Generalized Error Control in Multiple Hypothesis Testing Via Subsampling and Resampling

January 2008, New Orleans

Joseph P. Romano, Stanford University

http://www-stat.stanford.edu/~romano/joe.html

romano@stanford.edu

Recent Collaborators: Guo, Shaikh, Wolf

Virtually all empirical research seeks answers to multiple questions: how to counter effects of data snooping?

(i) Comparing educational achievement among states, leading to $\binom{50}{2} = 1225$ comparisons

(ii) Randomized trials with multiple outcomes, treatments, doses, subgroups, covariates, etc. (AIDS VaxGen vaccine)

(iii) Multiple Regression: which coefficients matter?

(iv) Comparing several strategies with a benchmark (White, 2000), any better? which ones?

(v) Genomics experiments with microarray data – multiple comparisons at the gene level

From the Ethical Guidelines for Statistical Practice at the A.S.A. website:

Recognize that any frequentist statistical test has a random chance of indicating significance when it is not really present. Running multiple tests on the same data set at the same stage of an analysis increases the chance of obtaining at least one invalid result. Selecting the one "significant" result from a multiplicity of parallel tests poses a grave risk of an incorrect conclusion. Failure to disclose the full extent of tests and their results in such a case would be highly misleading.

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.

Classical requirement: $\text{FWE}_P \leq \alpha \quad \forall P$.

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

Review of Subsampling for Single Inferences

$X_1, \ldots, X_n$ i.i.d. P.

Goal: inference for $\theta(P)$.

Suppose $\hat{\theta}_n$ is an estimator such that

$$\tau_n[\hat{\theta}_n - \theta(P)] \xrightarrow{L} J(\cdot, P) \tag{1}$$

under P, for some limit law $J(P)$.

Method: Fix a positive integer $b < n$ and let $Y_1, \ldots, Y_{N_n}$ be equal to the $N_n := \binom{n}{b}$ subsets of $\{X_1, \ldots, X_n\}$, ordered in any fashion.

Let $\hat{\theta}_{b,i}^{(a)}$ be equal to the statistic $\hat{\theta}_{n,i}$ evaluated at the data set Y_a, for $a = 1, \ldots, N_n$.

The subsampling estimator of $J(\cdot, P)$ is given by

$$\hat{J}_n(x) = \frac{1}{N_n} \sum_a I\{\tau_b[\hat{\theta}_b^{(a)} - \hat{\theta}_n] \leq x\} \ . \tag{2}$$

Theorem 1 (Basic Theorem) *Assume (1), $b/n \to 0$, $b \to \infty$, and $\tau_b/\tau_n \to 0$ as $n \to \infty$. If x is a continuity point of $J(\cdot, P)$, then*

$$\hat{J}_n(x) \xrightarrow{P} J(x, P) \ .$$

Why? $\hat{J}_n(x) \approx U_n(x) \equiv \frac{1}{N_n} \sum_a I\{\tau_b[\hat{\theta}_b^{(a)} - \theta(P)] \leq x\}$

But each $\tau_b[\hat{\theta}_b^{(a)} - \theta(P)]$ has exact distribution $J_b(\cdot, P)$ because $\hat{\theta}_b^{(a)}$ is the estimator computed from a *genuine* sample from P.

The empirical over all $\binom{n}{b}$ samples approximates $J_b(\cdot, P)$ (uniformly in P as well). But (1) implies J_b and J_n are close.

Starting point: Stepdown methods based on marginal p-values. Given p-value $\hat{p}_j$ for testing H_j, order them as

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

with corresponding $H_{(1)}, \ldots, H_{(s)}$.

Let $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_s$.

Method: Let j^* be the largest j: $\hat{p}_{(1)} \leq \alpha_1, \ldots, \hat{p}_{(j)} \leq \alpha_j$

and reject $H_{(1)}, \ldots, H_{(j^*)}$.

Bonferroni single-step: $\alpha_i = \alpha/s$ controls the FWE.

Holm stepdown: $\alpha_i = \alpha/(s - i + 1)$

While a big improvement over Bonferroni, still can be conservative.

Directions for Improving Holm

I. 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.* Also see Dudoit, Pollard and van der Laan (2004) and Romano and Wolf (2005).

II. 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 $R = 0$).

(i) *false discovery rate* (FDR) defined by $E(\text{FDP})$, popularized by Benjamini Hochberg (1995). Require $E(\text{FDP}) \leq \alpha$.

(ii) k-FWE: the probability that $F \geq k$. Require $P\{F \geq k\} \leq \alpha$.

(iii) Tail probability for FDP Given a value γ, require $P\{\text{FDP} > \gamma\} \leq \alpha$.

eg. FDP control with $\alpha = 1/2$ means $\text{med}(\text{FDP}) \leq \gamma$

(iv) $E(F) \leq \alpha$.

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; see Benjamini and Yekutieli (2001) and Romano and Shaikh (2006).

Here, we will combine I (incorporate dependence structure) and II (weaken error measure) to achieve greater power.

Goal: Derive stepwise procedures that control k-FWE and FDP which incorporate dependence structure among test statistics or p-values. Begin with k-FWE.

Theorem 2 *(Generalized Bonferroni) The method that rejects H_i if $\hat{p}_i \leq k\alpha/s$ controls the k-FWE.*

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

$$\alpha_i = \frac{k\alpha}{s+k-i} \quad \text{if } i > k \ . \tag{3}$$

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

Above results due to Hommel and Hoffman (1987) and elaborated on in Lehmann and Romano (2005).

The above results do not incorporate dependence structure. But we now argue it is vital to do so, especially for generalized error rates.

Under independence, one can improve the constant $k\alpha/s$ dramatically. Let

$$H_{k,s}(u) = \sum_{j=k}^{s} \binom{s}{j} u^j (1-u)^{s-j} \ . \tag{4}$$

Consider the (generalized Sidák) procedure that rejects any H_i whose corresponding p-value $\hat{p}_i$ is $\leq H_{k,s}^{-1}(\alpha)$.

This controls the k-FWE (Guo and Romano, 2007).

Further stepdown improvement: Let

$$\alpha_1 = \cdots = \alpha_k = H_{k,s}^{-1}(\alpha)$$

and, for $j > 0$,

$$\alpha_{k+j} = H_{k,s-j}^{-1}(\alpha) \ .$$

This controls the k-FWE.

How dramatic are these improvements? For $k = 1$, the ratio of critical values satisfies:

$$\lim_{s \to \infty} \frac{1 - (1 - \alpha)^{1/s}}{\alpha / s} \to \frac{-\log(1 - \alpha)}{\alpha} \ ,$$

which $= 1.026$ when $\alpha = 0.05$.

In general, if you use the cutoff $k\alpha/s$, then under independence,

$$k - \text{FWE} = O(\alpha^k) \qquad \text{as } \alpha \to 0, s \to \infty \ .$$

Table 1: Single step constants for k-FWE control with $s = 100$ and $\alpha = 0.05$

k	$A = k\alpha/s$	$B = C_{k,s}(\alpha)$	B/A
1	0.0005	0.00051	1.026
2	0.0010	0.00353	3.530
3	0.0015	0.00806	5.376
5	0.0025	0.01913	7.653
7	0.0035	0.03140	8.972
10	0.0050	0.05062	10.124

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).

Subsampling and appropriate bootstrap methods can handle this problem.

Burt Holland compiled data on the numerous papers in social science journals with sample correlation matrices accompanied by marginal test results. eg, in the Oct05 issue of *Academy of Management J.*, 8 out of 10 papers had them and none corrected for multiplicity.

Problem: how to construct the critical values at each step so that the k-FWE is controlled?

Idea: Reduce the multiple testing problem of controlling the k-FWE in a stepdown procedure to that of constructing single step procedures which control the probability of k or more false rejections.

Key features: Use of subsampling/resampling to capture dependence; provide error control and balance; no iterative resampling required.

Single-step methods based on inverting Balanced Simultaneous Confidence Regions

Fix k. Assume H_i specifies $\omega_i = \{P : \theta_i(P) = 0\}$. Let $\hat{\theta}_{n,i}$ be some estimate of $\theta_i(P)$. Tests of H_i (without regard to multiplicity) can be constructed by the usual duality between tests and confidence intervals, if one knows or can estimate the sampling distribution of $\hat{\theta}_{n,i} - \theta_i(P)$ under P. (Can also consider studentized "roots" and one-sided H_i.)

Let $H_{n,i}(\cdot, P)$ denote the c.d.f. of $\tau_n|\hat{\theta}_{n,i} - \theta_i(P)|$ under P, with inverse $H_{n,i}^{-1}(\cdot, P)$.

Assuming continuity of $H_{n,i}(\cdot, P)$, the confidence interval

$$\{\theta_i : \tau_n |\hat{\theta}_{n,i} - \theta_i| \leq H_{n,i}^{-1}(\gamma, P)\} \tag{5}$$

has coverage probability γ.

Problem: P unknown and would like to make a statement about the simultaneous coverage of the intervals.

To this end, let K denote an arbitrary subset of $\{1, \ldots, s\}$. We would like to make joint inferences for the parameters $\theta_i(P)$ simultaneously for $i \in K$.

Duality: Simultaneous confidence regions for $\{\theta_i(P) : i \in K\}$ of nominal level $1-\alpha$ can be used to construct tests of the hypotheses $H_i, i \in K$, by rejecting any H_i for which 0 is not included in the confidence interval for $\theta_i(P)$. In order to control the k-FWE, it is required to approximate the probability of

$$\{\tau_n|\hat{\theta}_{n,i} - \theta_i(P)| \leq H_{n,i}^{-1}(\gamma, P) \tag{6}$$

$$\text{for all but at most } (k-1) \text{ of the } i \in K\} \ .$$

But (6) can be rewritten as

$$\{H_{n,i}(\tau_n|\hat{\theta}_{n,i} - \theta_i(P)|, P) \leq \gamma \tag{7}$$

$$\text{for all but at most } (k-1) \text{ of the } i \in K\}$$

or

$$\{k\text{-}\max(H_{n,i}(\tau_n|\hat{\theta}_{n,i} - \theta_i(P)|, P), i \in K) \leq \gamma\} \ , \qquad (8)$$

where the k-max function is just the kth largest value. Then, take γ to be the $1-\alpha$ quantile of the distribution of

$$k\text{-}\max(H_{n,i}(\tau_n|\hat{\theta}_{n,i} - \theta_i(P)|, P), i \in K)$$

under P, which we denote by $L_{n,K}^{-1}(1-\alpha, k, P)$. This yields the joint confidence region

$$\{(\theta_i, i \in K) : \ \tau_n|\hat{\theta}_{n,i} - \theta_i| \leq H_{n,i}^{-1}(L_{n,K}^{-1}(1-\alpha, k, P), P)\} \ . \qquad (9)$$

(i) These regions covers all θ_i with $i \in K$, except for at most $k-1$ of them, with probability $1-\alpha$.

(ii) The intervals are balanced in the sense that the probability that $\theta_i(P)$ is covered does not depend on i.

(iii) Invert to get tests of $\theta_i = 0$.

(iv) But, P is unknown!

First solution: Bootstrap Replace P in (9) by an estimate $\hat{Q}_n$ to get

$$\{(\theta_i, i \in K) : \tau_n |\hat{\theta}_{n,i} - \theta_i| \leq H_{n,i}^{-1}(L_{n,K}^{-1}(1-\alpha, k, \hat{Q}_n), \hat{Q}_n)\} \, . \tag{10}$$

For $K \subset \{1, \ldots, s\}$, let $J_{n,K}(P)$ denote the joint distribution of $\{\tau_n[\hat{\theta}_{n,i} - \theta_i(P)],\ i \in K\}$. So, $J_{n,\{i\}}(P) = J_{n,i}(P)$ for a singleton subset $\{i\} \subset K$.

Assumption B1 $J_{n,\{1,\ldots,s\}}(P) \xrightarrow{L} J_{\{1,\ldots,s\}}(P)$.

Assumption B2 $J_i(P)$ has a continuous cdf for all i.

B1 implies that, for every $K \subset I(P)$, $L_{n,K}(k, P)$ has a continuous limiting distribution $L_K(k, P)$. Under an additional mild assumption, we can show that this limiting distribution is strictly increasing on its support, which will prove quite useful.

Assumption B3 The support of the limiting distribution $J_{\{1,\ldots,s\}}(P)$ is connected.

Assumption B3 is very weak and holds whenever the joint limiting distribution is multivariate Gaussian, as long as the diagonal entries of the covariance matrix are nonzero. The covariance matrix may even be singular (which happens in some simultaneous inference problems; e.g., pairwise comparisons of means). Finally, in order to show asymptotic validity of the bootstrap, we need a further assumption on the behavior of the estimator $\hat{Q}_n$ of P.

Assumption B4 For any metric ρ metrizing weak convergence on $\mathbf{R^s}$,

$$\rho\left(J_{n,\{1,\ldots,s\}}(P), J_{n,\{1,\ldots,s\}}(\hat{Q}_n)\right) \xrightarrow{P} 0 \ .$$

Theorem 4 *Suppose data is generated from P satisfying Assumptions B1–B3. Let $\hat{Q}_n$ be an estimator of P satisfying Assumption B4. Fix $K \subset \{1, \ldots, s\}$ and a positive integer k. Consider the joint confidence region given by (10), with the marginal interval $\hat{C}_{n,i}$ for $\theta_i(P)$ with $i \in K$ expressed as*

$$\hat{C}_{n,i} \equiv \hat{\theta}_{n,i} \pm \tau_n^{-1} H_{n,i}^{-1}(L_{n,K}^{-1}(1-\alpha, k, \hat{Q}_n), \hat{Q}_n) \ . \quad (11)$$

(i) For $i \in K$, the intervals $\hat{C}_{n,i}$, simultaneously cover all the corresponding true parameter values $\theta_i(P)$, except for at most $k-1$ of them, with asymptotic probability $1-\alpha$.

(ii) The intervals $\hat{C}_{n,i}$ are balanced in the sense that

$$\lim_{n \to \infty} P\{\theta_i(P) \in \hat{C}_{n,i}\} = \gamma, \text{ independent of } i \ , \quad (12)$$

where $\gamma = \gamma_K(1-\alpha, k, P)$ is the unique $1-\alpha$ quantile of the limiting distribution $L_K(k, P)$.

A General Subsampling Construction

We now detail the case of n i.i.d. observations $X_1, \ldots, X_n$ from P. The previous bootstrap estimators $H_{n,i}(\cdot, \hat{Q}_n)$ and $L_{n,K}(\cdot, k, \hat{Q}_n)$ are replaced by subsampling estimators as follows.

Fix a positive integer $b < n$ and let $Y_1, \ldots, Y_{N_n}$ be equal to the $N_n := \binom{n}{b}$ subsets of $\{X_1, \ldots, X_n\}$, ordered in any fashion.

Let $\hat{\theta}_{b,i}^{(a)}$ be equal to the statistic $\hat{\theta}_{n,i}$ evaluated at the data set Y_a, for $a = 1, \ldots, N_n$.

The subsampling estimator of $H_{n,i}(\cdot, P)$ is given by

$$\hat{H}_{n,i}(x) = \frac{1}{N_n} \sum_a I\{\tau_b |\hat{\theta}_{b,i}^{(a)} - \hat{\theta}_{n,i}| \leq x\} \ . \tag{13}$$

Also define

$$\hat{L}_{n,K}(x,k) = \frac{1}{N_n} \sum_a I\{k\text{-}\max(\hat{H}_{n,i}(\tau_b |\hat{\theta}_{b,i}^{(a)} - \hat{\theta}_{n,i}|) \leq x\} \ . \tag{14}$$

If we replace the bootstrap estimators by these subsampling estimators, we can prove a result analogous to Theorem 4, while removing the assumption B4.

Theorem 5 *Assume B1–B3. Fix* $K \subset \{1, \ldots, s\}$ *and a positive integer* k. *Let* $b \to \infty$, $b/n \to 0$ *and* $\tau_b/\tau_n \to 0$. *Consider the joint confidence region rectangle, with marginal intervals* $\tilde{C}_{n,i}$ *for* $\theta_i(P)$ *with* $i \in K$ *expressed as*

$$\tilde{C}_{n,i} \equiv \hat{\theta}_{n,i} \pm \tau_n^{-1} \hat{H}_{n,i}^{-1}(\hat{L}_{n,K}^{-1}(1-\alpha, k)) \ . \tag{15}$$

(i) For $i \in K$, *the intervals* $\tilde{C}_{n,i}$, *simultaneously cover all the corresponding true parameter values* $\theta_i(P)$, *except for at most* $k-1$ *of them, with asymptotic probability* $1-\alpha$.

(ii) The intervals $\tilde{C}_{n,i}$ are balanced in the sense that

$$\lim_{n\to\infty} P\{\theta_i(P) \in \tilde{C}_{n,i}\} = \gamma, \text{ independent of } i\ , \quad (16)$$

where $\gamma = \gamma_K(1-\alpha, k, P)$ is the unique $1-\alpha$ quantile of the limiting distribution $L_K(k, P)$.

Stepdown Control of the k-FWE

We now return to the general setup. Test statistics $T_{n,i}$ are available to test H_i (such as $\tau_n |\hat{\theta}_{n,i}|$). Given a single-step method, we will show how a stepdown improvement may be obtained. Suppose we have in mind critical values $\hat{c}_{n,K,i}(1-\alpha,k)$ which could be used to control the k-FWE at level α when testing the multiple hypotheses H_i with $i \in K$; that is, such a single step procedure would reject H_i if $T_{n,i} > c_{n,K,i}(1-\alpha,k)$.

eg, the subsampling critical values given by

$$\hat{c}_{n,K,i}(1-\alpha,k) = \hat{H}_{n,i}^{-1}(\hat{L}_{n,K}^{-1}(1-\alpha,k)) \ .$$

A stepdown method begins by first applying a single-step method, but then additional hypotheses may be rejected after this first stage by proceeding in a stepwise fashion, which we now describe. Begin by testing all null hypotheses $H_1, \ldots, H_s$. Any hypothesis H_i is rejected if $T_{n,i} > c_{n,\{1,\ldots,s\},i}(1-\alpha, k)$. If there are no rejections, then stop. If there are rejections, let A_2 be the set of hypotheses not yet rejected. Then, we compare $T_{n,i}$ for $i \in A_2$ with smaller critical values than used in the first stage, leading to the possibility of further rejections.

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

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

2. Let R_2 be the indices i of hypotheses H_i previously rejected, and let A_2 be the remaining hypotheses. If $|R_2| < k$, then stop. Otherwise, reject any H_i with $i \in A_2$ if $T_{n,i} > \hat{d}_{n,A_2,i}(1-\alpha, k)$, where

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

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

If there are no further rejections, stop.

⋮

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

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

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

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

⋮

And so on.

When $k = 1$, once a hypothesis is removed, it no longer enters into the algorithm.

More complex for $k > 1$. The reason is that we must acknowledge that when we consider a set of hypotheses not previously rejected, we may have gotten to that stage by rejecting true null hypotheses, but hopefully at most $k - 1$ of them. Since we do not know which of the hypotheses rejected thus far are true or false, we must maximize over subsets including some of those rejected, but at most $k - 1$ among the previously rejected ones.

Main point: If we can control the k-FWE at any stage of the algorithm, then the stepdown test will control the k-FWE.

In order to prove such an algorithm controls the k-FWE for suitable critical values $\hat{c}_{n,K,i}(1-\alpha, k)$, we assume monotonicity of the estimated critical values; that is, for any $K \supset I$,

$$\hat{c}_{n,K,i}(1-\alpha, k) \geq \hat{c}_{n,I,i}(1-\alpha, k)\ . \tag{17}$$

Under (17), it can be shown that k-FWE control of a stepdown procedure is reduced to that of a single-step method.

Theorem 6 *Consider Algorithm 1 with critical values $\hat{c}_{n,K,i}(1-\alpha,k)$ satisfying (17).*

(i) Then, k-$FWE_P \leq$

$$P\{T_{n,i} > \hat{c}_{n,I(P),i} \tag{18}$$

$$\text{for all but at most } k-1 \text{ of } i \in I(P)\} \ .$$

(ii) Therefore, if the critical values $\hat{c}_{n,I(P),i}$ control the k-FWE as a single-step procedure in the sense that the right side of (18) is $\leq \alpha$ (in finite samples or asymptotically), then k-$FWE_P \leq \alpha$ (in finite samples or asymptotically).

The monotonicity assumption (17) cannot be removed, as shown by Romano and Wolf (2005) even in the case $k = 1$; an analogous construction works for general k.

But, the general resampling and subsampling constructions **always** satisfy (17). Also applies to problems where randomization and permutations apply. A bunch of new results follow.

When testing multiple hypotheses, it seems natural that the critical values should satisfy the monotonicity condition, because larger critical values should be required when testing more hypotheses rather than a smaller subset of them.

Outside some parametric models, application of the Generic Stepdown Method can be computationally intensive, so we will also consider the following more streamlined algorithm. The basic idea is that at any stage, when testing whether or not to include further rejections, we need only look at the hypotheses not previously rejected together with the $k-1$ hypotheses that are least significant among those previously rejected. So, we avoid maximizing over all subsets of size $k-1$ of previously rejected hypotheses and just look at the least significant $k-1$ rejections. The arguments for such a procedure will be asymptotic.

If $k = 1$, at step j, no need to consider previously rejected hypotheses.

For $k > 1$, at step j, having made R_j rejections, one has to evaluate $\binom{R_j}{k-1}$ quantiles over which one maximizes.

Asymptotically, one need only consider the subset of $k - 1$ least significant hypotheses rejected.

Operative Method: Fix $N_{max} = 50$, say, and let M be the largest integer for which $\binom{M}{k-1} \leq N_{max}$. Consider at most the M most "recently" rejected hypotheses and maximize over subsets corresponding to those M hypotheses together with those not already rejected.

Generalizations:

- more general hypotheses
- other resampling schemes:

(i) permutations (can lead to finite sample control)

(ii) moving blocks and stationary bootstrap for dependent data

(iii) subsampling for dependent data (under weakest conditions)

- Also applies if $s = \infty$. Applications to partially identified econometric models (with Shaikh).

Simulations support

- good control of the k-FWE in finite samples
- increase in "power" over generalized Holm or methods based on marginal pvalues. For example, if $k = 1$, for s in the range 10–40, the stepdown method rejects between 20% and 50% more false hypotheses than Holm. Not surprisingly, increasing k rejects many more hypotheses.

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 a stepdown procedure such that at step i, we apply a k-FWE controlling procedure, where

$$k = k(i, \gamma) = \lfloor \gamma i \rfloor + 1 \ .$$

eg. Apply generalized Bonferroni/Holm constants. Leads to a stepdown method based on marginal p-values with critical values $\alpha_i = \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{s + \lfloor \gamma i \rfloor + 1 - i}$.

Theorem 7 *(Lehmann and Romano, 2005) 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)

By same reasoning, apply the bootstrap method to control the k-FWE at step i, where
$k = k(i, \gamma) = \lfloor \gamma i \rfloor + 1$.

Simulation results and applications presented in Romano and Wolf (Annals 07) and Romano, Shaikh and Wolf (Econometric Theory 07).

For FDR control, working paper available (RSW).

Conclusion: Asymptotic Theory and Simulations support the value of methods which account for dependence based on weaker measures of error control.

Caveats: Asymptotics, Increase in number of true rejections.