
Paper summary

Jinxi Liu

Contents

1	Lynch, Gavin, et al. "The control of the false discovery rate in fixed sequence multiple testing." arXiv preprint arXiv:1611.03146 (2016).	2
1.1	A conventional fixed sequence method that stops testing once an acceptance occurs	2
1.1.1	Under arbitrary dependence	2
1.1.2	Under negative dependence	2
1.2	Fixed Sequence Procedures that Allow More Acceptances	3
1.2.1	Under arbitrary dependence	3
1.2.2	Under the independence assumption	3
1.3	Data Driven Ordering	4
2	G'Sell, Max Grazier, et al. "Sequential selection procedures and false discovery rate control." Journal of the royal statistical society: series B (statistical methodology) 78.2 (2016): 423-444.	4
A	Positive regression dependence(PRDS)	5
B	Negative Association	5

1 Lynch, Gavin, et al. "The control of the false discovery rate in fixed sequence multiple testing." arXiv preprint arXiv:1611.03146 (2016).

In this paper, the authors focus on the fixed sequence structure where the testing order of the hypotheses has been strictly specified in advance. They develop step-down FDR controlling methods that fully exploit the fixed sequence structural information, in which hypotheses are tested from lowest-ranked to highest-ranked. They first consider a conventional fixed sequence method that stops testing once an acceptance occurs, and develop such a method controlling the FDR under both arbitrary and negative dependencies. To account for any potential mistakes in the ordering of the tests, they extend the conventional fixed sequence method to one that allows more but a given number of acceptances.

1.1 A conventional fixed sequence method that stops testing once an acceptance occurs

The authors present the developments of two simple conventional fixed sequence procedures controlling the FDR under both arbitrary dependence and negative dependence conditions on the p-values.

1.1.1 Under arbitrary dependence

Consider a conventional fixed sequence procedure with critical constants

$$\alpha_i^{(1)} = \min \left(\frac{m\alpha}{m-i+1}, 1 \right), i = 1, \dots, m.$$

- (i) This procedure strongly controls the FDR at level α under arbitrary dependence of the p-values.
- (ii) One cannot increase even one of the critical constants $\alpha_i^{(1)}$, $i = 1, \dots, m$, while keeping the remaining fixed without losing control of the FDR. This is true even when \vec{P} is assumed to be PRDS on \vec{P}_0 .

1.1.2 Under negative dependence

When the p-values are negatively associated, the critical constants of the conventional fixed sequence procedure can be further improved as in the following: The conventional fixed sequence method with critical constants

$$\alpha_i^{(2)} = \frac{i\alpha}{1 + (i-1)\alpha}, i = 1, \dots, m.$$

strongly controls the FDR at level α when the p-values are negatively associated on the true null p-values.

1.2 Fixed Sequence Procedures that Allow More Acceptances

There is a potential for loss of power in a conventional fixed sequence multiple testing method if the ordering of the hypotheses, particularly for the earlier ones, does not match with that of their true effect sizes, potentially leading to some earlier hypothesis being accepted and the follow-up hypotheses having no chance to be tested. To mitigate that, we consider generalizing the conventional fixed sequence multiple testing to one that allows more than one but a pre-specified number of acceptances, and develop such generalized fixed sequence multiple testing methods controlling the FDR under both arbitrary dependence and independence.

Fixed sequence method stopping on the k_{th} acceptance

- 1 If $P_1 < \alpha_1$, then reject H_1 ; otherwise, accept H_1 . If $k > 1$ or H_1 is rejected, then continue to test H_2 ; otherwise stop.
- 2 If $P_i < \alpha_i$, then reject H_i ; otherwise, accept H_i . If the number of accepted hypotheses so far is less than k , then continue to test H_{i+1} ; otherwise stop.

1.2.1 Under arbitrary dependence

The fixed sequence method stopping on the k_{th} acceptance with critical constants

$$\alpha_i^{(3)} = \begin{cases} \frac{\alpha}{k}, & \text{if } i = 1, \dots, k \\ \frac{(m-k+1)\alpha}{(m-1+1)\alpha}, & \text{if } i = k+1, \dots, m \end{cases}$$

strongly controls the FDR at level α arbitrary dependence of p-values.

1.2.2 Under the independence assumption

Under certain distributional assumptions on the p-values, the critical constants of the above procedure can potentially be improved. In particular, we have the following result.

Consider a fixed sequence method stopping on the k_{th} acceptance with critical constants

$$\alpha_i^{(4)} = \frac{(r_{i-1} + 1)\alpha}{k + (i - k)\alpha}, i = 1, \dots, m,$$

where r_i is the number of rejections among the first i tested hypotheses, with $r_0 = 0$, for $i = 1, \dots, m$. This procedure strongly controls the FDR at level α if the true null p-values are mutually independent and are independent of the false null p-values.

1.3 Data Driven Ordering

The applicability of the aforementioned fixed sequence methods depends on availability of natural ordering structure among the hypotheses. When the hypotheses cannot be pre-ordered, one can use pilot data available to establish a good ordering among the hypotheses in some cases. For example, in replicated studies, the hypotheses for the follow-up study can be ordered using the data from the primary study. However, when prior information is unavailable ordering information can usually be assessed from the data itself.

Assume that the variables of interest are $X_i, i = 1, \dots, m$, with n independent observations X_{i1}, \dots, X_{in} on each X_i . An ordering statistic, Y_i , and a test statistic, T_i , are determined for each $i = 1, \dots, m$. The Y_i s are used to order all of the hypotheses H_1, \dots, H_m , T_i is used to test the hypothesis H_i ; $i = 1, \dots, m$, and P_i is the corresponding p-value. In addition, Y_i is chosen such that it is independent of the T_i s under H_i and tends to be larger as the effect size increases. The approach is outlined below.

Example 2: Two sample T-test. Consider testing $H_i : \mu_i^{(1)} = \mu_i^{(2)}$ against $H'_i : \mu_i^{(1)} \neq \mu_i^{(2)}$ simultaneously using $n = n_1 + n_2$ data vectors. Suppose $X_{ij}^{(l)}, j = 1, \dots, n_l$, follows a $N(\mu_i^{(l)}, \sigma^2)$ distribution, for $l = 1, 2$. Then, the hypotheses can be tested using the two-sample t-test statistics T_i and are ordered through the values of the total sum of squares, which is $Y_i = \sum_{l=1}^2 \sum_{j=1}^{n_l} X_{ij}^{(l)2} / n$, for $i = 1, \dots, m$. The rationale behind this is independence between the Y_i s and T_i under H_i , and the following result: $E[Y_i] = (n-1)\sigma^2 + n_1 n_2 (\mu_i^{(1)} - \mu_i^{(2)})^2 / n$.

The fixed sequence multiple testing procedures introduced above can still strongly control the FDR at level α .

2 G'Sell, Max Grazier, et al. "Sequential selection procedures and false discovery rate control." Journal of the royal statistical society: series B (statistical methodology) 78.2 (2016): 423-444.

In this paper, the authors consider a multiple hypothesis testing setting where the hypotheses are ordered and one is only permitted to reject an initial contiguous block, H_1, \dots, H_k , of hypotheses. A rejection rule in this setting amounts to a procedure for choosing the stopping point k .

Their proposed methods start by transforming the sequence of p-values p_1, \dots, p_m into a monotone increasing sequence of statistics $0 \leq q_1 \leq \dots, q_m \leq 1$. They then prove that they achieve ordered FDR control by applying the original Benjamini-Hochberg procedure on the monotone test statistics q_i .

A Positive regression dependence(PRDS)

The vector of p-values \vec{P} is PRDS on the vector of null p-values $\vec{P}_0 = (\hat{P}_1, \dots, \hat{P}_{m_0})$ if for every increasing set D and for each $i = 1, \dots, m_0$, the conditional probability $\Pr(\vec{P} \in D | \hat{P}_i = p)$ non-decreasing in p.

B Negative Association

The vector of p-values \vec{P} is negatively associated with null p-values if for each $i = 1, \dots, m_0$, the following inequality holds:

$$\Pr(\hat{P}_i \leq p_{\mu_i}, P_j \leq p_j, j = 1, \dots, m, \text{ with } j \neq \mu_i) \leq \Pr(\hat{P}_i \leq p_{\mu_i}) \Pr(P_j \leq p_j, j = 1, \dots, m, \text{ with } j \neq \mu_i)$$

for all fixed p_j 's.

Several multivariate distributions possess the conventional negative association property, including multivariate normal with non-positive correlation, multinomial, dirichlet, and multivariate hypergeometric. Independence is a special case of negative dependence.