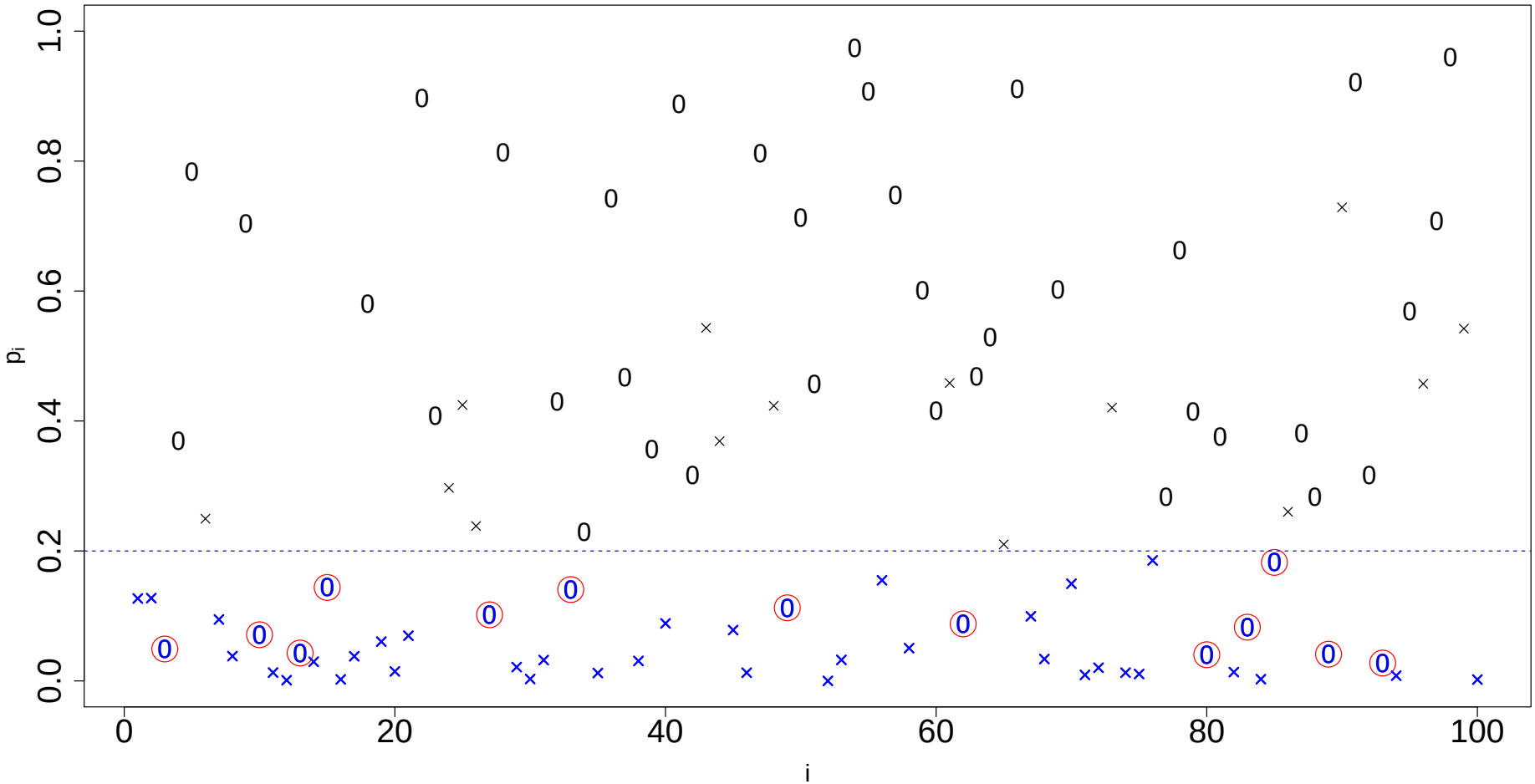
- Rejection set $R = \{1 \leq i \leq m \ : \ p_i(X) \leq t\}$
- Some false positives

Thresholding



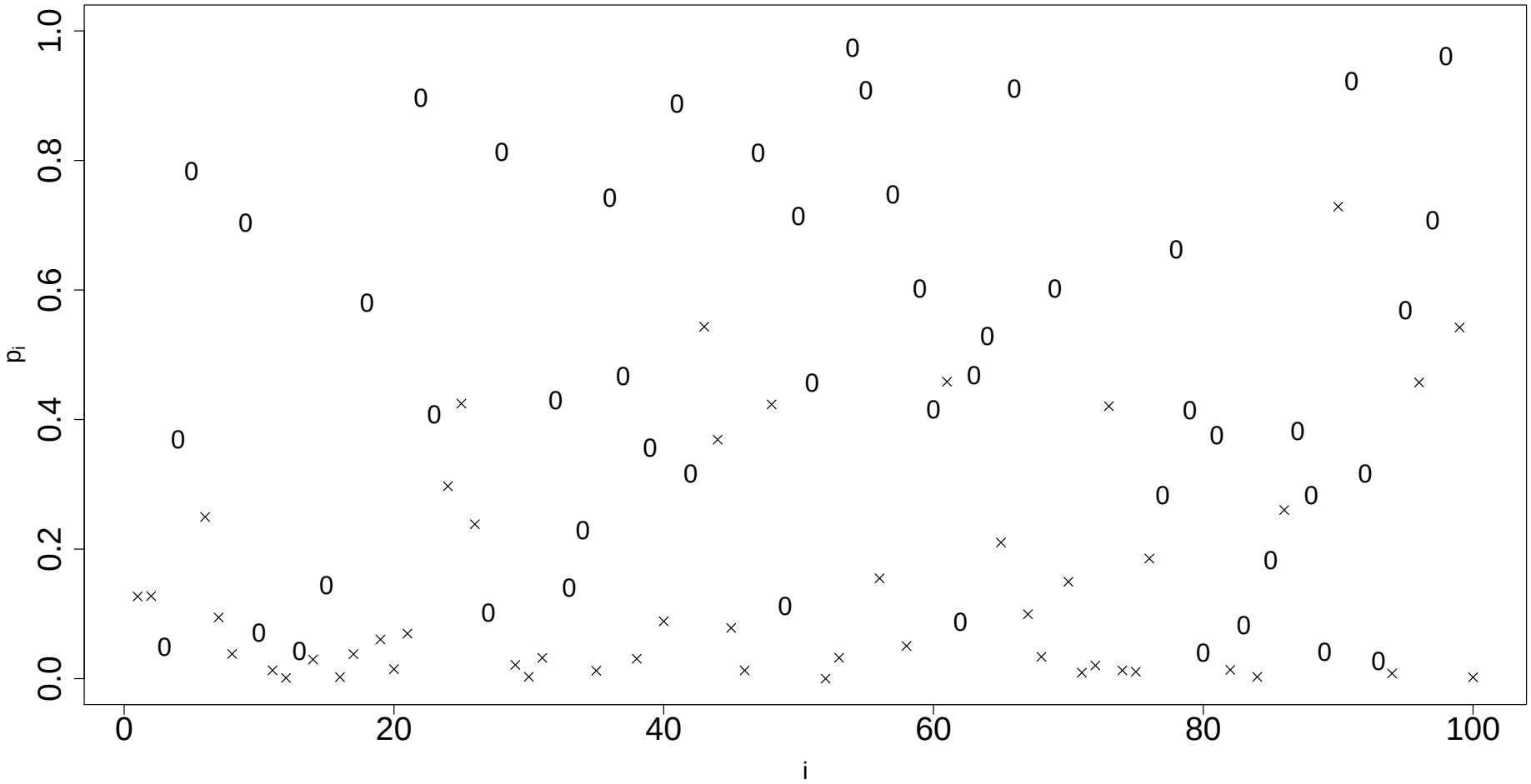
- Rejection set $R = \{1 \leq i \leq m \ : \ p_i(X) \leq t\}$
- Some false positives

Thresholding



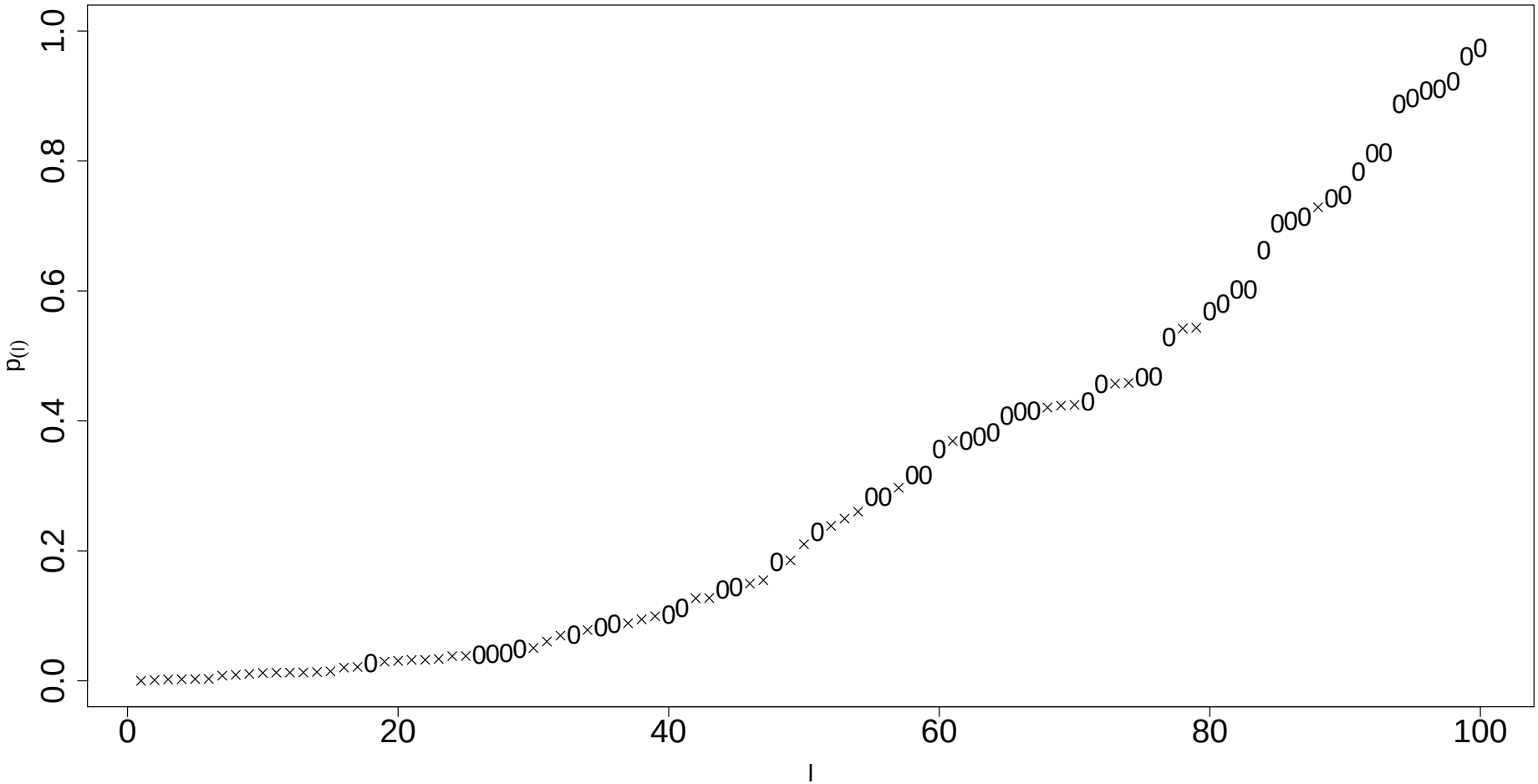
- Rejection set $R = \{1 \leq i \leq m \ : \ p_i(X) \leq t\}$
- Some false positives

Ordering *p*-values



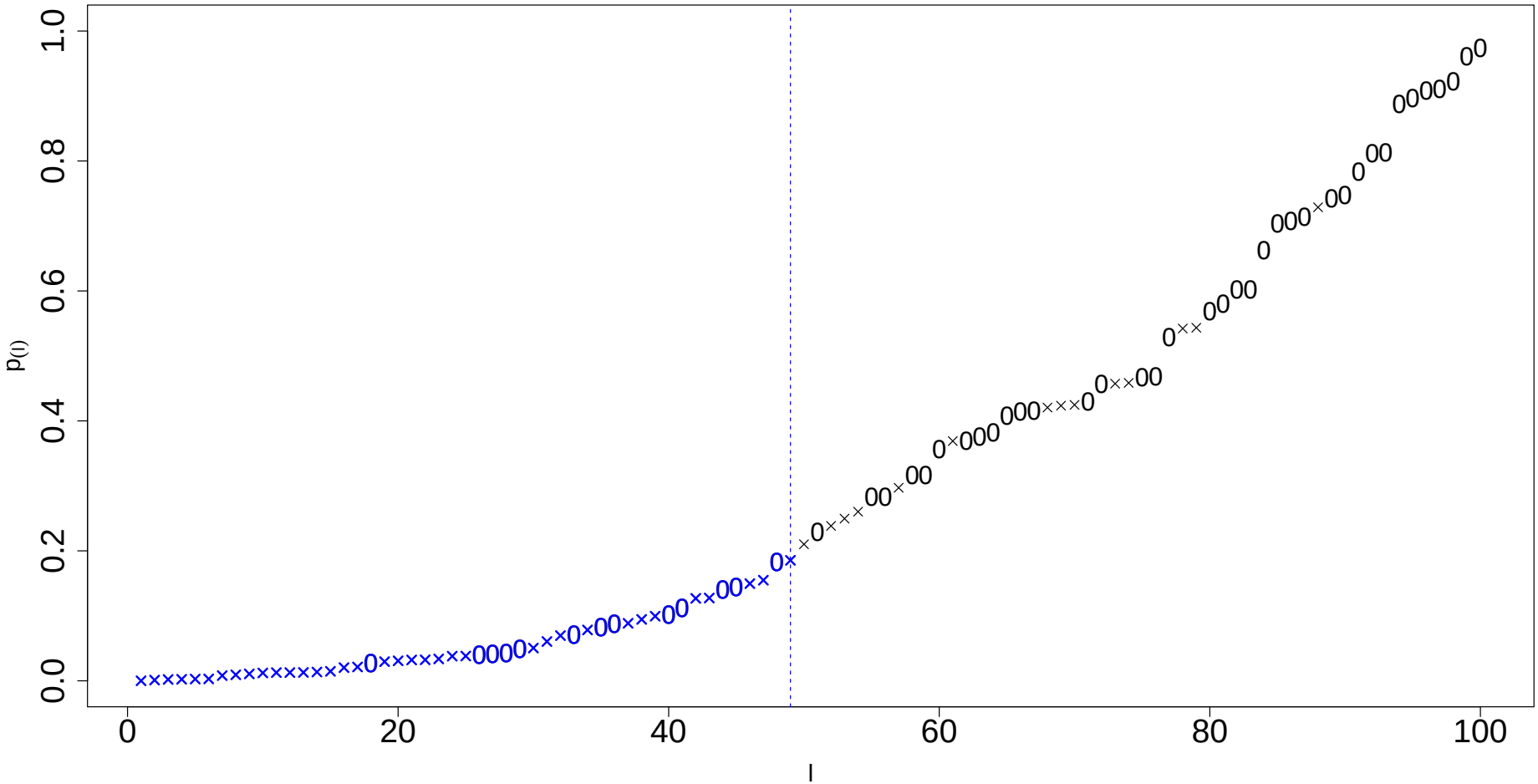
- Stopping rule
- Reject $\widehat{\ell}$ smallest *p*-values

Ordering *p*-values



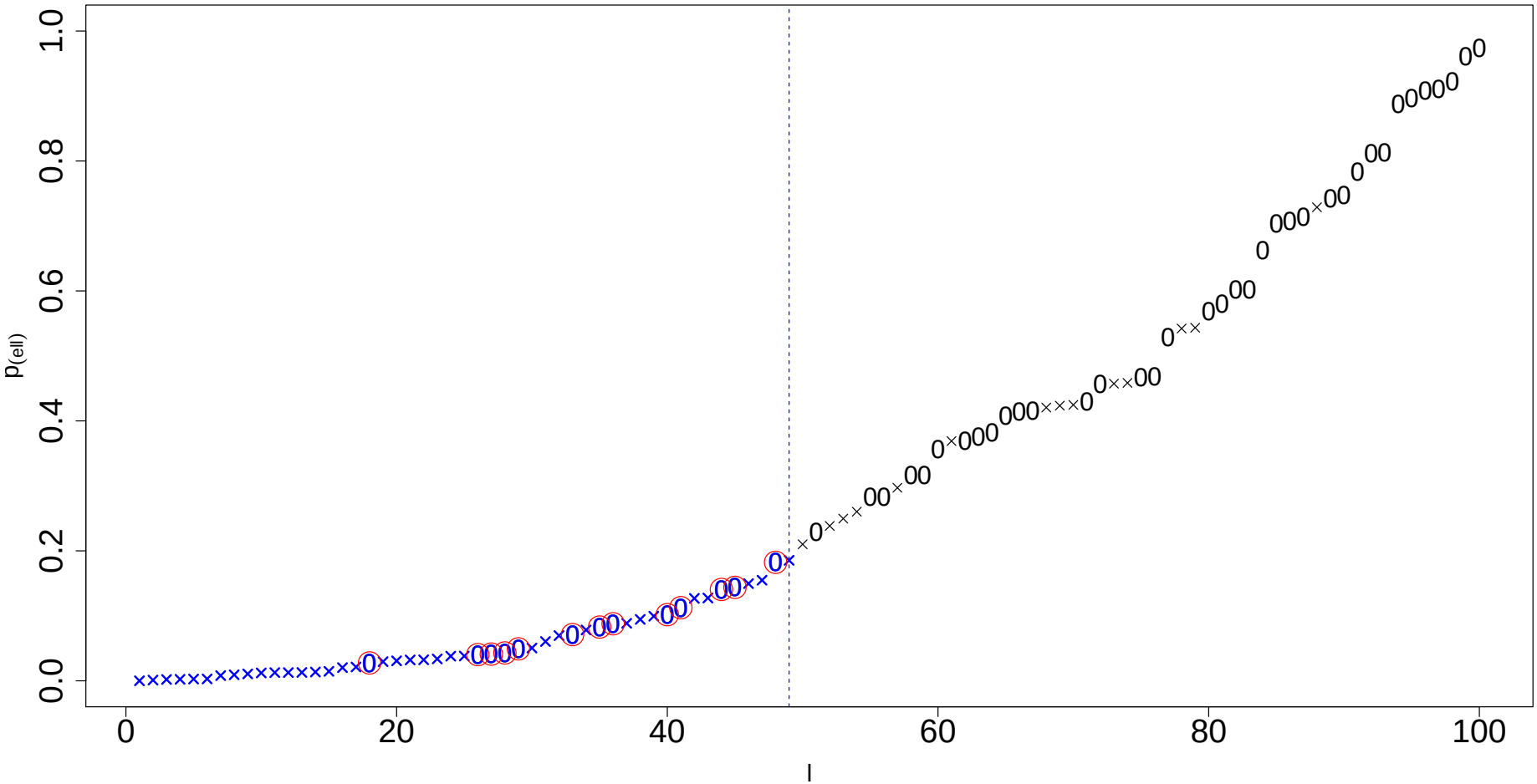
- Stopping rule
- Reject $\widehat{\ell}$ smallest *p*-values

Ordering *p*-values



- Stopping rule
- Reject $\widehat{\ell}$ smallest *p*-values

Ordering *p*-values



- Stopping rule
- Reject $\widehat{\ell}$ smallest *p*-values

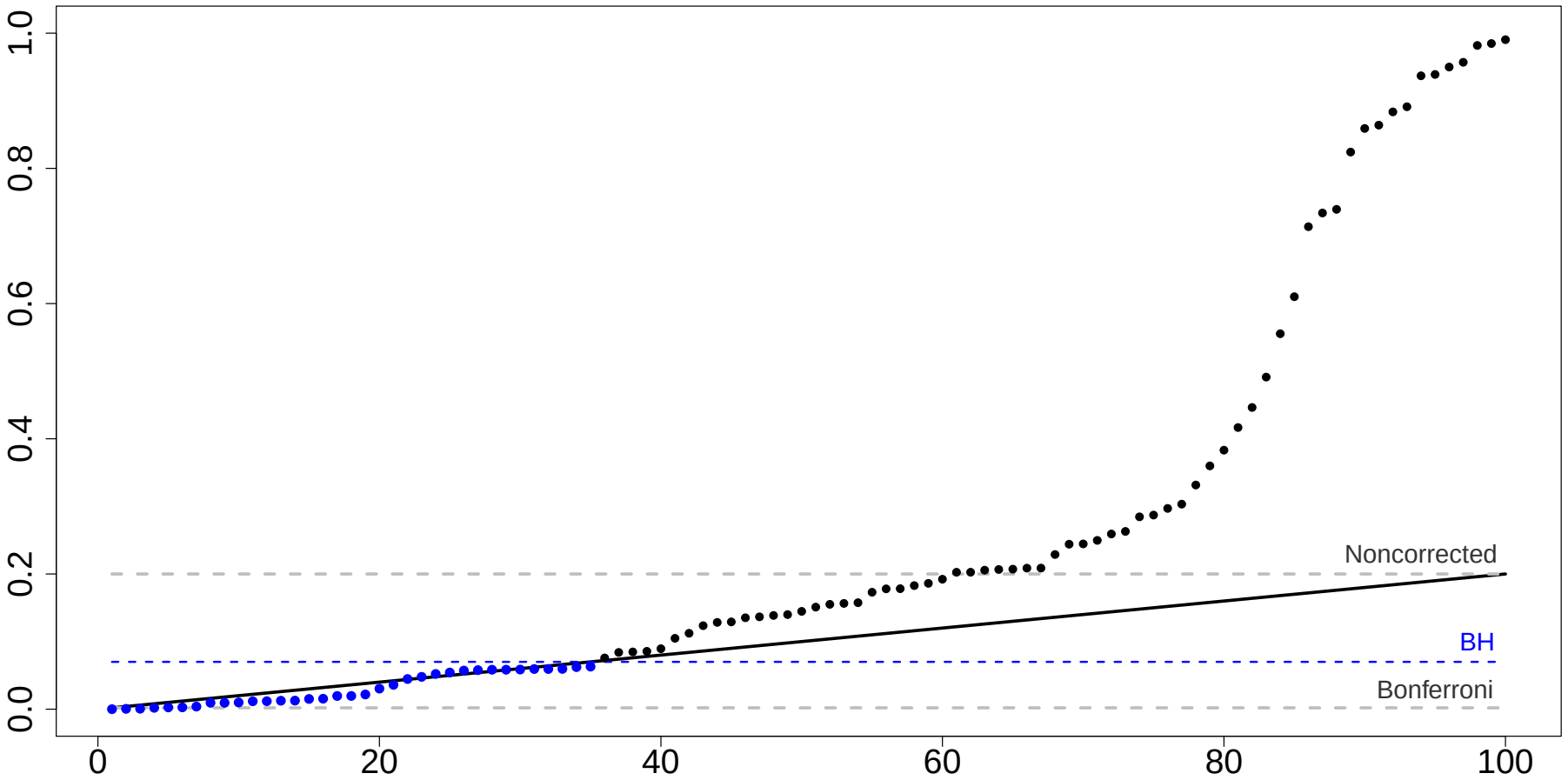
1 Introduction

2 BH procedure

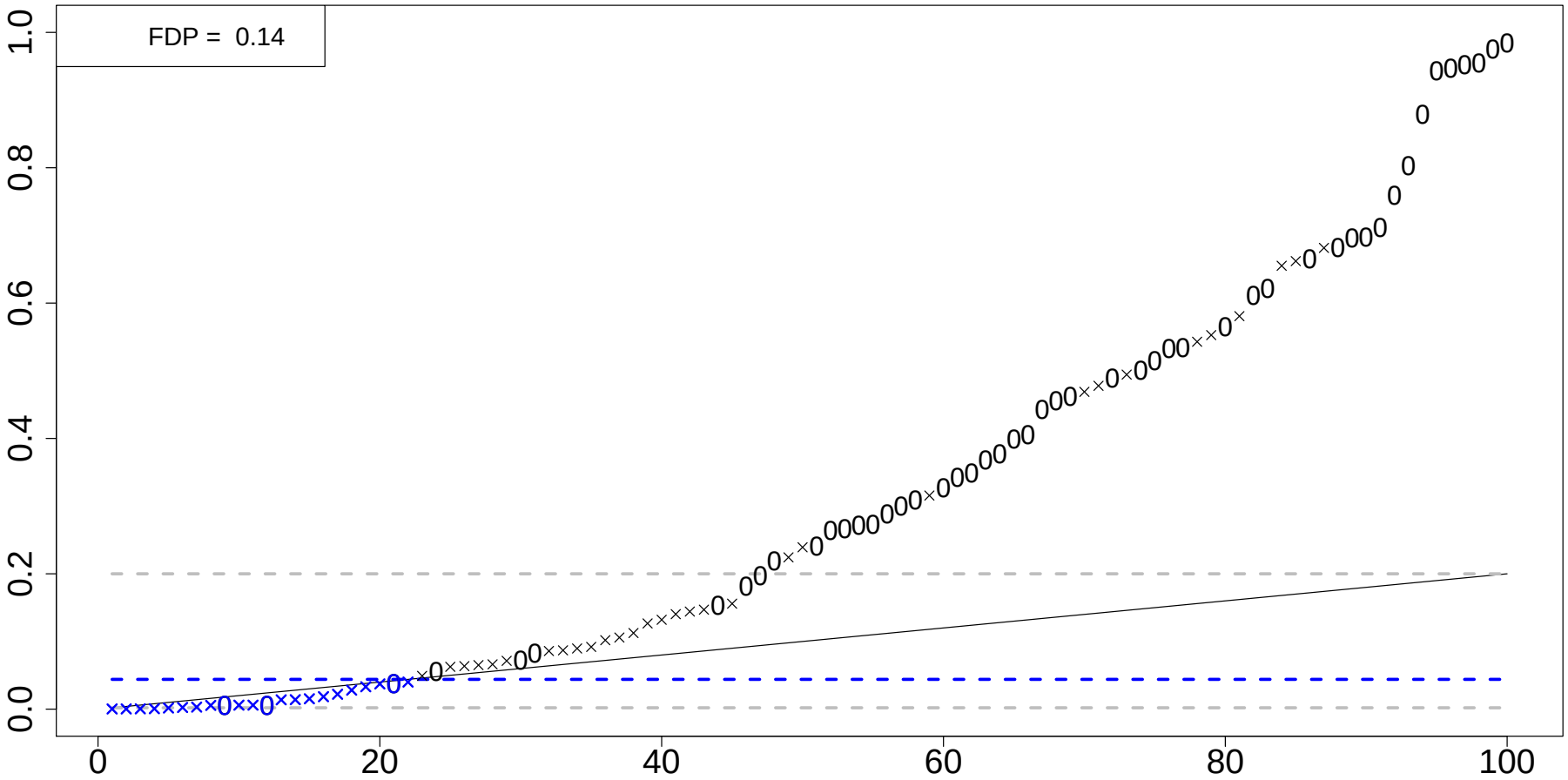
3 Bayesian tricks

BH procedure [Benjamini and Hochberg, 1995]

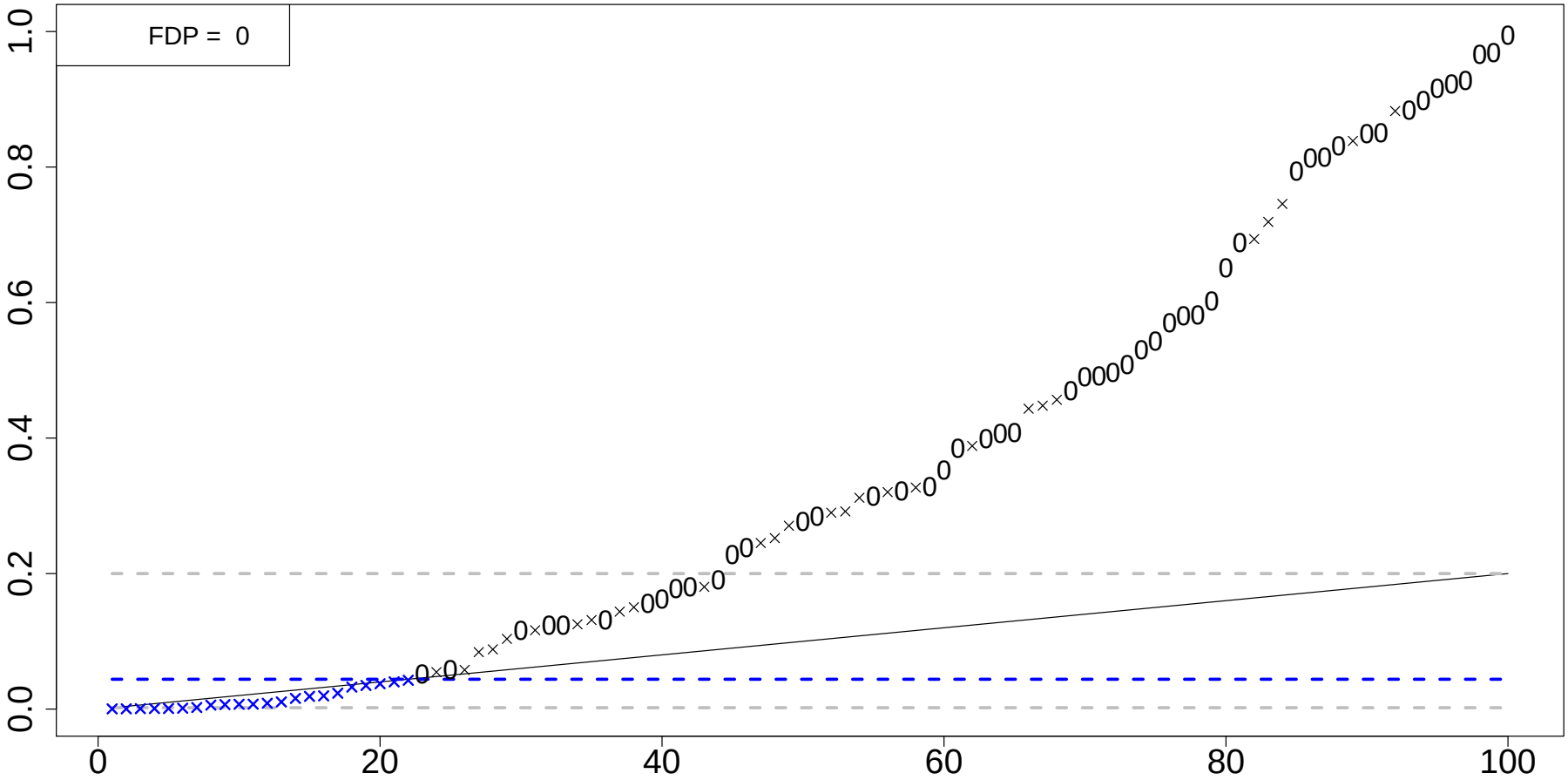
$$\widehat{t} = \alpha(\widehat{k} \vee 1)/m \quad \text{with} \quad \widehat{k} = \max\{0 \leq k \leq m \,:\, p_{(k)} \leq \alpha k/m\}$$



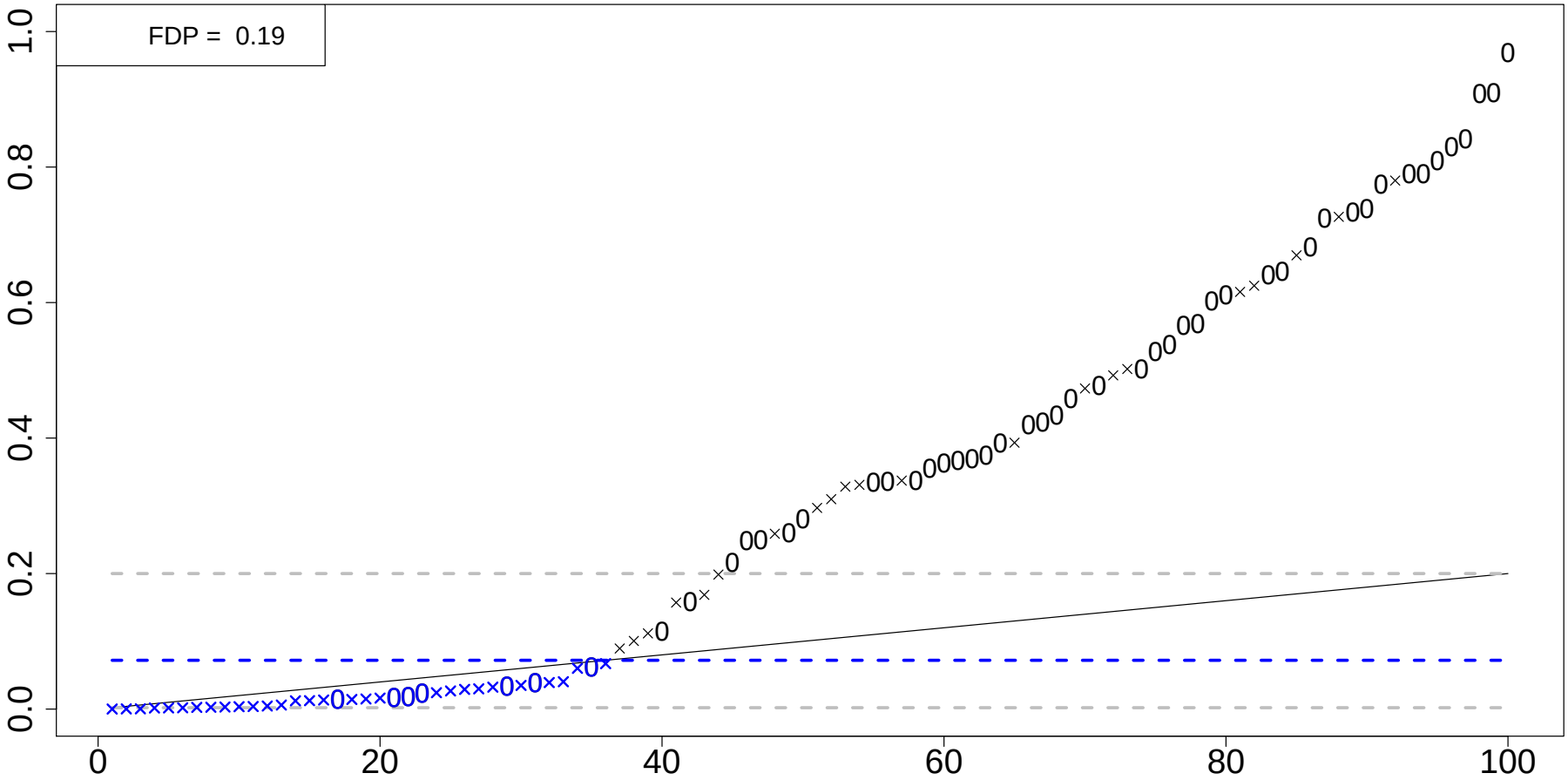
Illustration, $\alpha = 0.2$, $m = 100$



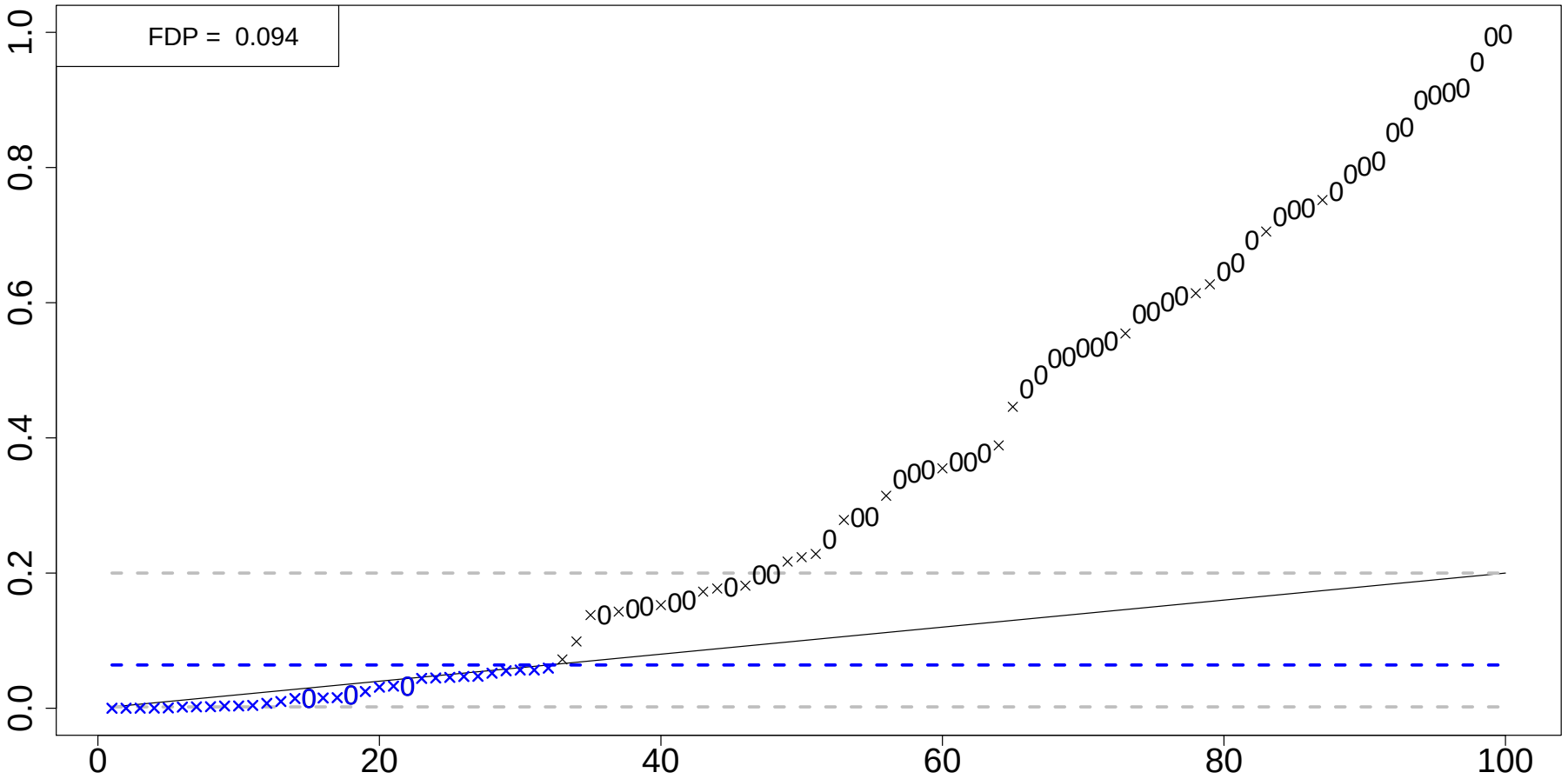
Illustration, $\alpha = 0.2$, $m = 100$



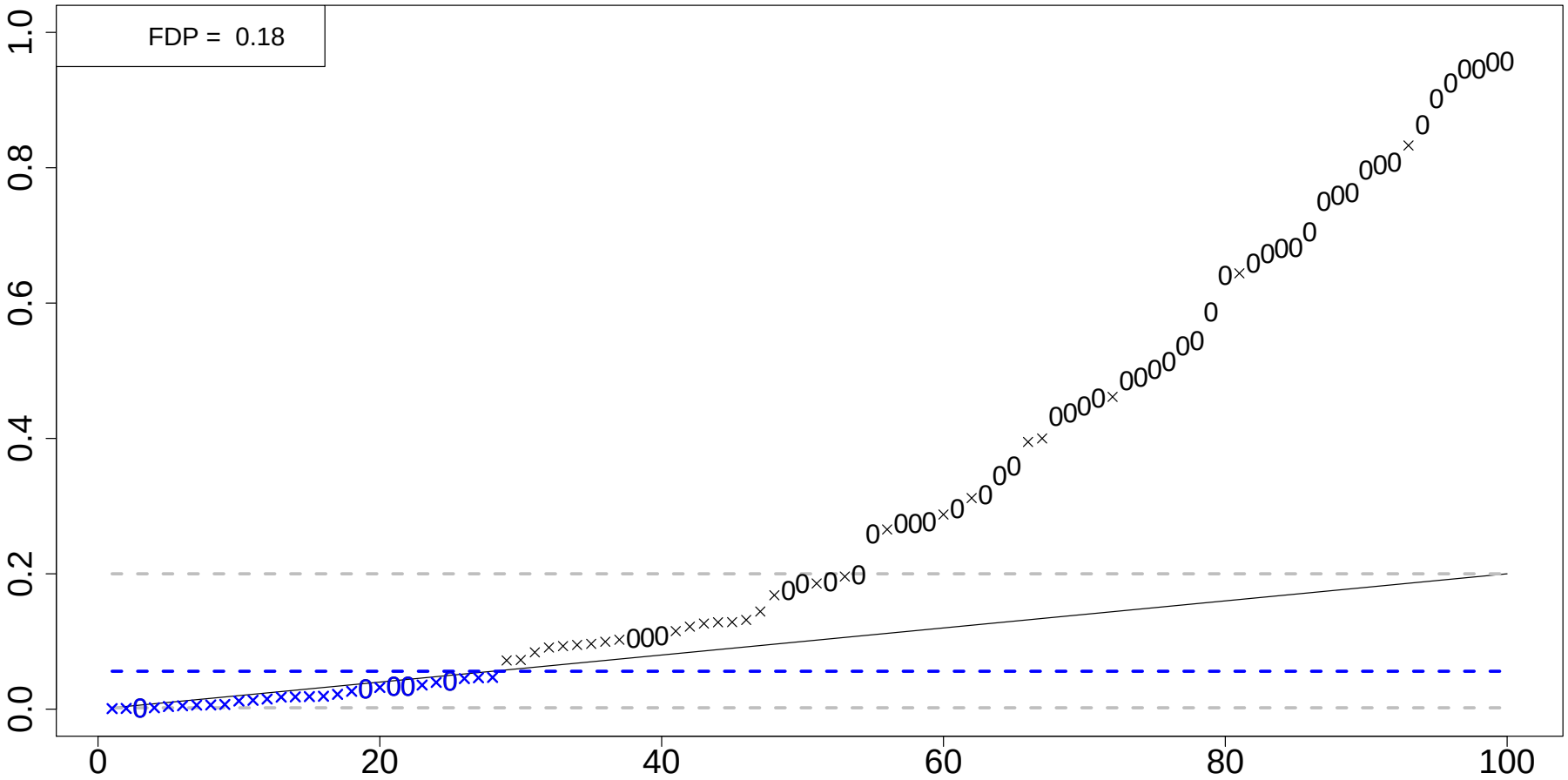
Illustration, $\alpha = 0.2$, $m = 100$



Illustration, $\alpha = 0.2$, $m = 100$



Illustration, $\alpha = 0.2$, $m = 100$



More formally ...

False discovery rate

For a threshold t, $V(t) = \sum_{i=1}^{m}(1-\theta_i)\mathbf{1}\{p_i \leq t\}$, $R(t) = \sum_{i=1}^{m}\mathbf{1}\{p_i \leq t\}$,

$$\text{FDR}(t,P) = \mathbf{E}[\text{FDP}(t,P)], \quad \text{FDP}(t,P) = \frac{V(t)}{R(t)} \qquad \left(\frac{0}{0} = 0\right)$$

here $P = (\theta, \mu)$ parameter.

Difficult point: random denominator

Theorem [Benjamini and Hochberg (1995)]

$$\text{for all } \theta,\ \mu, \quad \text{FDR}(\text{BH}_\alpha, (\theta,\mu)) \leq \alpha$$

Widely used

More formally ...

False discovery rate

For a threshold t, $V(t) = \sum_{i=1}^{m}(1-\theta_i)\mathbf{1}\{p_i \leq t\}$, $R(t) = \sum_{i=1}^{m}\mathbf{1}\{p_i \leq t\}$,

$$\mathrm{FDR}(t,P) = \mathbf{E}[\mathrm{FDP}(t,P)], \quad \mathrm{FDP}(t,P) = \frac{V(t)}{R(t)} \qquad \left(\frac{0}{0} = 0\right)$$

here $P = (\theta, \mu)$ parameter.

Difficult point: random denominator

Theorem [Benjamini and Hochberg (1995)]

$$\text{for all } \theta,\ \mu, \quad \mathrm{FDR}(\mathrm{BH}_\alpha, (\theta,\mu)) \leq \alpha$$

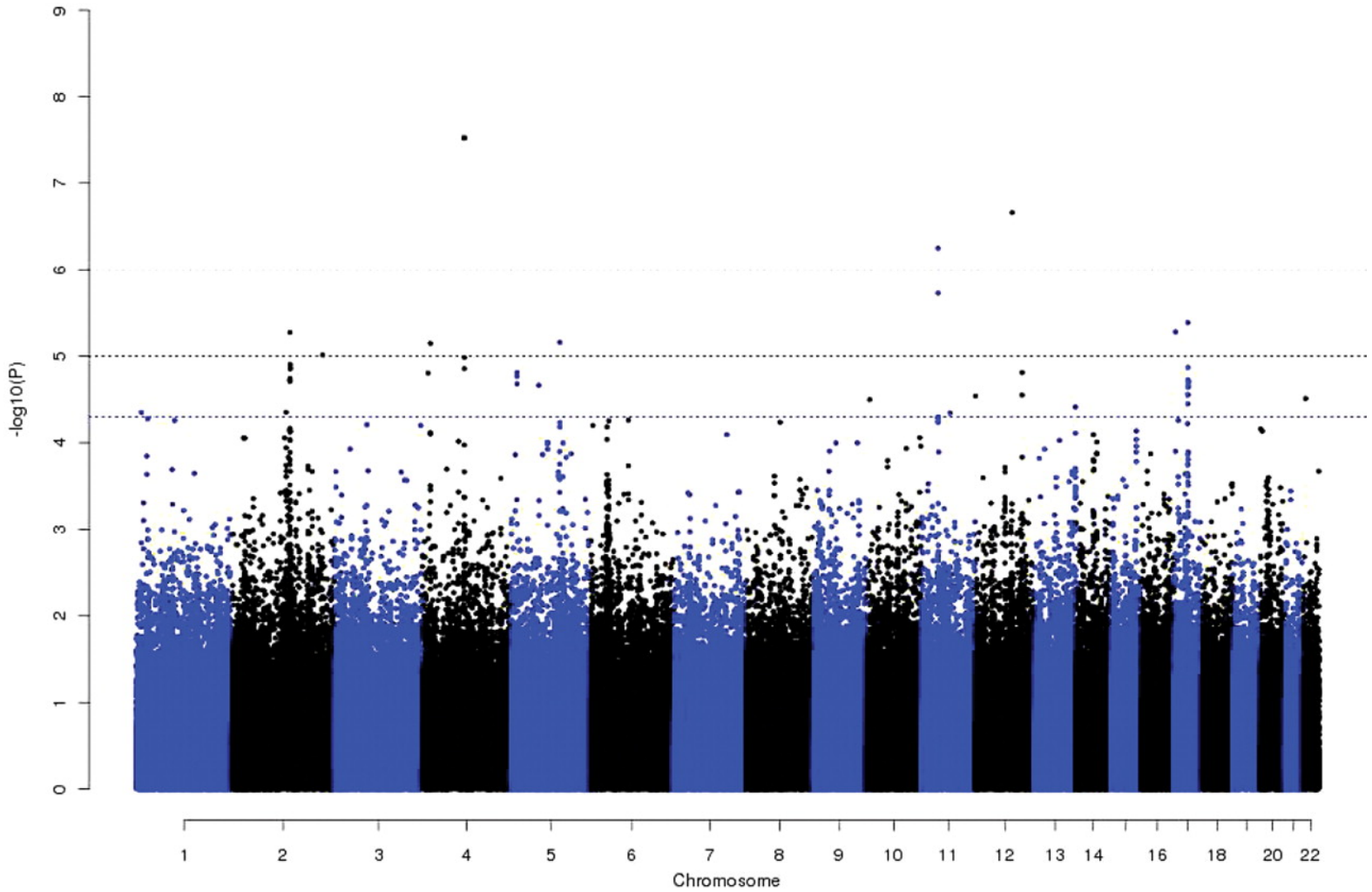
Widely used

1 Introduction

2 BH procedure

3 Bayesian tricks

Manhattan plot (GWAS data)



- Looking for structure on $\theta_1, \ldots, \theta_m$
- Structure = a priori

Bayesian multiple testing literature

[Efron and Tibshirani. *Empirical Bayes methods and false discovery rates for microarrays* (2002, Gen. Epi.)]

- a priori : $\theta_1, \ldots, \theta_m$ i.i.d. (unstructured)
- BH 'optimal' (empirical Bayes method)

[Sun and Cai. *Large-scale multiple testing under dependence* (2009, JRSSB)]

- $\theta_1, \ldots, \theta_m$ with a Markov structure
- New Bayesian multiple testing procedures
- BH suboptimal

(See also [Storey (2003)], [van der Laan et al. (2005)], [Scott and Berger (2005)], [Tang et al. (2007)], [Ghosal and Roy (2012)] . . .)

First trick: optimal classification

Model:

- $\theta = (\theta_1, \ldots, \theta_m) \in \{0,1\}^m$ Markov chain transition matrix A
- $X = (X_1, \ldots, X_m) \mid \theta$ mutually independent,

$$X_i \sim \mathcal{N}(0,1) \mid \theta_i = 0$$
$$X_i \sim \mathcal{N}(\mu,1) \mid \theta_i = 1$$

- p-values

$$p_i(X) = \overline{\Phi}(X_i), \quad 1 \leq i \leq m$$

- ℓ-values

$$\ell_i(X) = \mathbf{P}(\theta_i = 0 \mid X), \quad 1 \leq i \leq m$$

Theorem:

mis-classification risk minimum for the rule

$$\Phi_i(X) = \mathbf{1}\{\ell_i(X) \leq 1/2\}, \quad 1 \leq i \leq m$$

First trick: optimal classification

Model:

- $\theta = (\theta_1, \ldots, \theta_m) \in \{0,1\}^m$ Markov chain transition matrix A
- $X = (X_1, \ldots, X_m) \mid \theta$ mutually independent,

$$X_i \sim \mathcal{N}(0,1) \mid \theta_i = 0$$
$$X_i \sim \mathcal{N}(\mu,1) \mid \theta_i = 1$$

- p-values

$$p_i(X) = \overline{\Phi}(X_i), \quad 1 \leq i \leq m$$

- ℓ-values

$$\ell_i(X) = \mathbf{P}(\theta_i = 0 \mid X), \quad 1 \leq i \leq m$$

Theorem:

mis-classification risk minimum for the rule

$$\Phi_i(X) = \mathbf{1}\{\ell_i(X) \leq 1/2\}, \quad 1 \leq i \leq m$$

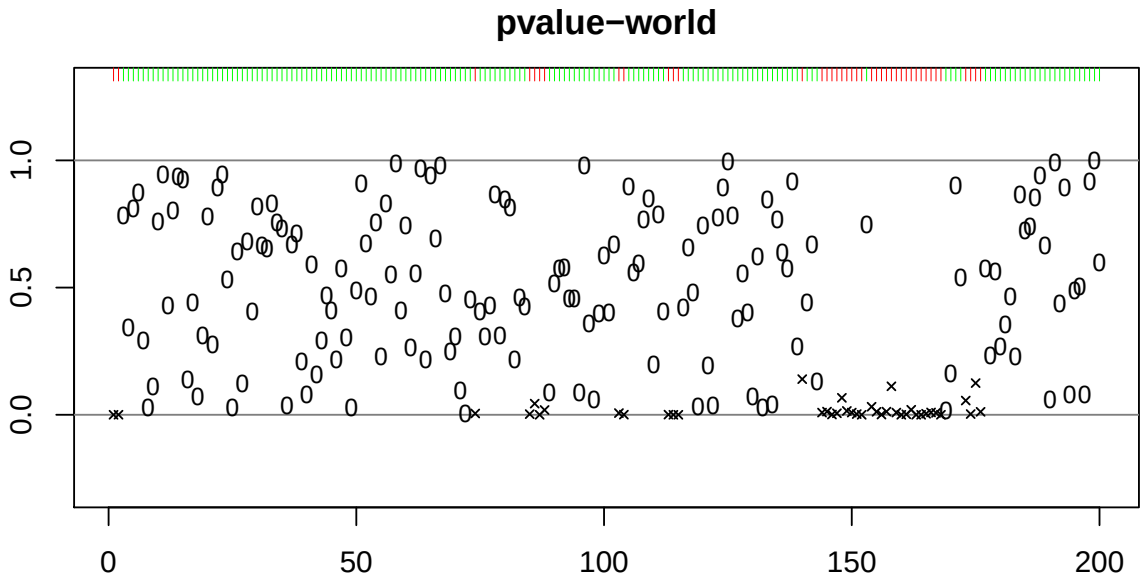
First trick: optimal classification

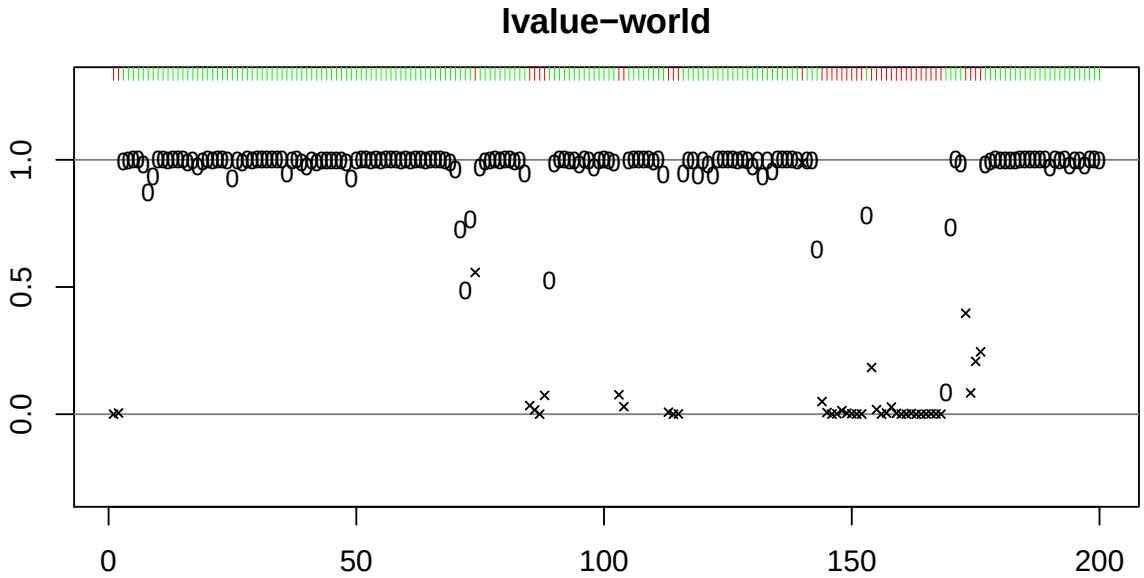
- $m = 200$
- Transition matrix

$$A = \begin{pmatrix} 0.95 & 0.05 \\ 0.25 & 0.75 \end{pmatrix}$$

- Gaussian

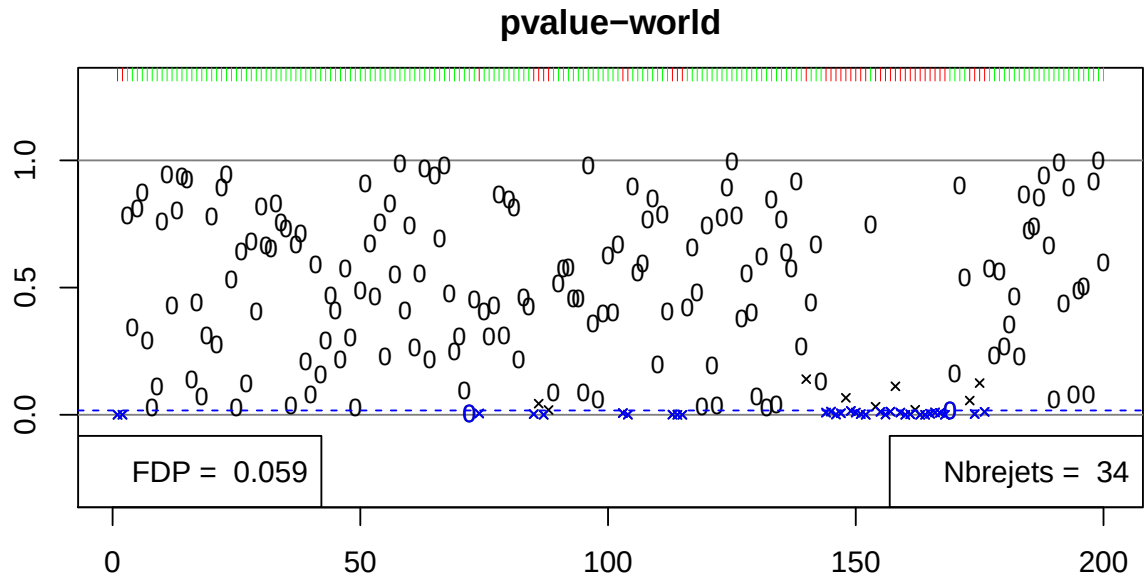
$\mu = 2.5$
$\sigma = 1$

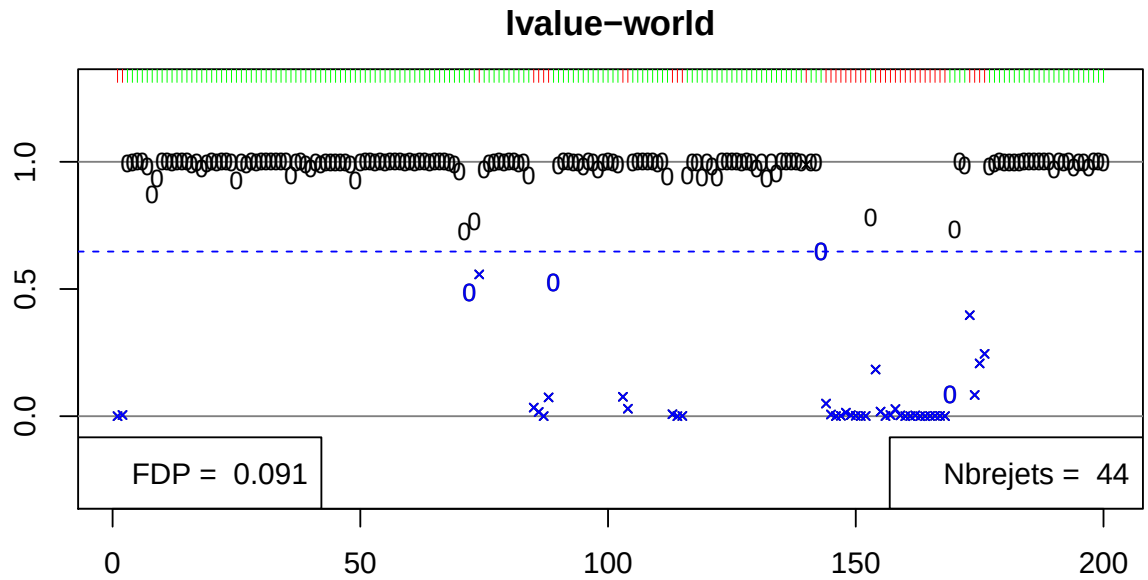




Second trick: Bayesian FDR

- $\alpha = 0.1$
- BH procedure
- Bayesian procedure





Second trick: Bayesian FDR

Bproc : reject $\widehat{k}$ smallest ℓ-values for

$$\widehat{k} = \max \left\{ k \ : \ \frac{1}{k} \sum_{k'=1}^{k} \ell_{(k')} \leq \alpha \right\}.$$

Theorem:

$$\text{for all } \mu, \quad \int_{\theta} \text{FDR}(\text{Bproc}, (\theta, \mu)) d\nu(\theta) \leq \alpha$$

Proof : "one line"

Discussions

Limitations

- FDR control is only valid under the prior
- ℓ-values to be computed
- procedure depends on hyper-parameters

If a priori assumed 'true' :

- Hyper-parameters have to be estimated (empirical Bayes)
- Plug-in step needs additional theory

If a priori just a tool:

frequentist FDR control of Bayesian multiple testing procedures?

$$\text{for all } \theta \in \Theta, \mu, \quad \text{FDR}(\text{Bproc}, (\theta, \mu)) \approx \alpha \quad \text{under sparsity}$$

Discussions

Limitations

- FDR control is only valid under the prior
- ℓ-values to be computed
- procedure depends on hyper-parameters

If a priori assumed 'true' :

- Hyper-parameters have to be estimated (empirical Bayes)
- Plug-in step needs additional theory

If a priori just a tool:

frequentist FDR control of Bayesian multiple testing procedures?

$$\text{for all } \theta \in \Theta,\ \mu, \quad \text{FDR}(\text{Bproc}, (\theta, \mu)) \approx \alpha \quad \text{under sparsity}$$

Discussions

Limitations

- FDR control is only valid under the prior
- ℓ-values to be computed
- procedure depends on hyper-parameters

If a priori assumed 'true' :

- Hyper-parameters have to be estimated (empirical Bayes)
- Plug-in step needs additional theory

If a priori just a tool:

frequentist FDR control of Bayesian multiple testing procedures?

for all $\theta \in \Theta$, μ, FDR(Bproc, (θ, μ)) $\approx \alpha$ under sparsity