Family Wise Separation Rates for multiple testing

Magalie Fromont (IRMAR, Université Rennes, France)

Joint work with Matthieu Lerasle and Patricia Reynaud-Bouret
(Université Nice Sophia Antipolis, France)

These are exciting times for statisticians [...]

The classical theory of hypothesis testing was fashioned for a scientific world of single inferences, small data sets, and slow computation. Exciting advances in technology have changed the equation.

Bradley Efron
(Doing thousands of hypothesis tests at the same time, 2007)

Setting of the problem
Goal of the present work

Testing problems: Type I error criteria

Single tests	Multiple tests
First Kind Error Rate or Size ER	Weak Family Wise Error Rate wFWER
	FWER
	kFWER, TPFDP (Lehmann, Romano 2005) (Romano, Wolf 2010...)
	FDR (Benjamini, Hochberg 1995)

Setting of the problem
Goal of the present work

Testing problems: Type II error criteria

Single tests	**Multiple tests**
Second Kind Error Rate Power	Minimal, Global, Average power (Romano, Wolf 2005...) Maximin (Lehmann, Romano, Schaffer 2005)
Separation Rates SR (for tests controlling the ER) (Ingster 1993, Baraud 2005)	

Setting of the problem
Goal of the present work

Testing problems: Type II error criteria

Single tests	Multiple tests
Second Kind Error Rate Power	Minimal, Global, Average power (Romano, Wolf 2005...) Maximin (Lehmann, Romano, Schaffer 2005)
Separation Rates SR (for tests controlling the ER) (Ingster 1993, Baraud 2005)	

Main concern: Definition of Family Wise Separation Rates for multiple tests strongly controlling the FWER.

Setting of the problem
Multiple testing

Let X be an observed r. v., with unknown distribution $P \in \mathcal{P}$.
Following Goeman and Solari (2010), a hypothesis H is defined as a subset of $\mathcal{P}$, and:

- H is true under $P \Leftrightarrow P \in H$,
- H is false under $P \Leftrightarrow P \notin H$.

Let $\mathcal{H}$ be a collection of hypotheses H, $\#\mathcal{H} = N$.

The aim is simultaneously testing, for all H in $\mathcal{H}$, H v.s. $\mathcal{P} \setminus H$
$\Leftrightarrow$ "H is true under P" v.s. "H is false under P",
$\Leftrightarrow$ "$P \in H$" v.s. "$P \notin H$".

Set of true hypotheses under P: $\mathcal{T}(P) = \{H \in \mathcal{H} / P \in H\}$,
Set of false hypotheses under P: $\mathcal{F}(P) = \{H \in \mathcal{H} / P \notin H\}$.

Setting of the problem
Multiple testing : $w\mathrm{FWER}$ and FWER

A multiple test is a collection of rejected hypotheses $\mathcal{R}^* \subset \mathcal{H}$, depending on X, whose goal is to infer $\mathcal{F}(P)$.

Weak Family-Wise Error Rate of $\mathcal{R}^*$

$w\mathrm{FWER}(\mathcal{R}^*) = \sup_{P/\mathcal{T}(P)=\mathcal{H}} P(\mathcal{R}^* \cap \mathcal{T}(P) \neq \emptyset)$.

(Strong) Family-Wise Error Rate of $\mathcal{R}^*$

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

Given a prescribed α in $(0,1)$, to construct a multiple test $\mathcal{R}^*$ such that $\mathrm{FWER}(\mathcal{R}^*) \leq \alpha \Rightarrow w\mathrm{FWER}(\mathcal{R}^*) \leq \alpha$.

$\hookrightarrow$ Classical examples: Bonferroni, Holm, min-p.

Separation Rates

Separation rates for single tests

Single test $\Phi \in \{0,1\}$ rejecting $H_0 \subset \mathcal{P}$ when $\Phi = 1$

- First kind error rate $\mathrm{ER}(\Phi) = \sup_{P \in H_0} P(\Phi = 1) \Rightarrow$ level-α test.
- Second kind error rate $P \notin H_0 \mapsto P(\Phi = 0)$.

Separation Rates

Separation rates for single tests

Single test $\Phi \in \{0,1\}$ rejecting $H_0 \subset \mathcal{P}$ when $\Phi = 1$

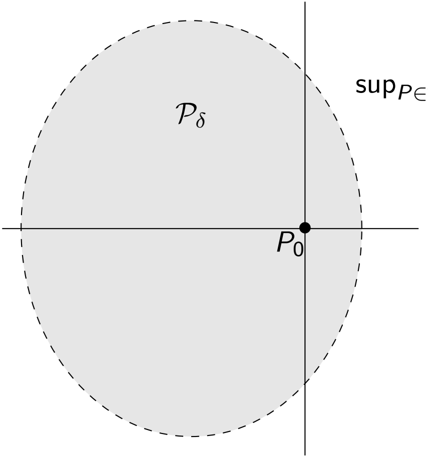
- First kind error rate $\mathrm{ER}(\Phi) = \sup_{P \in H_0} P(\Phi = 1) \Rightarrow$ level-α test.
- Second kind error rate $P \notin H_0 \mapsto P(\Phi = 0)$.

Separation rates $(\alpha, \beta \in (0,1),\ \mathcal{P}_\delta \subset \mathcal{P})$

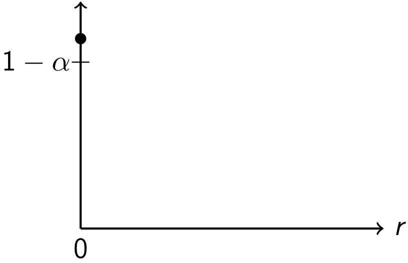
Given a distance d on $\mathcal{P}$, for $\mathcal{Q} \subset \mathcal{P}$, $d\left(P, \mathcal{Q}\right) = \inf_{Q \in \mathcal{Q}} d(P, Q)$.

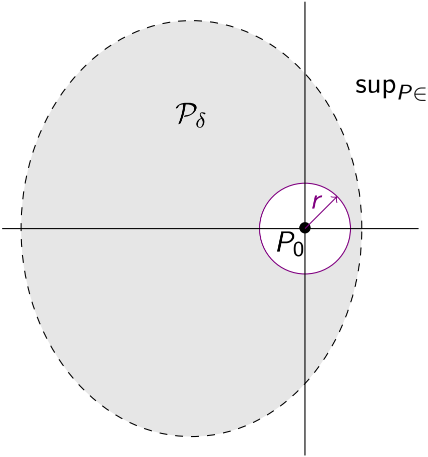
- The separation rate of a level α test Φ_α over $\mathcal{P}_\delta$ for β is

$$\mathrm{SR}_d\left(\Phi_\alpha, \mathcal{P}_\delta, \beta\right) = \inf\{r > 0 / \sup_{P \in \mathcal{P}_\delta / d(P, H_0) \geq r} P(\Phi_\alpha = 0) \leq \beta\}.$$

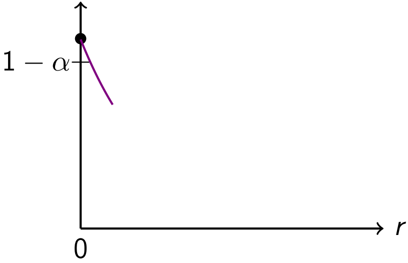


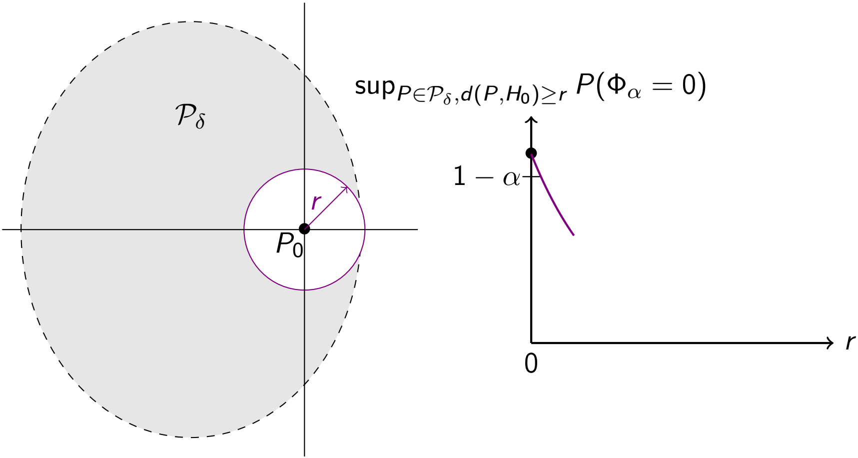
$$\sup_{P\in\mathcal{P}_\delta, d(P,H_0)\geq r} P(\Phi_\alpha = 0)$$

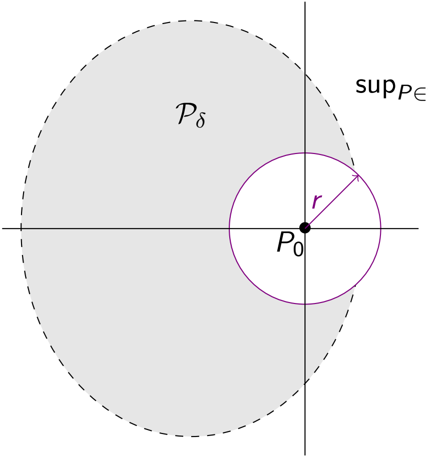




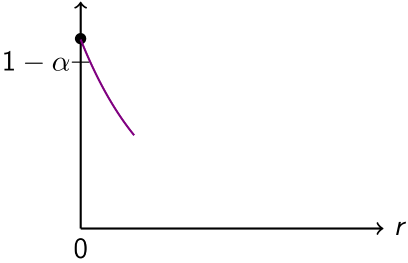
$\sup_{P\in\mathcal{P}_\delta, d(P,H_0)\geq r} P(\Phi_\alpha = 0)$

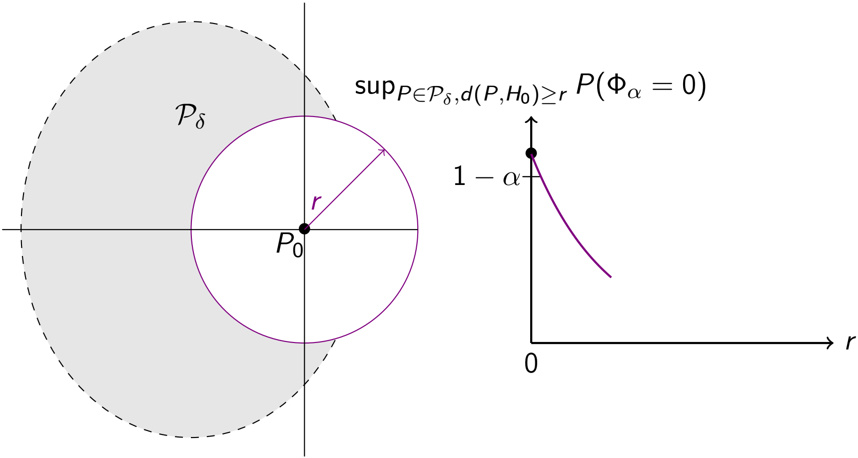


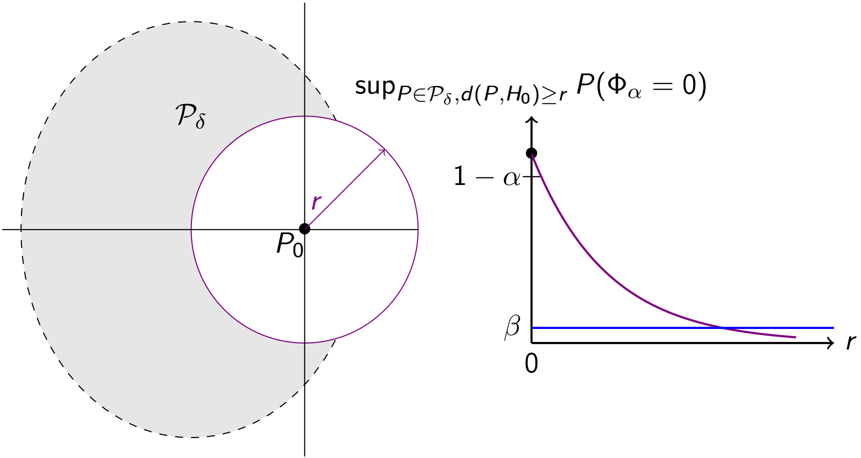

$\mathcal{P}_\delta$
r
P_0
$\sup_{P\in\mathcal{P}_\delta, d(P,H_0)\geq r} P(\Phi_\alpha = 0)$
$1-\alpha$
0
r

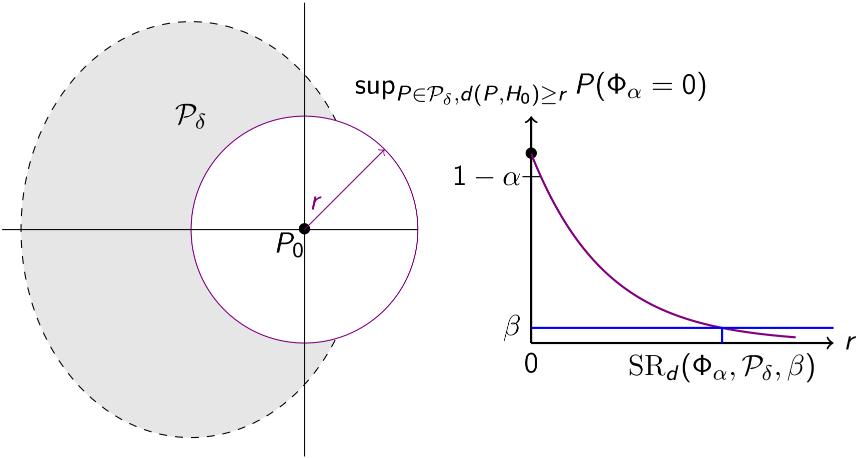


$\sup_{P\in\mathcal{P}_\delta, d(P,H_0)\geq r} P(\Phi_\alpha = 0)$




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


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


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

Separation Rates

Separation rates for single tests

Single test $\Phi \in \{0,1\}$ rejecting $H_0 \subset \mathcal{P}$ when $\Phi = 1$

- First kind error rate $\mathrm{ER}(\Phi) = \sup_{P \in H_0} P(\Phi = 1)$.
- Second kind error rate $P \notin H_0 \mapsto P(\Phi = 0)$.

Separation rates $(\alpha, \beta \in (0,1),\ \mathcal{P}_\delta \subset \mathcal{P})$

Given a distance d on $\mathcal{P}$, for $\mathcal{Q} \subset \mathcal{P}$, $d\left(P, \mathcal{Q}\right) = \inf_{Q \in \mathcal{Q}} d(P, Q)$.

- The separation rate of a level α test Φ_α over $\mathcal{P}_\delta$ for β is
$\mathrm{SR}_d\left(\Phi_\alpha, \mathcal{P}_\delta, \beta\right) = \inf\{r > 0 / \sup_{P \in \mathcal{P}_\delta / d(P,H_0) \geq r} P(\Phi_\alpha = 0) \leq \beta\}$.
- The minimax rate of testing over $\mathcal{P}_\delta$ for α and β is
$m\mathrm{SR}_d\left(\mathcal{P}_\delta, \alpha, \beta\right) = \inf_{\Phi'_\alpha / \mathrm{ER}(\Phi'_\alpha) \leq \alpha} \mathrm{SR}_d\left(\Phi'_\alpha, \mathcal{P}_\delta, \beta\right)$.
- Φ_α is minimax over $\mathcal{P}_\delta$, if
$\mathrm{SR}_d\left(\Phi_\alpha, \mathcal{P}_\delta, \beta\right) \simeq m\mathrm{SR}_d\left(\mathcal{P}_\delta, \alpha, \beta\right)$ (up to constants).

Separation Rates

Separation rates for single tests

Example in Gaussian regression (Baraud 2005)

$X \sim P_{\mathbf{f}} = \mathcal{N}_N(\mathbf{f}, \sigma^2 I_N)$, with $\mathbf{f} = (f_1, ..., f_N)'$ unknown, σ^2 known.

$$X_i = f_i + \sigma\varepsilon_i, \quad i = 1, ..., N, \quad \varepsilon_i\text{'s i.i.d. } \mathcal{N}(0,1).$$

Testing $H_0 = \{P_0\}$ v.s. $\mathcal{P} \setminus H_0$ that is "$\mathbf{f} = 0$" v.s. "$\mathbf{f} \neq 0$".

– $d_2(P_{\mathbf{f}}, P_{\mathbf{g}}) = \left(\sum_{i=1}^N (f_i - g_i)^2\right)^{1/2}$.

– $\mathcal{P}_\delta = \{P_{\mathbf{f}} / \#\{i / f_i \neq 0\} \leq \delta\}$ $(\delta = 1, \ldots, N)$.

Separation Rates
Separation rates for single tests

Theorem (Baraud, 2005)

(i) $m\mathrm{SR}_{d_2}\left(\mathcal{P}_\delta, \alpha, \beta\right) \geq \sigma\left(\delta \ln\left(1 + \frac{N}{\delta^2} \vee \sqrt{\frac{N}{\delta^2}}\right)\right)^{1/2}$, when $\alpha + \beta \leq 0.5$.

(ii) There exists a level α test $\Phi_{\alpha,\delta}$ s.t. $\mathrm{SR}_{d_2}\left(\Phi_{\alpha,\delta}, \mathcal{P}_\delta, \beta\right) \leq C'(\alpha,\beta)\sigma\left(\left(\delta \ln\left(e\frac{N}{\delta}\right)\right) \wedge \sqrt{N}\right)^{1/2}$, i.e. $\Phi_{\alpha,\delta}$ is minimax over $\mathcal{P}_\delta$ for any $\delta \in \{1, \ldots, N\}$.

- Problem: $\Phi_{\alpha,\delta,\sigma}$ depends on δ (unknown in practice).
- Aim: constructing a test Φ_α which is independent of δ, but still (almost) minimax over $\mathcal{P}_\delta$, for many δ's simultaneously

$\Rightarrow$ minimax adaptivity, aggregated test

Separation Rates

Separation rates for single tests

Level u_α single test of $H_i = \{ P_\mathbf{f} \ / \ f_i = 0 \}$:

$$\mathbb{1}_{\{X_i^2 > \sigma^2 F^{-1}(1-u_\alpha)\}}, \quad F = \text{c.d.f. of } \chi^2(1).$$

Level α aggregated test of $H_0 = \{ P_0 \} = \cap H_i$:

$$\Phi_\alpha = \mathbb{1}_{\{\max_{i=1,\dots,N} X_i^2 > \sigma^2 F^{-1}(1-u_\alpha)\}},$$

- $u_\alpha = \alpha / N$, or
- $F^{-1}(1 - u_\alpha) = (1 - \alpha)$ quantile of $\max_{i=1,\dots,N} \varepsilon_i^2$

$\hookrightarrow$ Minimax adaptive over the $\mathcal{P}_\delta$'s ($\delta \leq \sqrt{N}$):

$\hookrightarrow$ Generalization in other frameworks.

Separation Rates

Parallel between aggregated tests and multiple tests

From aggregated test to multiple test

$\Phi_\alpha = \mathbb{1}_{\left\{\max X_i^2 > \sigma^2 F^{-1}(1-u_\alpha)\right\}}$
$\Rightarrow \mathcal{R}^*(\Phi_\alpha) = \left\{ H_i, X_i^2 > \sigma^2 F^{-1}(1-u_\alpha) \right\}$ multiple test of the H_i's.

Proposition

- When $u_\alpha = \alpha/N$, $\mathcal{R}^*(\Phi_\alpha) =$ Bonferroni, first step of Holm proc.
- With the other choice, $\mathcal{R}^*(\Phi_\alpha) =$ first step of a min-p procedure.

Consequence: $\mathrm{FWER}(\mathcal{R}^*(\Phi_\alpha)) \leq \alpha$.

From multiple test to aggregated test

Conversely, $\mathcal{R}^* \Rightarrow \Phi(\mathcal{R}^*) = \mathbb{1}_{\{\mathcal{R}^* \neq 0\}}$ single test of $\cap H_i = \{P_0\}$.

Separation Rates

Weak Family Wise Separation Rate for multiple tests

General problem of simultaneously testing H, $H \in \mathcal{H}$

Notice $w\mathrm{FWER}(\mathcal{R}^*) = \mathrm{ER}\left(\mathbb{1}_{\{\mathcal{R}^* \neq \emptyset\}}\right)$ (as a single test of $\cap\mathcal{H}$).

$\hookrightarrow \mathrm{SR}_d\left(\mathbb{1}_{\{\mathcal{R}^* \neq \emptyset\}}, \mathcal{P}_\delta, \beta\right)$ as a definition of the weak Family Wise Separation Rate ?

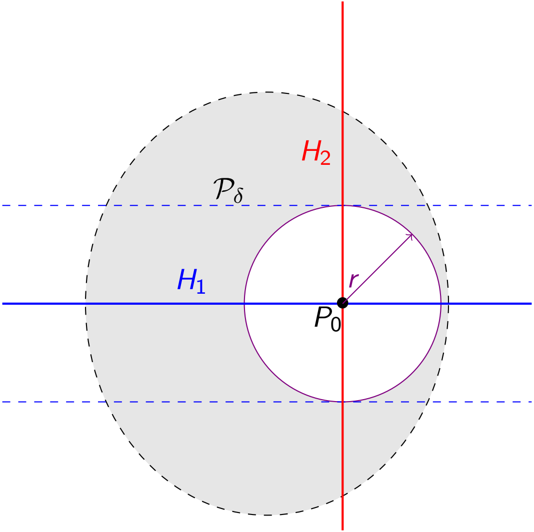
$$\mathrm{SR}_d\left(\mathbb{1}_{\{\mathcal{R}^* \neq \emptyset\}}, \mathcal{P}_\delta, \beta\right) = \inf\{r > 0 / \sup_{P \in \mathcal{P}_\delta / d(P, \cap\mathcal{H}) \geq r} P\left(\mathcal{R}^* = \emptyset\right) \leq \beta\}.$$

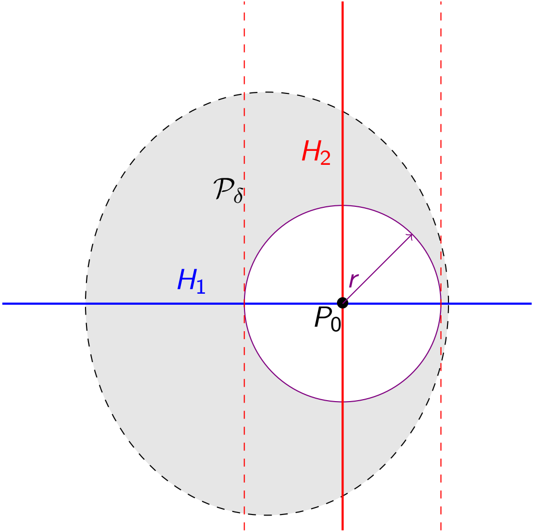
But an alternative to $P \in \cap\mathcal{H}$ should rather be for $r > 0$:

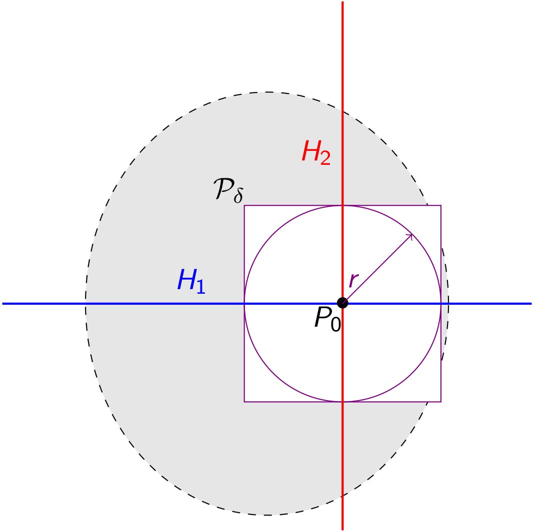
$\exists H \in \mathcal{H} / d(P, H) \geq r$ than $d(P, \cap\mathcal{H}) \geq r$.

$\hookrightarrow$ Set of false hypotheses under P at distance r from P:

$$\mathcal{F}_r(P) = \{ H \in \mathcal{H} \ / \ d(P, H) \geq r \}.$$


H_2
$\mathcal{P}_\delta$
H_1
r
P_0


H_2
$\mathcal{P}_\delta$
H_1
r
P_0


H_2
$\mathcal{P}_\delta$
H_1
r
P_0

Separation Rates
Weak Family Wise Separation Rate for multiple tests

Weak Family Wise Separation Rate

We define the weak Family Wise Separation Rate of a multiple test $\mathcal{R}^*$ over $\mathcal{P}_\delta$ for β, by

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

Proposition

- $w\mathrm{FWSR}_d\left(\mathcal{R}^*, \mathcal{P}_\delta, \beta\right) \leq \mathrm{SR}_d\left(\mathbb{1}_{\{\mathcal{R}^* \neq \emptyset\}}, \mathcal{P}_\delta, \beta\right)$.
- If $(\mathcal{A})$ $\forall r > 0, \quad d(P, \cap\mathcal{H}) \geq r \Leftrightarrow \mathcal{F}_r(P) \neq \emptyset$,

then $w\mathrm{FWSR}_d\left(\mathcal{R}^*, \mathcal{P}_\delta, \beta\right) = \mathrm{SR}_d\left(\mathbb{1}_{\{\mathcal{R}^* \neq \emptyset\}}, \mathcal{P}_\delta, \beta\right)$.

$\hookrightarrow$ Closed collections of hypotheses satisfy $(\mathcal{A})$

Separation Rates

Family Wise Separation Rate for multiple tests

(Strong) Family Wise Separation Rate

We define the Family Wise Separation Rate of a multiple test $\mathcal{R}^*$ over $\mathcal{P}_\delta$ for β, by

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

Minimax Family Wise Separation Rate

We define the minimax Family Wise Separation Rate over $\mathcal{P}_\delta$ for α and β, by

$$m\mathrm{FWSR}_d\left(\mathcal{P}_\delta, \alpha, \beta\right) = \inf_{\mathcal{R}^*_\alpha / \mathrm{FWER}(\mathcal{R}^*_\alpha) \leq \alpha} \mathrm{FWSR}_d\left(\mathcal{R}^*_\alpha, \mathcal{P}_\delta, \beta\right)$$

Separation Rates

Family wise Error Rate for multiple tests

Proposition

$$w\mathrm{FWSR}_d\left(\mathcal{R}^*, \mathcal{P}_\delta, \beta\right) \leq \mathrm{FWSR}_d\left(\mathcal{R}^*, \mathcal{P}_\delta, \beta\right)$$

Corollary

Under assumption $(\mathcal{A})$, $m\mathrm{FWSR}_d\left(\mathcal{P}_\delta, \alpha, \beta\right) \geq m\mathrm{SR}_d\left(\mathcal{P}_\delta, \alpha, \beta\right)$.

$\hookrightarrow$ The classical literature in minimax single testing may be used to obtain a lower bound for $m\mathrm{FWSR}_d\left(\mathcal{P}_\delta, \alpha, \beta\right)$ under $(\mathcal{A})$.

Examples in Gaussian regression
Unclosed collection of hypotheses

$X_i = f_i + \sigma\varepsilon_i,\ \varepsilon_i$'s i.i.d. $\mathcal{N}(0,1)$

$H_i = \{ P_{\mathbf{f}} \in \mathcal{P} / f_i = 0 \}$

Lower bound

$d_\infty(P_{\mathbf{f}}, P_{\mathbf{g}}) = \max_{i=1,\ldots,N} |f_i - g_i|,$

$d_s(P_{\mathbf{f}}, P_{\mathbf{g}}) = \left(\sum_{i=1}^N (f_i - g_i)^s \right)^{1/s} \quad (s \geq 1).$

From the above corollary and the results of Baraud,

Theorem

For every δ, $s \in [1, \infty]$, $m\mathrm{FWSR}_{d_s}(\mathcal{P}_\delta, \alpha, \beta) \geq \sigma\sqrt{\ln(1+N)}$.

Examples in Gaussian regression
Unclosed collection of hypotheses

Upper bound

Let $\mathcal{R}^*$ be the Bonferroni, Holm, or min-p procedure, based on the p-values $p_i(X) = 1 - F(\sigma^{-2} X_i^2)$, $\mathrm{FWER}(\mathcal{R}^*) \leq \alpha$.

Theorem

For every δ, $s \in [1, \infty]$,

$$\mathrm{FWSR}_{d_s}(\mathcal{R}^*, \mathcal{P}_\delta, \beta) \leq \sigma \left(\sqrt{2 \ln \left(\frac{\delta}{2\beta} \right)} + \sqrt{2 \ln \left(\frac{N}{\alpha} \right)} \right).$$

$\hookrightarrow$ $\mathcal{R}^*$ is minimax over all the $\mathcal{P}_\delta$'s simultaneously $\Rightarrow$ adaptivity.

$\hookrightarrow$ $m\mathrm{FWSR}_{d_2}(\mathcal{P}_N, \alpha, \beta)$ of order $\sigma\sqrt{\ln N}$ much smaller than $m\mathrm{SR}_{d_2}(\mathcal{P}_N, \alpha, \beta)$ of order $\sigma N^{1/4}$!

Examples in Gaussian regression
Closed collection of hypotheses

$$X_i = f_i + \sigma\varepsilon_i,\ \ \varepsilon_i\text{'s i.i.d. } \mathcal{N}(0,1)$$

$$\overline{H}_i = \{\, P_{\mathbf{f}} \in \mathcal{P} / f_1 = \ldots = f_i = 0 \,\}\,.$$

Lower bound

$$d_2(P_{\mathbf{f}}, P_{\mathbf{g}}) = \left(\sum_{i=1}^N (\, f_i - g_i \,)^2 \right)^{1/2}.$$

For every δ, when $\alpha + \beta \leq 0.5$,

$$m\mathrm{FWSR}_{d_2}\,(\mathcal{P}_\delta, \alpha, \beta) \geq \sigma \left(\delta \ln \left(1 + \frac{N}{\delta^2} \vee \sqrt{\frac{N}{\delta^2}} \right) \right)^{1/2}.$$

Examples in Gaussian regression
Closed collection of hypotheses

Upper bound

Let $\mathcal{H}_i = \{H_1, \ldots, H_i\}$, and $\mathcal{R}_i^*$ be the Bonferroni, Holm, or min-p procedure for the collection of hypotheses $\mathcal{H}_i$, based on the p-values $p_j(X) = 1 - F(\sigma^{-2} X_j^2)$ $(j = 1, \ldots, i)$.

$\Phi(\mathcal{R}_i^*) = \mathbb{1}_{\{\mathcal{R}_i^* \neq 0\}}$ single level α test of $\overline{H}_i$.

Based on the $\Phi(\mathcal{R}_i^*)$'s, multiple test for $\bar{\mathcal{H}} = \{\bar{H}_i, i = 1, ..., N\}$ obtained by the closure method of Marcus, Peritz, Gabriel:

$$\bar{\mathcal{R}}^* = \{\bar{H}_i, \ / \ \forall j \in \{i, \ldots, N\}, \mathcal{R}_j^* \neq \emptyset\}.$$

Examples in Gaussian regression

Closed collection of hypotheses

Theorem

For δ in $\{1, \ldots, n\}$, β in $(0, 0.5)$,

- $\mathrm{FWER}(\bar{\mathcal{R}}^*) \leq \alpha$,
- $\mathrm{FWSR}_{d_2}\left(\bar{\mathcal{R}}^*, \mathcal{P}_\delta, \beta\right) \leq \sigma\sqrt{\delta}\left(\sqrt{-2\ln(2\beta)} + \sqrt{2\ln(2N/\alpha)}\right)$.

$\hookrightarrow$ For $\delta = N^\gamma$, $\bar{\mathcal{R}}^*$ is minimax over all the $\mathcal{P}_\delta$'s simultaneously
$\Rightarrow$ adaptivity (with no price to pay)