Martingale methods for the FDR control of multiple tests

Arnold Janssen
(joint work with Philipp Heesen and Julia Benditkis)

Mathematical Institute
Heinrich-Heine-Universität Düsseldorf, Germany

Köln, Juni 2015

Outline

Outline

1. Introduction: Class-room notes
   - Benjamini and Hochberg Theorem
   - Dependent Normal Distributions
   - Bonferroni adjustment under dependency

2. Martingale Dependence
   - Examples of Martingale Dependence
   - Inequalities for the FDR
   - Sharp Results under Martingale Dependence

Notations

$H_1, \dots, H_n$	null-hypotheses
$p = (p_1, \dots, p_n)$	vector of corresponding p-values
$p_{1:n} \leqslant \dots \leqslant p_{n:n}$	corresponding order statistics
$I_0 \subset \{1, \dots, n\},\ n_0 := \lvert I_0 \rvert$	index-set of true null-hypotheses
$0 = a_0 < a_1 \leqslant a_2 \leqslant \dots \leqslant a_n < 1$	critical values

Step Up (SU) rejection bound

$$a = \sup\{a_i, i = 1, \dots, n : p_{i:n} \leqslant a_{i:n}\}$$

Reject H_i if $p_i \leqslant a$.

$$R = \#\{\ \text{rejected}\ \ H_i\} \qquad V = \#\{\ \text{rejected}\ \ H_i, H_i\ \ \text{true}\}.$$

Definition (Basic Independence Model (BI))

We say that p-values $p_1, \ldots, p_n$ fulfill the Basic Independence Model, if $(p_i)_{i \in I_0}$ are i.i.d. uniformly $U(0,1)$ distributed and independent from $(p_j)_{i \in I \setminus I_0}$.

$$\overbrace{\Big(\underbrace{\big((p_i)_{i\in I_0}\big)}_{\text{i.i.d. } U(0,1)}\ \underbrace{\big((p_i)_{i\in I\setminus I_0}\big)}_{\text{arbitrary}}\Big)}^{\text{independent}}$$

Linear SU

Aim: finite sample FDR control.

Let $\alpha \in (0, 1)$ be given. A SU Procedure with critical values

$$a_i = \frac{i\alpha}{n}, \; i = 1, ..., n,$$

is called Benjamini and Hochberg (BH) procedure.

Theorem (cf. Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001), Finner and Roters (2001))

Under BI we have $FDR = E\left[\frac{V}{\max(R,1)}\right] = \frac{\alpha n_0}{n}$.

A simple Proof of the BH Theorem

A simple proof was proposed by Heesen and J. [2015] and based on Fubini's Theorem. W.l.o.g. let $1 \in I_0$. Define the random variable R^0 by $R^0 = R(0, p_2, ..., p_n)$. Since $V = \sum_{i \in I_0} \mathbf{1}(p_i \leqslant a)$ we get:

$$\mathrm{FDR} = \sum_{i \in I_0} E\left[\frac{\mathbf{1}(p_i \leqslant \frac{\alpha R}{n})}{\max(R,1)}\right]$$

$$= n_0 E\left[\frac{\mathbf{1}(p_1 \leqslant \frac{\alpha R^0}{n})}{R^0}\right] = n_0 E\left[\frac{\frac{\alpha R^0}{n}}{R^0}\right] = \frac{\alpha n_0}{n}.$$

∎

Conclusion: Exact result under BI.

Example 1. BH procedure and positive (negative) normally distributed test statistics

Let X_1 and Y be i.i.d. standard normal random variables. Consider bivariate normals:

$$(X_1, X_2) = (X_1, \frac{1}{\sqrt{2}}X_1 + \frac{1}{\sqrt{2}}Y) \quad \text{positive dependence} \qquad (1)$$

$$(X_1, X_2) = (X_1, \frac{1}{\sqrt{2}}X_1 - \frac{1}{\sqrt{2}}Y) \quad \text{negative dependence} \qquad (2)$$

with related p-values $(p_1, p_2) = (\Phi(X_1), \Phi(X_2))$. Then for the FDR of the BH procedure at level $\alpha = 0.5$ and $n_0 = n = 2$ we get

$$\text{FDR} = \frac{7}{16} < \alpha \ \text{ under model (1) and}$$

$$\text{FDR} = \frac{9}{16} > \alpha \ \text{ under model (2).}$$

Example 2

Bonferroni: reject H_i if $p_i \leqslant \frac{\alpha}{n}$.

$$\text{FWER} = P(V > 0) \leqslant \alpha.$$

Example

Bonferroni can not be improved under arbitrary dependency because $\text{FWER} = \alpha$ may occur.

Reason: $P(\bigcup_{i=1}^{n} A_i) \leqslant \sum_{i=1}^{n} P(A_i)$ with $' =' $ if all A_i's are disjoint.

Example 2. Graphical representation.

Divide the [0,1]-Interval into n parts.

Example 2. Graphical representation.

Throw the point U in the interval and choose a subinterval

Example 2. Graphical representation.

Duplicate the subinterval with the point U n times and get uniformly distributed $p_1, \ldots, p_6$.

Formally:
We consider for $i = 1, ..., n$ and $U \sim U(0, 1)$ the following random variables

$$p_i = \left(U + \frac{i-1}{n} \right) \text{mod} 1 \tag{3}$$

For such random variables we get

$$P(p_{1:n} \leqslant \frac{\alpha}{n}) = P(\bigcup_{i=1}^{n} \{p_i \leqslant \frac{\alpha}{n}\}) = \sum_{i=1}^{n} P(p_i \leqslant \frac{\alpha}{n}) = \alpha,$$

because all the sets $\{p_i \leqslant \frac{\alpha}{n}\}$ are disjoint.

We get FWER $= \alpha$.

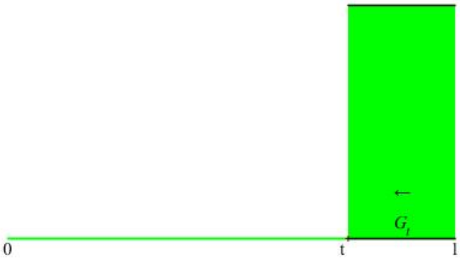
Conclusion: Specified dependence structure is needed for FDR and FWER control.

Outline

Martingale Dependence

Related to random variables $p_1, ..., p_n$ let us define the reverse filtrations

$$\mathcal{G}_t = \sigma(\mathbf{1}_{(s,1]}(p_i), 0 < t \leqslant s, i \in \{1, ..., n\}),\ 0 < t \leqslant 1.$$



Martingale Dependence

$$(M_t)_{t \in (0,1]} \text{ is a } \mathcal{G}_t - \text{ reverse martingale}$$
$$E\left[M_t | \mathcal{G}_s\right] = M_s, \text{for all } 1 \geqslant s \geqslant t > 1.$$

Definition (Reverse Martingale (RM))

We say that p-values $p_1, ..., p_n$ fulfill the Martingale Dependence Assumption RM if the process $\left(\frac{\mathbf{1}(p_i \leqslant t)}{t}\right)_{t \in (0,1]}$ is a $\mathcal{G}_t-$reverse martingale for all $i \in I_0$ ('true' p-values).

Examples of RM

Example

- Basic Independence Model fulfills RM.
- Let $X_0, X_1, ..., X_n$ be continuous, independent, real random variables, where $X_1, ..., X_n$ are i.i.d.. Consider $Z_i = \max(X_0, X_i),\ i = 1, ..., n$. Then the transformed true p-values of the form

$$p_i = H(Z_i), \text{ where } H(t) = P(Z_1 \leqslant t),\ i = 1, ..., n, \tag{4}$$

fulfill RM.
Interpretation: X_i is only observable over a random background level X_0.

Inequalities for the FDR under RM.

Theorem (Heesen and J. (2015))

Assume the reverse martingale model RM and consider the SU test with arbitrary deterministic critical values $0 < a_1 \leqslant \ldots \leqslant a_n < 1$. *Then we have*

$$\frac{n_0}{n}\left(\min_{i\leqslant n}\frac{na_i}{i}\right) \leqslant FDR \leqslant \frac{n_0}{n}\left(\max_{i\leqslant n}\frac{na_i}{i}\right). \quad (5)$$

Conclusion: BH : $a_i = \frac{\alpha i}{n} \Rightarrow \text{FDR} = \frac{n_0\alpha}{n}$.

Remark

Note that the lower bound from (5) is not valid for positive dependent true p-values (PRDS)(normal distribution).

Adaptive SU Tests with data dependent critical values

$$\text{Fix } 0 < \lambda < 1.$$

$$a_i = \min\left(\frac{i\alpha}{\hat{n}_0}, \lambda\right), \ 1 \leqslant i \leqslant n,$$

where $\frac{\hat{n}_0}{n}$ is an estimator of $\frac{n_0}{n}$.

Example (Storey's Estimator, cf. Storey (2002))

$$\hat{\pi}_0^{\text{Storey}} = \frac{\hat{n}_0}{n} = \frac{1 - \hat{F}_n(\lambda) + \frac{1}{n}}{(1-\lambda)},$$

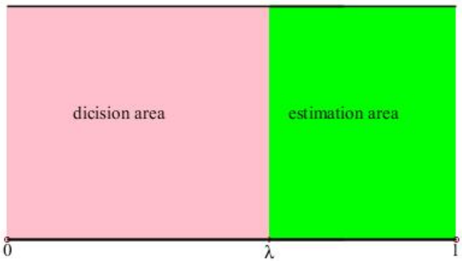
$$\text{FDR} \leqslant \alpha \ \text{ under BI.}$$

thereby $\hat{F}_n$ is edf of the p_i's.

General concept.

$$a_i = g_i((\hat{F}_n(t))_{t \geqslant \lambda}), \; i = 1, ..., n, \tag{6}$$

given by measurable functions g_i.



Result under RM

Theorem (Heesen and J. (2015), El. J. Stat.)

Let $0 < a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n \leqslant \lambda < 1$ be data dependent critical values (6) and introduce $a_0 = a_1$. Then

$$E\left[\frac{V}{na_R}\right] = \frac{n_0}{n}$$

holds for the corresponding SU tests under RM.

Key result for FDR control of adaptive SU tests under RM.
Example: Storey Type multiple tests

Result under RM

Theorem (Heesen and J. (2015),El. J. Stat.)

Let $0 < a_1 \leqslant a_2 \leqslant ... \leqslant a_n \leqslant \lambda < 1$ be data dependent critical values (6) and introduce $a_0 = a_1$. Then

$$E\left[\frac{V}{na_R}\right] = \frac{n_0}{n}$$

holds for the corresponding SU tests under RM.

Key result for FDR control of adaptive SU tests under RM.
Example: Storey Type multiple tests.

Lemma (Heesen, J.(2015),El. J. Stat.)

Let $V(\lambda) := \#\{p_i : p_i \leqslant \lambda, i \in I_0\}$. Under RM, the adaptive SU test with critical values (6) and a.s. positive estimator $\hat{n}_0 = g((\hat{F}_n(t))_{\lambda \leqslant t \leqslant 1})$ fulfills

$$E\left[\frac{V}{R}\right] = \frac{\alpha}{\lambda} E\left[V(\lambda) \min\left(\frac{1}{\hat{n}_0}, \frac{1}{n\hat{F}_n(\lambda)\alpha}\right)\right] \leqslant \frac{\alpha}{\lambda} E\left[\frac{V(\lambda)}{\hat{n}_0}\right]$$

Conclusion: BH Procedure is relative robust under dependence,

whereas Storey's Type procedures can be very liberal under dependence (also under RM)

block-wise dependence (k chromosomes)
(see Heesen and J.,(2015) Dynamic adaptive multiple tests with finite sample FDR control, arxiv: 1410.6296)

$$\frac{\hat{n}_0}{n} = \frac{1 - \hat{F}_n(\lambda) + \frac{\kappa}{n}}{1 - \lambda}, \ \kappa > 1 \text{ needed.}$$

1 Heesen, P. and Janssen, A. (2015) *Inequalities for the false discovery rate (FDR) under dependence.* El. J. Stat. 9, 679-716.
2 Heesen, P. and Janssen, A. (2015) *Dynamic adaptive multiple tests with finite sample FDR control .* Preprint ,arxiv: 1410.6296.
3 Benjamini, Y. and Hochberg, Y. (1995) *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* J. Roy. Statist. Soc. B **57** , 289-300.
4 Finner, H., Dickhaus, T. and Roters, M. (2009) *On the false discovery rate and an asymptotically optimal rejection curve.* Ann. Statist. **37**, 596-618.
5 Finner, H. and Roters, M. (2002) *On the false discovery rate and expected type I errors.* Biometrical Journal **43**, 985-1005.

Thank you for your attention!