Theoretical guidelines for (high-dimensional) data analysis

Christophe Giraud

Université Paris-Sud

M2 Data Science

Why this course?

Goal of the lectures

1. to provide some theoretical guidelines for (high-dimensional) data analysis;
2. to highlight some delicate issues;
3. to learn to read a research paper: find the take-home message and understand the limits of the message;
4. to learn to question research papers.

Maths or No-Maths inside?

- we will speak all along about maths results,
- but we will <u>not</u> prove maths results during the lectures.

You will learn to <u>question</u> and <u>understand</u> theoretical papers, not to <u>produce</u> them.

An interesting quote

"This is the first time that we two read an article in statistics on a state-of-the-art subject in detail. It was really not obvious at the beginning. We did not understand the notations and were not familiar with this domain, etc. But after reading it 4 or 5 times, the structure and the logic of the paper became clearer and clearer to us and we became more and more confident. So we would like to say that we are happy to have such experience of mini research in statistics. This will help us to be more confident when possible challenges in this domain occur to us in the future."

(Data Science 2016-17)

Organisation

Structures of the lectures

- Discussion of the paper from the previous session
- Lecture to explain the topic of the session and some related issues
- first (supervised) reading of a research paper

Between the lectures

Full reading of the research paper

Final "project"

Explain and discuss one of the paper exposed during the lectures (see below).

Please, ask questions!

Topics

1. False discoveries, multiple testing, online issue
2. Strength and weakness of the Lasso
3. Adaptive data analysis
4. Unsupervised dimension reduction: some limits
5. Robust learning

⚠ No deep learning inside!

Requirement

Download the papers before the lectures

http://www.math.u-psud.fr/~giraud/MSV/statsDS.html

Evaluation

Project

Due to mid-february

Mandatory

To attend to all lectures

Rapport à rendre: en binôme

The reports must be sent by email by February 15 in a zip file including:

1. the report in pdf format (10 to 20 pages)
2. if there is some numerics: the notebook (or source code)

Attendu

1) présenter le contexte et les principaux résultats du papier (moitié du rapport maximum).

Il ne s'agit pas de donner un panorama complet du papier, et encore moins un compte rendu littéral. Il s'agit de:

- sélectionner les résultats qui vous semblent les plus importants
- expliquer intuitivement les résultats et (si approprié) les idées sous-jacente à l'algorithme étudié
- commenter leurs implications

2) faire une analyse critique du papier.

- quelles portées des résultats? quelles limitations?
- quel message retenir?

Attendu (suite)

3) procéder à une exploration personnelle, de nature mathématique ou numérique

Côté maths: cela peut être

- expliquer les grandes lignes d'une preuve, les points cruciaux et proposer (de façon argumentée) des possibles extensions pour généraliser ou transposer les résultats.
- une étude théorique comparative des résultats à d'autres résultats récents de la littérature

Côté numérique: il s'agit d'explorer une ou plusieurs problématiques pratiques:

- définir la problématique, le plan d'expérience pour étudier cette problématique (justifier le plan);
- réaliser les expériences et rédiger un notebook explicatif (ou à défaut un code source bien annoté pour comprendre ce qui est fait)
- faire un choix pertinent des résultats à montrer et à commenter
- commenter les résultats et conclure

Critères d'évaluation

Evaluation

1. compréhension de l'article (contexte, motivation, apport, contresens, etc)
2. prise de recul (capacité à expliquer les idées et résultats, leurs implications et leur portée/limite)
3. analyse personnelle:
   - **maths:** compréhension et discernement des points importants, profondeur d'analyse et importance de la contribution personnelle
   - **numérique:** intérêt de la problématique étudiée, pertinence des expériences, qualités des résultats, de leur analyse et discussion

https://www.math.u-psud.fr/~giraud/MSV/statsDSevaluation.html

Projet

Projet

- en binôme
- prendre un des articles du cours et
  1. expliquer le contexte et son message
  2. en cerner/discuter les limites
  3. questionner/discuter numériquement ou théoriquement le papier
- A rendre pour le 15 février minuit.

The reports must be sent by email in a zip file including:

- the report in **pdf format**: 10 to 20 pages;
- the source code for the numerics.

Let's start!

False discoveries

Scientific and societal concern

THE MAGAZINE: SEPTEMBER 26, 2011

THE NEW YORKER

SUBSCRIBE | IN THE MAGAZINE | BLOGS | AUDIO & VIDEO | LISTINGS

POLITICS | PROFILES | THE TALK OF THE TOWN | COMMENT | TABLE OF CONTENTS | THE FINANCIAL PAGE

ANNALS OF SCIENCE

THE TRUTH WEARS OFF

Is there something wrong with the scientific method?

BY JONAH LEHRER

DECEMBER 13, 2010

On September 18, 2007, a few dozen neuroscientists, psychiatrists, and drug-company executives gathered in a hotel conference room in Brussels to hear some startling news. It had to do with a class of drugs known as atypical or second-generation antipsychotics, which came on the market in the early nineties. The drugs, sold under brand names such as Abilify, Seroquel, and Zyprexa, had been tested on schizophrenics in several large clinical trials, all of which had demonstrated a dramatic decrease in the subjects' psychiatric symptoms. As a result, second-generation antipsychotics had become one of the fastest-growing and most profitable pharmaceutical classes. By 2001, Eli Lilly's Zyprexa was generating more revenue than Prozac. It remains the company's top-selling drug.

Many results that are rigorously proved and accepted start shrinking in later studies.

KEYWORDS

Scientific Experiments; Decline Effect; Replicability; Scientists; Statistics; Jonathan Schooler; Scientific Theories

Lack of reproducibility

The Economist | World politics | Business & finance | Economics | Science & technology

Unreliable research

Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition

Timekeeper Like 22k Tweet

Systematic attemps to replicate widely cited priming experiments have failed

- Amgen could only replicate 6 of 53 studies they considered landmarks in basic cancer science
- HealthCare could only replicate about 25% of 67 seminal studies
- etc

What has gone wrong?

Main Flaws

- Statistical issues
- Publication Bias
- Lack of check
- Publish or Perish
- Exponential growth of publications
- Narcissism

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video

Archive > Volume 496 > Issue 7446 > Editorial > Article

NATURE | EDITORIAL

Announcement: Reducing our irreproducibility

24 April 2013

PDF | Rights & Permissions

Over the past year, *Nature* has published a string of articles that highlight failures in the reliability and reproducibility of published research (collected and freely available at go.nature.com/huhbyr). The problems arise in laboratories, but journals such as this one compound them when they fail to exert sufficient scrutiny over the results that they publish, and when they do not publish enough information for other researchers to assess results properly.

Be patient

I don't care, not an issue for my future start-up...

Are you sure?

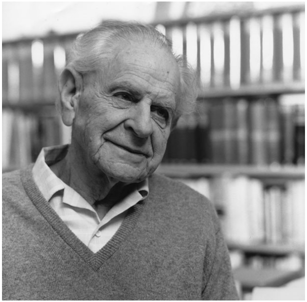
Back to the basics

Status of science

An hypothesis or theory can only be empirically <u>tested</u>.

Predictions are deduced from the theory and compared with the outcomes of experiments.

An hypothesis can be falsified or corroborated.



Karl Popper (1902-1994)

An historical example (1935)

The lady testing tea

A lady claims that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup.

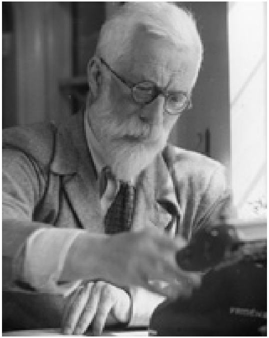
Experiment

8 cups are brought to the lady and she has to determine whether the milk or the tea was added first.

Test

Modeling: the success $X_1, \ldots, X_8$ are i.i.d. with $\mathcal{B}(\theta)$ distribution.

Test: $\mathcal{H}_0 : \theta = 1/2$ versus $\mathcal{H}_1 : \theta > 1/2$



R.A. Fisher (1890-1962)

Hypothesis testing

Testing statistics

We reject the hypothesis $\mathcal{H}_0$: "the lady cannot discriminate" if the number of success

$$\widehat{S} = X_1 + \ldots + X_8$$

is larger than some threshold s_{th}.

Distribution of the test statistics

Under $\mathcal{H}_0$ the distribution of $\widehat{S}$ is $\mathrm{Bin}(8, 1/2)$.

Choice of the threshold

We choose the threshold s_{th} such that the probability to reject wrongly $\mathcal{H}_0$ is smaller than α (e.g. 5%)

$$\mathbb{P}\left(\mathrm{Bin}(8, 1/2) \geq s_{th}\right) = \alpha.$$

p-values

p-value

The p-value of the observation $\widehat{S}(\omega_{obs})$, is the probability to observe $\widehat{S}$ larger than $\widehat{S}(\omega_{obs})$ when $\mathcal{H}_0$ is true

$$\hat{p}(\omega_{obs}) = G_{1/2}\left(\widehat{S}(\omega_{obs})\right), \quad \text{where } G_{1/2}(s) = \mathbb{P}\left(\mathrm{Bin}(8, 1/2) \geq s\right).$$

Remark

Since

$$\widehat{S}(\omega_{obs}) \geq s_{th}(\alpha) \iff \hat{p}(\omega_{obs}) \leq \alpha$$

we reject $\mathcal{H}_0$ if the p-value is smaller than α.

Foundations of science

Science is largely based on p-values. An hypothesis/theory is falsified or corroborated depending on the size of the p-value of the outcome of some experiment(s)/observation(s).

Where does-it go wrong?

Publications issues

- Publication bias
- Publishing pressure
- Lack of check: replication is not "recognized" and exponential growth of the number of scientific publications

Small sample size

Cost of adding individuals in experiments

Statistical issues

Collect data first $\longrightarrow$ ask (many) questions later

Issue of multiple testing (one aspect of the curse of dimensionality)

Multiple testing

Analyse différentielle

Question

Est-ce que le niveau d'expression d'un gène diffère entre une condition A (individu sain) et une condition B (individu malade)?

Données issues d'une expérience

Conditions	Mesures
A	$X_{A1}, \ldots, X_{Ar}$
B	$X_{B1}, \ldots, X_{Br}$

Objectif

Différentier entre les 2 hypothèses

$\mathcal{H}_0$: "la moyenne des X_{Ai} et des X_{Bi} sont les mêmes"

$\mathcal{H}_1$: "la moyenne des X_{Ai} et des X_{Bi} sont différentes"

Exemple de test

$Y_i = X_{Ai} - X_{Bi}$ pour $i = 1, \ldots, r$.

Rejet de $\mathcal{H}_0$ si

$$\widehat{S} := \frac{|\overline{Y}|}{\sqrt{\widehat{\sigma}^2 / r}} \geq s = \text{seuil à fixer}$$

avec $\widehat{\sigma}^2 = \overline{\mathrm{var}}(Y)$

Choix du seuil pour contrôler le risque de rejeter $\mathcal{H}_0$ à tort

$$\mathbb{P}_{\mathcal{H}_0}(\widehat{S} \geq s_\alpha) \leq \alpha$$

Test : $T = \mathbf{1}_{\widehat{S} \geq s_\alpha}$

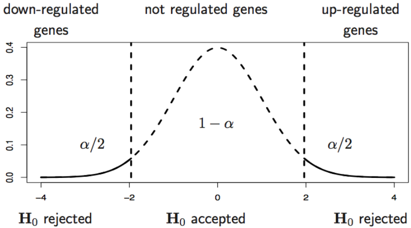
Modèle statistique

$$X_{Ai} \overset{i.i.d.}{\sim} \mathcal{N}(\mu_A, \sigma_A^2) \quad \text{and} \quad X_{Bi} \overset{i.i.d.}{\sim} \mathcal{N}(\mu_B, \sigma_B^2)$$

On a alors $\mathcal{H}_0 \; = \; \text{“}\mu_A = \mu_B\text{”}$.

Loi sous $\mathcal{H}_0$

$$\widehat{S} = \frac{\overline{Y}}{\sqrt{\widehat{\sigma}^2/r}} \overset{\mathcal{H}_0}{\sim} \mathcal{T}(r-1) \quad \text{(student à } r-1 \text{ degrés de liberté)}$$



Choix du seuil s_α

On prend s_α tel que $\mathbb{P}(|\mathcal{T}(r-1)| \geq s_\alpha) = \alpha$

Exemple : analyse différentielle de 1 gène

Data

i	X_A	X_B	Y
1	4.01	4.09	-0.08
2	0.84	0.97	-0.12
3	4.45	3.92	-0.53
4	4.73	6.01	1.28
5	6.16	6.01	0.15
6	4.23	6.48	-2.26
7	4.70	5.85	-1.15
8	10.65	11.02	-0.37
9	2.02	4.18	-2.16
10	3.96	5.19	-1.23
mean	4.58	5.37	-0.80
std	2.60	2.55	0.96

Test

r	10
$\overline{Y}$	-0.80
$\sqrt{\widehat{\sigma}^2}$	0.96
$\widehat{S}$	2.62
p-value	0.03

p-value d'un test

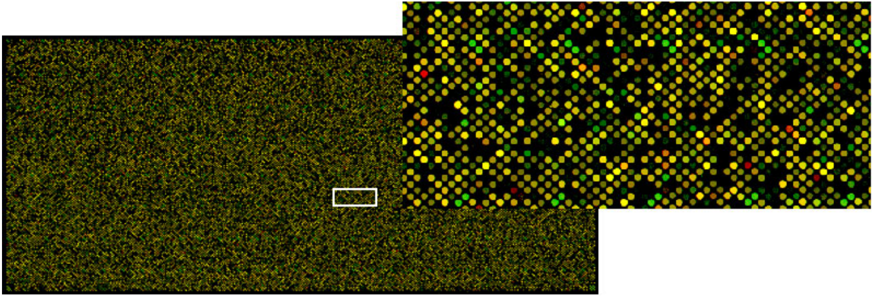
Valeur de α pour laquelle le test change de réponse ($s_{\widehat{p}} = \widehat{S}$)

Si p-value $\leq \alpha$: $s_\alpha \leq \widehat{S}$
le test rejette $\mathcal{H}_0$

Si p-value $> \alpha$: $s_\alpha > \widehat{S}$
le test accepte $\mathcal{H}_0$

Genomic data

We want to compare the gene expression levels for healthy/ill people.



Whole Human Genome Microarray covering over 41,000 human genes and transcripts on a standard 1" x 3" glass slide format

High-dimensional data

we measure 41,000 gene expression levels simultaneously!

Blessing?

Des nouvelles perspectives médicales

Objet

Personnaliser les traitements anti-cancer en combinant données cliniques et génomiques

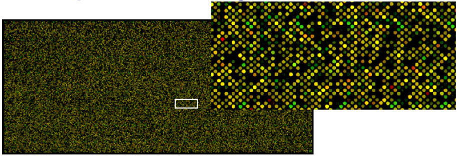
Moyens

RNAseq, puces CGH, etc

Questions

- Quelle prévision de survie?
- Quel "type" de cancer?
- Quel traitement adopter?
- etc

Comparaisons multiples : analyse différentielle de p gènes



Une puce microarray permet de comparer le niveau d'expression de milliers de gènes en même temps.

Résultat: liste de p-value classées par ordre croissant

gènes	p-value
2014	$< 10^{-16}$
1078	$6.66\ 10^{-16}$
123	$2.66\ 10^{-15}$
548	$1.02\ 10^{-11}$
3645	$3.09\ 10^{-10}$
$\vdots$	$\vdots$

Quels gènes sont statistiquement différentiellement exprimés?

Ceux qui ont une p-value $\leq 5\%$?

Combien de fausses découvertes?

An illustrative example

Assume that:

- 200 genes are differentially expressed
- you keep the p-values $\leq 5\%$

How many False Discoveries?

$$\mathbf{E}[\text{False Discoveries}] = \frac{5}{100} * (41000 - 200) = 2040$$

10 false discoveries for 1 discovery!

☹

Blessing?

☺ we can sense thousands of variables on each "individual" : potentially we will be able to scan every variables that may influence the phenomenon under study.

☹ the curse of dimensionality : separating the signal from the noise is in general almost impossible in high-dimensional data and computations can rapidly exceed the available resources.

Formalisation

Reminder: Tests

Tests

Let $\{\mathbb{P}_\theta : \theta \in \Theta\}$ be a family of probability distributions. We want to test

$$\mathcal{H}_0 : \theta \in \Theta_0 \quad \text{against} \quad \mathcal{H}_1 : \theta \in \Theta_1$$

(generalized) *p*-value

A (generalized) *p*-value is any (observable) random variable $\hat{p}$ fulfilling

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\hat{p} \leq u) \leq u : \quad \forall\, u \in [0, 1].$$

Canonical example

Test statistic

Assume that the test can be written as $\widehat{T} = \mathbf{1}_{\widehat{S} \geq \text{threshold}}$ where $\widehat{S}$ can be computed from the data.

Tail function

$$G_\theta(s) = \mathbb{P}_\theta(\widehat{S} > s) \quad \text{(non-increasing)}$$

Associated p-value

the p-value of the observation $\widehat{S}(\omega^{obs})$ is

$$\hat{p}(\omega^{obs}) := \sup_{\theta \in \Theta_0} G_\theta(\widehat{S}(\omega^{obs}))$$

(the larger $\widehat{S}(\omega^{obs})$, the smaller the p-value)

Canonical example (II)

Distribution

Under $\mathcal{H}_0$, the random variable $\hat{p}(\omega) = \sup_{\theta\in\Theta_0} G_\theta(\widehat{S}(\omega))$ is stochastically larger than the uniform distribution on $[0,1]$:

$$\sup_{\theta\in\Theta_0} \mathbb{P}_\theta(\hat{p} \leq u) \leq u : \quad \forall\, u \in [0,1].$$

Proof.

We set $F_\theta(t) = \mathbb{P}_\theta(\widehat{S} \leq t) = 1 - G_\theta(t)$, which is increasing with t.

For any $\theta_0 \in \Theta_0$, the random variable $U = F_{\theta_0}(\widehat{S})$ follows, under $\mathbb{P}_{\theta_0}$, a uniform distribution on $[0,1]$.

Then, under $\mathbb{P}_{\theta_0}$,

$$\hat{p} \;=\; \sup_{\theta\in\Theta_0} G_\theta(\widehat{S}) \;\geq\; G_{\theta_0}(\widehat{S}) \;=\; 1 - F_{\theta_0}(\widehat{S}) \;=\; 1 - U \stackrel{(d)}{=} U.$$

Reminder: p-values and level

Level

$$\text{level} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\text{test} = 1)$$

In order to have a test of level α we reject $\mathcal{H}_0$ when $\hat{p} \leq \alpha$.

Actually,

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\hat{p} \leq \alpha) \leq \alpha.$$

Multiple testing: statistical setting

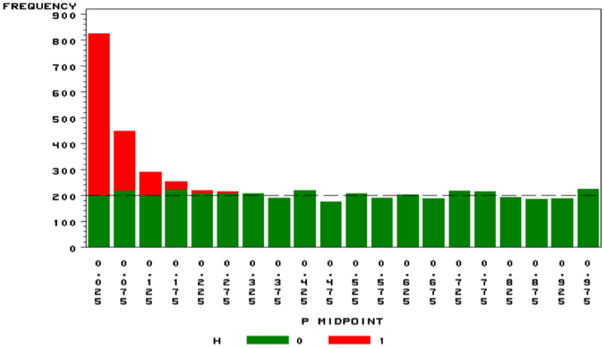
Multiple testing

we perform m tests simultaneously:

$$\begin{array}{lcl} \text{Test } 1 & : & \mathcal{H}_0^{(1)} : \theta \in \Theta_0^{(1)} \text{ against } \mathcal{H}_1^{(1)} : \theta \in \Theta_1^{(1)} \\ \dots & : & \qquad\qquad\qquad \dots \\ \text{Test } m & : & \mathcal{H}_0^{(m)} : \theta \in \Theta_0^{(m)} \text{ against } \mathcal{H}_1^{(m)} : \theta \in \Theta_1^{(m)} \end{array}$$

For these m tests, we collect the p-values, $\hat{p}_1, \dots, \hat{p}_m$.

Typical p-values distribution



Multiple testing procedure

Multiple testing procedure

$$R : (\widehat{p}_1, \ldots, \widehat{p}_m) \to \widehat{R} = R(\widehat{p}_1, \ldots, \widehat{p}_m) = \left\{ i : \mathcal{H}_0^{(i)} \text{ rejected} \right\} \subset \{1, \ldots, m\}$$

We set: $I_0 = \left\{ i \in \{1, \ldots, m\} : \mathcal{H}_0^{(i)} \text{ is true} \right\}$, and $m_0 =$Card(I_0).

False Positive

$$\mathrm{FP} = \mathrm{card}(\widehat{R} \cap I_0)$$

Example

We want to reject the smallest p-values.

Fixed level rejection set

A natural rejection set is

$$\widehat{R} = \{i : \widehat{p}_i \leq \tau\}$$

where τ is a fixed level.

- For this choice, each test has a level τ ☺
- Yet, on average, we will have $\tau \times m_0$ false discoveries. ☹

Example: for $\tau = 5\%$, $p = 41000$ and $m_0 = 40800$ we have around 2040 false discoveries!

Bonferroni correction

Bonferroni

We can always choose $\tau = \alpha/m$ since

$$\mathbb{E}[FP] = \sum_{i \in I_0} \mathbb{P}(\hat{p}_i \leq \tau) \leq m_0 \tau \leq m\tau \leq \alpha.$$

With this choice:

- The probability to (wrongly) reject one of the $\mathcal{H}_0^{(i)}$ is small ☺
- But the tests lack of power: we almost never detect $\mathcal{H}_1^{(i)}$. ☹

FDR

Motivation

The Bonferroni correction is too conservative. Instead of controlling $\mathbb{E}[FP]$, we control the proportion of false discoveries among all the discoveries.

FDR (False Discovery Rate)

The False Discovery Rate is

$$FDR = \mathbb{E}\left[\frac{|\widehat{R} \cap I_0|}{|\widehat{R}|}\,\mathbf{1}_{|\widehat{R}|>0}\right] = \mathbb{E}\left[\frac{FP}{|\widehat{R}|}\,\mathbf{1}_{|\widehat{R}|>0}\right]$$

What procedure to ensure $FDR \leq \alpha$?

Benjamini-Hocheberg

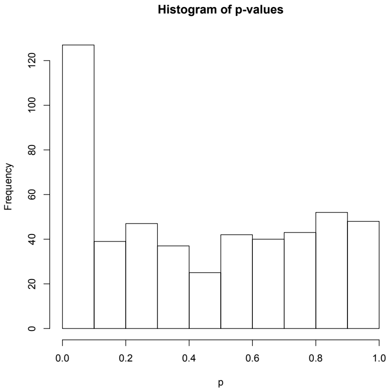
Benjamini-Hocheberg procedure

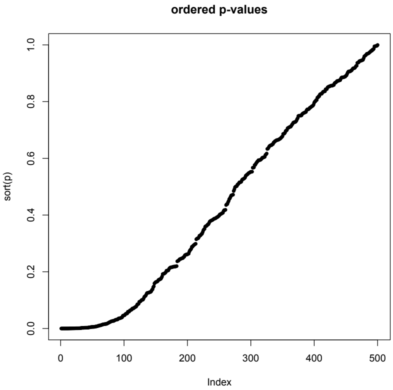
1. Write $\hat{p}_{(1)} \leq \ldots \leq \hat{p}_{(m)}$ for the ranked p-values;
2. Compute $\hat{k} = \max\left\{k : \hat{p}_{(k)} \leq \alpha k/m\right\}$;
3. Reject
$$\widehat{R} = \left\{i : \hat{p}_i \leq \alpha \frac{\hat{k}}{m}\right\}.$$

Remarks

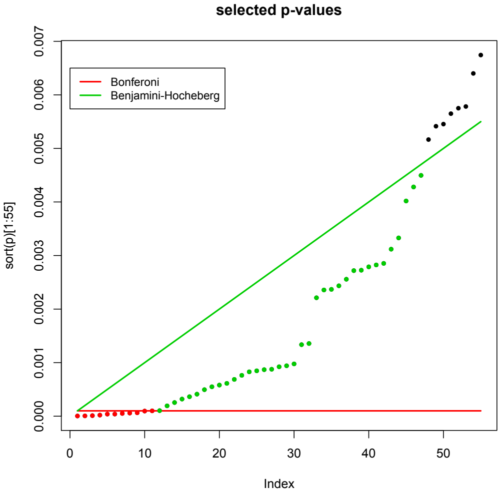
- widely used in science
- always less stringent than Bonferroni

Example : p-values





BH and Bonferroni ($\alpha = 5\%$)



Bonferroni in red and Benjamini-Hocheberg (BH) in green

Theory?

Under conditions

Under some conditions on the distributions of the p-values,

$$FDR(BH) \leq \alpha.$$

Without conditions?

Must replace α by $\alpha / \log(m)$ in the BH procedure.

Conclusion

Multiple testing issue

- a correction is needed when performing multiple testing, in order to avoid overwhelming false discoveries.
- *FDR* control is widely used in science (BH-procedure yet)
- Yet, false discoveries are still here....

Why?

Why?

- Bad statistical modelisation
- Publication bias
- Publishing pressure
- Lack of check...

Something else with multiple testing with data collections?

The problem of online FDR control

Re-use of data

Sustained effort for making easily available open-access data: the same data-set is used by many different users.
Amplified by open-data policies.

Issue

Each data-scientist controls the FDR for its own experiment. But there is no overall control of the FDR (reuse of data).

Online constraints

When computing p-values the data scientist ignores

- the nature and the total number of the future studies;
- the p-values of the future studies.

Online control of false discoveries

Online p-values

(infinite) sequence of p-values: $\widehat{p}_1, \widehat{p}_2, \ldots$

FWER control

Choose $\alpha_1, \alpha_2, \ldots \in [0,1]$ such that $\sum_{j\geq 1} \alpha_j = \alpha$ and reject if $\widehat{p}_j \leq \alpha_j$.

Example: $\alpha_j = \alpha/2^j$.

Issue

No rejection when j becomes large

Online control of false discoveries

Online levels

Reject ($R_j = 1$) if $\widehat{p}_j \leq \alpha_j(R_1, \ldots, R_{j-1})$.

Which $\alpha_j(R_1, \ldots, R_{j-1})$?

α-investing principle (D. Foster and R. Stine, 2008)

Update a wealth W_j which

- increases if $R_j = 1$,
- decreases if $R_j = 0$.

and at each step, choose α_j according to W_{j-1}.

Example: LORD algorithm

Input

- a decreasing sequence $(\gamma_\ell)_{\ell\geq 1}$ fulfilling $\sum_\ell \gamma_\ell = 1$;
- a level α and $W_0 = \alpha/2$.

Algorithm (A. Javanmard, A. Montanari, 2017)

For $j = 1, 2, \ldots$

- Set $\alpha_j = \gamma_{j-\tau_j} W_{\tau_j}$, where $\tau_j =$ last time t such that $R_t = 1$;
- Reject if $\widehat{p}_j \leq \alpha_j$;
- Update $W_j = W_{j-1} + R_j\alpha/2 - \alpha_j$.

(\alpha

FDR control

Under appropriate independence

$$\sup_{j\geq 1} FDR(R_1, \ldots, R_j) \leq \alpha.$$

Quality Preserving Data-bases

Issue

Even with LORD, after a (long) sequence of $R_j = 0$, it is very hard to reject again.

Quality Preserving Data-bases (E. Aharoni and S. Rosset, 2014)

Idea: Update Data-bases in order to keep the power (pay for testing)

Interesting idea.... But hard to change the practice ☹

ZZZZZZZZZZZZZ

Where is the link with my start-up?

Investigated paper

- Fanny Yang, Aaditya Ramdas, Kevin Jamieson and Martin J. Wainwright. *A framework for Multi-A(rmed)/B(andit) testing with online FDR control*. NIPS, 2017
  `https://arxiv.org/abs/1706.05378`

Typical organisation of a paper

Introduction (section 1)

- topic
- related literature
- contributions
- organisation of the paper

Exposition of the results

- setting, definitions, etc
- statement of the results
- numerics
- (additional results)

Proofs

Usually in the last section and appendix

Further references

Multiple testing and FDR

- Introduction to High-dimensional Statistics: Chapter 8.

Online FDR control and QDP

- E. Aharoni and S. Rosset. *Generalized α-investing: definitions, optimality results and application to public databases.* (2014) JR Statist. Soc. 76, pp. 771–794.
- S. Rosset, E. Aharoni, H. Neuvirth. *Novel statistical tools for management of public databases facilitate community-wide replicability and control of false discovery.* (2014) Genet Epidemiol.
- A. Javanmard, A. Montanari. *Online Rules for Control of False Discovery Rate and False Discovery Exceedance.* (2016) arXiv:1603.09000

Reliability of scientific findings?

- Summary and discussion of: "Why Most Published Research Findings Are False". Dallas Card and Shashank Srivastava (CMU).