On controlling the amount of false positives when making multiple tests

Part I : Introduction

① Setting

Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathcal{X}, \mathfrak{X}, P)$ observation (may be a vector, matrix, ...)

Consider $\begin{cases} P \in \mathcal{P} \text{ model} & (\text{distribution family on } (\mathcal{X}, \mathfrak{X})) \\ m \geq 1 \text{ null hypotheses on } P : & H_{0,j} : \text{"} P \in \mathcal{P}_{0,j} \text{"}, \quad 1 \leq j \leq m \end{cases}$ subsets of $\mathcal{P}$

(Called multiple testing setting)

Parameter of interest is $\theta = \theta(P) \in \{0,1\}^m$ defined by $\theta_j = 0 \Leftrightarrow P \in \mathcal{P}_{0,j}$
the j-th null is true for P

$\mathcal{H}_0(P) = \{ j \in \{1, \dots, m\} : \theta_j = 0 \}$ set of true nulls

Each $H_{0,j}$ tested with a test statistic $T_j(X)$ (expected large if $\theta_j = 1$)

Often : each $T_j(X)$ transformed into a p-value $p_j(X)$

Condition (*) there exists a family $(p_j(X), 1 \leq j \leq m)$ with the property :

$$\forall P \in \mathcal{P}, \quad p_j(X) \underset{\text{stoch}}{\succeq} \mathcal{U}(0,1) \quad \text{for each } i \text{ such that } \theta_i = 0$$

i.e. $P(p_j(X) \leq t) \leq t$ for all $t \in [0,1]$

$(p_j(X), 1 \leq j \leq m)$ called the p-value family (property p value property)

Aim : recover θ from X, $(T_j(X), 1 \leq j \leq m)$ or $(p_j(X), 1 \leq j \leq m)$

② Examples

(e.g., think about the standard linear model)

Gaussian case

$X \sim P = \mathcal{N}(\mu, \Gamma)$, $\underset{\text{unknown}}{\mu \in \mathbb{R}^m}$, Γ $m \times m$ covariance matrix (known or unknown) with $\Gamma_{jj} = 1 \ \forall j$

* One-sided testing : $\mathcal{P}_{0,j} = \{ P = \mathcal{N}(\mu, \Gamma) : \mu_j \leq 0 \}$ ie $H_{0,j}$: "$\mu_j \leq 0$"

$$\theta_j = \mathbb{1}\{\mu_j > 0\} ; \quad T_j(X) = X_j$$

$$p_j(X) = \bar{\Phi}(X_j) , \quad \text{pour } \bar{\Phi}(x) = \mathbb{P}(Z \geq x) \text{ pour } Z \sim \mathcal{N}(0,1)$$

p-value property (*) is ok if $\theta_j = 0$, $P(p_j(X) \leq t) = P(\bar{\Phi}(\underbrace{X_j}_{\leq X_j - \mu_j}) \leq t) \leq P(\bar{\Phi}(\underbrace{X_j - \mu_j}_{\sim \mathcal{N}(0,1)}) \leq t) \leq t$

* Two-sided testing : $\mathcal{P}_{0,j} = \{ P = \mathcal{N}(\mu, \Gamma) : \mu_j = 0 \}$ ie $H_{0,j}$: "$\mu_j = 0$"

$$\theta_j = \mathbb{1}\{\mu_j \neq 0\} ; \quad T_j(X) = |X_j| ; \quad p_j(X) = 2 \bar{\Phi}(|X_j|)$$

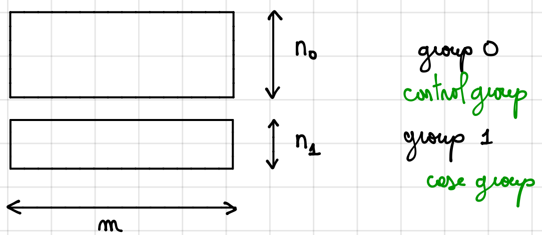
p-value property (*) is ok if $\theta_j = 0$, $P(p_j(X) \leq t) = P(2\bar{\Phi}(\underbrace{|X_j|}_{= |X_j - \mu_j|}) \leq t) = P(\bar{\Phi}(\underbrace{|X_j - \mu_j|}_{\sim \mathcal{N}(0,1)}) \leq \frac{t}{2})$

$$= P(|X_j - \mu_j| \geq \bar{\Phi}^{-1}(t/2)) = t$$

Two-group case

$$X = (X^{(1)}, \ldots, X^{(n)}) \in \mathbb{R}^m = (Y^{(1)}, \ldots, Y^{(n_0)}, Z^{(1)}, \ldots, Z^{(n_1)})$$

=



with $Y^{(i)} = \mu_0 + \xi_i$, $1 \leq i \leq n_0$ (effect group 0)
$Z^{(i)} = \mu_1 + \zeta_i$, $1 \leq i \leq n_1$ (effect group 1)

with ξ_i, ζ_i's iid $\sim Q$ some centered distribution on $\mathbb{R}^m$

$H_{0,j}$: " $\mu_{0,j} = \mu_{1,j}$ " (versus $\neq$) ; $\theta_j = \mathbb{1}\{\mu_{0,j} \neq \mu_{1,j}\}$

Here, only two-sided because we will permute items

$$T_j(X) = \frac{1}{\sqrt{1/n_0 + 1/n_1}} \frac{|\hat{\mu}_{0,j} - \hat{\mu}_{1,j}|}{\hat{\sigma}_j}$$ Student Stat

where $\hat{\mu}_{0,j} = \frac{1}{n_0} \sum_{i=1}^{n_0} Y_j^{(i)}$ and $\hat{\mu}_{1,j} = \frac{1}{n_1} \sum_{i=1}^{n_1} Z_j^{(i)}$

$*$ if $Q = \mathcal{N}(0, \Gamma)$ then $p_j(X) = 2\, \bar{F}(T_j(X))$ where $\bar{F}(x) = P(Z \geq x)$
$Z \sim \mathcal{T}(n-2)$

condition $(*)$ on pvalues can be checked as before

$*$ if Q unknown and arbitrary
we can obtain p-values by permutations!

$$p_j(X) = \frac{1}{B+1} \left(1 + \sum_{b=1}^{B} \mathbb{1}\{T_j(X^{\sigma_b}) \geq T_j(X)\}\right)$$

this breaks the group structure

where $\sigma_1 \dots \sigma_B$ iid uniform on $\mathfrak{S}_n$ and $X^\sigma = (X^{(\sigma(1))}, \dots, X^{(\sigma(n))})$

matrix X with the columns permuted by σ

condition $(*)$ on pvalues more subtle:

Proposition: condition $(*)$ holds for this p_value family

(single permutation testing)

Proof:

if $H_{0,j}$ is true, the n components of $X_j = (X_j^{(i)})_{1 \leq i \leq n}$ are iid ($\sim Q_j(\cdot - \mu_{0,j}) = Q_j(\cdot - \mu_{1,j})$)

hence $X_j^\sigma \sim X_j$ for any $\sigma \in \mathfrak{S}_n$
and thus also for any random permutations σ uniformly distributed on $\mathfrak{S}_n$ (taken indep. of the rest)

As a result, if $H_{0,j}$ is true

$$(X_j, X_j^{\sigma_1}, \dots, X_j^{\sigma_B}) \sim (X_j^\sigma, X_j^{\sigma_1 \circ \sigma}, \dots, X_j^{\sigma_B \circ \sigma})$$

(because true cond on $\sigma_1 \dots \sigma_B$)

$$\sim (X_j^\sigma, X_j^{\sigma_1}, \dots, X_j^{\sigma_B})$$

(because true cond on X_j and $(\sigma, \sigma_1 \circ \sigma, \dots, \sigma_B \circ \sigma) \sim (\sigma, \sigma_1, \dots, \sigma_B)$)

Hence if $H_{0,j}$ is true

$$(U_0, U_1, \dots, U_B) := (T_j(X), T_j(X^{\sigma_1}), \dots, T_j(X^{\sigma_B}))$$
$$\sim (T_j(X^\sigma), T_j(X^{\sigma_1}), \dots, T_j(X^{\sigma_B}))$$

is an exchangeable vector and $p_j(X) = \frac{1}{B+1} \sum_{b=0}^{B} \mathbb{1}\{U_b \geq U_0\}$ rank of U_0 in $(U_0, U_1, \dots, U_B)$ $\square$

③ Multiple testing procedure

A multiple testing procedure is some measurable function $R : (\mathcal{X}, X) \longrightarrow$ subsets of $\{1, \dots, m\}$

"$i \in R(X)$" means that the null H_{0i} is rejected by R

Thresholding based: reject large test stat. (or small p-values)

with test statistics $\quad R(X) = \{ j \in \{1, \dots, m\} : T_j(X) > s \}$ (mind the strict)

with p-values $\quad R(X) = \{ j \in \{1, \dots, m\} : p_j(X) \leq t \}$

Stopping rule: $\hat{\ell}(x) \in \{1, \dots, m\}$ and reject nulls corresponding to $p_{(1)}, \dots, p_{(\hat{\ell})}$
where $p_{(1)} \leq \dots \leq p_{(m)}$ are the ordered p-values

False positives are elements of $\mathcal{H}_0(P) \cap R$ (wrongly rejected by R)

Curse of multiplicity

if $R = \{ j \in \{1, \dots, m\} : p_j \leq \alpha \}$ (uncorrected)

only noise $\theta_j = 0$ for all j

and model with p_j iid $\sim \mathcal{U}(0,1)$ then probability to make a false positive is

$$P(|\mathcal{H}_0(P) \cap R| \neq 0) = P(\exists j \in \{1 \dots m\} : p_j \leq \alpha) = 1 - (1-\alpha)^m$$

quickly increasing to 1 as m grows.

Part II : FWER control

① FWER and Bonferroni procedure

In a general multiple testing framework with X, $(\mathcal{X}, \mathfrak{X})$, $P \in \mathcal{P}$, $\theta \in \mathbb{R}^m$, $\mathcal{H}_0(P)$

Let R being a multiple testing procedure $\qquad m_0(P) = |\mathcal{H}_0(P)|$ number of 'non signal'

The FWER of R is $\quad \mathrm{FWER}(R,P) = \mathbb{P}\left(|R(X) \cap \mathcal{H}_0(P)| \neq 0 \right)$ $\quad$ probability to make at least one false positive

Controlling FWER at level α means :

for some $\alpha \in (0,1)$, find $R = R_\alpha$ with $\forall P \in \mathcal{P}$, $\mathrm{FWER}(R,P) \leq \alpha$ $\quad$ (and as many rejections as possible)

The Bonferroni procedure $\quad R^{\mathrm{Bonf}} = \{ 1 \leq j \leq m : p_j(X) \leq \frac{\alpha}{m} \}$

Proposition: consider p_values satisfying (∗) in a general multiple testing setting

(i) $\forall P \in \mathcal{P}$, $\mathrm{FWER}(R^{\mathrm{Bonf}}, P) \leq \alpha \frac{m_0(P)}{m} \leq \alpha$

(ii) if any distribution in $[0,1]^m$ with uniform marginals corresponds to the distribution of the p-value family for some $P_0 \in \mathcal{P}$ with $\theta(P_0)_j = 0$ for all j $\quad$ under the 'full null'

$$\sup_{P \in \mathcal{P}} \{ \mathrm{FWER}(R^{\mathrm{Bonf}}, P) \} = \alpha$$

[Benditkis et al (2015)]

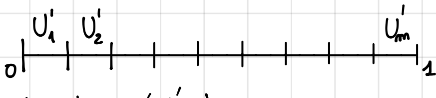
Bonf is sharp !

Proof : (i) $\mathrm{FWER}(R^{\mathrm{Bonf}}, P) \leq \sum_{j=1}^{m} (1-\theta_j) \mathbb{P}(p_j(X) \leq \frac{\alpha}{m}) \underset{\text{by } (*)}{\leq} \frac{\alpha}{m} \sum_{j=1}^{m} (1-\theta_j) = \frac{\alpha}{m} m_0$

(ii) Take $U_1, \dots, U_m$ iid $\sim \mathcal{U}(0,1)$, $\quad U'_j = \frac{j-1+U_j}{m} \in \left[\frac{j-1}{m}, \frac{j}{m}\right]$, $\quad 1 \leq j \leq m$

σ uniform on $\mathfrak{S}_n$ (indep)

Consider the distribution of $\left(U'_{\sigma(j)}, 1 \leq j \leq m \right)$



By assumption $\exists P_0$ with $\theta_j(P_0) = 0$ for all j and $(p_j(X))_j \sim (U'_{\sigma(j)})_j$

To conclude, we write

$$\mathrm{FWER}(R^{\mathrm{Bonf}}, P_0) = \mathbb{P}\left(\exists j \in \{1,\dots,m\} : U'_{\sigma(j)} \leq \frac{\alpha}{m}\right)$$

$$= \mathbb{P}\left(\bigcup_{j=1}^{m} \bigcup_{k=1}^{m} \left\{\sigma(j) = k,\ U'_k \leq \frac{\alpha}{m}\right\}\right)$$

$$= \mathbb{P}\Big(\bigcup_{k=1}^{m} \Big\{\underbrace{U'_k}_{\sim\, \mathcal{U}\left(\frac{(k-1)}{m}, \frac{k}{m}\right)} \leq \frac{\alpha}{m}\Big\}\Big) = \mathbb{P}\left(U'_1 \leq \frac{\alpha}{m}\right) = P(U_1 \leq \alpha) = \alpha$$

$k = 2 \dots m$ impossible

$\square$

Remark: if P_0 under the full null and p_j, $1 \leq j \leq m$ are independent

$$\mathrm{FWER}(R^{\mathrm{Bonf}}, P_0) = 1 - \left(1 - \frac{\alpha}{m}\right)^m \simeq \alpha$$ when α is small so 'almost sharp' under indep.

Lack of adaptiveness for some particular $P_0 \in \mathcal{P}$, $\mathrm{FWER}(R^{\mathrm{Bonf}}, P_0)$ can be much lower than α

R^{Bonf} 'too conservative'

for instances:

* <u>strong dependence between tests</u>:
  full null gaussian two-sided case with $\Gamma_{ij} = 1$ for all i,j
  gives $p_i(X) = p_j(X)$ for all i,j hence $\mathrm{FWER}(R^{\mathrm{Bonf}}, P_0) = \frac{\alpha}{m}$ $(\ll \alpha)$

* <u>many signal</u>: $\frac{m_0(P_0)}{m}$ not close to 1
  provides $\mathrm{FWER}(R^{\mathrm{Bonf}}, P) \leq \alpha \frac{m_0(P)}{m}$ not close to α

Adaptive control issue

How to build a new threshold $t = t(X)$ that incorporates the dependence or/and $\frac{m_0}{m}$

with $\begin{cases} t \text{ larger than } \alpha/m \\ \text{FWER still controlled by } \alpha \end{cases}$?

(Also remember that the dependence can be known or unknown)