Multiple testing: evaluating their theoretical performance

M. Fromont*, M. Lerasle‡, P. Reynaud-Bouret‡

* *IRMAR*, Univ. Rennes 2,
‡ *CNRS, Laboratoire J.A. Dieudonné*, Univ. Côte d'Azur,

Contents

Contents

Contents

Contents

One single test

The basic Gaussian example

For $i = 1, ..., n$, we observe the X_i's such that

$$X_i = f_i + \epsilon_i,$$

with ϵ_i i.i.d. $\mathcal{N}(0, \sigma^2)$ and known σ.

Not : X and f corresponding vectors, P_f the distribution of X, $\mathcal{P} = \{P_f, f \in \mathbb{R}^n\}$.

One single test

The basic Gaussian example

For $i = 1, ..., n$, we observe the X_i's such that

$$X_i = f_i + \epsilon_i,$$

with ϵ_i i.i.d. $\mathcal{N}(0, \sigma^2)$ and known σ.

Not : X and f corresponding vectors, P_f the distribution of X, $\mathcal{P} = \{P_f, f \in \mathbb{R}^n\}$.

The basic single test problem

Is $f = 0$?

One single test

The basic Gaussian example

For $i = 1, ..., n$, we observe the X_i's such that

$$X_i = f_i + \epsilon_i,$$

with ϵ_i i.i.d. $\mathcal{N}(0, \sigma^2)$ and known σ.

Not : X and f corresponding vectors, P_f the distribution of X, $\mathcal{P} = \{P_f, f \in \mathbb{R}^n\}$.

The basic single test problem

Test H_0 : "$f = 0$" versus "$f \neq 0$"

One single test

The basic Gaussian example

For $i = 1, ..., n$, we observe the X_i's such that

$$X_i = f_i + \epsilon_i,$$

with ϵ_i i.i.d. $\mathcal{N}(0, \sigma^2)$ and known σ.

Not : X and f corresponding vectors, P_f the distribution of X, $\mathcal{P} = \{P_f, f \in \mathbb{R}^n\}$.

The basic single test problem

$H_0 = \{P_0\}$, P is the distribution of the observation and test "$P \in H_0$" versus "$P \notin H_0$".

Classical solution

- a statistical test Δ, which only depends on X, with value 0 (accept H_0) or 1 (reject H_0) is said of level $\alpha \in (0,1)$ if

$$\text{Type I error} = \sup_{P \in H_0} P(\Delta = 1) \leq \alpha$$

Classical solution

- a statistical test Δ, which only depends on X, with value 0 (accept H_0) or 1 (reject H_0) is said of level $\alpha \in (0, 1)$ if

$$\text{Type I error } = \sup_{P \in H_0} P(\Delta = 1) \leq \alpha$$

- *Example :* $\Delta = \mathbf{1}_{\sum_i X_i^2 > \sigma^2 c_\alpha}$ *with* c_α, *chi-square quantile of order* $1 - \alpha$ *with* n *degrees of freedom.*

Classical solution

- a statistical test Δ, which only depends on X, with value 0 (accept H_0) or 1 (reject H_0) is said of level $\alpha \in (0,1)$ if

$$\text{Type I error} = \sup_{P \in H_0} P(\Delta = 1) \leq \alpha$$

- $\Delta = \mathbf{1}_{T \geq F^{-1}(1-\alpha)}$ where F is the known c.d.f. of the test statistics T under H_0.

Classical solution

- a statistical test Δ, which only depends on X, with value 0 (accept H_0) or 1 (reject H_0) is said of level $\alpha \in (0,1)$ if

$$\text{Type I error} = \sup_{P \in H_0} P(\Delta = 1) \leq \alpha$$

- $\Delta = \mathbf{1}_{T \geq F^{-1}(1-\alpha)}$ where F is the known c.d.f. of the test statistics T under H_0.
- $\Delta = \mathbf{1}_{p \leq \alpha}$, where p is the p-value of the previous test, that is $1 - F(T)$ (if F continuous).

Classical solution

- a statistical test Δ, which only depends on X, with value 0 (accept H_0) or 1 (reject H_0) is said of level $\alpha \in (0,1)$ if

$$\text{Type I error} = \sup_{P \in H_0} P(\Delta = 1) \leq \alpha$$

- $\Delta = \mathbf{1}_{T \geq F^{-1}(1-\alpha)}$ where F is the known c.d.f. of the test statistics T under H_0.
- $\Delta = \mathbf{1}_{p \leq \alpha}$, where p is the p-value of the previous test, that is $1 - F(T)$ (if F continuous).
- *NB1 : Not equivalent. OK if distribution is continuous.*

Classical solution

- a statistical test Δ, which only depends on X, with value 0 (accept H_0) or 1 (reject H_0) is said of level $\alpha \in (0,1)$ if

$$\text{Type I error} = \sup_{P \in H_0} P(\Delta = 1) \leq \alpha$$

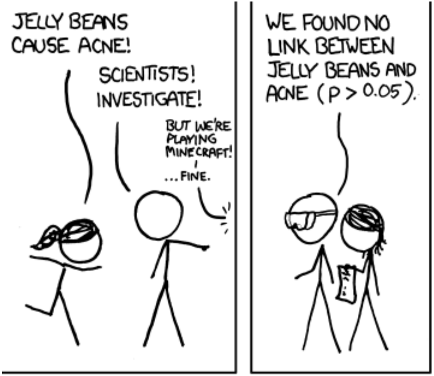
- $\Delta = \mathbf{1}_{T \geq F^{-1}(1-\alpha)}$ where F is the known c.d.f. of the test statistics T under H_0.
- $\Delta = \mathbf{1}_{p \leq \alpha}$, where p is the p-value of the previous test, that is $1 - F(T)$ (if F continuous).
- *NB1 : Not equivalent. OK if distribution is continuous.*
- *NB2: A p-value p always satisfies for any $P \in H_0$,*

$$\forall \alpha \in [0,1], P(p \leq \alpha) \leq \alpha.$$

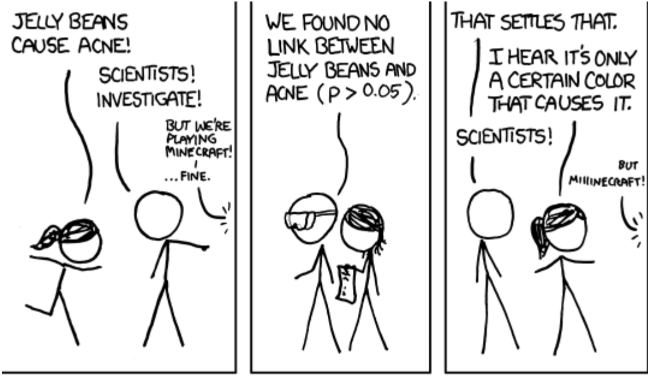
If $=$, p is uniform under H_0.

- Type II error $= \sup_{P \notin H_0} P(\Delta = 0)$

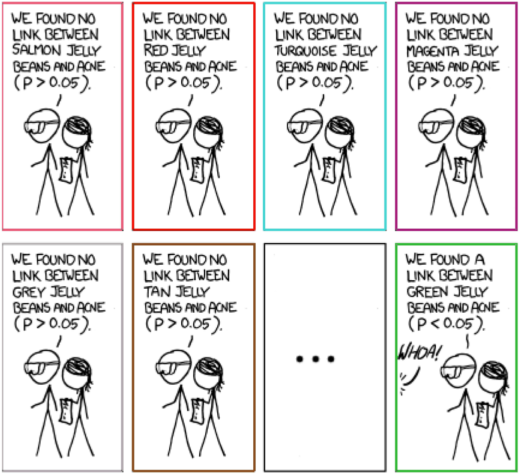
An intuition of multiple testing



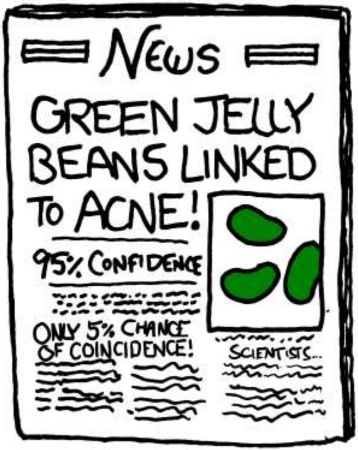
An intuition of multiple testing



An intuition of multiple testing



An intuition of multiple testing



Multiple testing (more seriously)

The basic multiple testing example

What are the non zero f_i?

Multiple testing (more seriously)

The basic multiple testing example

Test H_i: "$f_i = 0$" versus "$f_i \neq 0$" for all i

Multiple testing (more seriously)

The basic multiple testing example

Let $H_i = \{P_f / f_i = 0\}$ and test "$P \in H_i$" versus "$P \notin H_i$" for all i.

Multiple testing (more seriously)

The basic multiple testing example

Let $H_i = \{P_f / f_i = 0\}$ and test "$P \in H_i$" versus "$P \notin H_i$" for all i.

More generally,

- there is a family $\mathcal{H}$ of hypothesis H.
- Each hypothesis H is a subset of $\mathcal{P}$ the set of possible distributions.
- Outcome : a set of rejected hypothesis $\mathcal{R} \subset \mathcal{H}$, listing all the hypothesis that are not likely.
- Usually come from individual test for each H, with p-values p_H that are combined in a certain way.
- We hope that $\mathcal{R}$ is close to

$$\mathcal{F}(P) = \{H \in \mathcal{H} / P \notin H\},$$

the set of false hypothesis.

Aggregated tests

The basic aggregated test example

Several tests of H_0: "$f = 0$". For instance $\Delta_i = \mathbf{1}_{|X_i| > z_\alpha}$....
$\rightarrow$ How do we combine them ?

Aggregated tests

The basic aggregated test example

Several tests of H_0: "$f = 0$". For instance $\Delta_i = \mathbf{1}_{|X_i| > z_\alpha}$....
$\rightarrow$ How do we combine them ?

Link with multiple testing in general:

- $H_0 \subset \cap_{H \in \mathcal{H}} H$.
- a test per $H \in \mathcal{H}$, need to combine them to answer a single test question and control its type I error w.r.t. H_0 only.
- If multiple testing procedure gives an $\mathcal{R}$, then

reject H_0 if $\mathcal{R}$ non empty.

Aggregated tests

The basic aggregated test example

Several tests of H_0: "$f = 0$". For instance $\Delta_i = \mathbf{1}_{|X_i|>z_\alpha}$....
$\rightarrow$ How do we combine them ?

Link with multiple testing in general:

- $H_0 \subset \cap_{H\in\mathcal{H}} H$.
- a test per $H \in \mathcal{H}$, need to combine them to answer a single test question and control its type I error w.r.t. H_0 only.
- If multiple testing procedure gives an $\mathcal{R}$, then

reject H_0 if $\mathcal{R}$ non empty.

Control of Type II error for aggregated tests much more evolved than in multiple testing:

Our aim is to use control in aggregation to derive control in multiple testing.

First parallel between aggregation and multiple testing

Assume

- $H_0 = \cap_{H \in \mathcal{H}} H = \cap \mathcal{H}$,
- multiple testing procedure $\to \mathcal{R}$ in the family $\mathcal{H}$
- reject H_0 in an aggregated fashion, i.e. if $\mathcal{R}$ is non empty

First parallel between aggregation and multiple testing

Assume

- $H_0 = \cap_{H \in \mathcal{H}} H = \cap \mathcal{H}$,
- multiple testing procedure $\to \mathcal{R}$ in the family $\mathcal{H}$
- reject H_0 in an aggregated fashion, i.e. if $\mathcal{R}$ is non empty

Then

$$\text{Type I error of aggregated test} = \sup_{P \in \cap \mathcal{H}} P(\mathcal{R} \neq \emptyset).$$

First parallel between aggregation and multiple testing

Assume

- $H_0 = \cap_{H \in \mathcal{H}} H = \cap \mathcal{H}$,
- multiple testing procedure $\rightarrow \mathcal{R}$ in the family $\mathcal{H}$
- reject H_0 in an aggregated fashion, i.e. if $\mathcal{R}$ is non empty

Then

$$\text{Type I error of aggregated test} = \sup_{P \in \cap \mathcal{H}} P(\mathcal{R} \neq \emptyset).$$

$$= \text{weak Family-Wise Error Rate of } \mathcal{R} = w\mathrm{FWER}(\mathcal{R}).$$

It wants to guarantee a very weak Type I error :

$\mathcal{R}$ should be empty if $\mathcal{F}(P)$ is.

Family-Wise Error Rate

$$\mathrm{FWER}(\mathcal{R}) = \sup_{P \in \mathcal{P}} P(\mathcal{R} \cap \mathcal{T}(P) \neq \emptyset).$$

$$\text{with } \mathcal{T}(P) = \mathcal{F}(P)^c = \{H / P \in H\}.$$

- much stronger: control for all P and not just $P \in \cap\mathcal{H}$.
- if $\mathrm{FWER}(\mathcal{R}) \leq \alpha$, $\mathcal{R} \subset \mathcal{F}(P)$ very likely.
- does not say anything about $\mathcal{R} \simeq \mathcal{F}(P) \rightsquigarrow$ Type II error.

Some procedures with controlled FWER

Bonferroni

Perform a test Δ_H at level $\alpha/\#\mathcal{H}$ for each $H \in \mathcal{H}$,
$\mathcal{R} = \{H \ / \ \Delta_H \text{ rejects } \}$ is the set of rejected hypotheses.

$$P(\mathcal{R} \cap \mathcal{T}(P) \neq \emptyset) = P(\exists H \in \mathcal{T}(P), \Delta_H = 1)$$

$$\leq \sum_{H \in \mathcal{T}(P)} P(\Delta_H = 1) \leq \frac{\#\mathcal{T}(P)}{\#\mathcal{H}} \alpha \leq \alpha.$$

Some procedures with controlled FWER

Bonferroni

Perform a test Δ_H at level $\alpha/\#\mathcal{H}$ for each $H \in \mathcal{H}$,
$\mathcal{R} = \{H \ / \ \Delta_H \text{ rejects } \}$ is the set of rejected hypotheses.

- very conservative but always work.
- (almost) equivalent to say that one rejects all the H such that its p-value $p_H \leq \alpha/\#\mathcal{H}$.

Some procedures with controlled FWER

Bonferroni

Perform a test Δ_H at level $\alpha/\#\mathcal{H}$ for each $H \in \mathcal{H}$,
$\mathcal{R} = \{H \ / \ \Delta_H \text{ rejects } \}$ is the set of rejected hypotheses.

Min-p

If for all $\mathcal{G} \subset \mathcal{H}$, the c.d.f. $F_{\mathcal{G}}$ of $\min_{H\in\mathcal{G}} p_H$ does not depend on the $P \in \cap\mathcal{G}$, and if one knows $F_{\mathcal{H}}$, then

$$\mathcal{R} = \{H/p_H \leq F_{\mathcal{H}}^{-1}(\alpha)\}.$$

Some procedures with controlled FWER

Bonferroni

Perform a test Δ_H at level $\alpha/\#\mathcal{H}$ for each $H \in \mathcal{H}$,
$\mathcal{R} = \{H \ / \ \Delta_H \text{ rejects } \}$ is the set of rejected hypotheses.

Min-p

If for all $\mathcal{G} \subset \mathcal{H}$, the c.d.f. $F_{\mathcal{G}}$ of $\min_{H \in \mathcal{G}} p_H$ does not depend on the $P \in \cap\mathcal{G}$, and if one knows $F_{\mathcal{H}}$, then

$$\mathcal{R} = \{H/p_H \leq F_{\mathcal{H}}^{-1}(\alpha)\}.$$

$$\begin{aligned} P(\mathcal{R} \cap \mathcal{T}(P) \neq \emptyset) &= P(\exists H \in \mathcal{T}(P), p_H \leq F_{\mathcal{H}}^{-1}(\alpha)) \\ &= P(\min_{H \in \mathcal{T}(P)} p_H \leq F_{\mathcal{H}}^{-1}(\alpha)) \\ &= F_{\mathcal{T}(P)}(F_{\mathcal{H}}^{-1}(\alpha)) \leq F_{\mathcal{H}}(F_{\mathcal{H}}^{-1}(\alpha)) \leq \alpha. \end{aligned}$$

Some procedures with controlled FWER

Bonferroni

Perform a test Δ_H at level $\alpha/\#\mathcal{H}$ for each $H \in \mathcal{H}$,
$\mathcal{R} = \{H \ / \ \Delta_H \text{ rejects } \}$ is the set of rejected hypotheses.

Min-p

If for all $\mathcal{G} \subset \mathcal{H}$, the c.d.f. $F_\mathcal{G}$ of $\min_{H\in\mathcal{G}} p_H$ does not depend on the $P \in \cap\mathcal{G}$, and if one knows $F_\mathcal{H}$, then

$$\mathcal{R} = \{H/p_H \leq F_\mathcal{H}^{-1}(\alpha)\}.$$

- much less conservative and assumptions quite easy fulfilled in classical settings.
- aggregated version exists (with test statistics and not p-values) ...
- if one only knows that $F_\mathcal{H}$ does not depend on $P \in \mathcal{H}$, control of the $w\mathrm{FWER}$ only a priori.

Some procedures with controlled FWER

Bonferroni

Perform a test Δ_H at level $\alpha/\#\mathcal{H}$ for each $H \in \mathcal{H}$,
$\mathcal{R} = \{H \ / \ \Delta_H \text{ rejects } \}$ is the set of rejected hypotheses.

Min-p

If for all $\mathcal{G} \subset \mathcal{H}$, the c.d.f. $F_{\mathcal{G}}$ of $\min_{H \in \mathcal{G}} p_H$ does not depend on the $P \in \cap\mathcal{G}$, and if one knows $F_{\mathcal{H}}$, then

$$\mathcal{R} = \{H/p_H \leq F_{\mathcal{H}}^{-1}(\alpha)\}.$$

Step-Down

For both procedures, and many more, possible to reject a first set $\mathcal{R}_1$, do as if $\mathcal{H}_{new} = \mathcal{H} \setminus \mathcal{R}_1$ and iterate until no rejection anymore.

False Discovery Rate

$$\mathrm{FDR}(\mathcal{R}) = \left(\sup_{P\in\mathcal{H}}\right) \mathbb{E}_P\left[\frac{\#\mathcal{R}\cap\mathcal{T}(P)}{\#\mathcal{R}}\right],$$

with convention $0/0 = 0$.

False Discovery Rate

$$\mathrm{FDR}(\mathcal{R}) = \left(\sup_{P \in \mathcal{H}}\right) \mathbb{E}_P \left[\frac{\#\mathcal{R} \cap \mathcal{T}(P)}{\#\mathcal{R}} \right],$$

with convention $0/0 = 0$.

- if $\mathcal{F}(P) = \emptyset$, one "recovers" $w\mathrm{FWER}$, i.e.

$$w\mathrm{FWER}(\mathcal{R}) = \sup_{P \in \cap\mathcal{H}} P(\mathcal{R} \neq \emptyset) = \sup_{P \in \cap\mathcal{H}} \mathbb{E}_P \left[\frac{\#\mathcal{R} \cap \mathcal{T}(P)}{\#\mathcal{R}} \right].$$

- when $\mathcal{F}(P) \neq \emptyset$, less conservative :

$$\mathrm{FDR}(\mathcal{R}) \leq \mathrm{FWER}(\mathcal{R}).$$

Benjamini and Hochberg (BH) procedure

An exponentially increasing use and citations in very different domains since its parution in 1995.

1. sort the p_H: $p_{(1)} \leq p_{(2)} \leq \ldots$.
2. (Step-up algorithm) find the largest k, denoted $\hat{k}$, such that

$$p_{(k)} \leq \frac{k}{\#\mathcal{H}}\alpha.$$

3. $\mathcal{R} = \{H / p_H \leq p_{(\hat{k})}\}$.

Benjamini and Hochberg (BH) procedure

An exponentially increasing use and citations in very different domains since its parution in 1995.

1. sort the p_H: $p_{(1)} \leq p_{(2)} \leq \dots$.
2. (Step-up algorithm) find the largest k, denoted $\hat{k}$, such that

$$p_{(k)} \leq \frac{k}{\#\mathcal{H}}\alpha.$$

3. $\mathcal{R} = \{H / p_H \leq p_{(\hat{k})}\}$.

$$\mathrm{FDR}(\mathcal{R}) \leq \alpha \text{ if}$$

- the p_H are independent for all $H \in \mathcal{T}(P)$
- or PRDS property (positive regression dependency on each one from a subset): for Gaussian variables, positively correlated.

Benjamini and Hochberg (BH) procedure

An exponentially increasing use and citations in very different domains since its parution in 1995.

1. sort the p_H: $p_{(1)} \leq p_{(2)} \leq \dots$.
2. (Step-up algorithm) find the largest k, denoted $\hat{k}$, such that
$$p_{(k)} \leq \frac{k}{\#\mathcal{H}}\alpha.$$
3. $\mathcal{R} = \{H/p_H \leq p_{(\hat{k})}\}$.

$$\mathrm{FDR}(\mathcal{R}) \leq \alpha \text{ if}$$

- the p_H are independent for all $H \in \mathcal{T}(P)$
- or PRDS property (positive regression dependency on each one from a subset): for Gaussian variables, positively correlated.

Many variants exist... Most of the time one cannot prove the original BH works, but works in practice.

Many other Type I errors ...

- $k - \mathrm{FWER}$, PFER, FDP ...
- influence of the correlation structure, continuous versions etc
- for more, Etienne Roquain's Habilitation manuscript, or his review for SfdS journal.
- see also, Goeman and Solari (AoS 2010)

... Huge amount of papers interested in controlling that $\mathcal{R}$ does not intersect "too much" with $\mathcal{T}(P)$ (Type I error).

What about Type II errors ?

How do we measure that $\mathcal{R}$ is indeed a good approximation of $\mathcal{F}(P)$?

- the aggregated version (see also Romano, Shaikh and Wolf 2011)
  - transform $\mathcal{R}$ into a test of $\cap\mathcal{H}$ (rejected if $\mathcal{R} \neq \emptyset$)
  - say that the power of such test, i.e. $P(\mathcal{R} \neq \emptyset)$ if $P \notin \cap\mathcal{H}$ tends to 1.
- $P(\mathcal{R} \cap \mathcal{F}(P) \neq \emptyset)$ tends to 1 (Lehmann, Romano and Shaffer 2005).
- see $\mathcal{R}$ as a classification rule and measure its performance as a classifier (Neuvial, Roquain 2012)

Separation rates for a single test

Ingster (1993,...), Baraud (2005)

- Very efficient tool in large dimension or nonparametric problems to understand how much tests are powerful
- the "equivalent" of the risk theory in estimation
- give "rates" $\rightsquigarrow$ minimax theory
- depends on the smoothness of the class of alternatives

Separation rates for a single test

Ingster (1993,...), Baraud (2005)

- Very efficient tool in large dimension or nonparametric problems to understand how much tests are powerful
- the "equivalent" of the risk theory in estimation
- give "rates" $\rightsquigarrow$ minimax theory
- depends on the smoothness of the class of alternatives

Given a distance d on $\mathcal{P}$, α, β in $(0,1)$ and a smoothness class $\mathcal{Q} \subset \mathcal{P}$,

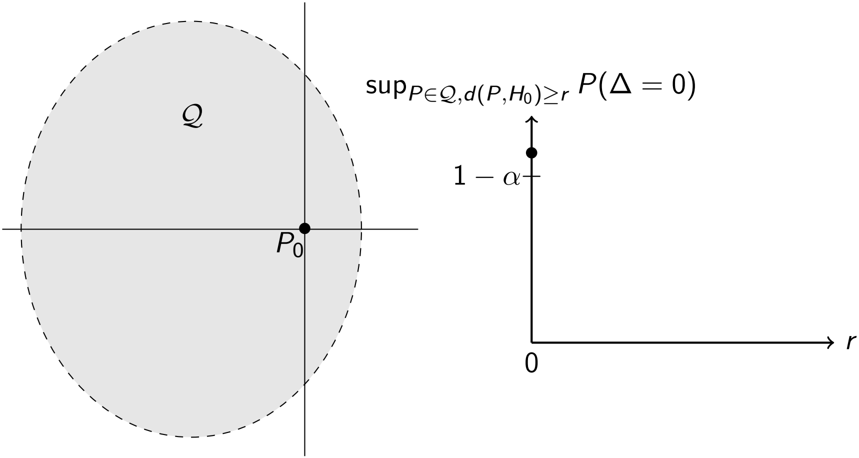
Uniform separation rate

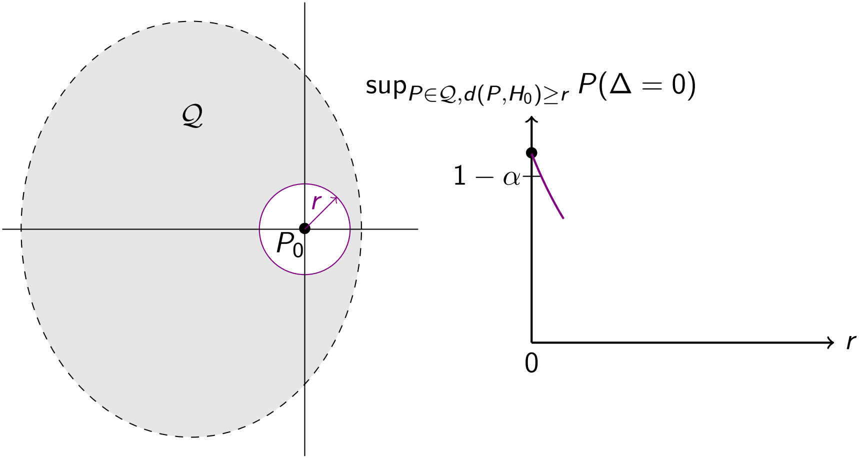
For a level α test Δ of H_0,

$$\mathrm{SR}_d^\beta(\Delta, \mathcal{Q}, H_0) = \inf\{r > 0 / \sup_{P \in \mathcal{Q} / d(P, H_0) \geq r} P(\Delta = 0) \leq \beta\}.$$

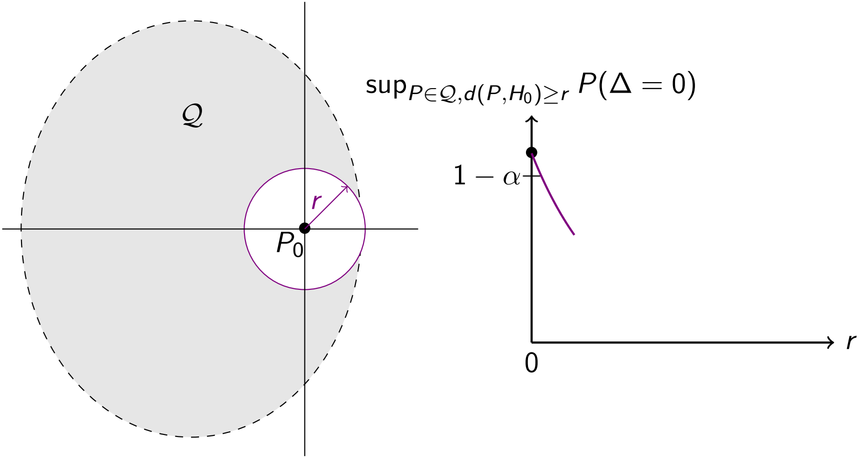
Minimax separation rate

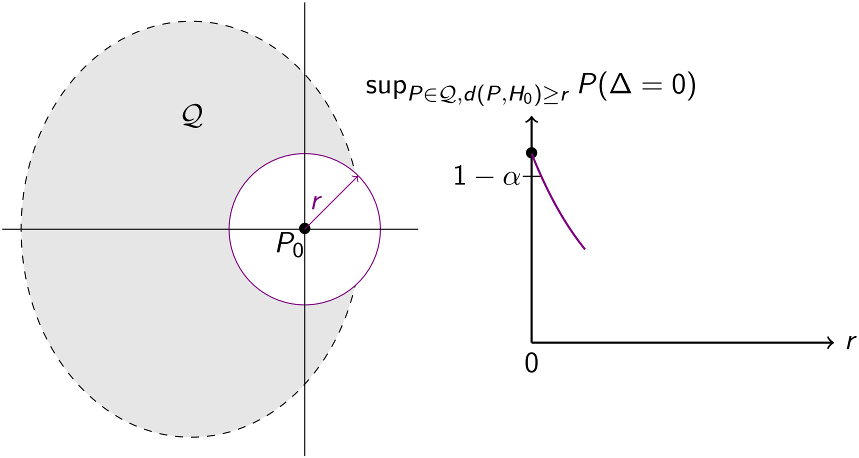
$$m\mathrm{SR}_d^{\alpha,\beta}(\mathcal{Q}, H_0) = \inf_{\Delta \text{with Type I error} \leq \alpha} \mathrm{SR}_d^\beta(\Delta, \mathcal{Q}, H_0).$$

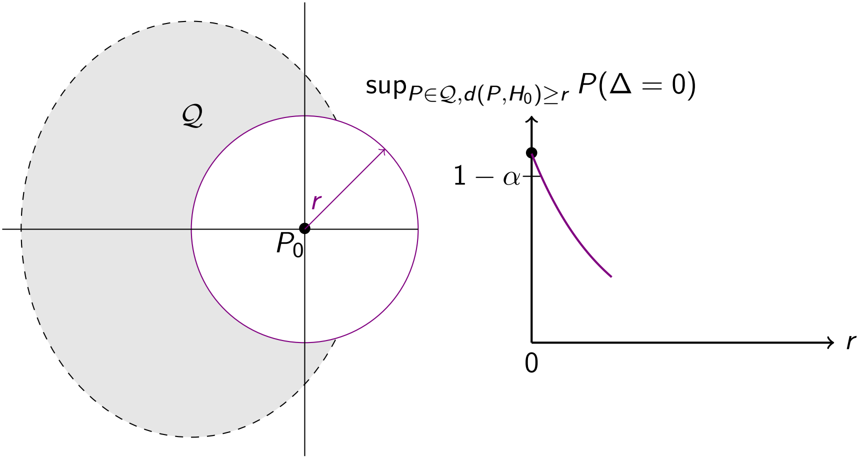

$\mathcal{Q}$
P_0
$\sup_{P\in\mathcal{Q},d(P,H_0)\geq r} P(\Delta = 0)$
$1-\alpha$
0
r

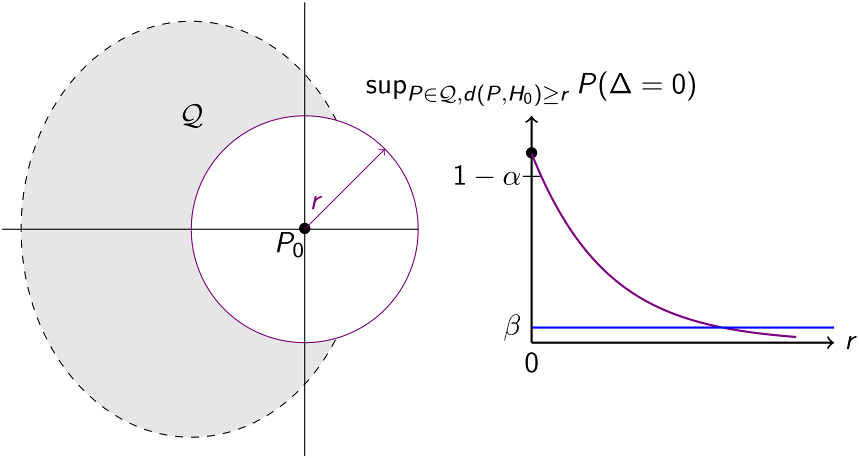

$\mathcal{Q}$
r
P_0
$\sup_{P \in \mathcal{Q}, d(P,H_0) \geq r} P(\Delta = 0)$
$1 - \alpha$
0
r

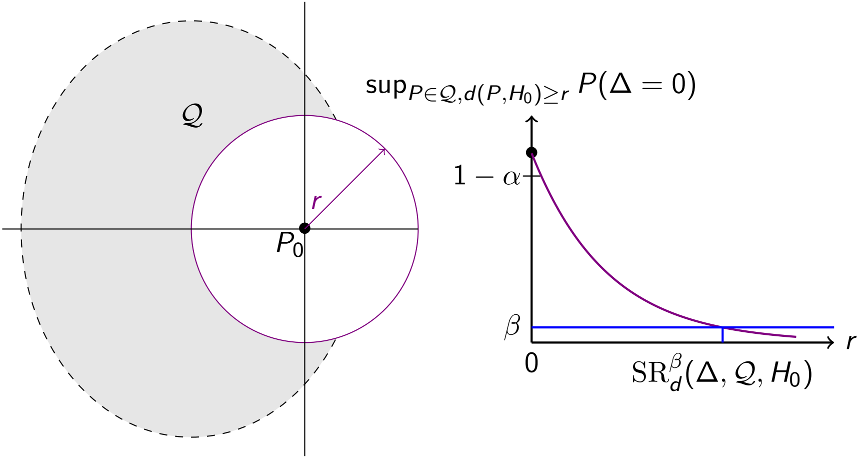
$\sup_{P \in \mathcal{Q}, d(P,H_0) \geq r} P(\Delta = 0)$




$\mathcal{Q}$
r
P_0
$\sup_{P\in\mathcal{Q},d(P,H_0)\geq r} P(\Delta = 0)$
$1-\alpha$
0
r

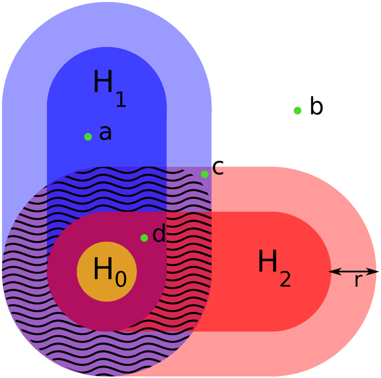

$\mathcal{Q}$
r
P_0
$\sup_{P \in \mathcal{Q}, d(P,H_0) \geq r} P(\Delta = 0)$
$1 - \alpha$
0
r


$\mathcal{Q}$
r
P_0
$\sup_{P\in\mathcal{Q},d(P,H_0)\geq r} P(\Delta = 0)$
$1-\alpha$
β
0
r


$\mathcal{Q}$
r
P_0
$\sup_{P\in\mathcal{Q},d(P,H_0)\geq r} P(\Delta = 0)$
$1-\alpha$
β
0
r
$\mathrm{SR}_d^\beta(\Delta,\mathcal{Q},H_0)$

Multiple testing version

$d(P, \cap\mathcal{H}) \geq r$ (single) vs $\mathcal{F}_r(P) = \{H/d(P,H) \geq r\}$ (multiple)



a: $\mathcal{T}(P) = \{H_1\}$ and $\mathcal{F}(P) = \mathcal{F}_r(P) = \{H_2\}$.
b: $\mathcal{T}(P) = \emptyset$ and $\mathcal{F}(P) = \mathcal{F}_r(P) = \{H_1, H_2\}$.
c: $\mathcal{T}(P) = \emptyset$, $\mathcal{F}(P) = \{H_1, H_2\}$, $\mathcal{F}_r(P) = \emptyset$, $d(P, H_1 \cap H_2) \geq r$.
d: $\mathcal{T}(P) = \{H_1, H_2\}$ and $\mathcal{F}(P) = \mathcal{F}_r(P) = \emptyset$, $P \notin H_0$.

Family-Wise Separation Rates (weak)

Weak Family Wise Separation Rate

$$w\mathrm{FWSR}_d^{\beta}(\mathcal{R},\mathcal{Q}) = \inf\{r > 0 / \sup_{P\in\mathcal{Q}/\mathcal{F}_r(P)\neq\emptyset} P(\mathcal{R}=\emptyset) \leq \beta\}$$

Family-Wise Separation Rates (weak)

Weak Family Wise Separation Rate

$$w\mathrm{FWSR}_d^{\beta}(\mathcal{R},\mathcal{Q}) = \inf\{r > 0 / \sup_{P \in \mathcal{Q}/\mathcal{F}_r(P) \neq \emptyset} P(\mathcal{R} = \emptyset) \leq \beta\}$$

Proposition

If $\Delta(\mathcal{R})$ is the aggregated test associated to $\mathcal{R}$,

$$w\mathrm{FWSR}_d^{\beta}(\mathcal{R},\mathcal{Q}) \leq \mathrm{SR}_d^{\beta}(\Delta(\mathcal{R}),\mathcal{Q},\cap\mathcal{H}).$$

Moreover, if

$$(\mathcal{A})\ \forall r > 0,\ \ d(P,\cap\mathcal{H}) \geq r \Leftrightarrow \mathcal{F}_r(P) \neq \emptyset,$$

then $w\mathrm{FWSR}_d^{\beta}(\mathcal{R},\mathcal{Q}) = \mathrm{SR}_d^{\beta}(\Delta(\mathcal{R}),\mathcal{Q},\cap\mathcal{H})$.

- $(\mathcal{A})$ true for closed family of hypotheses (in the sense of intersection).
- can also change the metric

Family-Wise Separation Rates (strong)

Family Wise Separation Rate

$$\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}) = \inf\{r > 0 / \sup_{P \in \mathcal{Q}} P(\mathcal{F}_r(P) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \emptyset) \leq \beta\}$$

Family-Wise Separation Rates (strong)

Family Wise Separation Rate

$$\mathrm{FWSR}_d^\beta(\mathcal{R},\mathcal{Q}) = \inf\{r>0 / \sup_{P\in\mathcal{Q}} P(\mathcal{F}_r(P)\cap(\mathcal{H}\setminus\mathcal{R})\neq\emptyset)\leq\beta\}$$

NB : Controlled FWER and FWSR guarantee that with large probability, for r larger than the FWSR

$$\mathcal{F}_r(P)\subset\mathcal{R}\subset\mathcal{F}(P).$$

- Never perfect with large probability in general : one cannot detect if too close.
- FWSR answers : how far away should P be from the H's so we can find those H's ?

Proposition

$$w\mathrm{FWSR}_d^\beta(\mathcal{R},\mathcal{Q})\leq\mathrm{FWSR}_d^\beta(\mathcal{R},\mathcal{Q})$$

A minimax theory

Minimax Family-Wise Separation Rate

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) = \inf_{\mathcal{R}\ /\ \mathrm{FWER}(\mathcal{R})\leq\alpha} \mathrm{FWSR}_d^{\beta}(\mathcal{R},\mathcal{Q}).$$

A minimax theory

Minimax Family-Wise Separation Rate

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) = \inf_{\mathcal{R} \ / \ \mathrm{FWER}(\mathcal{R}) \leq \alpha} \mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}).$$

Theorem

If $(\mathcal{A})$ holds, then $m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) \geq m\mathrm{SR}_d^{\alpha,\beta}(\mathcal{Q}, \cap\mathcal{H})$.

$\hookrightarrow$ natural idea that testing multiple hypotheses is more difficult than testing a single hypothesis.

$\hookrightarrow$ directly gives (tight) lower bounds for the minimax Family Wise Separation Rate over some $\mathcal{Q}$.

STOP $(\mathcal{A})$ is necessary! $\rightsquigarrow$ change of metric to make it work.

The classical Gaussian Example

The basic Gaussian example

For $i = 1, ..., n$, we observe the X_i's such that

$$X_i = f_i + \epsilon_i,$$

with ϵ_i i.i.d. $\mathcal{N}(0, \sigma^2)$ and known σ.

Not : X and f corresponding vectors, P_f the distribution of X, $\mathcal{P} = \{P_f, f \in \mathbb{R}^n\}$.

The classical Gaussian Example

The basic Gaussian example

For $i = 1, ..., n$, we observe the X_i's such that

$$X_i = f_i + \epsilon_i,$$

with ϵ_i i.i.d. $\mathcal{N}(0, \sigma^2)$ and known σ.

Not : X and f corresponding vectors, P_f the distribution of X, $\mathcal{P} = \{P_f, f \in \mathbb{R}^n\}$.

- d_s is the ℓ^s distance on the f's
- $\mathcal{Q} = \mathcal{P}_k = \{P_f / \#\{i, f_i \neq 0\} \leq k\}$: smoothness= sparsity (here)
- $H_0 = \{P_0\}$
- $m\mathrm{SR}_{d_2}^{\alpha,\beta}(\mathcal{P}_k, H_0) \sim \sigma\sqrt{k \ln(n)}$ if $k \leq n^\gamma$ with $\gamma \in (0, 1/2)$ (Baraud 2005)

Some applications (1)

If $H_i = \{P_f / f_i = 0\} \rightsquigarrow \mathcal{H}$,

- $m\mathrm{FWSR}_{d_2}^{\alpha,\beta}(\mathcal{P}_k) \sim \sigma\sqrt{\ln(n)}$ for all $k = 1, ..., n$.
- much smaller than $m\mathrm{SR}_{d_2}^{\alpha,\beta}(\mathcal{P}_k, H_0)$
- the "good" metric is d_∞, which guarantees $(\mathcal{A})$.
- achieved by Bonferroni, Min-p and their step-down versions based on single tests of the form $\mathbf{1}_{|X_i|/\sigma > ...}$
- methods for $m\mathrm{FWSR}$ have nothing to do with the methods used to achieve $m\mathrm{SR}$

Some applications (2)

If $H_i = \{P_f / f_1 = ... = f_i = 0\} \rightsquigarrow \mathcal{H}$ (closed family),

- $m\mathrm{FWSR}_{d_s}^{\alpha,\beta}(\mathcal{P}_k) \sim \sigma\sqrt{k\ln(n)}$
- achieved by variant of the closure method (Romano, Wolf 2005), with levels corrected in a Bonferroni fashion.

For more dependent structure, one can prove that Min-p version are strictly better in terms of rates Bonferroni like methods.

Conclusion

- many Type I errors for multiple testing in the literature
- few Type II errors studies
- we propose to use separation rates in a FamilyWise sense to better understand procedures guaranteeing FWER
- not possible right now to see the gain with step-down (maybe the constants ?)
- nothing yet for *FDR*

Thank you !

Fromont, M. Lerasle, M., Reynaud-Bouret, P. *Family Wise Separation Rates for multiple testing*, to appear in Annals of Statistics