Improving multiple testing procedures by estimating the proportion of true null hypotheses

Alain Celisse

[1]UMR 8524 CNRS - Université Lille 1

[2]SSB Group, Paris

joint work with Stéphane Robin

μTOSS

Berlin, February, 12 2010

Statistical setting

$\mathcal{H}_1, \ldots, \mathcal{H}_m$ m hypotheses such that

$$\mathcal{H}_i : \qquad \mathcal{H}_{i,0} \text{ is true} \quad vs \quad \mathcal{H}_{i,1} \text{ is true} \ .$$

Question:

Which hypotheses among $\{\mathcal{H}_1, \ldots, \mathcal{H}_m\}$ are true alternatives ?

- A test statistic is computed for each $\mathcal{H}_i$.
- $P_1, \ldots, P_m$ denote the coresponding p-values.
- π_0: unknown proportion of true nulls among $P_1, \ldots, P_m$.

Statistical setting

Decision rule:
Build a *rejection region*

$$\mathcal{R}(P_1, \ldots, P_m) \subset \{P_1, \ldots, P_m\} \ .$$

Type-I and II errors:

- $\mathcal{H}_i$ is a *false positive* if
$$\mathcal{H}_{i,0} \text{ is true and } P_i \in \mathcal{R}(P_1, \ldots, P_m).$$
- $\mathcal{H}_i$ is a *false negative* if
$$\mathcal{H}_{i,1} \text{ is true and } P_i \notin \mathcal{R}(P_1, \ldots, P_m).$$

Notation:

- FP: number of false positives,
- FN: number of false negatives,
- R: number of rejected hypoteses.

Control of type-I errors

Family Wise Error Rate (FWER)

$$FWER := \mathbb{P}(FP \geq 1) \quad .$$

Bonferroni procedure:

For $\alpha > 0$, $\mathcal{R}(P_1, \ldots, P_m) = [0, \alpha/m)^m$.

$$\Rightarrow \quad FWER[\mathcal{R}(P_1, \ldots, P_m)] \leq \sum_{i=1}^{m} \mathbb{P}(P_i \leq \alpha/m) \leq \alpha \quad .$$

$\longrightarrow$ Does not really take into account other p-values.

Control of type-I errors

False Discovery Rate (FDR)

$$FDR := \mathbb{E}\left[\frac{FP}{R} \mathbb{1}_{(R>0)} \right] \quad .$$

Linear step-up procedure: (BH (95))

For any $\alpha > 0$, $\mathcal{R}(P_1, \ldots, P_m) = \left\{ P_{(1)}, \ldots, P_{(\widehat{k})} \right\}$, with

$$\widehat{k} := \max \left\{ i \mid P_{(i)} \leq i\alpha / m \right\} .$$

$$\Rightarrow \quad FDR \left[\mathcal{R}(P_1, \ldots, P_m) \right] \leq \pi_0 \alpha \leq \alpha \quad .$$

$\longrightarrow$ Estimating π_0 would increase the power of the procedure.

Outline

1. Classical estimators of π_0
2. Cross-validation based π_0 estimator
    1. Density estimation by histograms
    2. Efficient cross-validation (closed-form expressions)
3. Control of the FDR
    1. New plug-in adaptive procedure
    2. Assessment of the procedure
4. Local FDR estimation
    1. Iterative algorithm
    2. *kerfdr* package

$\mathbb{I}$ Classical π_0 estimators

Distributional assumptions

Labels: For every i, let $H_i \sim \mathcal{B}(1 - \pi_0)$, with

$$H_i = 0, \quad \text{if } \mathcal{H}_{\mathrm{i},0} \text{ is true,}$$
$$H_i = 1, \quad \text{otherwise.}$$

Conditional distribution: For every i,

$$P_i \mid H_i = 0 \quad \sim \quad f_0 \text{ known,}$$
$$P_i \mid H_i = 1 \quad \sim \quad f_1 \text{ unknown.}$$

Mixture model:

- $f_0 = \mathcal{U}(0, 1)$ (continuous distribution),
- Assuming independence implies

$$P_i \overset{i.i.d.}{\sim} g(x) = \pi_0 + (1 - \pi_0)\, f_1(x), \quad \forall x \in [0, 1] \;.$$

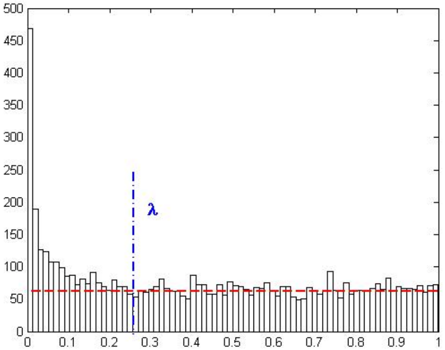
Assumptions on f_1

- **Assumption (NI):**

  f_1 is nonincreasing.

- **Assumption (V_λ):**

  f_1 vanishes on $[\lambda, 1]$.



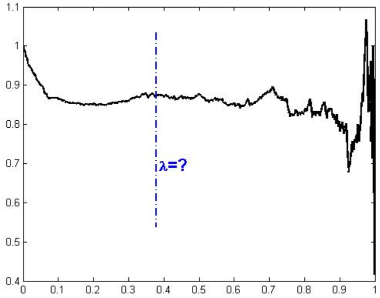
Remark:

Assumption (V_λ) entails the identifiability of π_0 (Genovese Wasserman (04)).

Classical π_0 estimators

Assumption (V_λ) with $\lambda < 1$:
Schweder and Spøtwoll (82)

$$\widehat{\pi}_0^{SS}(\lambda) := \frac{\mathrm{Card}\left(\{i \mid P_i > \lambda\}\right)}{m\,(1-\lambda)} \quad .$$



$\longrightarrow$ Requires to choose $\lambda \in (0,1)$ carefully.

Remark: Storey (02) uses bootstrap.

Classical π_0 estimators

Assumption (V_λ) with $\lambda = 1$:

Storey Tibshirani (03)

- $\widehat{\pi}_0^{SS}(\cdot)$ approximated by cubic spline $\longrightarrow \quad \widehat{\pi}_{0,\mathrm{approx}}^{SS}(\cdot)$.
-

$$\widehat{\pi}_0^{ST} := \widehat{\pi}_{0,\mathrm{approx}}^{SS}(1) \ .$$

Classical π_0 estimators

Assumption (V_λ) with $\lambda = 1$:

Storey Tibshirani (03)

- $\widehat{\pi}_0^{SS}(\cdot)$ approximated by cubic spline $\longrightarrow$ $\widehat{\pi}_{0,\mathrm{approx}}^{SS}(\cdot)$.
- $$\widehat{\pi}_0^{ST} := \widehat{\pi}_{0,\mathrm{approx}}^{SS}(1) \enspace .$$

Without Assumption (V_λ):

Scheid Spang (04)

- *Twilight*: A 'backward' approach yields $\mathcal{R}\left(P_1, \ldots, P_m\right)$.
- $$\widehat{\pi}_0^{Twil} := \frac{\mathrm{Card}\left(\mathcal{R}\left(P_1, \ldots, P_m\right)\right)}{m} \enspace .$$

$\longrightarrow$ Intensive computations are required.

Partial conclusion

Goal:

Build an estimator, which is

- fully data-driven (automatic choice of λ).
- not time consuming.
- also accurate in a wide range of realistic situations (not only under Assumption (V_λ)).

Idea:

Use

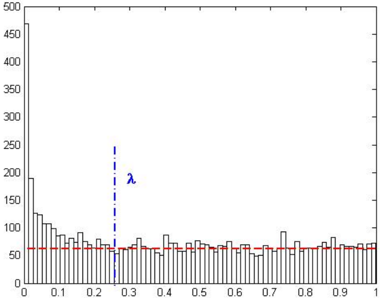
- Density estimation by histograms.
- Cross-validation to avoid unrealistic assumptions.

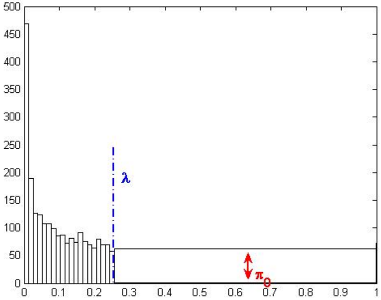
$\mathbb{III}$ CV-based π_0 estimator

(C. and Robin (09), arXiv:0804.1189)
(C. and Robin (08), CSDA)
(Arlot and C. (09), arXiv:0907.4728)

Density estimation by histograms (C. and Robin (08))

Idea: The choice of λ can be rephrased in tems of the choice of an histogram estimator $\widehat{s}_I$.

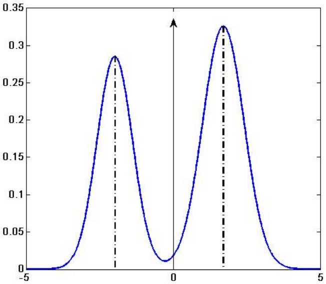


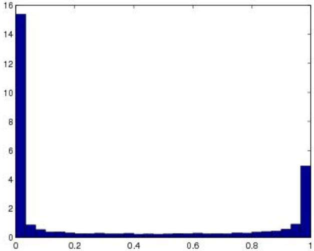


$$\widehat{\pi}_0 := \widehat{s}_I(x), \quad \forall x \in [\widehat{\lambda}, 1] \ ,$$
$$= \frac{\mathrm{Card}\,(i \mid \lambda \leq P_i \leq 1)}{m\,(1-\lambda)} \ .$$

Violation of Assumption (V_λ)

Pounds and Cheng (06) noticed ‘U-shape’ p-value density can occur in realistic situations.





It can occur

- with one-sided tests when the alternative is true.
- with a misspecified distribution of test statistics.
- under some dependence.

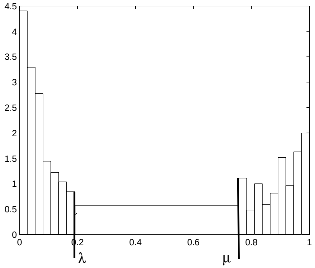
Relaxation of Assumption (V_λ)

Assumption $(V_{\lambda,\mu})$

$$f_1(x) = 0, \quad \forall x \in [\lambda, \mu], \quad \text{with } 0 < \lambda < \mu \leq 1 \ .$$

$$\widehat{\pi}_0 := \widehat{s}_I(x), \quad \forall x \in [\lambda, \mu] \ ,$$

$$= \frac{\text{Card}\left(\{i \mid \lambda \leq P_i \leq \mu\}\right)}{m\,(\mu - \lambda)} \ .$$



Collection of histograms estimators

- regular bins of width $1/N$ on $[0, \lambda]$ and $[\mu, 1]$.
- merge bins between λ and μ.

$\longrightarrow$ Choose the best histogram estimator.

Risk estimation

- $\mathcal{S} = \{\widehat{s}_I \mid I \in \mathcal{I}\}$: collection of histogram estimators.
- The best histogram:

$$I^* := \mathrm{Argmin}_{I \in \mathcal{I}} \{\|g - \widehat{s}_I\|_2\} \quad .$$

Cross-validation (CV)

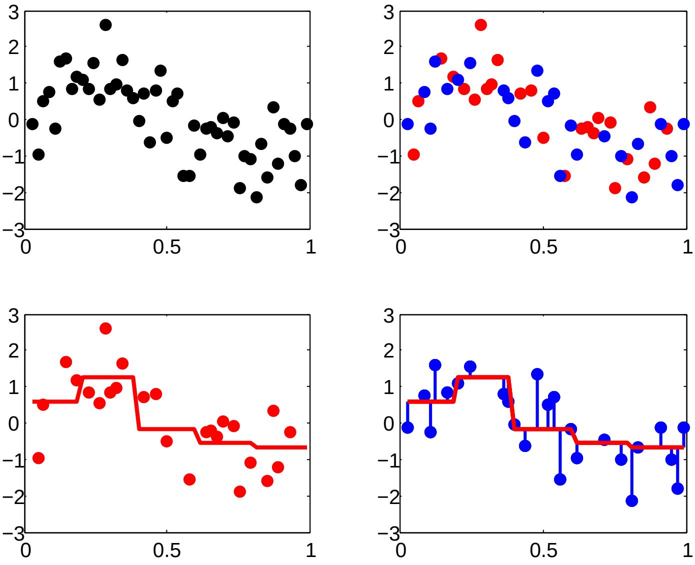
$$\widehat{I} := \mathrm{Argmin}_{I \in \mathcal{I}} \widehat{R}_{CV}(\widehat{s}_I),$$

where $\widehat{R}_{CV}(\widehat{s}_I)$ is the CV estimator of the risk of $\widehat{s}_I$.

Final histogram estimator

$$\widehat{s} = \widehat{s}_{\widehat{I}}.$$

Cross-validation principle



Explicit Leave-p-out cross-validation

Leave-p-out (LPO) $\forall 1 \leq p \leq m-1,$

$$\widehat{R}_p(\widehat{s}) = \binom{n}{p}^{-1} \sum_{D^{(t)} \in \mathcal{E}_p} \left[\frac{1}{p} \sum_{P_i \in D^{(v)}} \left\{ \| \widehat{s}^{\,D^{(t)}} \|_2^2 - 2 \widehat{s}^{\,D^{(t)}}(P_i) \right\} \right],$$

where $\mathcal{E}_p = \left\{ D^{(t)} \subset \{P_1, \ldots, P_n\} \mid \text{Card}\left(D^{(t)}\right) = n - p \right\}.$

Algorithmic complexity:

Exponential $\mathcal{O}\left(e^m\right)$.

$\longrightarrow$ CV in general (LPO) is expensive (intractable) to compute.

Efficient Leave-p-out

Histogram For $I = \{I_\lambda\}_\lambda$ partition of $[0, 1]$,

$$\widehat{s}_I(x) = \sum_\lambda \frac{n_\lambda}{m\,|I_\lambda|} \mathbb{1}_{I_\lambda}, \qquad \text{with} \quad n_\lambda := \mathrm{Card}\left(\{i \mid P_i \in I_\lambda\}\right) \quad \cdot$$

Closed-form expression For $p \in \{1, \ldots, m-1\}$,

$$\widehat{R}_p(\widehat{s}_I) = \frac{2m - p}{(m-1)(m-p)} \sum_\lambda \frac{n_\lambda}{m|I_\lambda|} - \frac{m(m-p+1)}{(m-1)(m-p)} \sum_\lambda \frac{1}{|I_\lambda|} \left(\frac{n_\lambda}{m}\right)^2 .$$

Computational complexity: $\mathcal{O}(m)$ instead of $\mathcal{O}(e^m)$.

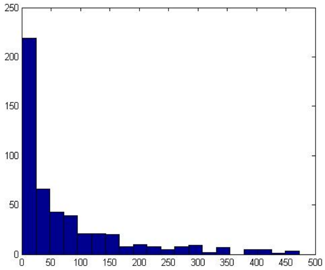
$\longrightarrow$ CV can be performed with no additional computation time.

Choice of p

For each partition I, choose $\widehat{p}(I)$ minimizing the MSE:

$$\widehat{p}(I) := \mathrm{Argmin}_{p \in \{1,\ldots,m-1\}} \widehat{\mathbb{E}} \left[\left(\widehat{R}_p(\widehat{s}_I) - \|g - \widehat{s}_I\|_2 \right)^2 \right] .$$

$\longrightarrow$ A closed-form expression is also available.



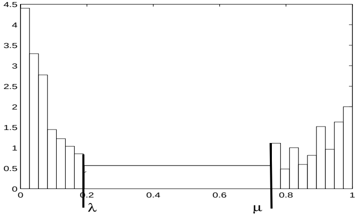
$\widehat{p}$ selected from 500 trials with $m = 1000$.

- A large amount of $\widehat{p}$ are larger than 50.
- The choice of $\widehat{p}$ is not time consuming.

CV-based estimator of π_0

Estimation procedure

1. For each partition $I \in \mathcal{I}$, define $\widehat{p}(I) = \mathrm{Argmin}_p \widehat{MSE}(I;p)$.
2. Find the best partition $\widehat{I} = \mathrm{Argmin}_{I \in \mathcal{I}} \widehat{R}_{\widehat{p}(I)}(I)$.
3. From $\widehat{I}$, get $(\widehat{\lambda}, \widehat{\mu})$.
4. Compute the estimator $\widehat{\pi}_0^{CV} = \frac{\mathrm{Card}\left\{i : P_i \in [\widehat{\lambda}, \widehat{\mu}]\right\}}{m(\widehat{\mu} - \widehat{\lambda})}$.



Consistency of $\widehat{\pi}_0^{CV}$

Theorem

- If Assumption $(V_{\lambda,\mu})$ is fulfilled for $0 \leq \lambda^* < \mu^* \leq 1$,
- if $[\lambda^*, \mu^*]$ is the widest interval such that g is constant,

then

$$\widehat{\pi}_0^{CV} \xrightarrow[m \to +\infty]{P} \pi_0 \quad .$$

Remarks:

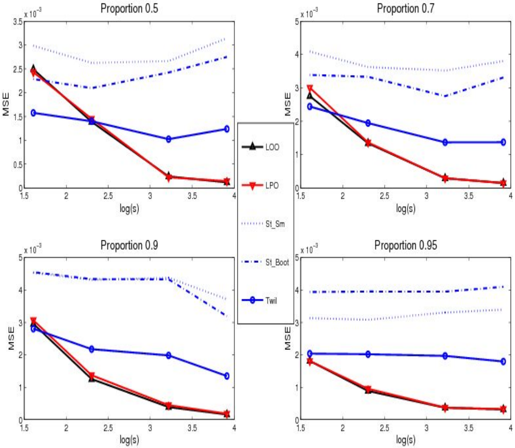
- This procedure is fully data-driven.
- It does not require any additional computational cost.

Simulation experiments

Assumption (V_1) fulfilled

Design

- $f_1(t) = s(1-t)^{s-1}$,
- $s \in \{5, 10, 25, 50\}$.
- $\pi_0 \in \{0.5, 0.7, 0.9, 0.95, 0.99\}$
- $m = 1000$ (sample size),
- 500 repetitions.



$\longrightarrow$ Best results are obtained by $\widehat{\pi}_0^{CV}$.

Simulation experiments

Assumption (V_1) fulfilled

π_0	0.7			0.99		
	Bias	Std	MSE	Bias	Std	MSE
LPO	1.4	3.4	**13.6** 10^{-2}	0.3	3.4	**11.4** 10^{-2}
St_{Sm}	-0.9	6.0	36.2 10^{-2}	-2.3	4.4	24.9 10^{-2}
St_{Boot}	-3.3	4.7	33.3 10^{-2}	-4.1	5.2	43.2 10^{-2}
Twil	-1.5	4.2	19.4 10^{-2}	-3.5	4.3	30.6 10^{-2}
ABH	27	2.4	7.6	1.0	0.1	0.9 10^{-2}

$(s = 10)$

Simulation experiments

Assumption $(V_{\lambda,\mu})$ fulfilled

Design

- Data are generated according to

$$\pi_0 \mathcal{N}(0, 2.510^{-2}) + \frac{1-\pi_0}{2}\left[\mathcal{N}(a,\theta^2) + \mathcal{N}(b,\nu^2)\right], \quad -a, b > 0 \ .$$

- For each $1 \leq i \leq m$,

$$\mathcal{H}_{i,0}\colon\ \mathbb{E}(Y_i) = 0 \quad \text{vs} \quad \mathcal{H}_{i,1}\colon\ \mathbb{E}(Y_i) > 0.$$

- $\pi_0 \in \{0.25, 0.5, 0.7, 0.8, 0.9\}$
- $m = 1000$
- 200 trials

Assumption $(V_{\lambda,\mu})$ fulfilled

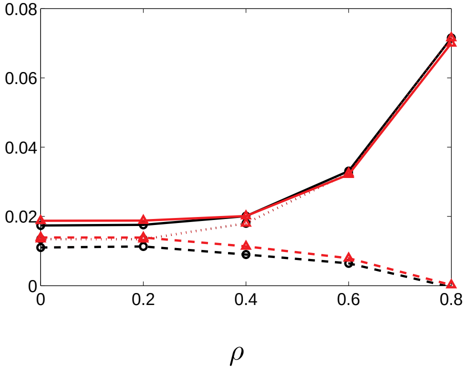
π_0	0.25			0.7			0.9		
	Bias	Std	MSE	Bias	Std	MSE	Bias	Std	MSE
LPO	5.5	6.2	**0.7**	5.3	4.4	**0.5**	4.2	2.7	**0.2**
St_{Sm}	75.0	0	56.0	30.0	0	9.0	9.9	0.2	**1.0**
St_{Boot}	43.2	3.2	18.7	17.4	1.6	3.0	5.4	1.6	**0.3**
Twil	73.2	2.5	53.6	27.4	2.3	8.0	8.0	1.3	**0.7**
ABH	45.5	5.4	21.0	19.8	3.1	4.0	7.4	1.3	**0.6**

- Except $\widehat{\pi}_0^{CV}$, every estimator overestimates π_0.
- This trends disappears as π_0 grows.

Simulation experiments

Dependence

- Data are split into b disjoint blocks.
- Correlation is generated using a *mixed-model*.
- Correlation intensity is given by $0 \leq \rho \leq 1$.



III FDR control

(C. and Robin (09), arXiv:0804.1189)
(C. and Robin (08), CSDA)

Plug-in adaptive procedure

Definition

Reject all hypotheses with p-values less than or equal to $T_\alpha\left(\widehat{\pi}_0^{CV}\right)$, where the threshold $T_\alpha(\cdot)$ is given by

$$T_\alpha(\theta) = \sup\{t \in (0,1) : \widehat{Q}_\theta(t) \leq \alpha\}, \quad \forall \theta \in [0,1] \ ,$$

$$\widehat{Q}_\theta(t) = \frac{m\,\theta\,t}{\mathrm{Card}\left(\{i \mid P_i \in \mathcal{R}\left(P_1, \ldots, P_m\right)\}\right)} \ .$$

Proposition

The step-up procedure $T_\alpha\left(\widehat{\pi}_0^{CV}\right)$ is equivalent to the BH-procedure with m replaced by $\widehat{\pi}_0^{CV} m$.

Asymptotic control of FDR

Theorem

- $\alpha \in [0, \pi_0[$.
- For $\delta > 0$, $\widehat{\pi}_0^{\delta} = \widehat{\pi}_0^{CV} + \delta$.
- Assumption (V_{λ^*,μ^*})
- f_1 is differentiable
- f_1 nonincreasing on $[0, \lambda^*]$, nondecreasing on $[\mu^*, 1]$.

Then

$$FDR\left(T_\alpha\left(\widehat{\pi}_0^{CV}\right)\right) \leq \alpha + o(1) \ .$$

Simulation experiments

Control of FDR and FNR

- $\alpha = 0.15$,
- *FNR* (between brackets) is the number of true alternatives missed by the procedure.
- *Oracle* is the plug-in procedure where the true π_0 is used.

s	π_0	$T_\alpha\left(\widehat{\pi}_0^{CV}\right)$	BH	Oracle
10	0.5	14.74 (25.69)	6.94 (96.83)	15.02 (23.22)
	0.7	15.14 (96.36)	10.29 (99.16)	15.12 (96.03)
	0.95	14.65 (99.76)	14.37 (99.77)	14.95 (99.74)
25	0.5	14.88 (0.88)	7.48 (17.72)	15.04 (0.79)
	0.7	14.69 (22.83)	10.47 (61.00)	14.84 (21.93)
	0.95	14.35 (99.16)	13.19 (99.23)	14.19 (99.14)

$\mathbb{IV}$ Local FDR estimation

(Robin et al. (2007), CSDA)

Local FDR and π_0

Local FDR (locFDR)

$$\forall 1 \leq i \leq m, \qquad locFDR\left(P_i\right) := \mathbb{P}\left[\mathcal{H}_{i,0} \text{ is true} \mid P_i\right] \quad .$$

- Unlike FDR, locFDR yields a local information about $\mathcal{H}_i$.
- With the mixture model:

$$locFDR\left(P_i\right) = \frac{\pi_0}{\pi_0 + (1-\pi_0) f_1\left(P_i\right)} = \frac{\pi_0}{g\left(P_i\right)} \quad .$$

$\longrightarrow$ Depends on π_0 and f_1.

Strategy

- Estimate π_0
- Estimate g, the density of the p-values.

Density estimation

Weighted kernel estimator

Use a *weighted kernel* as an estimator of f_1.

$$\forall h > 0, \quad \widehat{f}_{1,h}(P_i) := \sum_{j=1}^{m} \frac{\omega_j}{\sum_{k=1}^{m} \omega_k} K\left(\frac{P_j - P_i}{h}\right),$$

$$\forall i, \qquad \omega_i = 1 - locFDR\left(P_i\right) \ .$$

$$\left(locFDR(P_i) = \frac{\pi_0}{\pi_0 + (1-\pi_0) f_1(P_i)} \right)$$

$\longrightarrow$ Iterative algorithm to estimate *locFDR* and f_1.

Iterative algorithm

Algorithm

For a given π_0:

1. Initialize $\left(locFDR^0(P_1), \ldots, locFDR^0(P_m)\right)$,
2. Estimate f_1,
3. Estimate g
4. Update $\left(locFDR^1(P_1), \ldots, locFDR^1(P_m)\right)$.
5. Stopping rule:
   Repeat Step 2–3 until *locFDR* estimates are stable.

$\longrightarrow$ A preliminary estimate of π_0 must be plugged in.

Remark:

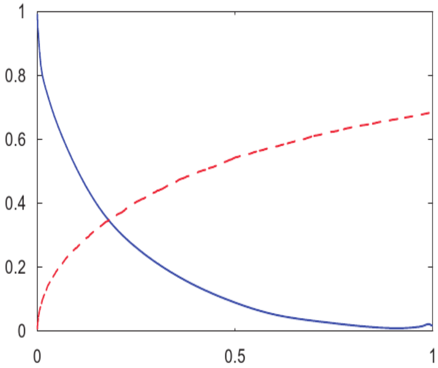
This algorithm has been proved to converge.

R-package kerfdr

- A R-package called *kerfdr* has been implemented.
- Available on the CRAN at:

http://cran.at.r-project.org/web/packages/kerfdr/index.html.

- Enables semi-supervised (or unsupervised) data.
- Allows to deal with discrete p-values (truncation problems).



Conclusion

- $\widehat{\pi}_0^{CV}$ is more accurate than several other existing ones.
- This estimator does not induce any additional computational cost.
- It is robust to various realistic assumptions on the p-value distribution.
- Enables yields a new plug-in procedure, which (asymptotically) controls FDR at the desired level.

Conclusion

- $\widehat{\pi}_0^{CV}$ is more accurate than several other existing ones.
- This estimator does not induce any additional computational cost.
- It is robust to various realistic assumptions on the p-value distribution.
- Enables yields a new plug-in procedure, which (asymptotically) controls FDR at the desired level.

Thank you.