ETH zürich

False Discovery Rate – Theory and Control

Luca Pedrazzini & Marco Hassan
26. October 2020, Zurich

The Table

TABLE 1

Number of errors committed when testing m null hypotheses

	Declared non-significant	*Declared significant*	*Total*
True null hypotheses	$\mathbf{U}$	$\mathbf{V}$	m_0
Non-true null hypotheses	$\mathbf{T}$	$\mathbf{S}$	$m - m_0$
	$m - \mathbf{R}$	$\mathbf{R}$	m

Quick Reminder and FDR Idea

Paper Claim:

"Often the control of FWER is not quite needed.

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."

Measure	Formally	Control at level α
Per Comparison Error Rate	$E\left(\frac{V}{m}\right)$	Individually at level α
Family Wise Error Rate	$P(V \geq 1)$	Individually at level $\frac{\alpha}{m}$
False Discovery Rate	$E\left(\frac{V}{R}\right)$	?

New Proposal: FDR

For many applications the question is: how many false positive are we ready to tolerate and not if *any* false positve occur.

Idea: Control the proportion of type I errors.

Definition - FDR:

$$Q_E = E(Q) = E\{\frac{V}{V+S}\} = E(\frac{V}{R})$$

where:

$$(V+S) = 0 \Rightarrow \mathrm{Q}_E = 0$$

FDR in Perspective

Measure	Formally	Control at level α
Per Comparison Error Rate	$\mathrm{E}\ (\frac{V}{m})$	Individually at level α
Family Wise Error Rate	$\mathrm{P}\ (\mathrm{V} \geq 1)$	Individually at level $\frac{\alpha}{m}$
False Discovery Rate	$\mathrm{E}\ (\frac{V}{R})$	?

Claims

Claim 1:

In the case of a global null, there is FWER control by the FDR (weak sense control).

Proof:

It follows immediately by noticing that v = r.

So that either Q = 0 or Q =1.

$$P(V \geq 1) = \mathrm{E}\ (\mathrm{Q}) = \mathrm{Q}_E$$

$\square$

Claims

Claim 2:

$$m_0 < m \implies FDR \leq FWER$$

Interpretation:

If you control FWER you automatically control FDR.

Proof:

Also straightforward after noticing that given the condition v $> 0 \Rightarrow \frac{v}{r} \leq 1$.

It follows that for the two random variables Ψ and Q:

$$\Psi_{(V \geq 1)} \geq Q$$

Such that in expectation FWER $\geq$ FDR. $\square$

Takeaway – Claim 2

FWER is more restrictive then FDR, you loose power. Always ask yourself when does it make sense to pay that price.

The higher the number of non-true Nulls the higher the price you pay.

Heuristics:

Set control of FWER and FDR at the same level α

Then obvious due to the more restrictive FWER that:

$$if\ m - m_0 \uparrow \Rightarrow \frac{S_{FDR}}{S_{FWER}} \uparrow \iff more\ power$$

FDR practice – strong Control

So far we only showed that it must hold theoretically that FDR $\leq$ FWER.

What if we are interested now at controlling the FDR at a given level.

I.e. how can we gain the theoretical demonstrated gain in power?

Measure	Formally	Control at level α
Per Comparison Error Rate	$\mathrm{E}\left(\frac{V}{m}\right)$	Individually at level α
Family Wise Error Rate	$\mathrm{P}\left(\mathrm{V} \geq 1\right)$	Individually at level $\frac{\alpha}{m}$
False Discovery Rate	$\mathrm{E}\left(\frac{V}{R}\right)$	?

FDR strong Control – Theorem

Notice - a **step-up procedure!**

Notice - strong parallel with Simes (1986) (treated by the first presented group!).

Consider testing $H_1, H_2, \ldots, H_m$ based on the corresponding p-values $P_1, P_2, \ldots, P_m$. Let $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(m)}$ be the ordered p-values, and denote by $H_{(i)}$ the null hypothesis corresponding to $P_{(i)}$. Define the following Bonferroni-type multiple-testing procedure:

$$\textit{let } k \textit{ be the largest } i \textit{ for which } P_{(i)} \leq \frac{i}{m} q^*;$$

then reject all $H_{(i)}$, with $i = 1, 2, \ldots, k$.

Proof of the Theorem

Notice – the proof presented in the paper overlaps extensively to the one presented by Simes and exposed by the first presenting group

We propose an alternative *proof by Candes and Barbers* to foster understanding in this group

Theorem 1: For independent test statistics and for any configuration of false null hypothesis, the above procedure controls the FDR at q*

Goal:

$$BH_q \Rightarrow FDR_q \leq q$$

Claim:

$$BH_q \Rightarrow FDR_q = \frac{m_0}{m} q$$

We will show the claim.

Assumption needed:

$$- p_i \; associated \; with \; true \; Nulls \; H_0 \; are \; independent \; from \; each \; other \; and \; from \; the \; Non - true \; Null \; Hypothesis$$

$$- p_i \; are \; uniformly \; distributed \; with \sim U(0,1)$$

Proof:
Step 1: Define

$$\mathrm{FDP} = \sum_{i \in H_0} \frac{V_i}{R \vee 1}.$$

with:

$$V_i = \mathbb{1}_{\{H_i \text{rejected}\}} \text{ for each null } i \in H_0$$

such that:

$$\mathrm{FDR} = \mathbb{E}[\mathrm{FDP}] = \sum_{i \in H_0} \mathbb{E}\left[\frac{V_i}{R \vee 1}\right]$$

Proof cont'd :

Step 2: Rewrite the Random variable $\frac{V_i}{R \vee 1}$, i.e. express the Random Variable *R* by an exhaustive search of the possible space:

$$\frac{V_i}{R \vee 1} = \sum_{k=1}^{m} \frac{V_i \mathbb{1}_{\{R=k\}}}{k}$$

Step 3: Leverage the properties of the BH step up procedure, such that, ***assuming***/***fixing*** *R=k* rejections it holds:

$$V_i = \mathbb{1}_{\{H_i \, rejected\}} \; with \; i \in H_0 = \mathbb{1}_{\left\{p_i \leq \frac{k}{m} q\right\}}$$

Proof cont'd :

Step 4: Assume you set $p_i = 0$, then you define the new number of rejections $R(p_i \to 0)$

$$\begin{cases} case\ 1\ if\ before\ \ p_i \leq \frac{k}{m} * q: \\ \\ case\ 2: if\ before\ \ p_i > \frac{k}{m} * q: \end{cases}$$

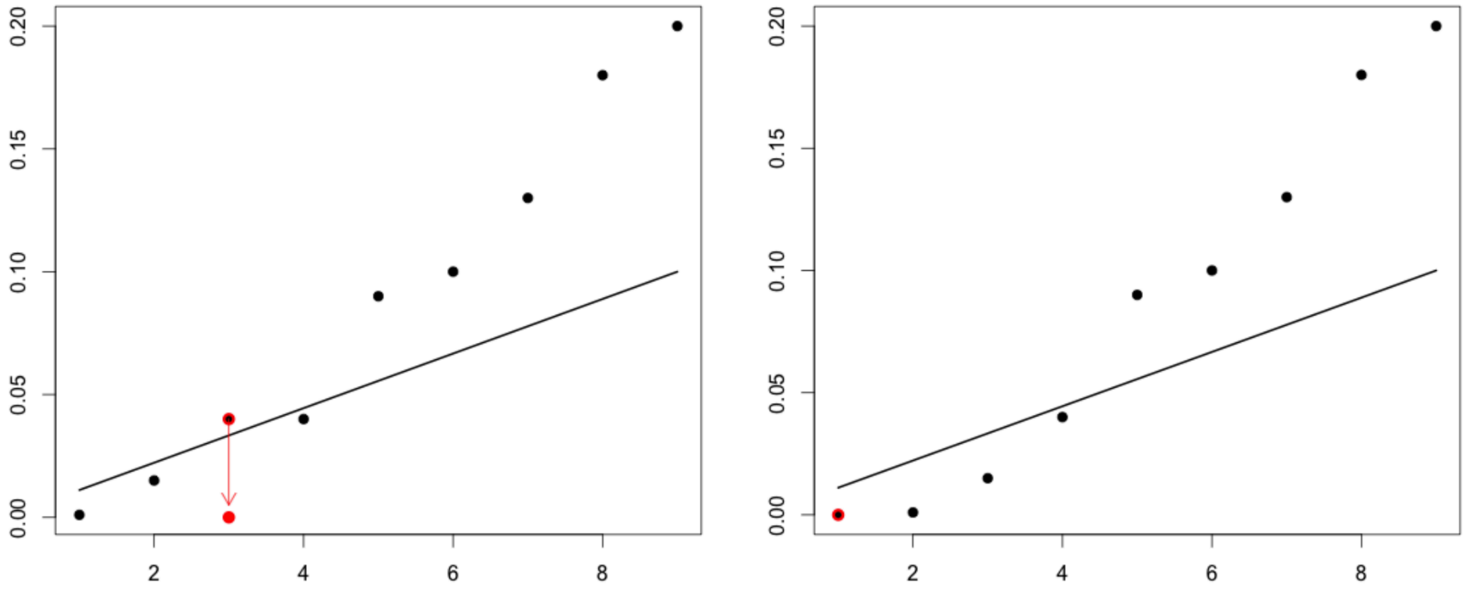
p_i was already rejected, so the number of rejections is the same.
despite the new p_i value. So nothing has changed- ***not dangerous***

p_{i0} was not rejected, so the number of rejections is increased.
Something changed - ***dangerous***

Proof cont‘d :

Illustration Step 4:

Nota bene the red dot at i=3 must be smaller than the one at i = 4 by definition of ordered p-values.



Proof cont'd :

Step 5: Combining Step 3 ($V_i = \mathbb{1}_{\{H_i rejected\}} = \mathbb{1}_{\left\{p_i \leq \frac{k}{m} q\right\}}$) and Step 4, we can easily see that:

$$V_i * \mathbb{1}_{\{R=k\}} = V_i * \mathbb{1}_{\{R(p_i \to 0) = k\}}$$

Simple idea: in the dangerous case previously mentioned we have by definition $V_i = 0$ so that our result is unaffected by a change in the indicator function.

Proof cont'd :

Step 6: It is now straightforward to see that:

Combining the observations above and taking the expectation conditional on all p-values expect for p_i, i.e., $\mathcal{F}_i = \{p_1, ..., p_{i-1}, ..., p_{i+1}, p_m\}$, we have

$$\mathbb{E}\left[\frac{V_i}{R \vee 1}\middle|\mathcal{F}_i\right] = \sum_{k=1}^{m} \frac{\mathbb{E}\left[\mathbb{1}_{\{p_i \le qk/m\}} \cdot \mathbb{1}_{\{R(p_i \to 0)=k\}} | \mathcal{F}_i\right]}{k} = \sum_{k=1}^{m} \frac{\mathbb{1}_{\{R(p_i \to 0)=k\}} qk/m}{k},$$

where the second equality holds because knowing $\mathcal{F}_i$ and $p_i = 0$ makes $\mathbb{1}_{\{R(p_i \to 0)=k\}}$ deterministic.

= recall the U(0,1) assumption for the p_{i0}

Proof cont'd :

Step 7: Trivial to see

$$\mathbb{E}\left[\frac{V_i}{R \vee 1}\middle|\mathcal{F}_i\right] = \frac{q}{m}\sum_{k=1}^{m} \mathbb{1}_{\{R(p_i \to 0)=k\}} = \frac{q}{m}.$$

With:

$$\sum_{k=1}^{m} \mathbb{1}_{\{R(p_i \to 0)=k\}} = 1$$

Proof cont'd :

Step 8: Final trivial result

$$\mathbb{E}\left[\frac{V_i}{R \vee 1}\middle|\mathcal{F}_i\right] = \frac{q}{m} \quad \Rightarrow \quad \mathrm{FDR} = \mathbb{E}[\mathrm{FDP}] = \sum_{i \in H_0} \mathbb{E}\left[\frac{V_i}{R \vee 1}\right] = \sum_{i \in H_0} \frac{q}{m} = \frac{m_0}{m} q.$$

$\square$

Introduction (part 2)

1) Benjamini-Hochberg
   - $FDR = E\left[\frac{V}{R \vee 1}\right]$
   - Step-up p-value method controls the FDR
   - Controls the FDR at a fixed predetermined level

2) Storey
   - $FDR(t) = E\left[\frac{V(t)}{R(t) \vee 1}\right]$, where $t \in [0, 1]$
   - Estimate of FDR, i.e. $\widehat{FDR}_\lambda(t)$ where λ is a tuning parameter
   - Controls the FDR for a fixed predetermined significance region $[0, t]$

Storey

Why do we use this new procedure?

1) Provide strong control of the FDR at level α.
2) With a replacement this procedure is equivalent to the BH procedure (new proof of BH using martingales).
3) For the asymptotic case, Storey procedure has a greater power than the BH procedure.

We assume that p-values corresponding to the true null hypotheses are:

- independent,
- uniformly distributed.

Drawback: By a theoretical point of view, Storey procedure is sensitive to assumptions.

What does FDR(t) mean?

FDR(t) denote the FDR when rejecting all H_0 with $p_i \leq t$ for $i = 1, ..., m$.

Empirical processes:

- $V(t) = \#\{\text{null true } p_i : p_i \leq t\}$,
- $S(t) = \#\{\text{non-true null } p_i : p_i \leq t\}$,
- $R(t) = V(t) + S(t) = \#\{p_i : p_i \leq t\}$.

Estimate of FDR(t)

The estimate of FDR(t) is defined as

$$\widehat{FDR}_\lambda(t) = \frac{\widehat{\pi}_0(\lambda)t}{\{R(t) \vee 1\}/m}.$$

The estimate $\hat{\pi}_0(\lambda)$ of $\pi_0 = m_0/m$ is defined as

$$\widehat{\pi}_0(\lambda) = \frac{W(\lambda)}{(1-\lambda)m},$$

where $W(\lambda) = m - R(\lambda)$.
Interpretation of $\widehat{\pi}_0(\lambda)$: $W(\lambda)/m \approx \pi_0(1-\lambda)$, since we expect $\pi_0(1-\lambda)$ of the p-values to lie in the interval $(\lambda, 1]$.

Conservative estimate of FDR(t)

Theorem 1: Suppose that the p-values corresponding to the true null hypotheses are independent and uniformly distributed. Then, for fixed $\lambda \in [0, 1)$, $E[\widehat{FDR}_\lambda(t)] \geq FDR(t)$.

How do we choose the tuning parameter λ?

We require the choice of the tuning parameter $\lambda \in [0, 1)$ in the estimate $\hat{\pi}_0(\lambda)$. The estimate $\hat{\pi}_0(\lambda)$ has a bias-variance trade-off situation, i.e.

- bias increases $\Leftrightarrow$ variance decreases
- bias decreases $\Leftrightarrow$ variance increases

Search for λ that finds a balance between bias and variance

Calculate this using an algorithm which uses the bootstrap procedure and the MSE of $\hat{\pi}_0(\lambda)$. [1]

In the theoretical we are going to work with a fixed λ.

Storey procedure

Need to find the largest t such that $\widehat{FDR}_\lambda(t) \leq \alpha$, which is defined as

$$t_\alpha(\widehat{FDR}_\lambda) = \sup\left\{0 \leq t \leq 1 : \widehat{FDR}_\lambda(t) \leq \alpha\right\}.$$

Final goal is

$$FDR\{t_\alpha(\widehat{FDR}_\lambda)\} \leq \alpha,$$

where we have

$$FDR\{t_\alpha(\widehat{FDR}_\lambda)\} = E\left[\frac{V\{t_\alpha(\widehat{FDR}_\lambda)\}}{R\{t_\alpha(\widehat{FDR}_\lambda)\} \vee 1}\right].$$

Equivalence with BH

Lemma 1: The p-value step-up method $t_\alpha(\widehat{FDR}_{\lambda=0})$ is equivalent to the BH procedure.

Lemma 2: The p-value step-up method $t_\alpha(\widehat{FDR}_\lambda)$ is equivalent to the BH procedure with m replace by $\hat{\pi}_0(\lambda)m$.

Remark: In other words we are showing $p_{k_\lambda} \leq t_\alpha(\widehat{FDR}_\lambda) < p_{k_\lambda+1}$, where $k_\lambda = \max\left\{1 \leq i \leq m : p_{(i)} \leq \frac{i}{\hat{\pi}_0(\lambda)m}\alpha\right\}$.

Proof (Lemma 1): For $\lambda = 0$ we have $\hat{\pi}_0(0) = \frac{W(0)}{(1-0)m} = \frac{m}{m} = 1$, because $W(0) = \#\{p_i > 0\} = m$. Using Lemma 2 we conclude the proof. $\square$

Proof of Lemma 2

Proof (Lemma 2): We need to show that $p_{k_\lambda} \leq t_\alpha(\widehat{FDR}_\lambda) < p_{k_\lambda+1}$, where $k_\lambda = \max\left\{1 \leq i \leq m : p_{(i)} \leq \frac{i}{\hat{\pi}_0(\lambda)m}\alpha\right\}$, i.e. BH procedure with m replace by $\hat{\pi}_0(\lambda)m$.

Case 1: $R(t) = 0$, trivial because both procedures reject no p-values.

Case 2: $R(t) \neq 0$.

$\widehat{FDR}_\lambda(p_{(i)}) = \frac{\hat{\pi}_0(\lambda)p_{(i)}}{\{R(p_{(i)}) \vee 1\}/m} = \frac{\hat{\pi}_0(\lambda)p_{(i)}m}{i}$, because $R(p_{(i)}) \vee 1 = R(p_{(i)}) = i$.

We can write it as $p_{(i)} = \frac{\widehat{FDR}_\lambda(p_{(i)})i}{\hat{\pi}_0(\lambda)m}$.

We insert this equation in $k_\lambda = \max\left\{1 \leq i \leq m : p_{(i)} \leq \frac{i}{\hat{\pi}_0(\lambda)m}\alpha\right\}$ and we get $k_\lambda = \max\left\{1 \leq i \leq m : \widehat{FDR}_\lambda(p_{(i)}) \leq \alpha\right\}$.

This shows the connection with the definition of $t_\alpha(\widehat{FDR}_\lambda)$. $\square$

Control the FDR for λ = 0

Theorem 2 (Benjamini and Hochberg, 1995): If the p-values corresponding to the true null hypotheses are independent, then

$$FDR\left\{t_\alpha(\widehat{FDR}_{\lambda=0})\right\} = \pi_0\alpha \leq \alpha.$$

Proof of Theorem 2 (definitions)

Def (filtration): A family of σ-algebras $\mathcal{F} = \{\mathcal{F}_t : t \in T\}$ is a *filtration* on $(\Omega, \mathcal{F})$ if $s, t \in T$ and $s \leq t$ imply $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$.

Def (martingale): The process $X = (X_t)_{t \in T}$ is a *martingale* (*backwards martingale*) with respect to a filtration $\mathcal{F}$ if $E(X_t | \mathcal{F}_s) = X_s$ $(E(X_s | \mathcal{F}_t) = X_t)$ $\forall s, t \in T$ with $s \leq t$.

Def (stopping time): Suppose that $\mathcal{F}$ is a filtration on $(\Omega, \mathcal{F})$. A random time τ is a *stopping time* with respect to $\mathcal{F}$ if $\{\tau \leq t\} \in \mathcal{F}_t$ $\forall t \in T$.

Proof of Theorem 2 (lemmas)

Lemma 3: If the p-values of the m_0 true null hypotheses are independent, then $V(t)/t$ for $0 \leq t < 1$ is a martingale with time running backwards with respect to the filtration $\mathcal{F}_t = \sigma(\mathbf{1}_{\{p_i \leq l\}})$, $t \leq l \leq 1, i = 1, \ldots, m$, i.e. for $s \leq t$, $E[V(s)/s|\mathcal{F}_t] = V(t)/t$.

Lemma 4: The random variable $t_\alpha(\widehat{FDR}_{\lambda=0})$ is a stopping time with respect to $\mathcal{F}_t$, with time running backwards.

Proof of Theorem 2 (optional stopping thm)

Theorem (optional stopping theorem): Let $X = (X_t)_{t \in T}$ be a martingale and τ a stopping time with respect to the filtration. Assume that there exists a constant c such that $|X_{t \wedge \tau}| \leq c$ a.s. for all $t \in T$. Then $E[X_\tau] = E[X_0]$.
Here we have $X_t = V(t)/t$, $\tau = t_\alpha(\widehat{FDR}_{\lambda=0})$ and $c = m/\alpha$. Backwards running time: $E[X_\tau] = E[X_1]$, for $t \in [0, 1]$.

Proof of Theorem 2

Proof:

$$FDR\{t_\alpha(\widehat{FDR}_{\lambda=0})\} = \frac{\alpha}{m} E\left[\frac{V\{t_\alpha(\widehat{FDR}_{\lambda=0})\}}{t_\alpha(\widehat{FDR}_{\lambda=0})}\right] = \frac{\alpha}{m} E\left[\frac{V(1)}{1}\right] = \frac{m_0}{m}\alpha.$$

$$\frac{mt_\alpha}{R(t_\alpha)} = \alpha$$

$V(t)/t$ martingale
$t_\alpha(\widehat{FDR}_{\lambda=0})$ stopping time
$E[V(t_\alpha)/t_\alpha] = E[V(1)/1]$

$\square$

Control the FDR for λ > 0

Theorem 3: If the p-values corresponding to the true null hypotheses are independent, then for $\lambda > 0$,

$$FDR\left\{t_\alpha(\widehat{FDR}^{\star}_{\lambda})\right\} \leq (1 - \lambda^{\pi_0 m})\alpha \leq \alpha.$$

Remark: A well chosen λ will tend to be larger than $t_\alpha(\widehat{FDR}^{\star}_{\lambda})$ and so there should be little difference between $t_\alpha(\widehat{FDR}^{\star}_{\lambda})$ and $t_\alpha(\widehat{FDR}_{\lambda})$.

Situation for λ > 0

Two modifications to $\widehat{FDR}_\lambda$:

- need to guarantee $\hat{\pi}_0(\lambda) > 0$, so we replace $\hat{\pi}_0(\lambda)$ with

$$\hat{\pi}_0^\star(\lambda) = \frac{W(\lambda)+1}{(1-\lambda)m}.$$

- need to limit the significance threshold to the region $[0, \lambda]$, so instead of $\widehat{FDR}_\lambda$ we get

$$\widehat{FDR}^\star_\lambda(t) = \begin{cases} \frac{\hat{\pi}_0^\star(\lambda)t}{\{R(t) \vee 1\}/m}, & \text{if } t \le \lambda \\ 1, & \text{if } t > \lambda. \end{cases}$$

Summary

Storey procedure

Step 1: Let α be the prechosen level at which to control the FDR.

Step 2: For any fixed significance region $[0, t]$, estimate $FDR(t)$ by either $\widehat{FDR}_\lambda(t)$ or $\widehat{FDR}^\star_\lambda(t)$.

Step 3: For m hypotheses, where the null p-values are independent, reject all null hypotheses corresponding to $p_i \leq t_\alpha(\widehat{FDR}^\star_\lambda)$ for $\lambda > 0$ and $p_i \leq t_\alpha(\widehat{FDR}_{\lambda=0})$ for $\lambda = 0$.

Recap

1) Benjamini-Hochberg procedure:
   - $k = \max\{1 \leq i \leq m : p_{(i)} \leq \alpha i/m\}$, $p_{(1)} \leq \ldots \leq p_{(m)}$
   - $FDR \leq FWER$
   - $FDR_q = m_0 q/m \leq q$, where $q = \alpha$

2) Storey procedure:
   - $\widehat{FDR}_\lambda(t)$ and $\widehat{FDR}^\star_\lambda(t) \rightsquigarrow t_\alpha$
   - $\lambda = 0$: equivalent to BH
     $\lambda > 0$: equivalent to BH with m replaced by $\hat{\pi}_0(\lambda)m$
   - $\lambda = 0$: $FDR\{t_\alpha(\widehat{FDR}_{\lambda=0})\} = \pi_0\alpha \leq \alpha$
     $\lambda > 0$: $FDR\{t_\alpha(\widehat{FDR}^\star_\lambda)\} \leq (1 - \lambda^{\pi_0 m})\alpha \leq \alpha$

References

Yoav Benjamini and Yosef Hochberg
Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing
J. R. Statist. Soc. B (1995) 57, No. 1, pp. 289-300

John D. Storey, Jonathan E. Taylor and David Siegmund
Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach
J. R. Statist. Soc. B (2004) 66, Part 1, pp. 187–205

John D. Storey
A direct approach to false discovery rates
J. R. Statist. Soc. B (2002) 64, Part 3, pp. 479–498