Resampling-Based Control of the FDR

Joseph P. Romano[1] Azeem S. Shaikh[2] and Michael Wolf[3]

[1]Departments of Economics and Statistics
Stanford University

[2]Department of Economics
University of Chicago

[3]Department of Economics
University of Zurich

Outline

Outline

General Set-Up & Notation

- Data $X = (X_1, \ldots, X_n)$ from distribution P
- Interest in parameter vector $\theta(P) = \theta = (\theta_1, \ldots, \theta_s)'$
- The individual hypotheses concern the elements θ_i, for $i = 1, \ldots, s$, and can be (all) one-sided or (all) two-sided

One-sided hypotheses:

$$H_i: \theta_i \leq \theta_{0,i} \quad \text{vs.} \quad H_i': \theta_i > \theta_{0,i}$$

Two-sided hypotheses:

$$H_i: \theta_i = \theta_{0,i} \quad \text{vs.} \quad H_i': \theta_i \neq \theta_{0,i}$$

General Set-Up & Notation

- Data $X = (X_1, \ldots, X_n)$ from distribution P
- Interest in parameter vector $\theta(P) = \theta = (\theta_1, \ldots, \theta_s)'$
- The individual hypotheses concern the elements θ_i, for $i = 1, \ldots, s$, and can be (all) one-sided or (all) two-sided

One-sided hypotheses:

$$H_i: \theta_i \leq \theta_{0,i} \quad \text{vs.} \quad H_i': \theta_i > \theta_{0,i}$$

Two-sided hypotheses:

$$H_i: \theta_i = \theta_{0,i} \quad \text{vs.} \quad H_i': \theta_i \neq \theta_{0,i}$$

- Test statistic $T_{n,i} = (\hat{\theta}_{n,i} - \theta_{0,i})/\hat{\sigma}_{n,i}$ or $T_{n,i} = |\hat{\theta}_{n,i} - \theta_{0,i}|/\hat{\sigma}_{n,i}$
- $\hat{\sigma}_{n,i}$ is a standard error for $\hat{\theta}_{n,i}$ or $\hat{\sigma}_{n,i} \equiv 1/\sqrt{n}$
- $\hat{p}_{n,i}$ is an individual p-value

The False Discovery Rate

Consider s individual tests H_i vs. H_i'.

False discovery proportion

F = # false rejections; R = # total rejections

$$\text{FDP} = \frac{F}{R} 1\{R > 0\} = \frac{F}{\max\{R, 1\}}$$

False discovery rate

- $\text{FDR}_P = \text{E}_P(\text{FDP})$

Goal: (strong) asymptotic control of the FDR at level α:

$$\limsup_{n \to \infty} \text{FDR}_P \leq \alpha \quad \text{for all } P$$

Outline

Benjamini and Hochberg (1995)

Stepup method:

- Ordered p-values: $\hat{p}_{n,(1)} \leq \hat{p}_{n,(2)} \leq \ldots \leq \hat{p}_{n,(s)}$
- Let $j^* = \max \left\{ j : \hat{p}_{n,(j)} \leq \alpha_j \right\}$, where $\alpha_j = j\alpha/s$
- Reject $H_{(1)}, \ldots, H_{(j^*)}$

Comments:

- Original proof assumes independence of the p-values
- Validity has been extended to certain dependence types (Benjamini and Yekutieli, 2001)

Modifications of BH (1995)

Storey, Taylor and Siegmund (2004):

- Under sufficient conditions for BH (1995):

$$\mathrm{FDR}_P \leq \frac{s_0}{s}\alpha \quad \text{where} \quad s_0 = |I(P)| = \#\{\text{true hypotheses}\}$$

- Instead of $\alpha_j = j\alpha/s$ use $\alpha_j = j\alpha/\hat{s}_0$ with

$$\hat{s}_0 = \frac{\#\{\hat{p}_{n,i} > \lambda\} + 1}{1 - \lambda} \quad \text{for some } 0 < \lambda < 1$$

- Proof assumes the $\hat{p}_{n,i}$ to be ‘almost independent’

Modifications of BH (1995)

Bejamini, Krieger and Yekutieli (2006):

Step 1:

- Apply the BH (1995) procedure at nominal level $\alpha^* = \alpha/(1+\alpha)$
- Let r denote the number of rejected hypotheses

(a) If $r = 0$, reject nothing, and stop

(b) If $r = s$, reject everything, and stop

(c) Otherwise, continue

Step 2:

- Apply the BH (1995) procedure at nominal level α
- Instead of $\alpha_j = j\alpha/s$ use $\alpha_j = j\alpha/\hat{s}_0$ with

$$\hat{s}_0 = s - r$$

- Proof assumes independence of the $\hat{p}_{n,i}$, but simulations show robustness against various dependence structures

Outline

Basic Idea (Troendle, 2000)

For any stepdown procedure with critical values $c_1, \ldots, c_s$:

$$\text{FDR}_P = \text{E}_P\left[\frac{F}{\max\{R,1\}}\right] = \sum_{1 \le r \le s} \frac{1}{r} \text{E}_P[F|R=r] P\{R=r\}$$

with $P\{R=r\} = P\{T_{n,(s)} \ge c_s, \ldots, T_{n,(s-r+1)} \ge c_{s-r+1}, T_{n,(s-r)} < c_{s-r}\}$

Basic Idea (Troendle, 2000)

For any stepdown procedure with critical values $c_1, \ldots, c_s$:

$$\mathrm{FDR}_P = \mathrm{E}_P\left[\frac{F}{\max\{R,1\}}\right] = \sum_{1 \le r \le s} \frac{1}{r} \mathrm{E}_P[F|R=r] P\{R=r\}$$

with $P\{R=r\} = P\{T_{n,(s)} \ge c_s, \ldots, T_{n,(s-r+1)} \ge c_{s-r+1}, T_{n,(s-r)} < c_{s-r}\}$

If all false hypotheses are rejected with $p. \to 1$, then with $p. \to 1$:

$$\mathrm{FDR}_P = \sum_{s-s_0+1 \le r \le s} \frac{r-s+s_0}{r} \tag{1}$$

$$\times P\{T_{n,s_0:s_0} \ge c_{s_0}, \ldots, T_{n,s-r+1:s_0} \ge c_{s-r+1}, T_{n,s-r:s_0} < c_{s-r}\}$$

Here $T_{n,r:t}$ is the rth largest of the test statistics $T_{n,1}, \ldots, T_{n,t}$, and we assume w.l.o.g. that $I(P) = \{1, \ldots, s_0\}$.

Basic Idea (continued)

Goal:

- Bound (1) above by α for any P, at least asymptotically
- In particular, this must be ensured for any $1 \leq s_0 \leq s$.

First, consider any P such that $s_0 = 1$:

- Then (1) reduces to $\frac{1}{s} P\{T_{n,1:1} \geq c_1\}$
- And so $c_1 = \inf\{x \in \mathbb{R} : \frac{1}{s} P\{T_{n,1:1} \geq x\} \leq \alpha\}$

Basic Idea (continued)

Goal:

- Bound (1) above by α for any P, at least asymptotically
- In particular, this must be ensured for any $1 \leq s_0 \leq s$.

First, consider any P such that $s_0 = 1$:

- Then (1) reduces to $\frac{1}{s} P\{T_{n,1:1} \geq c_1\}$
- And so $c_1 = \inf\{x \in \mathbb{R} : \frac{1}{s} P\{T_{n,1:1} \geq x\} \leq \alpha\}$

Next, consider any P such that $s_0 = 2$. Then (1) reduces to:

- $\frac{1}{s-1} P\{T_{n,2:2} \geq c_2, T_{n,1:2} < c_1\} + \frac{2}{s} P\{T_{n,2:2} \geq c_2, T_{n,1:2} \geq c_1\}$
- And so c_2 is the smallest $x \in \mathbb{R}$ for which $\frac{1}{s-1} P\{T_{n,2:2} \geq x, T_{n,1:2} < c_1\} + \frac{2}{s} P\{T_{n,2:2} \geq x, T_{n,1:2} \geq c_1\} \leq \alpha$

Basic Idea (continued)

Goal:

- Bound (1) above by α for any P, at least asymptotically
- In particular, this must be ensured for any $1 \leq s_0 \leq s$.

First, consider any P such that $s_0 = 1$:

- Then (1) reduces to $\frac{1}{s} P\{T_{n,1:1} \geq c_1\}$
- And so $c_1 = \inf\{x \in \mathbb{R} : \frac{1}{s} P\{T_{n,1:1} \geq x\} \leq \alpha\}$

Next, consider any P such that $s_0 = 2$. Then (1) reduces to:

- $\frac{1}{s-1} P\{T_{n,2:2} \geq c_2, T_{n,1:2} < c_1\} + \frac{2}{s} P\{T_{n,2:2} \geq c_2, T_{n,1:2} \geq c_1\}$
- And so c_2 is the smallest $x \in \mathbb{R}$ for which $\frac{1}{s-1} P\{T_{n,2:2} \geq x, T_{n,1:2} < c_1\} + \frac{2}{s} P\{T_{n,2:2} \geq x, T_{n,1:2} \geq c_1\} \leq \alpha$

And so forth ...

Estimation of the c_i

Since P is unknown, so are the 'ideal' critical values c_i.

We sugggest a bootstrap method to estimate the c_i:

- $\hat{P}_n$ is an *unrestricted* estimate of P with $\theta_i(\hat{P}_n) = \hat{\theta}_{n,i}$
- X^* is generated from $\hat{P}_n$ and the $T^*_{n,i}$ are computed from X^*, but centered at $\hat{\theta}_{n,i}$ rather than at $\theta_{0,i}$
- E.g., for one-sided testing: $T^*_{n,i} = (\hat{\theta}^*_{n,i} - \hat{\theta}_{n,i})/\hat{\sigma}^*_{n,i}$

Estimation of the c_i

Since P is unknown, so are the 'ideal' critical values c_i.

We sugggest a bootstrap method to estimate the c_i:

- $\hat{P}_n$ is an *unrestricted* estimate of P with $\theta_i(\hat{P}_n) = \hat{\theta}_{n,i}$
- X^* is generated from $\hat{P}_n$ and the $T^*_{n,i}$ are computed from X^*, but centered at $\hat{\theta}_{n,i}$ rather than at $\theta_{0,i}$
- E.g., for one-sided testing: $T^*_{n,i} = (\hat{\theta}^*_{n,i} - \hat{\theta}_{n,i})/\hat{\sigma}^*_{n,i}$

Important detail:

- The ordering of the original $T_{n,i}$ determines the 'true' null hypotheses in the bootstrap world
- The permutation $\{k_1, \ldots, k_s\}$ of $\{1, \ldots, s\}$ is defined such that $T_{n,k_1} = T_{n,(1)}, \ldots, T_{n,k_s} = T_{n,(s)}$
- Then $T^*_{n,r:t}$ is the rth smallest of the statistics $T^*_{n,k_1}, \ldots, T^*_{n,k_t}$

Estimation of the c_i (continued)

Start with c_1:

- $\hat{c}_1 = \inf\{x \in \mathbb{R} : \frac{1}{s}\hat{P}_n\{T^*_{n,1:1} \geq x\} \leq \alpha\}$

Estimation of the c_i (continued)

Start with c_1:

- $\hat{c}_1 = \inf\{x \in \mathbb{R} : \frac{1}{s}\hat{P}_n\{T^*_{n,1:1} \geq x\} \leq \alpha\}$

Then move on to c_2:

- $\hat{c}_2$ is the smallest $x \in \mathbb{R}$ for which

$$\frac{1}{s-1}\hat{P}_n\{T^*_{n,2:2} \geq x, T^*_{n,1:2} < \hat{c}_1\} + \frac{2}{s}\hat{P}_n\{T^*_{n,2:2} \geq x, T^*_{n,1:2} \geq \hat{c}_1\} \leq \alpha$$

Estimation of the c_i (continued)

Start with c_1:

- $\hat{c}_1 = \inf\{x \in \mathbb{R} : \frac{1}{s}\hat{P}_n\{T^*_{n,1:1} \geq x\} \leq \alpha\}$

Then move on to c_2:

- $\hat{c}_2$ is the smallest $x \in \mathbb{R}$ for which

$$\frac{1}{s-1}\hat{P}_n\{T^*_{n,2:2} \geq x, T^*_{n,1:2} < \hat{c}_1\} + \frac{2}{s}\hat{P}_n\{T^*_{n,2:2} \geq x, T^*_{n,1:2} \geq \hat{c}_1\} \leq \alpha$$

And so forth ...

Unlike Troendle (2000), monotonicity $\hat{c}_{i+1} \geq \hat{c}_i$ is not enforced.

Outline

Some Theory

Assumptions

(1) The sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ under P converges to a limit distribution with continuous marginals

(2) The bootstrap consistently estimates this limit distribution

(3) $\sqrt{n}\hat{\sigma}_{n,i}$ and $\sqrt{n}\hat{\sigma}^*_{n,i}$ converge to the same constant in probability (for $i = 1, \ldots, s$)

(4) The limiting joint distribution corresponding to the 'true' test statistics is exchangeable

Some Theory

Assumptions

(1) The sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ under P converges to a limit distribution with continuous marginals

(2) The bootstrap consistently estimates this limit distribution

(3) $\sqrt{n}\hat{\sigma}_{n,i}$ and $\sqrt{n}\hat{\sigma}^*_{n,i}$ converge to the same constant in probability (for $i = 1, \ldots, s$)

(4) The limiting joint distribution corresponding to the 'true' test statistics is exchangeable

Theorem

(i) *Any false H_i will be rejected with p. $\to 1$ as $n \to \infty$*

(ii) *The method asymptotically controls the FDR at level α*

Some Practice

Assumption (4) is somewhat restrictive
(though less restrictive than an assumption of independence)

But simulations indicate that the method appears robust to

- different limiting variances of the ‘true’ test statistics
- different limiting correlations of the ‘true’ test statistics

Outline

Set-Up I

Data generating process and testing problem:

- I.i.d. random vectors from $N(\theta,\Sigma)$, with $n = 100$
- $\theta_i = 0$ or $\theta_i = 0.2$, with $s = 50$
- Σ has constant correlation ρ
- $H_i : \theta_i \leq 0$ vs. $H_i' : \theta_i > 0$
- $T_{n,i}$ is the usual t-statistic

Methods considered:

- **(BH)** Benjamini and Hochberg (1995)
- **(STS)** Storey et al. (2004) with $\lambda = 0.5$
- **(BKY)** Benjamini et al. (2006)
- **(Boot)** Bootstrap method

Criteria:

- Empirical FDR (nominal level $\alpha = 10\%$)
- Average number of true rejections

Results I

	$\rho = 0$				$\rho = 0.9$			
	BH	STS	BKY	Boot	BH	STS	BKY	Boot
All $\theta_i = 0$								
Control	10.0	10.3	9.1	10.0	4.8	32.6	4.4	9.8
Rejected	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ten $\theta_i = 0.2$								
Control	7.6	9.5	7.3	7.3	5.0	26.5	5.8	10.0
Rejected	3.4	3.8	3.4	3.4	3.7	4.5	3.7	6.0
Twenty five $\theta_i = 0.2$								
Control	5.0	9.5	6.2	6.7	3.9	18.3	7.1	9.5
Rejected	13.2	17.4	14.5	14.9	12.6	14.2	12.7	16.6
All $\theta_i = 0.2$								
Control	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Rejected	34.8	49.7	44.9	48.2	32.1	47.3	32.1	36.4

Set-Up II

Data generating process and testing problem:

- I.i.d. random vectors from $N(\theta, \Sigma)$, with $n = 100$
- Three $\theta_i = 0$ and one $\theta_i = 0.2$ or 20, with $s = 4$
- Σ is a random correlation matrix: take 1,000 draws
- $H_i : \theta_i \leq 0$ vs. $H_i' : \theta_i > 0$
- $T_{n,i}$ is the usual t-statistic

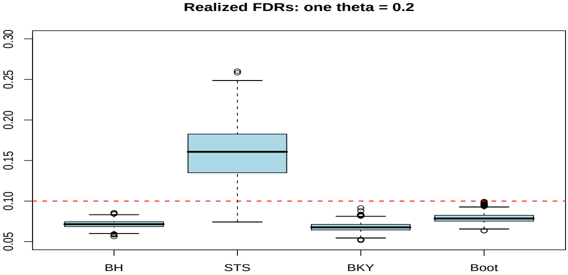
Methods considered:

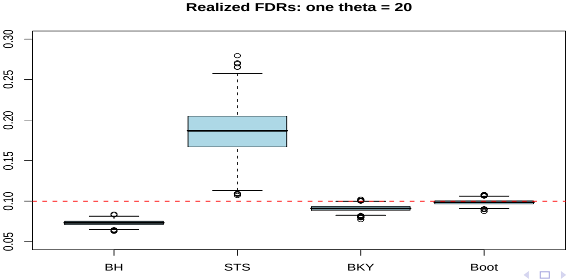
- **(BH)** Benjamini and Hochberg (1995)
- **(STS)** Storey et al. (2004) with $\lambda = 0.5$
- **(BKY)** Benjamini et al. (2006)
- **(Boot)** Bootstrap method

Criteria:

- Boxplot of empirical FDRs (nominal level $\alpha = 10\%$)

Results II


Realized FDRs: one theta = 0.2
0.05
0.10
0.15
0.20
0.25
0.30
BH
STS
BKY
Boot


Realized FDRs: one theta = 20
0.05
0.10
0.15
0.20
0.25
0.30
BH
STS
BKY
Boot

Outline

Conclusions

Methodology:

- Bootstrap method implicitly accounts for the dependence structure of the test statistics
- Extended the approach of Troendle (2000) to non-normal data

Conclusions

Methodology:

- Bootstrap method implicitly accounts for the dependence structure of the test statistics
- Extended the approach of Troendle (2000) to non-normal data

Advantages:

- Appears more powerful than current competitors
- At least compared to those, that are also robust against various dependence structures

Conclusions

Methodology:

- Bootstrap method implicitly accounts for the dependence structure of the test statistics
- Extended the approach of Troendle (2000) to non-normal data

Advantages:

- Appears more powerful than current competitors
- At least compared to those, that are also robust against various dependence structures

Disadvantage:

- Computationally more expensive than methods based on the individual p-values
- Should be considered negligible this day and age

Outline

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1):289–300.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4):1165–1188.

Storey, J. D., Taylor, J. E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society, Series B*, 66(1):187–205.

Troendle, J. F. (2000). Stepwise normal theory test procedures controlling the false discovery rate. *Journal of Statistical Planning and Inference*, 84(1):139–158.