

# Multiple Testing in Sequential and Non-Sequential Clinical Trials

Michael Barón

Department of Mathematics and Statistics  
American University, Washington DC, USA

[baron@american.edu](mailto:baron@american.edu)

with Shyamal De, Laurel MacMillan, Elaine Esther Oruk Opyene, Tian Zhao

Supported by the U.S. National Science Foundation

---

Basel Biometric Society

- Multiple testing and familywise error control
  - Our problem. Bonferroni. Holm. Wald.
- Sequential Bonferroni methods
- Sequential stepwise methods
- Large number of hypotheses
  - Generalized familywise error rates
  - Curtailing
- Further optimization: effect size and optimal error spending
  - Minimax problem and equalizer solution
- Clinical applications and further thoughts
  - Asymptotic optimality
  - Risk optimization
- Truncated designs
- Conclusions

# 1. Introduction

## Goal

Test multiple hypotheses in a sequential or non-sequential experiment and reach a decision on each individual test.

## Clinical Research

- Testing efficacy and safety
- Multiple endpoints
- Multiple populations
- Multiple treatments

## Quality Control

- Acceptance sampling with several criteria of acceptance.

## Multichannel change-point detection

- When a change occurs in one of the parameters, detect it as soon as possible, controlling for the rate of false alarms.

# Multiple testing in literature

## Selection of the best treatment

Drop inferior treatments at interim points

Jennison, Johnstone & Turnbull (1982), Paulson (1964)

## Multi-class classification

A null hypothesis  $H_0 : \theta \in \Theta_0$  is tested against several alternatives,  $H_1 : \theta \in \Theta_1$  vs ... vs  $H_k : \theta \in \Theta_k$ , where  $\theta$  is the common parameter of the observed sequence

Armitage (1950), Baum & Veeravalli (1994), Novikov (2009), Simons (1967)

These are not our problems.

## Inferences about a multivariate parameter

Inference about individual parameters is not required

Ghosh, Mukhopadhyay, and Sen (1997), sec. 6.8 and 7.5  
Sinha (1991)

In particular, testing a composite null hypothesis

$$H_0 : \theta_1 = \dots = \theta_k \text{ vs } H_A : \text{not } H_0$$

e.g. equality of treatment effects

Betensky (1996), Edwards (1987), Edwards & Hsu (1983),  
Hughes (1993), Jennison & Turnbull (2000), chap. 16, O'Brien &  
Fleming (1979), Siegmund (1993), Wilcox (2004), Zacks (2009),  
chap. 8

This is not our problem.

# Our problem - simultaneous testing of multiple hypotheses

## Goal:

Test multiple hypotheses in a sequential or non-sequential experiment and reach a decision for each individual test.

Test  $k$  individual hypotheses about parameters  $\theta^{(1)}, \dots, \theta^{(k)}$  of observed data  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  or  $\{\mathbf{X}_1, \mathbf{X}_2, \dots\}$ ,

$$\begin{array}{l} H_0^{(1)} : \theta_1 = \theta_0^{(1)} \quad \text{vs} \quad H_A^{(1)} : \theta_1 = \theta_1^{(1)}, \\ \dots \\ H_0^{(k)} : \theta_k = \theta_0^{(k)} \quad \text{vs} \quad H_A^{(k)} : \theta_k = \theta_1^{(k)}. \end{array}$$

# Our problem - simultaneous testing of multiple hypotheses

## Goal:

Test multiple hypotheses in a sequential or non-sequential experiment and reach a decision for each individual test.

Test  $k$  individual hypotheses about parameters  $\theta^{(1)}, \dots, \theta^{(k)}$  of observed data  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  or  $\{\mathbf{X}_1, \mathbf{X}_2, \dots\}$ ,

$$\begin{array}{l} H_0^{(1)} : \theta_1 \leq \theta_0^{(1)} \quad \text{vs} \quad H_A^{(1)} : \theta_1 \geq \theta_1^{(1)}, \\ \dots \\ H_0^{(k)} : \theta_k \leq \theta_0^{(k)} \quad \text{vs} \quad H_A^{(k)} : \theta_k \geq \theta_1^{(k)}. \end{array}$$

# Simultaneous testing of multiple hypotheses

- **Test**  $d$  individual hypotheses about parameters  $\theta^{(1)}, \dots, \theta^{(k)}$  of observed data  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  or  $\{\mathbf{X}_1, \mathbf{X}_2, \dots\}$ ,

$$\begin{array}{l} H_0^{(1)} : \theta_1 = \theta_0^{(1)} \quad \text{vs} \quad H_A^{(1)} : \theta_1 = \theta_1^{(1)}, \\ \dots \\ H_0^{(k)} : \theta_k = \theta_0^{(k)} \quad \text{vs} \quad H_A^{(k)} : \theta_k = \theta_1^{(k)}. \end{array}$$

- **Control**

Familywise Type I error rate =  $FWER_I$

$$= \max \mathbf{P} \{ \text{reject at least one true null hypothesis} \} \leq \alpha$$

Familywise Type II error rate =  $FWER_{II}$

$$= \max \mathbf{P} \{ \text{accept at least one false null hypothesis} \} \leq \beta$$

- **Minimize** the sample size  $n$ , expected sample size  $E(T)$ , expected cost, or risk under these conditions.

# Simultaneous testing of multiple hypotheses

- Control in the strong sense

$$FWER_I = \max_{\substack{(\theta_1, \dots, \theta_k) \\ \exists j : H_0^{(j)} \text{ is true}}} \mathbf{P} \left\{ \bigcup_{\substack{j=1, \dots, k \\ H_0^{(j)} \text{ is true}}} \text{reject } H_0^{(j)} \mid \theta_1, \dots, \theta_k \right\} \leq \alpha$$

$$FWER_{II} = \max_{\substack{(\theta_1, \dots, \theta_k) \\ \exists j : H_0^{(j)} \text{ is false}}} \mathbf{P} \left\{ \bigcup_{\substack{j=1, \dots, k \\ H_0^{(j)} \text{ is false}}} \text{accept } H_0^{(j)} \mid \theta_1, \dots, \theta_k \right\} \leq \beta$$

- Minimize** the sample size  $n$ , expected sample size  $E(T)$ , expected cost, or risk under these conditions.

# Simultaneous testing of multiple hypotheses

- Control in the strong sense

$$FWER_I = \max_{\substack{(\theta_1, \dots, \theta_k) \\ \exists j : H_0^{(j)} \text{ is true}}} \mathbf{P} \left\{ \bigcup_{\substack{j=1, \dots, k \\ H_0^{(j)} \text{ is true}}} \text{reject } H_0^{(j)} \mid \theta_1, \dots, \theta_k \right\} \leq \alpha$$

$$FWER_{II} = \max_{\substack{(\theta_1, \dots, \theta_k) \\ \exists j : H_0^{(j)} \text{ is false}}} \mathbf{P} \left\{ \bigcup_{\substack{j=1, \dots, k \\ H_0^{(j)} \text{ is false}}} \text{accept } H_0^{(j)} \mid \theta_1, \dots, \theta_k \right\} \leq \beta$$

## Related problems

Controlling FDR, FNR, FDP - [Bartroff \(2018\)](#), [Bartroff & Song \(2020\)](#)  
[Chen & Arias-Castro \(2021\)](#)

Bayesian formulation - [Kachiashvili \(2014, 2015\)](#)

# Bonferroni approach



Conduct each test at  $\alpha_j = \alpha/d$  and  $\beta_j = \beta/d$ .

Stop at

$$T = \min \{n : \text{all tests reached decision} \}$$

Error rates are controlled by the Bonferroni (Boole) inequality

$$\text{FWER-I} = \mathbf{P} \left\{ \bigcup_j \text{Type I error on } H_0^{(j)} \right\} \leq \sum_j \mathbf{P} \{ \alpha_j \} = \alpha$$

$$\text{FWER-II} = \mathbf{P} \left\{ \bigcup_j \text{Type II error on } H_0^{(j)} \right\} \leq \sum_j \mathbf{P} \{ \beta_j \} = \beta$$

*The inequality is not sharp, there is room for improvement.*

Pocock, Geller, Tsiatis (1987)

## Holm-Bonferroni approach

Based on p-values  $p_1, \dots, p_d$  for  $H_0^{(1)}, \dots, H_0^{(d)}$ .

Order:

$p_{(1)}$  ( most significant )  $\leq p_{(2)} \leq \dots \leq p_{(d)}$  ( least significant )

Compare  $\left\{ \begin{array}{lll} p_{(1)} & \text{against} & \alpha_1 = \alpha/d, \\ p_{(2)} & \text{against} & \alpha_2 = \alpha/(d-1), \\ \dots & & \\ p_{(d)} & \text{against} & \alpha_d = \alpha \end{array} \right.$

If  $p_{(j)} < \alpha_j$  then reject  $H_0^{(j)}$  and continue.

When  $p_{(j)} \geq \alpha_j$ , stop and accept all  $H_0^{(k)}$ ,  $k \geq j$ .

Comparing with Bonferroni -

$\alpha$ -levels are higher but FWER-I is still controlled!



Holm (1979)

# Non-sequential developments and generalizations

In a row of ordered p-values, start with some p-value, proceed in either direction.

- Holm and Hommel step-down methods controlling FWER-I  
Hommel (1988)
- Benjamini-Hochberg step-up and Guo-Sarkar method controlling false discovery rate –  
Benjamini & Hochberg (1995), Sarkar (2007)
- Fallback procedure: recycle  $\alpha$  – Wiens & Dmitrienko (2005)
- Use dependence – Shaffer (1986)
- Use independence – Sidak (1967)
- Use normality – Pocock, Geller, Tsiatis (1987), Tang, Geller, Pocock (1993)
- Fixed-sequence – Maurer et al (1995)
- For the modern overview, see Dmitrienko, Tamhane, and Bretz (2010)

# Wald's SPRT



Test  $H_0^{(j)}$  vs  $H_A^{(j)}$  by *Wald's SPRT*

with stopping boundaries

$$a_j = \ln \left( \frac{1 - \beta_j}{\alpha_j} \right), \quad b_j = \ln \left( \frac{\beta_j}{1 - \alpha_j} \right)$$

for the LLR statistic  $\Lambda_n^{(j)} = \ln \frac{f_j(X_1^{(j)}, \dots, X_n^{(j)} | \theta_1^{(j)})}{f_j(X_1^{(j)}, \dots, X_n^{(j)} | \theta_0^{(j)})}$ .

It controls **both errors**

$$\begin{cases} \mathbf{P} \{ \text{Type I error} \} & \leq \alpha_j \\ \mathbf{P} \{ \text{Type II error} \} & \leq \beta_j \end{cases}$$

for each  $j$



# Remark on Wald's approximation

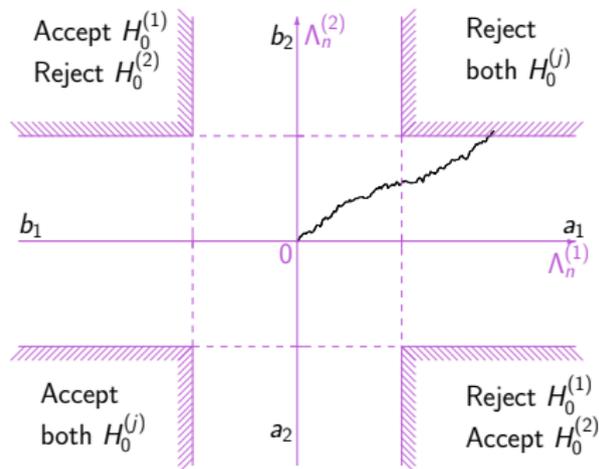
Wald's stopping boundaries are based on the Wald's approximation,  $\Lambda_{T_j}^{(j)} \approx a_j$  or  $\Lambda_{T_j}^{(j)} \approx b_j$  at stopping time  $T_j = \min \left\{ n : \Lambda_n^{(j)} \notin (b_j, a_j) \right\}$ .



Then

$$P \left\{ \Lambda_{T_j}^{(j)} \geq a_j \mid H_0^{(j)} \right\} \approx \alpha_j$$
$$P \left\{ \Lambda_{T_j}^{(j)} \leq b_j \mid H_A^{(j)} \right\} \approx \beta_j$$

This may be inaccurate in multiple testing.

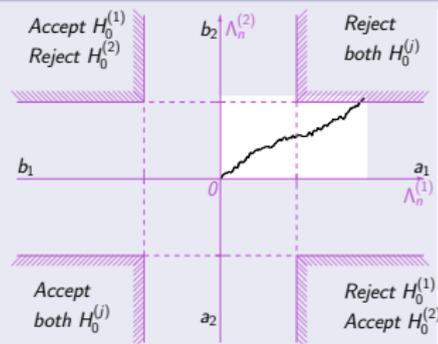


# Remark on Wald's approximation

For this stopping rule, Wald's approximation is inaccurate:  $\Lambda_{T_{int}}^{(1)} \not\approx a_1, b_1$

## Lemma

For  $\forall$  proper stopping time  $T$   
w.r.t. a vector sequence  $(\mathbf{X}_1, \mathbf{X}_2, \dots)$ ,  
such that  $\mathbf{P} \{ \Lambda_T \in (b, a) \mid \mathcal{T} \} = 0$   
for some  $b < 0 < a$  and any combination  
of true null hypotheses  $\mathcal{T}$  and any test  
that rejects  $H_0$  iff  $\Lambda_T \geq a$ ,



$$\mathbf{P} \{ \text{Type I error on } H_0 \} \leq \mathbf{P} \{ \Lambda_T \geq a \mid \theta_0 \} \leq e^{-a},$$

$$\mathbf{P} \{ \text{Type II error on } H_0 \} \leq \mathbf{P} \{ \Lambda_T \leq b \mid \theta_1 \} \leq e^b.$$

So, a sequential Bonferroni test with stopping boundaries  
 $a_j = -\ln(\alpha_j/d)$ ,  $b_j = \ln(\beta_j/d)$  controls FWER-I and FWER-II in  
the strong sense and rigorously, without an approximation.

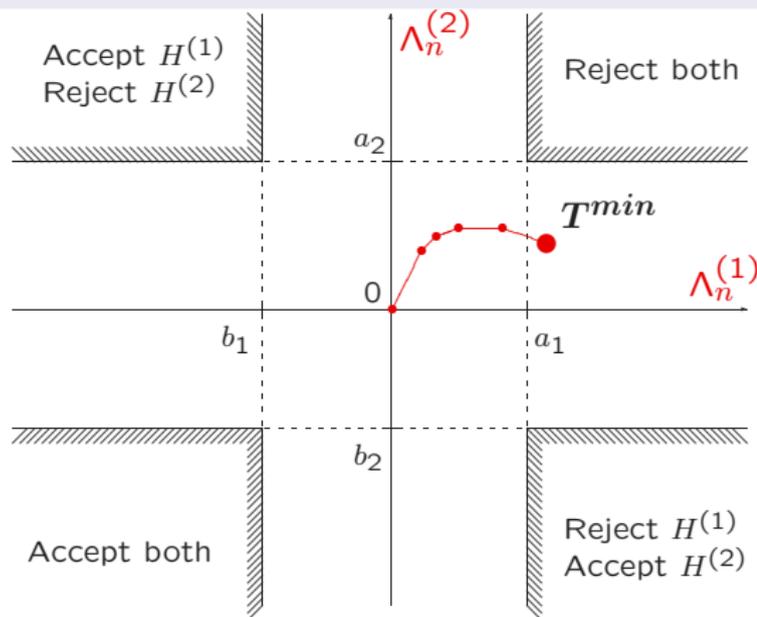
## 2. Sequential Bonferroni approaches

### “Tmin” rule

At the first acceptance or rejection, stop.

Conservatively accept all the remaining null hypotheses.

Jennison & Turnbull (2000), chap. 15



Bonferroni  
adjustment

$$\alpha_j = \alpha/d$$

$$\beta_j = \beta/d$$

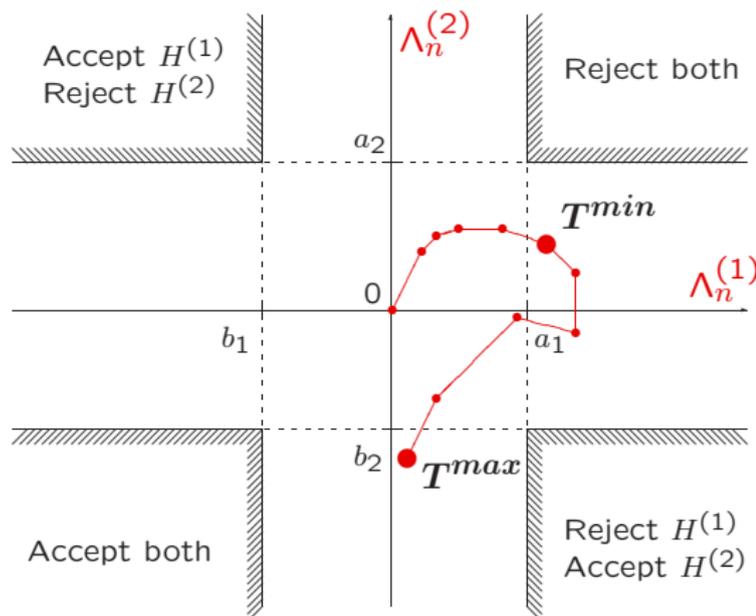
Controls FWER-I  
but not FWER-II.  
Low power

# Sequential Bonferroni approaches

## Sobel-Wald “Incomplete T<sub>max</sub>” rule

Continue beyond T<sub>min</sub>, until  $T^{max} = \max(T_1, \dots, T_d)$ ,

Stop testing  $H_0^{(j)}$  at  $T_j = \inf \left\{ n : \Lambda_n^{(j)} \notin (b_j, a_j) \right\}$ .



Sobel & Wald (1949)  
Bartroff & Lai (2010)

Controls **both**  
FWER-I, FWER-II.

Ignores data  
 $X_n^{(j)}$  for  $n > T_j$ .

# Sequential Bonferroni approaches

Ignore or keep resolved sequences?

What to do with the  $j$ -th sequence  $X_n^{(j)}$  when  $H_0^{(j)}$  has already resolved, but some  $H_0^{(k)}$  hasn't?

Cost per sampling unit (patient)

⇒ keep all data

⇒ this is our case

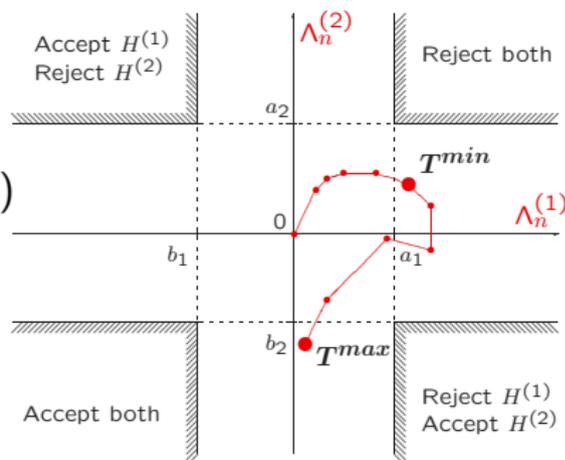
Cost per measurement (treatment)

⇒ terminate those resolved

⇒ this may also happen

Bartroff and Song (2014)  
and references therein

Both costs ⇒ ?



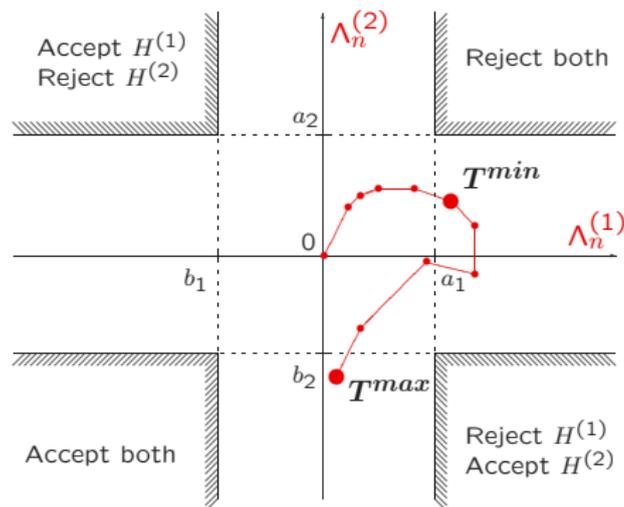
# Improvement: “Complete Tmax” rule

Use all the data from all the sampled units. Extra cost  $\approx 0!$

Satisfy the sufficiency principle.

Make decisions based on  $\left\{ \Lambda_{T^{max}}^{(1)}, \dots, \Lambda_{T^{max}}^{(d)} \right\}$ .

But... what to decide for  $\Lambda_{T^{max}}^{(j)} \in (b_j, a_j)$  ending inside the continue-sampling region?



*We could have rejected  $H_0^{(1)}$  at time  $T^{min}$  but sampling continued, and by the time  $T^{max}$ ,  $\Lambda_n^{(1)}$  returned to the sampling region.*

# Improvement: “Complete T<sub>max</sub>” rule

## Rao-Blackwellization

At time  $T_j$ , there was a decision on  $H_0^{(j)}$ .

At time  $T^{\max}$ , compute  $p_j^* = \mathbf{P} \left\{ \Lambda_{T_j}^{(j)} \geq a_j \mid \mathbf{\Lambda}_{T^{\max}}, T^{\max} \right\}$

Randomized decision rule  $\mathcal{D}_j^*$  rejects  $H_0^{(j)}$  at time  $T^{\max}$  with probability  $p_j^*$ .

## Theorem

*Rule  $\mathcal{D}^*$  strongly controls Type I and Type II familywise error rates at levels  $\alpha$  and  $\beta$ .*

*Food for thought:*

$\mathcal{D}_j^*$  may reject  $H_0^{(j)}$  when  $\Lambda_{T^{\max}}^{(j)} \leq b_j$  or accept when  $\Lambda_{T^{\max}}^{(j)} \geq a_j$ .



### 3. Sequential Holm-type stepwise testing

Sequential **step-down** (sequentially-rejective) procedure

Stopping boundaries  $a_j = -\log \alpha_j = -\log(\alpha / (d - j + 1))$  and arbitrary  $b_j < 0$  for  $j = 1, \dots, d$

Having observed  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , order log-likelihood ratio statistics,

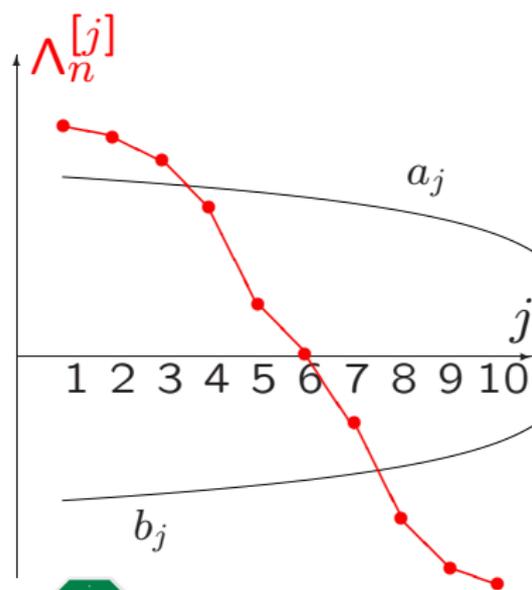
$$\Lambda_n^{[1]} \geq \Lambda_n^{[2]} \geq \dots \geq \Lambda_n^{[d]}.$$

Step 1. If  $\Lambda_n^{[1]} \leq b_1$  then accept all  $H_0^{[1]} \dots, H_0^{[d]}$   
If  $\Lambda_n^{[1]} \in (b_1, a_1)$  then continue; sample  $\mathbf{X}_{n+1}$   
If  $\Lambda_n^{[1]} \geq a_1$  then reject  $H_0^{[1]}$  and go to  $\Lambda_n^{[2]}$

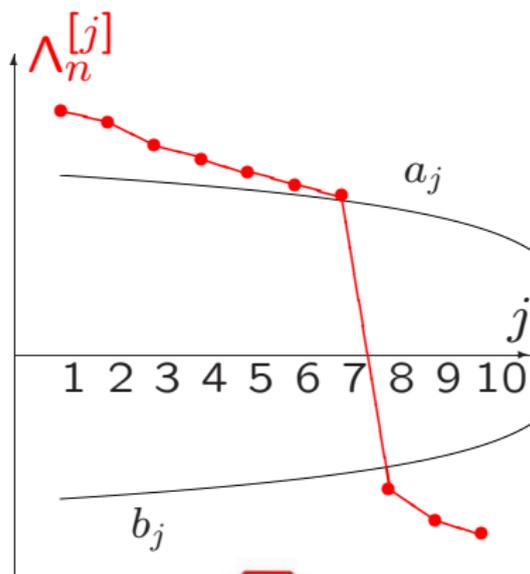
Step 2. If  $\Lambda_n^{[2]} \leq b_2$  then accept all  $H_0^{[2]} \dots, H_0^{[d]}$   
If  $\Lambda_n^{[2]} \in (b_2, a_2)$  then continue; sample  $\mathbf{X}_{n+1}$   
If  $\Lambda_n^{[2]} \geq a_2$  then reject  $H_0^{[2]}$  and go to  $\Lambda_n^{[3]}$

etc.

# Sequential **step-down** (sequentially-rejective) procedure



Continue sampling



$$\text{Stopping rule } T_1 = \inf \left\{ n : \bigcap_{j=1}^d \Lambda_n^{[j]} \notin (b_j, a_j) \right\}$$

# Sequential **step-down** (sequentially-rejective) procedure

## Theorem

- Stopping rule  $T_1 = \inf \left\{ n : \bigcap_{j=1}^d \Lambda_n^{[j]} \notin (b_j, a_j) \right\}$  is proper.
- Sequentially-rejective scheme strongly controls FWER-I.

That is, for any set  $\mathcal{T}$  of true hypotheses,

$$P \{ T_1 < \infty \mid \mathcal{T} \} = 1$$

$$P \{ \text{at least one Type I error} \mid \mathcal{T} \} \leq \alpha.$$

Combination of the stepwise idea for efficient multiple testing with Wald's sequential probability ratio testing of individual hypotheses.

Tighter stopping boundaries  $\Rightarrow$  reduced expected sample size  $E(T)$  under any combination of true and false hypotheses.

# Sequential stepwise testing

## Sequential **step-up** (sequentially-acceptive) procedure

Stopping boundaries  $b_j = \log \beta_j = \log(\beta/(d - j + 1))$  and **arbitrary**  $a_j > 0$  for  $j = 1, \dots, d$

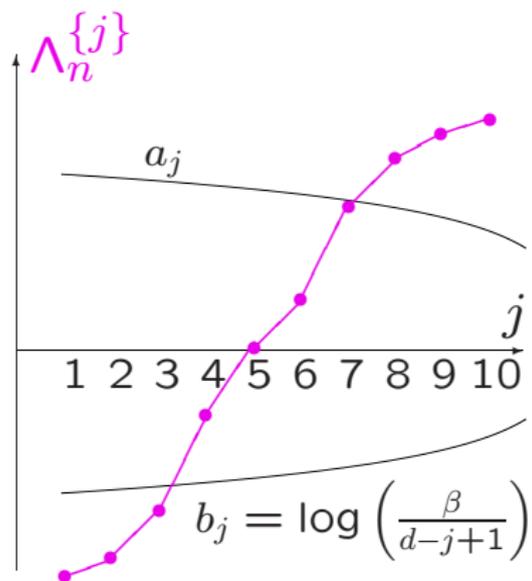
Having observed  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , order log-likelihood ratio statistics,

$$\Lambda_n^{\{1\}} \leq \Lambda_n^{\{2\}} \leq \dots \leq \Lambda_n^{\{d\}},$$

- Step 1. If  $\Lambda_n^{\{1\}} \geq a_1$  then reject all  $H_0^{\{1\}} \dots, H_0^{\{d\}}$   
If  $\Lambda_n^{\{1\}} \in (b_1, a_1)$  then continue; sample  $\mathbf{X}_{n+1}$   
If  $\Lambda_n^{\{1\}} \leq b_1$  then accept  $H_0^{\{1\}}$ , go to  $\Lambda_n^{\{2\}}$
- Step 2. If  $\Lambda_n^{\{2\}} \geq a_2$  then reject all  $H_0^{\{2\}} \dots, H_0^{\{d\}}$   
If  $\Lambda_n^{\{2\}} \in (b_2, a_2)$  then continue; sample  $\mathbf{X}_{n+1}$   
If  $\Lambda_n^{\{2\}} \leq b_2$  then accept  $H_0^{\{2\}}$ , go to  $\Lambda_n^{\{3\}}$

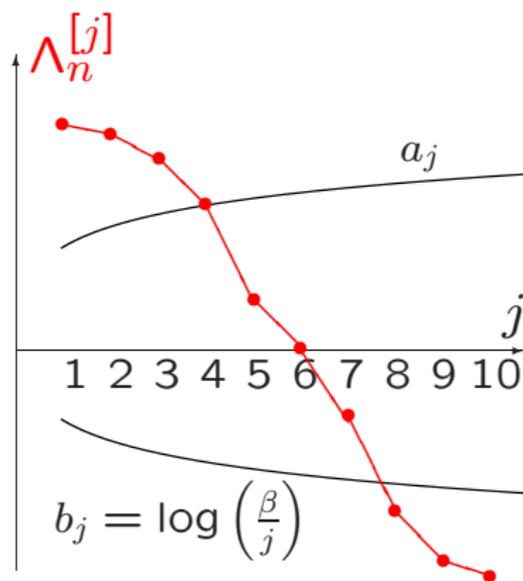
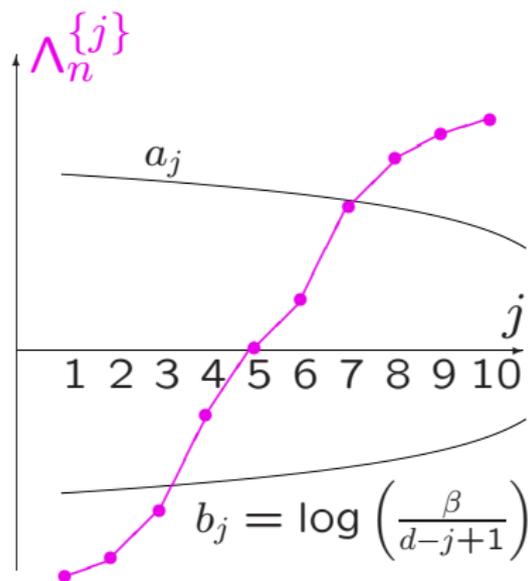
etc.

# Sequential **step-up** (sequentially-acceptive) procedure



$$\text{Stopping rule } T_2 = \inf \left\{ n : \bigcap_{j=1}^d \Lambda_n^{\{j\}} \notin (b_j, a_j) \right\}$$

# Sequential **step-up** (sequentially-acceptive) procedure



Stopping rule  $T_2 = \inf \left\{ n : \bigcap_{j=1}^d \Lambda_n^{\{j\}} \notin (b_j, a_j) \right\}$

# Sequential **step-up** (sequentially-acceptive) procedure

## Theorem

- Stopping rule  $T_2 = \inf \left\{ n : \bigcap_{j=1}^d \Lambda_n^{\{j\}} \notin (b_j, a_j) \right\}$  is proper.
- Sequentially-acceptive scheme strongly controls FWER-II.

That is, for set  $\mathcal{T}$  of true hypotheses,

$$P_{\mathcal{T}} \{ T_2 < \infty \} = 1$$

and

$$P_{\mathcal{T}} \{ \text{at least one Type II error} \} \leq \beta$$

# Sequential stepwise testing: Intersection Scheme

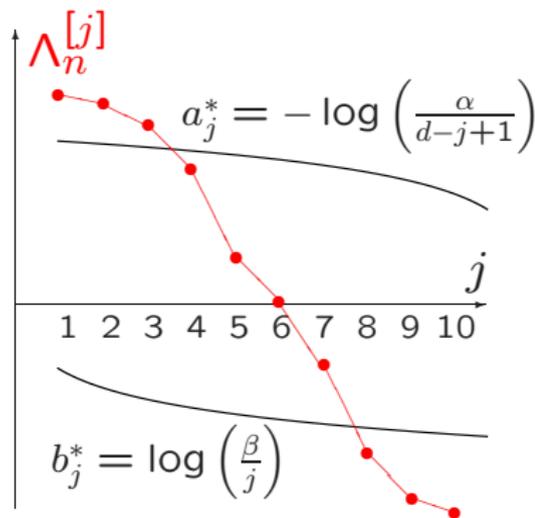
To control both FWER-I and FWER-II

Combine the *sequentially-rejective* and *sequentially-acceptive* schemes. Stop when each scheme stops.

Stopping rule:

$$T^* = \inf \left\{ n : \bigcap_{j=1}^d \Lambda_n^{[j]} \notin (b_j^*, a_j^*) \right\}$$

Tighter boundaries than Bonferroni's.  
Hence, smaller  $\mathbf{E}(T)$ .



# The Intersection Scheme

## Intersection

- Intersection of stopping conditions.  
This scheme stops when both step-up and step-down schemes stop.
- Intersection of acceptance regions and rejection regions.

$$\text{Acceptance Region}(T^*) = A.R.(T_1) \cap A.R.(T_2)$$

$$\text{Rejection Region}(T^*) = R.R.(T_1) \cap R.R.(T_2)$$

Being a special case of step-down and step-up schemes at the same time, the Intersection Scheme controls **both** FWER-I and FWER-II.

## Theorem

- *The stopping rule*

$$T^* = \inf \left\{ n : \bigcap_{j=1}^d \Lambda_n^{[j]} \notin \left( \log \frac{\beta}{j}, -\log \frac{\alpha}{d-j+1} \right) \right\} \text{ is}$$

*proper*

- *Strongly controls both familywise Type I and Type II error rates.*

*That is, for any set  $\mathcal{T}$  of true hypotheses,  $\mathbf{P}_{\mathcal{T}} \{T^* < \infty\} = 1$ ;*

$$\mathbf{P}_{\mathcal{T}} \{ \text{at least one Type I error} \} \leq \alpha,$$

$$\mathbf{P}_{\mathcal{T}} \{ \text{at least one Type II error} \} \leq \beta.$$

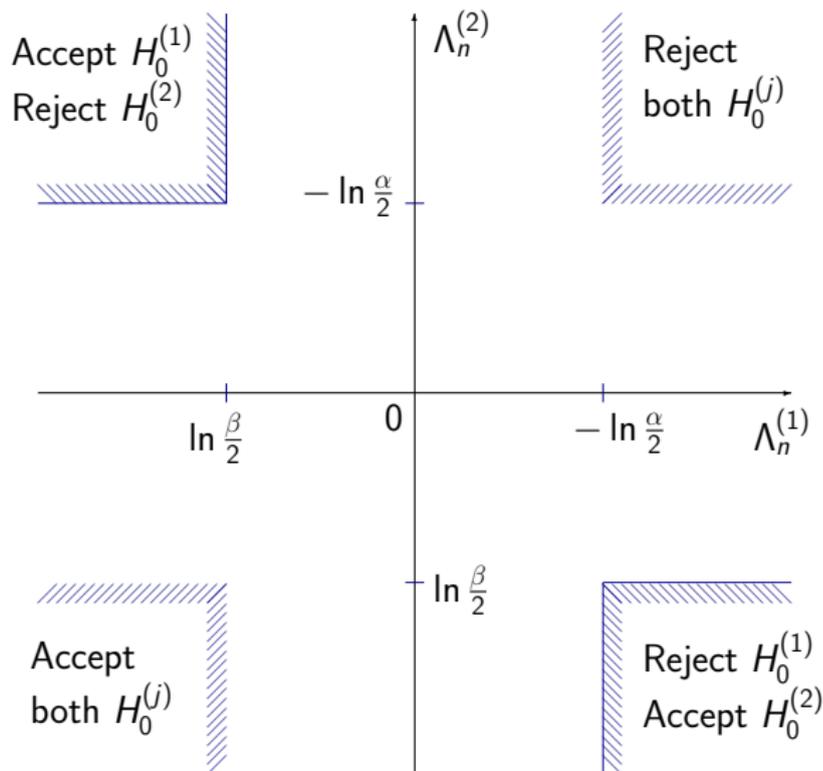
*De and Baron (2012b)*

*Further,  $T^*$  is asymptotically optimal.*

*Song and Fellouris (2017)*

# Comparison of savings

Sequential  
Bonferroni

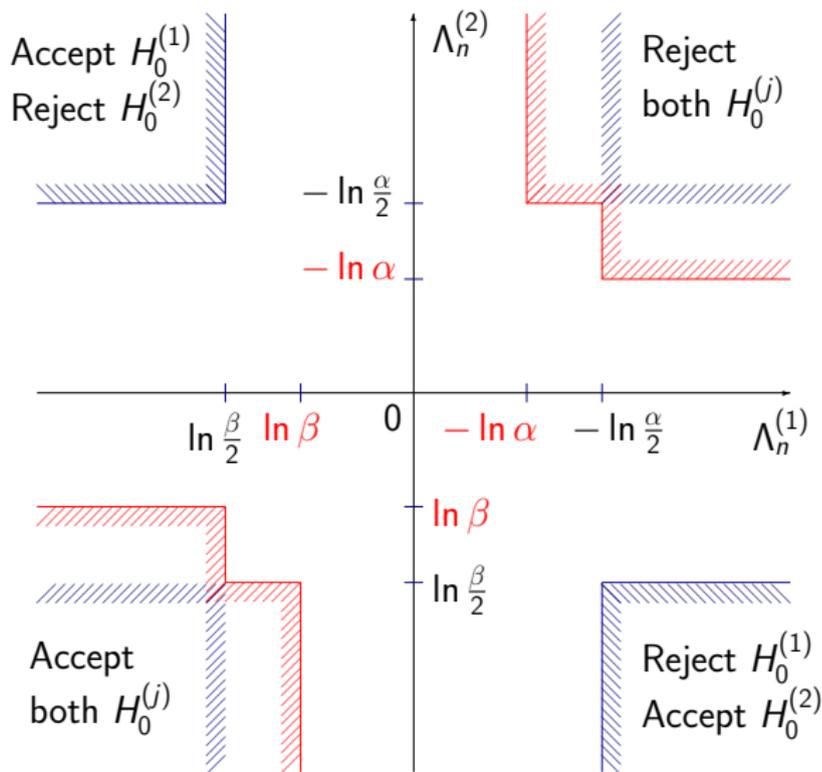


# Comparison of savings

Sequential  
Bonferroni

vs

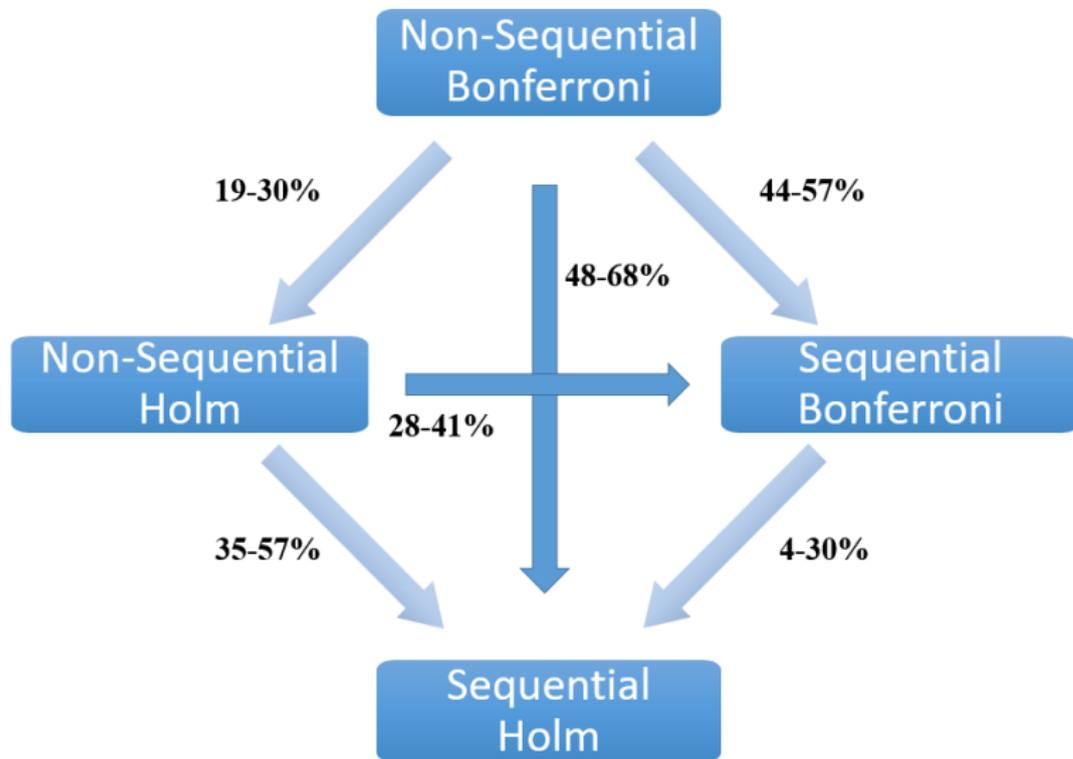
Sequential  
Holm



# Comparison of savings

	$H_a^1$	$H_a^2$	<b>Non-Seq Bonf</b>	<b>Non-Seq Holm</b>	<b>Seq Bonf</b>	<b>Seq Holm</b>
<b>D=2</b>	0.25	0.26	208	170	155	140
	0.25	0.3	208	169	139	126
	0.25	0.35	208	168	127	118
	0.25	0.4	208	168	122	113
	0.25	0.45	208	168	118	109
	0.25	0.5	208	169	116	108
	0.25	0.8	208	169	114	106
	0.25	1	208	168	113	106

# Comparison of savings



## 4. Large number of tests. Generalized error rates.

Situations with 100s-1000s of tests: biology, genomics, computer science, communications, and many other areas

Devlin & Roeder (1999); Siegmund, Yakir & Zhang (2011); Dmitrienko, Tamhane & Bretz (2010); Efron (2010)

Controlling FWERs is too stringent. Needs huge samples.  
Instead, control generalized error rates:

$$GFWER_I(k) = \max_{\substack{(\theta_1, \dots, \theta_d) \\ \exists j : H_0^{(j)} \text{ is true}}} \mathbf{P} \left\{ \sum_{\substack{j=1, \dots, d \\ H_0^{(j)} \text{ is true}}} I \left\{ \text{reject } H_0^{(j)} \right\} \geq \mathbf{k} \right\} \leq \alpha$$
$$GFWER_{II}(m) = \max_{\substack{(\theta_1, \dots, \theta_d) \\ \exists j : H_0^{(j)} \text{ is false}}} \mathbf{P} \left\{ \sum_{\substack{j=1, \dots, d \\ H_0^{(j)} \text{ is false}}} I \left\{ \text{accept } H_0^{(j)} \right\} \geq \mathbf{m} \right\} \leq \beta$$

## Method of Reduced Boundaries

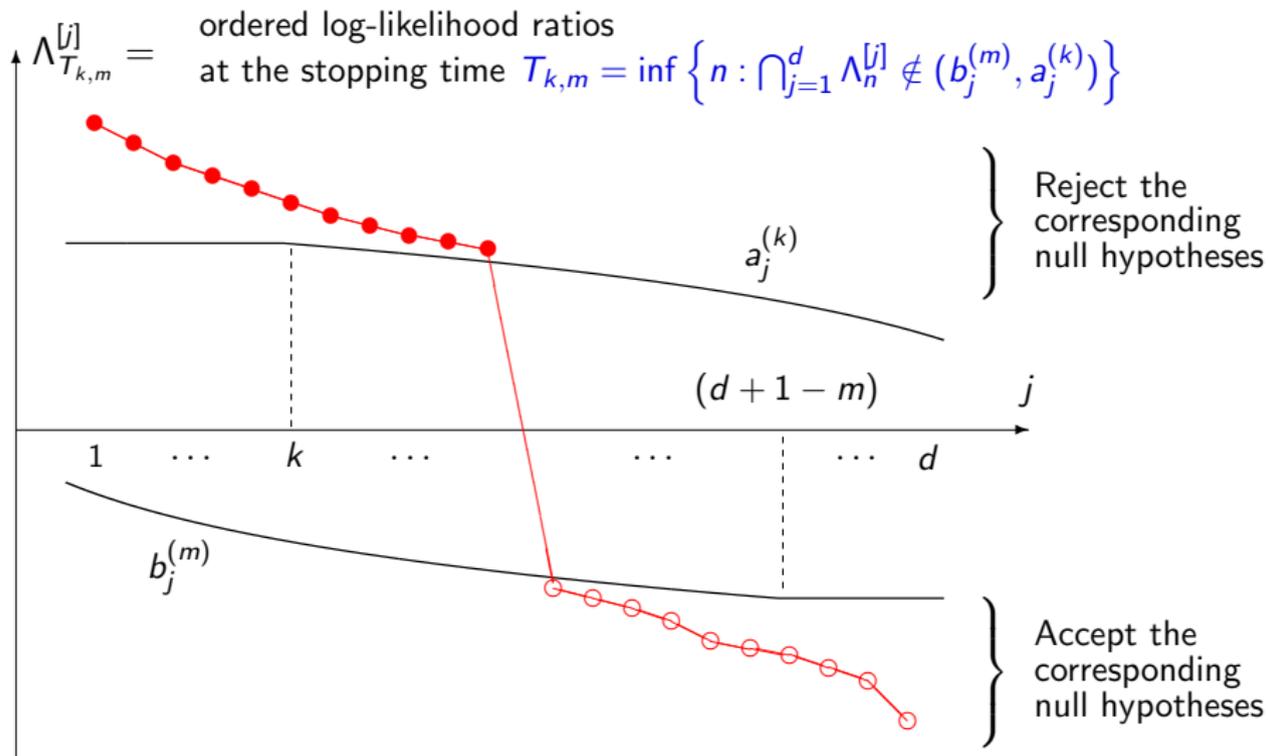
Increase the  $\alpha_j$  and  $\beta_j$  levels  $\Rightarrow$  shrink the continue-sampling region:

$$a_j^{(k)} = \begin{cases} \ln \frac{d+k-j}{\alpha k} & \text{for } j > k, \\ \ln \frac{d}{\alpha k} & \text{for } j \leq k, \end{cases}$$

$$b_j^{(m)} = \begin{cases} \ln \frac{m\beta}{m-1+j} & \text{for } j < d+1-m, \\ \ln \frac{m\beta}{d} & \text{for } j \geq d+1-m, \end{cases}$$

For non-sequential experiments, [Lehmann & Romano \(2005\)](#)

# Method of Reduced Boundaries



## Theorem

*For any  $k$ ,  $m$ ,  $\alpha$ , and  $\beta$ ,*

- *Stopping time  $T_{k,m}$  is proper.*
- *The Reduced Boundary sequential testing procedure controls  $GFWER_I(k)$  and  $GFWER_{II}(m)$  in the strong sense.*

Stopping rule

$$\tau = \tau_{k,m} = \inf \left\{ n : \sum_{j=1}^d I \left( \Lambda_n^{[j]} \in [B_n, A_n] \right) \leq k + m - 2 \right\}$$

where

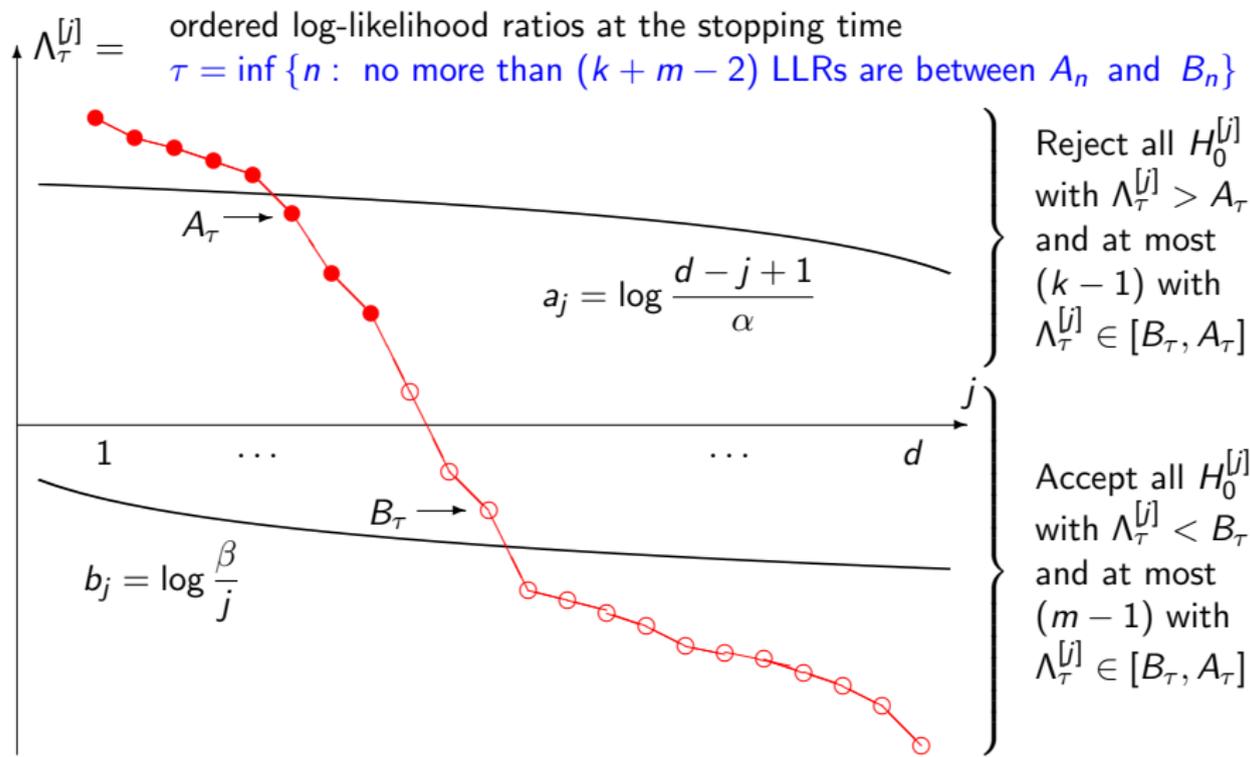
$$A_n = \sup \left\{ \Lambda_n^{[j]} : b_j < \Lambda_n^{[j]} < a_j \right\}$$

$$B_n = \inf \left\{ \Lambda_n^{[j]} : b_j < \Lambda_n^{[j]} < a_j \right\}$$

$$a_j = \log \frac{d - j + 1}{\alpha}$$

$$b_j = \log \frac{\beta}{j}, \quad \text{for } j = 1, \dots, d$$

# Curtailed Method



## Theorem

*For any  $k$ ,  $m$ ,  $\alpha$ , and  $\beta$ ,*

- *Stopping time  $\tau_{k,m}$  is proper.*
- *The Curtailed sequential testing procedure controls  $GFWER_I(k)$  and  $GFWER_{II}(m)$  in the strong sense.*

# Performance evaluation

Generate a sequence  $\mathbf{X}_1, \mathbf{X}_2, \dots \in \mathbf{R}^d$

$X_i^{(j)} \sim \text{Normal}(\theta_j, 1)$  for  $j = 1, \dots, d/2$

$X_i^{(j)} \sim \text{Bernoulli}(p_j)$  for  $j = d/2 + 1, \dots, d$

Test

$H_0^{(j)} : \theta_j = 0$  vs  $H_A^{(j)} : \theta_j = 0.5$  for  $j = 1, \dots, d/2$

$H_0^{(j)} : p_j = 0.5$  vs  $H_A^{(j)} : p_j = 0.75$  for  $j = d/2 + 1, \dots, d$

at the nominal familywise error rates  $\alpha = 0.05$  and  $\beta = 0.10$ .

Let  $H_0^{(1, \dots, d/4)} = \text{true}$ ,  $H_0^{(d/4+1, \dots, d/2)} = \text{false}$ ,  
 $H_0^{(d/2+1, \dots, 3d/4)} = \text{true}$ ,  $H_0^{(3d/4+1, \dots, d)} = \text{false}$ .

## How much is saved due to sequential sampling?

Number of tests	$k, m$	$n$	Lehmann-Romano		Sequential scheme		
			$GFWER_I$ (%)	$GFWER_{II}$ (%)	Scheme	$GFWER_I$ (%)	$GFWER_{II}$ (%)
$d = 12$	2, 2	85	0.47	0.05	Reduced	< 0.01	0.01
		58	0.53	3.33	Curtailed	0.17	0.45
$d = 40$	2, 2	130	0.42	< 0.01	Reduced	< 0.01	< 0.01
		101	0.80	0.14	Curtailed	0.09	0.19
$d = 60$	3, 3	142	0.11	< 0.01	Reduced	< 0.01	< 0.01
		102	0.07	0.05	Curtailed	0.01	0.04
$d = 100$	5, 5	156	< 0.01	< 0.01	Reduced	< 0.01	< 0.01
		105	< 0.01	< 0.01	Curtailed	< 0.01	< 0.01

\*In each case, the Lehmann-Romano procedure uses the same sample size as the expected sample size of each sequential scheme

# Performance evaluation

Adjusted boundaries to yield  $GFWER_I \approx 0.05$  and  $GFWER_{II} \approx 0.10$

No. of tests	$k, m$	$\rho$	Testing scheme	$\widehat{ET}$	$GFWER_I(k)$ (%)	$GFWER_{II}(m)$ (%)
$d = 12$	2	0	Lehmann-Romano	35	5.06	11.47
			Curtailed	28.4	4.92	9.72
		0.9	Lehmann-Romano	42	4.83	9.68
			Reduced	34.5	4.97	9.82
$d = 40$	2	0	Lehmann-Romano	65	5.09	10.9
			Curtailed	58.0	4.80	9.95
		0.9	Lehmann-Romano	62	4.90	9.56
			Curtailed	46.9	4.87	10.0
$d = 60$	3	0	Lehmann-Romano	55	4.03	8.56
			Curtailed	49.7	4.94	10.0
		0.9	Lehmann-Romano	55	4.84	9.5
			Curtailed	41.9	4.69	9.99
$d = 100$	5	0	Lehmann-Romano	48	4.39	9.44
			Curtailed	43.0	4.93	9.71
		0.9	Lehmann-Romano	51	5.01	10.16
			Curtailed	36.7	4.87	10.01

## How much is saved due to weaker constraints (GFWER)?

No. of tests	Testing scheme	$k, m$	$\widehat{ET}$	$GFWER_I(k)$ (%)	$GFWER_{II}(m)$ (%)
$d = 12$	Intersection	1, 1	91.8	0.62	0.91
	Reduced	2, 2	84.5	< 0.01	0.01
	Curtailed	2, 2	57.9	0.17	0.45
$d = 40$	Intersection	1, 1	137.9	0.24	0.29
	Reduced	2, 2	129.7	< 0.01	< 0.01
	Curtailed	2, 2	101.0	0.09	0.19
$d = 60$	Intersection	1, 1	154.2	0.21	0.19
	Reduced	3, 3	141.3	< 0.01	< 0.01
	Curtailed	3, 3	101.7	0.01	0.04
$d = 100$	Intersection	1, 1	175.7	0.09	0.12
	Reduced	5, 5	155.4	< 0.01	< 0.01
	Curtailed	5, 5	104.5	< 0.01	< 0.01

## 5. Further optimization - minimax solutions

### Bonferroni approach

Conduct each test at  $\alpha_j = \alpha/k$  and  $\beta_j = \beta/k$

But tests differ by effect size!



## 5. Further optimization - minimax solutions

### Bonferroni approach

Conduct each test at  $\alpha_j = \alpha/k$  and  $\beta_j = \beta/k$

But tests differ by effect size!



## 5. Further optimization - minimax solutions

### Bonferroni approach

Conduct each test at  $\alpha_j = \alpha/k$  and  $\beta_j = \beta/k$

But tests differ by effect size!



Example: test the following hypotheses at  $\alpha_j = 0.0125$ ,  $\beta_j = 0.025$ , to have  $FWER_I \leq 0.05$ ,  $FWER_{II} \leq 0.10$

- $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$   
based on  $X_1^{(1)}, X_2^{(1)}, \dots \sim \text{Normal}(\theta^{(1)}, 1)$
- $H_0^{(2)} : \theta^{(2)} = 0$  vs  $H_A^{(2)} : \theta^{(2)} = 0.5$   
based on  $X_1^{(2)}, X_2^{(2)}, \dots \sim \text{Normal}(\theta^{(2)}, 0.2)$
- $H_0^{(3)} : \theta^{(3)} = 0.5$  vs  $H_A^{(3)} : \theta^{(3)} = 0.75$   
based on  $X_1^{(3)}, X_2^{(3)}, \dots \sim \text{Bernoulli}(\theta^{(3)})$
- $H_0^{(4)} : \theta^{(4)} = 0.5$  vs  $H_A^{(4)} : \theta^{(4)} = 0.6$   
based on  $X_1^{(4)}, X_2^{(4)}, \dots \sim \text{Bernoulli}(\theta^{(4)})$

## 5. Further optimization

### Bonferroni approach, non-sequentially

Conduct each test at  $\alpha_j = \alpha/k$  and  $\beta_j = \beta/k$

But tests differ by effect size!



Example: test the following hypotheses at  $\alpha_j = 0.0125$ ,  $\beta_j = 0.025$ , to have  $FWER_I \leq 0.05$ ,  $FWER_{II} \leq 0.10$

- $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$  Required sample size  
based on  $\{X_1^{(1)}, \dots, X_n^{(1)}\} \sim \text{Normal}(\theta^{(1)}, 1)$   $n_1 = 71$
- $H_0^{(2)} : \theta^{(2)} = 0$  vs  $H_A^{(2)} : \theta^{(2)} = 0.5$  Required sample size  
based on  $\{X_1^{(2)}, \dots, X_n^{(2)}\} \sim \text{Normal}(\theta^{(2)}, 0.2)$   $n_2 = 3$
- $H_0^{(3)} : \theta^{(3)} = 0.5$  vs  $H_A^{(3)} : \theta^{(3)} = 0.75$  Required sample size  
based on  $\{X_1^{(3)}, \dots, X_n^{(3)}\} \sim \text{Bernoulli}(\theta^{(3)})$   $n_3 = 71$
- $H_0^{(4)} : \theta^{(4)} = 0.5$  vs  $H_A^{(4)} : \theta^{(4)} = 0.6$  Required sample size  
based on  $\{X_1^{(4)}, \dots, X_n^{(4)}\} \sim \text{Bernoulli}(\theta^{(4)})$   $n_4 = 442$

## 5. Further optimization

### Bonferroni approach, non-sequentially

Conduct each test at  $\alpha_j = \alpha/k$  and  $\beta_j = \beta/k$

But tests differ by effect size!



Example: test the following hypotheses at  $\alpha_j = 0.0125$ ,  $\beta_j = 0.025$ , to have  $FWER_I \leq 0.05$ ,  $FWER_{II} \leq 0.10$

- $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$   
based on  $\{X_1^{(1)}, \dots, X_n^{(1)}\} \sim \text{Normal}(\theta^{(1)}, 1)$   
Required sample size  $n_1 = 71$
- $H_0^{(2)} : \theta^{(2)} = 0$  vs  $H_A^{(2)} : \theta^{(2)} = 0.5$   
based on  $\{X_1^{(2)}, \dots, X_n^{(2)}\} \sim \text{Normal}(\theta^{(2)}, 0.2)$   
Required sample size  $n_2 = 3$
- $H_0^{(3)} : \theta^{(3)} = 0.5$  vs  $H_A^{(3)} : \theta^{(3)} = 0.75$   
based on  $\{X_1^{(3)}, \dots, X_n^{(3)}\} \sim \text{Bernoulli}(\theta^{(3)})$   
Required sample size  $n_3 = 71$
- $H_0^{(4)} : \theta^{(4)} = 0.5$  vs  $H_A^{(4)} : \theta^{(4)} = 0.6$   
based on  $\{X_1^{(4)}, \dots, X_n^{(4)}\} \sim \text{Bernoulli}(\theta^{(4)})$   
Required sample size  $n = 442$

## 5. Further optimization

### Bonferroni approach, sequentially

Conduct each test at  $\alpha_j = \alpha/k$  and  $\beta_j = \beta/k$

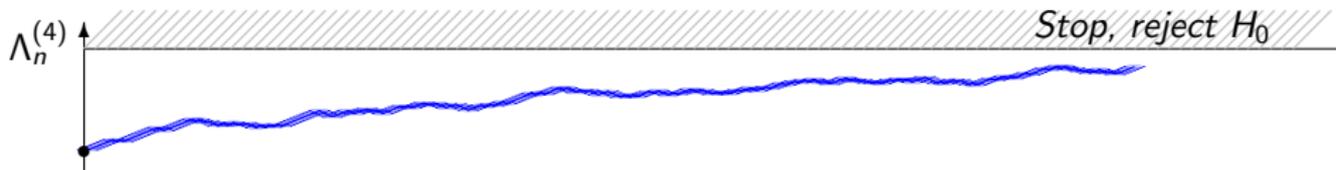
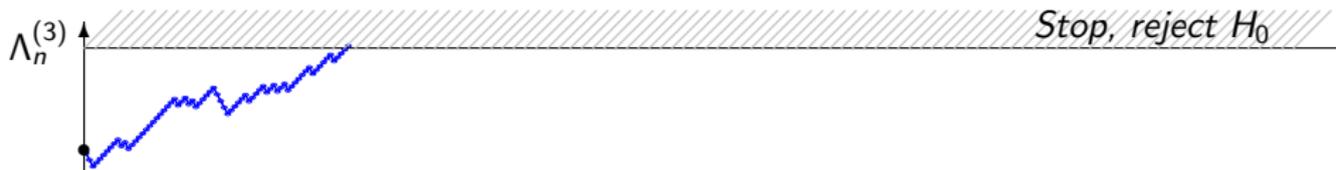
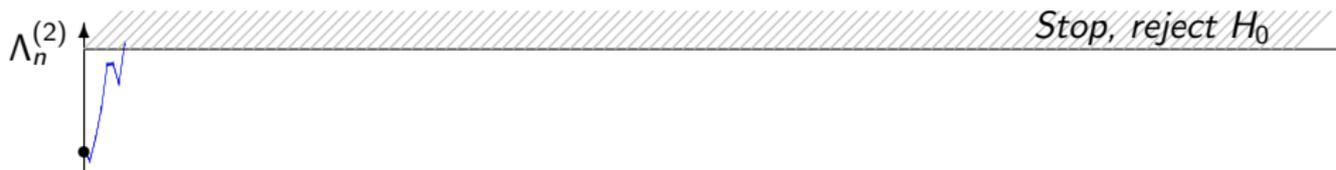
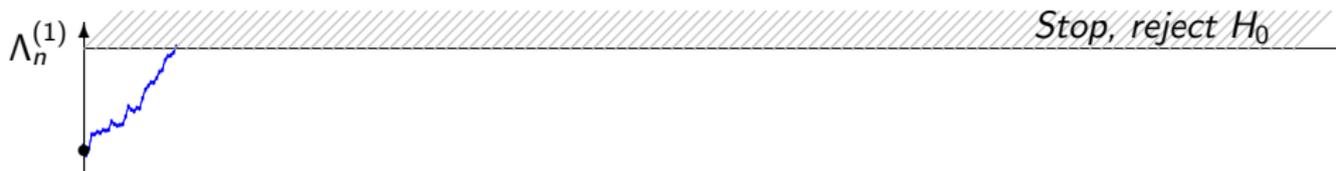
But tests differ by effect size!



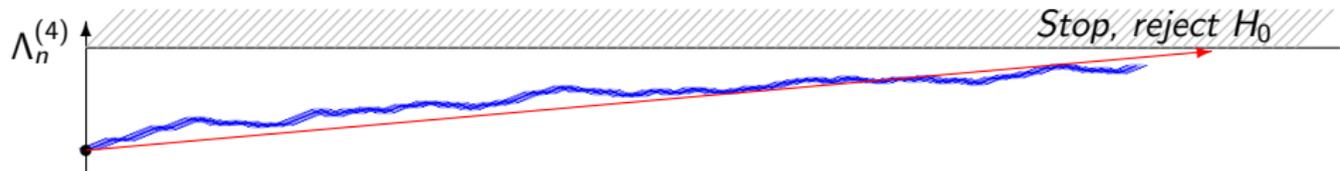
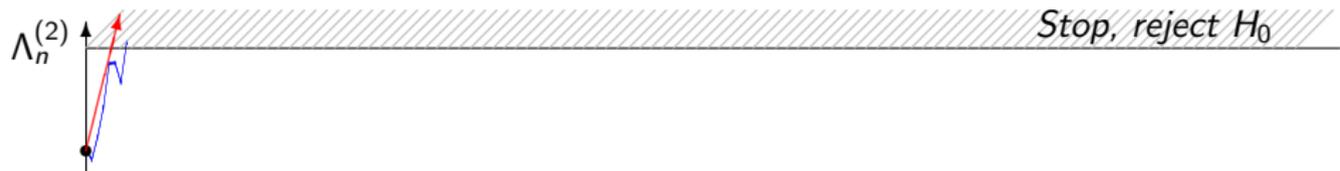
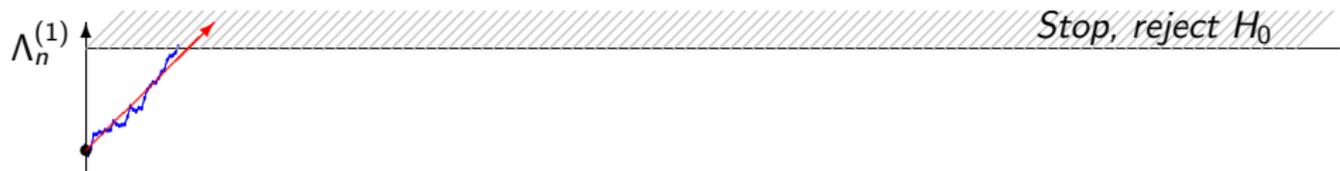
Example: test the following hypotheses at  $\alpha_j = 0.0125$ ,  $\beta_j = 0.025$ , to have  $FWER_I \leq 0.05$ ,  $FWER_{II} \leq 0.10$

- $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$   
based on  $\{X_1^{(1)}, X_2^{(1)}, \dots\} \sim \text{Normal}(\theta^{(1)}, 1)$   
Expected sample size  $E(T_1 | H_A) \approx 33.2$
- $H_0^{(2)} : \theta^{(2)} = 0$  vs  $H_A^{(2)} : \theta^{(2)} = 0.5$   
based on  $\{X_1^{(2)}, X_2^{(2)}, \dots\} \sim \text{Normal}(\theta^{(2)}, 0.2)$   
Expected sample size  $E(T_2 | H_A) \approx 1.3$
- $H_0^{(3)} : \theta^{(3)} = 0.5$  vs  $H_A^{(3)} : \theta^{(3)} = 0.75$   
based on  $\{X_1^{(3)}, X_2^{(3)}, \dots\} \sim \text{Bernoulli}(\theta^{(3)})$   
Expected sample size  $E(T_3 | H_A) \approx 31.8$
- $H_0^{(4)} : \theta^{(4)} = 0.5$  vs  $H_A^{(4)} : \theta^{(4)} = 0.6$   
based on  $\{X_1^{(4)}, X_2^{(4)}, \dots\} \sim \text{Bernoulli}(\theta^{(4)})$   
Expected sample size  $E(T_4 | H_A) \approx 206.4$

# Bonferroni approach, sequentially



# Bonferroni approach, sequentially



# Minimax problem

Minimax error spending:

$$\left\{ \begin{array}{ll} \text{Minimize } n = \max_{1, \dots, k}(n_j) & \text{or } E(T) = E \left\{ \max_{1, \dots, k}(T_j) \right\} \\ \text{(non-sequential version)} & \text{(sequential version)} \\ \\ \text{subject to } \sum_{j=1}^k \alpha_j = \alpha, & \\ & \sum_{j=1}^k \beta_j = \beta \end{array} \right.$$

## Minimaxity in non-sequential experiments

$$\text{Test} \quad \left\{ \begin{array}{l} H_0^{(1)} : \theta_1 = \theta_0^{(1)} \\ \dots \\ H_0^{(k)} : \theta_k = \theta_0^{(k)} \end{array} \right. \quad \text{vs} \quad \left\{ \begin{array}{l} H_A^{(1)} : \theta_1 = \theta_1^{(1)}, \\ \dots \\ H_A^{(k)} : \theta_k = \theta_1^{(k)}. \end{array} \right.$$

Tested at levels  $\alpha_j, \beta_j$ , with  $(\approx)$ Normal  $\hat{\theta}_j$ , test  $j$  requires

$$\text{sample size } n_j = \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} \right)^2, \text{ where}$$

$$\delta_j = \left| \frac{\theta_1^{(j)} - \theta_0^{(j)}}{\sigma_j} \right| \text{ is the } \textit{standardized effect size}, \text{ or Cohen's } d$$

# Minimaxity in non-sequential experiments

$$\text{Test} \quad \left\{ \begin{array}{l} H_0^{(1)} : \theta_1 = \theta_0^{(1)} \\ \dots \\ H_0^{(k)} : \theta_k = \theta_0^{(k)} \end{array} \right. \quad \text{vs} \quad \left\{ \begin{array}{l} H_A^{(1)} : \theta_1 = \theta_1^{(1)} \\ \dots \\ H_A^{(k)} : \theta_k = \theta_1^{(k)} \end{array} \right.$$

Tested at levels  $\alpha_j, \beta_j$ , with  $(\approx)$ Normal  $\hat{\theta}_j$ , test  $j$  requires

$$\text{sample size } n_j = \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} \right)^2, \text{ where}$$

$$\delta_j = \left| \frac{\theta_1^{(j)} - \theta_0^{(j)}}{\sigma_j} \right| \text{ is the } \textit{standardized effect size}, \text{ or Cohen's } d$$

$$n_j = \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} \right)^2 \quad \text{for a one-sample design}$$

$$n_j = 2 \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} \right)^2 \quad \text{for a two-sample design}$$

$$n_j = 2(1 - \rho) \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} \right)^2 \quad \text{for a matched-pairs design}$$

$$n_j = \frac{1}{2} \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} \right)^2 \quad \text{for a crossover design}$$

# Minimaxity in non-sequential experiments

$$\begin{aligned}n_j &= C_{\text{design}} \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} \right)^2 \\ &= \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j / \sqrt{C_{\text{design}}}} \right)^2\end{aligned}$$

Smaller effect size  $\delta_j$   
 $\Downarrow$   
more difficult test  
 $\Downarrow$   
requires a larger sample

## General approach

Information  $I_n$  is defined through  $\mathbf{E}(Z_n) = \theta\sqrt{I_n}$ .

In terms of  $I_n$ ,  $C_{\text{design}} = \sigma\sqrt{I_n/n}$ .

For the unified approach, design coefficient  $C_{\text{design}}$  can be merged with the effect size  $\delta_j$ .

Jennison & Turnbull (2000)

# Minimaxity in non-sequential experiments

$$\begin{aligned}n_j &= C_{\text{design}} \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} \right)^2 \\ &= \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j / \sqrt{C_{\text{design}}}} \right)^2\end{aligned}$$

Smaller effect size  $\delta_j$   
 $\Downarrow$   
more difficult test  
 $\Downarrow$   
requires a larger sample

Required sample size

$$n = \max_{1, \dots, k} (n_j) \sim \max_{1, \dots, k} \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} \right)^2 \hookrightarrow \text{minimize in } \{\alpha_j, \beta_j\}$$

subject to  $\sum_{j=1}^k \alpha_j = \alpha$ ,  $\sum_{j=1}^k \beta_j = \beta$ .

This is a minimax problem and its solution is an *equalizer*.

## Minimaxity in non-sequential experiments

$$\text{Equalizer} \Rightarrow \begin{cases} n_1 = \dots = n_k \\ \text{subject to } \sum_1^k \alpha_j = \alpha, \sum_1^k \beta_j = \beta \end{cases}$$

An equalizer is not unique:

$$\begin{cases} (2k) & \text{unknowns} & = \alpha_1, \dots, \alpha_k; \beta_1, \dots, \beta_k \\ (k+1) & \text{constraints} & = FWER_I, FWER_{II}, \text{ and } n_1 = \dots = n_k \end{cases}$$

## Minimaxity in non-sequential experiments

$$\text{Equalizer} \Rightarrow \begin{cases} n_1 = \dots = n_k & = \left( \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} \right)^2 \\ \text{subject to } \sum_1^k \alpha_j = \alpha, \sum_1^k \beta_j = \beta \end{cases}$$

An equalizer is not unique:

$$\begin{cases} (2k) & \text{unknowns} & = \alpha_1, \dots, \alpha_k; \beta_1, \dots, \beta_k \\ (k+1) & \text{constraints} & = FWER_I, FWER_{II}, \text{ and } n_1 = \dots = n_k \end{cases}$$

## Minimaxity in non-sequential experiments

$$\text{Equalizer} \Rightarrow \begin{cases} n_1 = \dots = n_k & = \left( \frac{\Phi^{-1}(\alpha_j)}{\delta_j} + \frac{\Phi^{-1}(\beta_j)}{\delta_j} \right)^2 \\ \text{subject to } \sum_1^k \alpha_j = \alpha, \sum_1^k \beta_j = \beta \end{cases}$$

An equalizer is not unique:

$$\begin{cases} (2k) & \text{unknowns} & = \alpha_1, \dots, \alpha_k; \beta_1, \dots, \beta_k \\ (k+1) & \text{constraints} & = FWER_I, FWER_{II}, \text{ and } n_1 = \dots = n_k \end{cases}$$

# Proportional error spending

$$\text{Equalizer} \Rightarrow \begin{cases} n_1 = \dots = n_k \\ \text{subject to } \sum_1^k \alpha_j = \alpha, \sum_1^k \beta_j = \beta \end{cases}$$

A convenient solution, close to being minimax, is to let

$$\frac{\Phi^{-1}(\alpha_j)}{\delta_j} = \text{const} = c_\alpha \quad \text{and} \quad \frac{\Phi^{-1}(\beta_j)}{\delta_j} = \text{const} = c_\beta.$$

Then

$$\alpha_j^{\text{prop}} = \Phi(-c_\alpha \delta_j), \quad \beta_j^{\text{prop}} = \Phi(-c_\beta \delta_j),$$

where  $c_\alpha, c_\beta$  are solutions of  $\begin{cases} \sum_j \Phi(-c_\alpha \delta_j) = \alpha \\ \sum_j \Phi(-c_\beta \delta_j) = \beta \end{cases}$ , and

$$n_j \equiv (c_\alpha + c_\beta)^2 = \text{const}.$$

# Minimax error spending (derivation)

Lagrangian ( $s = \sqrt{n}$ ):

$$\mathcal{L} = s + \sum_j \lambda_j \left( -\frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} - s \right) + \lambda_\alpha (\sum_j \alpha_j - \alpha) + \lambda_\beta (\sum_j \beta_j - \beta).$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_j} = -\frac{\lambda_j}{\delta_j \phi(\Phi^{-1}(\alpha_j))} + \lambda_\alpha = 0, \quad \frac{\partial \mathcal{L}}{\partial \beta_j} = -\frac{\lambda_j}{\delta_j \phi(\Phi^{-1}(\beta_j))} + \lambda_\beta = 0$$

The minimizer has

$$\frac{\phi(\Phi^{-1}(\alpha_j))}{\phi(\Phi^{-1}(\beta_j))} = \text{const} \quad \Rightarrow \quad (\Phi^{-1}(\alpha_j))^2 - (\Phi^{-1}(\beta_j))^2 = \text{const} = C$$

But also,  $-\frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} = s = \sqrt{n} = \text{const.}$

$$\Rightarrow \begin{cases} \frac{\Phi^{-1}(\alpha_j) + \Phi^{-1}(\beta_j)}{\delta_j} = \text{const} = s \\ (\Phi^{-1}(\alpha_j) - \Phi^{-1}(\beta_j))\delta_j = \text{const} = t \end{cases} \quad \Rightarrow \text{solution}$$

## Minimax error spending (solution)

Minimax error spending:

$$\alpha_j^{\text{opt}} = \Phi \left( -\frac{n\delta_j^2 + C}{2\delta_j\sqrt{n}} \right) \quad \text{and} \quad \beta_j^{\text{opt}} = \Phi \left( -\frac{n\delta_j^2 - C}{2\delta_j\sqrt{n}} \right).$$

where  $C$  and  $n$  are solutions of

$$\begin{cases} \sum_{j=1}^k \Phi \left( -\frac{n\delta_j^2 + C}{2\delta_j\sqrt{n}} \right) = \alpha \\ \sum_{j=1}^k \Phi \left( -\frac{n\delta_j^2 - C}{2\delta_j\sqrt{n}} \right) = \beta \end{cases}$$

Take  $n^{\text{prop}}$  and  $C^{\text{prop}} = (\Phi^{-1}(\alpha_j^{\text{prop}}))^2 - (\Phi^{-1}(\beta_j^{\text{prop}}))^2$  as initial guesses

# Performance comparison

## Example

### Test

•  $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$   
based on  $\{X_1^{(1)}, \dots, X_n^{(1)}\} \sim \text{Normal}(\theta^{(1)}, 1)$

Required sample size

$$n_1 = 71$$

•  $H_0^{(2)} : \theta^{(2)} = 0$  vs  $H_A^{(2)} : \theta^{(2)} = 0.5$   
based on  $\{X_1^{(2)}, \dots, X_n^{(2)}\} \sim \text{Normal}(\theta^{(2)}, 0.2)$

Required sample size

$$n_2 = 3$$

•  $H_0^{(3)} : \theta^{(3)} = 0.5$  vs  $H_A^{(3)} : \theta^{(3)} = 0.75$   
based on  $\{X_1^{(3)}, \dots, X_n^{(3)}\} \sim \text{Bernoulli}(\theta^{(3)})$

Required sample size

$$n_3 = 71$$

•  $H_0^{(4)} : \theta^{(4)} = 0.5$  vs  $H_A^{(4)} : \theta^{(4)} = 0.6$   
based on  $\{X_1^{(4)}, \dots, X_n^{(4)}\} \sim \text{Bernoulli}(\theta^{(4)})$

Required sample size

$$n_4 = 442$$

# Performance comparison

## Example

### Test

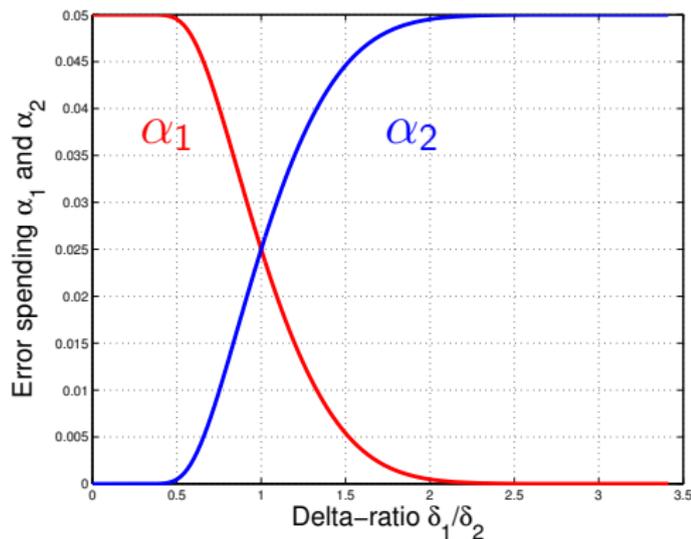
- $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$   
based on  $X_1^{(1)}, \dots, X_n^{(1)} \sim \text{Normal}(\theta^{(1)}, 1)$
- $H_0^{(2)} : \theta^{(2)} = 0$  vs  $H_A^{(2)} : \theta^{(2)} = 0.5$   
based on  $X_1^{(2)}, \dots, X_n^{(2)} \sim \text{Normal}(\theta^{(2)}, 0.2)$
- $H_0^{(3)} : \theta^{(3)} = 0.5$  vs  $H_A^{(3)} : \theta^{(3)} = 0.75$   
based on  $X_1^{(3)}, \dots, X_n^{(3)} \sim \text{Bernoulli}(\theta^{(3)})$
- $H_0^{(4)} : \theta^{(4)} = 0.5$  vs  $H_A^{(4)} : \theta^{(4)} = 0.6$   
based on  $X_1^{(4)}, \dots, X_n^{(4)} \sim \text{Bernoulli}(\theta^{(4)})$

### Error spending

Test	$\alpha_j$	$\beta_j$	$n_j$
<i>Standard Bonferroni, n = 442</i>			
1	0.0125	0.025	70.6
2	0.0125	0.025	2.8
3	0.0125	0.025	70.6
4	0.0125	0.025	441.3
<i>Proportional, n = 216</i>			
1	$1.95 \cdot 10^{-05}$	$6.36 \cdot 10^{-04}$	215.2
2	$2.80 \cdot 10^{-94}$	$1.07 \cdot 10^{-58}$	215.2
3	$1.95 \cdot 10^{-05}$	$6.36 \cdot 10^{-04}$	215.2
4	$4.99 \cdot 10^{-02}$	$9.87 \cdot 10^{-02}$	215.2
<i>Minimax, n = 215</i>			
1	$9.37 \cdot 10^{-05}$	$1.65 \cdot 10^{-04}$	214.6
2	$2.48 \cdot 10^{-75}$	$4.23 \cdot 10^{-75}$	214.6
3	$9.37 \cdot 10^{-05}$	$1.65 \cdot 10^{-04}$	214.6
4	$4.98 \cdot 10^{-02}$	$9.96 \cdot 10^{-02}$	214.6

# Finite-Sample Optimal Error Spending

Optimal  $\alpha_1$  and  $\alpha_2$ , the case of  $k = 2$  tests



$\delta_j$  = effect size, measure of difficulty of test  $H^{(j)}$

# We almost knew that... asymptotically

*Asymptotic result (Pitman alternative)* Consider an array of sequences of observed  $\mathbf{R}^d$  random vectors

$$\begin{array}{ccccccc} \mathbf{X}_{11}, & \mathbf{X}_{21}, & \cdots & \mathbf{X}_{i1}, & \cdots & & \\ & & & & & \cdots & \\ \mathbf{X}_{1\nu}, & \mathbf{X}_{2\nu}, & \cdots & \mathbf{X}_{i\nu}, & \cdots & & \\ & & & & & \cdots & \end{array}$$

In the  $\nu^{\text{th}}$  row,  $X_{i\nu}^{(j)} \sim f_j(\cdot | \theta_\nu^{(j)})$ .

Test

$$H_{0\nu}^{(j)} : \theta_\nu^{(j)} = \theta_0^{(j)} \quad \text{vs} \quad H_{A\nu}^{(j)} : \theta_\nu^{(j)} = \theta_0^{(j)} + \varepsilon_{j\nu}$$

for  $j = 1, \dots, d$ , where  $\varepsilon_{1\nu} \downarrow 0$  as  $\nu \rightarrow \infty$ ,  $\varepsilon_{1\nu} \ll \varepsilon_{j\nu}$  for  $j > 1$

so that the first test is the most difficult.

Let  $\alpha_\nu \downarrow 0$ ,  $\beta_\nu \downarrow 0$  are chosen so that

$$|\ln w_j \alpha_\nu| = o\left(\frac{\varepsilon_{j\nu}}{\varepsilon_{1\nu}}\right)^2 \quad \text{and} \quad |\ln w_j \beta_\nu| = o\left(\frac{\varepsilon_{j\nu}}{\varepsilon_{1\nu}}\right)^2$$

for  $w_j > 0$ ,  $\sum_2^d w_j = 1$ , and  $j = 2, \dots, d$ . Then, error probabilities

$$\begin{aligned} \alpha_{1\nu} &= \alpha - \alpha_\nu, & \beta_{1\nu} &= \beta - \beta_\nu, & \text{and,} \\ \alpha_{j\nu} &= w_j \alpha_\nu, & \beta_{j\nu} &= w_j \beta_\nu & \text{for } j > 1, \end{aligned}$$

are asymptotically optimal.

De and Baron (2012a)

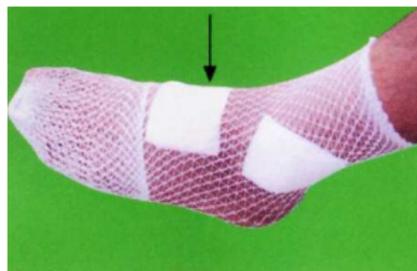
# Applications in Clinical Trials

## Clinical Trial of Flector<sup>®</sup> Patch

The trial was designed to establish that:

(1) Generic patch is as good as the brand name Flector<sup>®</sup> patch (*bioequivalence*)

(2) Both of them are more efficient than placebo (*efficacy*)



By the protocol, the total of 510 patients were randomized into three treatment groups,  $n = 170$  into each group, to attain a power of 0.99 on bioequivalence and 0.86 on efficacy under  $\alpha = 0.05$  for each test.

Cost: between \$1800 and \$3600 per patient (diagnostics, treatment, insurance, stipend, transportation, lodging, meals, documentation, and data management) + fixed costs  $\approx$  \$2.9M.

## Application: Clinical Trial of Flector Patch

Estimating standard deviations from earlier trials of similar products, the problem reduces to a multiple hypothesis testing

$$H_0^{(1)} : r = 0.8 \quad \text{vs} \quad H_A^{(1)} : r = 1;$$

$$H_0^{(2)} : \Delta = \mu_T - \mu_P = 0 \quad \text{vs} \quad H_A^{(2)} : \Delta = 4;$$

with

$$\sigma_1 = 0.373, \alpha_1 = 0.05, \beta_1 = 0.01; \quad \sigma_2 = 19.01, \alpha_2 = 0.05, \beta_2 = 0.14.$$

$$\text{We see that } \delta_1 = \left| \frac{r_A - r_0}{\sigma_1} \right| = 0.54 \quad \text{and} \quad \delta_2 = \left| \frac{\Delta_A - \Delta_0}{\sigma_2} \right| = 0.21,$$

so Test 2 is more difficult.

# Application: Clinical Trial of Flector Patch

## Design adjustments:

- Test of bioequivalence between two samples of size  $n$   
 $\Rightarrow$  Replace  $\delta_1$  with  $\delta_1^* = \frac{\delta_1}{\sqrt{2}} = 0.38$ .
- Test of efficacy between two samples of sizes  $2n$  and  $n$   
 $\Rightarrow$  Replace  $\delta_2$  with  $\delta_2^* = \delta_2 \sqrt{\frac{2}{3}} = 0.17$ .

Test 2 is still more difficult.

## Application: Clinical Trial of Flector Patch

- *Non-sequentially*, we can control FWE rates at the same levels with  $n = 163$  instead of  $n = 170$  in each treatment arm.
- Optimal error spending:  $\alpha = 0.10 = 0.0056 + 0.0944$ . The efficacy test is harder, so  $\alpha_2 > \alpha_1$
- *Sequentially*, the same error spending brings the saving between 21% and 45% on the expected sample size

True combination of null and alternative hypotheses	$H_0^{(1)}, H_0^{(2)}$	$H_A^{(1)}, H_0^{(2)}$	$H_0^{(1)}, H_A^{(2)}$	$H_A^{(1)}, H_A^{(2)}$
Expected sample size	92.7	114.7	117.8	134.1
Type I familywise error rate	0.0774	0.0682	0	0
Type II error probability, Test 1	0	0.0040	0	0.0022
Type II error probability, Test 2	0	0	0.1099	0.0868

# Sequential experiments

*Example:*

- $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$   
based on  $\{X_1^{(1)}, X_2^{(1)}, \dots\} \sim \text{Normal}(\theta^{(1)}, 1)$   
Expected sample size  $E(T_1 | H_A) \approx 33.2$
- $H_0^{(2)} : \theta^{(2)} = 0$  vs  $H_A^{(2)} : \theta^{(2)} = 0.5$   
based on  $\{X_1^{(2)}, X_2^{(2)}, \dots\} \sim \text{Normal}(\theta^{(2)}, 0.2)$   
Expected sample size  $E(T_2 | H_A) \approx 1.3$
- $H_0^{(3)} : \theta^{(3)} = 0.5$  vs  $H_A^{(3)} : \theta^{(3)} = 0.75$   
based on  $\{X_1^{(3)}, X_2^{(3)}, \dots\} \sim \text{Bernoulli}(\theta^{(3)})$   
Expected sample size  $E(T_3 | H_A) \approx 31.8$
- $H_0^{(4)} : \theta^{(4)} = 0.5$  vs  $H_A^{(4)} : \theta^{(4)} = 0.6$   
based on  $\{X_1^{(4)}, X_2^{(4)}, \dots\} \sim \text{Bernoulli}(\theta^{(4)})$   
Expected sample size  $E(T_4 | H_A) \approx 206.4$

# Sequential experiments

*Example:*

•  $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$   
based on  $\{X_1^{(1)}, X_2^{(1)}, \dots\} \sim \text{Normal}(\theta^{(1)}, 1)$

•  $H_0^{(2)} : \theta^{(2)} = 0$  vs  $H_A^{(2)} : \theta^{(2)} = 0.5$   
based on  $\{X_1^{(2)}, X_2^{(2)}, \dots\} \sim \text{Normal}(\theta^{(2)}, 0.2)$

•  $H_0^{(3)} : \theta^{(3)} = 0.5$  vs  $H_A^{(3)} : \theta^{(3)} = 0.75$   
based on  $\{X_1^{(3)}, X_2^{(3)}, \dots\} \sim \text{Bernoulli}(\theta^{(3)})$

•  $H_0^{(4)} : \theta^{(4)} = 0.5$  vs  $H_A^{(4)} : \theta^{(4)} = 0.6$   
based on  $\{X_1^{(4)}, X_2^{(4)}, \dots\} \sim \text{Bernoulli}(\theta^{(4)})$

**Bonferroni spending**

Expected sample size

$$E(T_1 | H_A) \approx 33.2$$

Expected sample size

$$E(T_2 | H_A) \approx 1.3$$

Expected sample size

$$E(T_3 | H_A) \approx 31.8$$

Expected sample size

$$E(T_4 | H_A) \approx 206.4$$

# Sequential experiments

*Example, with the same error spending:*

•  $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$   
based on  $\{X_1^{(1)}, X_2^{(1)}, \dots\} \sim \text{Normal}(\theta^{(1)}, 1)$

•  $H_0^{(2)} : \theta^{(2)} = 0$  vs  $H_A^{(2)} : \theta^{(2)} = 0.5$   
based on  $\{X_1^{(2)}, X_2^{(2)}, \dots\} \sim \text{Normal}(\theta^{(2)}, 0.2)$

•  $H_0^{(3)} : \theta^{(3)} = 0.5$  vs  $H_A^{(3)} : \theta^{(3)} = 0.75$   
based on  $\{X_1^{(3)}, X_2^{(3)}, \dots\} \sim \text{Bernoulli}(\theta^{(3)})$

•  $H_0^{(4)} : \theta^{(4)} = 0.5$  vs  $H_A^{(4)} : \theta^{(4)} = 0.6$   
based on  $\{X_1^{(4)}, X_2^{(4)}, \dots\} \sim \text{Bernoulli}(\theta^{(4)})$

## Minimax spending

Expected sample size

$$E(T_1 | H_A) \approx 74.2$$

Expected sample size

$$E(T_2 | H_A) \approx 55.0$$

Expected sample size

$$E(T_3 | H_A) \approx 70.9$$

Expected sample size

$$E(T_4 | H_A) \approx 118.3$$

# Sequential experiment: simultaneous SPRT

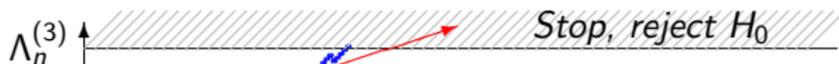


## Sequential log-likelihood ratio statistics



# Sequential experiment: simultaneous SPRT

## Sequential log-likelihood ratio statistics



Speed  $\sim$  slope  
 $= E(\Lambda_1^{(j)})$   
 $=$  Kullback  
information

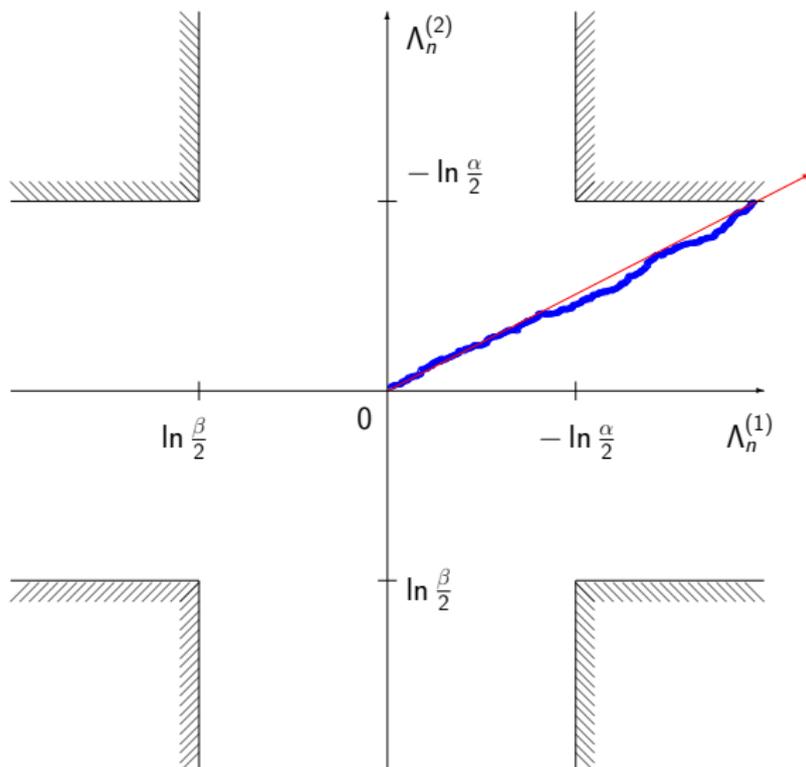


# Bonferroni vs Weighted Bonferroni

## Bonferroni

Uniform error spending

$$\begin{cases} \alpha = \sum_j \alpha/k \\ \beta = \sum_j \beta/k \end{cases}$$

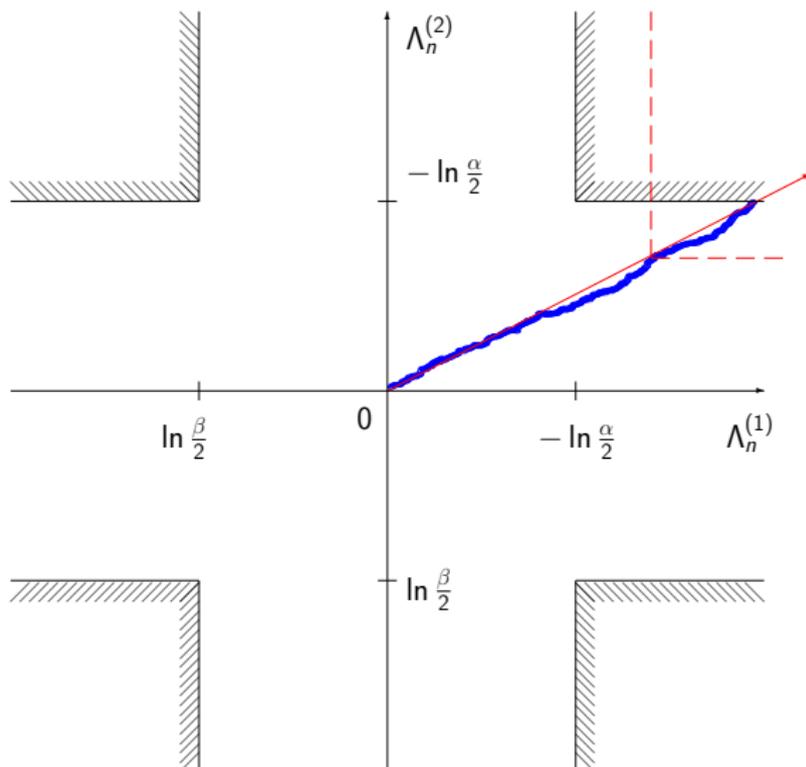


# Bonferroni vs Generalized Bonferroni

## Generalized Bonferroni

Optimal error spending

$$\left\{ \begin{array}{l} \ln \alpha_j \sim K(\theta_A^{(j)}, \theta_A^{(j)}) \\ \sum \alpha_j = \alpha \\ \ln \beta_j \sim K(\theta_0^{(j)}, \theta_0^{(j)}) \\ \sum \beta_j = \beta \end{array} \right.$$



# Minimax Error Spending, Sequential Bonferroni

Test  $H_0^{(j)} : \theta^{(j)} = \theta_0^{(j)}$  vs  $H_A^{(j)} : \theta^{(j)} = \theta_1^{(j)}$  with adjusted Wald stopping boundaries  $a_j = -\ln \alpha_j$ ,  $b_j = \ln \beta_j$ , for  $j = 1, \dots, k$ .

Expected sample size is

$$\mathbf{E}_{\theta^{(1)}, \dots, \theta^{(k)}}(T) \approx \max_j \left\{ a_j / K_A^{(j)} \text{ or } -b_j / K_0^{(j)} \right\} \leftrightarrow \text{minimize in } \alpha_j, \beta_j$$

subject to  $\sum_{j=1}^k \alpha_j = \alpha$ ,  $\sum_{j=1}^k \beta_j = \beta$ ,

where  $K_0^{(j)} = K(\theta_0^{(j)}, \theta_1^{(j)}) = -\mathbf{E}_0 \Lambda_1^{(j)}$ ,  $K_A^{(j)} = K(\theta_1^{(j)}, \theta_0^{(j)}) = \mathbf{E}_A \Lambda_1^{(j)}$  are Kullback-Leibler information numbers.

---

Asymptotic result for a multidimensional random walk:  $\mathbf{E}(T)$  is asymptotically proportional to the distance to the boundary along the direction of  $\mathbf{E}(\Lambda_1)$  and inversely proportional to  $\|\mathbf{E}\Lambda_1\|$ .

Borovkov and Mogulskii (2001)

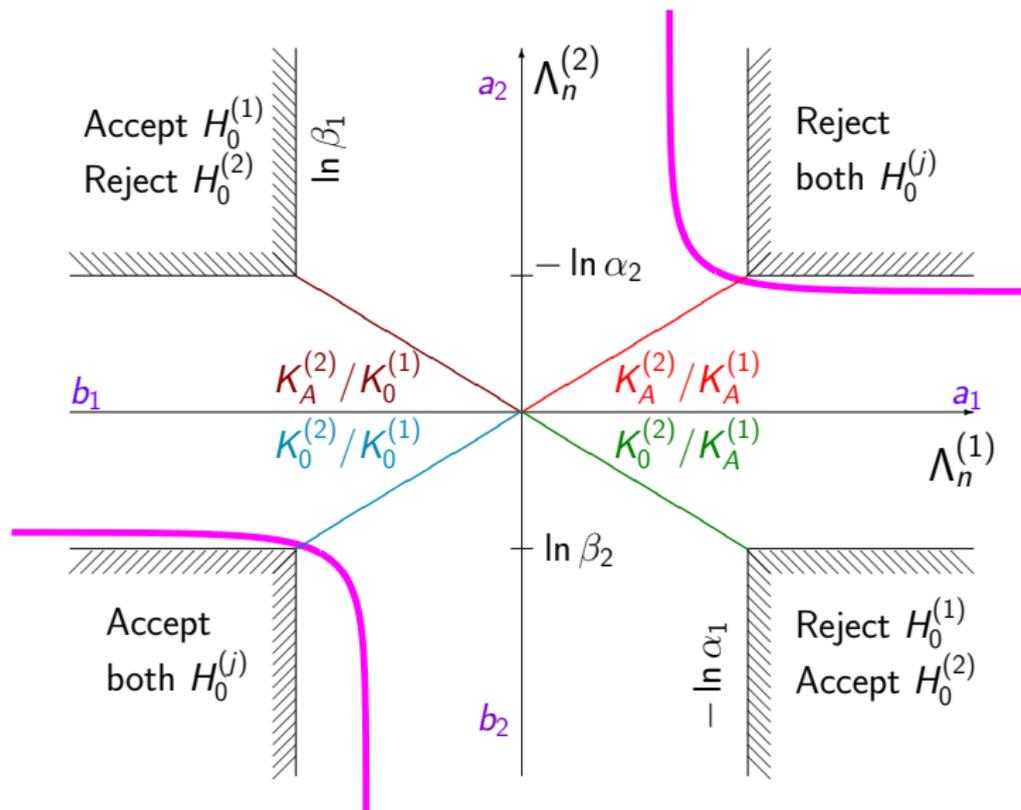
# Equalizer Solution for Sequential Bonferroni ( $k = 2$ )

Boundaries  
 $a_1, a_2; b_1, b_2$   
 solve

$$\begin{cases} a_1 = -\ln \alpha_1 \\ a_2 = -\ln \alpha_2 \\ \alpha_1 + \alpha_2 = \alpha \end{cases}$$

and

$$\begin{cases} b_1 = \ln \beta_1 \\ b_2 = \ln \beta_2 \\ \beta_1 + \beta_2 = \beta \end{cases}$$



# Equalizer Solution for Sequential Bonferroni (computation)

Minimax error spending and stopping boundaries

$$\begin{cases} a_j = c_\alpha K_A^{(j)} = -\ln \alpha_j \\ b_j = -c_\beta K_0^{(j)} = \ln \beta_j \end{cases}$$

where  $c_\alpha$  and  $c_\beta$  are unique solutions of

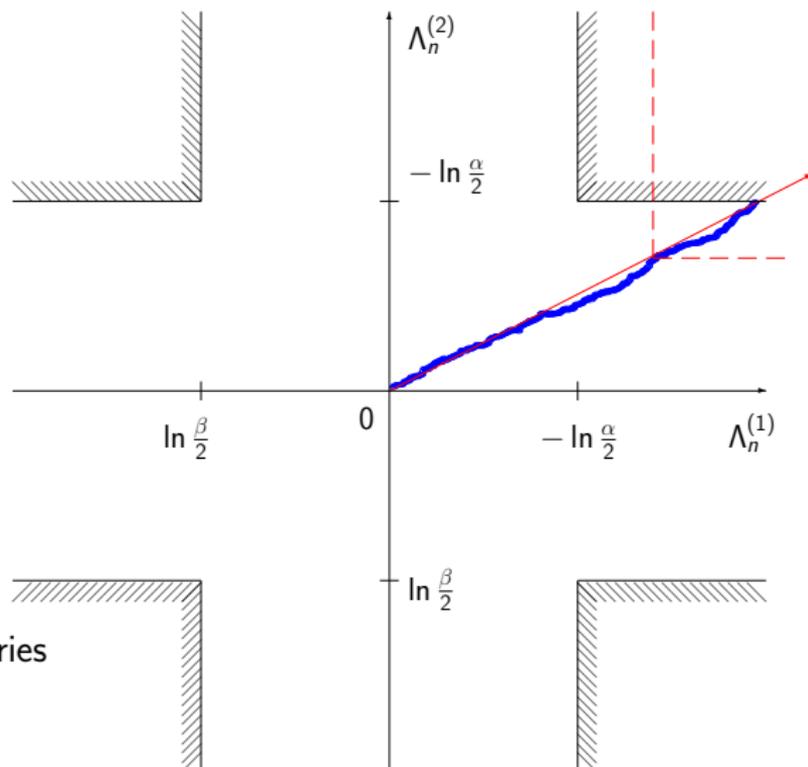
$$\begin{cases} \sum_j \exp \left\{ -c_\alpha K_A^{(j)} \right\} = \alpha \\ \sum_j \exp \left\{ -c_\beta K_0^{(j)} \right\} = \beta \end{cases}$$

# Equalizer in terms of stopping boundaries

Equalizer solution

$$\left\{ \begin{array}{l} a_j = (-\ln \alpha_j) K(\theta_A^{(j)}, \theta_A^{(j)}) \\ \sum \alpha_j = \alpha \\ b_j = (\ln \beta_j) K(\theta_0^{(j)}, \theta_0^{(j)}) \\ \sum \beta_j = \beta \end{array} \right.$$

$a_j, b_j =$  stopping boundaries  
for LLR  $\Lambda_n^{(j)}$



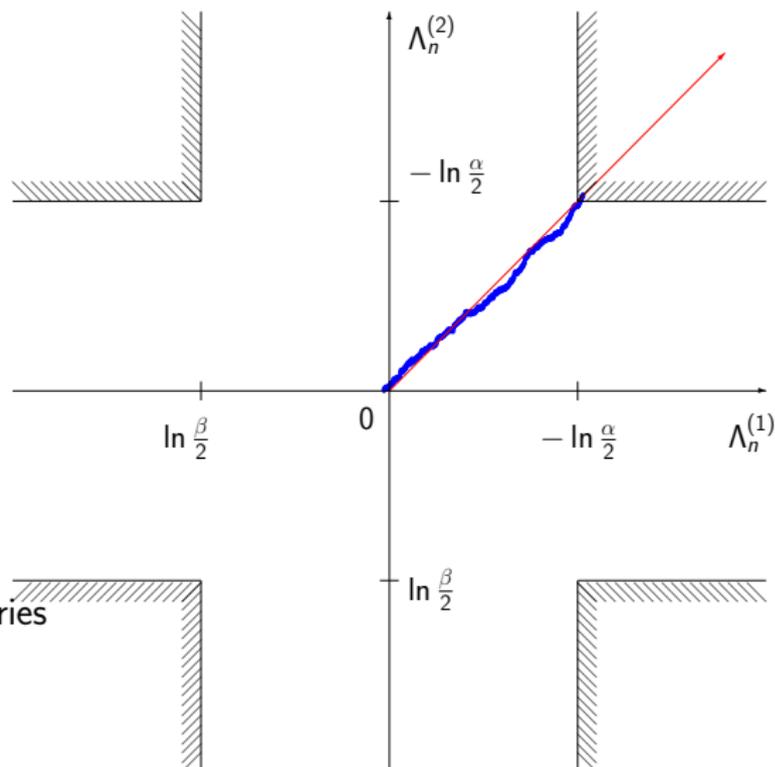
# The same equalizer in terms of LLR

Equalizer solution

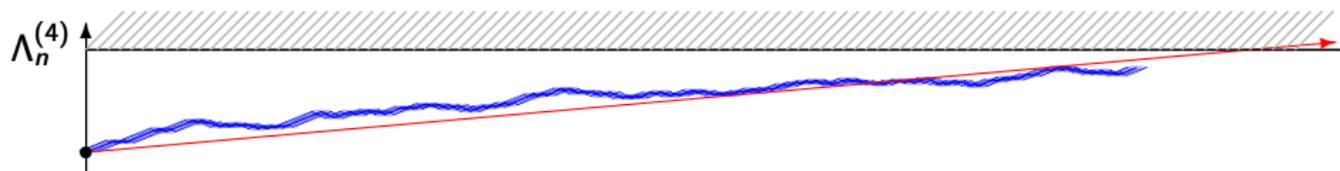
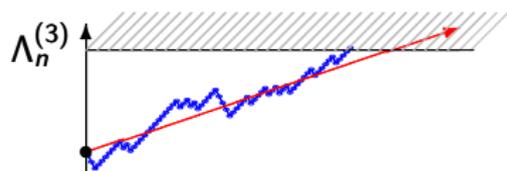
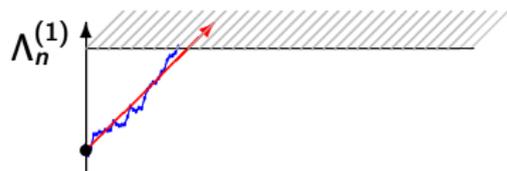
$$\left\{ \begin{array}{l} a_j = -\ln \alpha_j \\ \sum \alpha_j = \alpha \\ b_j = \ln \beta_j \\ \sum \beta_j = \beta \end{array} \right.$$

$a_j, b_j =$  stopping boundaries  
for **weighted** LLR

$$\lambda_n^{(j)} = \Lambda_n^{(j)} / K_j$$



# Equalizer Solution for Sequential Bonferroni



## Weighted LLR statistics

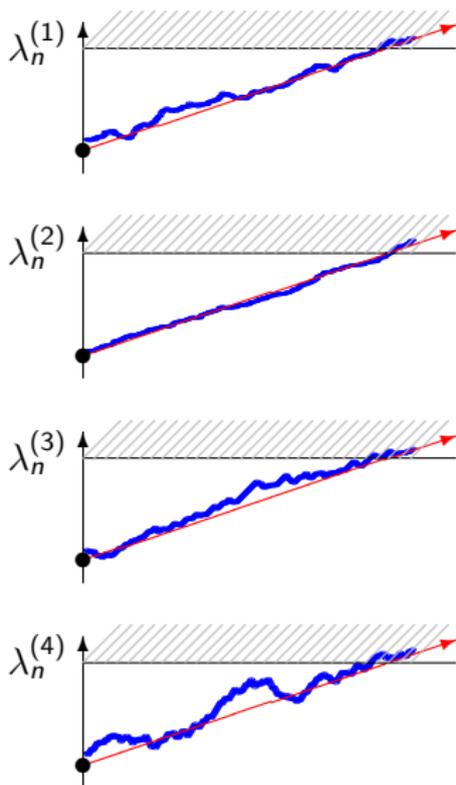
$$\text{Let } \lambda_n^{(j)} = \frac{\Lambda_n^{(j)}}{K_j}$$

“equalizing” slopes  
⇒ “equalizing” stopping times

## Minimax principle:

Sample until the *last* test is completed  
Minimax strategy = *equalizer*

# Equalizer Solution for Sequential Bonferroni



## Weighted LLR statistics

$$\text{Let } \lambda_n^{(j)} = \frac{\Lambda_n^{(j)}}{K_j}$$

“equalizing” slopes  
⇒ “equalizing” stopping times

## Minimax principle:

Sample until the *last* test is completed  
Minimax strategy = *equalizer*

# Minimax Error Spending with the Holm Method

*Original (non-sequential) Holm treats all tests equally*

Ordered p-values

$$p^{[1]} \leq p^{[2]} \leq \dots \leq p^{[d]}$$

have to be compared against

$$\frac{\alpha}{k}, \quad \frac{\alpha}{d-1}, \quad \dots, \quad \alpha.$$

so that in the proof,

$$\begin{aligned} \text{FWER} \leq \dots &\leq \mathbf{P} \left\{ \min \left( p^{[j]} \mid H_0^{[j]} \text{ is true} \right) \leq \frac{\alpha}{d - [j] + 1} \right\} \\ &\leq \mathbf{P} \left\{ \min \left( p^{[j]} \mid H_0^{[j]} \text{ is true} \right) \leq \frac{\alpha}{\# \text{ true } H_0^{(j)}} \right\} \\ &\leq \sum_{j: H_0^{(j)} \text{ is true}} \frac{\alpha}{\# \text{ true } H_0^{(j)}} = \alpha. \end{aligned}$$



# Minimax Error Spending with the Holm Method

Order rescaled LLR statistics

$$\lambda_n^{\{1\}} = \frac{\Lambda_n^{\{1\}}}{K_A^{\{1\}}} \leq \dots \leq \lambda_n^{\{d\}} = \frac{\Lambda_n^{\{1\}}}{K_A^{\{d\}}}; \quad \lambda_n^{[1]} = \frac{\Lambda_n^{[1]}}{K_0^{[1]}} \leq \dots \leq \lambda_n^{[d]} = \frac{\Lambda_n^{[1]}}{K_0^{[d]}}.$$

Test them against stopping boundaries:

$$\left\{ \begin{array}{l} a_j \text{ is a solution of } \sum_{m=1}^j \exp \left\{ -a_j K_A^{\{m\}} \right\} = \alpha \\ b_j \text{ is a solution of } \sum_{m=1}^{d-j+1} \exp \left\{ b_j K_0^{[m]} \right\} = \beta. \end{array} \right.$$

We can show that

$$a_1 < \dots < a_k$$

$$b_1 < \dots < b_k.$$

$$K_A^{\{m\}} := K_A^{(m:k)} \\ K_0^{[m]} := K_0^{(k-m+1:k)}$$

Stopping rule:  $T = \min \left\{ n : \lambda_n^{\{j\}} \geq a_j \text{ or } \lambda_n^{[j]} \leq b_j \forall j \right\}.$

Decision rule: reject  $H_0^{\{j\}}$  for  $j : \lambda_n^{\{j\}} \geq a_j.$

# Minimax Error Spending with the Holm Method

## Theorem.

The sequential stepwise multiple testing procedure based on weighted LLR controls  $FWER_I$  and  $FWER_{II}$  at levels  $\alpha$  and  $\beta$ .

Proof:

Let  $m = \max \{j : H_0^{\{j\}} \text{ is true} \}$  at time  $T$ .

Then, there are at most  $m$  true  $H_0$ 's,  $|\mathcal{T}| \leq m$ , and

$$\begin{aligned} FWER-I &= \mathbf{P} \left\{ \text{reject } H_0^{\{m\}} \right\} = \mathbf{P} \left\{ \lambda_T^{\{m\}} \geq a_m \right\} \leq \mathbf{P} \left\{ \lambda_T^{\{m\}} \geq a_{|\mathcal{T}|} \right\} \\ &= \mathbf{P} \left\{ \max_{j \in \mathcal{T}} \lambda_T^{(j)} \geq a_{|\mathcal{T}|} \right\} \leq \sum_{j \in \mathcal{T}} \mathbf{P} \left\{ \lambda_T^{(j)} \geq a_{|\mathcal{T}|} \right\} \\ &= \sum_{j \in \mathcal{T}} \mathbf{P} \left\{ \Lambda_T^{(j)} \geq a_{|\mathcal{T}|} K^{(j)} \right\} \leq \sum_{j \in \mathcal{T}} \exp \left\{ -a_{|\mathcal{T}|} K^{(j)} \right\} \\ &\leq \sum_{j=1}^{|\mathcal{T}|} \exp \left\{ -a_{|\mathcal{T}|} K^{\{j\}} \right\} = \alpha \end{aligned}$$

Similarly,  $FWER-II \leq \beta$ .

Baron and MacMillan (2019)

# Minimax Error Spending with the Holm Method

## *Special case*

For  $K_1 = \dots = K_k = K$ , the boundary values are solutions of:

$$j \exp\{-Ka_j\} = \alpha \quad \Rightarrow \quad a_j = \frac{-\ln(\alpha/j)}{K},$$

$$(d - j + 1) \exp\{Kb_j\} = \beta \quad \Rightarrow \quad b_j = \frac{\ln(\beta/(d - j + 1))}{K},$$

so we have boundaries  $\{\ln(\beta/(d - j + 1)), -\ln(\alpha/j)\}$  for  $\Lambda_T^{[j]}$ .

This is the Holm-type stepwise sequential multiple testing procedure. Special case!

De, Baron (2012b)

# Minimax Error Spending with the Holm Method

## Example.

Normal( $\theta^{(1)}, 1$ ), test  $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$   
 Normal( $\theta^{(1)}, 0.2$ ), test  $H_0^{(1)} : \theta^{(1)} = 0$  vs  $H_A^{(1)} : \theta^{(1)} = 0.5$   
 Bernoulli( $\theta^{(3)}$ ), test  $H_0^{(3)} : \theta^{(3)} = 0.5$  vs  $H_A^{(3)} : \theta^{(3)} = 0.75$   
 Bernoulli( $\theta^{(4)}$ ), test  $H_0^{(4)} : \theta^{(4)} = 0.5$  vs  $H_A^{(4)} : \theta^{(4)} = 0.6$

$E(T)$	Sequential Procedure			
# true hypotheses	Bonferroni	Stepwise	Generalized Bonferroni	Weighted Stepwise
4	211	150	113	110
3	215	178	111	97
2	212	196	115	105
1	213	193	114	94
0	181	120	121	103

# Non-sequential analogue of sequential Holm

Order  $K_0^{[1]} \leq \dots \leq K_0^{[d]}$  and  $K_A^{[1]} \leq \dots \leq K_A^{[d]}$

Let  $a_k$  be the unique solution of

$$\sum_{j=1}^{d-k+1} \exp \left\{ -a_k K_A^{[j]} \right\} = \alpha.$$

Define weighted p-values  $q^{(j)} = -\log(p^{(j)})/K_A^{(j)}$ ; order them,  $q^{[1]} \geq \dots \geq q^{[d]}$ .

Reject  $H_0^{[1]}, \dots, H_0^{[m]}$ , corresponding to  $q^{[1]}, \dots, q^{[m]}$ , where  $m = \max \{j : q^{[j]} \geq a_j\}$ ; accept all  $H_0^{(j)}$  if  $q^{[j]} < a_j$  for all  $j$ .

**Theorem.** The proposed step-down multiple testing procedure with critical values  $a_j$  for weighted p-values  $q^{[j]}$  controls the Type I familywise error rate,  $FWER_I \leq \alpha$ .

Baron, MacMillan (2019)

## Related problems – retrospectively

Considered problem:

$$\left\{ \begin{array}{l} \max_{\mathcal{T}} \mathbf{E}_{\mathcal{T}}(T) \\ \text{or } \max_j n_j \end{array} \right. \hookrightarrow \text{min under } FWER_I \leq \alpha, FWER_{II} \leq \beta$$

Retrospective error spending:

A sample of size  $n$  has been collected.

Partition  $\alpha = \sum \alpha_j$  to maximize power and minimize  $\beta = \sum \beta_j$ .

$$\text{Solution: } \left\{ \begin{array}{l} \alpha_j = \Phi^{-1} \left( -\frac{C + \delta_j^2 n}{2\delta_j \sqrt{n}} \right) \\ \text{where } C \text{ is the unique solution of} \\ \sum_j \Phi^{-1} \left( -\frac{C + \delta_j^2 n}{2\delta_j \sqrt{n}} \right) = \alpha \end{array} \right.$$

## Related problems – cost and loss

*Considered problem:*

$$\left\{ \begin{array}{l} \max_{\mathcal{T}} \mathbf{E}_{\mathcal{T}}(T) \\ \text{or } \max_j n_j \end{array} \right. \hookrightarrow \text{min under } FWER_I \leq \alpha, FWER_{II} \leq \beta$$

*Related problems:*

Let  $\left\{ \begin{array}{l} w_{\mathcal{T}} = \text{cost of each sampling unit under } \mathcal{T} \\ \pi(\mathcal{T}) = \text{prior probabilities} \end{array} \right.$

- Minimize  $\max_{\mathcal{T}} w_{\mathcal{T}} \mathbf{E}_{\mathcal{T}}(T) = \text{worst risk}$
- Minimize  $\sum_{\mathcal{T}} \pi(\mathcal{T}) w_{\mathcal{T}} \mathbf{E}_{\mathcal{T}}(T) = \text{Bayes risk}$
- Minimize  $\sum_{\mathcal{T}} \pi(\mathcal{T}) \mathbf{E}_{\mathcal{T}}(T) = \text{overall ASN}$

## Related problems – minimax risk optimization

$$\text{Loss function } \begin{cases} L_1^{(j)} & = \text{loss for a Type I error on } H^{(j)} \\ L_2^{(j)} & = \text{loss for a Type II error on } H^{(j)} \end{cases}$$

$$\text{Minimize } R_{\max}(\alpha, \beta) = \max_{j=1, \dots, k} \max \left\{ \alpha_j L_1^{(j)}, \beta_j L_2^{(j)} \right\}$$

$$\text{subject to } \sum \alpha_j = \alpha \text{ and } \sum \beta_j = \beta$$

$$\text{This risk splits } R_{\max}(\alpha, \beta) = \max \left\{ \max_j \alpha_j L_1^{(j)}, \max_j \beta_j L_2^{(j)} \right\}$$

$$\text{Minimize separately, } \min_{\alpha} \max_j \alpha_j L_1^{(j)} \text{ and } \min_{\beta} \max_j \beta_j L_2^{(j)}$$

$$\text{Equalizer solutions } \Rightarrow \boxed{\alpha_j = \frac{\alpha/L_1^{(j)}}{\sum_k (1/L_1^{(k)})}, \quad \beta_j = \frac{\beta/L_2^{(j)}}{\sum_k (1/L_2^{(k)})}}$$

## Risk-Optimal Holm Solution

Bonferroni suggests stopping boundaries for LLR  $\Lambda_n^{(j)}$

$$\begin{cases} a_j = -\ln \alpha = \ln L_1^{(j)} + C_1 \\ b_j = \ln \beta = -\ln L_2^{(j)} + C_2 \end{cases}$$

But for Holm,  $\sum_j \alpha_j > \alpha$ ,  $\sum_j \beta_j > \beta$ .

Define shifted LLR  $U_n^{(j)} = \Lambda_n^{(j)} - \ln L_1^{(j)}$ ,  $V_n^{(j)} = \Lambda_n^{(j)} + \ln L_2^{(j)}$ ;

order  $U_n^{\{1\}} \leq \dots \leq U_n^{\{d\}}$ ,  $V_n^{\{1\}} \leq \dots \leq V_n^{\{d\}}$ , at time  $n$ .

Test them against stopping boundaries:

$$a_j = -\log \alpha + \log \sum_{m=1}^j \frac{1}{L_1^{\{m\}}}, \quad b_j = \log \beta - \log \sum_{m=1}^{d-j+1} \frac{1}{L_2^{\{m\}}}$$

(  $L_1^{\{j\}}$  is the  $j$ -th smallest  $L_1^{(j)}$ , and  $L_2^{\{j\}}$  is the  $j$ -th smallest  $L_2^{(j)}$  )

# Risk-Optimal Holm Solution

Stopping rule

$$T^* = \min \left\{ n : \bigcap_{j=1}^k \left( U_n^{(j)} \geq a_{r(j)} \cup V_n^{(j)} \leq b_{s(j)} \right) \right\}$$

where  $r(j)$  = rank of  $U_T^{(j)}$  among  $U_T^{(1)}, \dots, U_T^{(k)}$ ;  $s(j)$  = rank of  $V_T^{(j)}$  among  $V_T^{(1)}, \dots, V_T^{(k)}$

Decision rule: reject  $H_0^{\{j\}}$  iff  $U_{T^*}^{\{j\}} \geq a_j$ .

## Theorem.

- The stopping time  $T^*$  is proper.
- This multiple testing procedure controls  $FWER_I$  and  $FWER_{II}$  at levels  $\alpha$  and  $\beta$ .

## 6. Truncated group sequential design for multiple testing

- Sample in groups of size  $m$ .
- Control

Familywise Type I error rate  $\leq \alpha$

Familywise Type II error rate  $\leq \beta$

- $P(T \leq K) = 1$ ,  
 $T$  is stopping time,  $K = \max$  allowed number of groups
- Do it at low expected cost

Jennison & Turnbull (2000)

# Bonferroni approach

Choose

$$\alpha_j \in (0, 1), \quad \sum \alpha_j = \alpha; \quad \beta_j \in (0, 1), \quad \sum \beta_j = \beta$$

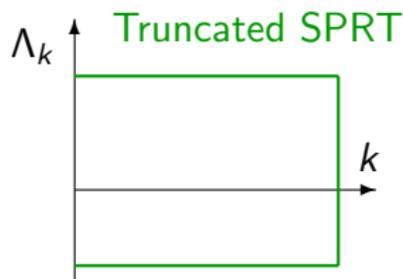
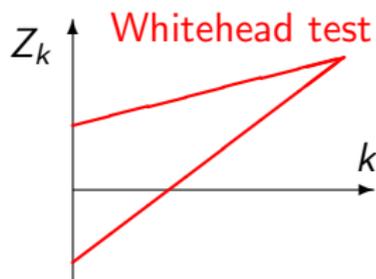
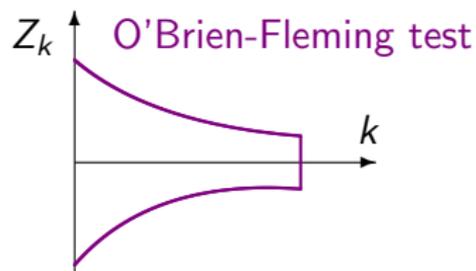
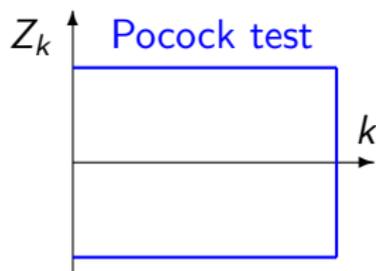
Test  $H_0^{(j)}$  at level  $\alpha_j$  and power  $(1 - \beta_j)$  by any known sequential procedure (e.g., Pocock, O'Brien-Fleming, Wang-Tsiatis, Whitehead triangular tests)

$$\begin{aligned} P \left( \bigcup_{j=1}^d \text{Type I error on } H_0^{(j)} \right) &\leq \sum P \left( \text{Type I error on } H_0^{(j)} \right) \\ &= \sum \alpha_j = \alpha \end{aligned}$$

$$\begin{aligned} P \left( \bigcup_{j=1}^d \text{Type II error on } H_A^{(j)} \right) &\leq \sum P \left( \text{Type II error on } H_A^{(j)} \right) \\ &= \sum \beta_j = \beta \end{aligned}$$

# Group sequential tests

For each  $H_0^{(j)}$ , use a group sequential  $(\alpha_j, \beta_j)$ -test



Next: improve Bonferroni using truncated SPRT

## Stopping Time and Decision Rule

Following SPRT approach for single hypothesis testing, truncated SPRT is based on log-likelihood ratios

$$\Lambda_n = \ln \frac{f(x|\theta_1)}{f(x|\theta_0)}$$

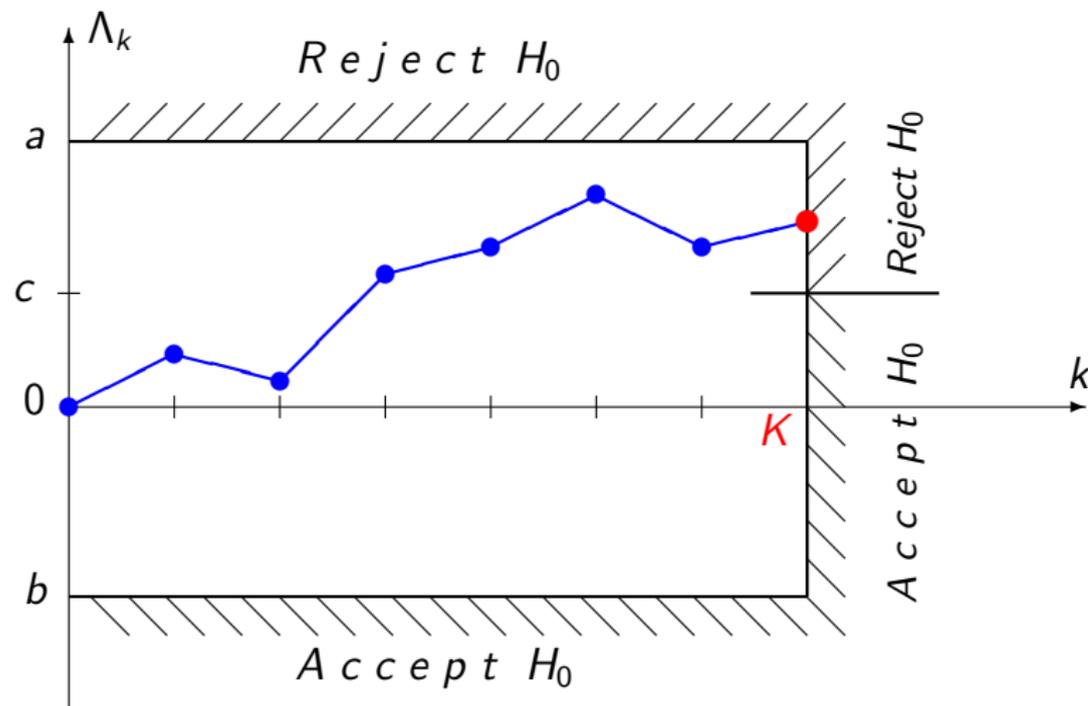
Stopping time

$$T = \min(K, \min \{n : \Lambda_n \notin (b, a)\}) \leq K$$

Decision rule,

$$\delta = \begin{cases} \text{reject } H_0 & \text{if } \Lambda_T \geq a \text{ or } \Lambda_T \geq c \cap T = K \\ \text{accept } H_0 & \text{if } \Lambda_T \leq b \text{ or } \Lambda_T < c \cap T = K \end{cases}$$

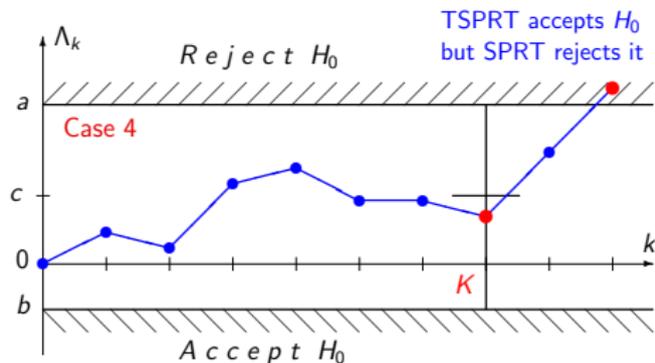
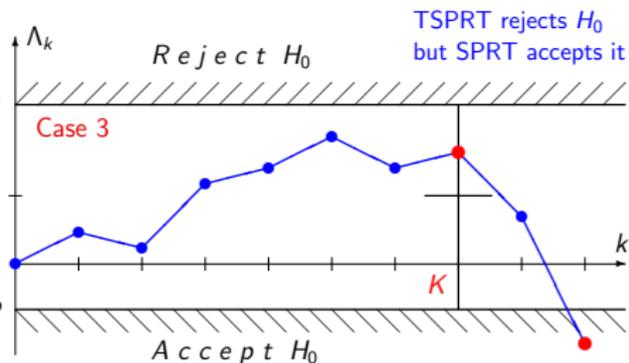
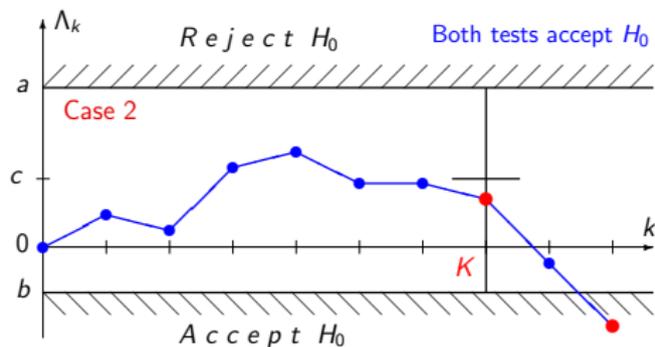
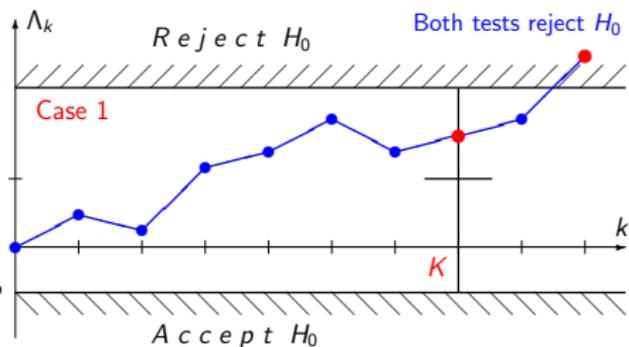
# Truncated SPRT (cont'd)



## Plan

- Evaluate performance of TSPRT, probabilities of Type I and Type II errors.
- Find **equations** of the stopping boundaries and the required group size to control  $P\{\text{Type I error}\}$  and  $P\{\text{Type II error}\}$ .
- Then develop Holm-type stepwise sequential procedures to test  $d$  hypotheses controlling **familywise error rates I and II**.

# Comparison of SPRT and TSPRT



# TSPRT and Error Probabilities

Start with  $(\alpha^*, \beta^*)$ -SPRT.

$P$  {Type I error by TSPRT}

$$\begin{aligned} &= P_{H_0} \{ \text{Type I error by SPRT} \} \\ &\quad + P_{H_0} (\{ \Lambda_1, \dots, \Lambda_{K-1} \in (b, a) \} \cap \{ \Lambda_K \in [c_1, a) \} \cap \{ \Lambda_k \leq b \text{ before } \Lambda_k \geq a \}) \\ &\quad - P_{H_0} (\{ \Lambda_1, \dots, \Lambda_{K-1} \in (b, a) \} \cap \{ \Lambda_K \in (b, c_1) \} \cap \{ \Lambda_k \geq a \text{ before } \Lambda_k \leq b \}) \\ &\leq \alpha^* + P_{H_0} \{ \text{Case 3} \} - P_{H_0} \{ \text{Case 4} \} \end{aligned}$$

where  $\Lambda_k = \sum_{j=1}^{km} \ln \frac{f(x_j | \theta^*)}{f(x_j | \theta_0)}$ .

Similarly,

$$P \{ \text{Type II error by TSPRT} \} \leq \beta^* + P_{H_A} \{ \text{Case 4} \} - P_{H_A} \{ \text{Case 3} \}$$

# Truncated SPRT and Optimal Error Spending

Choose  $c_1$  to attain  $P(\text{Type I error}) \leq \alpha$ ,

and  $c_2$  to attain  $P(\text{Type II error}) \leq \beta$ .

## Lemma

*For a sufficiently large group size  $m$ ,  $c_1 \leq c_2$ , and therefore, the union of the rejection region and the acceptance region for the statistic  $\Lambda_K$  at the truncation point  $K$  is  $\mathbf{R}$ .*

## Theorem

*The truncated group sequential procedure with error spending  $\alpha^*$ ,  $(\alpha - \alpha^*)$ ,  $\beta^*$ , and  $(\beta - \beta^*)$ , decision boundary  $c \in [c_1, c_2]$ , and the group size  $m$  guaranteed by Lemma controls the probability of Type I error at level  $\alpha$  and the probability of Type II error at level  $\beta$ .*

# TSPRT Performance under $H_0$

$(\alpha, \beta)$	$\delta$	Pollock-Golhar test			Zhao-Baron test		
		$m$	$E_{\theta_0}(mT)$	$P(T = K)$	$m$	$E_{\theta_0}(mT)$	$P(T = K)$
(0.01, 0.01)	0.5	55	69.85	0.270	49	57.03	0.164
(0.02, 0.02)	0.5	40	54.40	0.360	38	46.80	0.232
(0.05, 0.05)	0.5	24	35.52	0.480	25	32.98	0.319
(0.01, 0.02)	0.5	48	60.96	0.270	43	50.49	0.174
(0.05, 0.10)	0.5	22	31.02	0.410	21	26.68	0.270
(0.01, 0.01)	0.2	275	382.25	0.390	301	356.44	0.184
(0.02, 0.02)	0.2	215	309.60	0.440	236	292.52	0.239
(0.05, 0.05)	0.2	139	211.28	0.520	156	206.69	0.325
(0.01, 0.02)	0.2	274	361.68	0.320	267	315.54	0.182
(0.05, 0.10)	0.2	137	193.17	0.410	126	166.72	0.323

# TSPRT Performance under $H_A$

$(\alpha, \beta)$	$\delta$	Pollock-Golhar test			Zhao-Baron test		
		$m$	$E_{\theta_1}(mT)$	$P(T = K)$	$m$	$E_{\theta_1}(mT)$	$P(T = K)$
(0.01, 0.01)	0.5	55	69.85	0.270	49	57.03	0.164
(0.02, 0.02)	0.5	40	54.4	0.360	38	46.80	0.232
(0.05, 0.05)	0.5	24	35.52	0.480	25	33.07	0.323
(0.01, 0.02)	0.5	48	64.32	0.430	43	53.27	0.239
(0.05, 0.10)	0.5	22	33.22	0.510	20	28.53	0.427
(0.01, 0.01)	0.2	275	382.25	0.390	301	356.44	0.184
(0.02, 0.02)	0.2	215	309.60	0.440	236	292.52	0.239
(0.05, 0.05)	0.2	139	211.28	0.520	156	206.69	0.325
(0.01, 0.02)	0.2	274	380.86	0.390	268	332.92	0.242
(0.05, 0.10)	0.2	137	206.87	0.510	123	178.32	0.450

# Truncated Multiple Testing Scheme

## Stopping boundaries

$$a_j = -\log \frac{\alpha^*}{d-j+1}, \quad b_j = \log \frac{\beta^*}{j},$$

$$a_1 \geq \dots \geq a_d \geq b_1 \geq \dots \geq b_d, \text{ for } k < K, j = 1, \dots, d$$

Decision boundaries at  $k = K$ ,

$$c_1^{[1]} = -\log \frac{\alpha - \alpha^*}{d-j+1}, \quad c_2^{[j]} = \log \frac{\beta - \beta^*}{j}.$$

Group size  $m$  be the smallest satisfying  $\bigcap_{j=1}^d [c_1^{(j)}, c_2^{(j)}] \neq \emptyset$ ,

and in this case, the decision boundary at  $k = K$  is any number  $c$  that belongs to the intersection

$$c \in \bigcap_{j=1}^d [c_1^{(j)}, c_2^{(j)}].$$

# Truncated Multiple Testing Scheme

Stopping time  $T = \min \left\{ K, \min \left( k : \bigcap_{j=1}^d \Lambda_k^{[j]} \in (b_j, a_j) \right) \right\}$ ,

where  $\Lambda_k^{[j]}$  is the  $j$ -th largest log-likelihood ratio at time  $k$ .

If  $T < K$ , acceptance or rejection of each  $H_0^{(j)}$  is decided according to whether  $\Lambda_k^{[j]} \leq b_j$  or  $\Lambda_k^{[j]} \geq a_j$ .

At  $T = K$ , all the null hypotheses  $H_0^{(j)}$  corresponding to  $\Lambda_k^{(j)} \geq c$  are rejected, and all the others are accepted.

## Theorem

*The introduced stepwise truncated sequential multiple testing procedure guarantees simultaneous control of the Type I familywise error rate at level  $\alpha$  and the Type II familywise error rate at level  $\beta$  and stopping no later than at  $T = K$ .*

# Performance of Truncated Sequential Multiple Testing Scheme

Number of true null hypotheses	Expected number of groups	Expected sample size	Type I familywise error rate	Type II familywise error rate
Stepwise truncated group sequential procedure				
0	4.779	81.243	0	0.093
1	5.101	86.717	0.015	0.079
2	5.153	87.601	0.021	0.044
3	5.028	85.476	0.038	0.024
4	4.469	75.973	0.047	0
Bonferroni truncated group sequential procedure				
0	5.411	91.987	0	0.089
1	5.322	90.474	0.011	0.074
2	5.340	90.780	0.020	0.039
3	5.290	89.930	0.032	0.022
4	5.182	88.094	0.044	0

# Conclusions

Sequential schemes for multiple testing require smaller expected sample size.

Additional cost savings can be achieved by:

- Holm-type stepwise schemes
- Equalizer rules
- Optimal error spending and weighted statistics
- Controlling for generalized error rates
- Combination of the above

And still...

The actual FWER and GFWER for both sequential and non-sequential are still noticeably lower than the nominal levels due to the use of Markov/Bonferroni inequalities

- M. Baron and L. MacMillan. Weighted statistics for testing multiple endpoints in clinical trials. *Annals of Biostatistics and Biometric Applications*, 2(2):1–7, 2019.
- S. De and M. Baron. Sequential Bonferroni methods for multiple hypothesis testing with strong control of familywise error rates I and II. *Sequential Analysis*, 31(2):238–262, 2012.
- S. De and M. Baron. Step-up and step-down methods for testing multiple hypotheses in sequential experiments. *J. Statist. Plann. Inference*, 142:2059–2070, 2012.
- S. De and M. Baron. Sequential tests controlling generalized familywise error rates. *Statistical Methodology*, 23:88–102, 2015.
- T. Zhao and M. Baron. Multiple tests in group sequential clinical trials. *Jacobs Journal of Biostatistics*, 1(1), 2016.

- P. Armitage. Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis. *J. Roy. Statist. Soc. B*, 12:137–144, 1950.
- R. R. Bahadur. Rates of convergence of estimates and test statistics. *The Annals of Mathematical Statistics*, 38(2):303–324, 1967.
- J. Bartroff. Multiple hypothesis tests controlling generalized error rates for sequential data. *Statistica Sinica*, pages 363–398, 2018.
- J. Bartroff and T.-L. Lai. Multistage tests of multiple hypotheses. *Communications in Statistics - Theory and Methods*, 39:1597–1607, 2010.
- J. Bartroff and J. Song. Sequential tests of multiple hypotheses controlling type I and II familywise error rates. *Journal of Statistical Planning and Inference*, 153:100–114, 2014.
- J. Bartroff. Multiple hypothesis tests controlling generalized error rates for sequential data. *Statistica Sinica*, pages 363–398, 2018.
- C. W. Baum and V. V. Veeravalli. A sequential procedure for multihypothesis testing. *IEEE Trans. Inform. Theory*, 40:1994–2007, 1994.

- Y. Benjamini, F. Bretz, and S. Sarkar. *Recent Developments in Multiple Comparison Procedures*. IMS Lecture Notes - Monograph Series, Beachwood, Ohio, 2004.
- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc.*, 57(1):289–300, 1995.
- A. A. Borovkov and A. A. Mogulskii. Limit theorems in the boundary hitting problem for a multidimensional random walk. *Siberian Mathematical Journal*, 42(2):245–270, 2001.
- S. Chen and E. Arias-Castro. On the power of some sequential multiple testing procedures. *Annals of the Institute of Statistical Mathematics*, 73:311–336, 2021.
- A. Dmitrienko, A. C. Tamhane, and F. Bretz. *Multiple Testing Problems in Pharmaceutical Statistics*. CRC Press, Boca Raton, FL, 2010.
- B. Efron. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press, New York, 2010.

- M. Ghosh, N. Mukhopadhyay, and P. K. Sen. *Sequential Estimation*. Wiley, New York, 1997.
- S. Holm. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, 6:65–70, 1979.
- G. Hommel. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75(2):383–386, 1988.
- C. Jennison, I. M. Johnstone, and B. W. Turnbull. Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. In *S. S. Gupta and J. O. Berger, eds., Statistical Decision Theory and Related Topics III*, volume 2, pages 55–86, New York, 1982. Academic Press.
- C. Jennison and B. W. Turnbull. *Group sequential methods with applications to clinical trials*. Chapman & Hall, Boca Raton, FL, 2000.
- K. Kachiashvili. The methods of sequential analysis of bayesian type for the multiple testing problem. *Sequential Analysis*, 33(1):23–38, 2014.

- K. Kachiashvili. Constrained bayesian method for testing multiple hypotheses in sequential experiments. *Sequential Analysis*, 34(2):171–186, 2015.
- W. Maurer, L. Hothorn, and W. Lehmacher. Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypothesis. *Biomed. Chem.-Pharm. Ind.*, 6:3–18, 1995.
- A. Novikov. Optimal sequential multiple hypothesis tests. *Kybernetika*, 45(2):309–330, 2009.
- P. C. O'Brien and T. R. Fleming. A multiple testing procedure for clinical trials. *Biometrika*, 35:549–556, 1979.
- S. J. Pocock, N. L. Geller, and A. A. Tsiatis. The analysis of multiple endpoints in clinical trials. *Biometrics*, 43:487–498, 1987.
- S. K. Sarkar. Step-up procedures controlling generalized FWER and generalized FDR. *Ann. Stat.*, 35:2405–2420, 2007.
- J. P. Shaffer. Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, 81(395):826–831, 1986.

- Z. Sidak. Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Statist. Assoc.*, 62:626–633, 1967.
- D. Siegmund. A sequential clinical trial for comparing three treatments. *Ann. Stat.*, 21:464–483, 1993.
- D. Siegmund, B. Yakir, and N. R. Zhang. Detecting simultaneous variant intervals in aligned sequences. *Ann. Appl. Statist.*, 5:645–668, 2011.
- M. Sobel and A. Wald. A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *Ann. Math. Stat.*, 20(4):502–522, 1949.
- Y. Song and G. Fellouris. Asymptotically optimal, sequential, multiple testing procedures with prior information on the number of signals. *Electronic J. of Statistics*, 11:338–363, 2017.
- S. Zacks. *Stage-wise Adaptive Designs*. Wiley, Hoboken, NJ, 2009.



# INTERNATIONAL WORKSHOP IN SEQUENTIAL METHODOLOGIES

## JUNE 1 - 4, 2026

Focus areas include sequential and group sequential testing, adaptive designs, multi-stage inference, clinical trials, change-point detection, selection and ranking, machine learning, optimal stopping, reliability and survival analysis, stochastic approximation, and related fields of probability, statistics, and applications. Students and early-career researchers can apply for travel support. See the conference web site for details:

<https://www.american.edu/cas/iwsm2026/>

International Statistics  
conference in  
Washington DC

Latest developments in  
cutting-edge areas of  
sequential statistics and  
related fields

Applications in health,  
medicine, security,  
finance, genetics, and AI

Continuing a 20-year  
history of IWSM

Check the conference  
website for updates

### CONTACT THE ORGANIZING COMMITTEE:

Michael Baron

[baron@american.edu](mailto:baron@american.edu)

Yaakov Malinovsky

[yaakovm@umbc.edu](mailto:yaakovm@umbc.edu)

Sponsored  
by:



**2026 WNAR/IMS Annual Meeting**

NEW

**June 14–17, 2026. Pullman, Washington, USA****w** <https://wnar.org/wnar2026>

Washington State University in Pullman offers a beautiful summer setting for a conference, combining the energy of a vibrant campus with the charm of the surrounding Palouse region. Warm, sunny days highlight the rolling hills of wheat fields that create one of the most picturesque landscapes in the Pacific Northwest.

There will be short courses, a plenary lecture, invited and contributed sessions, young investigator events, and a Student Paper Award with oral sessions. Email [programchair@wnar.org](mailto:programchair@wnar.org) or [wnar@wnar.org](mailto:wnar@wnar.org) with questions.

**International Workshop in Sequential Methodologies****June 1–4, 2026****American University,  
Washington DC, USA****w** <https://www.american.edu/cas/iwsm2026/>*Now an IMS co-sponsored meeting.*

**Registration is open** for the ninth International Workshop in Sequential Methodologies (IWSM). This biannual conference will bring together researchers and practitioners to explore advances in sequential statistics, related areas of statistics and applied probability, and their many applications.

The technical program consists of theoretical and applied presentations in the areas of sequential testing, change-point detection, sequential estimation, selection and ranking, machine learning, artificial intelligence, clinical trials, adaptive design, stochastic quality and process control, optimal stopping, stochastic approximation, applied probability, mathematical finance, and related fields of probability, statistics, and applications.

The program features plenary lectures by leading experts in sequential statistics, including **Moshe Pollak** (Hebrew University), **Alexander Tartakovsky** (AGT StatConsult), **Dong-Yun Kim** (NIH), **Jay Bartroff** (University of Texas), and **Peihua Qiu** (University of Florida).

The regularly updated site has registration information, as well as on- and off-campus lodging reservations. **Early registration ends on April 1, 2026.**

Any questions? Innovative ideas, requests, or opportunities? Please contact the IWSM 2026 Organizing Committee: Michael Baron, American University ([baron@american.edu](mailto:baron@american.edu)), and Yaakov Malinovsky, University of Maryland, Baltimore County ([yaakovm@umbc.edu](mailto:yaakovm@umbc.edu)).

**SSP 2026: Seminar on Stochastic Processes**

UPDATED

**March 25–28, 2026. Union College, Schenectady, NY, USA****w** <https://www.math.union.edu/~marianop/SSP2026/>

SSP 2026 will be held March 26–28, 2026, with tutorial lectures delivered on March 25 by Timo Seppäläinen. The invited speakers will be **Saraí Hernández-Torres**, **Davar Khoshnevisan** (The Founders Lecturer), **Mateusz Kwaśnicki**, **Miklos Z. Racz**, and **Simon Tavaré**. **Registration is open now**, via the website above.

**International Symposium on Nonparametric Statistics (ISNPS 2026)****June 22–26, 2026, Thessaloniki, Greece****w** <https://easyconferences.eu/isnps2026/>

The International Symposium on Nonparametric Statistics (ISNPS 2026) will be held in Thessaloniki, Greece, June 22–26, 2026. This global forum will bring together researchers from around the world to exchange ideas, foster collaboration, and advance the fields of nonparametric statistics, data science and machine learning.

Building on the success of previous meetings, the 2026 symposium will feature plenary lectures, special invited sessions, contributed talks, and a dedicated student poster session. A student paper competition will be held within the poster session, with travel support awarded to the winners. Professor **Jianqing Fan** (Princeton University) will deliver the **Peter Hall Lecture**.

**The 4th Joint Conference on Statistics and Data Science (JCSDS 2026)**

NEW

**July 11–13, 2026. Guiyang, Guizhou China****w** <https://jcsds2026.scimeeting.cn/en/web/index/31392>

Jointly organized by the Chinese Association for Applied Statistics, Probability and Statistics Society of China, Association for Industrial Statistics Teaching, Business Statistics Society of China, the China Medical Association's Biostatistics Division and IMS–China. Since its inaugural meeting in 2023, JCSDS has become one of the world's largest gatherings in statistics and data science. The previous three meetings attracted 1800–2100 participants from 20+ countries. JCSDS typically has 6 keynote addresses, 100 invited sessions, and more than 50 contributed and poster sessions. **The 4th JCSDS will be staged together with the IMS–China biannual meeting**, with special sessions dedicated to the late Peter Hall, to mark his 10-year passing from us. In addition to the usual scholarly talks, it will have forums for Developing Statistics and Data Science in the era of AI, industry exhibitions, and extensive networking opportunities.

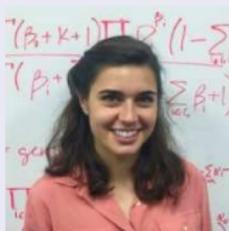
Important dates: Early-bird registration deadline May 16, 2026. Contributed talk & poster submission April 30, 2026. Accommodation booking deadline July 3, 2026.

We look forward to welcoming you to the beautiful “Forest City” of Guiyang in July 2026 for another unforgettable JCSDS!

# Acknowledgements

## *Thanks to:*

- Andreas Ziegler and Lilla Di Scala – for organizing and inviting
- Shyamal De, Laurel MacMillan, Elaine Esther Oruk Opyene, Tian Zhao – for collaboration and enthusiasm



- NSF - for sponsoring
- Everyone - for attending

Any questions?

# Non-sequential generalized Bonferroni, general case

## Bahadur efficiency

For the p-value  $p^{(j)}$  of the  $j$ -th test,

$$\liminf_{n_j \rightarrow \infty} \frac{1}{n_j} \log(p^{(j)}) \geq -K_A^{(j)}$$

with equality for LRT that rejects  $H_0^{(j)}$  when  $\Lambda_n^{(j)}$  is large.

Bahadur (1967)

Bonferroni multiple testing procedure based on  $\Lambda_n$ , with

$$\alpha_j = \exp(-c_\alpha K_A^{(j)}),$$

$c_\alpha$  being the unique solution of  $\sum_j \exp(-c_\alpha K_A^{(j)}) = \alpha$ , is asymptotically optimal in Bahadur sense, and it controls  $FWER_I \leq \alpha$ . Similarly, with  $\beta_j = \exp(-c_\beta K_0^{(j)})$ ,  $c_\beta$  solving  $\sum_j \exp(-c_\beta K_0^{(j)}) = \beta$ , controls  $FWER_{II} \leq \beta$ .