Repeated Significance Tests Controlling the False Discovery Rate

Sonja Zehetmayer

Martin Posch Peter Bauer

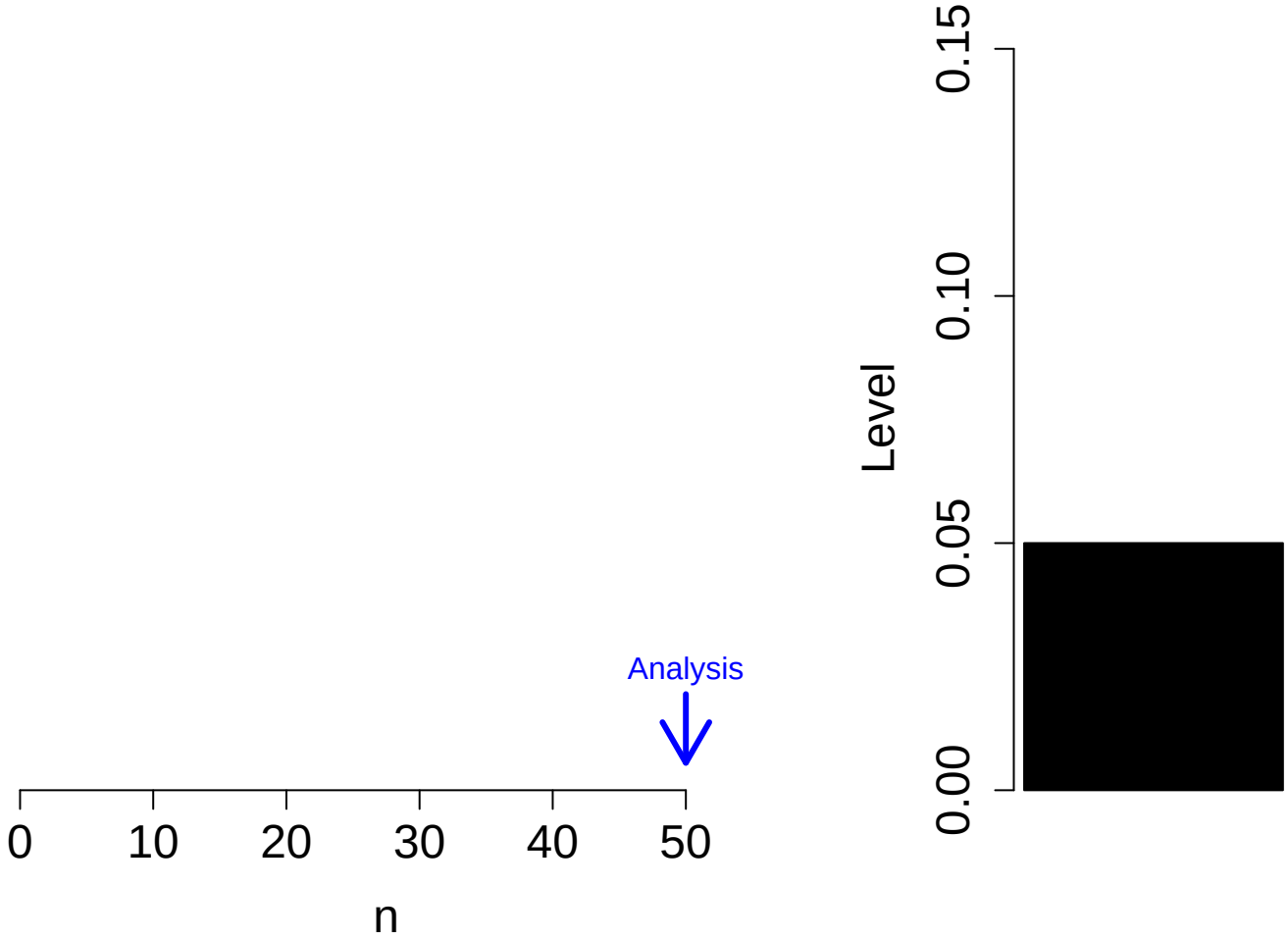
Section of Medical Statistics
Medical University of Vienna, Austria

Supported by FWF-Fund Nr. P18698-n15

ROeS, 2007

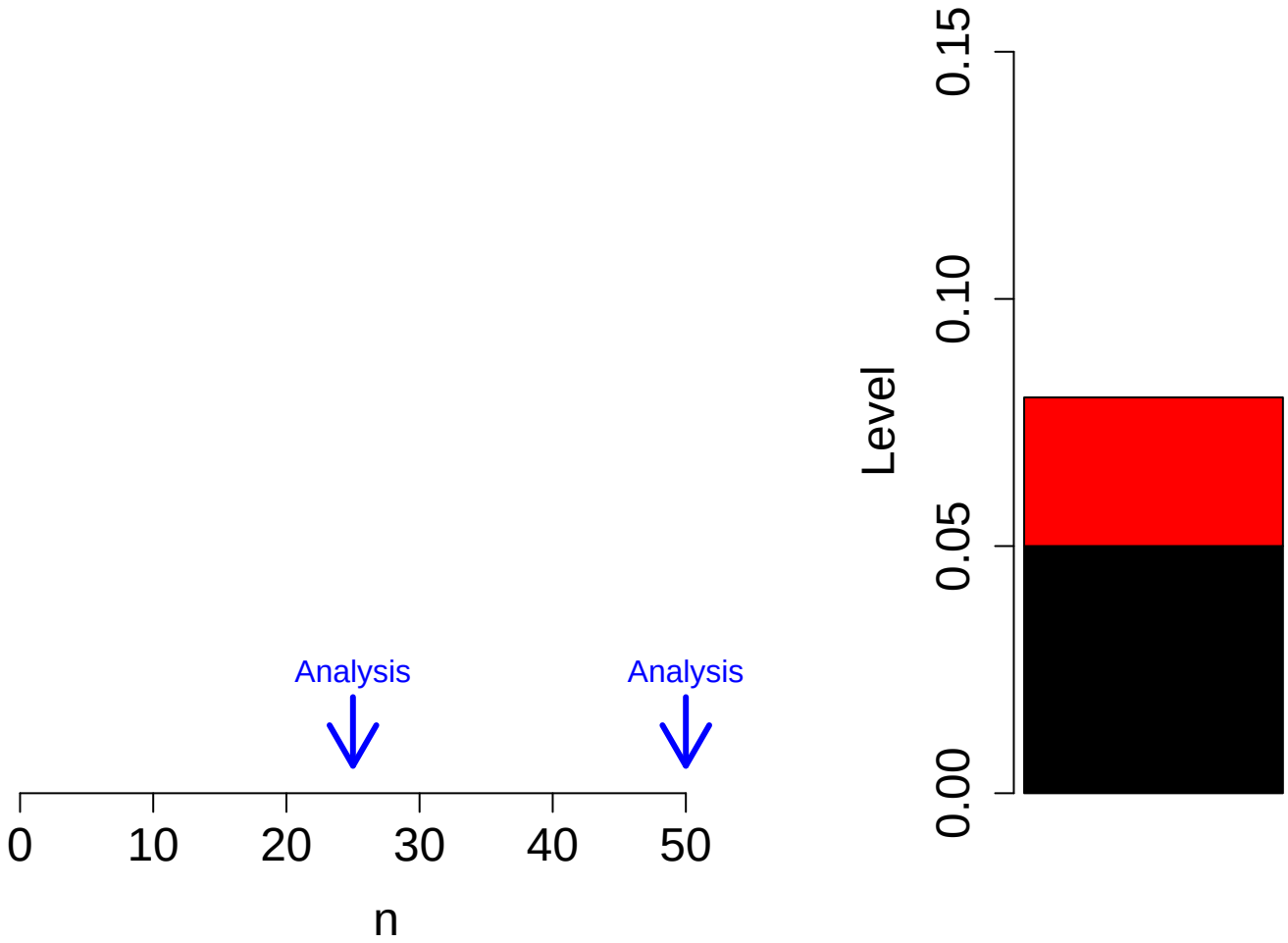
Hunting for Significance Tests for a Single Hypothesis

1 Analysis at level $\alpha = 0.05$



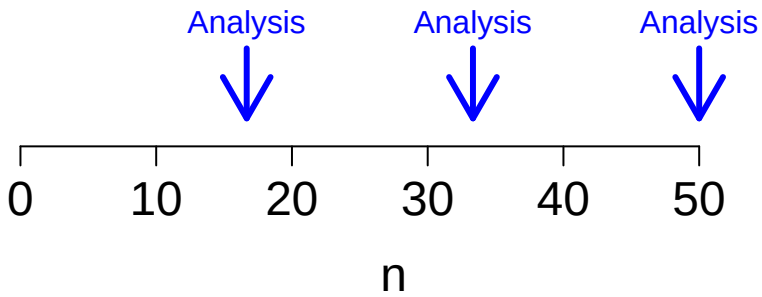
Hunting for Significance Tests for a Single Hypothesis

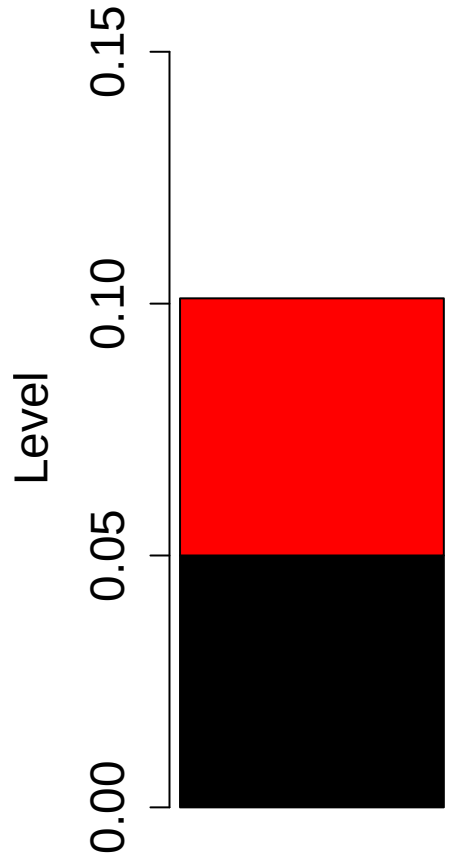
2 Analyses at the unadjusted level $\alpha = 0.05$



Hunting for Significance Tests for a Single Hypothesis

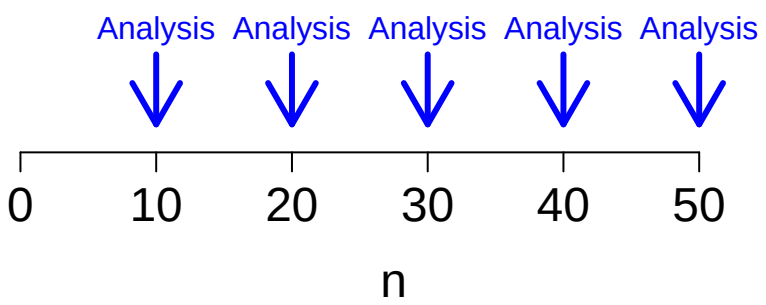
3 Analyses at the unadjusted level $\alpha = 0.05$

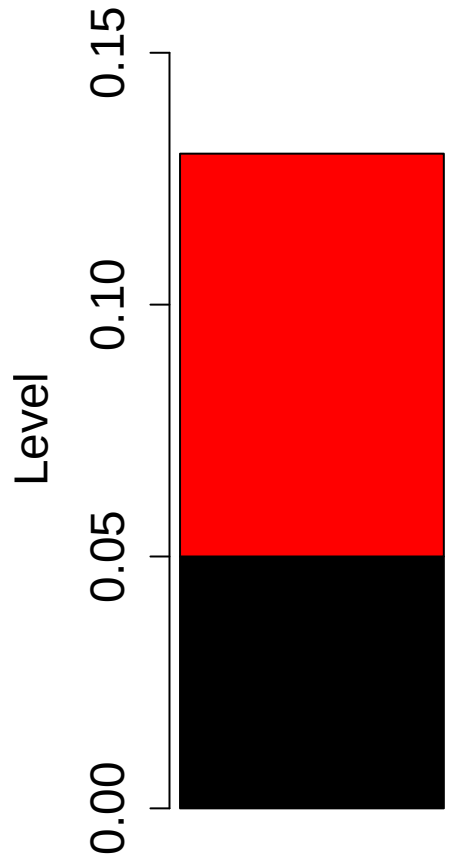


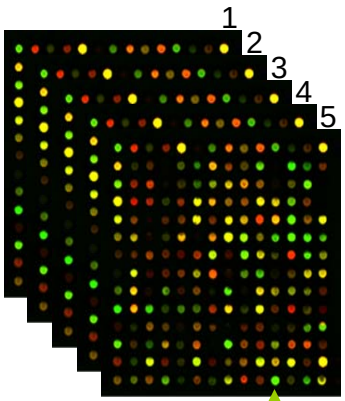


Hunting for Significance Tests for a Single Hypothesis

5 Analyses at the unadjusted level $\alpha = 0.05$



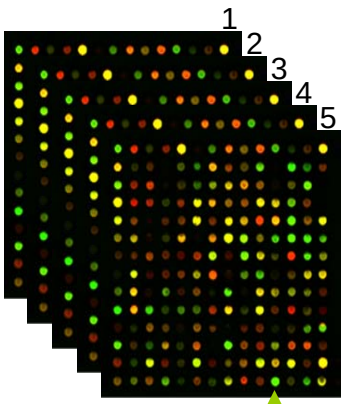




Analysis controlling the FDR

1 spot $\hat{=}$ 1 hypothesis

What is the effect of unadjusted repeated analyses on the FDR?



Analysis controlling the FDR

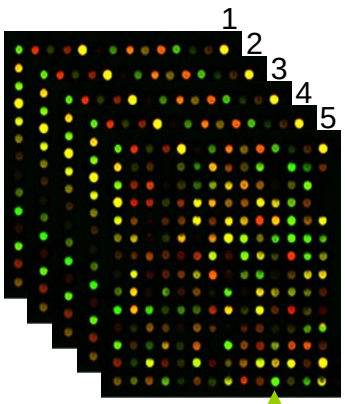
Stop the experiment.

Reject all significant hypotheses.

Accept all others.

1 spot $\hat{=}$ 1 hypothesis

What is the effect of unadjusted repeated analyses on the FDR?



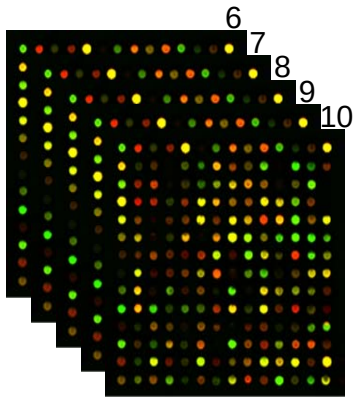
1 spot $\hat{=}$ 1 hypothesis

Analysis controlling the FDR

Stop the experiment.

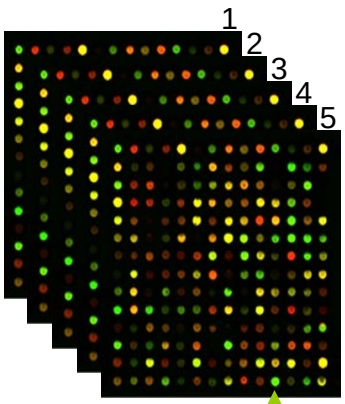
Reject all significant hypotheses.

Accept all others.



Analysis controlling the FDR

What is the effect of unadjusted repeated analyses on the FDR?



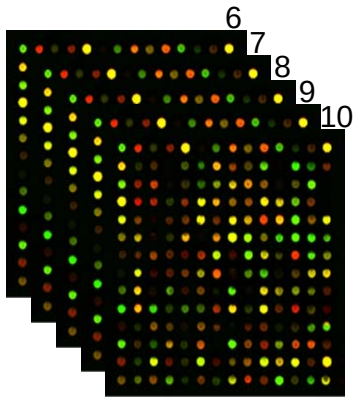
1 spot ≙ 1 hypothesis

Analysis controlling the FDR

Stop the experiment.

Reject all significant hypotheses.

Accept all others.



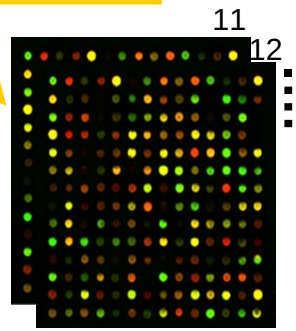
Analysis controlling the FDR

Stop …

Reject …

Accept …

.....



What is the effect of unadjusted repeated analyses on the FDR?

The Setup

- m independent hypotheses tests with

$$H_{0,i}: \ \theta/\sigma = 0 \quad \text{vs.} \quad H_{1,i}: \ \theta/\sigma > 0, \quad i = 1, \ldots, m$$

- m_0 true null hypotheses, m_1 true alternative hypotheses with $\delta/\sigma > 0$

- For each interim analysis: One-sided z-tests

$$p_{i,t} = 1 - \Phi(\sqrt{n_t}\, \bar{x}_{i,t}/\sigma),$$

$\bar{x}_{i,t}$ ... cumulated mean for hypothesis i at stage t
$n_t \leq 1 \ldots 1 \leq t \leq K$ ... cumulated sample size at stage t
$K < \infty$ ... maximum number of stages

Control of the False Discovery Rate (FDR)

Definition

E.G., BENJAMINI AND HOCHBERG (1995)

$$FDR = E\left(\frac{V}{max\{R, 1\}}\right)$$

V . . . number of erroneously rejected hypotheses

R . . . number of rejected hypotheses

Storey's (2002) procedure to control the FDR at level α*:*
Simple estimator searches for a critical boundary γ so that the FDR is controlled at α.

In the long run there is a fraction of not more than α% true null hypotheses among the rejected hypotheses.

False Discovery Rate for repeated analysis

In stage t, $t = 1, \ldots, K$, the FDR is given by

$$FDR_t(\gamma_t) = E\left(\frac{V_t(\gamma_t)}{max\{R_t(\gamma_t), 1\}}\right)$$

- $V_t(\gamma_t)$: number of erroneously rejected hypotheses in stage t at level γ_t
- $R_t(\gamma_t)$: number of rejected hypotheses in stage t at level γ_t

Control of the FDR

In case of $m_0/m = \pi_0 < 1$:

For $m \to \infty$, FDR is controlled asymptotically,
regardless of the stopping stage.

In case of $m_0/m = \pi_0 = 1$ (global H_0):

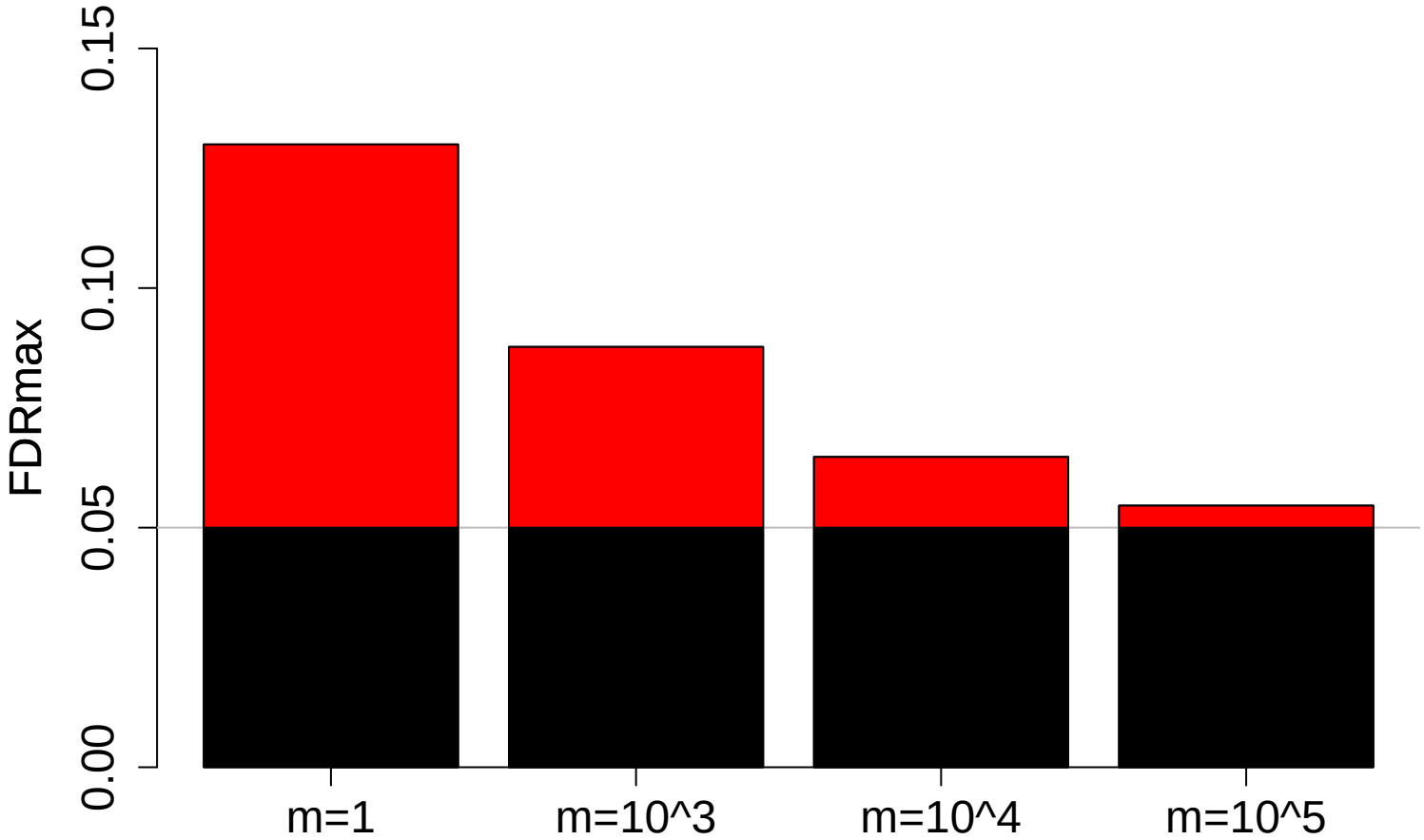
A constraint on the stopping rule has to be imposed:
Stop early only if at least a certain number $s(m)$ of hypotheses can be rejected.

Does asymptotically not lead to early stopping.
Then the FDR is controlled asymptotically.

Simulation Study

$N = 50$, 5 Analyses, $\delta/\sigma = 0.75$, $\pi_0 = 0.95$, s=10

Cheating rule: Stop at the interim analysis with the largest FDR.



Open question

When should the trial stop?

Goal:
Reach a certain power!

Stop as soon as the False Negative Rate is low enough

FNR $\hat{=}$ multiple Type II error ($=$ 1-"Power")

False Negative Rate

E.G. PAWITAN ET AL. (2005), NORRIS & KAHN (2006)

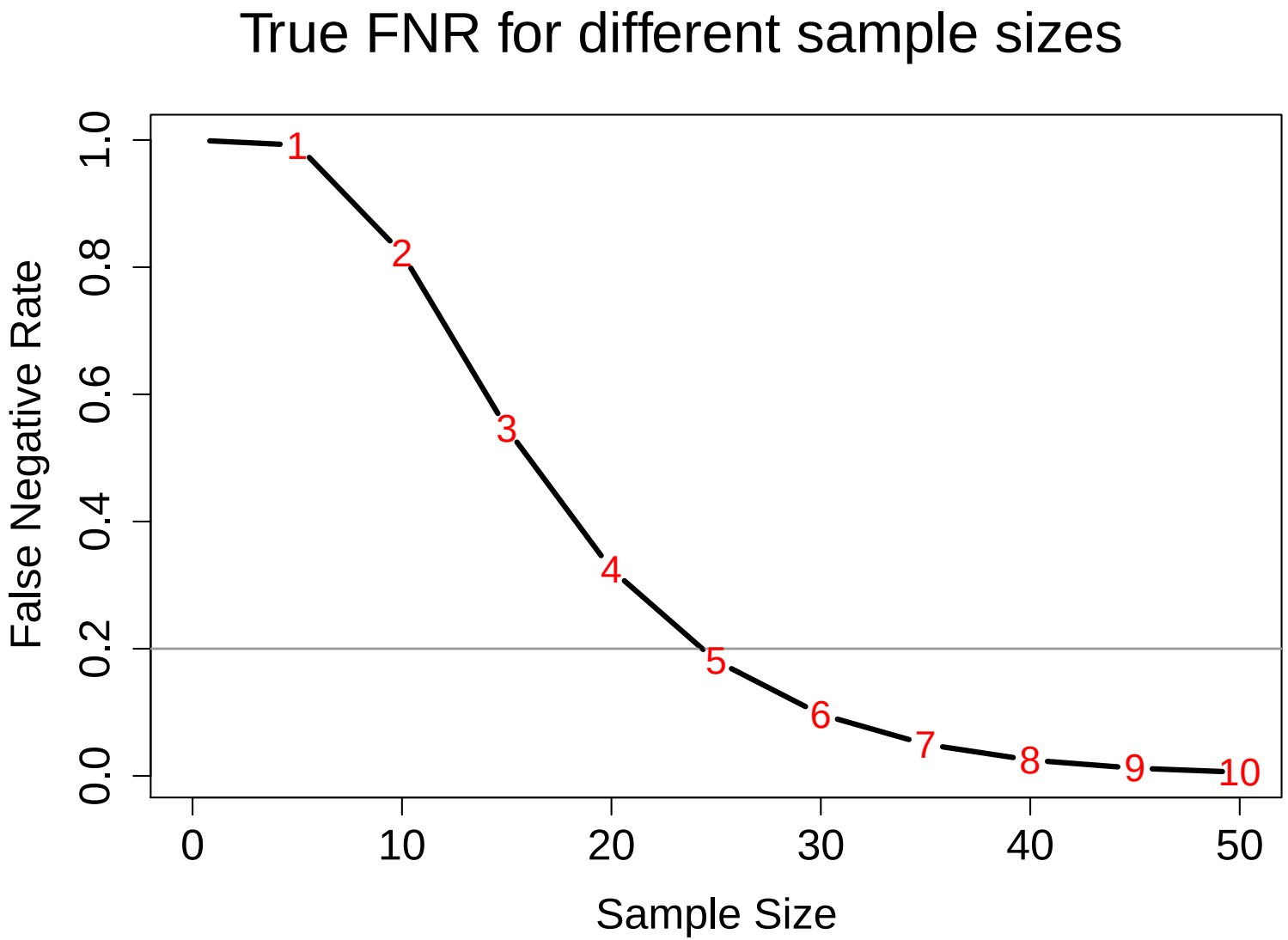
$FNR =$ "Proportion of accepted true alternative hypotheses"

$$\widehat{FNR}_t(\gamma_t) = 1 - \frac{R_t(\gamma_t) - m\gamma_t\hat{\pi}_{0t}}{m(1-\hat{\pi}_{0t})}.$$

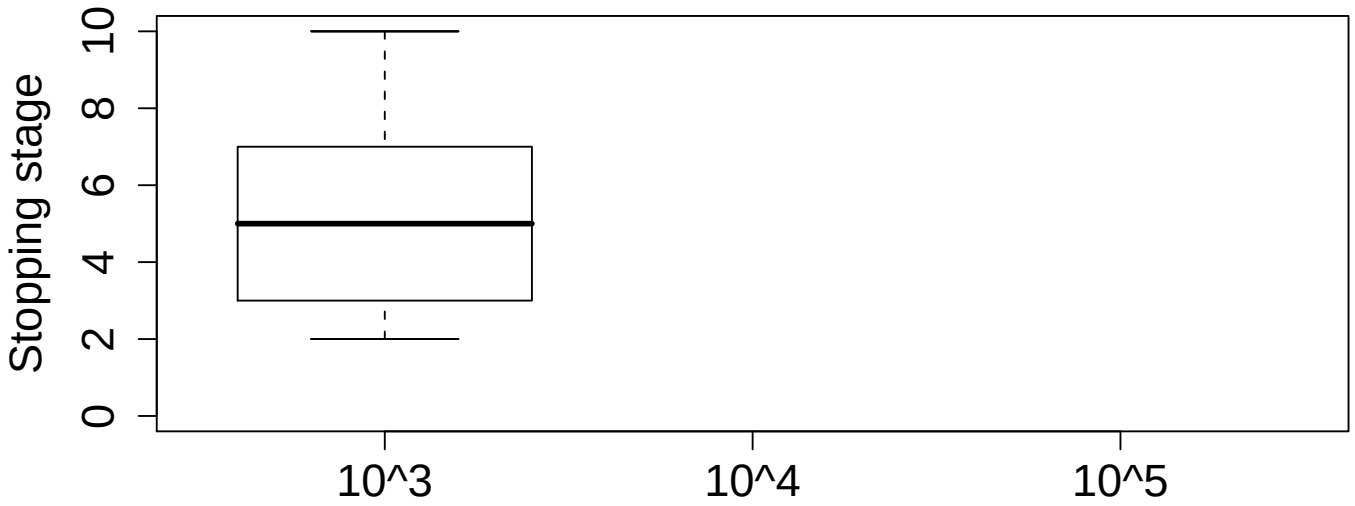
Stopping Rule τ

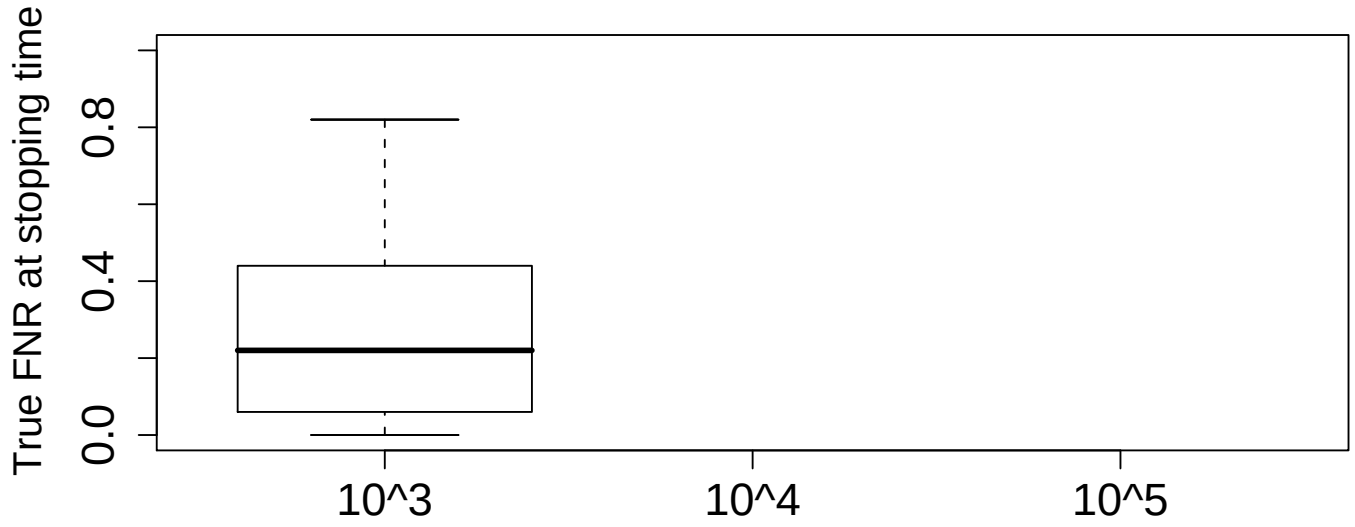
"Stop early, as soon as $\widehat{FNR}_t < \beta$ and at least s null hypotheses are rejected."

Example: $\pi_0 = 0.95$, $\theta/\sigma = 0.75$



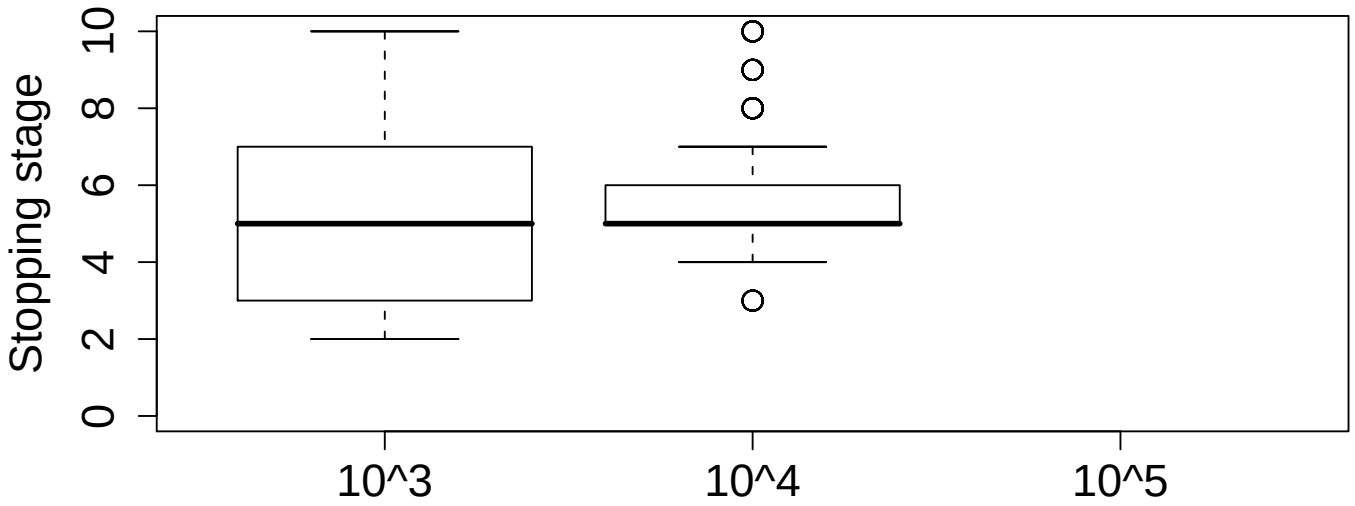
Simulation: $\pi_0 = 0.95$, $\theta/\sigma = 0.75$, $\beta = 0.2$

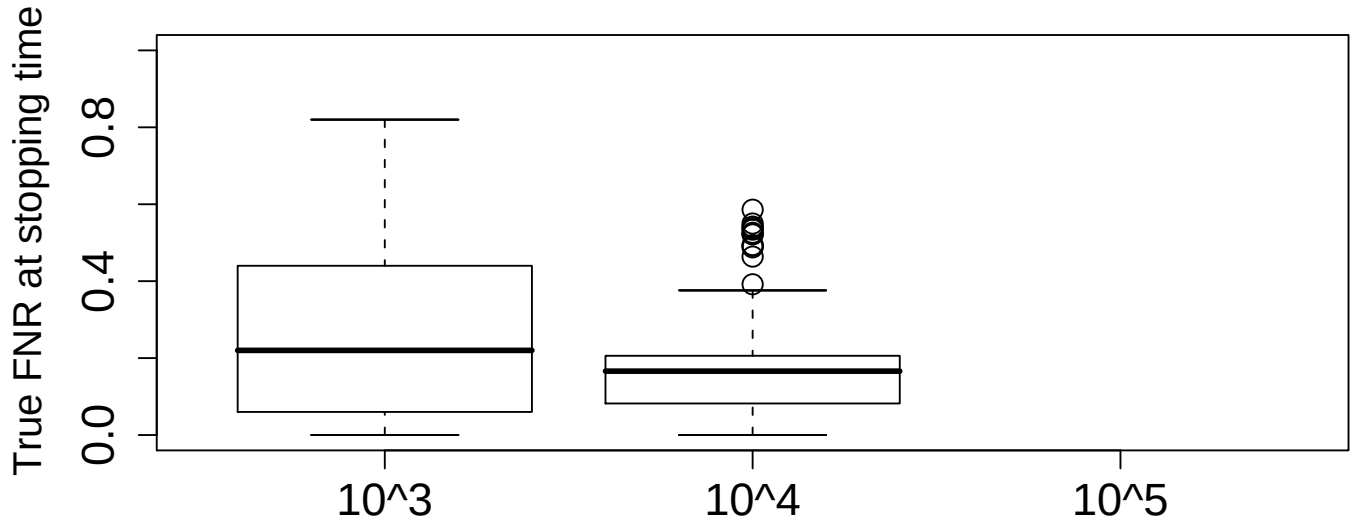




FDR<0.05

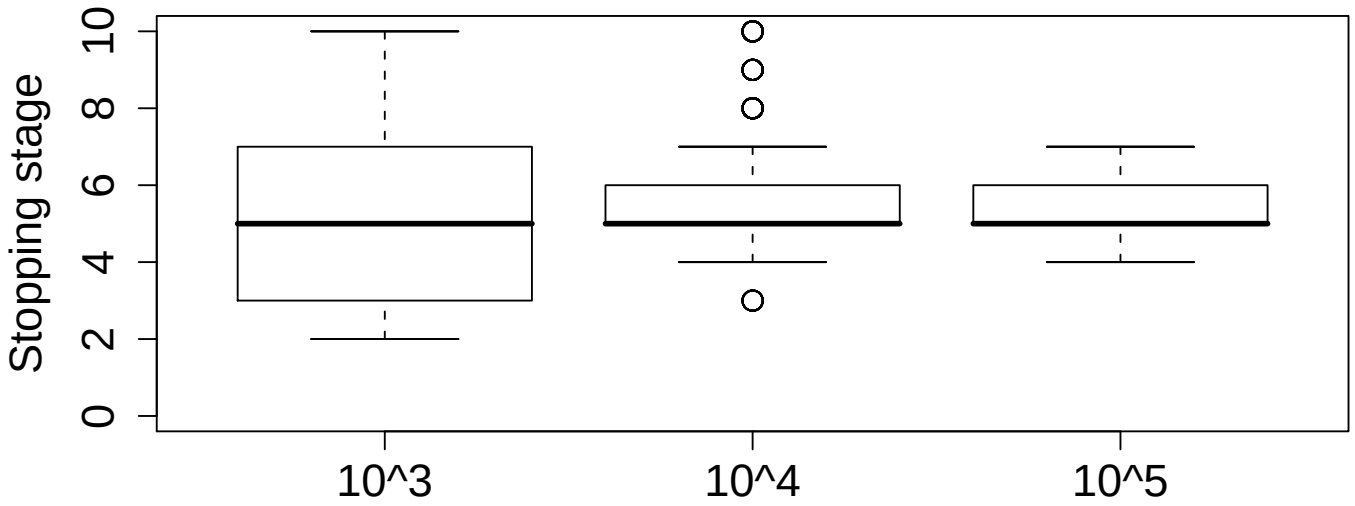
Simulation: $\pi_0 = 0.95$, $\theta/\sigma = 0.75$, $\beta = 0.2$

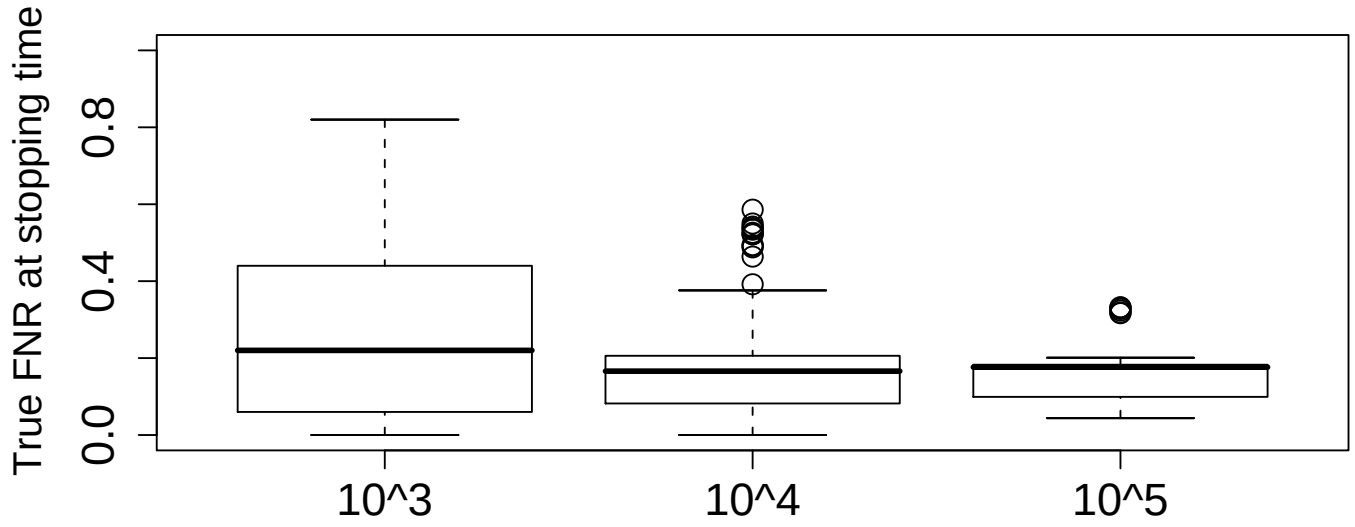




FDR<0.05

Simulation: $\pi_0 = 0.95$, $\theta/\sigma = 0.75$, $\beta = 0.2$





FDR<0.05

Extensions

- The application can be extended to
  - non-normal case
  - dependent test statistics
- Analytical estimates for the worst case FDR
- Stopping early for futility without any rejection at any time.
- Very early interim analyses are not recommendable.

Discussion

"Is it necessary to adjust for the number of looks?"

If the number of hypotheses is very large, multiple analyses hardly inflate the FDR.

Is this the solution to the sequential problem?

There are limitations

- result applies only for large m
- convergence rate depends on π_0 and the alternative
  $\Rightarrow$ can be handled by stopping early only if a certain number of hypotheses can be rejected.
- appropriate stopping rule

- Finner H., Roters, M. (2002) Multiple hypotheses testing and expected type I errors. *The Annals of Statistics* **30**, 220-238.
- Genovese C, Wasserman L. (2004) A stochastic process approach to false discovery control. *Annals of Statistics* **32**, 1035-1061.
- Norris A. W., Kahn C. R. (2006) Analysis of gene expression in pathophysiological states: Balancing false discovery and false negative rates *PNAS* **103**, 649 - 653.
- Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A (2005): False Discovery Rate and Sample Size for Microarray Studies *Bioinformatics* **21**, 3017-3024.
- Storey J. D., Taylor J. E. and Siegmund, D. (2004), Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B* **66** 187-205.