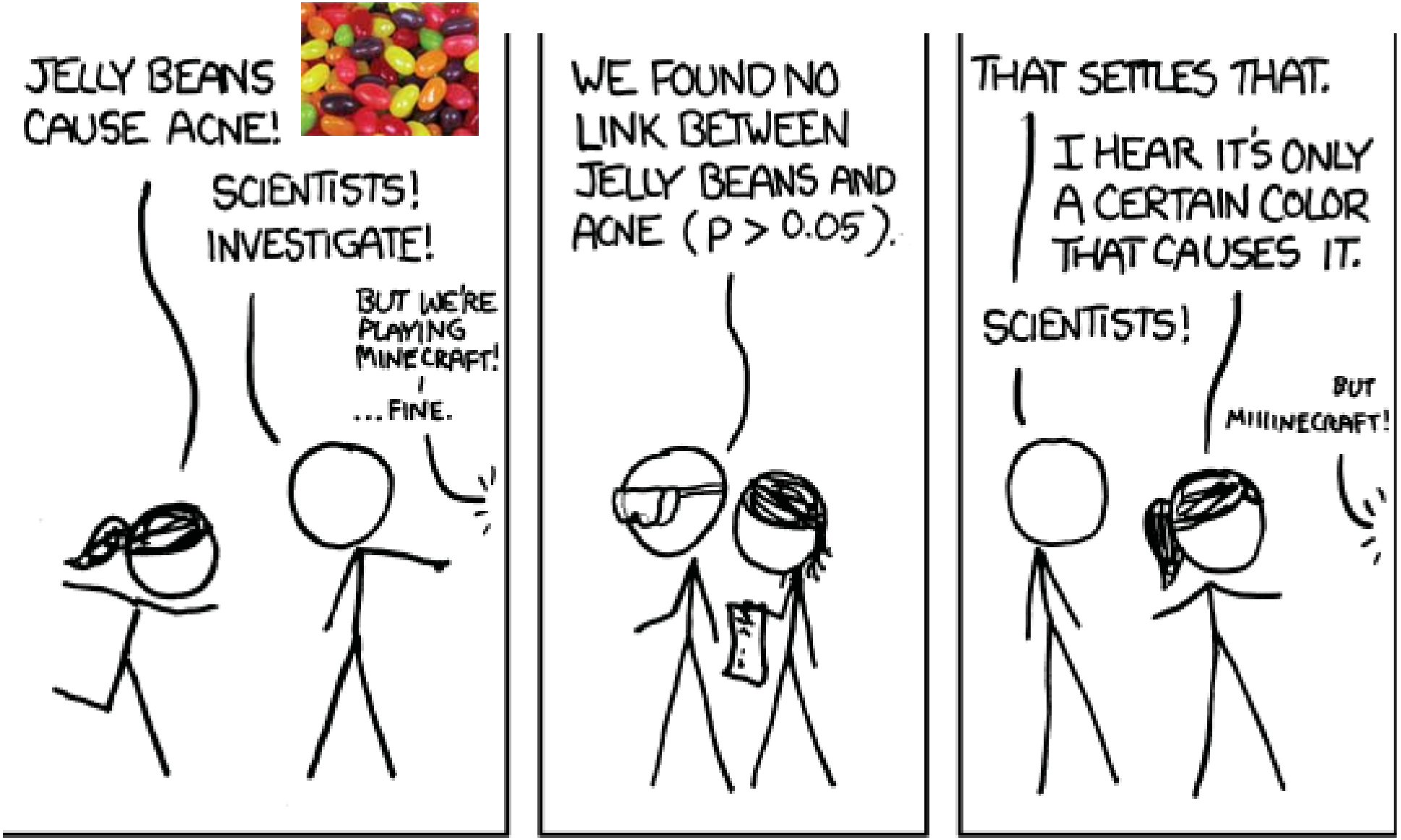
False Discovery Rate : enjeux et nouveaux défis

E. Roquain[1]
Joint work with S. Delattre[2]

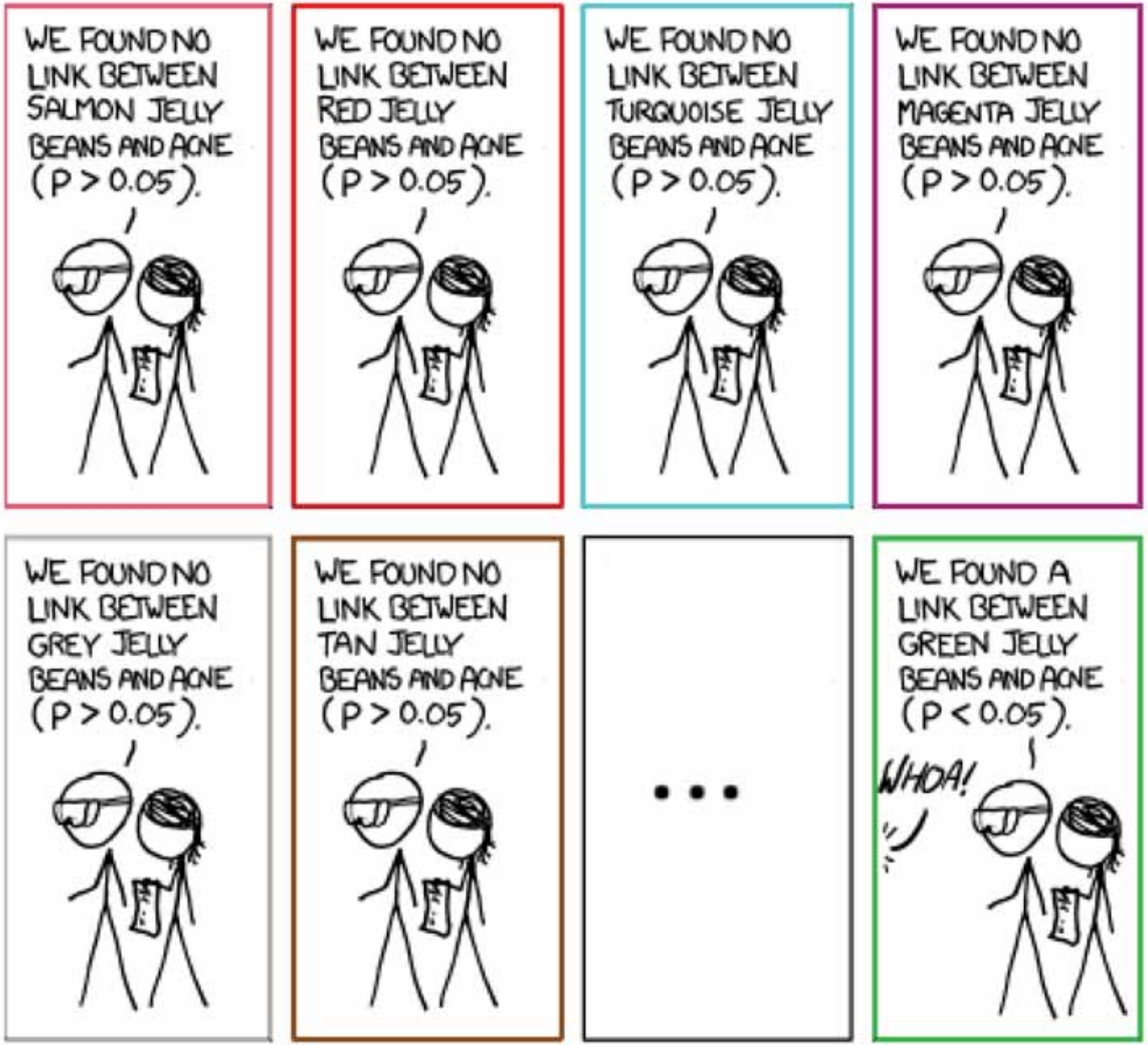
[1]Laboratory LPMA, Université Pierre et Marie Curie (Paris 6), France
[2]Laboratory LPMA, Université Paris Diderot (Paris 7), France

Journées MAS - SMAI, 27 Août 2014

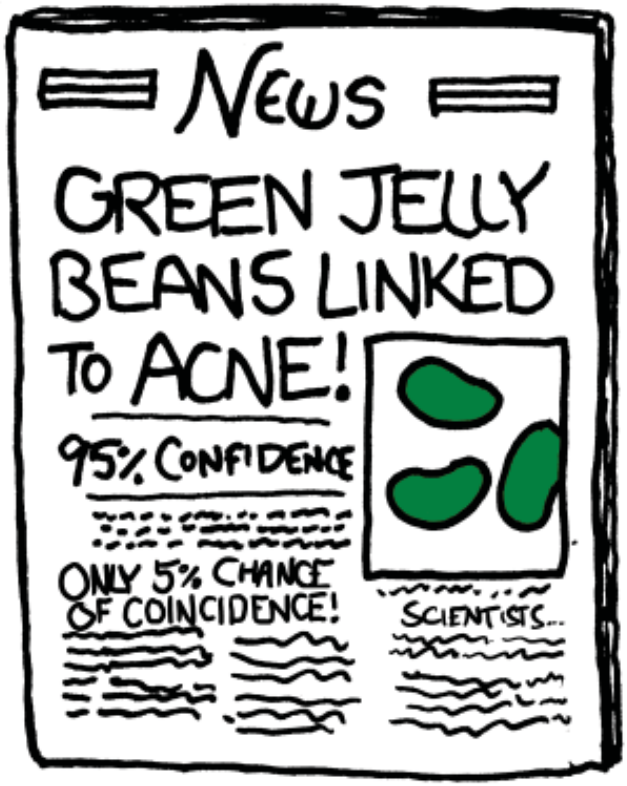
A “multiple testing joke" (http://xkcd.com)



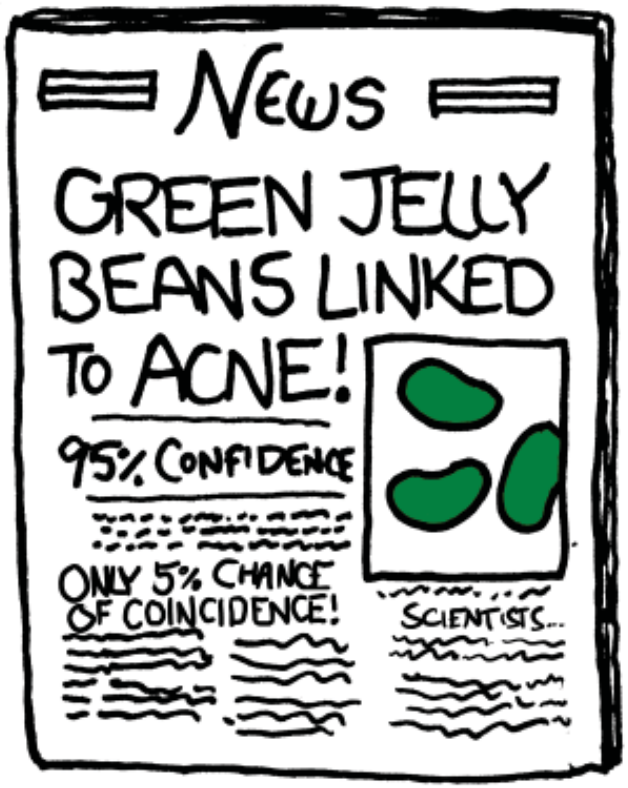
A “multiple testing joke" (http://xkcd.com)



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



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



Multiplicity problem

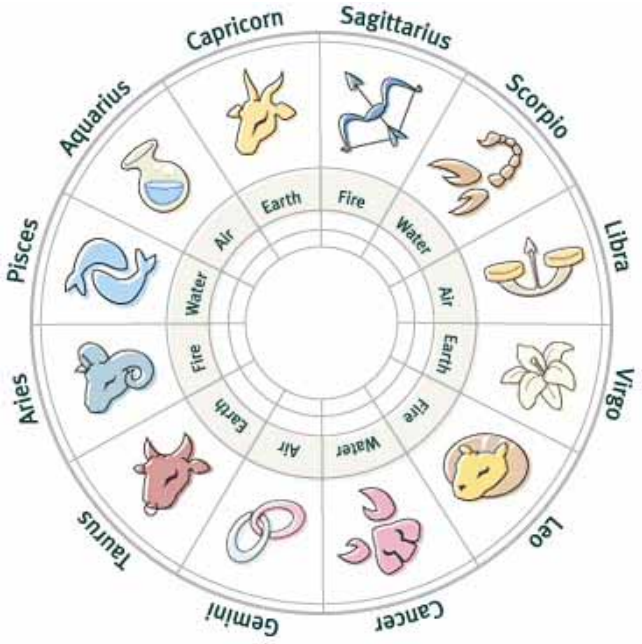
P(make at least one false discovery) $\gg$ **P**(the i-th is a false discovery)

A correction is needed to assess significancy!

Some other examples

Paradoxes due to large scale experiments

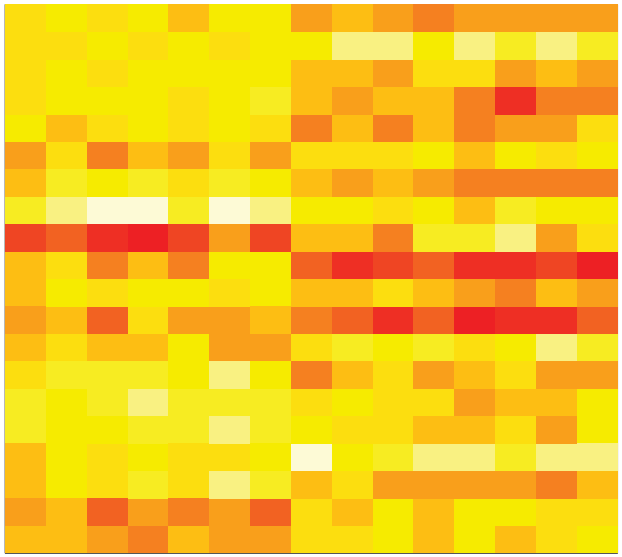
Probable facts appear significant



Multiplicity in microarray [Hedenfalk et al. (2001)]

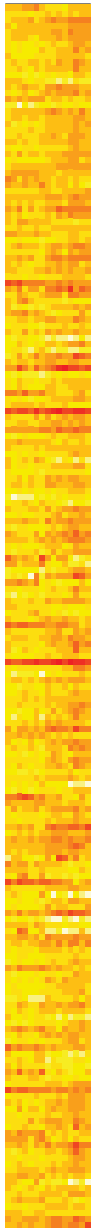
BRCA1 vs BRCA2

genes

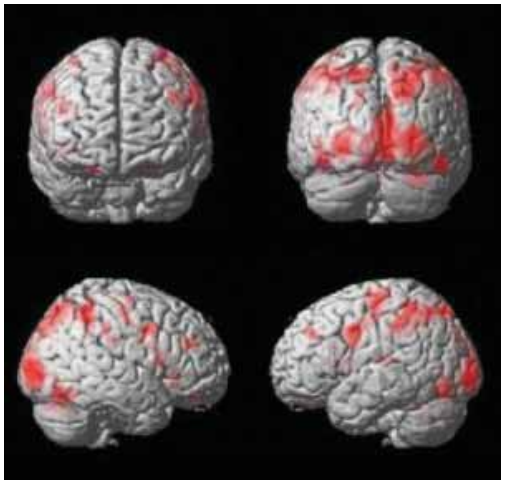


- expression level (activity)
- genes differentially activated?
- 1 test for each gene
- thousands of genes

- nb replications $\ll$ dimension
- correlations



Other applications



- Neuroimaging (FMRI)
  activated regions?
- Econometrics
  winning strategies?
- Astronomy
  directions with stars?

Canonical setting

- $X_i =$ avg group 2 - avg group 1 (rescaled) for genes i
- Gaussian model :

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} = \mu \begin{pmatrix} H_1 \\ H_2 \\ \vdots \\ H_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix},$$

  with $\mu > 0$, $H \in \{0,1\}^m$ (fixed) and $\varepsilon \sim \mathcal{N}(0, \Gamma)$ $(\Gamma_{i,i} = 1)$.

- Γ = dependence structure = I_m for now

Question: for each i, $H_i = 0$ or $H_i = 1$?

Multiple testing : favors the “0" decision

Individual decision and errors

- Test statistic: X_i
- p-value: $p_i = \overline{\Phi}(X_i)$, with $\overline{\Phi}(z) = \mathbf{P}(Z \geq z)$, $Z \sim \mathcal{N}(0,1)$

p_i such that

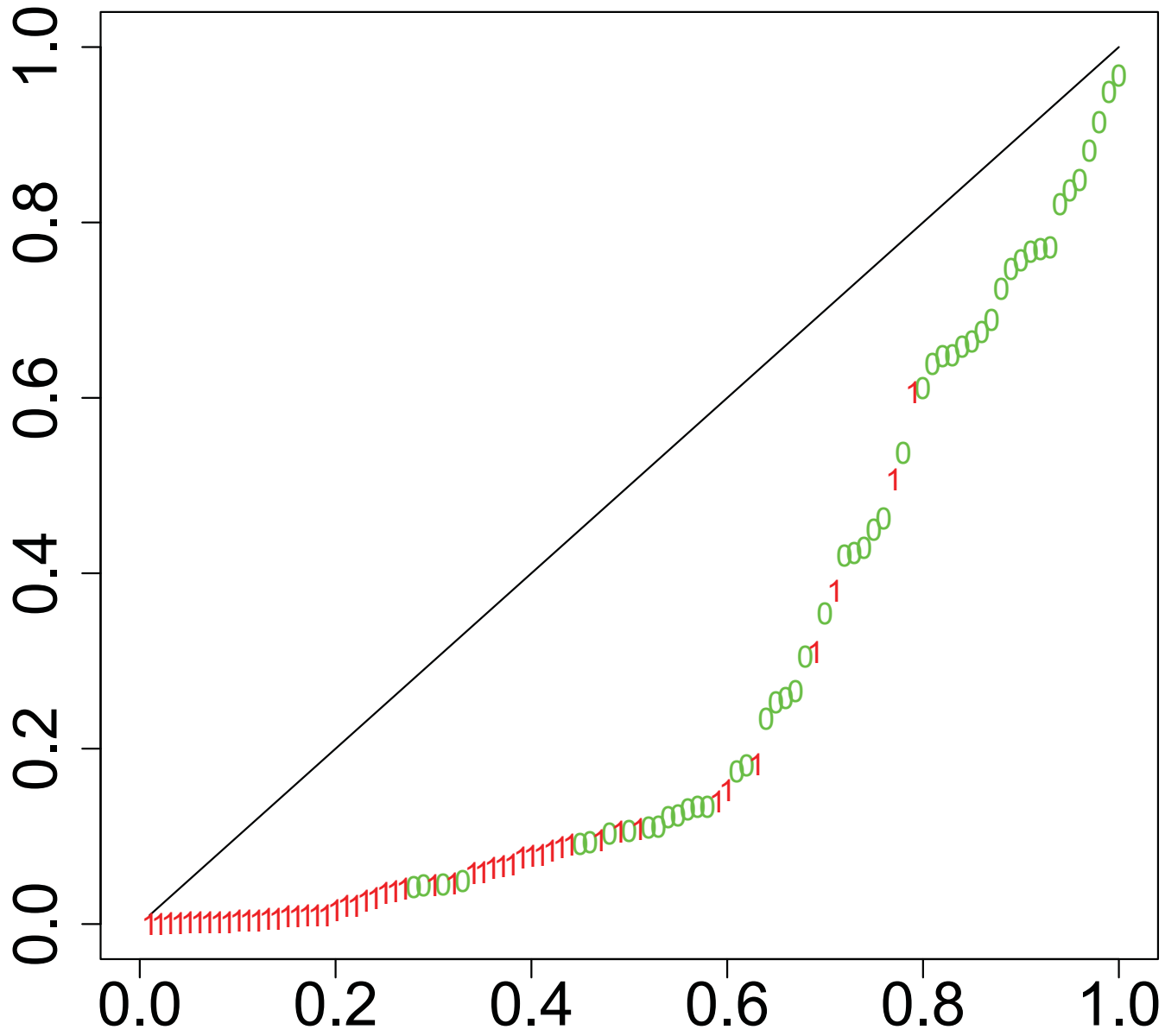
$$\text{if } H_i = 0, \quad p_i \sim U(0,1)$$
$$\text{if } H_i = 1, \quad p_i \sim \overline{\Phi}(\overline{\Phi}^{-1}(\cdot) - \mu)$$

- Choose $\widehat{H}_i = \mathbf{1}\{p_i \leq t\}$ for some threshold t
- Two errors:

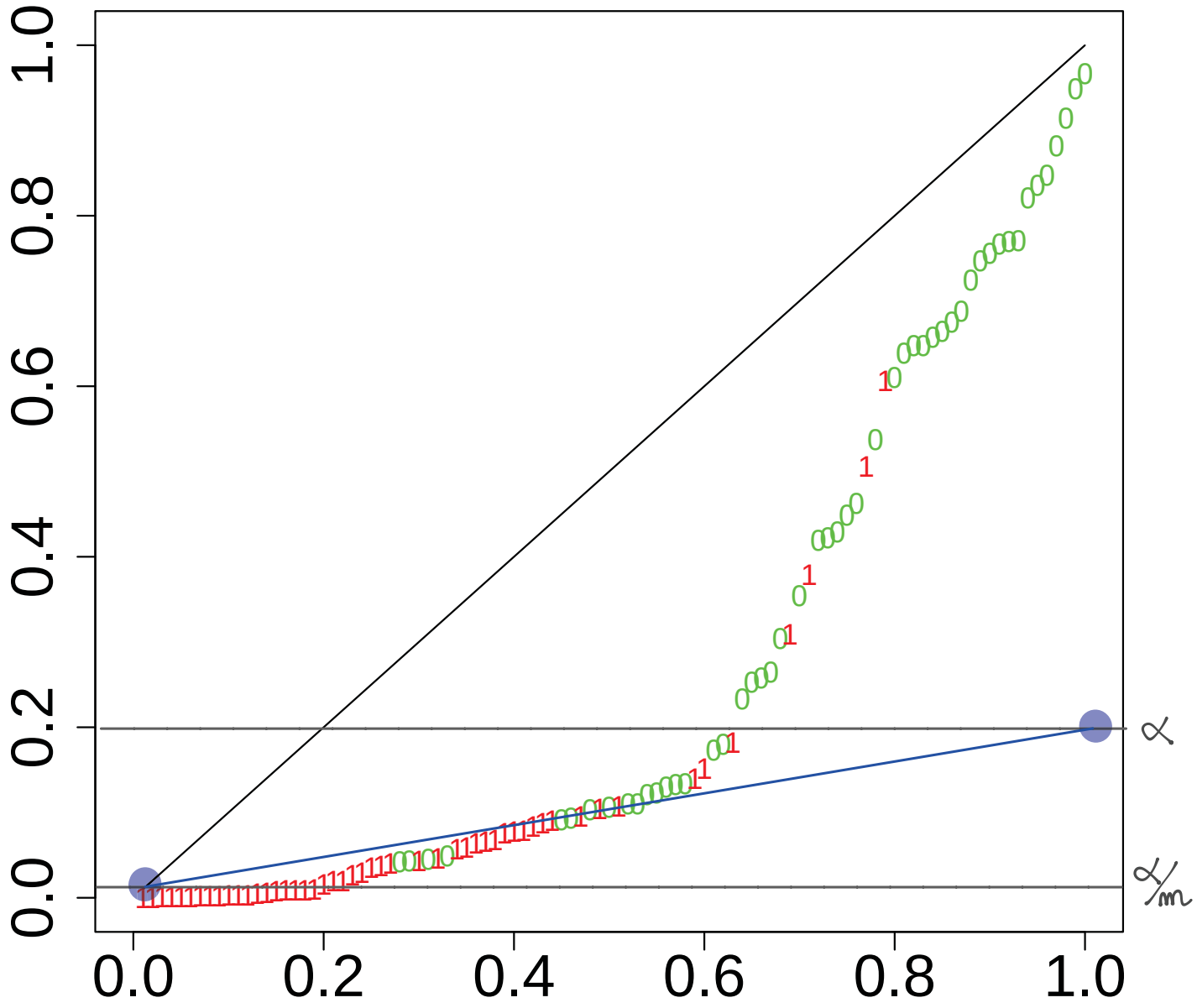
	$\widehat{H}_i = 0$	$\widehat{H}_i = 1$
$H_i = 0$	true negative	false positive
$H_i = 1$	false negative	true positive

- False positive more annoying

Data: $m = 100$; $m_0 = 50$; $\mu = 2$; $\Gamma = I_m$



BH thresholding



False discovery rate control

For a decision $\widehat{H}_i = \mathbf{1}\{p_i \leq \widehat{t}\}$ $(\forall i)$,

$$\mathrm{FDP}(\widehat{t}) = \frac{\#\{i \ : \ H_i = 0, \ \widehat{H}_i = 1\}}{\#\{i \ : \widehat{H}_i = 1\}} \qquad \left(\frac{0}{0} = 0\right)$$

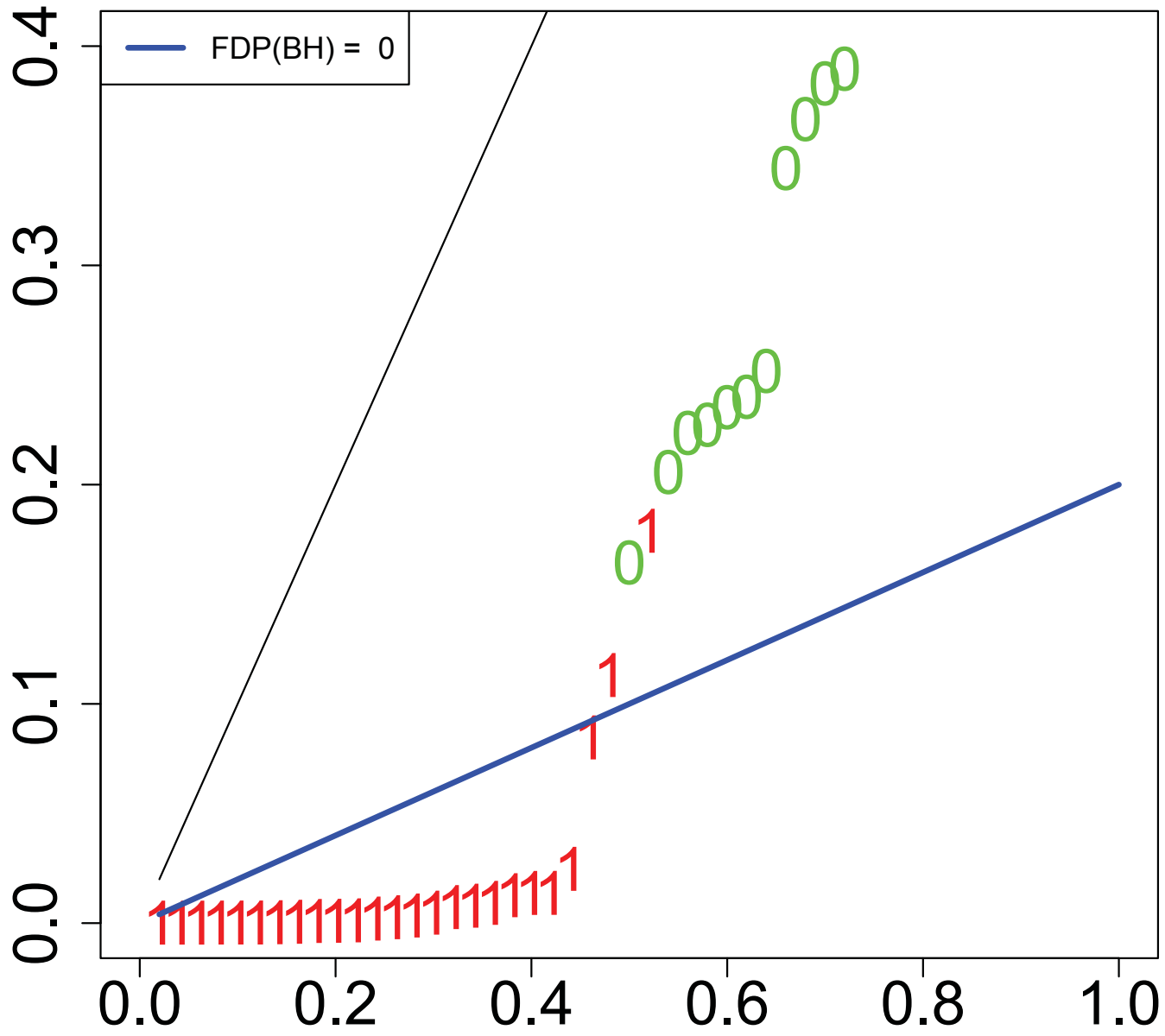
$$\mathrm{FDR}(\widehat{t}) = \mathbf{E}[\mathrm{FDP}(\widehat{t})]$$

Theorem [Benjamini and Hochberg (1995)] [Benjamini and Yekutieli (2001)]

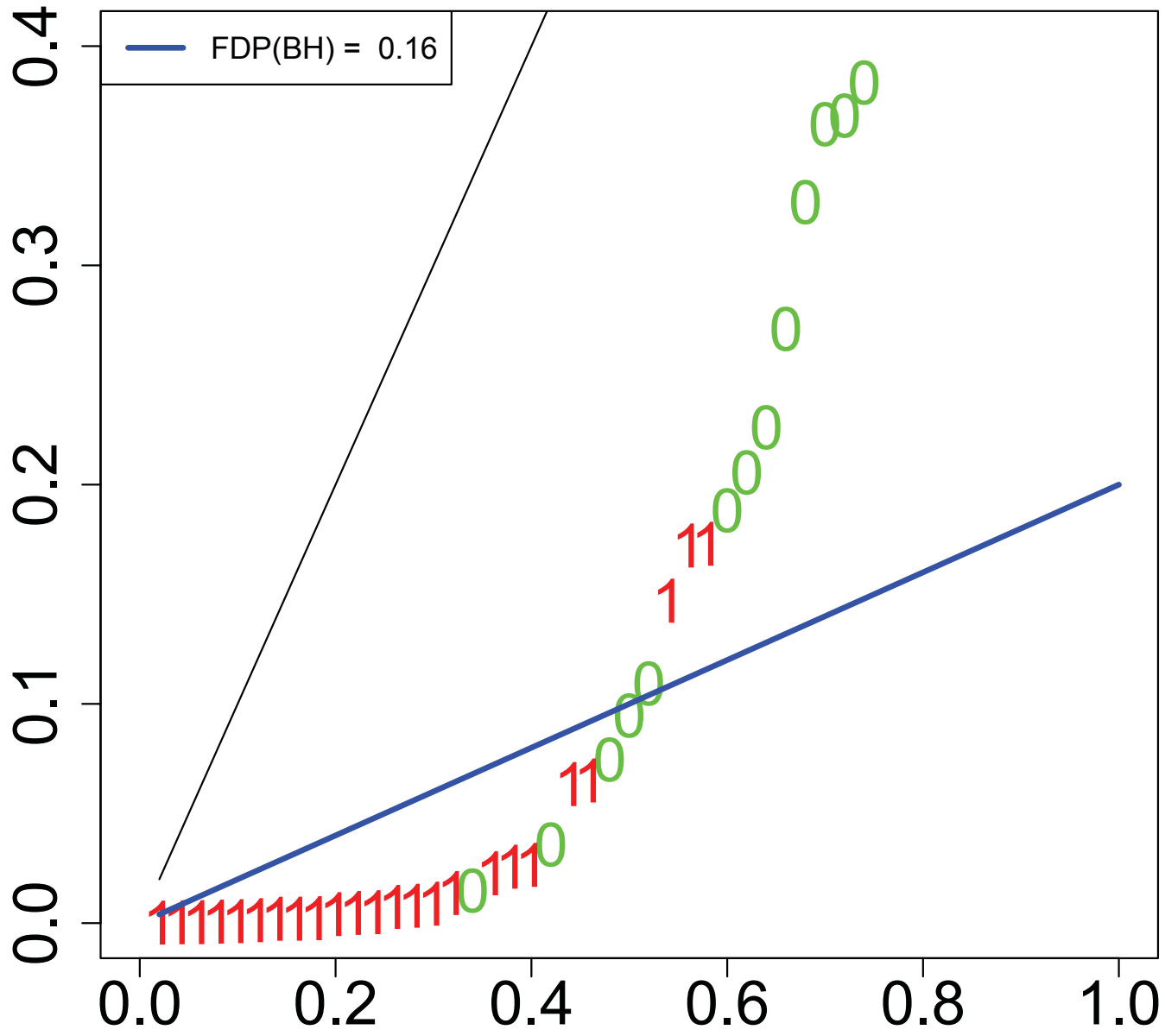
If $\Gamma = I_m$ and $\widehat{t}$ threshold of BH procedure, $\forall \mu, H$,

$$\mathrm{FDR}(\widehat{t}) = (m_0/m)\alpha \leq \alpha$$

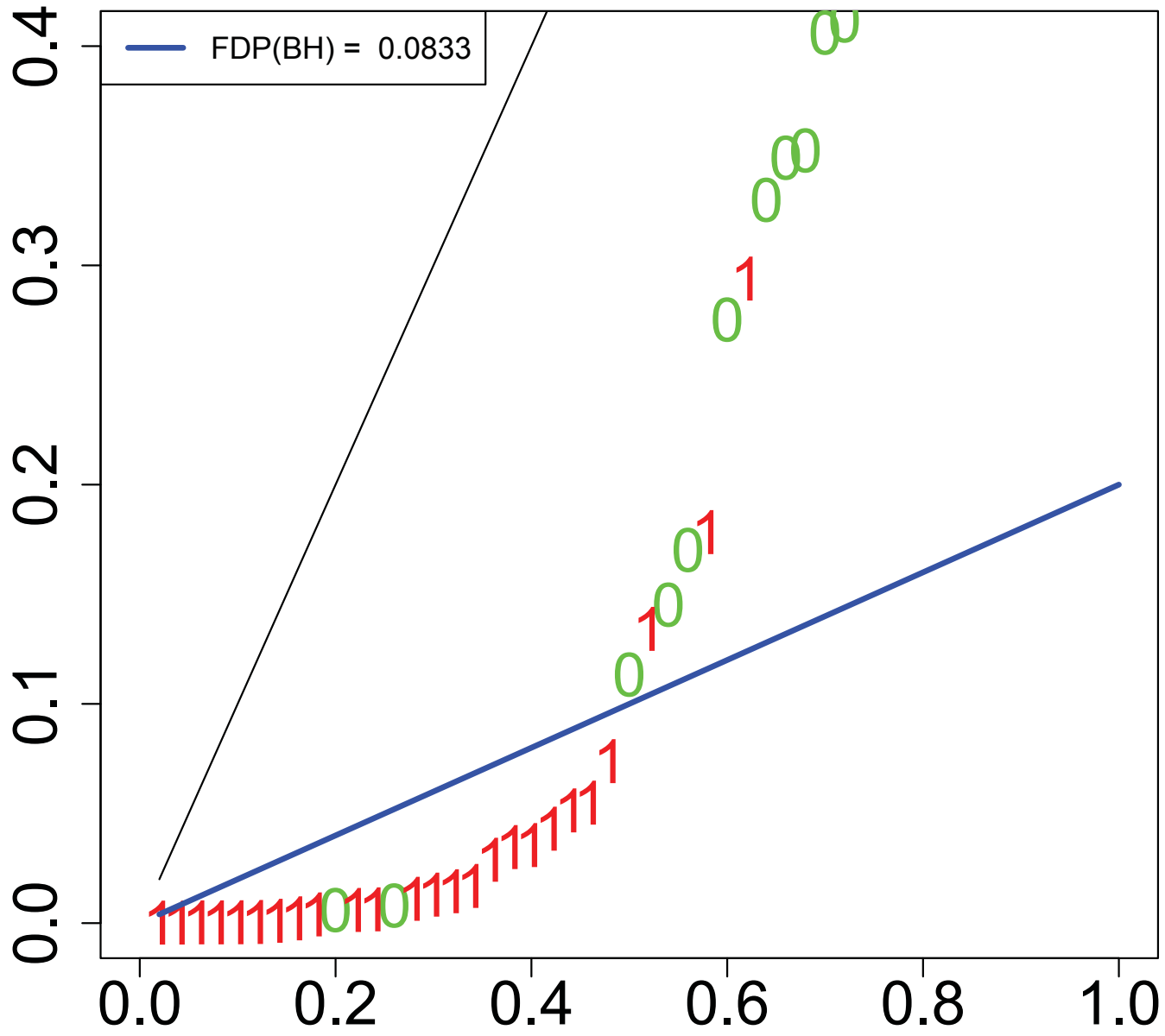
Simulations. $m = 50$; $m_0 = 25$; $\mu = 3$; $\Gamma = I_m$



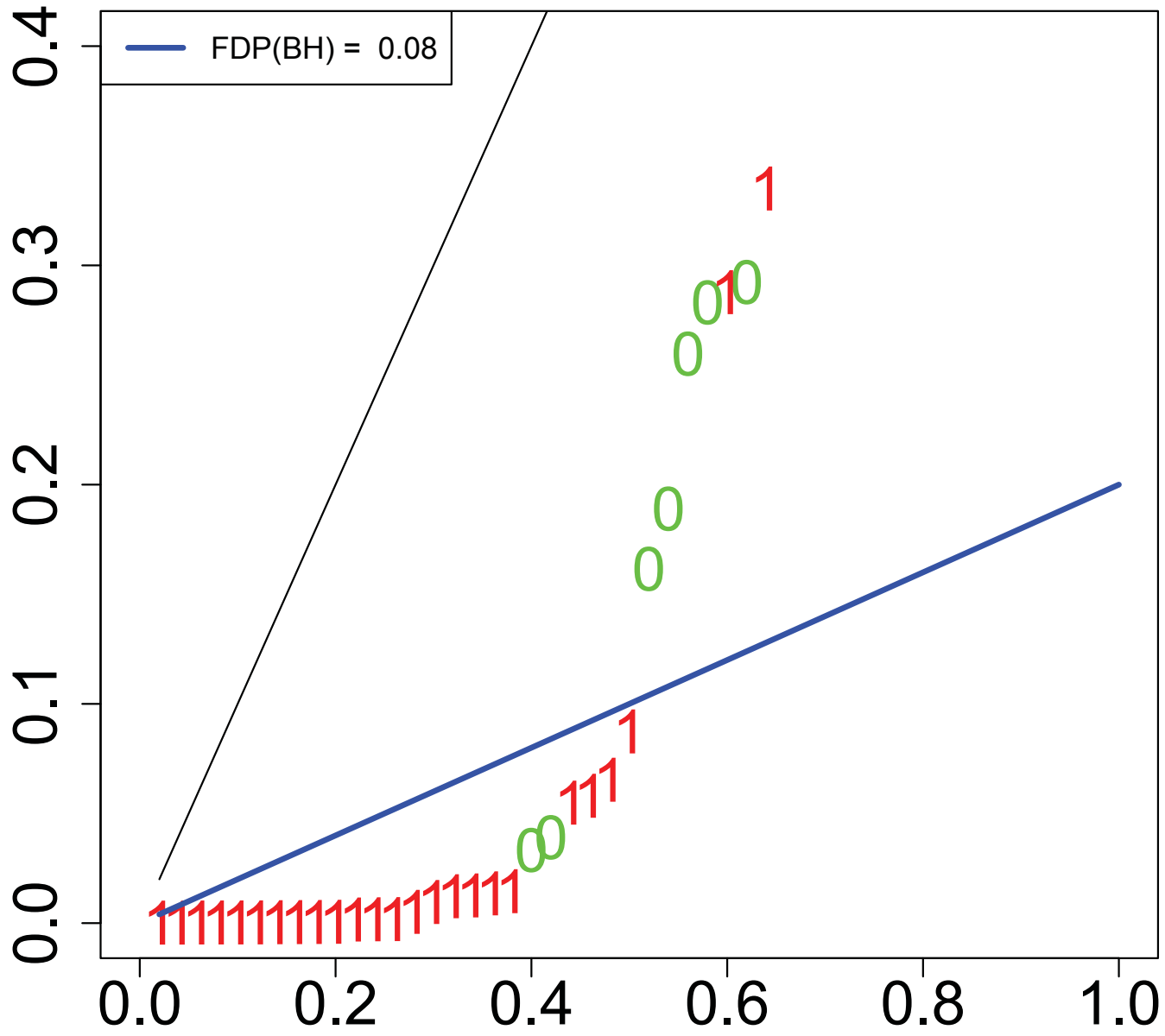
Simulations. $m = 50$; $m_0 = 25$; $\mu = 3$; $\Gamma = I_m$



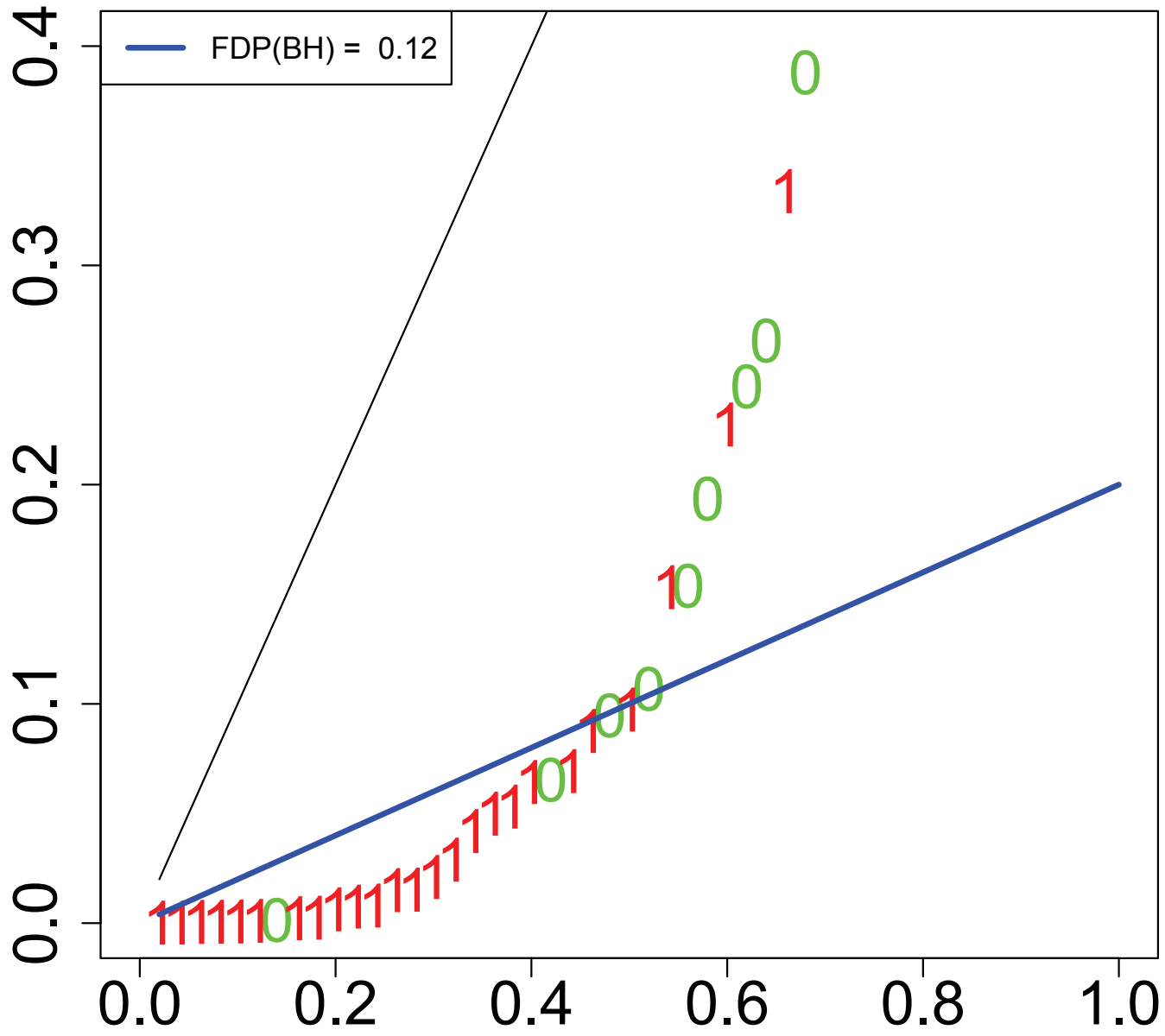
Simulations. $m = 50$; $m_0 = 25$; $\mu = 3$; $\Gamma = I_m$



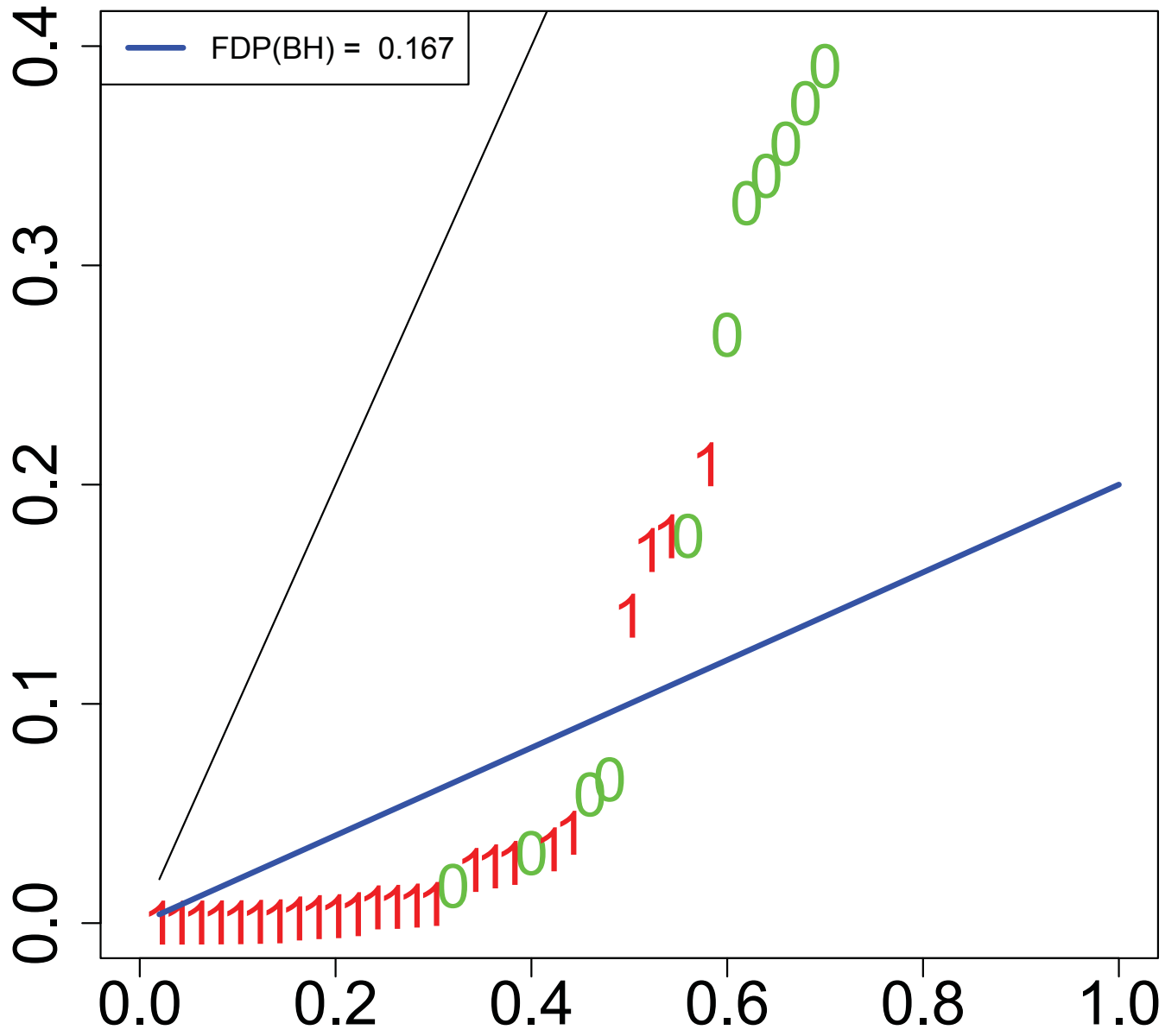
Simulations. $m = 50$; $m_0 = 25$; $\mu = 3$; $\Gamma = I_m$



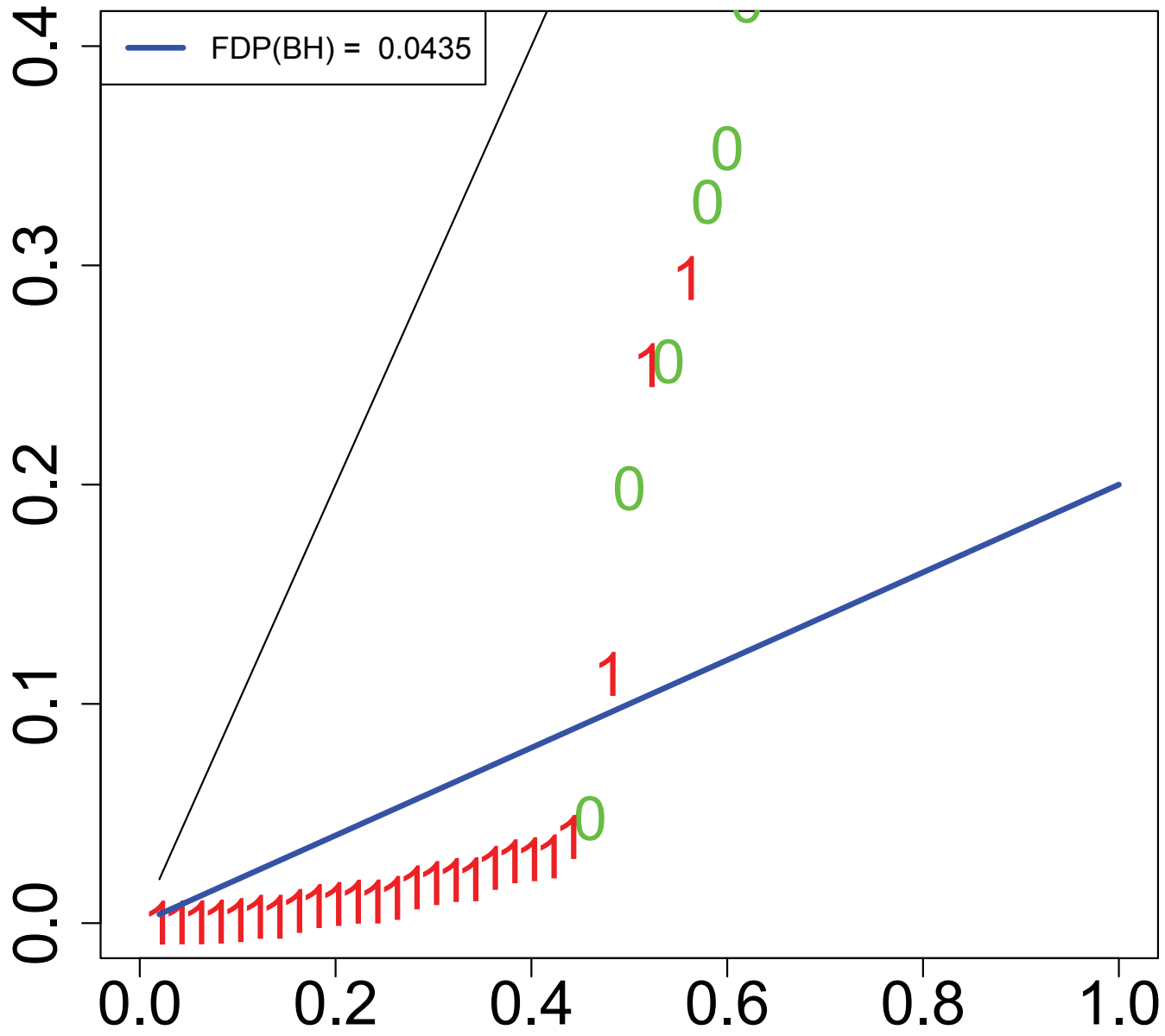
Simulations. $m = 50$; $m_0 = 25$; $\mu = 3$; $\Gamma = I_m$



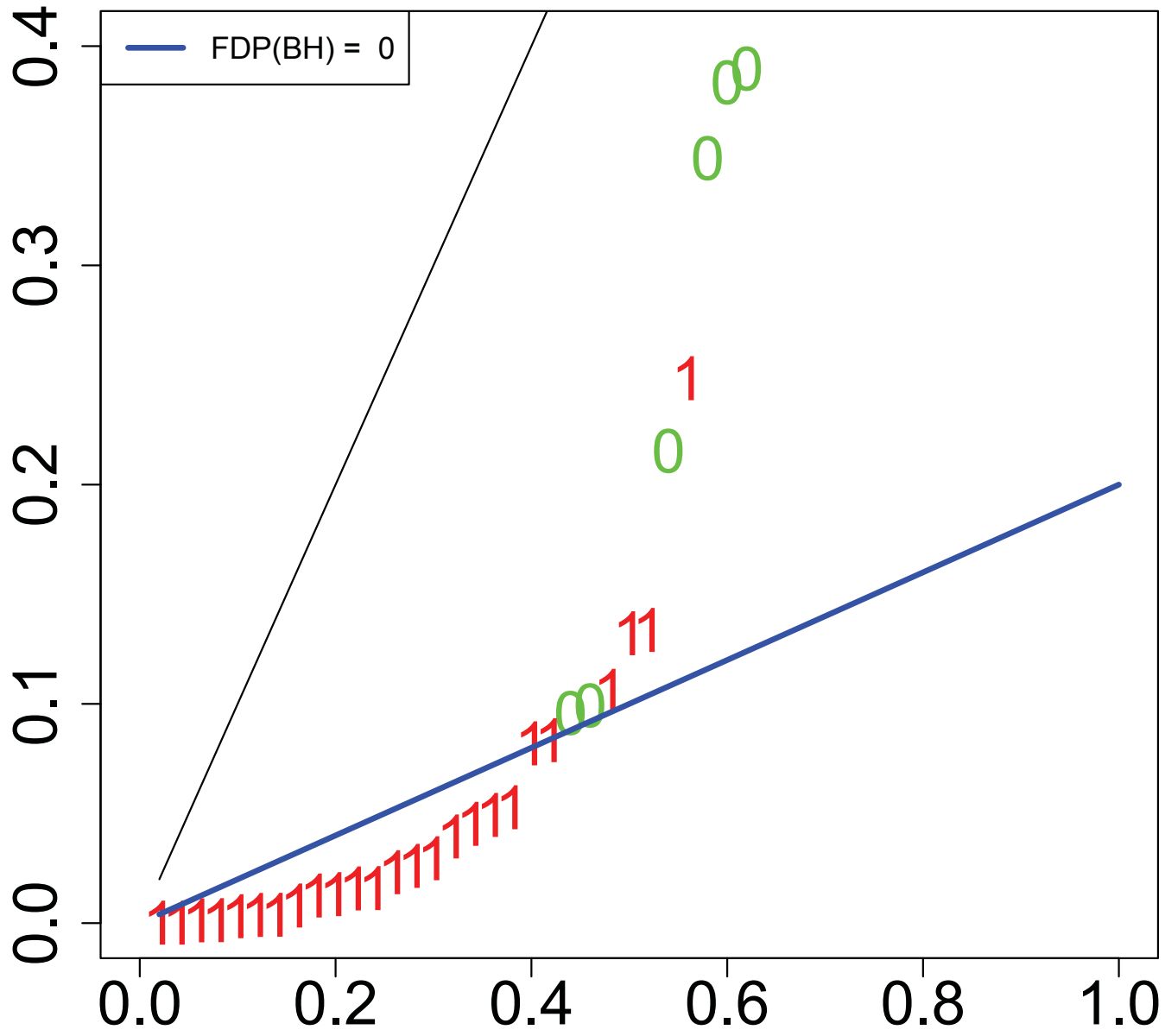
Simulations. $m = 50$; $m_0 = 25$; $\mu = 3$; $\Gamma = I_m$



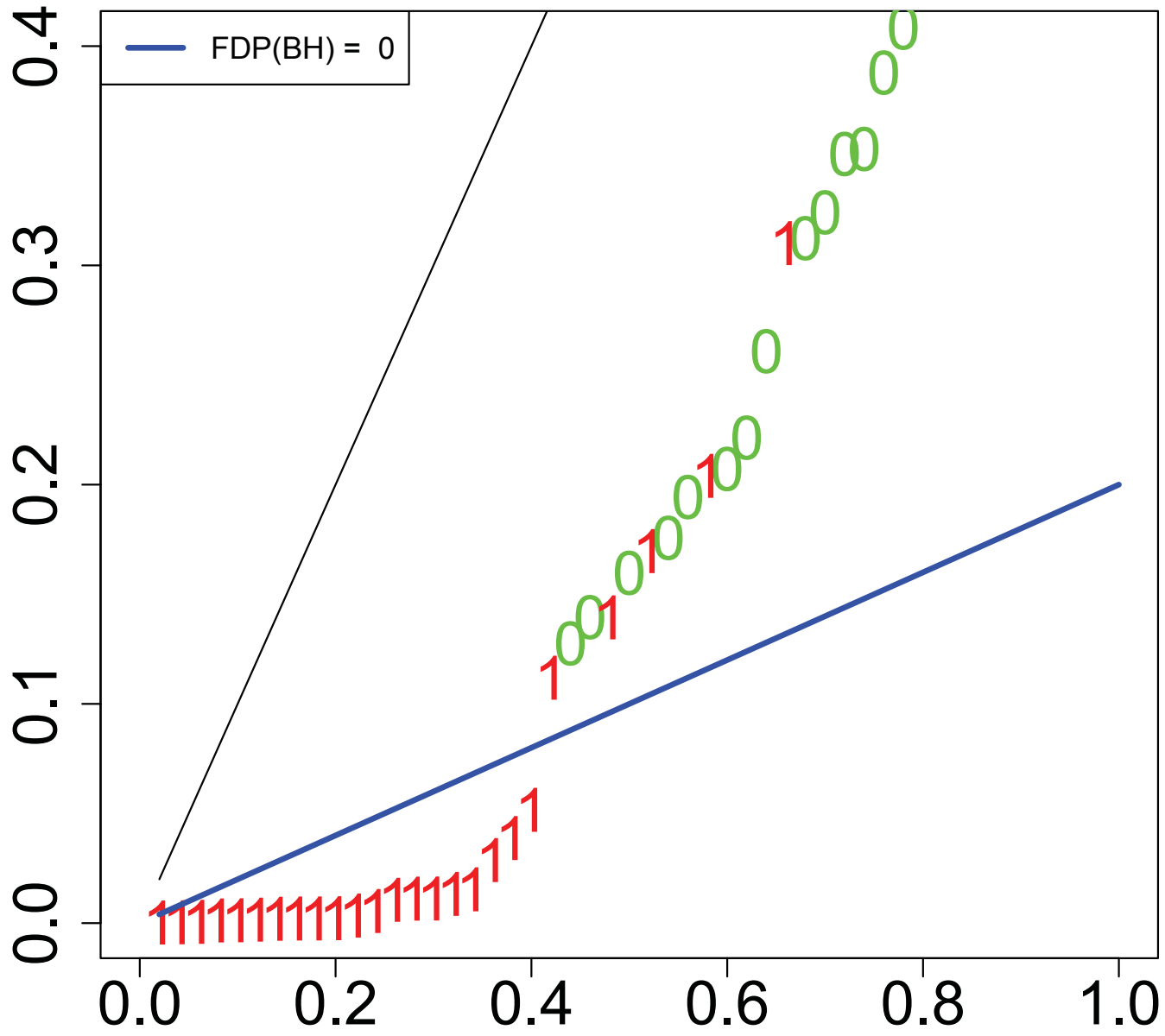
Simulations. $m = 50$; $m_0 = 25$; $\mu = 3$; $\Gamma = I_m$



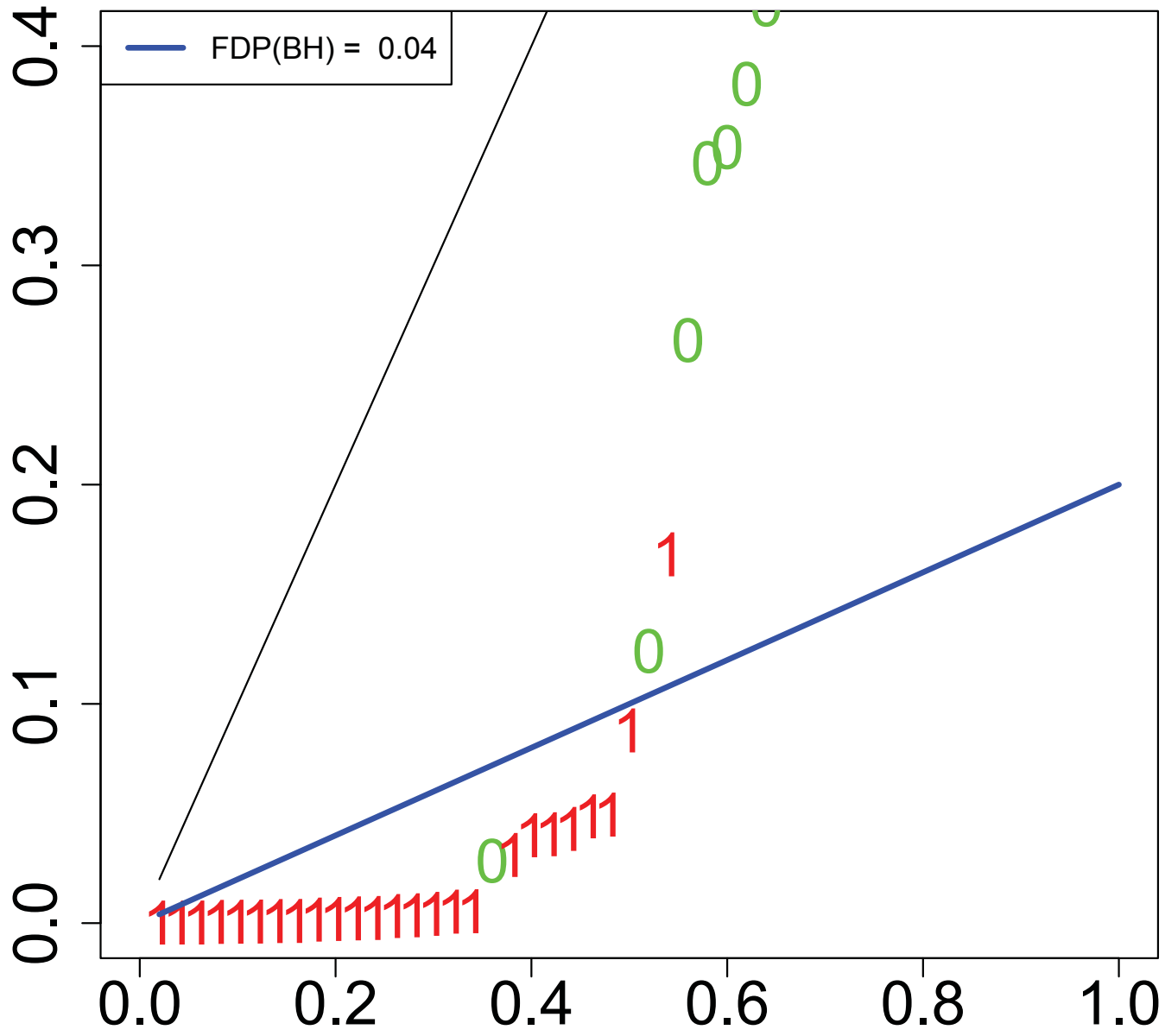
Simulations. $m = 50$; $m_0 = 25$; $\mu = 3$; $\Gamma = I_m$



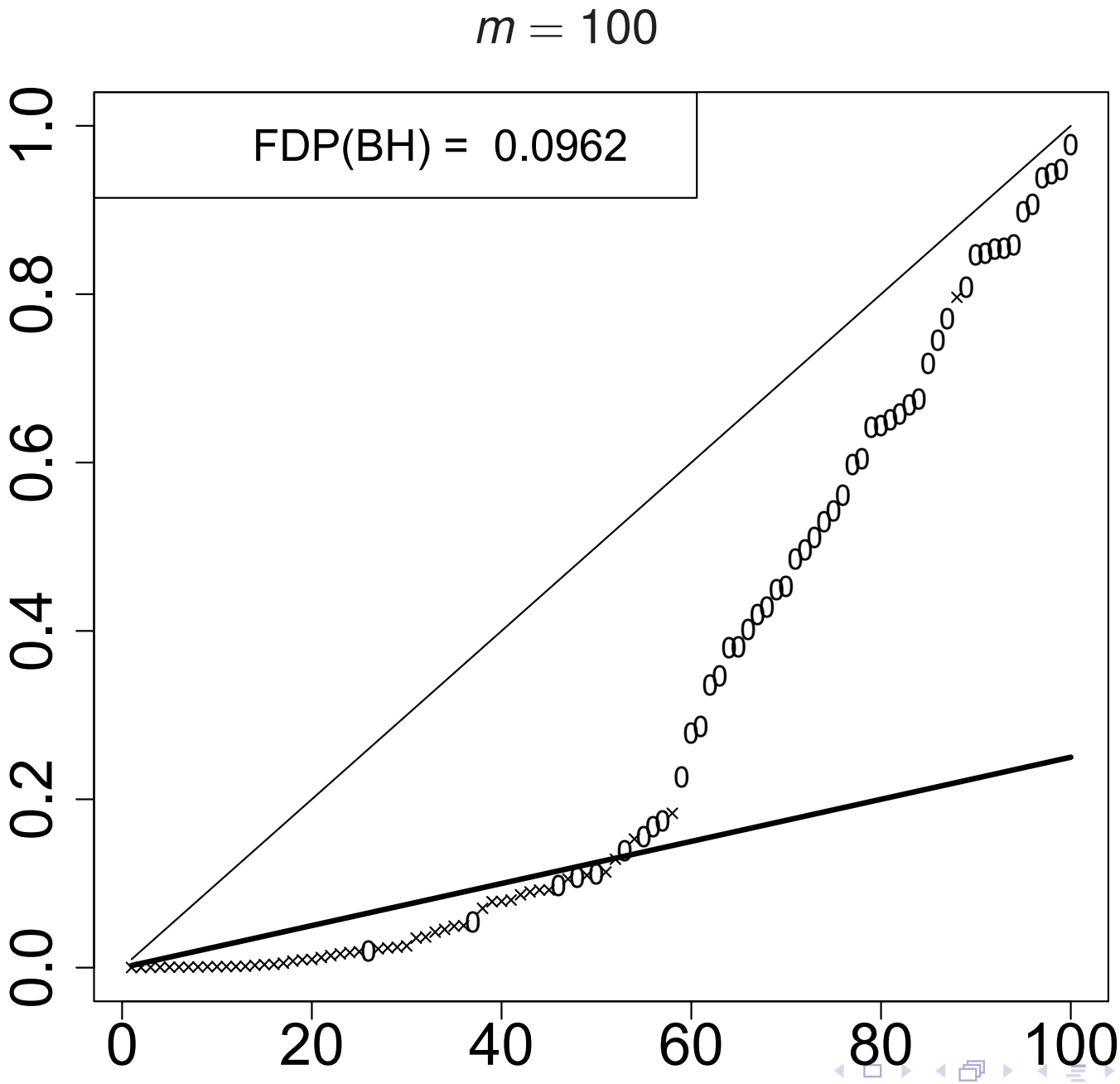
Simulations. $m = 50$; $m_0 = 25$; $\mu = 3$; $\Gamma = I_m$



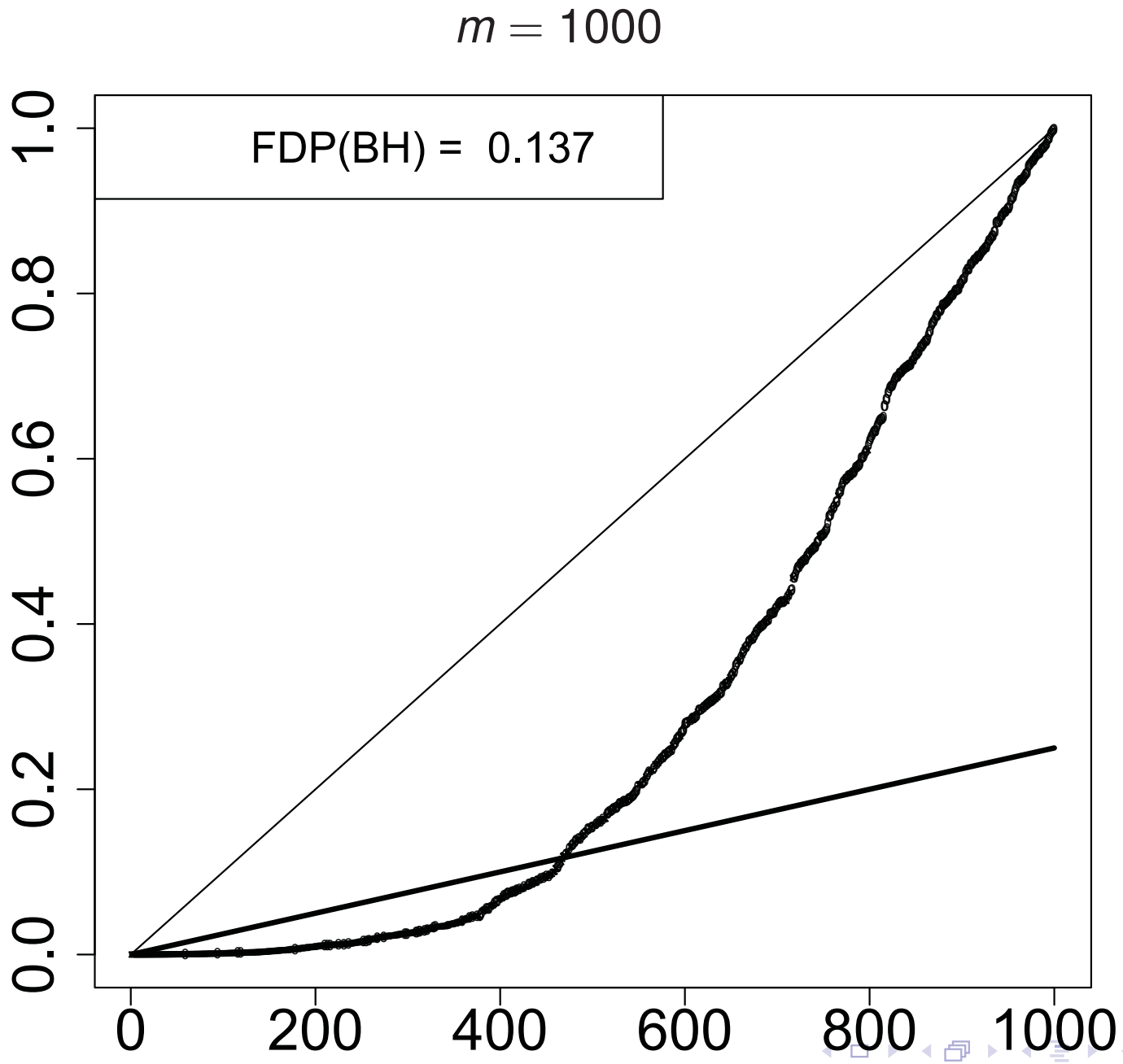
Simulations. $m = 50$; $m_0 = 25$; $\mu = 3$; $\Gamma = I_m$



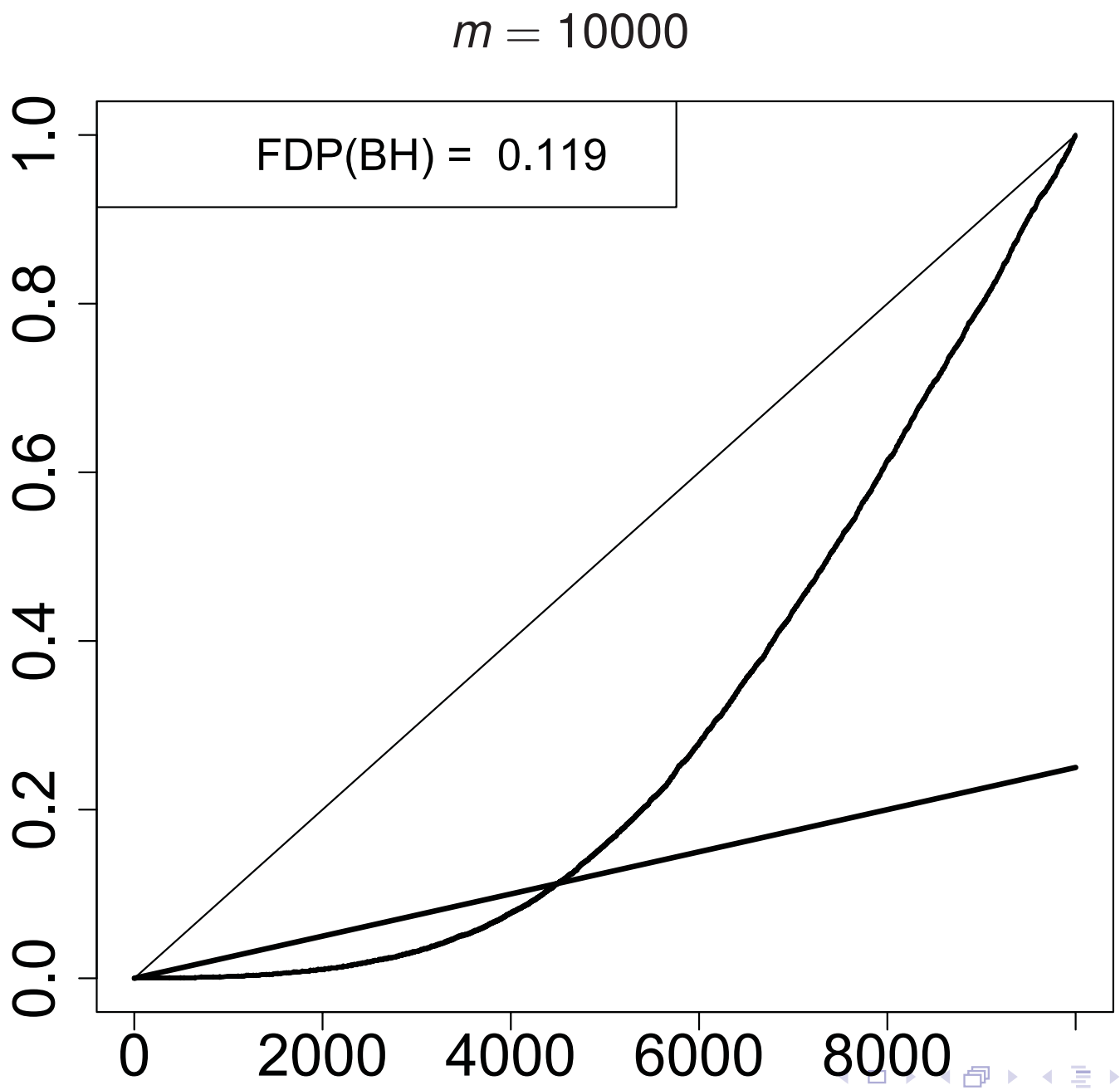
Scale invariance property $\pi_0 = 0.5$; $\mu = 2$



Scale invariance property $\pi_0 = 0.5$; $\mu = 2$



Scale invariance property $\pi_0 = 0.5$; $\mu = 2$



Benjamini and Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing

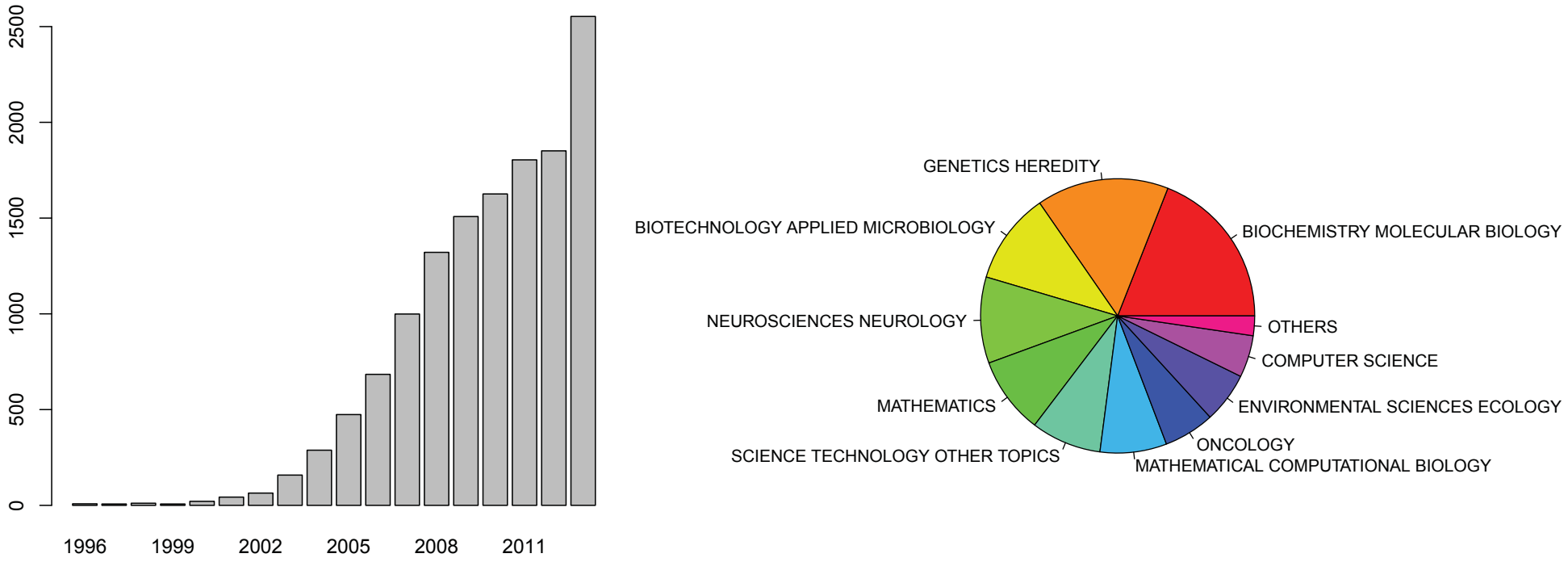


Figure: Statistics for the 13,427 papers citing this work from 1996 to 2013 and according to "the web of science". Left: per year. Right: per research fields.

FDR control under dependence

Theorem [Benjamini and Yekutieli (2001)]

If $\Gamma_{i,j} \geq 0$ for $i \neq j$ (PRDS) and $\widehat{t}$ threshold of BH procedure, $\forall \mu, H$,

$$\text{FDR}(\widehat{t}) \leq (m_0/m)\alpha \leq \alpha$$

Well-accepted property [Farcomeni (2006)], [Kim and van de Wiel (2008)],...

FDR(BH) $\leq$ or $\simeq \alpha$ for any “realistic" dependencies

End of the story?

Consider ρ-equicorrelation: $\Gamma = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{pmatrix}$ with $\rho \geq 0$

FDR control under dependence

Theorem [Benjamini and Yekutieli (2001)]

If $\Gamma_{i,j} \geq 0$ for $i \neq j$ (PRDS) and $\widehat{t}$ threshold of BH procedure, $\forall \mu, H$,

$$\mathrm{FDR}(\widehat{t}) \leq (m_0/m)\alpha \leq \alpha$$

Well-accepted property [Farcomeni (2006)], [Kim and van de Wiel (2008)],...

$\mathrm{FDR}(BH) \leq$ or $\simeq \alpha$ for any "realistic" dependencies

End of the story?

Consider ρ-equicorrelation: $\Gamma = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{pmatrix}$ with $\rho \geq 0$

FDR control under dependence

Theorem [Benjamini and Yekutieli (2001)]

If $\Gamma_{i,j} \geq 0$ for $i \neq j$ (PRDS) and $\widehat{t}$ threshold of BH procedure, $\forall \mu, H$,

$$\mathrm{FDR}(\widehat{t}) \leq (m_0/m)\alpha \leq \alpha$$

Well-accepted property [Farcomeni (2006)], [Kim and van de Wiel (2008)],...

$\mathrm{FDR}(BH) \leq$ or $\simeq \alpha$ for any "realistic" dependencies

End of the story?

Consider ρ-equicorrelation: $\Gamma = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{pmatrix}$ with $\rho \geq 0$

FDR control under dependence

Theorem [Benjamini and Yekutieli (2001)]

If $\Gamma_{i,j} \geq 0$ for $i \neq j$ (PRDS) and $\widehat{t}$ threshold of BH procedure, $\forall \mu, H$,

$$\mathrm{FDR}(\widehat{t}) \leq (m_0/m)\alpha \leq \alpha$$

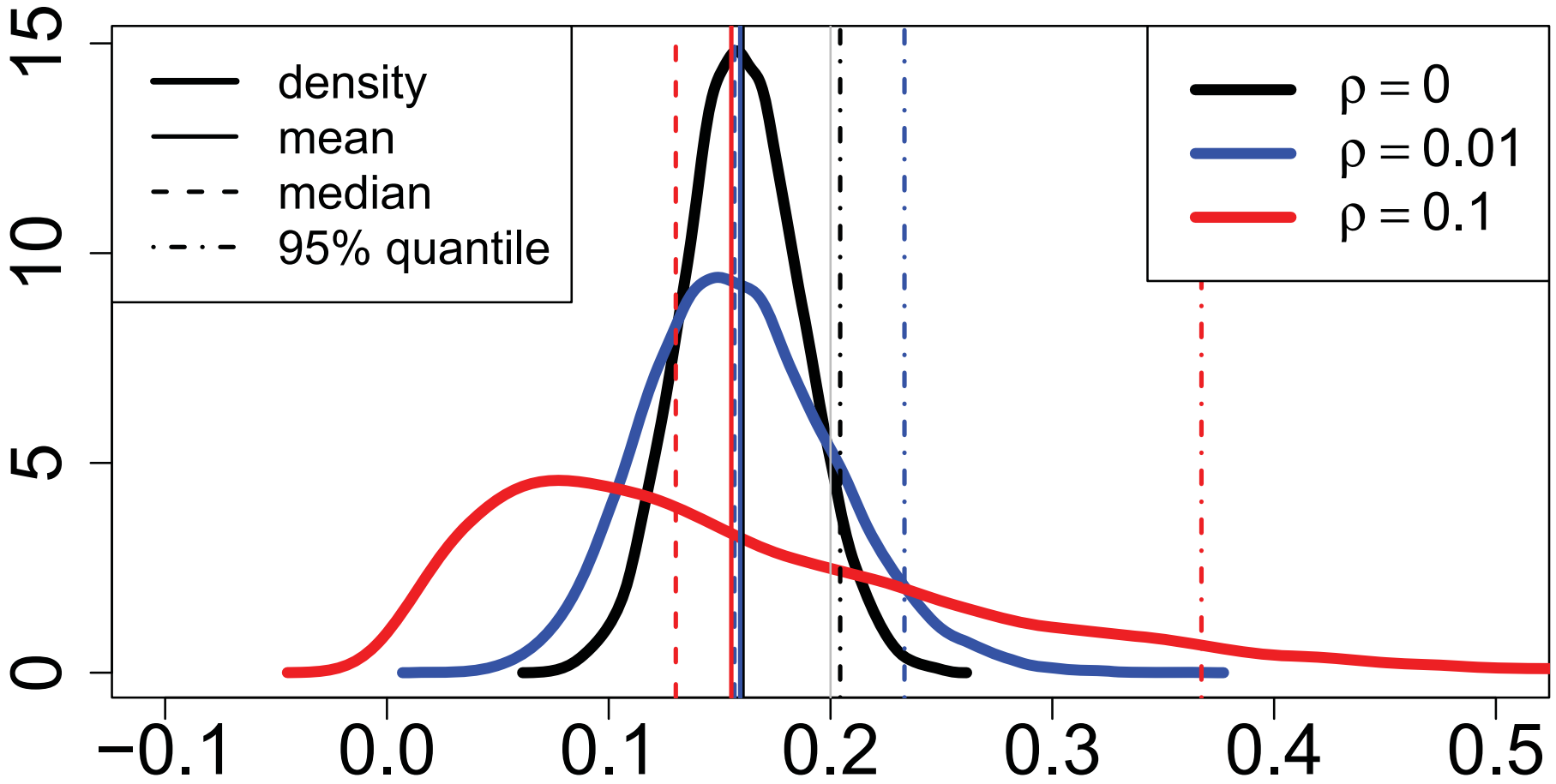
Well-accepted property [Farcomeni (2006)], [Kim and van de Wiel (2008)],...

$\mathrm{FDR}(BH) \leq$ or $\simeq \alpha$ for any "realistic" dependencies

End of the story?

Consider ρ-equicorrelation: $\Gamma = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{pmatrix}$ with $\rho \geq 0$

True underlying FDP(BH) $m = 1000$, $\pi_0 = 0.8$, $\alpha = 0.2$



True underlying FDP(BH) $m = 1000$, $\pi_0 = 0.8$, $\alpha = 0.2$

$$\text{FDR} = \mathbf{E}[\text{FDP}] \leq \alpha$$

does not provide

$$\text{FDP} \leq \alpha \text{ or even FDP} \simeq \alpha!$$

New challenges under dependence

- **Task 1**: behavior of FDP(BH) under "weak" dependence

  [Genovese Wasserman (2004)], [Farcomeni (2007,2008)], [Neuvial (2008)]
  FDP(BH)=$\Psi\left(\widehat{F}_0, \widehat{F}_1\right)$

- **Task 2**: control the FDP: for $\alpha, \zeta \in (0,1)$, get $\widehat{t} = \widehat{t}(\Gamma)$ s.t.

$$\mathbf{P}(\mathrm{FDP}(\widehat{t}) \leq \alpha) \geq 1 - \zeta.$$

  [Lehmann and Romano (2005)], [Romano and Shaikh (2006a, 2006b)], [Romano and Wolf (2007)], [Guo and Romano (2007)], [Guo, He and Sarkar (2013)]

- **Task 3**: Build new test statistics $X_i^\star \neq X_i$ that try to remove Γ.

  [Friguet et al. (2009), Fan et al. (2012)]

Our starting point: ρ-equicorrelation

New challenges under dependence

- **Task 1**: behavior of FDP(BH) under "weak" dependence

  [Genovese Wasserman (2004)], [Farcomeni (2007,2008)], [Neuvial (2008)]
  FDP(BH)=$\Psi\left(\widehat{F}_0, \widehat{F}_1\right)$

- **Task 2**: control the FDP: for $\alpha, \zeta \in (0,1)$, get $\widehat{t} = \widehat{t}(\Gamma)$ s.t.

$$\mathbf{P}(\mathrm{FDP}(\widehat{t}) \leq \alpha) \geq 1 - \zeta.$$

  [Lehmann and Romano (2005)], [Romano and Shaikh (2006a, 2006b)], [Romano and Wolf (2007)], [Guo and Romano (2007)], [Guo, He and Sarkar (2013)]

- **Task 3**: Build new test statistics $X_i^\star \neq X_i$ that try to remove Γ.

  [Friguet et al. (2009), Fan et al. (2012)]

Our starting point: ρ-equicorrelation

New results under weak dependence

Theorem (weak equicorrelation)

$\rho = \rho_m \to 0$, $m_0/m \to \pi_0 \in (0,1)$, $r_m = (m^{-1} + |\rho_m|)^{-1/2}$ then

$$r_m\big(\mathrm{FDP}(BH) - \pi_0\alpha\big) \rightsquigarrow \mathcal{N}(0, v)\ ;\ \ v \in (0,\infty)$$

In particular,

$$\mathbf{P}(\mathrm{FDP}(BH) \leq \alpha) \to 1 \geq 1 - \zeta$$

- BH works well under weak dep (asymp)
- convergence rate can be slow
- Relaxation to general weak dependence

$$\frac{1}{m(m-1)} \sum_{i \neq j} \Gamma_{i,j}^2 \to 0$$

New results under weak dependence

Theorem (weak equicorrelation)

$\rho = \rho_m \to 0$, $m_0/m \to \pi_0 \in (0,1)$, $r_m = (m^{-1} + |\rho_m|)^{-1/2}$ then

$$r_m\big(\mathrm{FDP}(BH) - \pi_0\alpha\big) \rightsquigarrow \mathcal{N}(0,v)\ ;\ \ v \in (0,\infty)$$

In particular,

$$\mathbf{P}(\mathrm{FDP}(BH) \leq \alpha) \to 1 \geq 1-\zeta$$

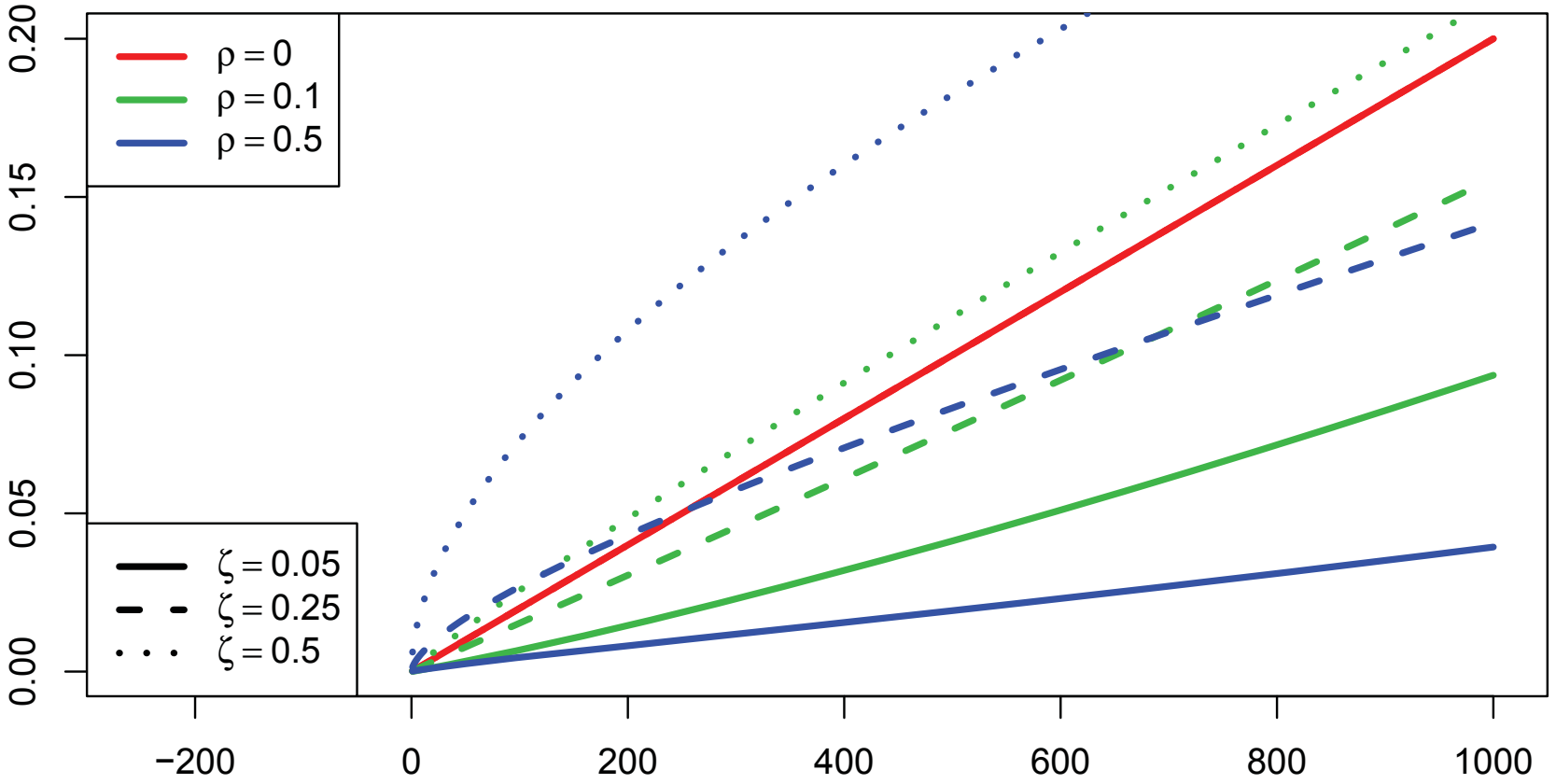
- BH works well under weak dep (asymp)
- convergence rate can be slow
- Relaxation to general weak dependence

$$\frac{1}{m(m-1)} \sum_{i \neq j} \Gamma_{i,j}^2 \to 0$$

New results under strong dependence

New BH type critical values under ρ-equicorrelation:

$$\tau_\ell = \Phi\left(\rho^{1/2}\Phi^{-1}(\zeta) + (1-\rho)^{1/2}\Phi^{-1}(\alpha\ell/m)\right), \quad 1 \leq \ell \leq m.$$



New results under strong dependence

Theorem (strong equicorrelation)

$\rho \in (0,1)$ (fixed!), $m_0/m \to \pi_0 \in (0,1)$, new critical values

$$\tau_\ell = \Phi\left(\rho^{1/2}\Phi^{-1}(\zeta) + (1-\rho)^{1/2}\Phi^{-1}(\alpha\ell/m)\right), \quad 1 \leq \ell \leq m.$$

Then

$$\liminf_m \mathbf{P}(\mathrm{FDP}(\mathit{New}) \leq \alpha) \geq 1 - \zeta.$$

- Plug-in possible: $(\log m)(\widehat{\rho}_m - \rho_m)^2 = o_P(1)$
- Relaxation to general dependence (Romano-Wolf's heuristic)

New results under strong dependence

Theorem (strong equicorrelation)

$\rho \in (0,1)$ (fixed!), $m_0/m \to \pi_0 \in (0,1)$, new critical values

$$\tau_\ell = \Phi\left(\rho^{1/2}\Phi^{-1}(\zeta) + (1-\rho)^{1/2}\Phi^{-1}(\alpha\ell/m)\right), \quad 1 \leq \ell \leq m.$$

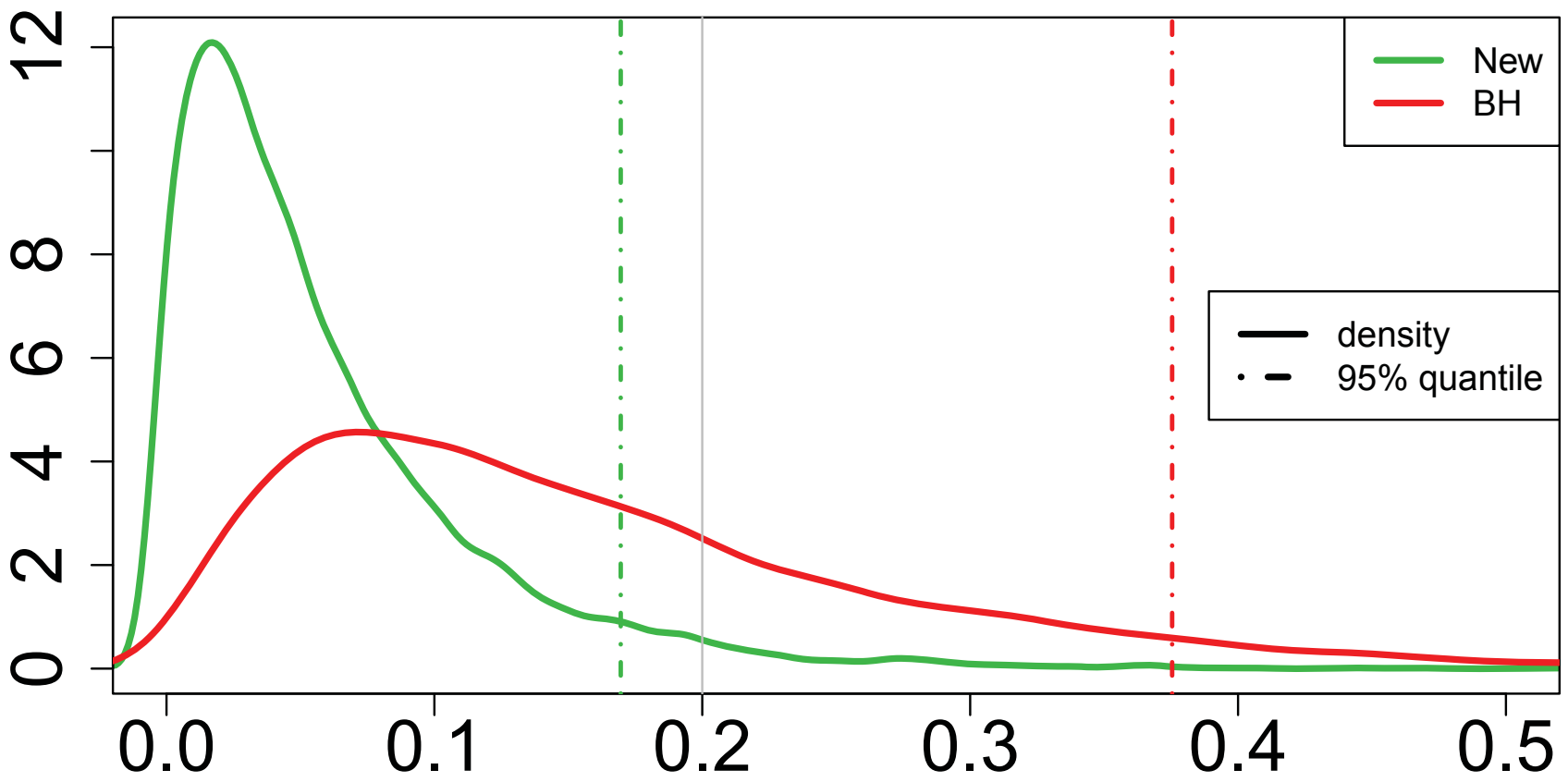
Then

$$\liminf_m \mathbf{P}(\mathrm{FDP}(\mathit{New}) \leq \alpha) \geq 1 - \zeta.$$

- Plug-in possible: $(\log m)(\widehat{\rho}_m - \rho_m)^2 = o_P(1)$
- Relaxation to general dependence (Romano-Wolf's heuristic)

Distribution of FDP(New)

$m = 1000$, $\alpha = 0.2$, $\zeta = 0.05$, $\rho = 0.1$



Outlook

Some take home messages

- BH procedure does not suffer from the curse of dimensionality
- “works" under weak dependence
- when equicorrelated strong: new critical values using ρ

Open problems

- FDP control in general, when Γ unknown (resampling)?
- when dependence strong and “well structured": change test statistics?
- optimality issues

Outlook

Some take home messages

- BH procedure does not suffer from the curse of dimensionality
- “works" under weak dependence
- when equicorrelated strong: new critical values using ρ

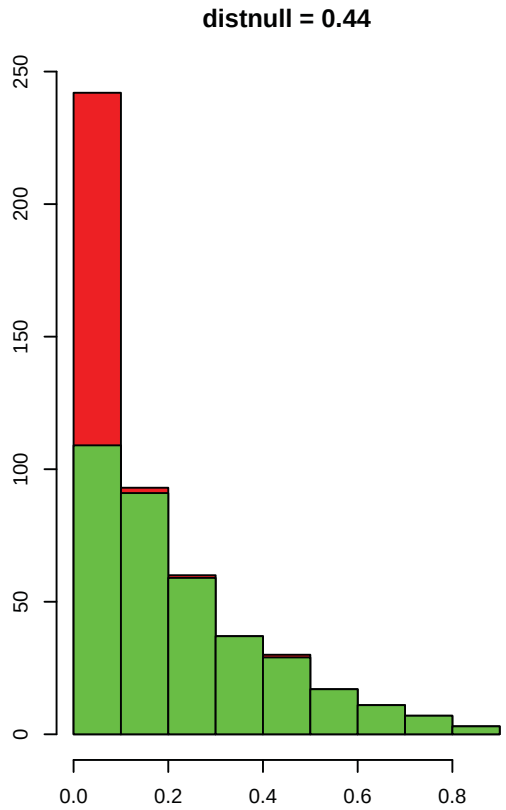
Open problems

- FDP control in general, when Γ unknown (resampling)?
- when dependence strong and “well structured": change test statistics?
- optimality issues

Removing strong dependence

Equicorrelation

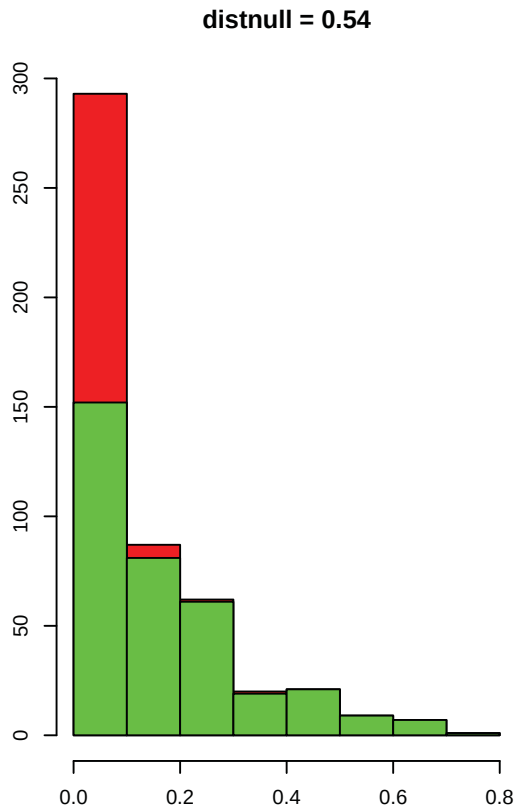
$$X_i = \mu H_i + U + \sqrt{1-\rho}\,\xi_i,$$



Removing strong dependence

Equicorrelation

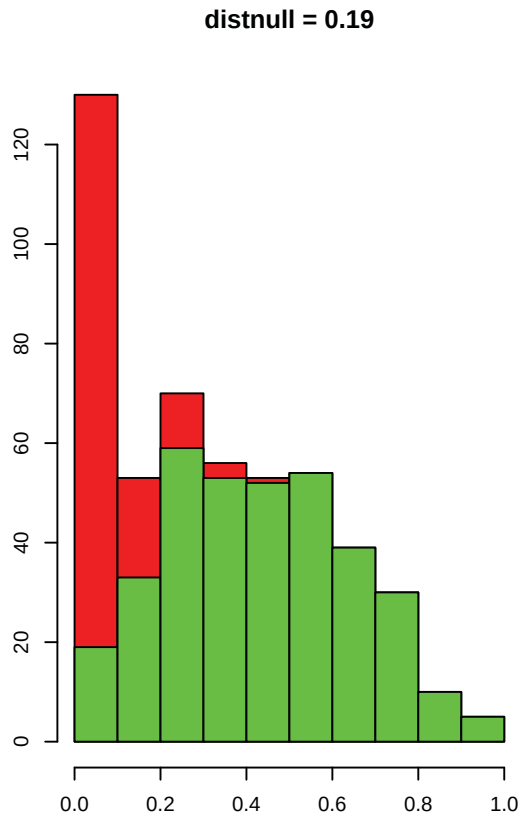
$$X_i = \mu H_i + U + \sqrt{1-\rho}\,\xi_i,$$



Removing strong dependence

Equicorrelation

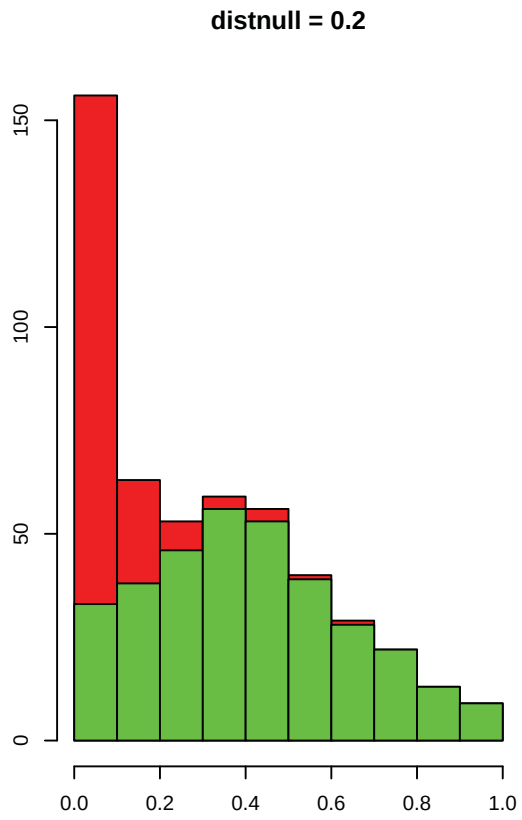
$$X_i = \mu H_i + U + \sqrt{1-\rho}\,\xi_i,$$



Removing strong dependence

Equicorrelation

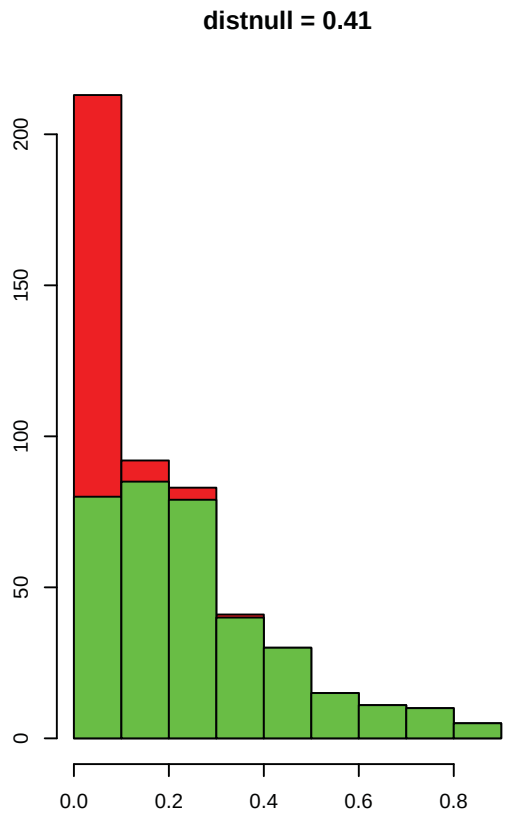
$$X_i = \mu H_i + U + \sqrt{1-\rho}\,\xi_i,$$



Removing strong dependence

Equicorrelation

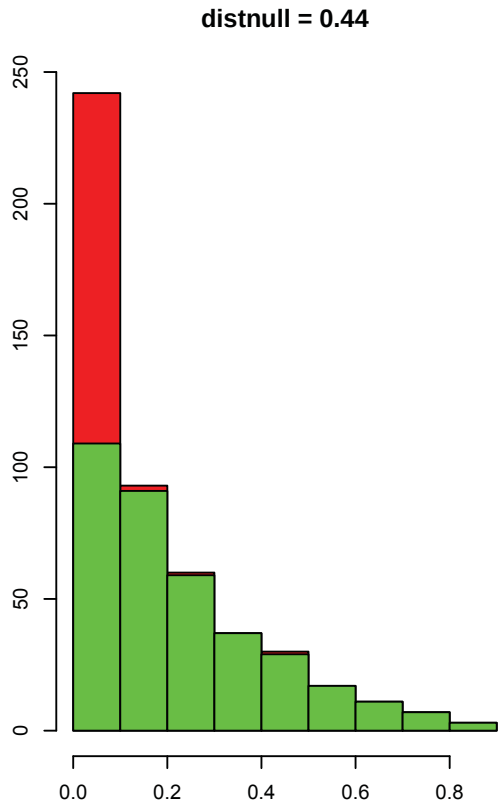
$$X_i = \mu H_i + U + \sqrt{1-\rho}\,\xi_i,$$



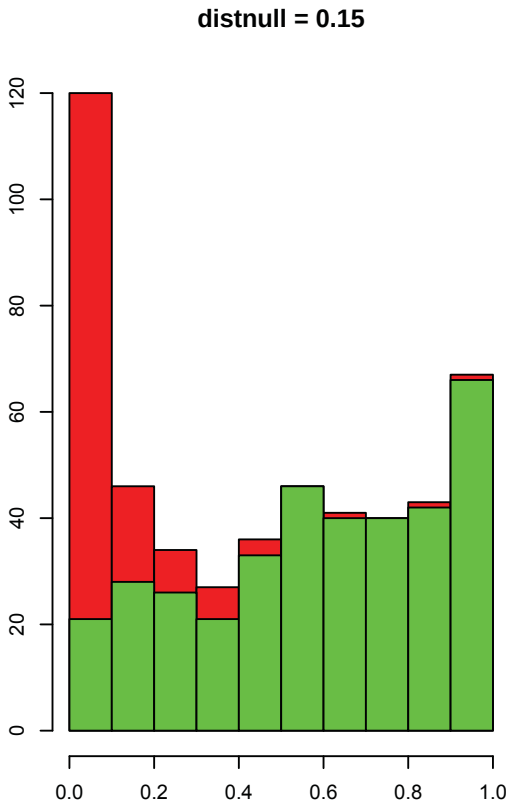
Removing strong dependence

Equicorrelation

$$X_i = \mu H_i + U + \sqrt{1-\rho}\,\xi_i, \qquad X_i^\star = X_i - \widehat{U}$$

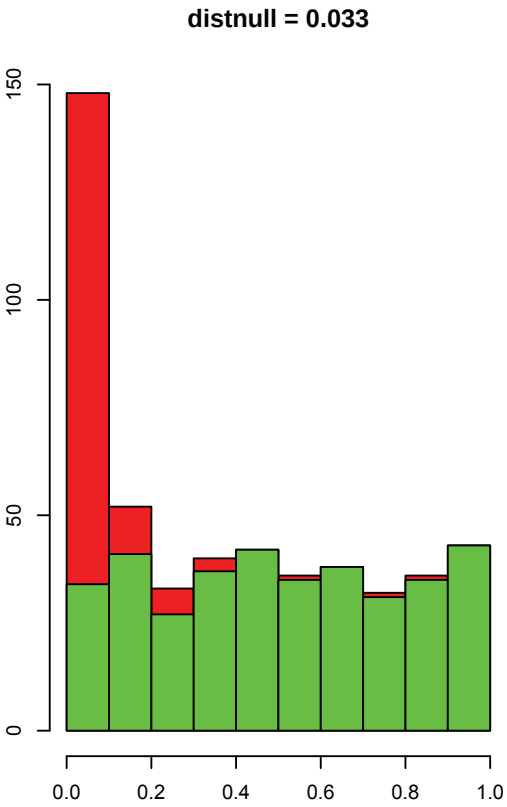


Original



$\widehat{U}_1 = X_{(0.475m)}$

[Fan et al. (2012)]

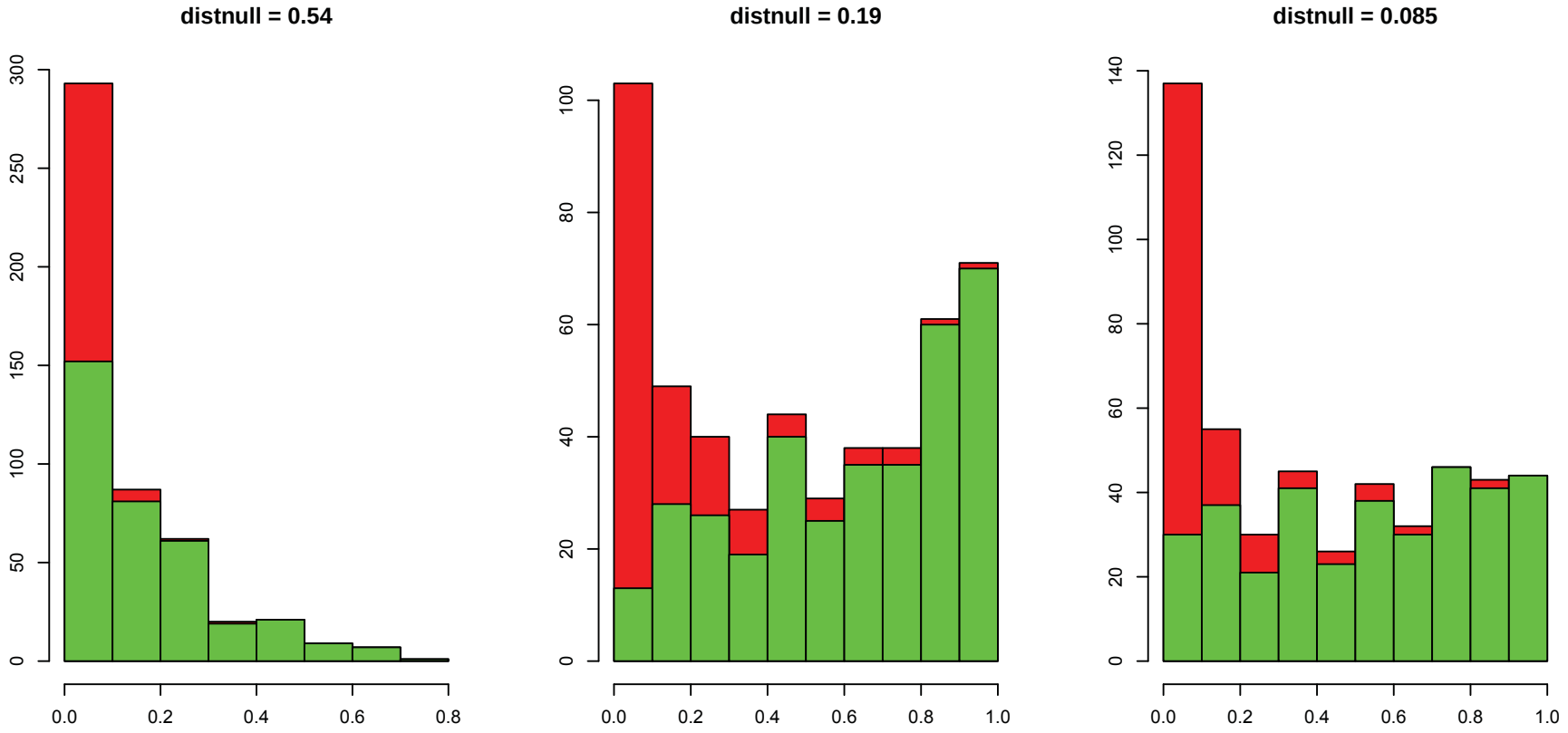


$\widehat{U}_2 = \text{Mean}_{k \le m^{1/2}}(X_{(k)} - e_k)$

Removing strong dependence

Equicorrelation

$$X_i = \mu H_i + U + \sqrt{1-\rho}\,\xi_i, \qquad X_i^\star = X_i - \widehat{U}$$



Original

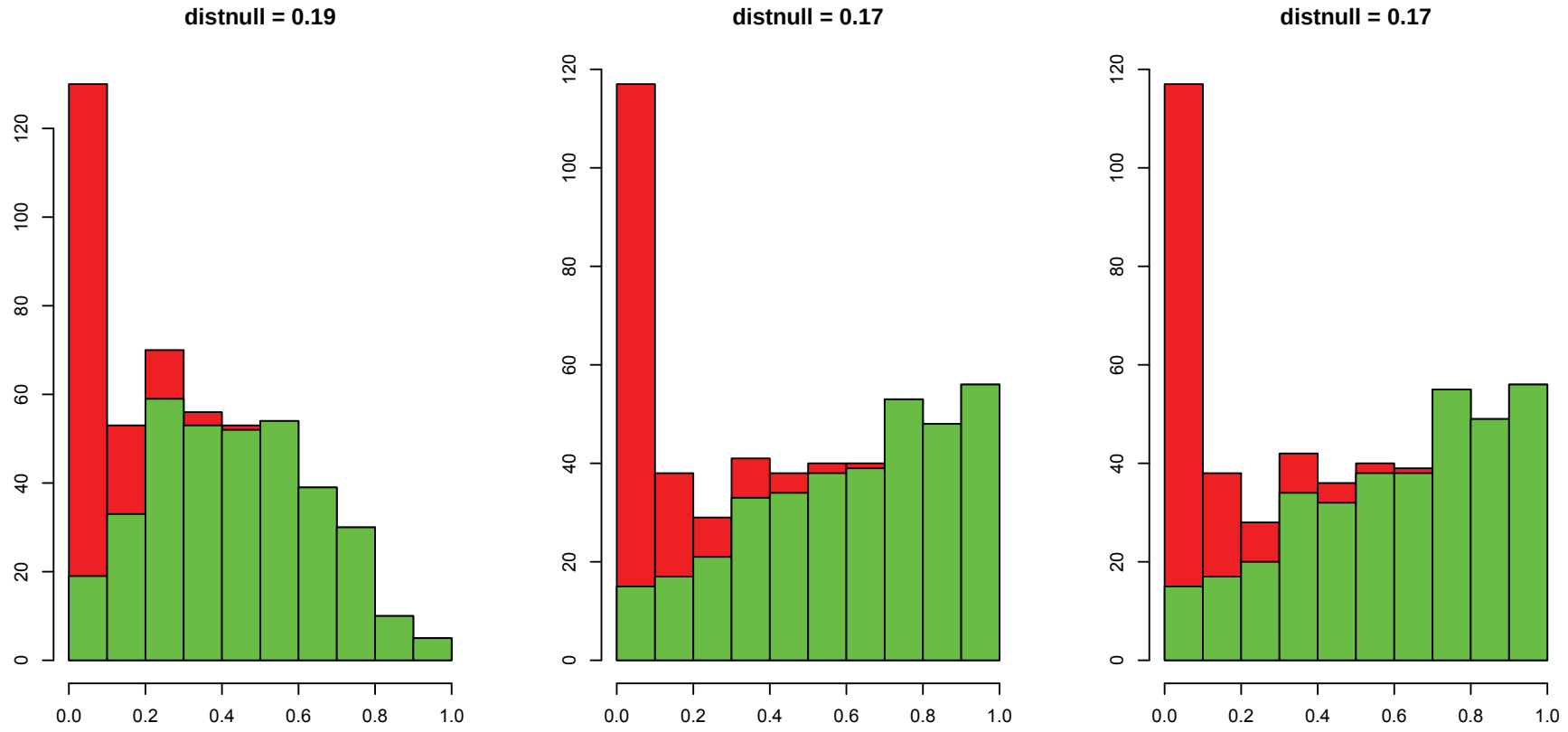
$\widehat{U}_1 = X_{(0.475m)}$ [Fan et al. (2012)]

$\widehat{U}_2 = \text{Mean}_{k \le m^{1/2}}(X_{(k)} - e_k)$

Removing strong dependence

Equicorrelation

$$X_i = \mu H_i + U + \sqrt{1-\rho}\,\xi_i, \qquad X_i^\star = X_i - \widehat{U}$$



Original

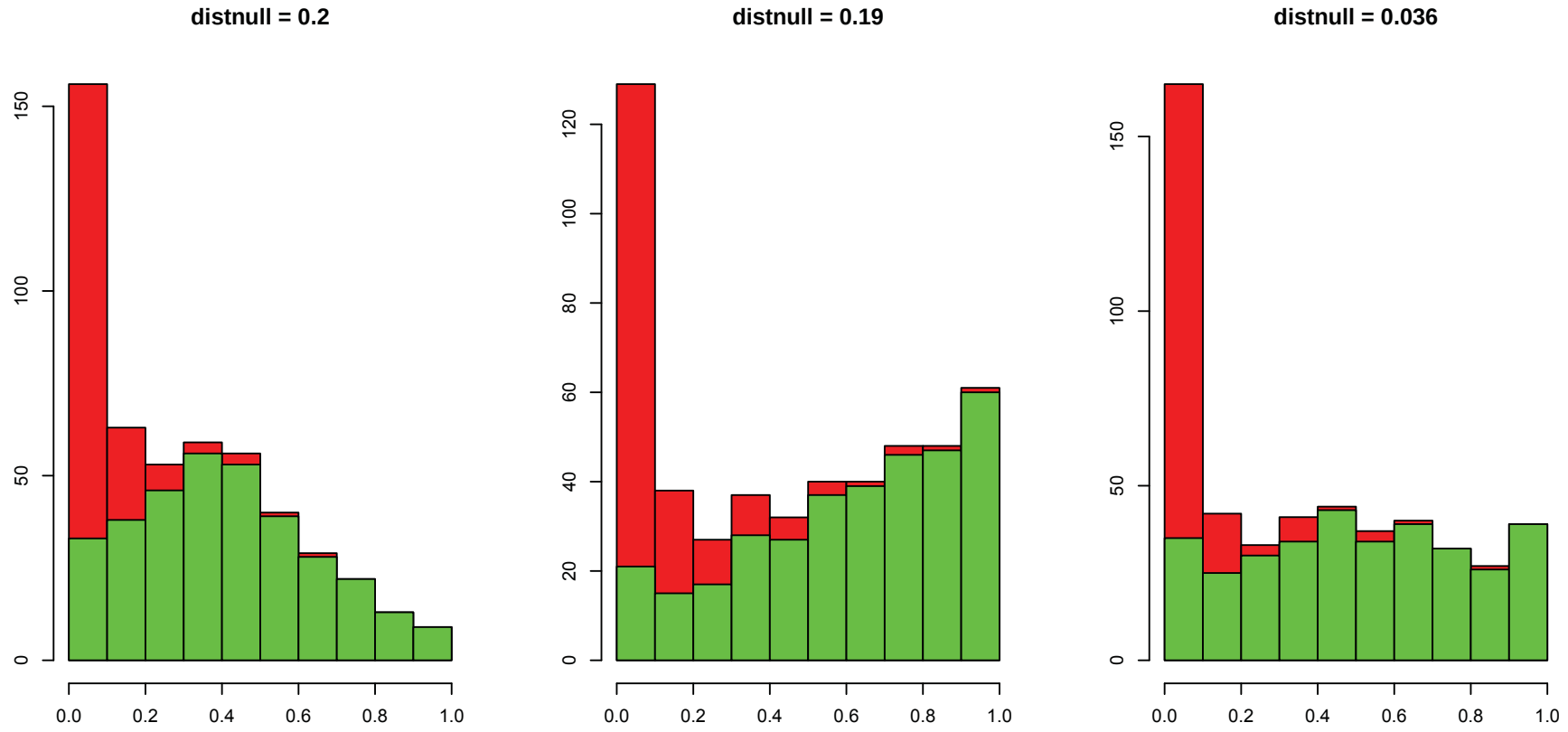
$\widehat{U}_1 = X_{(0.475m)}$ [Fan et al. (2012)]

$\widehat{U}_2 = \text{Mean}_{k \le m^{1/2}}(X_{(k)} - e_k)$

Removing strong dependence

Equicorrelation

$$X_i = \mu H_i + U + \sqrt{1-\rho}\,\xi_i, \qquad X_i^\star = X_i - \widehat{U}$$



Original

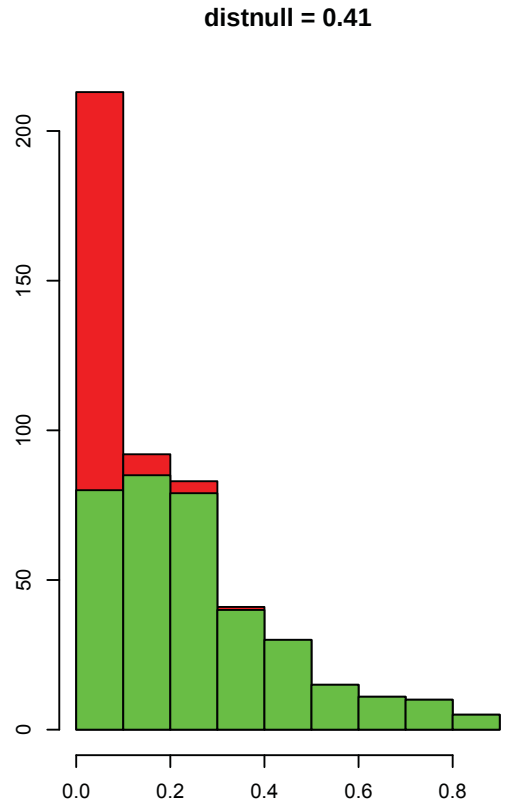
$\widehat{U}_1 = X_{(0.475m)}$ [Fan et al. (2012)]

$\widehat{U}_2 = \text{Mean}_{k \le m^{1/2}}(X_{(k)} - e_k)$

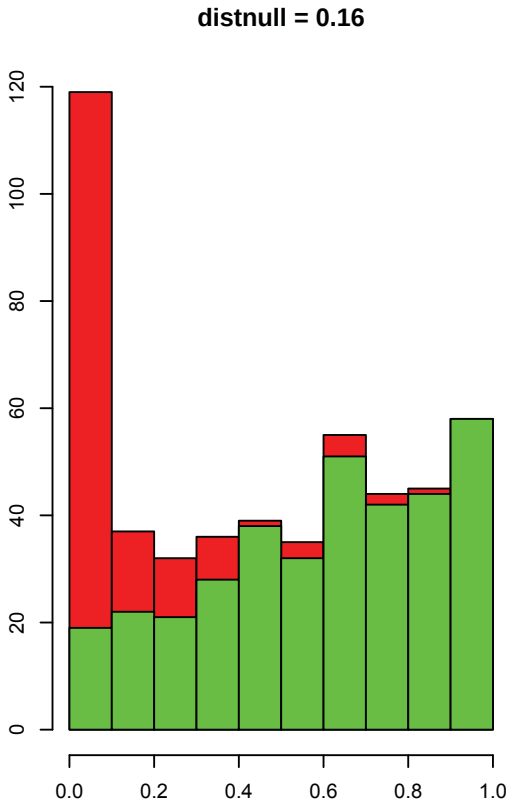
Removing strong dependence

Equicorrelation

$$X_i = \mu H_i + U + \sqrt{1-\rho}\,\xi_i, \qquad X_i^\star = X_i - \widehat{U}$$

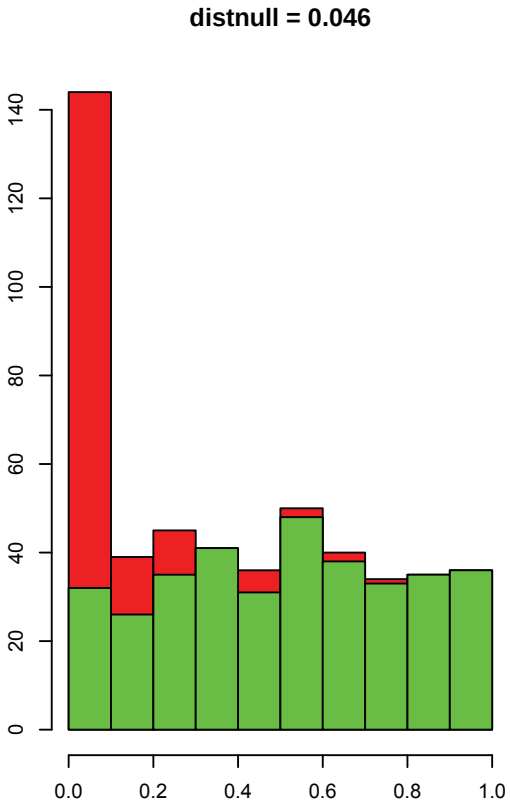


Original



$\widehat{U}_1 = X_{(0.475m)}$

[Fan et al. (2012)]



$\widehat{U}_2 = \text{Mean}_{k \le m^{1/2}}(X_{(k)} - e_k)$