FDR Control & Other Topics in Multiple Testing

Spencer Woody

UT-Austin, SDS 190

spencer.woody@utexas.edu

March 22, 2017

Overview

Notation

We are testing m hypotheses. Denote these hypotheses as $H_1, H_2, \ldots, H_m$ with associated p-values $p_1, p_2, \ldots, p_m$. Let $p_{(i)}$ denote the ith ordered p-value, and its corresponding hypothesis is $H_{(i)}$. A "discovery" occurs when a test statistic is declared significant.

Table: Errors in testing m hypotheses

	Non-discoveries	**Discoveries**	**Total**
True null hypotheses	U	V	m_0
Non-true null hypotheses	T	S	m_1
	$m - R$	R	m

Note that capital letters represent random variables, and let the lowercase letters u, v, t, s, and r be the respective realizations of these random variables.

The Multiplicity Problem

Fixing a confidence level at α gives a type-I error probability of α. With multiple tests, the probability of making *at least one* type-I error is

$$1 - (1 - \alpha)^m.$$

With $\alpha = 0.05$ and $m = 45$, this probablity is approximately 0.90. Consider an example of monitoring the effect of a drug on mutiple symptoms. Treatment differences may be exaggerated. This is sometimes called the *look-elsewhere effect*.

Per-comparison error rate (PCER)

The expected proportion of all hypotheses which are false positives.

$$\text{PCER} := \text{E}(V/m)$$

This essentially amounts to ignoring the multiplicity problem. Then we reject all H_i such that

$$\{i : p_i \leq \alpha\}.$$

Problem: Depending on how "sparse" the alternative hypotheses are, we may be much more likely to have a false positive than a true positive (false positive paradox).

Family-wise error rate (FWER)

The probability of having *at least one* type-I error.

$$\text{FWER} := \text{P}(V \geq 1) = 1 - \text{P}(V = 0)$$

The FWER may be controlled at a target level α with the *Bonferroni correction*, which is to reject all H_i such that

$$\left\{ i : p_i \leq \frac{\alpha}{m} \right\}$$

Problem: For large m, the cut-off point for the p-values becomes very small, so we will not capture many true effects (low power).

Definition of FDR

Concept first conceived by Soric (1989), and then formally defined by Benjamini & Hochberg (1995).

Define the random variable $Q = V/R$, the proportion of rejected hypotheses which are *erroneously* rejected. For the case of $R = 0$, we define $Q \equiv 0$. Otherwise, the FDR, notated as Q_c, is the expectation of Q,

$$Q_c = \mathrm{E}(Q) = E(V/R).$$

Note: this is distinct from the false positive rate, $\mathrm{FPR} = V/m_0$.

Some properties of the FDR

(1) If all null hypotheses are true, the FDR is equivalent to the FWER; $s = 0$, $v = r$, so $v = 0 \Rightarrow Q = 0$ and $v > 0 \Rightarrow Q = 1$, so

$$P(V \geq 1) = E(Q) = Q_c.$$

This means that controlling the FDR controls the FWER in a *weak sense* (i.e., only for this case)

(2) When $m_0 < m$ and if $v > 0$, then the FDR is less than or equal to the FWER.

$$\begin{aligned} 1 &\geq v/r \\ \Rightarrow \mathbf{1}(V \geq 1) &\geq Q \\ \Rightarrow P(V \geq 1) &\geq Q_c \end{aligned}$$

Alternative definitions

(1) The *false discovery rate* (FDR) (Benjamini & Hochberg 1995):

$$\text{FDR} := \text{E}\left(V/R|R > 0\right)\text{P}(R > 0)$$

(2) The *positive false discovery rate* (pFDR) (Storey 2003):

$$\text{pFDR} := \text{E}\left(V/R|R > 0\right)$$

(3) The *marginal false discovery rate* (mFDR) (Storey 2007):

$$\text{mFDR} := \text{E}(V)/\text{E}(R)$$

When is FDR preferable to FWER?

From Benjamini & Hochberg (1995), which mostly addresses mulitple testing in the context of clinical studies:

"The control of the FWER is important when a conclusion from the various individual inferences is likely to be erroneous when at least one of them is.... However, a treatment group and a control group are often compared by testing various aspects of the effect.... The overall conclusion that the treatment is superior need not be erroneous even if some of the null hypotheses are falsely rejected."

Some examples

- Making a hollistic decision based on multiple inferences.
- Making multiple separate decisions without an overall decision being required, which is the case in many high-throughput biological studies.
- Screening problems, such as in drug development, where we want to identify as many discoveries as possible, subject to FDR constraint.

Benjamini-Hochberg Procedure

First given by Simes (1986), and shown to control FDR by Benjamini & Hochberg (1995). Define

$$\hat{k} := \max \left\{ i : p_{(i)} \leq \frac{i}{m} q^\star \right\},$$

then, if $\hat{k}$ exists, reject all hypotheses

$$\left\{ H_{(j)} : j = 1, 2, \ldots, \hat{k} \right\}.$$

Otherwise, do not reject any hypotheses.

Theorem (FDR control)

For independent test statistics corresponding to the true null hypotheses and any configuration of false null hypotheses, the procedure described above controls the FDR at $q^\star$.

The theorem follows from the following lemma.

Lemma

For any $0 \leq m_0 \leq m$ independent p-values corresponding to the *true* null hypotheses, and for any values that the $m_1 = m - m_0$ p-values corresponding to the *false* null hypotheses, *which need not be independent*, the procedure on the previous slide satisfies

$$\mathrm{E}(Q | P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) \leq \frac{m_0}{m} q^\star.$$

Whatever the joint distribution of $P''_1, \ldots, P''_{m_1}$, integrating the inequality above gives

$$\mathrm{E}(Q) \leq \frac{m_0}{m} q^\star \leq q^\star.$$

Proof of Lemma

Proof by induction. For $m = 1$, either $m_0 = 0$, in which case $Q \equiv 0$, or $m_0 = 1$ and

$$\mathrm{E}(Q) = \mathrm{P}(V = 1) = P(U \leq q^\star) = q^\star = \frac{m_0}{m} q^\star,$$

where U represents a standard uniform random variable. Therefore, for our proof we begin by assuming that the lemma is correct for any $m' \leq m$ and show that it holds for $m + 1$.

Proof of Lemma (ctd.)

Let P'_i, $i = 1, 2, \ldots, m_0$ be the p-values corresponding to the true null hypotheses. These are iid standard uniform random variables. Denote the largest of these p-values as $P'_{(m_0)}$, which is a random variable with PDF

$$f_{P'_{(m_0)}}(p) = m_0 p^{m_0 - 1}, \ 0 < p < 1.$$

Proof of Lemma (ctd.)

Assume that the m_1 p-values corresponding to the false null hypotheses take on $p_1 \leq p_2 \leq \ldots \leq p_{m_1}$. Define

$$j_0 := \max \left\{ j : p_j \leq \frac{m_0 + j}{m + 1} q^\star \right\}$$

and

$$p'' := \frac{m_0 + j_0}{m + 1} q^\star.$$

Proof of Lemma (ctd.)

By the law of iterated expectations,

$$\begin{aligned}
&\mathrm{E}(Q|P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) \\
&= \int_0^{p''} \mathrm{E}(Q|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) f_{P'_{(m_0)}}(p) dp + \\
&\ldots \int_{p''}^1 \mathrm{E}(Q|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) f_{P'_{(m_0)}}(p) dp
\end{aligned}$$

Proof of Lemma (ctd.) [first integral]

For the first integral, $p \leq p''$, so all $m_0 + j_0$ hypotheses are rejected, therefore $Q \equiv \frac{m_0}{m_0+j_0}$ and the integral then may be evaluated

$$\int_0^{p''} \mathrm{E}(Q|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) f_{P'_{(m_0)}}(p)dp$$

$$= \int_0^{p''} \frac{m_0}{m_0 + j_0} \cdot m_0 p^{m_0-1} dp$$

$$= \frac{m_0}{m_0 + j_0} \left(p''\right)^{m_0}$$

$$= \frac{m_0}{m_0 + j_0} \cdot \frac{m_0 + j_0}{m+1} q^\star \left(p''\right)^{m_0-1}$$

$$= \frac{m_0}{m+1} q^\star \left(p''\right)^{m_0-1}$$

Proof of Lemma (ctd.) [second integral]

Consider separately each $p_{j_0} < p_j \leq P'_{(m_0)} = p < p_{j+1}$. No hypothesis may be rejected as a result of $P'_{(m_0)} = p, p_{j+1}, p_{j+2}, \ldots, p_{m_1}$, leaving now $m_0 + j - 1$ number of p-values which may be rejected. A hypothesis $H_{(i)}$ may only be rejected if $\exists\ k\ , i \leq k \leq m_0 + j - 1$ for which

$$p_{(k)} \leq \frac{k}{m+1} q^\star,$$

or equivalently,

$$\frac{p_{(k)}}{p} \leq \frac{k}{m_0 + j - 1} \cdot \frac{m_0 + j - 1}{(m+1)p} q^\star$$

Proof of Lemma (ctd.) [second integral]

When we condition on $P'_{(m_0)} = p$, the P'_i/p for $i = 1, 2, \ldots, m_0 - 1$ are distributed as standard uniform random variables, and p_i/p for $i = 1, 2, \ldots, j$ are numbers in the range $[0, 1]$ corresponding to the false null hypotheses. Using the inequality

$$\frac{p_{(k)}}{p} \leq \frac{k}{m_0 + j - 1} \cdot \frac{m_0 + j - 1}{(m+1)p} q^\star$$

is equivalent to using our original procedure, with $\frac{m_0+j-1}{(m+1)p} q^\star$ taking the place of $q^\star$. Using the induction hypothesis gives

$$\begin{aligned} \mathrm{E}(Q | P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) &\leq \frac{m_0 - 1}{m_0 + j - 1} \cdot \frac{m_0 + j - 1}{(m+1)p} q^\star \\ &= \frac{m_0 - 1}{(m+1)p} q^\star \end{aligned}$$

Proof of Lemma (ctd.) [second integral]

$$\mathrm{E}(Q|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) \leq \frac{m_0 - 1}{(m+1)p} q^\star,$$

so

$$\begin{aligned}
&\int_{p''}^1 \mathrm{E}(Q|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) f_{P'_{(m_0)}}(p) dp \\
\leq &\int_{p''}^1 \frac{m_0 - 1}{(m+1)p} q^\star \cdot m_0 p^{m_0 - 1} dp \\
= &\frac{m_0}{m+1} q^\star \left[1 - (p'')^{m_0 - 1}\right].
\end{aligned}$$

Combining these two results completes the lemma.

Power demonstration

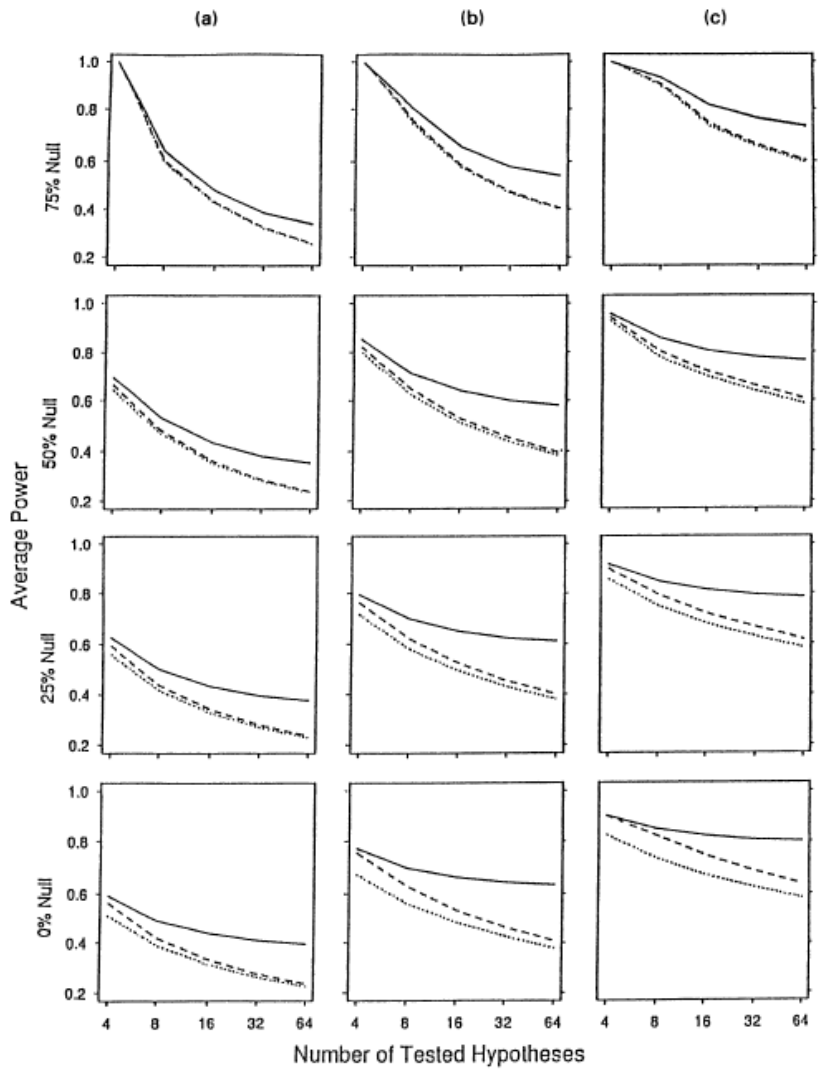

Fig. 1. Simulation-based estimates of the average power (the proportion of the false null hypotheses which are correctly rejected) for two FWER controlling methods, the Bonferroni (·····) and Hochberg's (1988) (– – – –) methods, and the FDR controlling procedure (———): (a) decreasing; (b) equally spread; (c) increasing

FDR estimation

The B-H procedure provides a way of *controlling* FDR by estimating a rejection region. However, another objective may be to *estimate* FDR by first fixing the rejection region, which falls into the standard context of point estimation. Define

$$V(t) = \#\left\{\text{true null } p_i : p_i \leq t\right\}$$
$$R(t) = \#\left\{p_i : p_i \leq t\right\},$$

then

$$\text{FDR}(t) := \begin{cases} \text{E}\left(\frac{V(t)}{R(t)}\right) & \text{if } R(t) \neq 0 \\ 0 & \text{if } R(t) = 0 \end{cases} \tag{1}$$

FDR estimation (ctd.)

Storey (2002) provides a family of conservatively biased point estimates of FDR(t),

$$\widehat{\text{FDR}}(t) = \frac{\hat{m}_0(\lambda) \cdot t}{R(t)},$$

where the $\hat{m}_0(\lambda)$ term is an estimate of m_0 and depends on the tuning parameter λ,

$$\hat{m}_0(\lambda) = \frac{m - R(\lambda)}{1 - \lambda}. \tag{2}$$

Storey & Tibshirani (2003) provide motivation for this estimator and a way of choosing the tuning parameter.

The q-value

The p-value is a measure of significance in terms of type-I error. The q-value (Storey 2001) is a measure of significance in terms of FDR. It is the minimum FDR at which the test is called significant,

$$q\text{-value}(p_i) = \min_{t \geq p_i} \text{FDR}(t),$$

and a plug-in estimate may be used

$$\hat{q}\text{-value}(p_i) = \min_{t \geq p_i} \widehat{\text{FDR}}(t).$$

For example, we expect 2% of hypotheses with a q-value of 0.02 to be false discoveries.

Thank you!

Questions

- Email: `spencer.woody@utexas.edu`
- Presentation: `github.com/spencerwoody`

References

- Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B* **85**: 289-300.
- Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance, *Biometrika* **73**: 751-754.
- Soric, B. (1989). Statistical discoveries and effect-size estimation, *Journal of the American Statistical Association* **84**: 608-610.
- Storey, J. D. (2001). The positive discovery rate: A Bayesian interpretation and the *q*-value. Technical Report 2001-12, Department of Statistics, Stanford University.
- Storey, J. D. (2002). A direct approach to false discovery rates, *Journal of the Royal Statistical Society, Series B* **64**: 479-498.
- Storey, J. D. & Tibshirani, R. (2003). Statistical significance for genome-wide studies, *Proceedings of the National Academy of Sciences* **100**: 9440-9445.