Multiple Testing Procedures with Applications to Genomics

Part II. Methodology

Sandrine Dudoit

PB HLTH 240D – Spring 2007

Outline

These lecture notes are based on the forthcoming book by Dudoit and van der Laan (2007).

Related articles and tech reports may be downloaded from Sandrine Dudoit's website

www.stat.berkeley.edu/~sandrine

and Mark van der Laan's website

www.stat.berkeley.edu/~laan.

Outline: Part I. Motivation and Overview

- Multiple Hypothesis Testing Problems in Genomics.
- Multiple Hypothesis Testing Framework.
- Data Generating Distribution.
- Parameters.
- Null and Alternative Hypotheses.
- Test Statistics.
- Multiple Testing Procedures.
- Rejection Regions.

Outline: Part I. Motivation and Overview

- Errors in Multiple Hypothesis Testing: Type I, Type II, and Type III errors.
- Type I Error Rates.
- Power.
- Unadjusted and Adjusted p-Values.
- Examples of Multiple Testing Procedures.

[Dudoit and van der Laan (2007, Chapter 1)]

Outline: Part II. Methodology

- Test Statistics Null Distribution [Chapter 2].
- Single-Step Multiple Testing Procedures for Controlling General Type I Error Rates, $\Theta(F_{V_n})$ [Chapter 4].
- Step-Down Multiple Testing Procedures for Controlling the Family-Wise Error Rate [Chapter 5].
- Augmentation Multiple Testing Procedures for Controlling Generalized Tail Probability Error Rates [Chapter 6].
- Resampling-Based Empirical Bayes Multiple Testing Procedures for Controlling Generalized Tail Probability Error Rates [Chapter 7].
- Appendix: Miscellaneous Mathematical and Statistical Results.

Outline: Part III. Applications to Genomics and Software Implementation

- Identification of Differentially Expressed and Co-Expressed Genes in High-Throughput Gene Expression Experiments [Chapter 9].
- Multiple Tests of Association with Biological Annotation Metadata [Chapter 10].
- HIV-1 Sequence Variation and Viral Replication Capacity [Chapter 11].
- Genetic Mapping of Complex Human Traits Using Single Nucleotide Polymorphisms: The ObeLinks Project [Chapter 12].
- Software Implementation [Chapter 13].

Test Statistics Null Distribution

One of the main tasks in specifying a multiple testing procedure is to derive rejection regions for the test statistics such that Type I errors are probabilistically controlled at a user-supplied level.

However, one is immediately faced with the problem that the distribution of the test statistics is usually unknown.

In practice, the test statistics true distribution $Q_n = Q_n(P)$ is replaced by a null distribution Q_0 (or estimator thereof, Q_{0n}) in order to derive rejection regions and resulting adjusted p-values.

The choice of a proper null distribution is crucial in order to ensure that (finite sample or asymptotic) control of the Type I error rate under the assumed null distribution does indeed provide the desired control under the true distribution.

Resampling procedures (e.g., bootstrap and permutation) are particularly useful in this context.

Test Statistics Null Distribution

Given a random M-vector $Z = (Z(m) : m = 1, \ldots, M)$, with joint distribution Q, and an M-vector of cut-offs $c = (c(m) : m = 1, \ldots, M) \in I\!R^M$, denote the numbers of rejected hypotheses and Type I errors by

$$R(c|Q) \equiv \sum_{m=1}^{M} \mathrm{I}\left(Z(m) > c(m)\right) \text{ and } V(c|Q) \equiv \sum_{m \in \mathcal{H}_0} \mathrm{I}\left(Z(m) > c(m)\right), \tag{1}$$

respectively.

For a given cut-off vector c, adopt the following shorthand notation, for the special cases where Q corresponds to the test statistics true distribution Q_n and null distribution Q_0,

$$\begin{aligned} R_n &\equiv R(c|Q_n), & R_0 &\equiv R(c|Q_0), \\ V_n &\equiv V(c|Q_n), & V_0 &\equiv V(c|Q_0). \end{aligned} \tag{2}$$

Test Statistics Null Distribution

A multiple testing procedure (MTP) $\mathcal{R}_n$ is said to control the Type I error rate $\Theta(F_{V_n,R_n})$, under the test statistics true distribution Q_n, at actual level $\alpha \in (0,1)$, if

$$\begin{aligned} \Theta(F_{V_n,R_n}) &\leq \alpha \qquad \text{[finite sample control]} \\ \limsup_{n\to\infty} \Theta(F_{V_n,R_n}) &\leq \alpha \qquad \text{[asymptotic control].} \end{aligned} \tag{3}$$

Test Statistics Null Distribution

Note that the actual Type I error rate $\Theta(F_{V_n,R_n})$ of a multiple testing procedure typically differs from its nominal level α, i.e., the level at which it claims to control Type I errors.

A testing procedure is said to be conservative/anti-conservative if the nominal Type I error level α is greater/less than the actual Type I error rate, that is,

$$\begin{array}{lrcl} \text{Conservative} & \Theta(F_{V_n,R_n}) & < & \alpha \\ \text{Anti-conservative} & \Theta(F_{V_n,R_n}) & > & \alpha. \end{array} \tag{4}$$

Discrepancies between actual and nominal Type I error levels can be attributed to a number of factors, including the choice of a test statistics null distribution Q_0 and the type of rejection regions for a given choice of Q_0.

Test Statistics Null Distribution

For proper control, the Type I error rate under the null distribution Q_0 must dominate the Type I error rate under the true distribution Q_n.

That is, the null distribution Q_0 must satisfy

$$\begin{aligned} \Theta(F_{V_n,R_n}) &\leq \Theta(F_{V_0,R_0}) \qquad \text{[finite sample control]} \\ \limsup_{n\to\infty} \Theta(F_{V_n,R_n}) &\leq \Theta(F_{V_0,R_0}) \qquad \text{[asymptotic control]}, \end{aligned} \tag{5}$$

where V_0 and R_0 denote, respectively, the numbers of Type I errors and rejected hypotheses under the null distribution Q_0.

Test Statistics Null Distribution

Controlling the number of Type I errors V_n, i.e., Type I error rates of the form $\Theta(F_{V_n})$, can be achieved by the three-step road map of Procedure 1, below.

This road map provides intuition behind the general characterization and explicit construction of a proper test statistics null distribution Q_0.

It also provides a template for $\Theta(F_{V_n})$-controlling joint single-step common-cut-off Procedure 5 and common-quantile Procedure 8.

The main idea is to substitute control of the unknown parameter $\Theta(F_{V_n})$, for the true distribution F_{V_n} of the number of Type I errors, by control of the corresponding known parameter $\Theta(F_{R_0})$, for the null distribution F_{R_0} of the number of rejected hypotheses.

Test Statistics Null Distribution

Procedure 1 [Three-step road map for controlling Type I error rates $\Theta(F_{V_n})$]

1. *Null domination conditions for the Type I error rates $\Theta(F_{V_n})$ and $\Theta(F_{V_0})$. Select a test statistics null distribution Q_0 such that the following null domination assumption for the Type I error rates is satisfied.*

$$\Theta(F_{V_n}) \leq \Theta(F_{V_0}) \quad \text{[finite sample control]}$$

$$\limsup_{n\to\infty} \Theta(F_{V_n}) \leq \Theta(F_{V_0}) \quad \text{[asymptotic control].} \tag{NDΘ}$$

2. *Monotonicity of the Type I error rate mapping Θ.* *Under monotonicity Assumption $\mathsf{M}\Theta$ for the Type I error rate mapping Θ, one has*

$$\Theta(F_{V_0}) \leq \Theta(F_{R_0}). \tag{6}$$

3. *Control of $\Theta(F_{R_0})$.* *Select rejection regions $\mathcal{C}_n(m) = \mathcal{C}(m; T_n, Q_0, \alpha)$ so that the following Type I error constraint is satisfied,*

$$\Theta(F_{R_0}) \leq \alpha. \tag{7}$$

Test Statistics Null Distribution

Combining Steps 1–3 provides the desired control of the actual Type I error rate $\Theta(F_{V_n})$ at level $\alpha \in (0,1)$, that is,

$$\begin{aligned} \Theta(F_{V_n}) \leq \Theta(F_{V_0}) \leq \Theta(F_{R_0}) \quad &\leq \quad \alpha \qquad \text{[finite sample control]} \\ \limsup_{n\to\infty} \Theta(F_{V_n}) \leq \Theta(F_{V_0}) \leq \Theta(F_{R_0}) \quad &\leq \quad \alpha \qquad \text{[asymptotic control]}. \end{aligned} \tag{8}$$

Note that the road map of Procedure 1 is conservative in two ways: (i) from the null domination of the Type I error rate in Step 1, $\Theta(F_{V_n}) \leq \Theta(F_{V_0})$; (ii) from controlling $\Theta(F_{R_0}) \geq \Theta(F_{V_0})$ in Step 3.

Step 1 is often the most problematic and requires a judicious choice for the test statistics null distribution Q_0.

Test Statistics Null Distribution

For certain families of Type I error rate mappings Θ and rejection regions $\mathcal{C}_n$, Θ-specific Type I error rate null domination Assumption $\mathsf{ND}\Theta$, in Step 1 of the road map, can be shown to hold under the following alternate forms of null domination.

- Null domination for the distributions F_{V_n} and F_{V_0} of the number of Type I errors.
- Null domination for the joint distributions $Q_{n,\mathcal{H}_0}$ and $Q_{0,\mathcal{H}_0}$ of the $\mathcal{H}_0$-specific subvector $(T_n(m) : m \in \mathcal{H}_0)$ of test statistics for the true null hypotheses $\mathcal{H}_0$.

Test Statistics Null Distribution

Null domination conditions for the numbers of Type I errors V_n and V_0. The number of Type I errors V_0, under the null distribution Q_0, is stochastically greater than the number of Type I errors V_n, under the true distribution Q_n for the test statistics T_n.

That is, for each $x \in \{0, \ldots, M\}$,

$$\begin{aligned} F_{V_n}(x) &\geq F_{V_0}(x) && \text{[finite sample control]} \\ \liminf_{n\to\infty} F_{V_n}(x) &\geq F_{V_0}(x) && \text{[asymptotic control].} \end{aligned} \quad \text{(NDV)}$$

Test Statistics Null Distribution

Joint null domination conditions for the $\mathcal{H}_0$-specific test statistics $(T_n(m) : m \in \mathcal{H}_0)$. The null distribution $Q_{0,\mathcal{H}_0}$, of the $\mathcal{H}_0$-specific subvector of test statistics $(T_n(m) : m \in \mathcal{H}_0)$, is stochastically greater than the corresponding true distribution $Q_{n,\mathcal{H}_0} = Q_{n,\mathcal{H}_0}(P)$. That is, for all $z \in I\!R^{h_0}$,

$$\begin{aligned} Q_{n,\mathcal{H}_0}(z) &\geq Q_{0,\mathcal{H}_0}(z) && \text{[finite sample control]} \\ \liminf_{n\to\infty} Q_{n,\mathcal{H}_0}(z) &\geq Q_{0,\mathcal{H}_0}(z) && \text{[asymptotic control],} \end{aligned} \tag{jtNDT}$$

where, for the asymptotic statement, the null distribution $Q_{0,\mathcal{H}_0}$ is further required to be continuous.

Test Statistics Null Distribution

Assumption jtNDT: Joint null domination for $\mathcal{H}_0$-specific test statistics

$$Q_{n,\mathcal{H}_0} \geq Q_{0,\mathcal{H}_0}.$$

$$\Downarrow$$

Assumption NDV: Null domination for number of Type I errors, for one-sided rejection regions of the form $\mathcal{C}(m) = (c(m), +\infty)$,

$$F_{V_n} \geq F_{V_0}.$$

$$\Downarrow$$

Assumption NDΘ: Null domination for Type I error rate, under Assumptions MΘ and CΘ,

$$\Theta(F_{V_n}) \leq \Theta(F_{V_0}).$$

Test Statistics Null Distribution

The general characterization of a proper null distribution in terms of null domination conditions leads to the explicit construction of the following two main types of test statistics null distributions.

- Null shift and scale-transformed test statistics null distribution, based on user-supplied upper bounds for the means and variances of the test statistics for the true null hypotheses.
- Null quantile-transformed test statistics null distribution, based on user-supplied marginal test statistics null distributions.

Test Statistics Null Distribution

Null shift and scale-transformed test statistics null distribution.
The first original null distribution of Dudoit et al. (2004b), van der Laan et al. (2004a), and Pollard and van der Laan (2004), is defined as the asymptotic distribution of the M-vector Z_n of null shift and scale-transformed test statistics,

$$Z_n(m) \equiv \sqrt{\min\left\{1, \frac{\tau_0(m)}{\mathrm{Var}[T_n(m)]}\right\}}\,(T_n(m) - \mathrm{E}[T_n(m)]) + \lambda_0(m), \tag{9}$$

where $\lambda_0(m)$ and $\tau_0(m)$ are, respectively, user-supplied upper bounds for the means and variances of the $\mathcal{H}_0$-specific test statistics.

Test Statistics Null Distribution

- Role of null shift values λ_0. The construction in Equation (9) is inspired by joint null domination Assumption jtNDT, for the $\mathcal{H}_0$-specific subvector of test statistics $(T_n(m) : m \in \mathcal{H}_0)$.

The purpose of the null shift values $\lambda_0(m)$ is to generate $\mathcal{H}_0$-specific statistics $(Z_n(m) : m \in \mathcal{H}_0)$ that are asymptotically stochastically greater than the original test statistics $(T_n(m) : m \in \mathcal{H}_0)$.

Thus, for one-sided rejection regions of the form $\mathcal{C}_n(m) = (c_n(m), +\infty)$, the number of Type I errors V_0, under the null distribution Q_0, is asymptotically stochastically greater than the number of Type I errors V_n, under the true distribution Q_n.

The null distribution therefore satisfies asymptotic null domination Assumption NDV, for the number of Type I errors, and Θ-specific asymptotic null domination Assumption NDΘ, under monotonicity Assumption MΘ and continuity Assumption CΘ at F_{V_0}.

Test Statistics Null Distribution

- Role of null scale values τ_0. In contrast, the null scale values $\tau_0(m)$ are not needed for Type I error control.

The purpose of $\tau_0(m)$ is to avoid a degenerate null distribution and infinite cut-offs for the false null hypotheses ($m \in \mathcal{H}_1$), an important property for power considerations.

This scaling is needed, in particular, for F-statistics that have asymptotically infinite means and variances for non-local alternative hypotheses.

Test Statistics Null Distribution

- Estimation of null values λ_0 and τ_0. The null values $\lambda_0(m)$ and $\tau_0(m)$ only depend on the marginal distributions of the test statistics $T_n(m)$ for the true null hypotheses $\mathcal{H}_0$ and are generally known from single hypothesis testing.

For instance, for the test of single-parameter null hypotheses using t-statistics, the null values are $\lambda_0(m) = 0$ and $\tau_0(m) = 1$.

For testing the equality of K population mean vectors using F-statistics, the null values are $\lambda_0(m) = 1$ and $\tau_0(m) = 2/(K-1)$, under the assumption of equal variances in the different populations.

The null values $\lambda_0(m)$ and $\tau_0(m)$ may depend on the unknown data generating distribution P (e.g., F-statistics when population variances are unequal). In such a situation, one may replace the parameters $\lambda_0(m)$ and $\tau_0(m)$ by consistent estimators thereof.

Test Statistics Null Distribution

- t-statistics: Gaussian null distribution. For a broad class of testing problems, such as the test of single-parameter null hypotheses using t-statistics, the null distribution $Q_0 = Q_0(P)$ is an M-variate Gaussian distribution, with mean vector zero and covariance matrix $\sigma^* = \Sigma^*(P)$, that is, $Q_0 = \mathrm{N}(0, \sigma^*)$.

For tests where the parameter of interest is the M-dimensional mean vector $\Psi(P) = \psi = \mathrm{E}[X]$, the estimator ψ_n is simply the M-vector of empirical means and $\sigma^* = \Sigma^*(P) = \mathrm{Cor}[X]$ is the correlation matrix of $X \sim P$, that is, $Q_0(P) = \mathrm{N}(0, \mathrm{Cor}[X])$. More generally, for an asymptotically linear estimator ψ_n, $\Sigma^*(P)$ is the correlation matrix of the vector influence curve.

This situation covers standard one-sample and two-sample t-statistics for tests of means, but also test statistics for correlation coefficients and regression coefficients in linear and non-linear models.

Test Statistics Null Distribution

- F-statistics: Gaussian quadratic form null distribution. For testing the equality of K population mean vectors using F-statistics, an F-statistic-specific null distribution Q_0^F may be defined as the joint distribution of an M-vector of quadratic forms of Gaussian random variables.

- Estimation of the test statistics null distribution. In practice, the test statistics null distribution $Q_0 = Q_0(P)$ is unknown, as it depends on the unknown data generating distribution P.

  Resampling procedures may be used to conveniently obtain consistent estimators Q_{0n} of the null distribution Q_0 and of the corresponding test statistic cut-offs and adjusted p-values.

Test Statistics Null Distribution

Theorem 2 [Null shift and scale-transformed test statistics null distribution]
Suppose there exist known M-vectors $\lambda_0 \in I\!R^M$ and $\tau_0 \in I\!R^{+M}$ of null values, so that, for each $m \in \mathcal{H}_0$,

$$\begin{aligned} \limsup_{n\to\infty} \mathrm{E}[T_n(m)] &\leq \lambda_0(m) \qquad (10)\\ \limsup_{n\to\infty} \mathrm{Var}[T_n(m)] &\leq \tau_0(m). \end{aligned}$$

Let

$$\nu_{0,n}(m) \equiv \sqrt{\min\left\{1, \frac{\tau_0(m)}{\mathrm{Var}[T_n(m)]}\right\}} \qquad (11)$$

and define an M-vector of null shift and scale-transformed test statistics $Z_n = (Z_n(m) : m = 1, \ldots, M)$ by

$$Z_n(m) \equiv \nu_{0,n}(m)\left(T_n(m) - \mathrm{E}[T_n(m)]\right) + \lambda_0(m), \qquad m = 1, \ldots, M. \qquad (12)$$

Suppose that the random M-vector Z_n converges weakly to a random M-vector Z, with continuous joint distribution $Q_0 = Q_0(P)$,

$$Z_n \stackrel{\mathcal{L}}{\Rightarrow} Z \sim Q_0(P). \tag{13}$$

Then, the asymptotic joint distribution Q_0 satisfies asymptotic joint null domination Assumption **jtNDT** *for the $\mathcal{H}_0$-specific subvector of test statistics $(T_n(m) : m \in \mathcal{H}_0)$. That is, for all $z \in I\!R^{h_0}$,*

$$\liminf_{n \to \infty} Q_{n,\mathcal{H}_0}(z) \geq Q_{0,\mathcal{H}_0}(z).$$

In addition, for all $c = (c(m) : m = 1, \ldots, M) \in I\!R^M$ and $x \in \{0, \ldots, M\}$,

$$\liminf_{n \to \infty} \Pr_{Q_n} \left(\sum_{m \in \mathcal{H}_0} \mathrm{I}\left(T_n(m) > c(m)\right) \leq x \right) \geq \Pr_{Q_0} \left(\sum_{m \in \mathcal{H}_0} \mathrm{I}\left(Z(m) > c(m)\right) \leq x \right).$$

Thus, for one-sided rejection regions of the form $\mathcal{C}_n(m) = (c_n(m), +\infty)$*, the null distribution* Q_0 *satisfies asymptotic null domination Assumption* **NDV** *for the number of Type I errors,*

$$\liminf_{n\to\infty} F_{V_n}(x) \geq F_{V_0}(x), \qquad \forall\, x \in \{0, \ldots, M\}.$$

If one further assumes that the Type I error rate mapping Θ *meets monotonicity Assumption* **M**Θ *and continuity Assumption* **C**Θ *at* F_{V_0}*, then the null distribution* Q_0 *also satisfies asymptotic null domination Assumption* **ND**Θ *for the Type I error rate,*

$$\limsup_{n\to\infty} \Theta(F_{V_n}) \leq \Theta(F_{V_0}).$$

Test Statistics Null Distribution

Proof of Theorem 2. The proof is straightforward and is based on an intermediate random vector $\widetilde{Z}_n = (\widetilde{Z}_n(m) : m = 1, \ldots, M)$, defined as

$$\widetilde{Z}_n(m) = T_n(m) + \max\left\{0, \lambda_0(m) - \mathrm{E}[T_n(m)]\right\}, \qquad m = 1, \ldots, M. \quad (14)$$

First, note that $T_n(m) \leq \widetilde{Z}_n(m)$ for each $m = 1, \ldots, M$. Next, for $m \in \mathcal{H}_0$, since $\limsup_n \mathrm{E}[T_n(m)] \leq \lambda_0(m)$ and $\limsup_n \mathrm{Var}[T_n(m)] \leq \tau_0(m)$, then $\lim_n \nu_{0,n}(m) = 1$ and the $\mathcal{H}_0$-specific subvectors $(\widetilde{Z}_n(m) : m \in \mathcal{H}_0)$ and $(Z_n(m) : m \in \mathcal{H}_0)$ have the same asymptotic joint distribution. That is,

$$(\widetilde{Z}_n(m) : m \in \mathcal{H}_0) \stackrel{\mathcal{L}}{\Rightarrow} (Z(m) : m \in \mathcal{H}_0) \sim Q_{0,\mathcal{H}_0}.$$

Thus, asymptotic joint null domination Assumption jtNDT follows from the definition of weak convergence to a continuous limit distribution Q_0 (Equation (57)). That is, for each $z \in I\!R^{h_0}$ and corresponding

h_0-dimensional rectangle $(-\infty, z] \subseteq I\!R^{h_0}$,

$$\begin{aligned}
\liminf_{n\to\infty} Q_{n,\mathcal{H}_0}(z) &= \liminf_{n\to\infty} \Pr\left((T_n(m) : m \in \mathcal{H}_0) \in (-\infty, z]\right) \\
&\geq \liminf_{n\to\infty} \Pr\left((\widetilde{Z}_n(m) : m \in \mathcal{H}_0) \in (-\infty, z]\right) \\
&= \Pr\left((Z(m) : m \in \mathcal{H}_0) \in (-\infty, z]\right) \\
&= Q_{0,\mathcal{H}_0}(z).
\end{aligned}$$

In addition, for all $c = (c(m) : m = 1, \ldots, M) \in I\!R^M$, the Continuous Mapping Theorem (Theorem 18), applied to the function $\ell((z(m) : m \in \mathcal{H}_0)) = \sum_{m\in\mathcal{H}_0} \mathrm{I}\left(z(m) > c(m)\right)$, implies that

$$\begin{aligned}
\ell((\widetilde{Z}_n(m) : m \in \mathcal{H}_0)) &= \sum_{m\in\mathcal{H}_0} \mathrm{I}\left(\widetilde{Z}_n(m) > c(m)\right) \\
&\overset{\mathcal{L}}{\Rightarrow} \sum_{m\in\mathcal{H}_0} \mathrm{I}\left(Z(m) > c(m)\right) = \ell((Z(m) : m \in \mathcal{H}_0)).
\end{aligned}$$

Asymptotic null domination Assumption NDV then follows from Proposition 1. That is, for all $c = (c(m) : m = 1, \ldots, M) \in I\!R^M$ and

$x \in \{0, \ldots, M\}$,

$$\begin{aligned}
\liminf_{n\to\infty} F_{V_n}(x) &= \liminf_{n\to\infty} \Pr\left(\sum_{m\in\mathcal{H}_0} \mathrm{I}\left(T_n(m) > c(m)\right) \leq x\right) \\
&\geq \liminf_{n\to\infty} \Pr\left(\sum_{m\in\mathcal{H}_0} \mathrm{I}\left(\widetilde{Z}_n(m) > c(m)\right) \leq x\right) \\
&= \Pr\left(\sum_{m\in\mathcal{H}_0} \mathrm{I}\left(Z(m) > c(m)\right) \leq x\right) \\
&= F_{V_0}(x).
\end{aligned}$$

$\square$

Test Statistics Null Distribution

Null quantile-transformed test statistics null distribution. The second and most recent proposal of van der Laan and Hubbard (2006) is defined as the asymptotic distribution of the M-vector $\breve{Z}_n$ of null quantile-transformed test statistics,

$$\breve{Z}_n(m) \equiv q_{0,m}^{-1} Q_{n,m}^{\Delta}(T_n(m)), \tag{15}$$

where $q_{0,m}$ are user-supplied marginal test statistics null distributions that satisfy marginal null domination Assumption mgNDT, below.

According to the generalized quantile-quantile function transformation of Yu and van der Laan (2002), $Q_{n,m}^{\Delta}(z) \equiv \Delta Q_{n,m}(z) + (1-\Delta)Q_{n,m}(z^-)$ and the random variable Δ is uniformly distributed on the interval $[0,1]$, independently of the data $\mathcal{X}_n$.

Test Statistics Null Distribution

Marginal null domination conditions for the $\mathcal{H}_0$-specific test statistics $(T_n(m) : m \in \mathcal{H}_0)$. For each $m \in \mathcal{H}_0$ and $z \in I\!R$,

$$
\begin{aligned}
Q_{n,m}(z) &\geq q_{0,m}(z) \qquad \text{[finite sample control]} \\
\liminf_{n\to\infty} Q_{n,m}(z) &\geq q_{0,m}(z) \qquad \text{[asymptotic control].}
\end{aligned}
\tag{mgNDT}
$$

Test Statistics Null Distribution

The above two test statistics null distributions are defined as the asymptotic distributions of null-transformed test statistics.

Finite sample versions of both null distributions and corresponding Type I error control results have also been derived.

Test Statistics Null Distribution

As previously noted, the two proposed test statistics null distributions $Q_0 = Q_0(P)$ depend on the typically unknown data generating distribution P.

Although in some cases marginal test statistics null distributions may be known from single hypothesis testing, the dependence structure among the test statistics is usually unknown. In practice, one therefore needs to estimate the joint null distribution Q_0.

Consistent estimators Q_{0n} of the test statistics null distribution Q_0 and of the corresponding test statistic cut-offs and adjusted p-values may be obtained according to the following three main approaches: (i) general direct bootstrap estimation; (ii) test statistic-specific estimation (e.g., for t-statistics, χ^2-statistics, F-statistics); (iii) data generating null distribution estimation.

Test Statistics Null Distribution

Let $P_n^\star$ denote an estimator of the true data generating distribution P.

For the non-parametric bootstrap, $P_n^\star$ is simply the empirical distribution P_n, that is, samples of size n are drawn at random, with replacement from the observed data $\mathcal{X}_n = \{X_i : i = 1, \ldots, n\}$.

For the model-based bootstrap, $P_n^\star$ belongs to a model $\mathcal{M}$ for the data generating distribution P, such as a family of multivariate Gaussian distributions.

Test Statistics Null Distribution

Procedure 3 [Bootstrap estimation of the null shift and scale-transformed test statistics null distribution]

1. *Generate B bootstrap samples, $\mathcal{X}_n^b \equiv \{X_i^b : i = 1, \ldots, n\}$, $b = 1, \ldots, B$. For the bth sample, the X_i^b, $i = 1, \ldots, n$, are n IID copies of a random variable $X^{\#} \sim P_n^{\star}$.*
2. *For each bootstrap sample $\mathcal{X}_n^b$, compute an M-vector of test statistics, $T_n^B(\cdot, b) = (T_n^B(m, b) : m = 1, \ldots, M)$, that can be arranged in an $M \times B$ matrix, $\mathbf{T}_n^B = \big(T_n^B(m, b) : m = 1, \ldots, M; b = 1, \ldots, B\big)$, with rows corresponding to the M null hypotheses and columns to the B bootstrap samples.*
3. *Compute row means and variances of the matrix $\mathbf{T}_n^B$, to yield estimators of the means, $\mathrm{E}[T_n(m)]$, and variances, $\mathrm{Var}[T_n(m)]$, of the test statistics under the true data generating distribution P.*

That is, compute

$$\mathrm{E}[T_n^B(m,\cdot)] \equiv \frac{1}{B}\sum_{b=1}^{B} T_n^B(m,b), \tag{16}$$

$$\mathrm{Var}[T_n^B(m,\cdot)] \equiv \frac{1}{B}\sum_{b=1}^{B} \left(T_n^B(m,b) - \mathrm{E}[T_n^B(m,\cdot)]\right)^2.$$

4. *Obtain an* $M \times B$ *matrix,* $\mathbf{Z}_n^B = \big(Z_n^B(m,b) : m = 1,\ldots,M;\, b = 1,\ldots,B\big)$, *of null shift and scale-transformed bootstrap test statistics* $Z_n^B(m,b)$, *as in Theorem 2, by row-shifting and scaling the matrix* $\mathbf{T}_n^B$ *using the bootstrap estimators of* $\mathrm{E}[T_n(m)]$ *and* $\mathrm{Var}[T_n(m)]$ *and the user-supplied null values* $\lambda_0(m)$ *and* $\tau_0(m)$. *That is, define*

$$Z_n^B(m,b) \equiv \sqrt{\min\left\{1, \frac{\tau_0(m)}{\mathrm{Var}[T_n^B(m,\cdot)]}\right\}}\Big(T_n^B(m,b) - \mathrm{E}[T_n^B(m,\cdot)]\Big) + \lambda_0(m). \tag{17}$$

5. *The bootstrap estimator* Q_{0n} *of the null distribution* Q_0 *from Theorem 2 is the empirical distribution of the* B *columns* $\{Z_n^B(\cdot, b) : b = 1, \ldots, B\}$ *of matrix* $\mathbf{Z}_n^B$.

Test Statistics Null Distribution

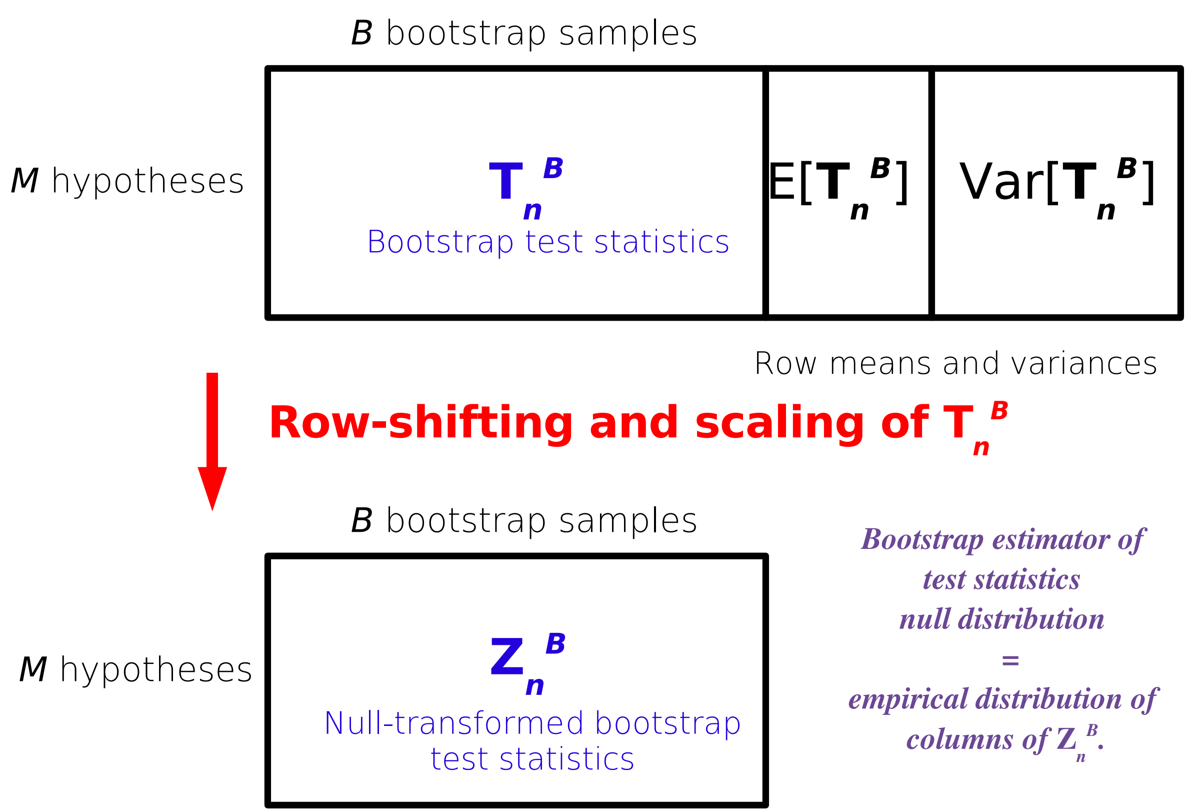


Figure 1: *Bootstrap estimation of the null shift and scale-transformed test statistics null distribution Q_0 (Procedure 3).*

Test Statistics Null Distribution

Procedure 4 [Bootstrap estimation of the null quantile-transformed test statistics null distribution]

1. *Generate B bootstrap samples, $\mathcal{X}_n^b \equiv \{X_i^b : i = 1, \ldots, n\}$, $b = 1, \ldots, B$. For the bth sample, the X_i^b, $i = 1, \ldots, n$, are n IID copies of a random variable $X^\# \sim P_n^\star$.*
2. *For each bootstrap sample $\mathcal{X}_n^b$, compute an M-vector of test statistics, $T_n^B(\cdot, b) = (T_n^B(m, b) : m = 1, \ldots, M)$, that can be arranged in an $M \times B$ matrix, $\mathbf{T}_n^B = \big(T_n^B(m, b) : m = 1, \ldots, M; b = 1, \ldots, B\big)$, with rows corresponding to the M null hypotheses and columns to the B bootstrap samples.*
3. *Define M bootstrap marginal cumulative distribution functions*

$Q^B_{n,m}$, *as the empirical CDFs of the rows of matrix* $\mathbf{T}^B_n$, *that is,*

$$Q^B_{n,m}(z) \equiv \frac{1}{B} \sum_{b=1}^{B} \mathrm{I}\left(T^B_n(m,b) \leq z \right). \qquad (18)$$

4. *Obtain an* $M \times B$ *matrix,* $\mathbf{Z}^B_n = \left(Z^B_n(m,b) : m = 1, \ldots, M; b = 1, \ldots, B \right)$, *of null quantile-transformed bootstrap test statistics* $Z^B_n(m,b)$, *defined as*

$$Z^B_n(m,b) \equiv q^{-1}_{0,m} Q^{B,\Delta}_{n,m}(T^B_n(m,b)), \qquad (19)$$

   where $Q^{B,\Delta}_{n,m}(z) \equiv \Delta Q^B_{n,m}(z) + (1-\Delta) Q^B_{n,m}(z^-)$ *and the random variable* Δ *is uniformly distributed on the interval* $[0,1]$, *independently of the data* $\mathcal{X}_n$.

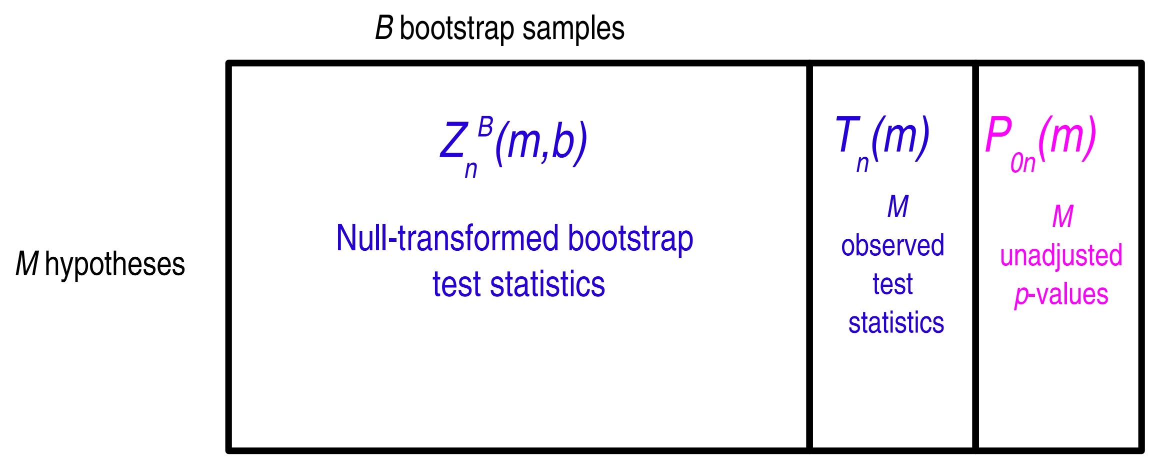
5. *The bootstrap estimator* $\breve{Q}_{0n}$ *of the null distribution* $\breve{Q}_0$ *is the empirical distribution of the* B *columns* $\{Z^B_n(\cdot,b) : b = 1, \ldots, B\}$ *of matrix* $\mathbf{Z}^B_n$.

Test Statistics Null Distribution

For one-sided rejection regions of the form $\mathcal{C}_n(m) = (c_n(m), +\infty)$, bootstrap estimators of the unadjusted p-values $P_{0n}(m)$ may be obtained from the matrix $\mathbf{Z}_n^B = \left(Z_n^B(m,b)\right)$ of Procedure 3 or 4 by recording, for each row m, the proportion of null-transformed bootstrap test statistics $Z_n^B(m,b)$ that are greater than or equal to the observed test statistic $T_n(m)$. That is,

$$P_{0n}(m) = \frac{1}{B}\sum_{b=1}^{B} \mathrm{I}\left(Z_n^B(m,b) \geq T_n(m)\right), \qquad m = 1, \ldots, M. \quad (20)$$

Test Statistics Null Distribution



For each row m, the unadjusted p-value $P_{0n}(m)$ is the proportion of the B null-transformed bootstrap test statistics $Z_n^B(m,b)$ that exceed the observed test statistic $T_n(m)$.

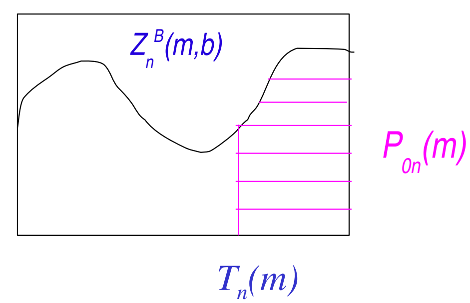


Figure 2: *Bootstrap estimation of the unadjusted p-values $P_{0n}(m)$.*

Test Statistics Null Distribution

There is no obvious general recommendation for the number of bootstrap samples B.

However, note that bootstrap unadjusted p-values are discrete tail probabilities, with steps of size $1/B$. Thus, for estimating very small p-values (e.g., of the order of 10^{-9}), one clearly needs a very large B in order to get enough resolution in the tails.

In general, the user needs to find a balance between estimation accuracy and computational cost.

Test Statistics Null Distribution

Given an estimator Q_{0n} of the null distribution Q_0, we have derived algorithms for estimating test statistic cut-offs and adjusted p-values for $\Theta(F_{V_n})$-controlling single-step MTPs and FWER-controlling step-down MTPs.

For a consistent estimator of the null distribution, the resulting MTPs can be shown to asymptotically control the Type I error rate.

Details in Dudoit and van der Laan (2007, Chapters 4 and 5).

Test Statistics Null Distribution

The following two main points distinguish our approach from existing approaches to Type I error control and the choice of a test statistics null distribution (e.g., Hochberg and Tamhane (1987) and Westfall and Young (1993)).

- Firstly, we are only concerned with control of the Type I error rate under the true data generating distribution P, i.e., under the joint distribution $Q_n = Q_n(P)$, implied by P, for the test statistics T_n.

The concepts of weak and strong control of a Type I error rate are therefore irrelevant in our context.

Test Statistics Null Distribution

- Secondly, we propose a null distribution for the test statistics ($T_n \sim Q_0$) rather than a data generating null distribution ($X \sim P_0 \in \cap_{m=1}^{M} \mathcal{M}(m)$).

The latter practice does not necessarily provide proper Type I error control under the true distribution P.

Indeed, the test statistics assumed null distribution $Q_n(P_0)$ and their true distribution $Q_n(P)$ may have different dependence structures for the true null hypotheses $\mathcal{H}_0$ and, as a result, may violate the required null domination condition for the Type I error rate (Assumption $\mathsf{ND}\Theta$).

Test Statistics Null Distribution

We stress the generality of the two test statistics null distributions: Type I error control does not rely on restrictive assumptions such as subset pivotality and holds for general data generating distributions (with arbitrary dependence structures among variables), null hypotheses, and test statistics.

The null quantile-transformed test statistics null distribution has the additional advantage that the marginal test statistics null distributions may be set to the optimal (i.e., most powerful) null distributions one would use in single hypothesis testing (e.g., permutation marginal null distributions, Gaussian or other parametric marginal null distributions).

We provide resampling procedures (e.g., bootstrap) to conveniently obtain consistent estimators of the null distribution and the resulting test statistic cut-offs and adjusted p-values.

Multiple Testing Procedures

Having selected a proper test statistics null distribution Q_0 (or estimator thereof, Q_{0n}), there remains the main task of specifying rejection regions for each null hypothesis, so that the resulting procedure probabilistically controls Type I errors. We have developed four main classes of multiple testing procedures (MTP).

- Joint single-step common-cut-off and common-quantile procedures for controlling general Type I error rates $\Theta(F_{V_n})$, defined as arbitrary parameters of the distribution of the number of Type I errors V_n.
  E.g. Generalized family-wise error rate (gFWER), $gFWER(k) = 1 - F_{V_n}(k) = \Pr(V_n > k)$.
- Joint step-down common-cut-off (maxT) and common-quantile (minP) procedures for controlling the family-wise error rate (FWER), $FWER = gFWER(0) = 1 - F_{V_n}(0) = \Pr(V_n > 0)$.

Multiple Testing Procedures

- (Marginal/joint single-step/stepwise common-cut-off/common-quantile) augmentation multiple testing procedures (AMTP) for controlling generalized tail probability (gTP) error rates, $gTP(q, g) = \Pr(g(V_n, R_n) > q)$, based on an initial gFWER-controlling procedure.
  E.g. gFWER with $g(v, r) = v$, TPPFP with $g(v, r) = v/r$.
- Joint resampling-based empirical Bayes procedures for controlling generalized tail probability error rates.

Multiple Testing Procedures

N.B. Our proposed multiple testing procedures (asymptotically) control Type I error rates $\Theta(F_{V_n,R_n})$, for general

- data generating distributions, with arbitrary dependence structures among variables;
- null hypotheses, defined in terms of submodels for the data generating distribution;
- test statistics, e.g., t-statistics, χ^2-statistics, F-statistics.

$\Theta(F_{V_n})$-Controlling Single-Step Procedures

Inspired by the three-step road map of Procedure 1, we have derived joint single-step procedures for controlling Type I error rates defined as arbitrary parameters $\Theta(F_{V_n})$ of the distribution of the number of Type I errors V_n (Dudoit et al., 2004b; Pollard and van der Laan, 2004).

Given a test statistics null distribution Q_0 and nominal Type I error level α, the main idea is to substitute control of the unknown parameter $\Theta(F_{V_n})$, for the true distribution of the number of Type I errors V_n, by control of the corresponding known parameter $\Theta(F_{R_0})$, for the null distribution of the number of rejected hypotheses R_0.

$\Theta(F_{V_n})$-Controlling Single-Step Procedures

Specifically, consider single-step procedures of the form

$$\mathcal{R}_n = \{m : T_n(m) > c(m)\},$$

where the cut-offs $c(m)$ are chosen so that

$$\Theta(F_{R_0}) \leq \alpha, \tag{21}$$

for $R_0 = R(c|Q_0) = \sum_{m=1}^{M} \mathrm{I}(Z(m) > c(m))$ and $Z \sim Q_0$.

Among the family of MTPs that satisfy the Type I error constraint $\Theta(F_{R_0}) \leq \alpha$, two types of procedures are proposed:
Procedure 5, based on a common cut-off for all test statistics;
Procedure 8, with common-quantile cut-offs for the test statistics.

$\Theta(F_{V_n})$-Controlling Single-Step Procedures

Procedure 5 [$\Theta(F_{V_n})$-controlling single-step common-cut-off procedure]
For controlling the Type I error rate $\Theta(F_{V_n})$ at level α, the set of rejected null hypotheses is of the form $\mathcal{R}_n = \{m : T_n(m) > c_0\}$, where the common cut-off c_0 is the smallest (i.e., least conservative) value for which $\Theta(F_{R_0}) \leq \alpha$. Adjusted p-values are given by

$$\widetilde{P}_{0n}(m) = \Theta(F_{R(T_n(m)^{(M)}|Q_0)}), \qquad m = 1, \ldots, M, \qquad (22)$$

where $T_n(m)^{(M)}$ denotes an M-vector of common cut-offs equal to $T_n(m)$.

$\Theta(F_{V_n})$-Controlling Single-Step Procedures

Two special cases of interest are

- gFWER control, i.e., $\Theta(F_{V_n}) = 1 - F_{V_n}(k)$: single-step $T(k+1)$ Procedure 6;
- FWER control, i.e., $\Theta(F_{V_n}) = 1 - F_{V_n}(0)$: single-step maxT Procedure 7.

$\Theta(F_{V_n})$-Controlling Single-Step Procedures

Procedure 6 [gFWER-controlling single-step common-cut-off $T(k+1)$ procedure]
For controlling $gFWER(k)$, the single-step $T(k+1)$ procedure is based on the $(k+1)$st largest test statistic. Adjusted p-values are given by

$$\widetilde{p}_{0n}(m) = \Pr_{Q_0}\left(Z^{\circ}(k+1) \geq t_n(m)\right), \qquad m = 1, \ldots, M, \qquad (23)$$

where $Z^{\circ}(m)$ denotes the mth largest element of the M-vector $Z = (Z(m) : m = 1, \ldots, M) \sim Q_0$, so that $Z^{\circ}(1) \geq \cdots \geq Z^{\circ}(M)$.

$\Theta(F_{V_n})$-Controlling Single-Step Procedures

Procedure 7 [FWER-controlling single-step common-cut-off maxT procedure]
For controlling the FWER, the single-step maxT procedure is based on the maximum test statistic, $Z^\circ(1) \equiv \max_m Z(m)$, *for the* M*-vector* $Z = (Z(m) : m = 1, \ldots, M) \sim Q_0$*. Adjusted p-values are given by*

$$\widetilde{p}_{0n}(m) = \Pr_{Q_0}\left(\max_{m=1,\ldots,M} Z(m) \geq t_n(m)\right), \qquad m = 1, \ldots, M. \tag{24}$$

$\Theta(F_{V_n})$-Controlling Single-Step Procedures

Procedure 8 [$\Theta(F_{V_n})$-controlling single-step common-quantile procedure]
For controlling the Type I error rate $\Theta(F_{V_n})$ at level α, the set of rejected null hypotheses is of the form $\mathcal{R}_n = \{m : T_n(m) > c_0(m)\}$, where $c_0(m) = Q_{0,m}^{-1}(\delta_0) = \inf\{z \in I\!R : Q_{0,m}(z) \geq \delta_0\}$ is the δ_0-quantile of the marginal null distribution $Q_{0,m}$. The common quantile probability δ_0 is chosen as the smallest (i.e., least conservative) value for which $\Theta(F_{R_0}) \leq \alpha$. Adjusted p-values are given by

$$\widetilde{P}_{0n}(m) = \Theta(F_{R(q_0^{-1}(1-P_{0n}(m)))|Q_0}), \qquad m = 1, \ldots, M, \tag{25}$$

where $P_{0n}(m)$ is the unadjusted p-value for null hypothesis $H_0(m)$,

$$P_{0n}(m) = \bar{Q}_{0,m}(T_n(m)) = 1 - Q_{0,m}(T_n(m)), \tag{26}$$

and $q_0^{-1}(\delta) = (Q_{0,m}^{-1}(\delta) : m = 1, \ldots, M)$ denotes an M-vector of δ-quantiles for the marginal null distributions $Q_{0,m}$.

$\Theta(F_{V_n})$-Controlling Single-Step Procedures

Two special cases of interest are

- gFWER control, i.e., $\Theta(F_{V_n}) = 1 - F_{V_n}(k)$: single-step $P(k+1)$ Procedure 9;
- FWER control, i.e., $\Theta(F_{V_n}) = 1 - F_{V_n}(0)$: single-step minP Procedure 10.

$\Theta(F_{V_n})$-Controlling Single-Step Procedures

Procedure 9 [gFWER-controlling single-step common-quantile $P(k+1)$ procedure]
For controlling $gFWER(k)$, the single-step $P(k+1)$ procedure is based on the $(k+1)$st smallest unadjusted p-value. Adjusted p-values are given by

$$\widetilde{p}_{0n}(m) = \Pr_{Q_0}\left(P_0^\circ(k+1) \leq p_{0n}(m)\right), \qquad m = 1, \ldots, M, \tag{27}$$

where $P_0(m) = \bar{Q}_{0,m}(Z(m))$ denote unadjusted p-values under the test statistics null distribution Q_0, i.e., for $Z = (Z(m) : m = 1, \ldots, M) \sim Q_0$, and $P_0^\circ(m)$ denotes the mth smallest unadjusted p-value, so that $P_0^\circ(1) \leq \cdots \leq P_0^\circ(M)$.

$\Theta(F_{V_n})$-Controlling Single-Step Procedures

Procedure 10 [FWER-controlling single-step common-quantile minP procedure]
For controlling the FWER, the single-step minP procedure is based on the minimum unadjusted p-value, $P_0^\circ(1) \equiv \min_m P_0(m)$*, where* $P_0(m) = \bar{Q}_{0,m}(Z(m))$ *denote unadjusted p-values under the test statistics null distribution* Q_0*, i.e., for* $Z = (Z(m) : m = 1, \ldots, M) \sim Q_0$*. Adjusted p-values are given by*

$$\widetilde{p}_{0n}(m) = \Pr_{Q_0}\left(\min_{m=1,\ldots,M} P_0(m) \leq p_{0n}(m)\right), \qquad m = 1, \ldots, M. \tag{28}$$

FWER-Controlling Step-Down Procedures

In stepwise procedures, the decision to reject a particular null hypothesis depends on the outcome of the tests of other hypotheses.

That is, the (single-step) testing procedure is applied to a sequence of successively smaller nested random (i.e., data-dependent) subsets of null hypotheses, defined by the ordering of the test statistics (common-cut-off MTPs) or unadjusted p-values (common-quantile MTPs).

- Step-down MTPs start with the most significant null hypothesis; as soon as one fails to reject a null hypothesis, no further hypotheses are rejected.
- Step-up MTPs start with the least significant null hypothesis; as soon as one null hypothesis is rejected, all remaining more significant hypotheses are rejected.

FWER-Controlling Step-Down Procedures

FWER-controlling step-down MTPs are similar in spirit to their above single-step counterparts, with the important step-down distinction that null hypotheses are considered successively, from most significant to least significant, with further tests depending on the outcome of earlier ones (van der Laan et al., 2004a).

Single-step maxT Procedure 7 and minP Procedure 10 are based, respectively, on the distributions of the maximum test statistic and minimum unadjusted p-value over all M null hypotheses.

In contrast, step-down maxT Procedure 11 and minP Procedure 12 are based, respectively, on the distributions of maxima of test statistics and minima of unadjusted p-values over successively smaller nested random subsets of ordered null hypotheses.

FWER-Controlling Step-Down Procedures

Procedure 11 [FWER-controlling step-down common-cut-off maxT procedure]
Let $O_n(m)$ denote the indices for the ordered test statistics, $T_n^\circ(m) \equiv T_n(O_n(m))$, so that $T_n(O_n(1)) \geq \cdots \geq T_n(O_n(M))$. The step-down maxT procedure is based on the distributions of maxima of test statistics over the nested subsets of ordered null hypotheses $\overline{\mathcal{O}}_n(h) \equiv \{O_n(h), \ldots, O_n(M)\}$. The adjusted p-values are given by

$$\widetilde{p}_{0n}(o_n(m)) = \max_{h=1,\ldots,m} \left\{ \Pr_{Q_0}\left(\max_{l \in \overline{\mathcal{O}}_n(h)} Z(l) \geq t_n(o_n(h)) \right) \right\}, \quad (29)$$

where $Z = (Z(m) : m = 1, \ldots, M) \sim Q_0$.

FWER-Controlling Step-Down Procedures

Procedure 12 [FWER-controlling step-down common-quantile minP procedure]
Let $O_n(m)$ denote the indices for the ordered unadjusted p-values, $P^{\circ}_{0n}(m) \equiv P_{0n}(O_n(m))$, so that $P_{0n}(O_n(1)) \leq \cdots \leq P_{0n}(O_n(M))$. The step-down minP procedure is based on the distributions of minima of unadjusted p-values over the nested subsets of ordered null hypotheses $\overline{\mathcal{O}}_n(h) \equiv \{O_n(h), \ldots, O_n(M)\}$. The adjusted p-values are given by

$$\widetilde{p}_{0n}(o_n(m)) = \max_{h=1,\ldots,m} \left\{ \Pr_{Q_0}\left(\min_{l \in \overline{\mathcal{O}}_n(h)} P_0(l) \leq p_{0n}(o_n(h)) \right) \right\}, \quad (30)$$

where $P_0(m) = \bar{Q}_{0,m}(Z(m))$ denote unadjusted p-values under the test statistics null distribution Q_0, i.e., for $Z = (Z(m) : m = 1, \ldots, M) \sim Q_0$.

Multiple Testing Procedures

Multiple testing problems currently encountered in biomedical and genomic research are characterized by a large number of hypotheses (in the thousands, or even millions), concerning high-dimensional multivariate distributions, with complex and unknown dependence structures among variables.

For instance, for the identification of differentially expressed or co-expressed genes, microarray datasets typically consist of thousands of expression measures for fewer than one hundred observational units.

MTPs controlling the proportion of false positives among the rejected hypotheses are particularly appealing for large-scale testing problems, compared to procedures controlling the number of false positives.

Multiple Testing Procedures

However, only a handful of approaches are currently available for controlling the proportion of false positives.

Furthermore, existing procedures tend to suffer from the following limitations.

- They are based solely on the marginal distributions of the test statistics, i.e., on unadjusted p-values.
- They rely on assumptions concerning the joint distribution of the test statistics, e.g., independence, Simes' Inequality.
- They err on the conservative side by using Bonferroni-like adjustments.

Multiple Testing Procedures

Motivated by these observations, we have developed two new classes of MTPs for controlling a broad class of Type I error rates, defined as generalized tail probabilities and expected values for arbitrary functions $g(V_n, R_n)$ of the numbers of Type I errors V_n and rejected hypotheses R_n.

- Augmentation multiple testing procedures (Dudoit et al., 2004a; van der Laan et al., 2004b).
- Resampling-based empirical Bayes procedures (Dudoit and van der Laan, 2007; van der Laan et al., 2005).

Unlike previously-proposed MTPs, our procedures take into account the *joint* distribution of the test statistics and provide Type I error control for general data generating distributions (with arbitrary dependence structures among variables), null hypotheses, and test statistics.

gTP-Controlling Augmentation Procedures

In order to control a new target Type I error rate, an augmentation multiple testing procedure (AMTP) adds suitably chosen null hypotheses to the set of hypotheses already rejected by an initial MTP (Dudoit et al., 2004a; van der Laan et al., 2004b).

Given any initial $gFWER$-controlling MTP, we have derived AMTPs for controlling generalized tail probability (gTP) error rates, $gTP(q, g) = \Pr(g(V_n, R_n) > q)$, for arbitrary functions $g(V_n, R_n)$ of the numbers of Type I errors V_n and rejected hypotheses R_n.

$gFWER(k_0)$	AMTP	$gTP(q, g)$
$\Pr(V_n > k_0) \leq \alpha$	$\Longrightarrow$	$\Pr(g(V_n, R_n) > q) \leq \alpha$

E.g. gFWER: $g(V_n, R_n) = V_n$, number of false positives; TPPFP: $g(V_n, R_n) = V_n/R_n$, proportion of false positives among the rejected hypotheses.

gTP-Controlling Augmentation Procedures

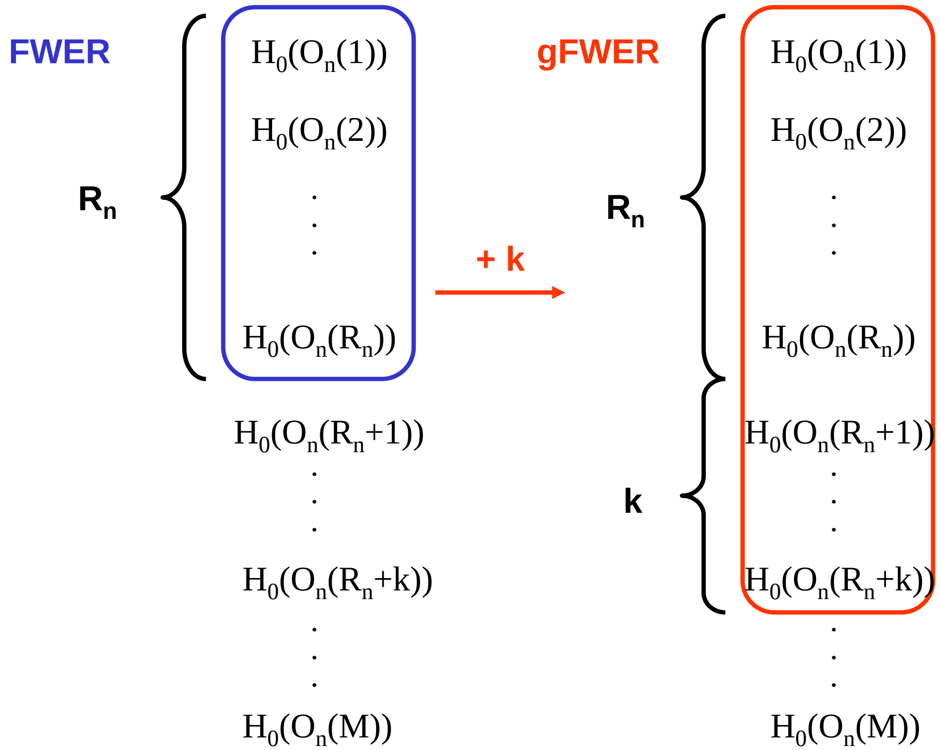


Figure 3: *Augmentation multiple testing procedures.* $gFWER(k)$-control via FWER control.

gTP-Controlling Augmentation Procedures

Procedure 13 [gFWER-controlling augmentation multiple testing procedure]
Consider any FWER-controlling procedure $\mathcal{R}_n(\alpha)$*, with adjusted p-values* $\widetilde{P}_{0n}(m)$ *and indices* $O_n(m)$ *so that* $\widetilde{P}_{0n}(O_n(1)) \leq \cdots \leq \widetilde{P}_{0n}(O_n(M))$*. This initial FWER-controlling procedure rejects the following* $R_n(\alpha) \equiv |\mathcal{R}_n(\alpha)|$ *null hypotheses,*

$$\mathcal{R}_n(\alpha) \equiv \left\{m : \widetilde{P}_{0n}(m) \leq \alpha\right\} = \{O_n(m) : m = 1, \ldots, R_n(\alpha)\}.$$

For controlling $gFWER(k)$ *at level* α*, the augmentation multiple testing procedure rejects the* $R_n(\alpha)$ *null hypotheses specified by the initial FWER-controlling MTP, as well as the next* $A_n(\alpha)$ *most*

significant null hypotheses, where

$$A_n(\alpha) \equiv \min\{k, M - R_n(\alpha)\}. \tag{31}$$

The set of rejected null hypotheses for the gFWER-controlling AMTP is

$$\mathcal{R}_n^+(\alpha) \equiv \{O_n(m) : m = 1, \ldots, R_n(\alpha) + A_n(\alpha)\} \tag{32}$$

and the adjusted p-values are

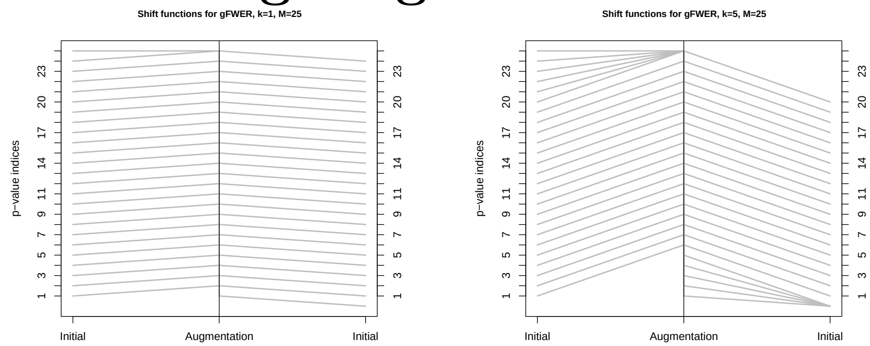
$$\widetilde{P}_{0n}^+(O_n(m)) = \begin{cases} 0, & \text{if } m \le k \\ \widetilde{P}_{0n}(O_n(m-k)), & \text{if } m > k \end{cases}. \tag{33}$$

gTP-Controlling Augmentation Procedures

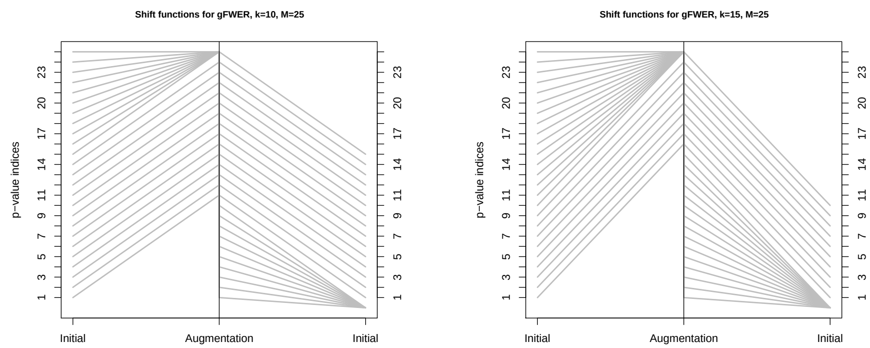
The AMTP guarantees at least k rejected null hypotheses.

The adjusted p-values for the AMTP are simply k-shifted versions of the adjusted p-values of the initial FWER-controlling MTP, with the first k adjusted p-values set to zero.

gTP-Controlling Augmentation Procedures



Panel (a): $k = 1$ Panel (b): $k = 5$



Panel (c): $k = 10$ Panel (d): $k = 15$

Figure 4: *gFWER-controlling AMTP.* Adjusted p-value shift and inverse shift functions for a $gFWER(k)$-controlling AMTP based on an initial FWER-controlling MTP, for the test of $M = 25$ null hypotheses.

gTP-Controlling Augmentation Procedures

Procedure 14 [TPPFP-controlling augmentation multiple testing procedure]
Consider any FWER-controlling procedure $\mathcal{R}_n(\alpha)$*, with adjusted p-values* $\widetilde{P}_{0n}(m)$ *and indices* $O_n(m)$ *so that* $\widetilde{P}_{0n}(O_n(1)) \leq \cdots \leq \widetilde{P}_{0n}(O_n(M))$*. This initial FWER-controlling procedure rejects the following* $R_n(\alpha) \equiv |\mathcal{R}_n(\alpha)|$ *null hypotheses,*

$$\mathcal{R}_n(\alpha) \equiv \left\{ m : \widetilde{P}_{0n}(m) \leq \alpha \right\} = \{O_n(m) : m = 1, \ldots, R_n(\alpha)\}.$$

For controlling $TPPFP(q)$ *at level* α*, the augmentation multiple testing procedure rejects the* $R_n(\alpha)$ *null hypotheses specified by the initial FWER-controlling MTP, as well as the next* $A_n(\alpha)$ *most*

significant null hypotheses, where

$$\begin{aligned} A_n(\alpha) &\equiv \max\left\{m \in \{0,\ldots,M-R_n(\alpha)\} : \frac{m}{R_n(\alpha)+m} \leq q\right\} \\ &= \min\left\{\left\lfloor \frac{qR_n(\alpha)}{1-q} \right\rfloor, M - R_n(\alpha)\right\}. \end{aligned} \tag{34}$$

The set of rejected null hypotheses for the TPPFP-controlling AMTP is

$$\mathcal{R}_n^+(\alpha) \equiv \{O_n(m) : m = 1,\ldots,R_n(\alpha)+A_n(\alpha)\} \tag{35}$$

and the adjusted p-values are

$$\widetilde{P}_{0n}^+(O_n(m)) = \widetilde{P}_{0n}(O_n(\lceil (1-q)m \rceil)), \qquad m = 1,\ldots,M, \tag{36}$$

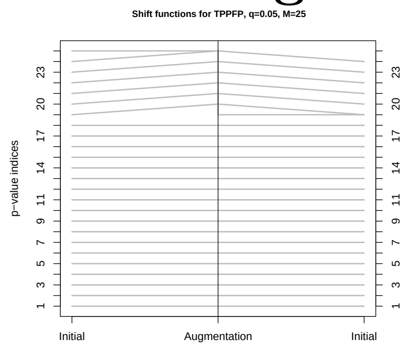
where the ceiling $\lceil x \rceil$ *denotes the least integer greater than or equal to* x*, i.e.,* $\lceil x \rceil \in \mathbb{Z}$ *and* $\lceil x \rceil - 1 < x \leq \lceil x \rceil$*.*

gTP-Controlling Augmentation Procedures

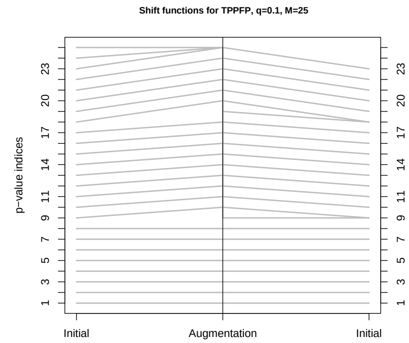
The AMTP keeps rejecting null hypotheses until the ratio of additional rejections to the total number of rejections reaches the allowed proportion q of false positives.

The adjusted p-values for the AMTP are simply mq-shifted versions (up to a ceiling integer transformation) of the adjusted p-values of the initial FWER-controlling MTP.

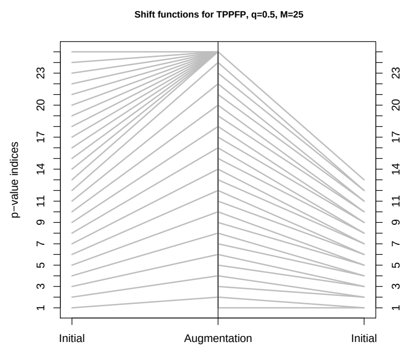
gTP-Controlling Augmentation Procedures



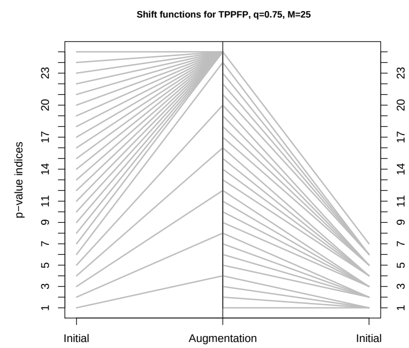
Panel (a): $q = 0.05$



Panel (b): $q = 0.10$



Panel (c): $q = 0.50$



Panel (d): $q = 0.75$

Figure 5: *TPPFP-controlling AMTP.* Adjusted p-value shift and inverse shift functions for a $TPPFP(q)$-controlling AMTP based on an initial FWER-controlling MTP, for the test of $M = 25$ null hypotheses.

gTP-Controlling Augmentation Procedures

Consider a multiple testing procedure $\mathcal{R}_n(\alpha)$, that controls the generalized family-wise error rate at level α, i.e., such that $gFWER(k_0) = \Pr(V_n(\alpha) > k_0) \leq \alpha$.

Given the initial MTP $\mathcal{R}_n(\alpha)$, we wish to derive an augmentation multiple testing procedure $\mathcal{R}_n^+(\alpha)$, that controls the generalized tail probability error rate also at level α, i.e., such that $gTP(q, g) = \Pr(g(V_n^+(\alpha), R_n^+(\alpha)) > q) \leq \alpha$.

gTP-Controlling Augmentation Procedures

Consider a function g and an initial MTP $\mathcal{R}_n(\alpha)$ that satisfy the following three assumptions.

Assumption AMTP.MgV. **[Monotonicity of g]** The function $v \rightarrow g(v, r)$ is non-decreasing for any given r.

Assumption AMTP.MgA. **[Monotonicity of g]** The function $a \rightarrow g(v + a, r + a)$ is non-decreasing for any given (v, r).

Assumption AMTP.gTP0. **[Initial gTP control]** One has

$$\Pr(g(k_0, R_n(\alpha)) \leq q) = 1, \tag{37}$$

so that the initial gFWER-controlling MTP also controls the target gTP error rate.

gTP-Controlling Augmentation Procedures

Monotonicity Assumption AMTP.MgV is used to prove gTP control by an AMTP such as Procedure 15, below.

Monotonicity Assumption AMTP.MgA guarantees that the cardinality of the augmentation set for Procedure 15 increases with the bound q for false positives (Equation (38)).

Finally, Assumption AMTP.gTP0 ensures that the initial gFWER-controlling procedure also controls the target gTP error rate at level α.

gTP-Controlling Augmentation Procedures

Procedure 15 [gTP-controlling augmentation multiple testing procedure]
Consider any $gFWER(k_0)$-controlling procedure $\mathcal{R}_n(\alpha)$, with adjusted p-values $\widetilde{P}_{0n}(m)$ and indices $O_n(m)$ so that $\widetilde{P}_{0n}(O_n(1)) \leq \cdots \leq \widetilde{P}_{0n}(O_n(M))$. This initial gFWER-controlling procedure rejects the following $R_n(\alpha) \equiv |\mathcal{R}_n(\alpha)|$ null hypotheses,

$$\mathcal{R}_n(\alpha) \equiv \left\{m : \widetilde{P}_{0n}(m) \leq \alpha\right\} = \{O_n(m) : m = 1, \ldots, R_n(\alpha)\}.$$

For controlling $gTP(q, g)$ at level α, the augmentation multiple testing procedure rejects the $R_n(\alpha)$ null hypotheses specified by the initial gFWER-controlling MTP, as well as the next $A_n(\alpha)$ most

significant null hypotheses, where

$$A_n(\alpha) \equiv \max\{m \in \{0,\ldots,M-R_n(\alpha)\} : g(k_0+m, R_n(\alpha)+m) \leq q\}. \tag{38}$$

The set of rejected null hypotheses for the gTP-controlling AMTP is

$$\mathcal{R}_n^+(\alpha) \equiv \{O_n(m) : m = 1,\ldots,R_n(\alpha)+A_n(\alpha)\} \tag{39}$$

and the adjusted p-values satisfy

$$\widetilde{P}_{0n}(O_n(m)) = \widetilde{P}_{0n}^+(O_n(S_n(m))), \tag{40}$$

where $S_n : \{1,\ldots,M\} \rightarrow \{1,\ldots,M\}$ *is an integer shift function defined by*

$$S_n(m) \equiv R_n^+(\widetilde{P}_{0n}(O_n(m))) = m + A_n(\widetilde{P}_{0n}(O_n(m))). \tag{41}$$

gTP-Controlling Augmentation Procedures

The AMTP keeps rejecting null hypotheses until $g(k_0 + m, R_n + m)$ reaches the bound q for false positives.

The adjusted p-values for the AMTP are shifted versions of the adjusted p-values of the initial gFWER-controlling MTP.

gTP-Controlling Augmentation Procedures

- Any gFWER-controlling procedure (marginal/joint single-step/stepwise common-cut-off/common-quantile) provides immediately and trivially multiple testing procedures controlling a broad class of Type I error rates, such as the gFWER and TPPFP.
- One can build on the large pool of available FWER-controlling MTPs, such as the single-step and step-down maxT and minP procedures.
- Adjusted p-values for an AMTP are simply shifted versions of the ordered adjusted p-values for the initial MTP.

gTP-Controlling Augmentation Procedures

- $gFWER(k)$-controlling augmentation Procedure 13 guarantees at least k rejected null hypotheses.
- AMTPs augment the set of null hypotheses rejected by an initial MTP conservatively, in the sense that every additional rejected hypothesis is counted as a false positive.
- Unlike existing procedures controlling the proportion of false positives, which assume either independence or specific dependence structures for the joint distribution of the test statistics, our AMTPs provide gTP control for general data generating distributions, i.e., arbitrary joint distributions for the test statistics.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

Augmentation multiple testing procedures (AMTP) provide a simple and general approach for controlling generalized tail probability error rates, that can account for the joint distribution of the test statistics.

However, even joint AMTPs tend to be conservative in finite sample situations, as they count every additional rejected hypothesis as a false positive.

The simulation studies in Dudoit et al. (2004a) and van der Laan et al. (2005) suggest that, although AMTPs compare favorably to TPPFP-controlling marginal procedures, they become more conservative as the number of tested hypotheses increases.

The latter feature is problematic for the large-scale testing questions commonly-encountered in genomics.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

Motivated by these observations, van der Laan et al. (2005) propose a new resampling-based empirical Bayes approach for controlling TPPFP, which, as does the augmentation method, provides asymptotic Type I error control for general data generating distributions, but is less conservative for finite samples.

Dudoit and van der Laan (2007) extend the TPPFP-controlling approach of van der Laan et al. (2005) to a broad class of Type I error rates, defined as generalized tail probability (gTP) error rates, $gTP(q, g) = \Pr(g(V_n, S_n) > q)$, for arbitrary functions $g(V_n, S_n)$ of the numbers of false positives V_n and true positives $S_n = R_n - V_n$.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

The gTP-controlling resampling-based empirical Bayes procedure detailed below involves specifying:

- a null distribution Q_0 (or estimator thereof, Q_{0n}) for M-vectors of null test statistics T_{0n};
- a distribution $Q_{0n}^{\mathcal{H}}$ for random guessed sets of true null hypotheses $\mathcal{H}_{0n}$.

By randomly sampling null test statistics T_{0n} and guessed sets of true null hypotheses $\mathcal{H}_{0n}$, one obtains a distribution for a random variable $G(c; \mathcal{H}_{0n}, T_{0n}, T_n)$ representing the guessed g-specific function of the numbers of false positives and true positives (given the empirical distribution P_n), for any given cut-off vector c.

Cut-offs can then be chosen to control tail probabilities for this distribution at a user-supplied level α.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

Consider functions g that satisfy the following two monotonicity assumptions.

Assumption EB.MgV. [Monotonicity of g] The function $g_{S=s} : v \rightarrow g(v, s)$ is continuous and strictly increasing for any given s.

Assumption EB.MgS. [Monotonicity of g] The function $g_{V=v} : s \rightarrow g(v, s)$ is continuous and non-increasing for any given v.

The functions $g(v, s) = v$ and $g(v, s) = v/(v + s)$, defining, respectively, the gFWER and TPPFP, clearly satisfy these two monotonicity assumptions.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

Define the following g-specific function of the numbers of false positives and true positives,

$$\begin{aligned} G(c;\mathcal{H},Z_0,Z) &\equiv g(V(c;\mathcal{H},Z_0),S(c;\mathcal{H},Z)) \qquad (42) \\ &\equiv g\left(\sum_{m\in\mathcal{H}} \mathrm{I}\left(Z_0(m)>c(m)\right), \sum_{m\notin\mathcal{H}} \mathrm{I}\left(Z(m)>c(m)\right)\right), \end{aligned}$$

where $c=(c(m):m=1,\ldots,M)\in I\!R^M$ denotes an M-dimensional cut-off vector that defines one-sided rejection regions of the form $\mathcal{C}(m)=(c(m),+\infty)$, $\mathcal{H}\subseteq\{1,\ldots,M\}$ denotes a set of null hypotheses, and $Z_0=(Z_0(m):m=1,\ldots,M)$ and $Z=(Z(m):m=1,\ldots,M)$ denote random M-vectors of test statistics.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

Given a user-supplied Type I error level $\alpha \in (0, 1)$, one seeks cut-offs $c_n = (c_n(m) : m = 1, \ldots, M)$, for the test statistics $T_n = (T_n(m) : m = 1, \ldots, M) \sim Q_n$, so that $gTP(q, g) = \Pr(g(V_n, S_n) > q)$ is controlled at level α. That is, one seeks cut-offs that satisfy the gTP Type I error constraint

$$
\begin{aligned}
\Pr(G(c_n; \mathcal{H}_0, T_n, T_n) > q) &\leq \alpha \qquad \text{[finite sample control]} \\
\limsup_{n \to \infty} \Pr(G(c_n; \mathcal{H}_0, T_n, T_n) > q) &\leq \alpha \qquad \text{[asymptotic control].}
\end{aligned}
\tag{43}
$$

However, one is immediately faced with the problem that the distribution of $G(c_n; \mathcal{H}_0, T_n, T_n) = g(V_n, S_n)$ depends on the unknown data generating distribution P.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

The gTP-controlling resampling-based empirical Bayes approach replaces the true unknown g-specific function of the numbers of false positives and true positives $G(c; \mathcal{H}_0, T_n, T_n)$ by the corresponding guessed function $G(c; \mathcal{H}_{0n}, T_{0n}, T_n)$, where $T_n \sim Q_n$ is the M-vector of observed test statistics, $T_{0n} \sim Q_{0n}$ is an M-vector of null test statistics, and $\mathcal{H}_{0n} \sim Q_{0n}^{\mathcal{H}}$ is a guessed set of true null hypotheses.

The null test statistics T_{0n} and the guessed sets $\mathcal{H}_{0n}$ are sampled independently, given the empirical distribution P_n, from distributions Q_{0n} and $Q_{0n}^{\mathcal{H}}$, chosen conservatively so that the guessed function $G(c; \mathcal{H}_{0n}, T_{0n}, T_n)$ is asymptotically stochastically greater than the corresponding true function $G(c; \mathcal{H}_0, T_n, T_n)$.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

Common marginal non-parametric mixture model. A proposed working model for the distribution pair $(Q_{0n}, Q_{0n}^{\mathcal{H}})$ posits M identically distributed pairs of test statistics and null hypotheses $((T_n(m), H_0(m)) : m = 1, \ldots, M)$.

Test statistics are assumed to have the following common marginal non-parametric mixture distribution,

$$T_n(m) \sim f \equiv \pi_0 f_0 + (1 - \pi_0) f_1, \qquad m = 1, \ldots, M, \tag{44}$$

where $\pi_0 \equiv \Pr(H_0(m) = 1)$ denotes the prior probability of a true null hypothesis, f_0 the marginal null density of the test statistics, and f_1 the marginal alternative density of the test statistics, i.e., $T_n(m)|\{H_0(m) = 1\} \sim f_0$ and $T_n(m)|\{H_0(m) = 0\} \sim f_1$.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

A parameter of interest, for generating guessed sets of true null hypotheses under the marginal non-parametric mixture model of Equation (44), is the local q-value function, i.e., the posterior probability function for a true null hypothesis $H_0(m)$, given the corresponding test statistic $T_n(m)$,

$$\pi_0(t) \equiv \Pr(H_0(m) = 1 | T_n(m) = t) = \frac{\pi_0 f_0(t)}{f(t)}, \qquad m = 1, \ldots, M. \tag{45}$$

Empirical Bayes q-values are similar in spirit to frequentist p-values: the smaller the q-value $\pi_0(T_n(m))$, the stronger the evidence against the corresponding null hypothesis $H_0(m)$.

Approaches for estimating local q-value functions are detailed in Dudoit and van der Laan (2007, Chapter 7).

gTP-Controlling Resampling-Based Empirical Bayes Procedures

Procedure 16 [gTP-controlling resampling-based empirical Bayes procedure]

Consider the simultaneous test of M null hypotheses $H_0(m)$, $m = 1, \ldots, M$, based on an M-vector of test statistics $T_n = (T_n(m) : m = 1, \ldots, M)$, with true distribution $Q_n = Q_n(P)$. Given a function g, that satisfies monotonicity Assumptions EB.MgV *and* EB.MgS*, and a Type I error bound q, the following resampling-based empirical Bayes procedure may be used to control the generalized tail probability error rate, $gTP(q, g) = \Pr(g(V_n, S_n) > q)$.*

1. *Generate B pairs, $\{(T_{0n}^b, \mathcal{H}_{0n}^b) : b = 1, \ldots, B\}$, of null test statistics T_{0n}^b and random guessed sets $\mathcal{H}_{0n}^b$ of true null hypotheses, as follows.*

   (a) *The M-vectors of null test statistics T_{0n}^b have a null distri-*

bution Q_{0n}, such as the bootstrap distributions of Procedures 3 and 4, i.e., T^b_{0n} is a column of an $M \times B$ matrix $\mathbf{Z}^B_n$ of null-transformed bootstrap test statistics.

(b) *The random guessed sets of true null hypotheses $\mathcal{H}^b_{0n}$ have a distribution $Q^{\mathcal{H}}_{0n}$ that corresponds to M independent Bernoulli random variables with parameters $\pi_{0n}(T_n(m))$. That is, generate binary random M-vectors $H^b_{0n} = (H^b_{0n}(m) : m = 1, \ldots, M)$ of null hypotheses as*

$$H^b_{0n}(m) \stackrel{\perp}{\sim} Bernoulli\left(\pi_{0n}(T_n(m))\right), \qquad m = 1, \ldots, M, \tag{46}$$

and define sets

$$\mathcal{H}^b_{0n} \equiv \left\{ m : H^b_{0n}(m) = 1 \right\}. \tag{47}$$

Here, $\pi_{0n}(t)$ is an estimated true null hypothesis posterior probability function, such as the estimated local q-value *function*

$$\pi_{0n}(t) = \min\left\{1, \frac{\pi_{0n} f_{0n}(t)}{f_n(t)}\right\}, \tag{48}$$

corresponding to the marginal non-parametric mixture model of Equation (44)*.*

(c) *Null test statistics* T^b_{0n} *and guessed sets* $\mathcal{H}^b_{0n}$ *are independent, given the empirical distribution* P_n*.*

2. *For any given test statistic cut-off vector* $c = (c(m) : m = 1, \ldots, M)$*, compute, for each of the* B *pairs* $(T^b_{0n}, \mathcal{H}^b_{0n})$*, the corresponding guessed g-specific function of the numbers of false positives and true positives,*

$$\begin{aligned} G^b_n(c) &\equiv G(c; \mathcal{H}^b_{0n}, T^b_{0n}, T_n) \\ &= g(V(c; \mathcal{H}^b_{0n}, T^b_{0n}), S(c; \mathcal{H}^b_{0n}, T_n)). \end{aligned} \tag{49}$$

3. *For user-supplied Type I error bound* q *and Type I error level*

$\alpha \in (0,1)$, *derive a cut-off vector* c_n *that satisfies the empirical gTP Type I error constraint*

$$\frac{1}{B}\sum_{b=1}^{B} \mathrm{I}\left(G_n^b(c_n) > q\right) \leq \alpha. \tag{50}$$

Common-cut-off procedure. *The common cut-off* γ_n *is the smallest (i.e., least conservative) value* γ *for which the gTP constraint in Equation* (50) *is satisfied. That is,*

$$\gamma_n \equiv \inf\left\{\gamma \in I\!R : \frac{1}{B}\sum_{b=1}^{B} \mathrm{I}\left(G_n^b(\gamma^{(M)}) > q\right) \leq \alpha\right\}, \tag{51}$$

where $\gamma^{(M)}$ *denotes the* M*-vector with all elements equal to* γ*, i.e.,* $\gamma^{(M)}(m) = \gamma$, $\forall\, m = 1, \ldots, M$.

The adjusted p-values may be approximated as

$$\widetilde{p}_{0n}(o_n(m)) \approxeq \min_{h \in \overline{\mathcal{O}}_n(m)} \frac{1}{B} \sum_{b=1}^{B} \mathrm{I}\left(G_n^b((t_n(h))^{(M)}) > q \right), \quad (52)$$

where $O_n(m)$ denote the indices for the ordered test statistics $T_n(O_n(m))$, so that $T_n(O_n(1)) \geq \cdots \geq T_n(O_n(M))$, and $\overline{\mathcal{O}}_n(m) \equiv \{O_n(m), \ldots, O_n(M)\}$.

Common-quantile procedure. *The common quantile probability δ_n, corresponding to the test statistics null distribution Q_{0n}, is the smallest (i.e., least conservative) value δ for which the gTP constraint in Equation (50) is satisfied. That is,*

$$\delta_n \equiv \inf \left\{ \delta \in [0,1] : \frac{1}{B} \sum_{b=1}^{B} \mathrm{I}\left(G_n^b(q_{0n}^{-1}(\delta)) > q \right) \leq \alpha \right\}, \quad (53)$$

where $q_{0n}^{-1}(\delta) = (Q_{0n,m}^{-1}(\delta) : m = 1, \ldots, M)$ denotes the M-

vector of δ-quantiles for the null distribution Q_{0n}.
The adjusted p-values may be approximated as

$$\widetilde{p}_{0n}(o_n(m)) \approx \min_{h \in \overline{\mathcal{O}}_n(m)} \frac{1}{B} \sum_{b=1}^{B} \mathrm{I}\left(G_n^b(q_{0n}^{-1}(1 - p_{0n}(h))) > q\right), \tag{54}$$

where $p_{0n}(m) = 1 - Q_{0n,m}(t_n(m))$ is the unadjusted p-value for null hypothesis $H_0(m)$, $O_n(m)$ denote the indices for the ordered unadjusted p-values $P_{0n}(O_n(m))$, so that $P_{0n}(O_n(1)) \leq \cdots \leq P_{0n}(O_n(M))$, and $\overline{\mathcal{O}}_n(m) \equiv \{O_n(m), \ldots, O_n(M)\}$.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

gTP-controlling resampling-based empirical Bayes Procedure 16 is very general.

- It controls generalized tail probability error rates for arbitrary functions $g(V_n, S_n)$ of the numbers of false positives and true positives.
- Unlike most MTPs controlling the proportion of false positives, it provides Type I error control for general data generating distributions, with arbitrary dependence structures among variables.
- It can be applied to any distribution pair $(Q_{0n}, Q_{0n}^{\mathcal{H}})$ for the null test statistics and guessed sets of true null hypotheses, i.e., the common marginal non-parametric mixture model of Equation (44) is only one among many reasonable working models that does not assume independence of the test statistics.

gTP-Controlling Resampling-Based Empirical Bayes Procedures

The simulation studies of van der Laan et al. (2005) reveal that TPPFP-controlling resampling-based empirical Bayes procedures tend to be more powerful than existing TPPFP-controlling MTPs.

Note that the empirical Bayes approach can be extended to control a broader class of Type I error rates, defined as parameters $\Theta(F_{g(V_n,S_n)})$ of the distribution of functions $g(V_n, S_n)$ of the numbers of false positives V_n and true positives S_n, where the function g satisfies monotonicity Assumptions EB.MgV and EB.MgS and the mapping Θ satisfies monotonicity Assumption MΘ.

Such error rates include generalized expected value (gEV) error rates, $gEV(g) = \mathrm{E}[g(V_n, S_n)]$, and, in particular, the false discovery rate, with $g(v, s) = v/(v + s)$.

Details in Dudoit and van der Laan (2007, Chapter 7).

Appendix: Miscellaneous Mathematical and Statistical Results

Consider J-dimensional random vectors $\{X_n\}$ and X, with respective joint cumulative distribution functions (CDF) $\{F_n\}$ and F, defined as

$$
\begin{aligned}
F_n(x) &\equiv \Pr(X_n \in (-\infty, x]) = \Pr(\cap_j \{X_n(j) \leq x(j)\}) \quad (55)\\
F(x) &\equiv \Pr(X \in (-\infty, x]) = \Pr(\cap_j \{X(j) \leq x(j)\}),
\end{aligned}
$$

for $x \in I\!R^J$ and corresponding J-dimensional rectangle

$$
\begin{aligned}
(-\infty, x] &\equiv \prod_j (-\infty, x(j)] = (-\infty, x(1)] \times \cdots \times (-\infty, x(J)] \\
&= \{y \in I\!R^J : y(j) \leq x(j),\ j = 1, \ldots, J\}.
\end{aligned}
$$

Appendix: Miscellaneous Mathematical and Statistical Results

Theorem 17 [Weak convergence] *The following are equivalent definitions of weak convergence of the sequence $\{X_n\}$ to X.*
(i) For every bounded continuous function $\ell : \mathbb{R}^J \rightarrow \mathbb{R}$,

$$\lim_{n\rightarrow\infty} \mathrm{E}[\ell(X_n)] = \mathrm{E}[\ell(X)]. \tag{56}$$

(ii) For every continuity point x of F, that is, for every $x \in \mathbb{R}^J$ such that the J-dimensional rectangle $(-\infty, x]$ has null boundary probability $\Pr(X \in \partial(-\infty, x])$ under F,

$$\lim_{n\rightarrow\infty} F_n(x) = F(x). \tag{57}$$

[Theorem 29.1, p. 390, Billingsley (1986)]

Weak convergence may also be referred to as convergence in distribution or convergence in law and may be denoted by $X_n \stackrel{\mathcal{L}}{\Rightarrow} X$, $X_n \Rightarrow X$, or $F_n \Rightarrow F$.

Appendix: Miscellaneous Mathematical and Statistical Results

Proposition 1 [Weak convergence for discrete distributions] *Consider discrete integer-valued random variables* $\{X_n\}$ *and* X*, with respective cumulative distribution functions* $\{F_n\}$ *and* F*. If the sequence* $\{X_n\}$ *converges weakly to* X*, then* $\lim_n F_n(x) = F(x)$ *for every* $x \in I\!R$*.*

Theorem 18 [Continuous Mapping Theorem] *Consider a function* $\ell : I\!R^J \rightarrow I\!R^K$*, with discontinuity set* $\mathcal{D}_\ell \equiv \{x \in I\!R^J : \ell \textit{ is discontinuous at } x\}$*. If* $X_n \stackrel{\mathcal{L}}{\Rightarrow} X$ *and* $\Pr(X \in \mathcal{D}_\ell) = 0$*, then*

$$\ell(X_n) \stackrel{\mathcal{L}}{\Rightarrow} \ell(X). \tag{58}$$

[Theorem 29.2, p. 391, Billingsley (1986)]

Appendix: Miscellaneous Mathematical and Statistical Results

Theorem 19 [Central Limit Theorem] *Suppose $\{X_n\}$ is a sequence of independent and identically distributed (IID) J-dimensional random vectors, with J-dimensional mean vector μ and $J \times J$ covariance matrix σ. Then,*

$$\sqrt{n}(\bar{X}_n - \mu) \stackrel{\mathcal{L}}{\Rightarrow} N(0, \sigma), \tag{59}$$

where $\bar{X}_n \equiv \sum_i X_i/n$ denotes the empirical mean vector. [Theorem 2.9.1, p. 51, Mardia et al. (1979)]

References

P. Billingsley. *Probability and Measure*. Probability and Mathematical Statistics. Wiley, New York, 2nd edition, 1986.

S. Dudoit and M. J. van der Laan. *Multiple Testing Procedures with Applications to Genomics*. Springer, New York, 2007. (In preparation).

S. Dudoit, M. J. van der Laan, and M. D. Birkner. Multiple testing procedures for controlling tail probability error rates. Technical Report 166, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7360, 2004a. URL `www.bepress.com/ucbbiostat/paper166`.

S. Dudoit, M. J. van der Laan, and K. S. Pollard. Multiple testing. Part I. Single-step procedures for control of general Type I error rates. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 13, 2004b. URL `www.bepress.com/sagmb/vol3/iss1/art13`.

Y. Hochberg and A. C. Tamhane. *Multiple Comparison Procedures*. Probability and Mathematical Statistics. Wiley, New York, 1987.

K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, London, New York, 1979.

K. S. Pollard and M. J. van der Laan. Choice of a null distribution in resampling-based multiple testing. *Journal of Statistical Planning and Inference*, 125(1–2):85–100, 2004.

M. J. van der Laan and A. E. Hubbard. Quantile-function based null distribution in resampling based multiple testing. *Statistical Applications in Genetics and Molecular Biology*, 5(1):Article 14, 2006. URL `www.bepress.com/sagmb/vol5/iss1/art14`.

M. J. van der Laan, S. Dudoit, and K. S. Pollard. Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 14, 2004a. URL `www.bepress.com/sagmb/vol3/iss1/art14`.

M. J. van der Laan, S. Dudoit, and K. S. Pollard. Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 15, 2004b. URL `www.bepress.com/sagmb/vol3/iss1/art15`.

M. J. van der Laan, M. D. Birkner, and A. E. Hubbard. Empirical Bayes and resampling based multiple testing procedure controlling tail probability of the proportion of false positives. *Statistical Applications in Genetics and Molecular Biology*, 4(1):Article 29, 2005. URL `www.bepress.com/sagmb/vol4/iss1/art29`.

P. H. Westfall and S. S. Young. *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment*. Wiley, New York, 1993.

Z. Yu and M. J. van der Laan. Construction of counterfactuals and the G-computation formula. Technical Report 122, Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720-7360, 2002. URL `www.bepress.com/ucbbiostat/paper122`.