Asymptotic control of the False Discovery Rate under Dependence

Controllo asintotico del False Discovery Rate nel caso di dipendenza

Alessio Farcomeni[12]
Dipartimento di Statistica, Probabilità e Statistiche Applicate
Universita' "La Sapienza"
alessio.farcomeni@uniroma1.it

Riassunto: Si dimostra che il metodo classico di controllo del False Discovery Rate (FDR) rimane asintoticamente valido nel caso di dipendenza tra le statistiche test, sotto ampie ipotesi. Seguendo Genovese and Wasserman (2001), proponiamo inoltre un metodo di aumento della potenza dei test, utilizzando uno stimatore iterativo della proporzione di ipotesi nulle false. I risultati vengono illustrati tramite simulazione.

Keywords: Multiple Hypothesis Testing, False Discovery Rate, α-mixing

1. Introduction

In a breakthrough paper, Benjamini and Hochberg (1995) proposed a new error measure, the False Discovery Rate (FDR) to be used in multiple testing; and a method of controlling this new measure. If the test statistics are independent, they proved the method allows for a sufficiently low number of false rejections, while being more powerful than other classical methods, like the well known Bonferroni correction. Benjamini and Yekutieli (2001) discovered conditions under which the technique works under dependence. Unfortunately, these are not easy to check. In this paper we prove that, as the number of tests grows, very broad conditions on the dependence of the variables are enough to make the classic procedure work. We also propose a way for improving the classic method under dependence, using an iterative estimator of the proportion of false null hypotheses.

2. Background

Consider a multiple testing situation in which m tests are being performed. Suppose m_0 of the m hypotheses are true, and m_1 are false. Table 1 shows a categorization of the outcome of the tests. R is the number of rejections. Note that $N_{1|0}$ and $N_{0|1}$ are the exact (unknown) number of errors committed, respectively false positives and false negatives. $N_{1|1}$ and $N_{0|0}$ are respectively the number of hypotheses correctly rejected and correctly retained.

[1]Piazzale Aldo Moro, 5 00185 Roma
[2]This is joint work with Larry Wasserman and Chris Genovese, CMU, Pittsburgh, PA, USA.

Table 1: *Categorization of the outcome*

	H_0 **not rejected**	H_0 **rejected**	**Total**
H_0 **True**	$N_{0\|0}$	$N_{1\|0}$	m_0
H_0 **False**	$N_{0\|1}$	$N_{1\|1}$	m_1
Total	$m - R$	R	m

In the frequentist framework, one controls the probability of Type I error (i.e., false rejection) of each test at a certain level (say, 0.05). When more than one hypothesis is verified simultaneously, this can result in an unsatisfactory number of false rejections. Hence the simultaneity must be taken into account, and a different error measure must be considered. Traditional methods for multiple testing, like the well known Bonferroni correction, attempted control on the Family-wise error rate (FWER), i.e., $P(N_{1|0} > 0)$. FWER proves to be too strict as an error measure. Especially when the number of tests is big, the power of each one will be very low: in order not to make false rejections, less and less are made, even if the evidence against the null hypothesis is strong. In particular, as m grows to infinity, no hypothesis will be rejected, no matter whether it is true or false. Note that there are many areas of application in which usually the number of hypotheses is in the order of the thousands, like fMRI (brain imaging) or analysis of DNA micro-arrays.

Benjamini and Hochberg (1995) define the FDR as the expected proportion of incorrect rejections, i.e., $E[\frac{N_{1|0}}{R+1_{\{R=0\}}}]$, where $1_{\{\cdot\}}$ is the indicator function. They argue that control on the FDR is also a loose control on the FWER (it actually is FWER when $R = 1$), and that each test will have a much higher power. Since then, many procedures have been proposed to control the FDR.

The setting is as follows: call p_i the test statistic computed for the i-th hypothesis. The test statistic will usually be a p-value (i.e., the probability of an opportune statistic being bigger than its observed value under the null). Suppose we reject the i-th hypothesis if p_i is smaller than a certain $t \in (0, 1)$, so that $I_i = 1_{\{p_i<t\}}$ will be equal to 1. The problem here is to set a threshold t that will ensure that the FDR is below a pre-specified level ρ. Typically, ρ will be taken to be 0.05. Let $H_i = 1$ if the i^{th} null hypothesis is false, and $a = m_1/m$. Note that the proportion of incorrect rejections as a function of the threshold t defines the process $\Gamma(t) = \frac{\sum(1-H_i)I_i}{\sum I_i+\prod(1-I_i)}$. Define $\widehat{G}(t) = \frac{1}{m}\sum I_i$, the empirical distribution of the p-values. The methodology was put under the empirical process framework by Genovese and Wasserman (2002). The classical Benjamini and Hochberg (BH) method fixes the threshold $t = T_{BH}$ as $\sup\{s : \widehat{G}(s) = \frac{s}{\rho}\}$, where ρ is the desired upper bound for the FDR. The plug-in method fixes $t = T_{PI}$ as $\sup\{s : \widehat{G}(s) = \frac{(1-\widehat{a})s}{\rho}\}$, where $\widehat{a}$ is a suitable estimator for a. Obviously, if a is known, the optimal plug-in threshold is $t = T_O = \sup\{s : \widehat{G}(s) = \frac{(1-a)s}{\rho}\}$. The most common estimator for a is Storey's estimator, proposed in Storey (2002), defined as: $\widehat{a} = \frac{\widehat{G}(t_0)-t_0}{1-t_0}$ for some $t_0 \in (0, 1)$. The BH method yields an $FDR \leq (1-a)\rho$; while the plug-in method is asymptotically less conservative, with an $FDR \leq \rho$, thus being more powerful. Note that these methods are all distribution free, and assume the test statistics are independent. In many applications the assumption of independence is not easily made: for instance, in fMRI, it is well known that the neuronal activity in a spot of the brain must be dependent on the activity in the proximities. Storey and Siegmund (2003) present several theorems that all require almost sure point-wise convergence separately of the empirical distributions of the null

p-values and alternative p-values. Here we use broad conditions on the whole sequence of p-values to prove the validity of the FDR control methods under dependence, and provide asymptotic distributional results.

3. Asymptotic Validity of the Plug-in Method Under Dependence

We will show that under conditions of α-mixing for the p-value sequence, the plug-in procedure (with a good estimator of a) will control the FDR at the desired level ρ. See Doukan (1994) for an extensive discussion on mixing. This is a generalization of some of the results from Genovese and Wasserman (2001).

Theorem 1 *Let $\{p_i\}_{i \in \mathcal{N}}$ be a random sequence of p-values from tests. Assume $\Pr(p_i < t | H_i = 1) \sim F(t)$, where F is arbitrary on $[0, 1]$ apart from Uniform, and $m_1/m \to a$ as $m \to \infty$. Let $\alpha(n)$ be the α-mixing coefficients for the p-value sequence. Assume there exists $\delta > 0$ such that*

$$\alpha(n) \leq C n^{-3-\delta}, \tag{1}$$

for some constant C. Assume the quantity a is known.

Then, $E[\Gamma(T_O)] = \rho + o_P(1)$. [3]

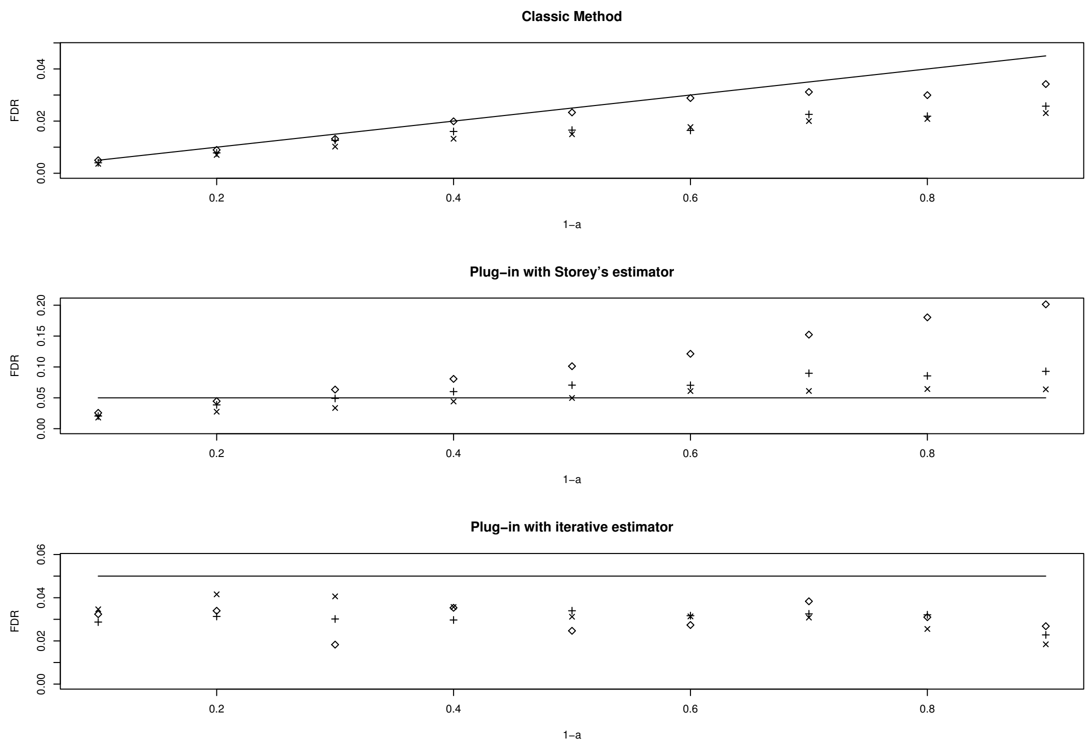
Note that if a is unknown and $\widehat{a}$ is a consistent estimator of a, under the same assumptions of Theorem 1, $E[\Gamma(T_{PI})] = \rho + o_P(1)$. If $\widehat{a}$ is conservative as m grows (i.e., $\widehat{a} \leq a$), $E[\Gamma(T_{PI})] \leq \rho + o_P(1)$. This implies the result for the BH method, for which clearly $\widehat{a} = 0$. Note that α-mixing is a broad hypothesis on the dependence of a sequence of random variables, which will be implied by other well known kinds of dependence; like m-dependency, gaussianity with decaying covariance, etc. Figure 1 shows the FDR simulated for arrays of $m = 10000$, 1600 and 100 normal random variables, with slow decaying correlation, as a function of the proportion of true nulls[4]. We want to see here if, as it is suggested by the Theorem, the procedures work under dependence; and what is the effect of the number of tests m. The first panel shows the outcome of the classic method. We can see that the simulated FDR is always well below level $(1 - a)\rho$. The second panel shows that under dependence the plug-in method, using Storey's estimator to estimate the proportion of false nulls a, doesn't always bring FDR below level ρ. We propose here an iterative, estimator; computed in this way: at the first step, take $\widehat{a} = 0$. In the following steps, apply a plug-in method using the proportion of rejected hypotheses in the previous step as an estimator of a; until the estimator is the same in two subsequent iterations. It is easy to prove that the number of iterations is finite almost surely. The third panel shows the results for the plug-in method using our iterative estimator. The FDR is always well below level ρ. This implies that our iterative estimator is, if not consistent, at least robust with respect to dependence. In summary, in case of weak dependence, both the classical BH method and the plug-in method (with our iterative estimator) asymptotically control the FDR at the desired level.

---

[3] Note that T_O is a random value, so that $\Gamma(T_O)$ is a random process evaluated at a random point.

[4] Same results for all three cases in Figure 1 have been observed simulating Student's T random variables.

Figure 1: *FDR simulations under dependence (line = FDR level, x =* 10000 *r.v., + =* 1600 *r.v.,* $\diamondsuit = 100$ *r.v.)*



References

Benjamini Y. and Hochberg Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B*, 57, 289–300.

Benjamini Y. and Yekutieli D. (2001) The control of the false discovery rate in multiple testing under dependency, *Annals of Statistics*, 29, 1165–1188.

Doukan P. (1994) *Mixing*, Lectures Notes in Statistics, 85; Springer-Verlag.

Genovese C. and Wasserman L. (2001) False discovery rates, *Technical Report, CMU Department of Statitics*.

Genovese C. and Wasserman L. (2002) Operating characteristics and extensions of the fdr procedure, *Journal of the Royal Statistical Society B*, 64, 499–518.

Storey J. (2002) A direct approach to false discovery rates, *Journal of the Royal Statistical Society B*, 64, 479–498.

Storey J.D. Taylor J. and Siegmund D. (2003) Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach, *Journal of the Royal Statistical Society B (to appear)*.