Decision analytic approach to subgroup analysis

Willi Maurer, Frank Bretz and Xiaolei Xun,

ADAPTIVE DESIGNS AND MULTIPLE COMPARISON PROCEDURES, June 25, Köln

NOVARTIS

Motivation

Familywise Error Rate (FWER)

- When testing a family of hypotheses, the conventional frequentist approach is to apply a multiple test procedure such that the FWER is strongly controlled at level $\alpha \in (0,1)$
  - FWER = P(incorrectly reject at least one true null hypothesis)

- Is controlling FWER always appropriate?

  Example: Compare two treatments with a control (e.g. two doses and placebo)

  - The trial is a success if either treatment is efficacious
  - FWER control seems appropriate

NOVARTIS

Motivation

Is controlling FWER always appropriate?

Example: Two disjoint subgroups (pre-specified)

- Global trial with different primary clinical endpoint in the US/EU
- Positive / negative biomarker
- Females / males

- FWER control seems to be inappropriate
  - The risk of a false decision is strictly restricted to each subgroup

Motivation

Controlling FWER is not always appropriate

- FWER ignores the differential impact / consequences of incorrect decisions (e.g. on patients, health care costs, society, ...)
- Incorrect and correct decisions may cause different losses and gains for different stakeholders
- Using a decision-analytic approach, incorrect rejections can be assigned different loss values
- This presentation focuses on subgroup analyses, but principles apply more generally

NOVARTIS

Decision Analytic Approach

Key components/steps

- Decision rules
- Loss and gain functions
- Expected losses and gains
- Optimal decision rules
  - Minimax regret decision rules
  - Bayesian decision rules

NOVARTIS

Decision Analytic Approach

Subgroup Analysis

- Consider testing for m subgroups the null hypotheses $H_i = \{\theta_i \leq 0\}$ against $K_i = \{\theta_i > 0\}$, $i \in M = \{1, \ldots, m\}$.
  - Indicator $\boldsymbol{r}$ of true state:

$$\boldsymbol{r}(\boldsymbol{\theta}) = \big(r(\theta_1), \ldots, r(\theta_m)\big):\ r(\theta_i) = \begin{cases} 0, & \text{if } \theta_i \in H_i, \\ 1, & \text{if } \theta_i \in K_i, \end{cases} \quad i \in M.$$

  - Decision $\boldsymbol{d}$ based on outcome

$$\boldsymbol{d} = (d_1, \ldots, d_m): \qquad d_i = \begin{cases} 0, & \text{if } H_i \text{ is retained}, \\ 1, & \text{if } H_i \text{ is rejected}, \end{cases} \quad i \in M.$$

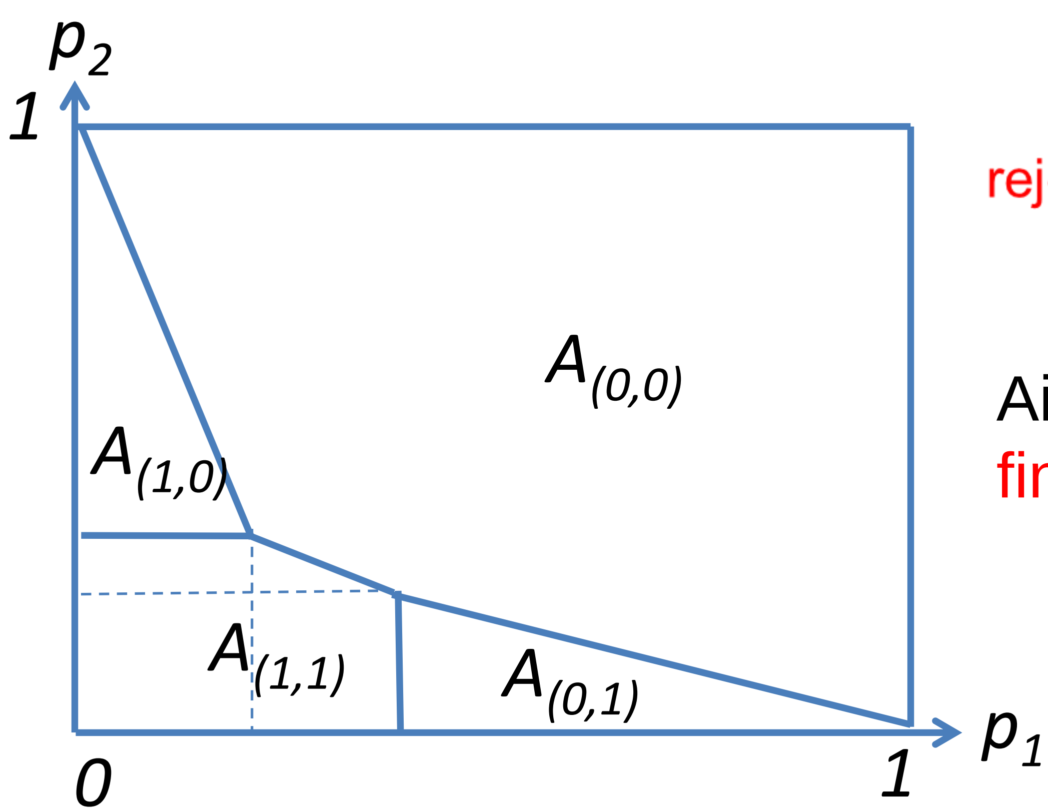
  - Outcomes summarized by p-values $\boldsymbol{p} = (p_1, \ldots, p_m)$
  - Decision rule $\boldsymbol{d}(\boldsymbol{p})$: $(0,1)^m \mapsto \{0,1\}^m$
  - A decision $d_i(\boldsymbol{p})$ in general depends on all p-values
  - If $d_i(\boldsymbol{p}) = d_i(p_i)$ for all $i \in M$, then we call $\boldsymbol{d}(\boldsymbol{p})$ separable.

NOVARTIS

Decision Rules for Two Subgroups

Based on P-values

- Monotonicity requirement for multiple test procedures
  - If H_i is rejected for $\boldsymbol{p} = (p_1, \dots, p_m)$, then it should also be rejected based on any $\boldsymbol{q}$ with $q_i \leq p_i$, $i \in M$. (Hommel and Bretz, 2008)

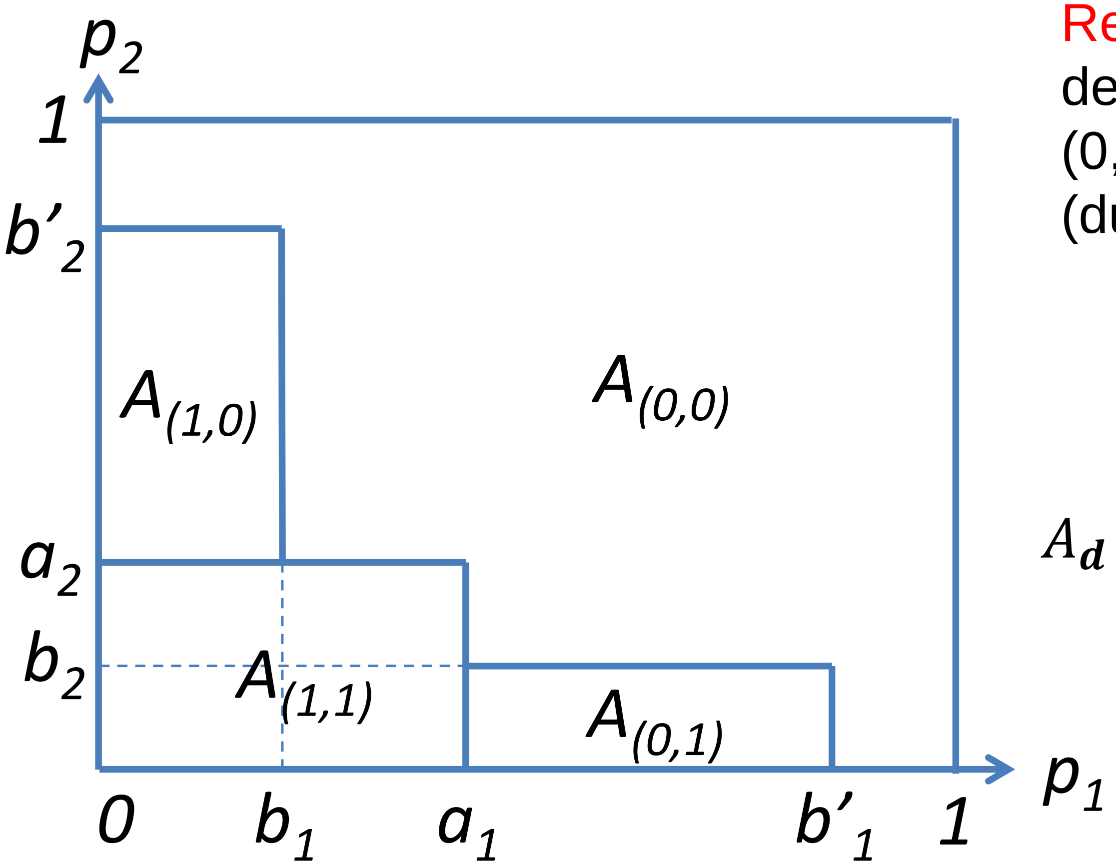


rejection areas $A_d = \{\boldsymbol{p}: \boldsymbol{d}(\boldsymbol{p}) = \boldsymbol{d}\}$

Aim:
find optimal decision rules

Rectangular Decision Rules for Two Subgroups

Based on P-values



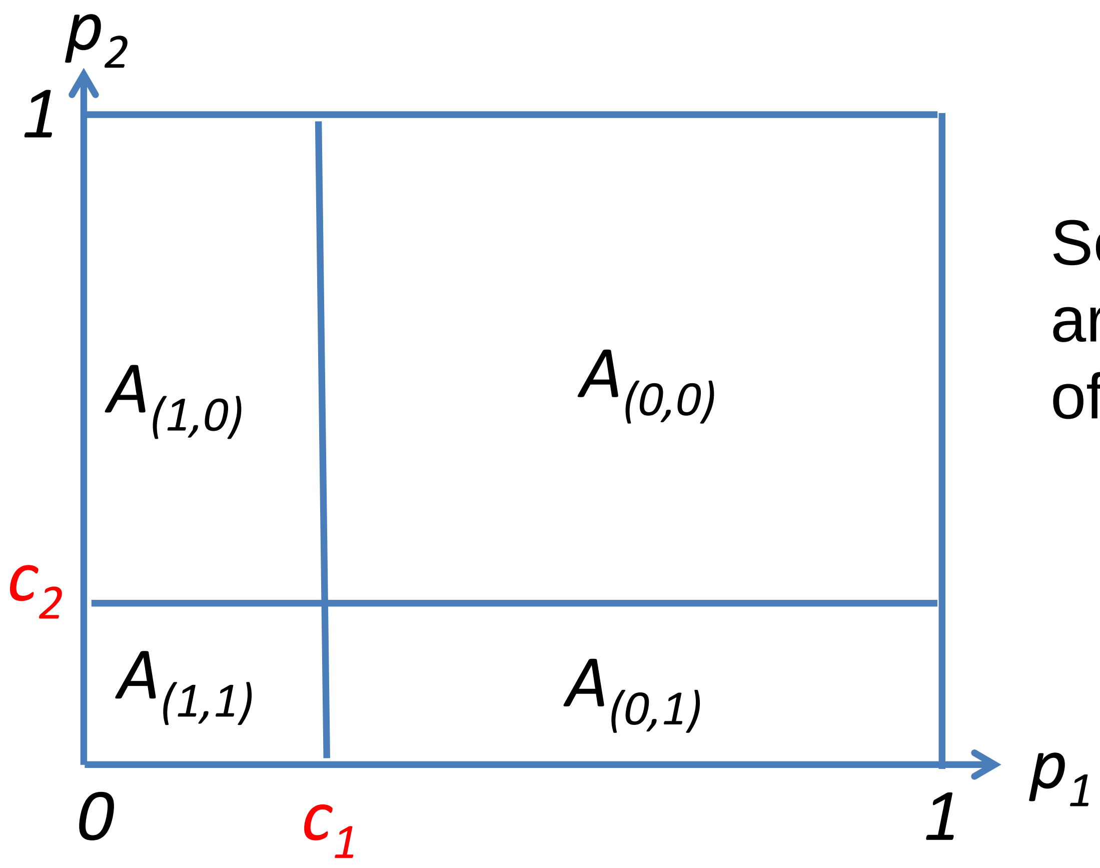
Rectangular Decision rules defined by $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{b}'$ **, where** $(0,0) \leq \boldsymbol{b} \leq \boldsymbol{a} \leq \boldsymbol{b}' \leq (1,1)$ (due to monotonicity).

$$A_{\boldsymbol{d}} = \begin{cases} A_{(1,1)}\text{: reject both} \\ A_{(1,0)}\text{: reject } H_1 \text{ only} \\ A_{(0,1)}\text{: reject } H_2 \text{ only} \\ A_{(0,0)}\text{: no rejection} \end{cases}$$

NOVARTIS

Separable Decision Rule for Two Subgroups

- Separable rules (rejection of H_i depends on p_i only)



Separable rules are a special case of rectangular rules

NOVARTIS

Loss and Gain functions

- Score step functions: $s\big(\boldsymbol{d}, \boldsymbol{r}(\boldsymbol{\theta})\big)$
  - depend on $\boldsymbol{\theta}$ only through $\boldsymbol{r}(\boldsymbol{\theta}) \in \{0,1\}$
- Loss step functions: a loss is incurred if $d = 1$ and $r = 0$
  - FWER $\ell\big(\boldsymbol{d}, \boldsymbol{r}(\boldsymbol{\theta})\big) = \max(d_i(1 - r(\theta_i))$
  - Additive loss function $\ell(\boldsymbol{d}, \boldsymbol{r}(\boldsymbol{\theta})) = \sum_{i=1}^{m} \ell_i \, d_i(1 - r(\theta_i))$
  - Standardization: $\max\Big(\ell\big(\boldsymbol{d}, \boldsymbol{r}(\boldsymbol{\theta})\big)\Big) = 1 \ \Rightarrow \sum_{i=1}^{m} \ell_i \ = 1$
- Gain step functions: A gain results if $d = 1$ and $r = 1$
  - Additive gain function $g\big(\boldsymbol{d}, \boldsymbol{r}(\boldsymbol{\theta})\big) = \sum_{i=1}^{m} g_i \, d_i r(\theta_i)$
- For subgroups: additive; loss and gain proportional to group size:
  - $\ell_i = \frac{n_i}{N}, \ g_i = c\frac{n_i}{N}$

NOVARTIS

Expected Losses and Gains

- **Expected score**:
- $E_{\boldsymbol{\theta}}[s(\boldsymbol{d}(\boldsymbol{p}), \boldsymbol{r}(\boldsymbol{\theta}))] = \sum_{\boldsymbol{d} \in A} s\,(\boldsymbol{d}, \boldsymbol{r}(\boldsymbol{\theta}))\; P_{\boldsymbol{\theta}}(A_{\boldsymbol{d}})$

  where $A = \{0,1\}^m$ and $A_{\boldsymbol{d}} = \{\boldsymbol{p}: \boldsymbol{d}(\boldsymbol{p}) = \boldsymbol{d}\}$
- **Expected loss and gain for additive step functions**
- $E_{\boldsymbol{\theta}}\left[\ell\left(\boldsymbol{d}(\boldsymbol{p}), \boldsymbol{r}(\boldsymbol{\theta})\right)\right] = \sum_{i=1}^{m} \ell_i\,(1 - r(\theta_i)) \sum_{\boldsymbol{d}: d_i = 1} P_{\boldsymbol{\theta}}(A_{\boldsymbol{d}})$
- $E_{\boldsymbol{\theta}}\left[g\left(\boldsymbol{d}(\boldsymbol{p}), \boldsymbol{r}(\boldsymbol{\theta})\right)\right] = \sum_{i=1}^{m} g_i\, r(\theta_i) \sum_{\boldsymbol{d}: d_i = 1} P_{\boldsymbol{\theta}}(A_{\boldsymbol{d}})$
- **Disjoint subgroups and rectangular regions $A_{\boldsymbol{d}}$:**
  - $P_{\boldsymbol{\theta}}(A_{\boldsymbol{d}})$ is the product of marginal probabilities dependent on $\boldsymbol{\theta}$, as well as on the univariate test statistics, variance and sample sizes
  - under the null H_i the marginal probability is the length of the interval on the p_i - axis
  - for separable rules the expected score only depends on the marginal distributions

Valid and Exhaustive Decision Rules

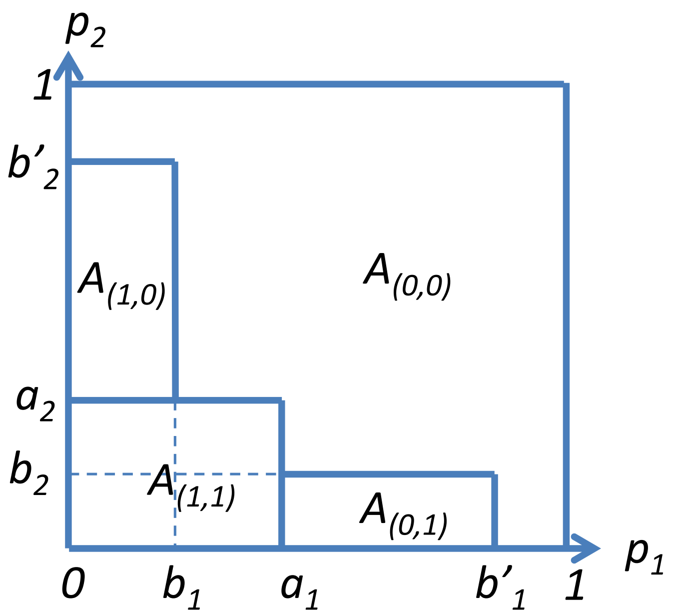
- In analogy to strong control of FWER, a decision rule $\boldsymbol{d}(\boldsymbol{p})$
  - is valid if the expected loss $E_{\boldsymbol{\theta}}\left[\ell\left(\boldsymbol{d}(\boldsymbol{p}), \boldsymbol{r}(\boldsymbol{\theta})\right)\right]$ does not exceed a pre-specified threshold α for any configuration of $\boldsymbol{\theta}$,
  - is exhaustive if there is a $\boldsymbol{\theta}$ for which $E_{\boldsymbol{\theta}}\left[\ell\left(\boldsymbol{d}(\boldsymbol{p}), \boldsymbol{r}(\boldsymbol{\theta})\right)\right] = \alpha$
  - Is inadmissible if there is a rule $\boldsymbol{d}'(\boldsymbol{p})$ with expected gain at least as large as that of $\boldsymbol{d}(\boldsymbol{p})$ and larger for at least one $\boldsymbol{\theta}$.
    - An inadmissible rule cannot be optimal

- Lemma: For monotonic decision rules and additive step loss and gain functions with positive weights l_i and g_i a non-exhaustive valid decision rule is inadmissible
  - Can be generalized to non-decreasing step functions:

$$s\left(\boldsymbol{d}', \boldsymbol{r}(\boldsymbol{\theta})\right) \geq s\left(\boldsymbol{d}, \boldsymbol{r}(\boldsymbol{\theta})\right) \text{ if } \boldsymbol{d}' > \boldsymbol{d}$$

Valid and exhaustive rectangular decision rules, m=2

- The rule $\boldsymbol{d}(\boldsymbol{p})$, parametrized by $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{b'}$ is valid if
  - $a_1 \leq \alpha/\ell_1, a_2 \leq \alpha/\ell_2$ and $a_1 a_2 + \ell_1 b_1 (b'_2 - a_2) + \ell_2 b_2 (b'_1 - a_1) \leq \alpha$
  - and exhaustive if at least one of the equations is strict.



In search of optimal rules we look at subclasses:

Symmetric rules B_{sym}

$a_1 = a_2 = x\alpha$ and $b_1 = b_2 = y\alpha$, $b'_1 = b'_2$

Valid: $0 < y \leq x \leq 1/\max(\ell_1, \ell_2)$

Asymmetric rules B_{asym}

$a_i = x\alpha/2\ell_i$, $b_i = y\alpha/2\ell_i$, $b'_1 = b'_2$

Valid: $0 < y \leq x \leq 2$

"Skew" rules $a_i = \alpha/\ell_i$, $b_i \leq a_i$, $b'_1 = b'_2 = 1$

Valid and Exhaustive Decision Rules

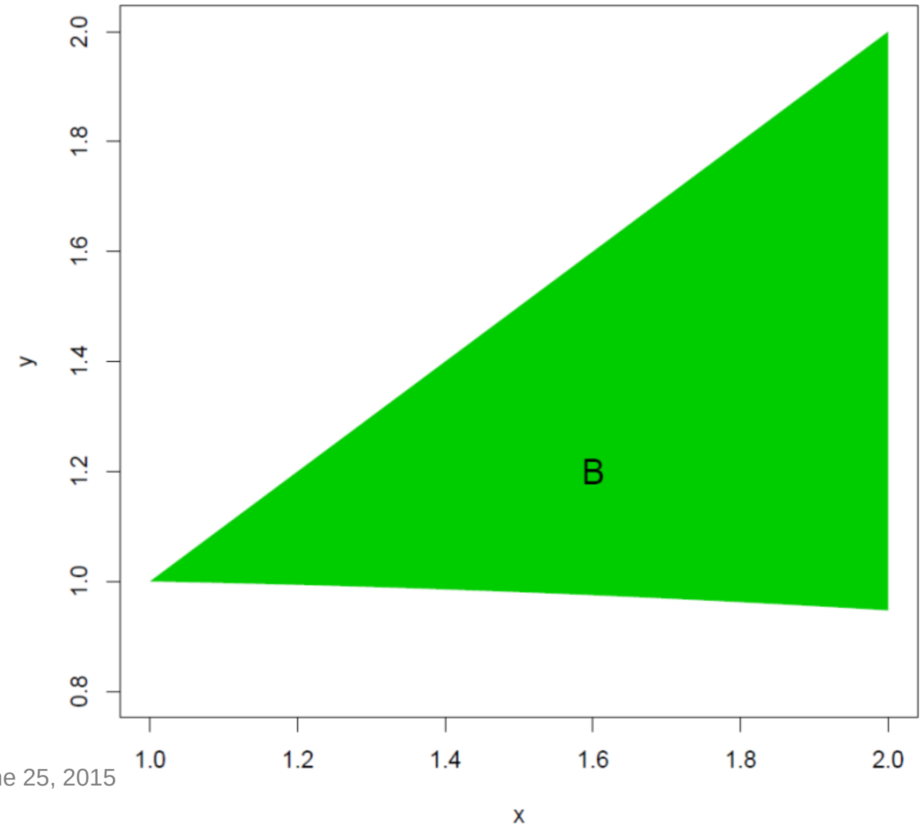
For $\ell_1 = \ell_2 = 0.5$; symmetric and asymmetric rule coincide

- For $a_1 = a_2 = x\alpha$ and $b_1 = b_2 = y\alpha$, $b_1' = b_2'$

  the set $B = \left\{(x,y): 1 \le x \le \frac{1}{\max(\ell_1,\ell_2)}, \frac{1-x^2\alpha}{1-x\alpha} \le y \le x\right\}$

  contains all valid and exhaustive rules

B for $\ell_1 = \ell_2 = 0.5$



- The search for optimal rules can be restricted to B since rules outside are inadmissible.

Optimal Decision Rule for m=2

Simple Hypotheses

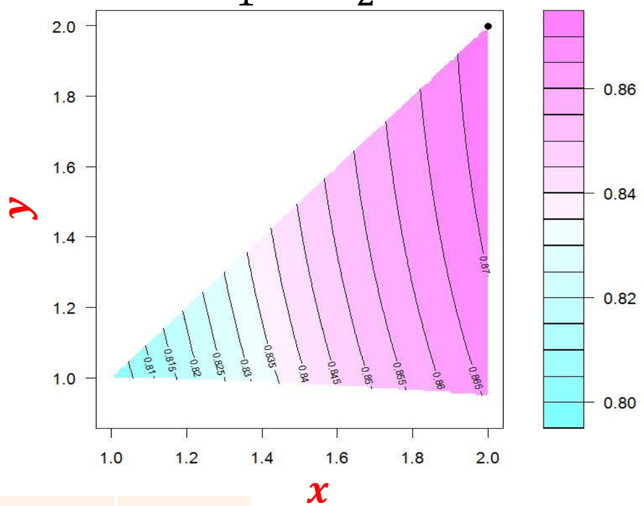
- For a class B of valid exhaustive decision rules parametrized by a vector $v = (x, y)$
  - let $v_{max}(\boldsymbol{\theta}^+)$ denote the vector that characterizes the rule that maximizes the expected gain for a fixed $\boldsymbol{\theta}^+ = (\theta_1^+, \theta_2^+)$ with $\theta_i^+ > 0$ for at least one $i = 1,2$.
  - The optimal rule is invariant under linear transformation of the gain function
    - There is no need that loss and gain are measured on the same scale
    - We standardize the maximum gain to 1
- We assume that the distribution of the two stochastically independent test statistics Z_i can be approximated by a normal distribution with variance 1 and noncentrality parameter $\delta_i(\theta_i^+, \sigma_i, n_i)$.

Examples of Optimal Decision Rule

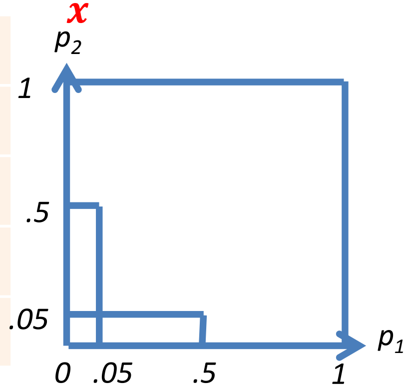
Simple Alternative $\boldsymbol{\theta}^+ = (\theta_1^+, \theta_2^+)$, $\delta_2(\theta_2^+, \sigma_2, n_2) = 2.80$, $P(p_2 < .025) = 0.8$

- Scenario: $\alpha = 0.025$, gains $g_1 = g_2 = 0.5$, losses $\ell_1 = \ell_2 = 0.5$

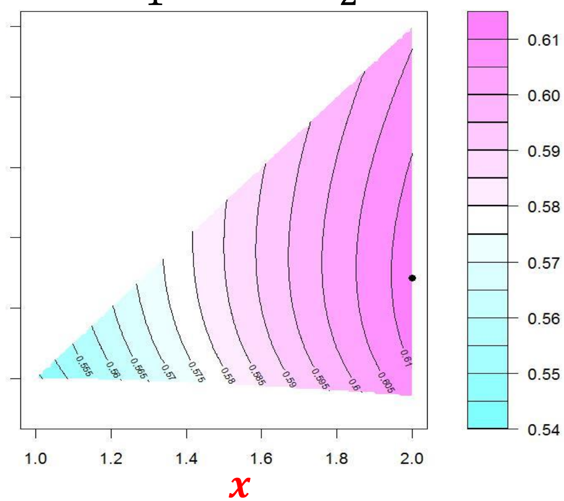
$\theta_1^+ = \theta_2^+$



x =	2
y =	2
a =	0.05
b =	0.05
b' =	.5

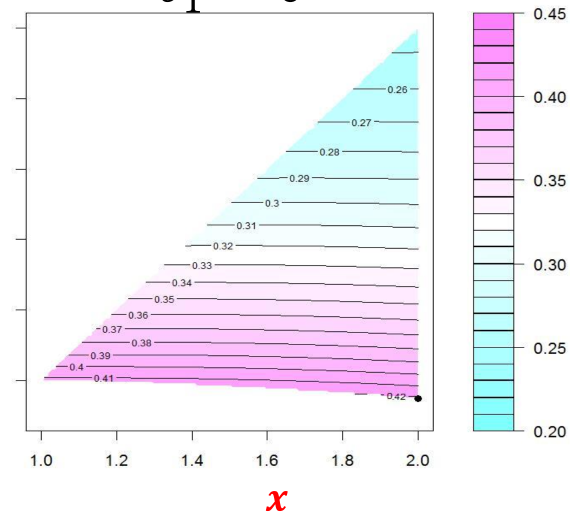


$\theta_1^+ = 0.5\ \theta_2^+$



x =	2
y =	1.28
a =	0.05
b =	0.032
b' =	0.794

$\theta_1^+ = 0$



x =	2
y =	0.95
a =	0.05
b =	0.023
b' =	1

NOVARTIS

Optimal Decision Rules for Unknown $\boldsymbol{\theta}$

Composite Hypotheses – Minimax Regret Decision Rule

- We use the abbreviation $w_{\boldsymbol{\theta}}(\boldsymbol{v}) = E_{\boldsymbol{\theta}}[g(\boldsymbol{d_v}(\boldsymbol{p}), \boldsymbol{\theta})]$

- Regret function for a decision rule $\boldsymbol{d_v}(\boldsymbol{p})$ given $\boldsymbol{\theta}$:

$$R_{\boldsymbol{\theta}}(\boldsymbol{d_v}(\boldsymbol{p})) = \max_{\boldsymbol{u} \in B} w_{\boldsymbol{\theta}}(\boldsymbol{u}) - w_{\boldsymbol{\theta}}(\boldsymbol{v})$$

- assigns for for a given $\boldsymbol{\theta}$ to each decision rule $\boldsymbol{d_v}(\boldsymbol{p})$ a non-negative value, expressing the regret for not having chosen the optimal procedure

- Minimax principle: the optimal decision rule that minimizes the maximum regret over all $\boldsymbol{\theta}$, then is characterized by

$$\boldsymbol{v}_{\text{opt.regret}} = \arg \min_{\boldsymbol{v} \in B} \max_{\boldsymbol{\theta} \in \Theta} R_{\boldsymbol{\theta}}(\boldsymbol{d_v}(\boldsymbol{p}))$$

Optimal Decision Rule

Composite Hypotheses – Bayes Decision Rule

- Bayes expected gain: for an assumed prior distribution $\Omega(\boldsymbol{\theta})$, the Bayesian expected gain of a decision $\boldsymbol{d}$ is

$$\tilde{G}(\boldsymbol{d}_{\boldsymbol{v}}(\boldsymbol{p})) = \int w_{\boldsymbol{\theta}}\,(\boldsymbol{v})\Omega(\boldsymbol{\theta})d$$

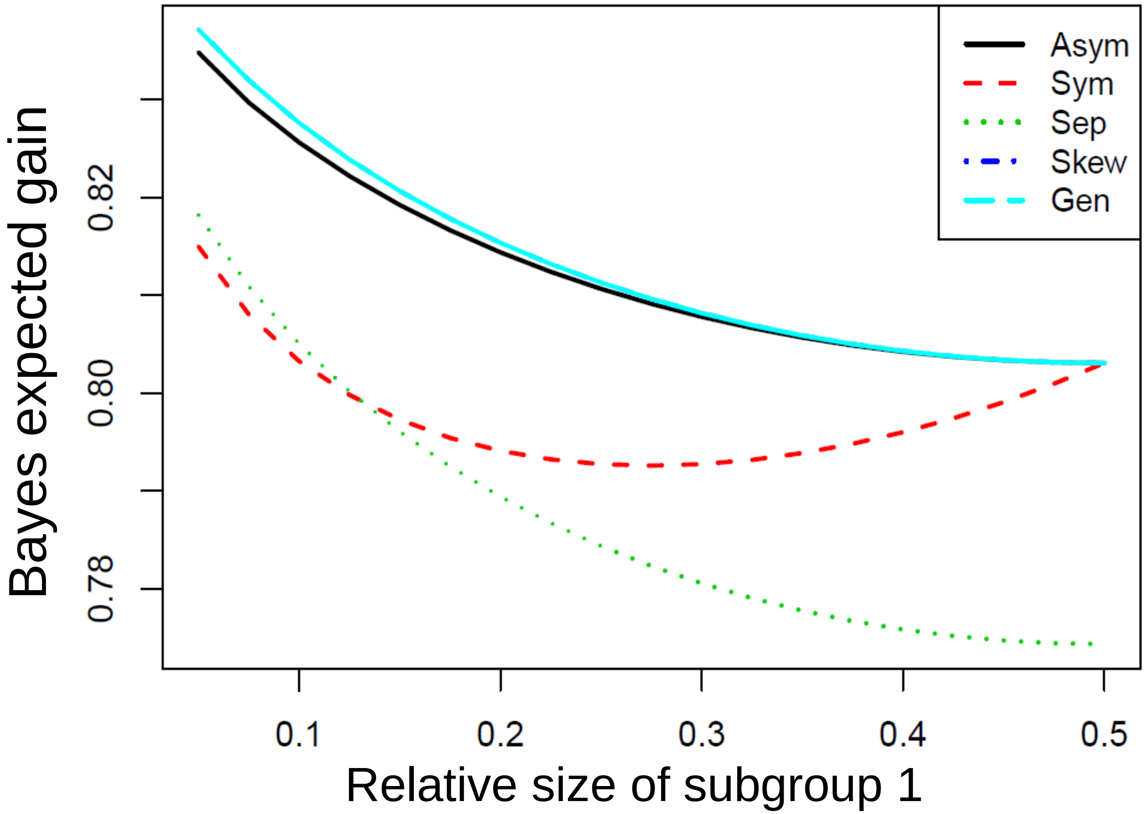
- Bayes principle: the optimal decision rule $\boldsymbol{v}_{\text{opt.Bayes}}$ then maximizes the above, i.e.,

$$\boldsymbol{v}_{\text{opt.Bayes}} = \arg \max_{\boldsymbol{v} \in B} \tilde{G}\,(\boldsymbol{d}_{\boldsymbol{v}}(\boldsymbol{p}))$$

Optimal Decision Rule

Composite Hypotheses – Bayes Approaches

- Optimal Bayes and minimax regret rules are almost identical if $P(\theta_1 > \theta_2) = 0.5$



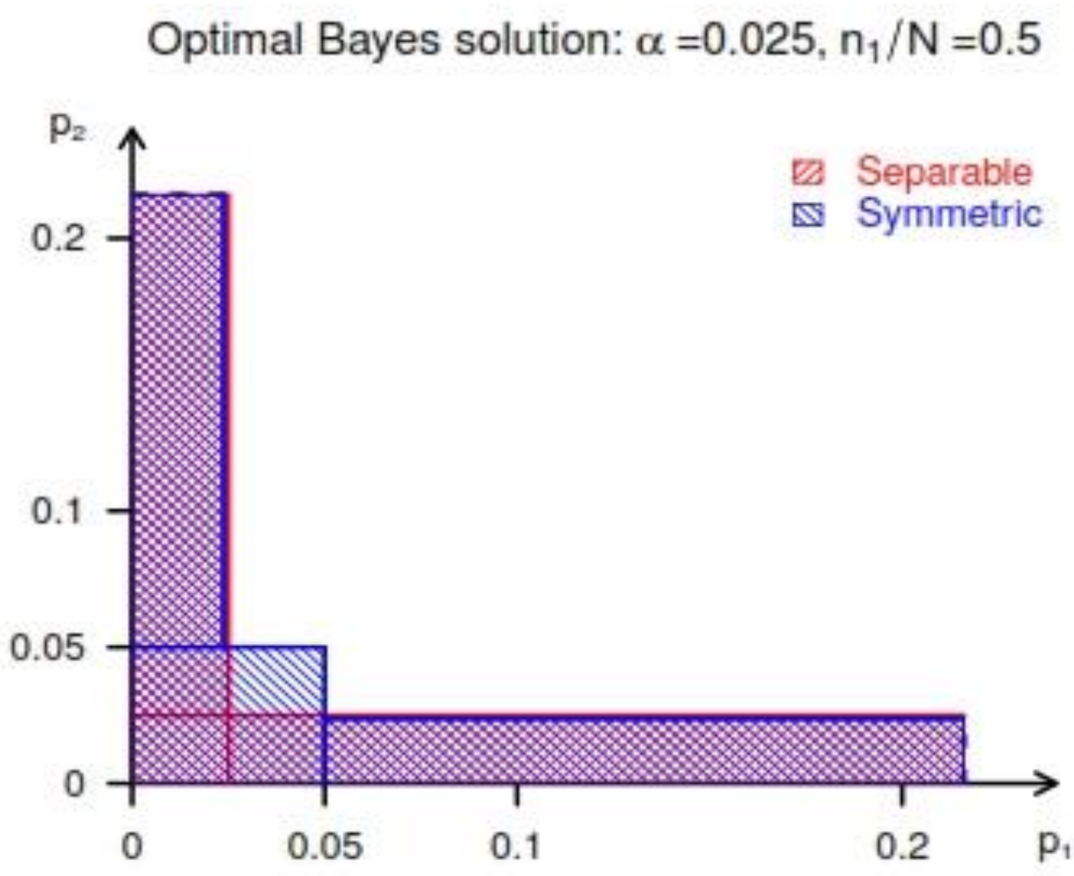
The unrestricted optimal Bayes rule (Gen)
is the optimal skew rule.

Separable and symmetric rules are clearly worse

NOVARTIS

Optimal Rule $n_1 = n_2$

Symmetric= Asymmetric = Skew are optimal

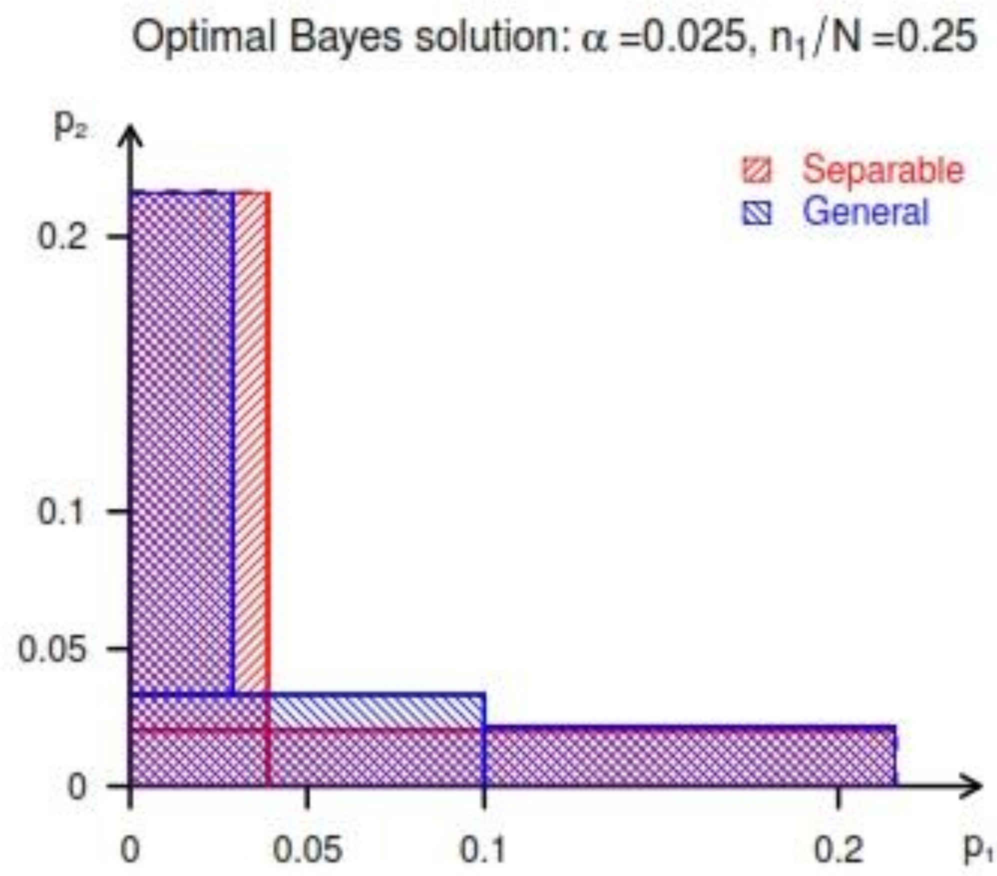


Optimal rule is similar to a Hochberg test but with boundaries about double as large !

NOVARTIS

Optimal Rule $n_1 = n_2/3$; $l_1 = ¼$ $l_2 = ¾$

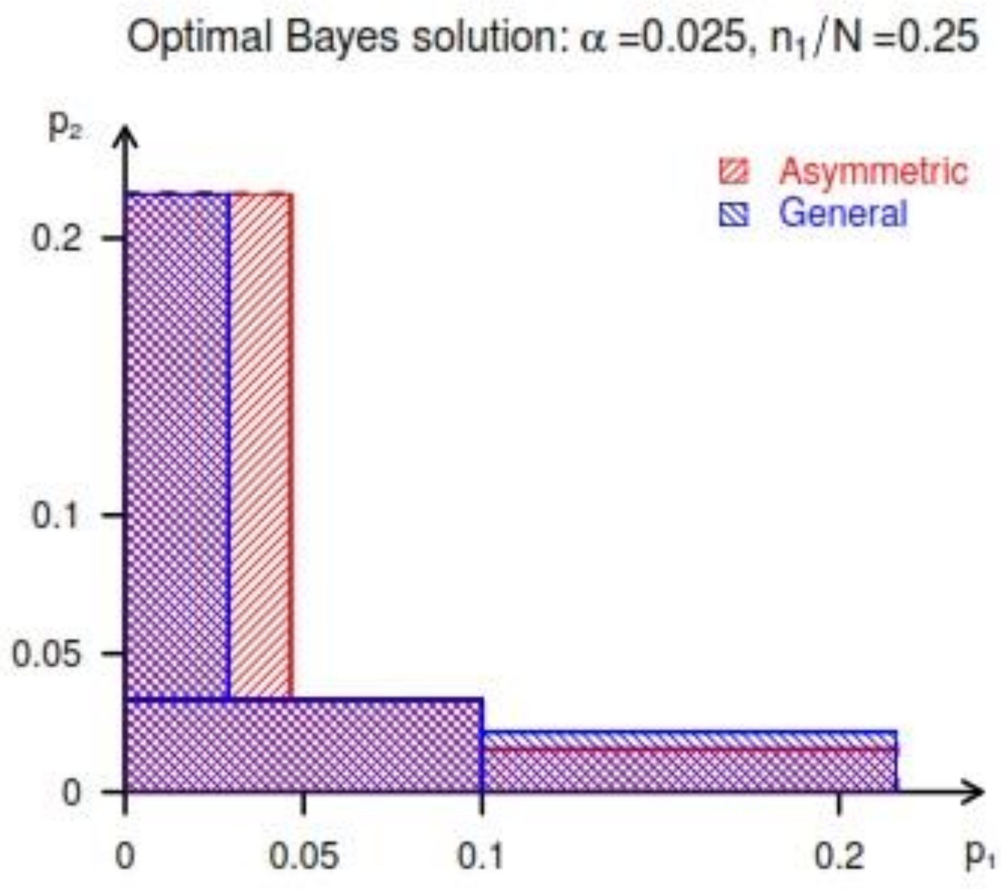
Generally optimal = Skew rule > asymmetric > symmetric > separable



The optimal rule allows to reject both hypotheses if the p of the larger subgroup is < 0.033 and p of the smaller subgroup is < 0.1

Optimal rule $n_1 = n_2 / 3$

Comparison of optimal asymmetric rule and generally optimal rule



Discussion

More Remarks

- The proposed generalized error rates can be applied to other situations, e.g.
  - testing two unequally important endpoints
  - granting conditional approval for a new drug
- The idea of constraining the expected loss by a pre-specified α and maximizing the expected gain is in line with current framework in regulatory decision making
- Future considerations
  - More general non-additive gain and loss functions
  - Relation to closed test principle and its generalization
  - Restricted prior distributions (e.g. $\theta_1^+ < \theta_2^+$)

NOVARTIS

Backup slides

Outline

- Motivation
- Generalized error rates and decision rules
- Examples
- Discussions

Loss Functions

For Incorrectly Rejecting H_i

- An additive loss function $\ell(\boldsymbol{d},\boldsymbol{\theta})$, where $\ell_1 + \ell_2 = 1$

$$\ell(\boldsymbol{d},\boldsymbol{r}(\boldsymbol{\theta})) = \sum_{i=1}^{2} \ell_i \, d_i\big(1 - r(\theta_i)\big)$$

Decision $\boldsymbol{d}$	Reject H_2	Retain H_2
Reject H_1	$\ell_1 + \ell_2 = 1$	ℓ_1
Retain H_1	ℓ_2	0

Example: Losses for all possible decisions when both nulls are true

- Similar tables need to be considered if exactly one null is true

Gain Functions

For Correctly Rejecting H_i

- An additive gain function $g(\boldsymbol{d}, \boldsymbol{\theta})$, where $g_1 + g_2 = 1$

$$g(\boldsymbol{d}, \boldsymbol{\theta}) = \sum_{i=1}^{2} g_i \, d_i r(\theta_i)$$

Decision $\boldsymbol{d}$	Reject H_2	Retain H_2
Reject H_1	$g_1 + g_2 = 1$	g_1
Retain H_1	g_2	0

Example: Gains for all possible decisions when both nulls are false

- Similar tables need to be considered if exactly one null is false

Expected Losses and Gains

- Define $\pi_{\boldsymbol{d}}(\boldsymbol{\theta}) = P(\boldsymbol{p} \in A_{\boldsymbol{d}})$, i.e. the probability that decision $\boldsymbol{d}$ is taken
  - Probability $\pi_d(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$, as well as on the univariate test statistics, variance and sample sizes
  - We consider z-tests
- Expected loss of the decision rule $D(\boldsymbol{p})$

$$E[\ell(D(\boldsymbol{p}), \boldsymbol{\theta})|\boldsymbol{\theta}] = \sum_{\boldsymbol{d}} \ell(\boldsymbol{d}, \boldsymbol{\theta}) \pi_{\boldsymbol{d}}(\boldsymbol{\theta})$$

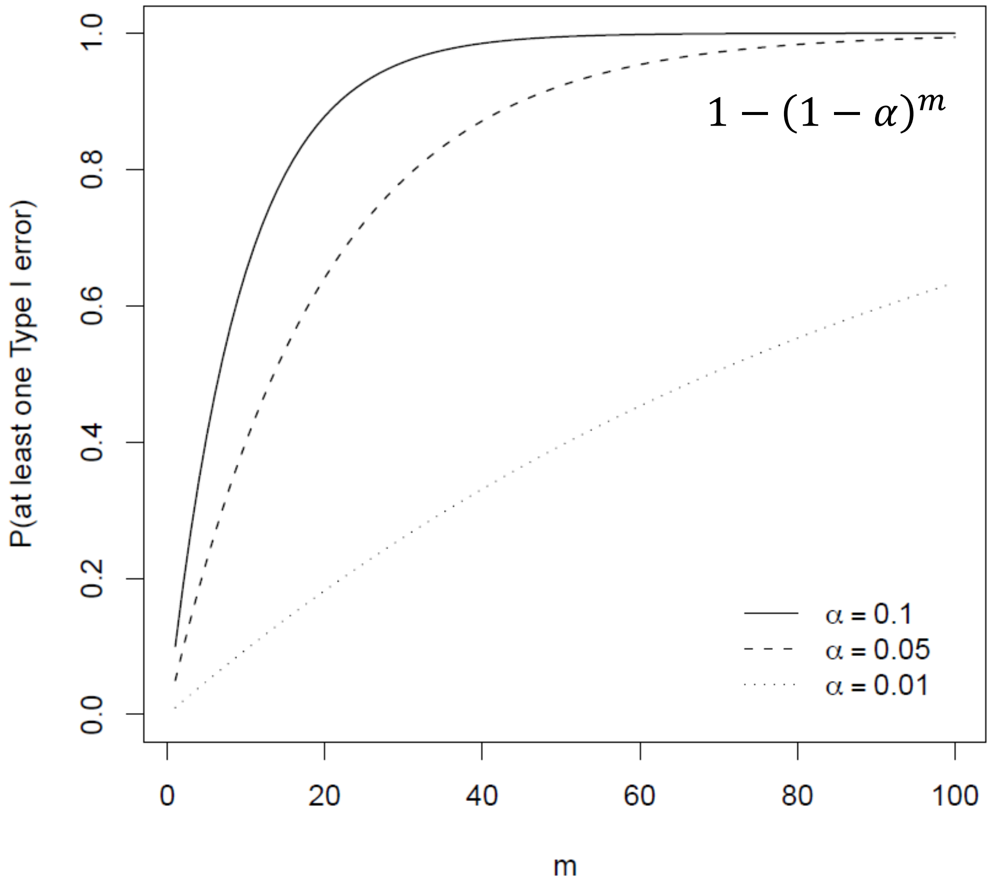
- Expected gain of the decision rule $D(\boldsymbol{p})$

$$E[g(D(\boldsymbol{p}), \boldsymbol{\theta})|\boldsymbol{\theta}] = \sum_{\boldsymbol{d}} g(\boldsymbol{d}, \boldsymbol{\theta}) \pi_{\boldsymbol{d}}(\boldsymbol{\theta})$$

Introduction

Type I Error Rate Inflation

Probability of at least one Type I error
For testing m independent hypotheses



- Probability of making Type I error increases as m or α increases
- For large m we almost surely reject incorrectly at least one of the true null hypotheses

Discussion

Relationship to FWER

- Non-additive loss function under $H_1 \cap H_2$:

Decision $\boldsymbol{d}$	Reject H_2	Retain H_2
Reject H_1	1	1
Retain H_1	1	0

- The expected loss is equivalent to FWER:

$$E[\ell(D(\boldsymbol{p}), \boldsymbol{\theta})|\boldsymbol{\theta}] = P(\text{reject at least one true null})$$