Generalized Error Control in Multiple Hypothesis Testing

March, 2006

Joseph P. Romano, Stanford University

http://www-stat.stanford.edu/∼/romano/welcome.html

romano@stanford.edu

Collaborators: E. Lehmann, J. Shaffer, A. Shaikh, M. Wolf

The Basic Setup

Observe data $X = (X_1, \ldots, X_n)$ from P.

Test hypotheses $H_1, \ldots, H_s$: $H_j \equiv P \in \omega_j$

Let $I = I(P) \subset \{1, \ldots, s\}$ denote the indices of the set of true hypotheses: $j \in I$ if and only $P \in \omega_j$. The familywise error rate (FWE_P) is the probability under P that any H_j with $j \in I$ is rejected.

Require $\text{FWE}_P \leq \alpha \quad \forall P$.

Suppose H_j is rejected for large values of $T_{n,j}$, or small p-value $\hat{p}_j$.

Starting point: Bonferroni Given p-value $\hat{p}_j$ for testing H_j, order them as

$$\hat{p}_{(1)} \leq \cdots \leq \hat{p}_{(s)}$$

with corresponding $H_{(1)}, \ldots, H_{(s)}$. Reject any H_i if $\hat{p}_i \leq \alpha/s$. This controls FWE_P.

Holm's Stepdown Method: Reject $H_{(1)}, \ldots, H_{(j)}$ if

$$\hat{p}_{(1)} \leq \frac{\alpha}{s}, \ldots, \hat{p}_{(j)} \leq \frac{\alpha}{s-j+1} \, , \tag{1}$$

using the largest such j in (**??**). This is an example of a *stepdown* method. A *stepup* method starts with the largest p-value. While a big improvement over Bonferroni, still can be conservative.

Directions for Improving Holm

I. Constructing optimal procedures. Some success in parametric models; see Lehmann, Romano, Shaffer (2005).

- Key montonicity restriction. A decision rule D based on p-values $\hat{p} = (\hat{p}_1, \ldots, \hat{p}_s)$ states for each possible value $\hat{p}$ the subset $I = I_{\hat{p}}$ of $\{1, \ldots, s\}$ of values i for which the hypothesis H_i is rejected. A decision rule D is *monotone* if

$$\hat{q}_i \leq \hat{p}_i \quad \text{for} \quad i \in I_{\hat{p}} \quad \text{but} \quad \hat{p}_i < \hat{q}_i \quad \text{for} \quad i \notin I_{\hat{p}}$$

implies that $I_{\hat{p}} = I_{\hat{q}}$.

- Resulting optimal procedures are stepwise procedures!

II. Incorporating or estimating the dependence structure of p-values. This is the approach taken in Westfall and Young (1993), *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment.* However, a crucial assumption is *subset pivotality*. This assumption can be removed.

III. Relax control of the FWE. Given a multiple testing decision rule, let $F = \#$ false rejections, $R = \#$ rejections. Define the *false discovery proportion* (FDP) as F/R (defined to be 0 if $R = 0$).

As a measure of error control, Benjamini & Hochberg (1995) the *false discovery rate* defined by

(i) $E(FDP)$.

We'll also consider

(ii) k-FWE: the probability that $F \geq k$.

(iii) Given a value γ, $P\{FDP > \gamma\}$.

Given p-values for individual tests, stepdown methods exist for controlling these at level α with no assumptions about the dependence structure of the p-values.

IV. Combine II and III

II. A general construction of stepdown tests under weak assumptions

Let P be the true probability, $P \in \Omega$.

H_i specified by $\omega_i \subset \Omega$.

$i \in I(P)$ if and only $P \in \omega_i$.

For $K \subset \{1, \ldots, s\}$, let H_K denote the intersection hypothesis that all H_i with $i \in K$ are true; that is, H_K is equivalent to $P \in \bigcap_{i \in K} \omega_i$.

Let

$$T_{n,r_1} \geq T_{n,r_2} \geq \cdots \geq T_{n,r_s}$$

denote the observed ordered test statistics, and $H_{(1)}$, $H_{(2)}, \ldots, H_{(s)}$ the corresponding hypotheses.

Motivating Example: Correlations

$X_1, \ldots, X_n$ are i.i.d. random vectors in $\mathbb{R}^d$, with $X_i = (X_{i,1}, \ldots, X_{i,d})$.

Assume $E|X_{i,j}|^2 < \infty$ and $Var(X_{i,j}) > 0$, so that the correlation between $X_{1,i}$ and $X_{1,j}$, namely $\rho_{i,j}$ is well-defined.

$H_{i,j}: \ \rho_{i,j} = 0$, $(s = \binom{d}{2})$

Let $T_{n,i,j}$ = sample correlation between variables i and j. (Note we are indexing hypotheses and test statistics by 2 indices i and j.)

By Aitken (1969, 71), if $d = 3$, $H_{1,2}$ and $H_{1,3}$ are true but $H_{2,3}$ is false, the limiting distribution of $n^{1/2}(T_{n,1,2}, T_{n,1,3})$ is biv. normal: means 0, variances 1, and correlation $\rho_{2,3}$. Subset pivotality fails, as noted by WY (1993).

A stepdown procedure begins with the most significant test statistic. First, test the joint null hypothesis $H_{\{1,\ldots,s\}}$ that all hypotheses are true. This hypothesis is rejected if T_{n,r_1} is large. If it is not large, accept all hypotheses; otherwise, reject the hypothesis corresponding to the largest test statistic. Once a hypothesis is rejected, remove it and test the remaining hypotheses by rejecting for large values of the maximum of the remaining test statistics, and so on.

Let H_K denote the *intersection* hypothesis that all H_i with $i \in K$ are true.

Suppose critical values $\hat{c}_{n,K}(1-\alpha)$ are used for testing the intersection hypothesis H_K (possibly data-dependent).

Procedure 1 *[Generic Stepdown Method]*

1. Let $K_1 = \{1, \dots, s\}$. If $T_{n,r_1} \le \hat{c}_{n,K_1}(1-\alpha)$, then accept all hypotheses and stop; otherwise, reject $H_{(1)}$ and continue.

2. Let K_2 be the indices of the hypotheses not previously rejected. If $T_{n,r_2} \le \hat{c}_{n,K_2}(1-\alpha)$, then accept all remaining hypotheses and stop; otherwise, reject $H_{(2)}$ and continue.

$\vdots$

j. Let K_j be the indices of the hypotheses not previously rejected. If $T_{n,r_j} \le \hat{c}_{n,K_j}(1-\alpha)$, then accept all remaining hypotheses and stop; otherwise, reject $H_{(j)}$ and continue.

$\vdots$

s. If $T_{n,s} \le \hat{c}_{n,K_s}(1-\alpha)$, then accept $H_{(s)}$; otherwise, reject $H_{(s)}$.

Problem: how to construct the $\hat{c}_{n,K}(1-\alpha)$ so that the FWE is controlled.

Idea: Reduce the multiple testing problem of controlling the FWE to that of constructing single tests that control the probability of a Type 1 error.

A resampling construction. Assume H_i specifies by $\{P : \ \theta_i(P) = 0\}$ for some real-valued θ_i. Let $\hat{\theta}_{n,i}$ be an estimate of θ_i. Let $T_{n,i} = \tau_n |\hat{\theta}_{n,i}|$ for some nonrandom sequence $\tau_n \to \infty$; e.g., $\tau_n = n^{1/2}$.

First, consider testing H_K. Let $b_{n,K}(1-\alpha, P)$ denote a $1-\alpha$ quantile under P of $\max\{\tau_n |\hat{\theta}_{n,i} - \theta_i(P)| : \ i \in K\}$.

Let $\hat{Q}_n$ be some estimate of P. Then, a nominal $1-\alpha$ level bootstrap confidence region for the subset of parameters $\{\theta_i(P) : \ i \in K\}$ is given by

$$\{(\theta_i : \ i \in K) : \max_{i \in K} \tau_n |\hat{\theta}_{n,i} - \theta_i|$$

$$\leq b_{n,K}(1-\alpha, \hat{Q}_n)\} \ .$$

For those unfamiliar with the bootstrap, $b_{n,K}(1-\alpha, \hat{Q}_n)$ approximated by Monte Carlo:

For $b = 1, \dots B$, let $X^*(b)$ be a sample drawn from $\hat{Q}_n$. Based on $X^*(b)$, compute estimates $\hat{\theta}^*_{n,i}(b)$. Let

$$m_b = \max_{i \in K} \tau_n |\hat{\theta}^*_{n,i}(b) - \hat{\theta}_{n,i}|$$

A $1-\alpha$ quantile of the empirical distribution of the B values $m_1, \dots, m_B$ approximates $b_{n,K}(1-\alpha, \hat{Q}_n)$.

- Same set of resamples for any K.

A value of 0 for $\theta_i(P)$ falls outside the region iff $\tau_n|\hat{\theta}_{n,i}| > b_{n,K}(1-\alpha, \hat{Q}_n)$.

Usual duality suggests using

$$\hat{c}_{n,K}(1-\alpha) = b_{n,K}(1-\alpha, \hat{Q}_n) \ . \tag{2}$$

In words, begin by constructing a joint region for all θ_i and reject any hypothesis whose corresponding θ_i doesn't include 0.

Throw out the rejected hypotheses. Construct a joint confidence region for the remaining hypotheses by the same recipe.

Keep rejecting hypotheses until no more are rejected.

Theorem 1 *For Procedure* **??** *based on critical values* $\hat{c}_{n,K}(1-\alpha)$ *satisfying the monotonicity requirement: for any* $K \supset I(P)$,

$$\hat{c}_{n,K}(1-\alpha) \geq \hat{c}_{n,I(P)}(1-\alpha) \ . \tag{3}$$

(i) Then, $FWE_P \leq$

$$P\{\max(T_{n,j} : \ j \in I(P)) > \hat{c}_{n,I(P)}(1-\alpha)\} \ .$$

(ii) If testing the intersection hypothesis H_K *using* $\hat{c}_{n,K}(1-\alpha)$ *is level* α *when* $K = I(P)$ *then* $FWE_P \leq \alpha$.

(iii) If $\limsup_n \alpha_n(P) \leq \alpha$, *then* $\limsup_n FWE_P \to \alpha$ *as* $n \to \infty$.

(iv) If $\alpha_n(P) = \alpha + O(\epsilon_n)$, *then*

$$FWE_P = \alpha + O(\epsilon_n) \ .$$

Under monotonicity (**??**), the multiplicity problem is effectively reduced to testing a single intersection hypothesis at a time. The problem is to construct intersection tests whose critical values are monotone and asymptotically controls type 1 error.

The previous concrete bootstrap construction *always* satisfies the monotonicity requirement. Since, when $K = I(P)$, the error probability is controlled (at least asymptotically), bootstrap consistency theorems follow.

Generalizes to:

• more general hypotheses

• other resampling schemes: permutations, subsampling, moving blocks bootstrap. etc/

Eg 1 *[Multivariate Mean]* Assume $X_i = (X_{i,1}, \ldots, X_{i,s})$ are n i.i.d. random vectors with $E(|X_i|^2) < \infty$ and mean vector $\mu = (\mu_1, \ldots, \mu_s)$.

H_i specifies $\mu_i = 0$ and $T_{n,i} = n^{-1/2} |\sum_{j=1}^n X_{j,i}|$. Asymptotics follow from Theorem **??**.

Or, consider the studentized test statistic $t_{n,i} = T_{n,i}/S_{n,i}$, where $S_{n,i}^2$ is the sample variance of the ith components of the data.

Note: The same set of resamples can be used for all steps, i.e., for all K when testing H_K. ■

Eg 2 *[Testing Correlations]* Suppose $X_1, \ldots, X_n$ are i.i.d. random vectors in $\mathbb{R}^d$, so that $X_i = (X_{i,1}, \ldots, X_{i,d})$. Assume $E|X_{i,j}|^2 < \infty$ and $Var(X_{i,j}) > 0$.

$H_{i,j}$ specifies $\rho_{i,j} = 0$. Let $T_{n,i,j}$ = sample correlation between variables i and j.

The conditions for the bootstrap hold because correlations are smooth functions of means. ■

Eg 3 *[s-variate 2-sample Problem]* $Y_1, \cdots, Y_{n_Y}$ i.i.d. P_Y,

$Z_1, \cdots, Z_{n_Z}$ i.i.d. P_Z.

P_Y and P_Z are distributions on $\mathbf{R}^s$, with jth components denoted $P_{Y,j}$ and $P_{Z,j}$. Assume H_j implies $P_{Y,j} = P_{Z,j}$. Permutation tests apply and yield exact control.

Simulations support

- good control of the FWE in finite samples
- increase in "power" over Holm. For example, for s in the range 10–40, the stepdown method rejects between 20% and 50% more false hypotheses.

Extends to time series models using block resampling methods.

Also applies if $s = \infty$ (applications to underidentified econometric models).

Balance and Error Allocation May be desirable to have $P\{\text{reject } H_i\}$ independent of i.

This can be achieved by using studentized statistics, or p-values.

Of course, one can use the bootstrap to convert each $T_{n,i}$ into a p-value $\hat{p}_{n,i}$ and then apply the basic algorithm to $T'_{n,i} = -\hat{p}_{n,i}$. This would involve a double bootstrap. Actually, the bootstrap can be used to achieve balance automatically, by a generalization of the basic algorithm, and by doing only a single bootstrap.

$X \sim P$. Test $H_i: \ P \in \omega_i$, $i = 1, \ldots, s$

III. Stepdown and Stepup Methods Controlling

• k-FWE $= P\{F \geq k\}$, where $F = \#$ *false* rejections

• $P\{FDP > \gamma\}$, where $FDP = F/(R \vee 1)$, where

$R =$ number of total rejections

• $FDR = E(FDP)$

Given p-values $\hat{p}_i$ satisfying

$$P\{\hat{p}_i \leq u\} \leq u \quad , P \in \omega_i\}$$

Order them as

$$\hat{p}_{(1)} \leq \cdots \leq \hat{p}_{(s)}$$

Stepdown Methods Let $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_s$ be constants. Start with most significant $\hat{p}_i$. If $\hat{p}_{(1)} > \alpha_1$, reject no null hypotheses. Otherwise, if

$$\hat{p}_{(1)} \leq \alpha_1, \ldots, \hat{p}_{(r)} \leq \alpha_r \ , \tag{4}$$

reject hypotheses $H_{(1)}, \ldots, H_{(r)}$ where the largest r satisfying (**??**) is used.

Stepup Methods If $\hat{p}_{(s)} \leq \alpha_s$, then reject all null hypotheses; otherwise, reject hypotheses $H_{(1)}, \ldots, H_{(r)}$ where r is the smallest index satisfying

$$\hat{p}_{(s)} > \alpha_s, \ldots, \hat{p}_{(r+1)} > \alpha_{r+1} \ . \tag{5}$$

If, for all r, $\hat{p}_{(r)} > \alpha_r$, then reject no hypotheses.

Benjamini-Hochberg *stepup* FDR controlling procedure: uses

$$\alpha_i = i\alpha/s$$

under *independence* of p-values.

Benjamini-Yekutieli: If you divide each α_i by

$$C_s = \sum_{j=1}^{s} \frac{1}{j}$$

the FDR is controlled with no dependence assumptions on p-values.

Goal: Derive stepwise procedures that control k-FWE and FDP under no assumptions on p-values.

Theorem 2 *(Generalization of Bonferroni) For testing $H_i : P \in \omega_i$, $i = 1, \ldots, s$, suppose $\hat{p}_i$ satisfies*

$$P\{\hat{p}_i \le u\} \le u \quad \text{any } u \in (0,1),\ P \in \omega_i\ .$$

(i) Consider the procedure that rejects any H_i for which $\hat{p}_i \le k\alpha/s$; this controls the k-FWE, i.e. for all P,

$$P\{\text{reject} \ge k \text{ or more true } H_i\} \le \alpha\ . \tag{6}$$

So, if each of the hypotheses is tested at level $k\alpha/s$, then the k-FWE is controlled.

*(ii) The constant $k\alpha/s$ is sharp in the sense that there exists a joint distribution for $(\hat{p}_1, \ldots, \hat{p}_s)$ for which equality holds in (**??**).*

Proof of (i) Fix any P and suppose H_i with $i \in I(P)$ are true and the remainder false. Let N be the number of false rejections. By Markov's inequality,

$$P\{N \geq k\} \leq \frac{E(N)}{k} =$$

$$\frac{E\left[\sum_{i \in I(P)} I\{\hat{p}_i \leq k\alpha/s\}\right]}{k} =$$

$$\sum_{i \in I(P)} \frac{P\{\hat{p}_i \leq k\alpha/s\}}{k}$$

$$\leq \sum_{i \in I(P)} \frac{k\alpha/s}{k} = |I(P)|\frac{\alpha}{s} \leq \alpha \ .$$

Theorem 3 *(Generalization of Holm) Let* $\alpha_i = k\alpha/s$ *if* $i \leq k$ *and*

$$\frac{k\alpha}{s+k-i} \quad \text{if } i > k \ . \tag{7}$$

The stepdown procedure with above choice of α_i *controls the k-FWE.*

Also, it is not possible to increase any α_i *without violating control of the k-FWE, i.e. there exists for each* $i \geq k$ *a joint distribution of the p-values for which*

$$P\left\{\bigcap_{j=1}^{i}\{\hat{p}_{(j)}\} \leq \alpha_i\right\} = \alpha \ .$$

Comparison with Augmentation procedures of vanderLaan, Dudoit and Pollard (2004)

Idea: To control the k-FWER, use a 1-FWER controlling procedure and then reject $k-1$ more hypotheses.

eg. Use Holm for the 1-FWER procedure: $\hat{p}_{(1)}$ is compared with α/s. Our procedure: $\hat{p}_{(1)}$ is compared with $k\alpha/s$.

Augmentation doesn't take full advantage of measure of error control.

Plus, can reject hypotheses with large p-values.

Control of the FDP

Recall $F = \#$ false rejections, $R = \#$ rejections. Define the *false discovery proportion* (FDP) as F/R (defined as 0 if $R = 0$). Given a value γ, require

$$P\{FDP > \gamma\} \leq \alpha$$

Basic idea: At step i, having rejected $i-1$ hypotheses, we want to guarantee $F/i \leq \gamma$, i.e. $F \leq \lfloor \gamma i \rfloor$, where $\lfloor x \rfloor$ is the greatest integer $\leq x$. So, if $k = \lfloor \gamma i \rfloor + 1$, then $F \geq k$ should have probability no greater than α; that is, we must control the number of false rejections to be $\leq k$. Therefore, we use the stepdown constant α_i with this choice of k (which now depends on i).

$$\alpha_i = \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{s + \lfloor \gamma i \rfloor + 1 - i} \ . \tag{8}$$

Theorem 4 *Under weak dependence assumptions, the stepdown method with these α_i controls the FDP.*

e.g. the family of distributions is positively dependent and is characterized by the multivariate positive of order two condition. (Sarkar, 1998)

Theorem 5 *Consider testing s null hypotheses, with $|I|$ of them true. Let $\hat{q}_{(1)} \leq \cdots \leq \hat{q}_{(|I|)}$ denote their corresponding ordered p-values. Set*
$M = \min(\lfloor \gamma s \rfloor + 1, |I|)$.

*(i) For the stepdown procedure with α_i given by (**??**),*

$$P\{FDP > \gamma\} \leq P\{\bigcup_{i=1}^{M} \{\hat{q}_{(i)} \leq \frac{i\alpha}{|I|}\}\} \ .$$

(ii) Therefore, if the joint distribution of the p-values of the true null hypotheses satisfies Simes inequality; that

is,

$$P\left\{\bigcup_{i=1}^{|I|}\{\hat{q}_{(i)} \leq \frac{\alpha i}{|I|}\}\right\} \leq \alpha \ ,$$

then $P\{FDP > \gamma\} \leq \alpha.$

One can remove the dependence assumptions entirely.

How? Need to bound the union of events about ordered p-values. Next result generalizes Hommel (1983).

Lemma 1 *Suppose* $\hat{p}_1, \ldots, \hat{p}_t$ *are p-values, ordered as* $\hat{p}_{(1)} \le \cdots \le \hat{p}_{(t)}$*. Let* $0 = \beta_0 \le \beta_1 \le \beta_2 \le \cdots \le \beta_m \le 1$ *for some* $m \le t$*.*

(i). Then,

$$P\{\bigcup_{i=1}^{m} \{\hat{p}_{(i)} \le \beta_i\}\} \le t \sum_{i=1}^{m} (\beta_i - \beta_{i-1})/i \ . \qquad (9)$$

*(ii). As long as the right side of (**??**) is* ≤ 1*, the bound is sharp in the sense that there exists a joint distribution for the p-values for which the inequality is an equality.*

Let $C_j = \sum_{i=1}^{j} 1/i$.

Theorem 6 *Let*

$$\alpha_i' = \alpha_i / C_{\lfloor \gamma s + 1 \rfloor}$$

Using α_i' controls the FDP.

Note: $C_n \approx \log(n + .5) + \zeta_E$, with $\zeta_E \approx 0.5772156649$ is known as Euler's constant.

• Can be improved by replacing $C_{\lfloor \gamma s \rfloor + 1}$ with a much smaller $D = D(\gamma, \alpha, s)$. This procedure will also control the FDP but reject more hypotheses and hence be more powerful.

• $\exists$ general theorem for any α_i.

A convenient formula is not available for $D(\gamma, s)$, but it's simple to program. For example, if $s = 100$ and $\gamma = 0.1$, then $D = 2.0385$. In contrast, $C_{\lfloor \gamma s \rfloor + 1} = C_{11} = 3.0199$. In this case, the value of $|I|$ that maximizes S to yield D is 55.

We've tabled many values. Roughly, the resulting constants are generally about 50 percent larger than those of Theorem **??**.

Constants nearly optimal (to within a factor 1.06).

Stepup procedures Control of the k-FWE

Theorem 7 *Consider the stepup procedure with critical values* $\alpha'_i = \alpha\alpha_i / D_1(k, s)$*, where* D_1 *is easily computed.*

(i) Then, $k\text{-}FWER \leq \alpha$.

(ii) For any stepup procedure with critical values $\tilde{\alpha}_i = \alpha\alpha_i / D'$ *for some constant* D' *that satisfies* $k\text{-}FWER \leq \alpha$*, we have for each* i *that* $\alpha'_i \geq \tilde{\alpha}_i$.

Comparison of Stepdown and Stepup Take $\alpha_i = k\alpha/s$ if $i \le k$ and

$$\alpha_i = \frac{\alpha k}{(s+k-i)} \text{ if } i > k$$

A stepdown method using these α_i controls the k-FWE.

For same α_i applied in a steupup fashion, and large s, the constants α_i must be approximately halved to ensure control of the k-FWE This factor 2 is independent of k! If $k=1$, the value 2.1314 works for all s.

Stepup Control of the FDP Let $m(j) = \lfloor \gamma j \rfloor + 1$, where $\lfloor x \rfloor$ is the greatest integer $\leq x$, and $x \vee y = \max(x, y)$

Given $\alpha_1 \leq \cdots \leq \alpha_s$, define

$$S_2 = S_2(\gamma, s, |I|) = |I|\alpha_1 + \tag{10}$$

$$|I| \sum_{k \in J} \frac{\alpha_{s-|I|+k} - \alpha_{s-|I|+k-1}}{k \vee m(s - |I| + k)}$$

where J is the set of inidices k:

$$|I| - s + 1 < k \leq |I|, \; |I| \geq m(s - |I| + k)$$

Let

$$D_2 = D_2(\gamma, s) = \max_{1 \leq |I| \leq s} S_2(\gamma, s, |I|) \ . \tag{11}$$

Theorem 8 *Consider the stepup procedure with critical values* $\alpha_i'' = \alpha\alpha_i / D_2(\gamma, s)$, *where* $D_2(\gamma, s)$ *is easily computed.*

(i) Then, $P\{FDP > \gamma\} \leq \alpha$.

(ii) Moreover, for any stepup procedure with critical values of the form $\tilde{\alpha}_i = \alpha\alpha_i / D'$ *for some constant* D' *that controls the FDP, we have for each* i *that* $\alpha_i'' \geq \tilde{\alpha}_i$.

If $P\{FDP > \gamma\} \leq \alpha$, you are controlling the quantiles of the FDP.

eg. FDP control with $\alpha = 1/2$ means

$$\text{median}(FDP) \leq \gamma$$

Can compare Benjamini Yekutieli's method for control of $E(FDP)$ with control of $\text{med}(FDP)$.

See following graph: control of the median allows for larger critical values α_i for most values of i and nearly the same otherwise.

IV. Combine II and III Account for dependence structure while using generalized measures of error control. Related work by Korn, Troendle, McShane and Simin (2004), van der Laan, Dudoit and Pollard (2004).

Notation: If $\{y_i, \ i \in K\}$ is a collection of numbers indexed by a finite set K having $|K|$ elements. Then, for $k \leq |K|$, $k\text{-}\max_{i \in K}(y_i)$ is used to denote the kth largest value of the y_i with $i \in K$.

Suppose $H_i \equiv \theta_i(P) = 0$. let $K_0 = \{1, \ldots, s\}$. For any $K \subset K_0$, let $c_{n,K}(\alpha, k, P)$ denote an α-quantile of the distribution of $k\text{-}\max_{j \in K} |\hat{\theta}_{n,j} - \theta_j(P)|$ under P. Then, rejecting any H_j for which $|\hat{\theta}_{n,j}|$ exceeds $c_{n,K_0}(1 - \alpha, k, P)$ controls the k-FWE. Since P is unknown, replace P by $\hat{Q}_n$.

In the following algorithm designed for control of the k-FWE, suppose $\hat{c}_{n,K}(1 - \alpha, k)$ are used to test H_K, $K \subset \{1, \ldots, s\}$.

Algorithm 1 **Generic Stepdown Method For Control of the k-FWE** Let $A_1 = \{1, \ldots, s\}$.

1. If $\max_{i \in A_1} T_{n,i} \leq \hat{c}_{n,A_1}(1-\alpha, k)$, then accept all hypotheses and stop; otherwise, reject any H_i for which $T_{n,i} \leq \hat{c}_{n,A_1}(1-\alpha)$ and continue.

2. Let R_2 be the indices i of hypothesis H_i previously rejected, and let A_2 be the remaining hypotheses. If $R_2 < k$, stop. Otherwise, let

$$\hat{d}_{n,A_2}(1-\alpha, k) = \max\{c_{n,K}(1-\alpha, k) :$$

$$K = A_2 \bigcup I,\ I \subset R_2,\ |I| = k-1\}\ .$$

Then, reject any $T_{n,i}$ with $i \in A_2$ satisfying $T_{n,i} > \hat{d}_{n,A_2}(1-\alpha, k)$. If no further rejections, stop.

$\vdots$

j. Let R_j be the indices i of hypotheses previously rejected, and let A_j be the remaining hypotheses. Let

$$\hat{d}_{n,A_j}(1-\alpha,k) = \max\{c_{n,K}(1-\alpha,k) :$$

$$K = A_j \bigcup I,\ I \subset R_j,\ |I| = k-1\} \ .$$

Then, reject any $T_{n,i}$ with $i \in A_j$ satisfying $T_{n,i} > \hat{d}_{n,A_j}(1-\alpha,k)$. If no further rejections, stop.

$\vdots$

And so on.

Theorem 9 *Using above algorithm with critical values $\hat{c}_{n,K}(1-\alpha, k)$ satisfying*

$$\hat{c}_{n,K}(1-\alpha, k) \geq \hat{c}_{n,I(P)}(1-\alpha, k) \ ,$$

$$k\text{-}FWE_P \leq P\{k\text{-}\max(T_{n,j} : \ j \in I(P)) >$$
$$\hat{c}_{n,I(P)}(1-\alpha, k)\}$$

So, if last expression $\leq \alpha$, then k-$FWE_P \leq \alpha$.

• Resampling methods can be used to satisfy monotonicity requirement, and the last requirement typically holds at least asymptotically. Results for FDP as well.

A Bootstrap Construction Assume H_i specifies $\theta_i = 0$. A $1-\alpha$ level bootstrap joint confidence region for the subset of parameters $\{\theta_i(P) : \ i \in K\}$ is given by $(\theta_i : \ i \in K)$:

$$\max(|\hat{\theta}_{n,i} - \theta_i| : i \in K) \leq b_{n,K}(1-\alpha, k, \hat{Q}_n)$$

where $b_{n,K}$ is a bootstrap estimate of the distribution of k- $\max|\hat{\theta}_{n,i} - \theta_i|$. This region contains all $\theta_i \in K$, except for at most $k-1$ of them, with asymptotic prob $1-\alpha$. So a value of 0 for $\theta_i(P)$ falls outside the region if and only if $|\hat{\theta}_{n,i}| > b_{n,K}(1-\alpha, 1, \hat{Q}_n)$. So apply previous algorithm with $\hat{c}_{n,K}(1-\alpha, k) = b_{n,K}(1-\alpha, k, \hat{Q}_n)$.

Many open problems in multiple testing

- Optimality: two-sided hypotheses
- Type 3 or directional errors
- Asymptotics as $s \to \infty$
- Higher-order accuracy of proposed methods
- FDR-controlling methods that take into account dependency structure
- Stepup methods