Multiple Hypothesis Testing

FWER, FDR control, and negative dependence

Purushottam Saha (BS2119)

B.Stat. 3rd Year
Indian Statistical Institute, Kolkata

Table of Contents

Table of Contents

Multiple Hypothesis Testing Problem

Due to certain statistical testing requirements, we may require testing multiple statistical hypotheses at the same time. Additionally, the problem might require either accepting them all together or rejecting even if one hypothesis is false. If the hypothesis is accepted, we know that each of the individual hypothesis are true. When we reject them we might be interested in knowing which of the individual hypothesis were true. Thus, we want to take a decision by considering all the hypothesis and allow individual hypothesis analysis too. The problem lies in developing an effective method which allows us to do so.

Multiple Hypothesis Testing Framework

- Suppose we want to test k Hypotheses, say $H_1, H_2, \ldots, H_k$, simultaneously.
- Denote $\{1, 2, \ldots, k\} = \mathcal{K}$, $\mathcal{N} \subset \mathcal{K}$ is the index set of true null hypotheses. $N_0 = |\mathcal{N}|$ is the number of true null hypotheses.
- $p_1, p_2, \ldots, p_k$ are the k p-values of the k individual test statistics testing the aforementioned hypothesis.
- $\mathcal{D} : [0, 1]^k \mapsto \{2\}^{\mathcal{K}}$ is the testing procedure, taking k p-values as the inputs, and returning the indices of the rejected hypothesis, often referred to as **Discoveries**.

Different errors to minimise

There are several methods for controlling errors in this setting, some of the popular choices are:

- Uncorrected Testing
- Family Wise Error Rate Control (FWER)
- False Discovery Rate Control (FDR)
- False Discovery Exceedance Rate Control (FDX)

Here we mainly discuss FWER control and FDR control procedures.

Uncorrected Testing

Here the we control probability of false rejection for each hypothesis, by testing individual hypothesis at level α, when the multiple hypothesis problem is to be tested at level α.
The error here is under the independence of hypothesis set-up the Type-I error adds up and the hypothesis is no longer tested at level α. A proof of the statement can be obtained by simulating.

Table of Contents

FWER control

In single hypothesis testing problem we try to control the probability of a single false rejection. Under multiple hypothesis set-up we try to control the probability of one or more false rejections. The probability of one or more false rejection is Family Wise Error Rate(FWER). Thus testing a multiple hypothesis problem at level α is making sure

$$\text{P(One or more false rejections)} \leq \alpha$$

Bonferroni Method

- It is a single-step procedure with a common cutoff value of $\frac{\alpha}{n}$, to bound the FWER to α.
- Every p-value is tested against the common threshold $\frac{\alpha}{n}$, hypotheses with p-value less than the threshold are rejected, where as others are failed to reject, or in simple terms, accepted.
- The FWER control for this procedure is guaranteed by Bonferroni Inequality, for proof see this.
  Problem: When n is very large, $\frac{\alpha}{n}$ is small, the level of significance against which each hypothesis is tested is very small, leading to a high false acceptance rate or increased probability of Type-II error.
  Thus our multiple testing problem reduces to finding $\alpha_1, \alpha_2, \ldots, \alpha_k$, the level of significance of each hypothesis such that $\frac{\alpha}{n} < \alpha_i < \alpha$, for $i = 1(1)n$ and the probability of Type-I and Type-II error are balanced.

Simes Procedure

- Simes (1986) proposed an improvement on the Bonferroni Procedure and showed for some dependence structures a significant improvement can be made to control FWER, or in some sense, for the global null test.
- Merging the p-values with the Simes function $S(\underset{\sim}{p})$, we get :

$$S(\underset{\sim}{p}) = \min_{k \in \mathcal{K}} \frac{K}{k} p_{k:K}$$

and we test this merged form of p-values at the significance level of α, i.e. the test becomes

$$S(\underset{\sim}{p}) \leq \alpha$$

- Simes proved that $\mathbb{P}(S(\underset{\sim}{p}) \leq \alpha) \leq \alpha$ holds for a wide range of distributions, and hence considered a better version of Bonferroni for where it applies.
- The conditions generally hold under the independence model and also under the PRDS dependence structure which is discussed later.

Critical values and Rejection Curve

- For single-step procedures and without distributional assumption, not much more can be done. But, instead of single-step procedures, we may try to test different p-values with different critical values, or more specifically we may try to test the order statistic of p-values with a sequence of increasing critical values, i.e.

$$0 < \alpha_{1:n} \leq \alpha_{2:n} \leq \ldots \leq \alpha_{n:n} < 1$$

- Fitting these critical values to a continuous function $f : [0, 1] \mapsto [0, 1]$, such that $f(\frac{i}{n}) = \alpha_{i:n}$, gives us the rejection function.
- Sometimes, while working with general critical values, and hence the rejection curve, assumptions like the convexity of the rejection curve may give something to work for.

Step Up and Step Down Procedures

- Step-wise procedures are a probable improvement to Single-step procedures, allowing us to test different p-values at different significance levels, i.e. we test the ordered p-values according to a deterministic set/sequence of critical values.
- Step-wise procedures can be broadly classified into two groups, i.e.:
  1. **Step Down Tests** are tested from the most significant p-value, i.e. if $p_{1:n} < \alpha_{1:n}$ we reject hypothesis with p-value $p_{1:n}$, then we test for $p_{2:n} < \alpha_{2:n}$. Once we accept one hypothesis, we accept the rest.
  2. **Step Up Tests** are similar but opposite, instead of starting from the most significant p-value, we start from the least significant p-value, and accept it for $p_{n:n} \geq \alpha_{n:n}$, once we reject, we reject the rest.

Holm's Procedure

- Holm's Procedure is a Step-Down test procedure, using critical values

$$\alpha_{i:n} = \frac{\alpha}{n+1-i} \text{ for all } i = 1(1)n$$

.

- Holm's Procedure controls the FWER at level α.
- It is always at least as good as Bonferroni, as

$$\alpha_{i:nHolm} \geq \frac{\alpha}{n} = \alpha_{Bonferroni}$$

- It can be shown that this is the best we can do without any distributional assumptions (See this). But by incorporating distributional assumption, we can indeed perform better tests, the idea behind being exhausting the significance level of α for the specific distribution class.

Conservative nature of FWER

- Though controlling the False Discoveries with great precision, these tests are conservative and hence lack power, i.e. type-II error probability is much higher.
- For most of today's applications of multiple hypothesis testing, power is critical as the most interesting effects are usually at the edge of detection.
- The problem is to choose a procedure that balances the competing demands of sensitivity and specificity.

Table of Contents

FDR (False Discovery Rate)

- To overcome the conservative nature of FWER control, we introduce the idea of False Discovery Proportion, and False Discovery Rate control, in short, FDR control.
- The main idea behind this is, that 5 false discoveries out of 10 discoveries is a more serious error than 10 false discoveries out of 50 discoveries, a discovery being rejecting the null.
- In many recent applications of Multiple Hypothesis testing, the interesting behaviours are on the edge of acceptance. Hence, the conservative nature of FWER control does not capture those areas, which FDR control is capable of capturing.

BH Procedure for FDR control

- Benjamini and Hochberg (1995) introduced FDR control and showed Simes procedure, here on referred to as BH procedure, controls the same.
- The test is as follows:
  order the p-values as $p_{1:k} \leq p_{2:k} \leq \ldots \leq p_{k:k}$, and reject $H_{(i)}$ if $p_{i:k} < \alpha_{i:k} = \frac{i}{k}\alpha$.
- They showed under the assumption of independence of p-values (or reverse martingale model, or PRDS model also), the procedure controls the FDR at α level of significance. For the proof, check this.

Table of Contents

Generalizations: far from Independence

- Often in applications, the independence assumption really doesn't hold for the p-values, and hence the requirement of considering dependence structures is felt.
- For example, while considering the tests for genes/diseases based on them, on a particular chromosome, the chromosome being the same, the genes may not be independent and hence one's effect on others or at least on the whole of the chromosome is felt, here independence assumptions on p-values, i.e. the test statistics is not a good assumption to make, and the dependence structure has to be assumed/figured out.

Dependence Structures

- Here we discuss the most important and well-studied dependence structures, and next discuss the FDR control under these dependencies.
  1. Basic Independence Model: If any p-value is from null or not, and the values associated with them if the hypothesis is null, are independent, that sub-model for p-values is referred to as Basic Independence Model. If n p-values for a multiple testing scenario are defined as (U_i, e_i, E_i), i.e. $e_i = 0/1$ based on the p-value is from null or not, and U_i is the realisation of the p-variable if the null is true and E_i is the realisation of the p-variable if the null is false, then BI Model says $(e_i, E_i)_{i=1}^n$ and $(U_i)_{i=1}^n$ are independent.

Dependence Structures (Contd.)

2. PRDS (Positive Regression Dependence Structure): This notion of positive correlation is defined in 2 ways:
   - Strong PRDS: For the strong PRDS model, we require for any Increasing Set C, $\mathbb{P}\{\mathbf{p} \in C \mid p_j \leq x\}$ is non decreasing in x, for all $i \in \mathcal{N}$, i.e. for p-values.
   - Weak PRDS: For the strong PRDS model, we require for any Decreasing set C, $\mathbb{P}\{\mathbf{p} \in C \mid p_j \leq x\}$ is non increasing in x, for all $i \in \mathcal{K}$, i.e. for all true null p-values.
3. Reverse Martingale Model: This is another sub-model where for all true null, $\frac{\mathbb{1}\{p_i \leq t\}}{t}$, $0 < t \leq 1$ follows a reverse martingale model. This dependency structure is generally seen in various financial models.

BH under Reverse Martingale Model and PRDS

- Benjamini and Yukiteli proved that the BH procedure controls FDR even under **Reverse Martingale model** and **PRDS model**, and the independence structure is a special case of these results.
- In fact, for the Reverse Martingale model, the FDR level of α is attained, whereas, for the PRDS model, the FDR is bounded above by α, the bound may or may not be achieved.

Table of Contents

Inequalities on FDR under dependence structures

We start with arbitrary non decreasing critical values

$$0 < \alpha_{1:n} \leq \alpha_{2:n} \leq \ldots \alpha_{n:n} < 1.$$

- One of the major results [2] show that, Under Reverse Martingale Model, and for SU tests,

$$\frac{\mathbb{E}(N)}{n}(\min_{i \leq n} \frac{n\alpha_{i:n}}{i}) \leq FDR \leq \frac{\mathbb{E}(N)}{n}(\max_{i \leq n} \frac{n\alpha_{i:n}}{i})$$

  where $E(N)$ is the expected number of true nulls.
- Under PRDS model for SU tests, we again get

$$FDR \leq \frac{\mathbb{E}(N)}{n}(\max_{i \leq n} \frac{n\alpha_{i:n}}{i})$$

Inequalities on FDR under dependence structures (Contd.)

- Another important result states, again for SU tests, and critical values $\alpha_{j:n}$, $j = 1(1)n$, satisfying $j \mapsto \frac{\alpha_{j:n}}{j}$ is non-decreasing (concave rejection function), such that FDR is controlled at α for all Basic Independence Model of p-values, we have
  1. A necessary condition for the above to satisfy is $\alpha_{j:n} \leq \frac{j\alpha}{n+1-j}$ for all j.
  2. If for any $k < n$, $\alpha_{k:n} < \frac{k\alpha}{n+1-k}$ holds, then the strict inequality holds for all j.
- An important corollary of this result says that $\alpha_{1:n} \leq \frac{\alpha}{n}$ always holds, with $\alpha_{1:n} = \frac{\alpha}{n}$ case referring to the BH procedure, and for others, $\alpha_{1:n} < \frac{\alpha}{n}$.
- It also holds that, if $\alpha_{1:n} = \frac{\beta}{n}$, $FDR \geq \frac{\beta\,\mathbb{E}(N)}{n}$.

Other Optimal Threshold Values

- k-adjusted
- A threshold sequence for SU tests defined as: $\alpha_{(i)} = \frac{i\alpha}{n+b-ia}$ for non-negative a and b, are frequently discussed in the literature. For this to be a valid set of threshold values, i.e. $\alpha_{n:n} < 1$, we have $a \leq 1 - \alpha$, and for FDR controlled at α, we will see from the next a few results, that, $a \leq b$ is required. Moreover, for non-zero a, $b > a$ is required, which follows from the same result. For optimal values of a, b we wait for some time.

Table of Contents

Problem with Negative Dependence

- Suppose X_1 and Y are iid N(0,1) variables.
  Now consider $X_2 = \frac{-X_1+Y}{\sqrt{2}}$.
- X_2 also follows N(0,1), but is neither independent of X_1, nor has positive regression dependence, rather clearly has negative regression dependence.
- Consider the multiple testing problem H_1, H_2, such that, $H_i : X_i$ has mean 0. The corresponding p-values are $p_i = 1 - \Phi(X_i)$ for i = 1,2.
- It is shown that for significance level $\alpha = 0.5$, FWER = FDR(2) = $\frac{9}{16} > 0.5 = \alpha$, where FDR(2) is the conditioned FDR that $N_0 = 2$.
- Similar real-life examples come from testing of Multinomial models, constant/random-sum games testing, various time series models etc., where independence or PRDS model assumptions don't really hold.

BH under arbitrary dependence

- Benjamini and Yuketeli proposed a correction to the original BH procedure to control the FDR for arbitrary dependence.
- They showed, the BH procedure controls the FDR level at αl_k, where $l_k = \sum_{i=1}^{k} \frac{1}{i} \approx \log(k)$, for arbitrary dependence.
- Though being a very interesting result, The power for this test is even worse than the Bonferroni Procedure and hence is not used in a practical sense.

Negative Dependence Structures

As arbitrary dependence results are very weak, we try to focus on negative dependencies. For the same, we define some Negative Dependence Structures to work with, from weakest to strongest.
Let $\mathbf{X} \equiv (X_1, X_2, ..., X_K)$ be a random vector, with F_k being the CDF for X_k, $k \in \mathcal{K}$.

- $\mathbf{X}$ is called **Weakly Negative Dependent** if for all $A \subseteq \mathcal{K}$ and $p \in (0, 1)$,

$$\mathbb{P}\left(\bigcap_{k \in A} \{X_k \leq F_k^{-1}(p)\}\right) \leq \prod_{k \in A} \mathbb{P}(X_k \leq F_k^{-1}(p)) \text{ (lower) or,}$$

$$\mathbb{P}\left(\bigcap_{k \in A} \{X_k > F_k^{-1}(p)\}\right) \leq \prod_{k \in A} \mathbb{P}(X_k > F_k^{-1}(p)) \text{ (upper).}$$

Negative Dependence Structures (Contd.)

- **X** is called **Negatively Orthant Dependent** if for all $A \subseteq \mathcal{K}$ and $p \in (0, 1)$,

$$\mathbb{P}\left(\bigcap_{k \in A} \{X_k \leq x_k\}\right) \leq \prod_{k \in A} \mathbb{P}(X_k \leq x_k) \text{ (lower) or,}$$

$$\mathbb{P}\left(\bigcap_{k \in A} \{X_k > x_k\}\right) \leq \prod_{k \in A} \mathbb{P}(X_k > x_k) \text{ (upper).}$$

- **X** is called **Negatively Associated** if for any disjoint subsets $A, B \subseteq \mathcal{K}$, and any real-valued, coordinate-wise increasing functions f, g, we have,

$$Cov(f(\mathbf{X}_A), g(\mathbf{X}_B)) \leq 0$$

where $\mathbf{X}_A = (X_k)_{k \in A}$ and $\mathbf{X}_B = (X_k)_{k \in B}$ assuming that $f(\mathbf{X}_A)$ and $g(\mathbf{X}_B)$ have finite second moments.

Negative Dependence Structures (Contd.)

- A random vector $\mathbf{X}$ is said to be stochastically decreasing in $\mathbf{Y}$ if $\mathbb{E}[g(\mathbf{X})|\mathbf{Y} = y]$ is decreasing in y whenever g is a coordinate-wise increasing function such that the conditional expectation exists. Then, $\mathbf{X}$ is **Negative Regression Dependent** if

$$\mathbf{X}^{-i} \textit{ is stochastically decreasing in } X_i \textit{ for every } i,$$

  where $\mathbf{X}^{-i}$ is the vector formed by deleting the i-th coordinate of $\mathbf{X}$.
- A random vector $\mathbf{X}$ is Gaussian dependent if there exist increasing functions (or decreasing functions) $f_1, \ldots, f_K$ and a Gaussian vector $(Y_1, \ldots, Y_K)$ such that$X_k = f_k(Y_k)$ *for* $k \in K$.

Negative Dependence Structures (Contd.)

- Finally, a random vector $\mathbf{X}$ is counter-monotonic if each pair of its components is counter-monotonic, that is,

$$(X_i(\omega) - X_i(\omega^{'}))(X_j(\omega) - X_j(\omega^{'})) \leq 0 \text{ for almost every } (\omega, \omega^{'}) \in \Omega^2$$

$$\text{and every distinct } i, j.$$

  In simpler form, for real-valued random variables, (X, Y) is said to be counter-monotonic if there exist increasing functions f, g and a random variable Z such that $(X, Y) = (f(Z), -g(Z))$ almost surely. This is the strongest notion of negative dependence.

In the next page, a chart is given which shows the dependence structures and which dependence structures imply other weaker dependence structures. For example, Counter-monotonicity implies Negative Associativity and Negative Regression Dependence, and hence subsequently Negative Orthant Dependence. For further details refer [1].

Negative Dependence Structures (Contd.)

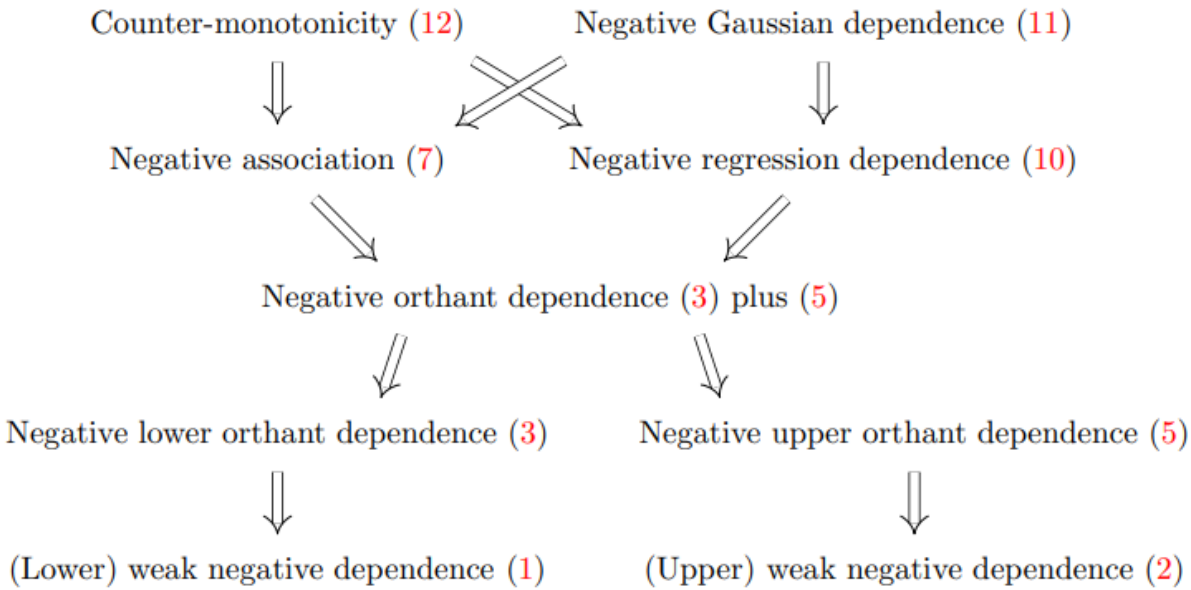

Figure: Negative Dependence Structures and Relations

Merging p-values under neg. dependence

We again consider the Simes function for merging p-values, but under negative dependence, it does not satisfy the good-old bound of $\mathbb{P}(S_K(P) \leq \alpha) \leq \alpha$ is not valid anymore. Ramdas and Wang [1] propose a non-trivial upper bound on $\mathbb{P}(S_K(P) \leq \alpha)$, which is:

for Weakly Negatively Dependent p-values P,

$$\mathbb{P}(S_K(P) \leq \alpha) \leq \alpha + \sum_{k=2}^{K} \binom{K}{k} \left(\frac{ak}{K}\right)^k$$

A simplified, and non-K-dependent estimate for the additive error bound will be:

$$\mathbb{P}(S_K(P) \leq \alpha) \leq \alpha + 2\alpha^2 + \frac{9}{2}\alpha^3 + \frac{1}{\sqrt{8\pi}} \frac{(e\alpha)^4}{1 - e\alpha} \text{ for all } \alpha \in (0, \frac{1}{e}) \text{ or,}$$

$$\mathbb{P}(S_K(P) \leq \alpha) \leq \alpha + 2\alpha^2 + 6\alpha^3 \text{ for all } \alpha \in (0, 0.1]$$

Merging p-values under neg. dependence (Contd.)

Similarly trying for a multiplicative error to the same bound gives us,

$$\mathbb{P}(S_K(P) \leq \alpha) \leq 1.26\alpha \text{ for all } \alpha \in (0, 0.1] \text{ or,}$$

$$\mathbb{P}(S_K(P) \leq \alpha) \leq 3.4\alpha \text{ for all } \alpha \in (0, 1)$$

where this 3.4 can be replaced by $l_K \wedge 3.4$, as for small K's, it's just better to use the bound for arbitrary dependence by Benjamini Yuketeli sometimes.
A weighted version of this merging p-values is also discussed, as some p-values may contribute more to the global null. In that case, replacing the p-values p_i, by the corresponding weighted p-value $c_i = \frac{w_i p_i}{\sum_{i=1}^{K} w_i}$ and the previous results discussed again hold true.

E-values

- An e-variable E is an "extended"[1] random variable on probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $E : \Omega \to [0, 1]$ with $\mathbb{E}_{\mathbb{P}}[E] \leq 1$. The values taken by e-variables are referred to as e-values.
- Now, a p-variable is a random variable $P : \Omega \to [0, 1]$ satisfying

$$\forall \epsilon \in (0, 1) : \mathbb{P}(P \leq \epsilon) \leq \epsilon.$$

  The values taken by p-variables are what we know as p-values.
- A decreasing function $f : [0, 1] \to [0, \infty]$ is called a calibrator (more specifically p to e calibrator, but we abuse the notation), if for any probability space, and a p-variable P in that space, $f(P)$ is an e-variable.
- It is shown [5] that a decreasing function $f : [0, 1] \to [0, \infty]$ is a calibrator *if and only if* $\int_0^1 f \leq 1$.

[1] Extended random variables are rvs that are allowed to take ∞ as a value.

E-values (Contd.)

- For two calibrators f, g, g is said to be dominated by f if $f \geq g$. The domination is strict if $f(x) > g(x)$ for some $x \in [0, 1]$. If a calibrator is not strictly dominated by any other calibrator, it is said to be *admissible*.
- It is also shown [5] that a calibrator f is admissible if and only if it is left-continuous, $f(0) = \infty$ and $\int_0^1 f = 1$. For example, $f(x) = \frac{1}{\sqrt{x}} - 1$ is an admissible calibrator.
- An e to p calibrator is a function, that turns a e-variable into a p-variable. It is shown [5] that the only admissible e to p calibrator is $f(x) = \min\{\frac{1}{x}, 1\}$.
- This connection between e-values and p-values, as well as the easy structure of e-values makes it a good candidate to perform the hypothesis testing on, rather than p-values.

Merging e-values: Testing for Global Null

- Vovk and Wang [5] showed the process of merging e-values, to test for the global null.
- Multiplying all the e-values to merge them and testing for global null is a popular choice, which is described here as well.
- 2 improving algorithms are also proposed to boost the performance of the e-values, which are recommended.
- Another approach of merging E-values involving the idea of the Simes function (use calibrations and then the form of the Simes function for p-values) is referred to in [1], but it is dominated by the simple average of the e-values (which is also an e-value), and hence is not much discussed.

e-BH Procedure

- Wang and Ramdas (2020) [6] proposed a method generalizing BH procedure for e-values, making the original BH procedure as a special case of this e-BH procedure using p-values, by transforming e-values using the e to p calibrator.
- One of the crucial results [6] is that this testing procedure controls FDR under arbitrary dependence, without correction.
- One may prefer using E-values for merging e-values for the ease of forming e-values and the result on arbitrary dependence, but transforming p-values to e-values and then using the e-BH procedure will not give better results, similarly transforming e-values to p-values and using regular BH procedure will not give better results. So using p-values for B, procedure when we get p-values directly, and using e-values in other cases should be the best approach. For further details regarding these, see [1].

FDR control under negative dependence

Based on a result of Su [4], Ramdas [1] showed that, if the null p-values are weakly negatively dependent, then the BH procedure at level α has FDR at most $\alpha((-log\alpha + 3.18) \wedge l_N)$ where $l_N = \Sigma_{k=1}^{N} \frac{1}{k} \approx logN$ under the stronger condition that the null p-values are iid uniform on [0, 1], there is an upper bound $FDR \leq \alpha(-log\alpha + 1)$. Better bounds on FDR under stronger negative dependence setups are some of the open problems.

Group level FDR control

- Lastly, we discuss the Group level FDR control. Sometimes, instead of individual testing of hypotheses, we may want to test several hypotheses, or a group of hypotheses together, i.e. if any one of the hypotheses in a group is rejected, the group hypothesis is rejected. An example can be stated as the data on genes are given, and we want to reject based on the grouping of the same chromosome, i.e. if a gene is defective, the chromosome associated with it is considered defective.
- Several works on these when the hypotheses are at least PRDS for each null group, are done, and standard BH procedure controls the FDR at the desired level, but for negative dependence, the bounds have to be relaxed.

Group level FDR control (Contd.)

- [1] proposes a method to tackle this problem, by first grouping the p-values by Simes' function and then applying the BH procedure. This two-stage procedure (referred to as Simes+BH procedure) of testing bounds the FDR to $3.4\alpha(-\log(3.4\alpha) + 3.18)$. This is easily seen by using 3.4α instead of α for the BH procedure, which is the bound on Simes' function merging of p-values under negative dependence.
- Though the bound is not very good, it is one of the first results, and given more information on the dependence structure, several improvements are possible.

Table of Contents

Conclusion

Throughout the whole discussion of the literature on Multiple Testing, FDR and especially under negative dependence situations, many improvements and open problems have been discussed. Though many results mentioned here are not very practical to use, these lay out the foundation for the work upcoming work to be done, and expose us to new ideas (e.g. e-values), which show us different ways to think about the same age-old problem of testing hypotheses.

References

[1] Ziyu Chi, Aaditya Ramdas, and Ruodu Wang. Multiple testing under negative dependence, 2023.

[2] Philipp Heesen and Arnold Janssen. Inequalities for the false discovery rate (fdr) under dependence, 2015.

[3] Erich Leo Lehmann, Joseph P Romano, and George Casella. *Testing statistical hypotheses*, volume 3. Springer, 1986.

[4] Weijie J. Su. The fdr-linking theorem, 2018.

[5] Vladimir Vovk and Ruodu Wang. E-values: Calibration, combination and applications. *The Annals of Statistics*, 49(3), June 2021.

[6] Ruodu Wang and Aaditya Ramdas. False discovery rate control with e-values, 2021.

Some Required Definitions

Let $\mathcal{B}([0,1]^n)$ denote the Borel sets of $[0,1]^n$.

- Decreasing Sets : A set $C \in \mathcal{B}([0,1]^n)$ is said to be Decreasing iff, $c_0 \in C$ and $c \leq c_0$ component-by-component imply $c \in C$.
- Increasing Sets : A set $C \in \mathcal{B}([0,1]^n)$ is said to be Increasing iff, $c_0 \in C$ and $c \geq c_0$ component-by-component imply $c \in C$..

Appendix : Bonferroni Procedure

Suppose hypotheses H_i with $i \in I$ are true and the remainder false, with $|I|$ denoting the cardinality of I. From the Bonferroni inequality it follows that

$$FWER = \mathbb{P}\{\text{reject any } H_i \text{ with } i \in I\} \leq \sum_{i \in I} \mathbb{P}\{\text{reject } H_i$$

$$= \sum_{i \in I} \mathbb{P}\{\hat{p}_i \leq \frac{\alpha}{s}\} \leq \sum_{i \in I} \frac{\alpha}{s} \leq \frac{|I|\alpha}{s} \leq \alpha. \quad \square$$

Appendix: Optimality of Holm's Procedure

Theorem 9.2.4 [3] : For a General Step-Down Procedure, the critical values $\alpha_{i:n}$ has the range

$$\frac{\alpha}{s-i+1} \leq \alpha_{i:n} \leq \alpha.$$

Furthermore, the upper bound α is attained when $T_{1:n} = T_{2:n} = ... = T_{n:n}$, i.e., when there really is no multiplicity. The lower bound $\frac{\alpha}{s-i+1}$ is attained when the distribution of $T_{1:n}, T_{2:n}, ..., T_{s-i+1:n}$ is given as follows :
Let U be uniformly distributed on $(0, 1)$, and suppose when all $H_1, H_2, ..., H_n$ are true, then

$$Y_1 = U, Y_2 = U + \frac{1}{s}(mod 1), ..., Y_n = U + \frac{s-1}{s}(mod 1).$$

$T_i = 1 - X_i, where (X_1, X_2, ..., X_n) = (Y_{\pi(1)}, Y_{\pi(2)}, ..., Y_{\pi(n)})$, to support the exchangeability condition.

Appendix : Algorithm for Merging E-values

- **Algorithm 1 :** take cumulative average of most significant e-values in order.
- **Algorithm 2 :** Take the product of all the e-values < 1, and multiply this multiplier with all the e-values.

For further reference, check [5].