2

Test Statistics Null Distribution

2.1 Introduction

2.1.1 Motivation

A key feature of our proposed multiple testing procedures (MTP) is the *test statistics null distribution* (rather than a data generating null distribution) used to obtain rejection regions (i.e., cut-offs) for the test statistics, confidence regions for the parameters of interest, and adjusted p-values. Indeed, whether testing single or multiple hypotheses, one needs the (joint) distribution of the test statistics in order to derive a procedure that probabilistically controls Type I errors. In practice, however, the true distribution of the test statistics is unknown and replaced by a null distribution. 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*. This issue is particularly relevant for large-scale testing problems, such as those encountered in biomedical and genomic research (Chapters 9–12), which concern high-dimensional multivariate distributions, with complex and unknown dependence structures among variables.

Common approaches use a data generating distribution, such as a permutation distribution, that satisfies the *complete null hypothesis* that *all* null hypotheses are true. Procedures based on such a *data generating null distribution* typically rely on the *subset pivotality* assumption, stated in Westfall and Young (1993, p. 42–43), to ensure that Type I error control under the data generating null distribution leads to the desired control under the true data generating distribution. However, subset pivotality is violated in many important testing problems, because a data generating null distribution may result in a joint distribution for the test statistics that has a different dependence structure than their true distribution. In fact, in most problems, there does not exist a data generating null distribution that correctly specifies the joint distribution of the test statistics corresponding to the true null hypotheses.

Indeed, subset pivotality fails for two types of testing problems that are highly relevant in biomedical and genomic data analysis: tests concerning correlation coefficients and tests concerning regression coefficients (Chapter 8; Pollard et al. (2005a); Pollard and van der Laan (2004)).

We have formulated a general characterization of a test statistics null distribution for which the multiple testing procedures of Chapters 3–7 provide proper Type I error control. Our general characterization is based on the intuitive notion of *null domination*, whereby the number of Type I errors is stochastically greater under the test statistics' null distribution than under their true distribution. Null domination conditions lead us to the explicit construction of two main types of test statistics null distributions. The first original proposal of Dudoit et al. (2004b), van der Laan et al. (2004a), and Pollard and van der Laan (2004), defines the null distribution as the asymptotic distribution of a vector of *null shift and scale-transformed test statistics*, based on user-supplied upper bounds for the means and variances of the test statistics for the true null hypotheses. The second and most recent proposal of van der Laan and Hubbard (2006) defines the null distribution as the asymptotic distribution of a vector of *null quantile-transformed test statistics*, based on user-supplied marginal test statistics null distributions. Resampling procedures (e.g., non-parametric or model-based bootstrap) are provided to conveniently obtain consistent estimators of the null distribution and of the corresponding test statistic cut-offs, parameter confidence regions, and adjusted p-values.

We stress the generality of these 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 (defined in terms of submodels for the data generating distribution), and test statistics (e.g., t-statistics, χ^2-statistics, F-statistics). In particular, the proposed null distributions allow one to address testing problems that cannot be handled by existing approaches, such as tests concerning correlation coefficients and parameters in general regression models (e.g., linear regression models where the covariates and error terms are allowed to be dependent, logistic regression models, Cox proportional hazards models; Chapter 8; Pollard et al. (2005a)). The latest proposal of van der Laan and Hubbard (2006) 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).

As illustrated in the simulation studies of Chapter 8 and articles by van der Laan and Hubbard (2006), Pollard et al. (2005a), and Pollard and van der Laan (2004), the choice of null distribution can have a substantial impact on the Type I error and power properties of a given multiple testing procedure. In particular, Pollard et al. (2005a) show that procedures based on our general non-parametric bootstrap null shift and scale-transformed test statistics null distribution typically control the Type I error rate "on target" at the nominal level. In contrast, comparable procedures, based on parameter-specific bootstrap data generating null distributions, can be severely anti-conservative (bootstrapping residuals for testing regression coefficients) or conservative (independent bootstrap for testing correlation coefficients). van der Laan and Hubbard (2006) further illustrate that, for finite samples, the new null quantile-transformed test statistics null distribution provides more accurate Type I error control and is more powerful than the original null shift and scale-transformed null distribution.

Finally, note that the null shift and scale-transformed and null quantile-transformed test statistics null distributions are only two among a family of null distributions that satisfy null domination conditions for a given testing problem. The explicit construction of null distributions with good Type I error control and power properties still represents an open and important research avenue.

2.1.2 Outline

Section 2.2 outlines the main features of our approach to Type I error control and the key choice of a test statistics null distribution based on the notion of null domination. Section 2.3 discusses in detail our first proposal of a null shift and scale-transformed test statistics null distribution. Section 2.4 introduces our most recent null quantile-transformed test statistics null distribution. Section 2.5 considers the choice of a null distribution for transformations of the test statistics, such as the absolute value transformation. Sections 2.6 and 2.7 focus on two particular examples of testing problems covered by our framework: the test of single-parameter null hypotheses using t-statistics (e.g., tests of means, correlation coefficients, regression coefficients in linear and non-linear models) and the test of multiple-parameter null hypotheses using F-statistics. The last two sections are devoted to contrasting our proposed methodology with existing approaches. Specifically, Section 2.8 revisits the notions of weak and strong control of a Type I error rate and the related assumption of subset pivotality. We stress that such conditions are made irrelevant by our general approach, which is only concerned with control of the Type I error rate under the *true data generating distribution* and is based on a *test statistics null distribution* rather than a data generating null distribution. Finally, Section 2.9 examines test statistics null distributions based on bootstrap and permutation data generating distributions.

2.2 Type I error control and choice of a test statistics null distribution

2.2.1 Type I error control

As in Section 1.2, consider the simultaneous test of M null hypotheses $H_0(m)$, $m = 1, \ldots, M$, based on test statistics $T_n = (T_n(m) : m = 1, \ldots, M)$, with finite sample joint distribution $Q_n = Q_n(P)$, under the data generating distribution P. We wish to derive rejection regions for the test statistics $T_n(m)$, such that Type I errors are probabilistically controlled at a user-supplied level α (see Section 1.2.9 for definitions of Type I error rates). In practice, however, the true distribution $Q_n(P)$ of the test statistics is unknown and replaced by a null distribution Q_0 (or estimator thereof, Q_{0n}). As in Section 1.2.6, let $\mathcal{C}_n(m) = \mathcal{C}(m; T_n, Q_0, \alpha)$, $m = 1, \ldots, M$, and $\mathcal{R}_n = \mathcal{R}(T_n, Q_0, \alpha)$ denote, respectively, the M rejection regions and corresponding set of rejected null hypotheses, for a MTP with nominal Type I error level α. That is,

$$\mathcal{R}(T_n, Q_0, \alpha) = \{m : T_n(m) \in \mathcal{C}(m; T_n, Q_0, \alpha)\}\,. \tag{2.1}$$

Given a random M-vector $Z = (Z(m) : m = 1, \ldots, M)$, with joint distribution Q, and a collection of M rejection regions $\mathcal{C} = \{\mathcal{C}(m) : m = 1, \ldots, M\}$[1], denote the numbers of rejected hypotheses and Type I errors by

$$R(\mathcal{C}|Q) \equiv \sum_{m=1}^{M} \mathrm{I}\left(Z(m) \in \mathcal{C}(m)\right) \tag{2.2}$$

and

$$V(\mathcal{C}|Q) \equiv \sum_{m \in \mathcal{H}_0} \mathrm{I}\left(Z(m) \in \mathcal{C}(m)\right),$$

respectively. For given rejection regions $\mathcal{C}$, 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,

$$\begin{aligned} R_n &\equiv R(\mathcal{C}|Q_n), & R_0 &\equiv R(\mathcal{C}|Q_0), \\ V_n &\equiv V(\mathcal{C}|Q_n), & V_0 &\equiv V(\mathcal{C}|Q_0). \end{aligned} \tag{2.3}$$

For one-sided rejection regions of the form $\mathcal{C}(m) = (c(m), +\infty)$, based on an M-vector of cut-offs $c = (c(m) : m = 1, \ldots, M) \in I\!R^M$, further denote the numbers of rejected hypotheses and Type I errors by $R(c|Q)$ and $V(c|Q)$, respectively.

Rejection regions are typically derived so that the Type I error rate $\Theta(F_{V_0,R_0})$, *under the test statistics null distribution* Q_0, is controlled at *nominal level* $\alpha \in (0, 1)$[2], that is,

[1] N.B. In stepwise procedures, the rejection regions $\mathcal{C}$ may be random, i.e., may depend on Z.

[2] N.B. Without loss of generality, we focus for simplicity on Type I error rates $\Theta(F_{V_n,R_n}) \in [0, 1]$.

$$\Theta(F_{V_0,R_0}) \leq \alpha. \tag{2.4}$$

The multiple testing procedure $\mathcal{R}_n$ 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

$$\begin{aligned} \Theta(F_{V_n,R_n}) &\leq \alpha \qquad &&\text{[finite sample control]} \\ \limsup_{n\to\infty} \Theta(F_{V_n,R_n}) &\leq \alpha \qquad &&\text{[asymptotic control].} \end{aligned} \tag{2.5}$$

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

$$\begin{array}{ll} \text{Conservative} & \Theta(F_{V_n,R_n}) < \alpha \\ \text{Anti-conservative} & \Theta(F_{V_n,R_n}) > \alpha. \end{array} \tag{2.6}$$

The choice of a suitable test statistics null distribution Q_0 is crucial in order to ensure that (finite sample or asymptotic) control of the Type I error rate under this assumed null distribution does indeed provide the desired control under the true distribution Q_n. 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

$$\begin{aligned} \Theta(F_{V_n,R_n}) &\leq \Theta(F_{V_0,R_0}) \qquad &&\text{[finite sample control]} \\ \limsup_{n\to\infty} \Theta(F_{V_n,R_n}) &\leq \Theta(F_{V_0,R_0}) \qquad &&\text{[asymptotic control].} \end{aligned} \tag{2.7}$$

Chapter 8 and articles by van der Laan and Hubbard (2006), Pollard et al. (2005a), and Pollard and van der Laan (2004), present simulation studies investigating the impact of the null distribution on the Type I error control and power properties of a MTP.

2.2.2 Sketch of proposed approach to Type I error control

The following discussion motivates our general approach to the problem of Type I error control and highlights important considerations in choosing a test statistics null distribution. We focus on Type I error rates defined as arbitrary parameters $\Theta(F_{V_n})$ of the distribution of the number of Type I errors V_n (Section 1.2.9).

Recall that the distribution F_{V_n}, for the number of Type I errors $V_n = |\mathcal{R}_n \cap \mathcal{H}_0| = |\mathcal{R}(T_n, Q_0, \alpha) \cap \mathcal{H}_0(P)|$, depends on the following: the true distribution $Q_n = Q_n(P)$ of the test statistics T_n; the test statistics null distribution Q_0, used to derive the rejection regions $\mathcal{C}_n(m) = \mathcal{C}(m; T_n, Q_0, \alpha)$; the nominal Type I error level α of the MTP; and the set $\mathcal{H}_0 = \mathcal{H}_0(P)$ of true null hypotheses. Type I error control is therefore a statement about the true, unknown data generating distribution P, via $Q_n(P)$ and $\mathcal{H}_0(P)$.

Control of Type I error rates of the form $\Theta(F_{V_n})$ can be achieved by the three-step road map of Procedure 2.1, below. This road map provides intuition behind the general characterization (Section 2.2.3) and explicit construction (Sections 2.3 and 2.4) of a proper test statistics null distribution Q_0. It also provides a template for Θ-controlling joint single-step common-quantile Procedure 4.1 and common-cut-off Procedure 4.2. 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.

Procedure 2.1. [Three-step road map for controlling Type I error rates $\boldsymbol{\Theta(F_{V_n})}$]

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

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

2. **Monotonicity of the Type I error rate mapping $\boldsymbol{\Theta}$.** Note that the number of Type I errors is always less than or equal to the total number of rejected hypotheses (i.e., $V_0 \leq R_0$), so that $F_{V_0} \geq F_{R_0}$. Hence, under monotonicity Assumption $\mathsf{M}\Theta$ for the Type I error rate mapping Θ, one has
$$\Theta(F_{V_0}) \leq \Theta(F_{R_0}). \tag{2.8}$$

3. **Control of $\boldsymbol{\Theta(F_{R_0})}$.** 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,
$$\Theta(F_{R_0}) \leq \alpha. \tag{2.9}$$

That is, control the known parameter $\Theta(F_{R_0})$, corresponding to the number of rejected hypotheses $R_0 = \sum_{m=1}^{M} \mathrm{I}(T_n(m) \in \mathcal{C}_n(m))$, under the null distribution Q_0, i.e., assuming $T_n \sim Q_0$.

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,

$$\begin{aligned} \Theta(F_{V_n}) \leq \Theta(F_{V_0}) \leq \Theta(F_{R_0}) \leq \alpha \qquad & \text{[finite sample control]} \\ \limsup_{n\to\infty} \Theta(F_{V_n}) \leq \Theta(F_{V_0}) \leq \Theta(F_{R_0}) \leq \alpha \qquad & \text{[asymptotic control].} \end{aligned} \tag{2.10}$$

Note that the road map of Procedure 2.1 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.

2.2.3 Characterization of test statistics null distribution in terms of null domination conditions

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 $\mathsf{ND}\boldsymbol{\Theta}$, 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 $\mathcal{H}_0$.

Null domination conditions for the numbers of Type I errors V_n and V_0

One can specify *null domination* conditions in terms of the distributions of the *numbers of Type I errors* V_n and V_0, as follows. For each $x \in \{0, \ldots, M\}$,

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

That is, 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.

For Type I error rate mappings Θ that satisfy monotonicity Assumption $\mathsf{M}\Theta$ and continuity Assumption $\mathsf{C}\Theta$ at F_{V_0}, null domination Assumption NDV for the number of Type I errors implies null domination Assumption $\mathsf{ND}\boldsymbol{\Theta}$ for the Type I error rate.

Joint null domination conditions for the $\mathcal{H}_0$-specific test statistics $(T_n(m) : m \in \mathcal{H}_0)$

One can also specify multivariate null domination conditions in terms of the joint distribution of the test statistics $(T_n(m) : m \in \mathcal{H}_0)$ for the true null hypotheses $\mathcal{H}_0$, based on the notion of multivariate stochastic order (Kamae et al., 1977, p. 899). Below are three equivalent *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 said to be stochastically greater than the corresponding true distribution $Q_{n,\mathcal{H}_0} = Q_{n,\mathcal{H}_0}(P)$, if, for all bounded componentwise increasing functions $\ell : I\!R^{h_0} \rightarrow I\!R$,

$$\begin{aligned} \mathrm{E}_{Q_{n,\mathcal{H}_0}}\left[\ell((T_n(m) : m \in \mathcal{H}_0))\right] &\leq \mathrm{E}_{Q_{0,\mathcal{H}_0}}\left[\ell((Z(m) : m \in \mathcal{H}_0))\right] \\ \limsup_{n\rightarrow\infty} \mathrm{E}_{Q_{n,\mathcal{H}_0}}\left[\ell((T_n(m) : m \in \mathcal{H}_0))\right] &\leq \mathrm{E}_{Q_{0,\mathcal{H}_0}}\left[\ell((Z(m) : m \in \mathcal{H}_0))\right], \end{aligned} \tag{2.11}$$

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

An alternate formulation of joint null domination is that, for all Borel sets $\mathcal{B} \subseteq I\!R^{h_0}$ with componentwise increasing indicator function $\mathrm{I}_{\mathcal{B}} : z \in I\!R^{h_0} \rightarrow \mathrm{I}(z \in \mathcal{B}) \in \{0, 1\}$,

$$\begin{aligned} \mathrm{Pr}_{Q_{n,\mathcal{H}_0}}\left((T_n(m) : m \in \mathcal{H}_0) \in \mathcal{B}\right) &\leq \mathrm{Pr}_{Q_{0,\mathcal{H}_0}}\left((Z(m) : m \in \mathcal{H}_0) \in \mathcal{B}\right) \\ \limsup_{n\rightarrow\infty} \mathrm{Pr}_{Q_{n,\mathcal{H}_0}}\left((T_n(m) : m \in \mathcal{H}_0) \in \mathcal{B}\right) &\leq \mathrm{Pr}_{Q_{0,\mathcal{H}_0}}\left((Z(m) : m \in \mathcal{H}_0) \in \mathcal{B}\right), \end{aligned} \tag{2.12}$$

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

A third, more compact formulation of joint null domination, in terms of the joint cumulative distribution functions of the test statistics $(T_n(m) : m \in \mathcal{H}_0)$, is that, for all $z \in I\!R^{h_0}$,

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

where, for the asymptotic statement, the null distribution $Q_{0,\mathcal{H}_0}$ is further required to be continuous. Note that Assumption jtNDT corresponds to Equation (2.12), with sets $\mathcal{B} = (-\infty, z]^c$ defined in terms of h_0-dimensional rectangles $(-\infty, z] = \prod_{m=1}^{h_0}(-\infty, z(m)] \subseteq I\!R^{h_0}$.

For ease of notation, we may simply refer to the finite sample and asymptotic joint null domination conditions as $Q_{n,\mathcal{H}_0} \geq Q_{0,\mathcal{H}_0}$ and $\liminf_n Q_{n,\mathcal{H}_0} \geq Q_{0,\mathcal{H}_0}$, respectively.

Relationships between null domination Assumptions jtNDT, NDV, and NDΘ

For one-sided rejection regions of the form $\mathcal{C}(m) = (c(m), +\infty)$, joint null domination Assumption jtNDT for the test statistics implies null domination Assumption NDV for the number of Type I errors. Indeed, for given $c = (c(m) : m = 1, \ldots, M) \in I\!R^M$ and $x \in \{0, \ldots, M\}$, one may apply Equation (2.11), with the bounded componentwise increasing function $\ell : I\!R^{h_0} \rightarrow I\!R$ defined such that

$$\ell((Z(m) : m \in \mathcal{H}_0)) = \mathrm{I}\left(\sum_{m \in \mathcal{H}_0} \mathrm{I}\left(Z(m) > c(m)\right) > x\right) = \mathrm{I}\left(V(c|Q) > x\right),$$

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

$$\Pr\left(V(c|Q_n) > x\right) \leq \Pr\left(V(c|Q_0) > x\right)$$
$$\limsup_{n \to \infty} \Pr\left(V(c|Q_n) > x\right) \leq \Pr\left(V(c|Q_0) > x\right).$$

Noting that $\Pr(V(c|Q_n) > x) = 1 - F_{V_n}(x)$ and $\Pr(V(c|Q_0) > x) = 1 - F_{V_0}(x)$ yields Assumption NDV.

Monotonicity Assumption MΘ and continuity Assumption CΘ at F_{V_0} then imply null domination Assumption NDΘ for the Type I error rate.

Note that, for the asymptotic versions of null domination in Equations (2.11), (2.12), and (jtNDT), one could relax the continuity assumption on Q_0, by requiring, for example, that the cut-offs c be continuity points of Q_0.

To summarize, one has the following relationships among the three types of null domination assumptions introduced thus far. Under these assumptions, the road map of Procedure 2.1 provides (finite sample or asymptotic) control of general Type I error rates of the form $\Theta(F_{V_n})$.

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,

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

Note that null domination is only a statement about the joint distribution of the subvector of test statistics $(T_n(m) : m \in \mathcal{H}_0)$ for the true null hypotheses $\mathcal{H}_0$.

More specific (i.e., less stringent) forms of null domination may be derived for given definitions of the Type I error rate mapping Θ and rejection regions (e.g., null domination conditions for FWER-controlling step-down common-cut-off and common-quantile MTPs in Chapter 5 and van der Laan et al. (2004a)).

General joint null domination Assumption jtNDT, for the $\mathcal{H}_0$-specific test statistics, provides a template for deriving test statistics null distributions that lead to proper Type I error control: identify a collection of M functions, $\ell_m : I\!R \rightarrow I\!R$, such that the joint distribution of the transformed test statistics $(\ell_m(T_n(m)) : m \in \mathcal{H}_0)$ dominates the joint distribution of the original test statistics $(T_n(m) : m \in \mathcal{H}_0)$. Based on this general characterization, Sections 2.3 and 2.4, below, provide two explicit constructions for a proper test statistics null distribution Q_0: the asymptotic distribution of a vector of null shift and scale-transformed test statistics, based on user-supplied upper bounds for the means and variances of the $\mathcal{H}_0$-specific test statistics (Section 2.3; Dudoit et al. (2004b); van der Laan et al. (2004a); Pollard and van der Laan (2004)) and the asymptotic distribution of a vector of null quantile-transformed test statistics, based on user-supplied marginal test statistics null distributions (Section 2.4; van der Laan and Hubbard (2006)).

Either test statistics null distribution may be used in any of the multiple testing procedures proposed in Chapters 3–7 of this book, as they both satisfy the key property of joint null domination for the $\mathcal{H}_0$-specific test statistics (Assumption jtNDT). Specifically, the null shift and scale-transformed null distribution (or a consistent estimator thereof) provides Type I error control for: $\Theta(F_{V_n})$-controlling joint single-step common-cut-off and common-quantile procedures (Chapter 4; Dudoit et al. (2004b)); FWER-controlling joint step-down common-cut-off (maxT) and common-quantile (minP) procedures (Chapter 5; van der Laan et al. (2004a)); gTP-controlling (marginal/joint single-step/stepwise) augmentation multiple testing procedures (Chapter 6; Dudoit et al. (2004a); van der Laan et al. (2004b)); gTP-controlling joint resampling-based empirical Bayes procedures (Chapter 7; van der Laan et al. (2005)). van der Laan and Hubbard (2006) argue that the above results also hold for the new null quantile-transformed test statistics null distribution.

2.2.4 Contrast with other approaches

One of our main contributions is the general characterization (Section 2.2.3) and explicit construction (Sections 2.3 and 2.4) of proper null distributions Q_0 (or estimators thereof, Q_{0n}) for the test statistics T_n. As detailed in Section 2.8, 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)).

Type I error control under the true data generating distribution

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.

In particular, our *null domination* Assumptions jtNDT, NDV, and $\mathsf{ND}\boldsymbol{\Theta}$, introduced in Section 2.2.3, differ from the standard *subset pivotality* assumption of Westfall and Young (1993, p. 42–43), in the following senses: (i) null domination is only concerned with the true data generating distribution P, i.e., the subset $\mathcal{H}_0(P)$ of true null hypotheses and not all possible 2^M subsets $\mathcal{J}_0 \subseteq \{1, \ldots, M\}$ of null hypotheses; (ii) null domination does not require equality of the joint distributions $Q_{0,\mathcal{H}_0}$ and $Q_{n,\mathcal{H}_0}(P)$, for the $\mathcal{H}_0$-specific test statistics $(T_n(m) : m \in \mathcal{H}_0)$, but the weaker domination of $Q_{n,\mathcal{H}_0}(P)$ by $Q_{0,\mathcal{H}_0}$.

Null distribution for the test statistics

Secondly, we propose a *null distribution for the test statistics* ($T_n \sim Q_0$) rather than a *data generating null distribution* ($X \sim P_0$). A common choice of data generating null distribution P_0 is one that satisfies the complete null hypothesis, $H_0^C = \prod_{m=1}^M H_0(m) = \prod_{m=1}^M \mathrm{I}(P \in \mathcal{M}(m)) = \mathrm{I}(P \in \cap_{m=1}^M \mathcal{M}(m))$, that all M null hypotheses are true, i.e., $P_0 \in \cap_{m=1}^M \mathcal{M}(m)$. The data generating null distribution P_0 then implies a null distribution $Q_n(P_0)$ for the test statistics.

As discussed in Pollard et al. (2005a) and Pollard and van der Laan (2004), procedures based on $Q_n(P_0)$ do not necessarily provide proper Type I error control under the true distribution P. Indeed, the assumed null distribution $Q_{n,\mathcal{H}_0}(P_0)$ and the true distribution $Q_{n,\mathcal{H}_0}(P)$, of the $\mathcal{H}_0$-specific test statistics $(T_n(m) : m \in \mathcal{H}_0)$, may have different dependence structures and, as a result, may violate the required null domination condition for the Type I error rate (Assumption $\mathsf{ND}\boldsymbol{\Theta}$, in Step 1 of the road map of Procedure 2.1). For instance, for test statistics with Gaussian asymptotic distributions (Section 2.6), the asymptotic covariance matrix of the $\mathcal{H}_0$-specific test statistics $\Sigma_{\mathcal{H}_0}(P)$, under the true distribution P, may be different from the corresponding covariance matrix $\Sigma_{\mathcal{H}_0}(P_0)$, under the complete null distribution P_0. For the two-sample test of means, based on difference statistics and the commonly-used permutation data generating null distribution P_0, Pollard and van der Laan (2004) show that $\Sigma_{\mathcal{H}_0}(P) = \Sigma_{\mathcal{H}_0}(P_0)$ if and only if (i) the two populations have the same covariance matrices or (ii) the population frequencies are equal (Section 2.9).

Consequently, approaches based on permutation or other data generating null distributions P_0 (e.g., Korn et al. (2004), Troendle (1995, 1996), and Westfall and Young (1993)) are only valid under certain assumptions for the true data generating distribution P. In fact, in most testing problems, there does not exist a data generating null distribution $P_0 \in \cap_{m=1}^{M} \mathcal{M}(m)$ that correctly specifies a joint null distribution for the $\mathcal{H}_0$-specific test statistics $(T_n(m) : m \in \mathcal{H}_0)$, i.e., such that the required null domination condition for the Type I error rate is satisfied (Assumption $\mathsf{ND}\Theta$).

In summary, unlike current methods that can only be applied to a limited set of multiple testing problems, the general constructions of Sections 2.3 and 2.4 lead to joint single-step and stepwise procedures that provide the desired Type I error control for general data generating distributions, null hypotheses, and test statistics. Our proposed test statistics null distributions can be used in testing problems that cannot be handled by traditional approaches based on a data generating null distribution and the associated assumption of subset pivotality. Such problems include tests for correlation coefficients and regression coefficients in linear and non-linear models where covariates and error terms are allowed to be dependent (Chapter 8; Pollard et al. (2005a)).

2.3 Null shift and scale-transformed test statistics null distribution

2.3.1 Explicit construction for the test statistics null distribution

Following Dudoit et al. (2004b), van der Laan et al. (2004a), and Pollard and van der Laan (2004), our first proposal for a test statistics null distribution is the asymptotic distribution of a vector of null shift and scale-transformed test statistics, based on user-supplied upper bounds for the means and variances of the $\mathcal{H}_0$-specific test statistics.

Theorem 2.2. [Null shift and scale-transformed test statistics null distribution]
Asymptotic 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 $m \in \mathcal{H}_0$,*

$$\limsup_{n\to\infty} \mathrm{E}[T_n(m)] \leq \lambda_0(m) \tag{2.13}$$

and

$$\limsup_{n\to\infty} \mathrm{Var}[T_n(m)] \leq \tau_0(m).$$

Let

$$\nu_{0,n}(m) \equiv \sqrt{\min\left\{1, \frac{\tau_0(m)}{\mathrm{Var}[T_n(m)]}\right\}} \tag{2.14}$$

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

$$Z_n(m) \equiv \nu_{0,n}(m)\,(T_n(m) - \mathrm{E}[T_n(m)]) + \lambda_0(m), \qquad m = 1, \ldots, M. \tag{2.15}$$

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

$$Z_n \stackrel{\mathcal{L}}{\Rightarrow} Z \sim Q_0(P). \tag{2.16}$$

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

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

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 MΘ *and continuity Assumption* CΘ *at F_{V_0}, then the null distribution Q_0 also satisfies asymptotic null domination Assumption* **NDΘ** *for the Type I error rate,*

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

Finite sample test statistics null distribution. *Suppose there exists a known M-vector $\lambda_{0,n} \in I\!R^M$ of null values, so that, for each $m \in \mathcal{H}_0$,*

$$\mathrm{E}[T_n(m)] \leq \lambda_{0,n}(m). \tag{2.17}$$

Define an M-vector of null shift-transformed test statistics $Z_n = (Z_n(m)\colon m = 1, \ldots, M)$ *by*

$$Z_n(m) \equiv T_n(m) - \mathrm{E}[T_n(m)] + \lambda_{0,n}(m), \qquad m = 1, \ldots, M. \tag{2.18}$$

Then, the finite sample joint distribution $Q_{0,n} = Q_{0,n}(P)$ of Z_n satisfies the finite sample versions of null domination Assumptions jtNDT, NDV, *and* NDΘ.

The asymptotic distribution Q_0, of the null shift and scale-transformed test statistics Z_n, generalizes the null distribution proposed in Pollard and van der Laan (2004) for the test of single-parameter null hypotheses based on t-statistics. In this special case, the null distribution Q_0 turns out to be a Gaussian distribution with mean vector zero (Section 2.6).

Dudoit et al. (2004b) and van der Laan et al. (2004a) prove that joint single-step and step-down procedures based on the null distribution of Theorem 2.2 (or a consistent estimator thereof) do indeed provide the desired asymptotic control of the Type I error rate $\Theta(F_{V_n})$, for general data generating distributions (with arbitrary dependence structures among variables), null hypotheses (defined in terms of submodels for the data generating distribution), and test statistics (e.g., t-statistics, χ^2-statistics, F-statistics).

As seen in Sections 2.6 and 2.7, the null distribution Q_0 is continuous for a broad class of testing problems. Otherwise, one could relax the continuity assumption on Q_0, by requiring, for example, that the cut-offs c be continuity points of Q_0.

Proof of Theorem 2.2.

Asymptotic test statistics null distribution. 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

$$\widetilde{Z}_n(m) = T_n(m) + \max\left\{0, \lambda_0(m) - \mathrm{E}[T_n(m)]\right\}, \qquad m = 1, \ldots, M. \tag{2.19}$$

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 (B.7)). 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 B.3), 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 B.2. That is, for all $c = (c(m) : m = 1, \ldots, M) \in I\!R^M$ and $x \in \{0, \ldots, M\}$,

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

Finite sample test statistics null distribution. The finite sample results follow immediately by noting that, under Equation (2.17), $Z_n(m) \geq T_n(m)$ for $m \in \mathcal{H}_0$.

□

Remarks

1. **Role of null shift values λ_0.** The construction of the null distribution Q_0 in Theorem 2.2 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 Q_0 therefore satisfies asymptotic null domination Assumption NDV, for the number of Type I errors, and also Θ-specific asymptotic null domination Assumption $\mathsf{ND}\boldsymbol{\Theta}$, for any Type I error rate mapping Θ that meets monotonicity Assumption $\mathsf{M}\Theta$ and continuity Assumption $\mathsf{C}\Theta$ at F_{V_0}.

2. **Role of null scale values $\boldsymbol{\tau_0}$.** 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 (Section 2.7).
3. **Estimation of null values $\boldsymbol{\lambda_0}$ and $\boldsymbol{\tau_0}$.** 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 $\tau_0(m) = 1$ (Section 2.6). 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 (Section 2.7). More generally, the null values $\lambda_0(m)$ and $\tau_0(m)$ may depend on the unknown data generating distribution P, as is the case for F-statistics when population variances are unequal (Equation (2.54)). In such a situation, one may replace the parameters $\lambda_0(m)$ and $\tau_0(m)$ by consistent estimators thereof.
4. ***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, $Q_0 = \mathrm{N}(0, \sigma^*)$ (Section 2.6). 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.
5. ***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 (Section 2.7).

6. **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. As detailed in Section 2.3.2, below, resampling procedures, such as the bootstrap procedures proposed in Dudoit et al. (2004b), van der Laan et al. (2004a), and Pollard and van der Laan (2004), 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.

2.3.2 Bootstrap estimation of the test statistics null distribution

As noted above, the test statistics null distribution $Q_0 = Q_0(P)$ proposed in Theorem 2.2 depends 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.

Given an estimator Q_{0n} of the null distribution Q_0, Procedures 4.20 and 4.21 provide algorithms for estimating cut-offs and adjusted p-values for $\Theta(F_{V_n})$-controlling joint single-step common-quantile Procedure 4.1 and common-cut-off Procedure 4.2, respectively. Similar algorithms are proposed in Procedure 5.15 for FWER-controlling joint step-down maxT Procedure 5.1 and minP Procedure 5.6.

General direct bootstrap estimation

As discussed below, bootstrap procedures provide a very general approach for obtaining consistent estimators of the test statistics null distribution Q_0 proposed in Theorem 2.2. The method may be summarized as follows and is illustrated in Figure 2.1.

1. Given B bootstrap samples of the data $\mathcal{X}_n$, obtain an $M \times B$ matrix of test statistics, $\mathbf{T}_n^B = \big(T_n^B(m,b)\big)$, with rows corresponding to the M null hypotheses and columns to the B bootstrap samples.
2. Estimate the expected values, $\mathrm{E}[T_n(m)]$, and variances, $\mathrm{Var}[T_n(m)]$, of the test statistics (under the true data generating distribution P) by taking row means and variances of the matrix $\mathbf{T}_n^B$.
3. Row-shift and scale the matrix of bootstrap test statistics $\mathbf{T}_n^B$, with the user-supplied null values $\lambda_0(m)$ and $\tau_0(m)$, to produce an $M \times B$ matrix $\mathbf{Z}_n^B = \big(Z_n^B(m,b)\big)$.

4. Estimate the null distribution Q_0 by the empirical distribution Q_{0n} of the B columns of matrix $\mathbf{Z}_n^B$.

The remainder of this section provides details on the (non-parametric or model-based) bootstrap estimation of the null distribution Q_0 of Theorem 2.2. Specifically, 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.

A *bootstrap sample* consists of n IID copies, $\mathcal{X}_n^\# \equiv \{X_i^\# : i = 1, \ldots, n\}$, of a random variable $X^\# \sim P_n^\star$. Denote the M-vector of test statistics computed from such a bootstrap sample by $T_n^\# = (T_n^\#(m) : m = 1, \ldots, M)$. The null distribution Q_0 proposed in Theorem 2.2 can be estimated by the distribution of the null shift and scale-transformed bootstrap test statistics,

$$Z_n^\#(m) \equiv \sqrt{\min\left\{1, \frac{\tau_0(m)}{\mathrm{Var}_{P_n^\star}[T_n{}^\#(m)]}\right\}} \left(T_n^\#(m) - \mathrm{E}_{P_n^\star}[T_n{}^\#(m)]\right) + \lambda_0(m). \tag{2.20}$$

In practice, one can only approximate the distribution of $Z_n^\# = (Z_n^\#(m) : m = 1, \ldots, M)$ by an empirical distribution over B bootstrap samples drawn from $P_n^\star$, as described next in Procedure 2.3.

Procedure 2.3. [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, $\mathbf{T}_n^B = \left(T_n^B(m, b) : m = 1, \ldots, M;\ b = 1, \ldots, B\right)$, 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

$$\mathrm{E}[T_n^B(m,\cdot)] \equiv \frac{1}{B}\sum_{b=1}^{B} T_n^B(m,b), \tag{2.21}$$

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

4. Obtain an $M \times B$ matrix, $\mathbf{Z}_n^B = (Z_n^B(m,b) : m=1,\ldots,M; b=1,\ldots,B)$, of null shift and scale-transformed bootstrap test statistics $Z_n^B(m,b)$, as in Theorem 2.2, 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

$$Z_n^B(m,b) \equiv \sqrt{\min\left\{1, \frac{\tau_0(m)}{\mathrm{Var}[T_n^B(m,\cdot)]}\right\}} \times \left(T_n^B(m,b) - \mathrm{E}[T_n^B(m,\cdot)]\right) + \lambda_0(m). \tag{2.22}$$

5. The bootstrap estimator Q_{0n} of the null distribution Q_0 from Theorem 2.2 is the empirical distribution of the B columns $\{Z_n^B(\cdot,b) : b = 1,\ldots,B\}$ of matrix $\mathbf{Z}_n^B$.

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 $\mathbf{Z}_n^B = \left(Z_n^B(m,b)\right)$ by recording, for each row m, the proportion of null shift and scale-transformed bootstrap test statistics $Z_n^B(m,b)$ that are greater than or equal to the observed test statistic $T_n(m)$ (Section 1.2.12). That is,

$$P_{0n}(m) = \frac{1}{B}\sum_{b=1}^{B} \mathrm{I}\left(Z_n^B(m,b) \geq T_n(m)\right), \qquad m = 1,\ldots,M. \tag{2.23}$$

Figures 2.1 and 2.2 provide, respectively, graphical summaries of the bootstrap estimation of the null distribution Q_0 and of the corresponding unadjusted p-values $P_{0n}(m)$.

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 addition, according to the definition in Equation (2.23), unadjusted p-values are often zero, even for moderate numbers of bootstrap samples B. In order to deal with the discreteness of the bootstrap distribution, the marginal null distributions $Q_{0n,m}$ obtained from the matrix $\mathbf{Z}_n^B$ may be replaced by Gaussian approximations or smoothed (e.g., using kernel density estimation methods). Specific algorithms for accurate estimation of tail probabilities are beyond the scope of this book. In general, the user needs to find a balance between estimation accuracy and computational cost.

Test statistic-specific estimation: *t*-statistics and *F*-statistics

For certain types of test statistics T_n (e.g., t-statistics, χ^2-statistics, F-statistics) one may exploit the known parametric form of the null distribution Q_0 of Theorem 2.2. An advantage of test statistic-specific parametric estimation approaches, such as those discussed in Sections 2.6 and 2.7, is that they yield continuous null distributions, which do not suffer from the discreteness of the non-parametric bootstrap null distribution described above.

t-statistics

As detailed in Section 2.6, for the test of single-parameter null hypotheses using t-statistics, a t-statistic-specific null distribution $Q_0^t = Q_0^t(P)$ is the M-variate Gaussian distribution $\mathrm{N}(0, \sigma^*)$, where $\sigma^* = \Sigma^*(P)$ is the correlation matrix of the M-dimensional vector influence curve, $IC(X|P) = (IC(X|P)(m) : m = 1, \ldots, M)$, for an asymptotically linear estimator ψ_n of the parameter M-vector ψ (Section 1.2.5).

In this case, one can estimate Q_0^t by $Q_{0n}^t = \mathrm{N}(0, \sigma_n^*)$, where $\sigma_n^* = \hat{\Sigma}^*(P_n)$ is a consistent estimator of the correlation matrix σ^*. For example, one could use the correlation matrix σ_n^* corresponding to the following estimator of the $M \times M$ influence curve covariance matrix,

$$\sigma_n = \hat{\Sigma}(P_n) = \frac{1}{n} \sum_{i=1}^{n} IC_n(X_i) IC_n^\top(X_i), \tag{2.24}$$

where $IC_n(X) = (IC_n(X)(m) : m = 1, \ldots, M)$ is an estimator of the M-vector influence curve $IC(X|P)$.

Influence curves can be derived straightforwardly for simple parameters such as means. For example, when estimating the mean vector $\psi = \mathrm{E}[X]$, for a random M-vector $X \sim P$, using the corresponding empirical mean vector $\psi_n = \bar{X}_n$, the influence curves are $IC(X|P)(m) = X(m) - \psi(m)$ and corresponding estimators are $IC_n(X)(m) = X(m) - \psi_n(m)$, where $\psi_n(m) = \bar{X}_n(m) = \sum_i X_i(m)/n$, $m = 1, \ldots, M$. Then, σ_n^* is simply the empirical correlation matrix. Influence curves for estimators of correlation coefficients and regression coefficients are given in Section 2.6.

In cases where the influence curves are not readily available, the correlation matrix σ^* may be estimated with the bootstrap.

F-statistics

As detailed in Section 2.7, for testing the equality of K population mean vectors using F-statistics, an F-statistic-specific null distribution

$Q_0^F = Q_0^F(P_1, \ldots, P_K)$ can be defined in terms of a simple quadratic function of K independent Gaussian M-vectors, $Y_k \sim \mathrm{N}(0, \sigma_k)$, where $\sigma_k = \Sigma(P_k)$ denotes the covariance matrix for the kth population, $k = 1, \ldots, K$.

An estimator Q_{0n}^F of the null distribution Q_0^F can be obtained by estimating each population covariance matrix σ_k by the corresponding empirical covariance matrix or by using the bootstrap.

Data generating null distribution estimation

In certain testing problems, one may define a test statistics null distribution $Q_n(P_0)$, in terms of a data generating distribution P_0 that satisfies the complete null hypothesis $H_0^C = \prod_{m=1}^M H_0(m)$ that all M null hypotheses are true. Such a null distribution may be estimated by $Q_{0n} = Q_n(P_{0n})$, where, for example, P_{0n} is a bootstrap- or permutation-based estimator of P_0.

Test statistics null distributions based on bootstrap and permutation data generating distributions are discussed in Section 2.9. Parameter-specific bootstrap data generating null distributions are described in Chapter 8 for tests concerning regression coefficients and correlation coefficients (Procedures 8.4 and 8.6, respectively).

However, as discussed in Pollard et al. (2005a) and Pollard and van der Laan (2004), approaches based on a data generating null distribution can fail in important testing problems, as the assumed null distribution $Q_{n,\mathcal{H}_0}(P_0)$ and the true distribution $Q_{n,\mathcal{H}_0}(P)$, of the $\mathcal{H}_0$-specific test statistics $(T_n(m) : m \in \mathcal{H}_0)$, may have different dependence structures and, as a result, may violate the required null domination condition for the Type I error rate (Assumption $\mathsf{ND}\boldsymbol{\Theta}$, in Step 1 of the road map of Procedure 2.1).

Indeed, the simulation studies of Chapter 8 show that bootstrap data generating null distributions can lead to severely anti-conservative (bootstrapping residuals for testing regression coefficients) or conservative (independent bootstrap for testing correlation coefficients) procedures.

2.4 Null quantile-transformed test statistics null distribution

Following van der Laan and Hubbard (2006), our second proposal for a test statistics null distribution is the asymptotic distribution of a vector of null quantile-transformed test statistics, based on user-supplied marginal test statistics null distributions. Because this promising approach represents a very recent development in our ongoing research on multiple testing, this book only introduces the main features of the null quantile-transformed null distribution. The reader is referred to van der Laan and Hubbard (2006) for formal theorems and proofs, a detailed treatment of tests based on t-statistics and χ^2-statistics, simulation studies, and an application to tests of association between non-Hodgkin lymphoma (NHL) subclass and single nucleotide polymorphisms (SNP) in the ghrelin (`GHRL`) and neuropeptide Y (`NPY`) genes.

This latest construction has the 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). The preliminary results in van der Laan and Hubbard (2006) indeed illustrate that, for finite samples, the new null quantile-transformed null distribution provides more accurate Type I error control and is more powerful than the null shift and scale-transformed null distribution of Section 2.3.

2.4.1 Explicit construction for the test statistics null distribution

Marginal null domination conditions for the $\mathcal{H}_0$-specific test statistics $(T_n(m) : m \in \mathcal{H}_0)$

The main ingredients of the new null quantile-transformed test statistics null distribution are user-supplied marginal test statistics null distributions $q_{0,m}$, $m = 1, \ldots, M$, that satisfy the following *marginal null domination* condition[3]. For each $m \in \mathcal{H}_0$ and $z \in I\!R$,

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

That is, the test statistics $(T_n(m) : m \in \mathcal{H}_0)$, for the true null hypotheses $\mathcal{H}_0$, are marginally stochastically greater under the null distributions $q_{0,m}$ than under the true distributions $Q_{n,m}$. Note that the above marginal null domination Assumption mgNDT is implied by the stronger joint null domination Assumption jtNDT.

Finite sample test statistics null distribution

Given marginal null distributions $q_{0,m}$, $m = 1, \ldots, M$, that satisfy marginal null domination Assumption mgNDT, the proposed finite sample joint null distribution is based on the *generalized quantile-quantile function transformation* of Yu and van der Laan (2002). Specifically, let $\breve{Q}_{0,n} = \breve{Q}_{0,n}(P)$ denote the joint distribution of the M-vector of *null quantile-transformed test statistics* $\breve{Z}_n = (\breve{Z}_n(m) : m = 1, \ldots, M)$ defined as

$$\breve{Z}_n(m) \equiv q_{0,m}^{-1} Q_{n,m}^{\Delta}(T_n(m)), \qquad m = 1, \ldots, M, \tag{2.25}$$

[3] N.B. In practice, user-supplied marginal null distributions, such as permutation distributions, depend on the sample size n. However, for simplicity, references to the sample size n are omitted from the notation $q_{0,m}$.

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

One can easily verify that the marginal distributions $\breve{Q}_{0,n,m}$, corresponding to the proposed joint null distribution $\breve{Q}_{0,n}$, are indeed equal to the user-supplied marginal null distributions $q_{0,m}$. For continuous user-supplied marginal null distributions $q_{0,m}$ and continuous true marginal distributions $Q_{n,m}$, one has $Q^{\Delta}_{n,m}(z) = Q_{n,m}(z)$ for each $z \in I\!R$ and, hence,

$$\begin{aligned}\breve{Q}_{0,n,m}(z) &= \Pr\left(\breve{Z}_n(m) \leq z\right)\\ &= \Pr\left(q^{-1}_{0,m}Q_{n,m}(T_n(m)) \leq z\right)\\ &= \Pr\left(Q_{n,m}(T_n(m)) \leq q_{0,m}(z)\right)\\ &= q_{0,m}(z),\end{aligned}$$

where the last equality follows from Proposition 1.2.

In cases where the marginal distributions $Q_{n,m}$ and $q_{0,m}$ are not necessarily continuous, Lemma 2.4 of Yu and van der Laan (2002) ensures that the marginal distributions $\breve{Q}_{0,n,m}$ are indeed equal to the user-supplied marginal null distributions $q_{0,m}$.

Result 1 in van der Laan and Hubbard (2006) establishes that the finite sample joint null distribution $\breve{Q}_{0,n}$ satisfies null domination Assumption NDV for the number of Type I errors. That is, for each $c = (c(m) : m = 1, \ldots, M) \in I\!R^M$ and $x \in \{0, \ldots, M\}$,

$$\Pr\left(V(c|\breve{Q}_{0,n}) \leq x\right) - \Pr\left(V(c|Q_n) \leq x\right) \leq 0 \tag{2.26}$$

$$\limsup_{n\to\infty}\left(\Pr\left(V(c|\breve{Q}_{0,n}) \leq x\right) - \Pr\left(V(c|Q_n) \leq x\right)\right) \leq 0.$$

In other words, the number of Type I errors $V_{0,n} = V(c|\breve{Q}_{0,n}) = \sum_{m\in\mathcal{H}_0} \mathrm{I}\left(\breve{Z}_n(m) > c(m)\right)$, under the null distribution $\breve{Q}_{0,n}$, is stochastically greater than the number of Type I errors $V_n = V(c|Q_n) = \sum_{m\in\mathcal{H}_0} \mathrm{I}\left(T_n(m) > c(m)\right)$, under the true distribution Q_n. Null domination Assumption $\mathsf{ND}\boldsymbol{\Theta}$ for the Type I error rate follows for mappings Θ that satisfy monotonicity Assumption $\mathsf{M}\Theta$ and uniform continuity Assumption $\mathsf{C}\Theta$.

Asymptotic test statistics null distribution

As in van der Laan and Hubbard (2006), further assume that the finite sample joint null distribution $\breve{Q}_{0,n} = \breve{Q}_{0,n}(P)$ converges weakly to an asymptotic joint null distribution $\breve{Q}_0 = \breve{Q}_0(P)$.

Result 2 in van der Laan and Hubbard (2006) is an analogue for $\breve{Q}_0$ of Result 1 for $\breve{Q}_{0,n}$. That is, the asymptotic joint null distribution $\breve{Q}_0$ satisfies null domination Assumption NDV for the number of Type I errors.

In general, proofs of null domination properties for the new null quantile-transformed null distribution are similar to those for the original null shift and scale-transformed null distribution (e.g., Theorem 2.2).

2.4.2 Bootstrap estimation of the test statistics null distribution

As for our original null shift and scale-transformed test statistics null distribution $Q_0 = Q_0(P)$ (Section 2.3), neither the finite sample null distribution $\breve{Q}_{0,n} = \breve{Q}_{0,n}(P)$ nor the asymptotic null distribution $\breve{Q}_0 = \breve{Q}_0(P)$ is known, as they both depend on the true, unknown data generating distribution P. van der Laan and Hubbard (2006) propose in their Section 2 a bootstrap procedure, similar to Procedure 2.3, for estimating the asymptotic null distribution $\breve{Q}_0$.

Procedure 2.4. [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 $Q_{n,m}^B$, as the empirical CDFs of the rows of matrix $\mathbf{T}_n^B$, that is,
$$Q_{n,m}^B(z) \equiv \frac{1}{B} \sum_{b=1}^{B} \mathrm{I}\left(T_n^B(m, b) \leq z\right). \tag{2.27}$$
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 quantile-transformed bootstrap test statistics $Z_n^B(m, b)$, defined as
$$Z_n^B(m, b) \equiv q_{0,m}^{-1} Q_{n,m}^{B,\Delta}(T_n^B(m, b)), \tag{2.28}$$
where $Q_{n,m}^{B,\Delta}(z) \equiv \Delta Q_{n,m}^B(z) + (1 - \Delta) Q_{n,m}^B(z^-)$ 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 $\{Z_n^B(\cdot, b) : b = 1, \ldots, B\}$ of matrix $\mathbf{Z}_n^B$.

From Lemma 2.4 in Yu and van der Laan (2002), the generalized quantile-quantile function transformation $q_{0,m}^{-1} Q_{n,m}^{B,\Delta}(z)$ ensures that the margins $\breve{Q}_{0n,m}$, of the estimator $\breve{Q}_{0n}$ based on a finite number B of bootstrap samples, are equal to the user-supplied marginal null distributions $q_{0,m}$.

As discussed in Section 2.3.2, in the context of the null shift and scale-transformed null distribution, one could also envisage estimation approaches that are test statistic-specific (e.g., for t-statistics, χ^2-statistics, F-statistics) or based on a data generating null distribution. The reader is referred to Section 4 in van der Laan and Hubbard (2006) for a detailed treatment of null distributions for tests based on t-statistics and χ^2-statistics.

2.4.3 Comparison of null shift and scale-transformed and null quantile-transformed null distributions

This section compares our two main constructions for a test statistics null distribution. Recall from Section 2.3 that the first null distribution $Q_0 = Q_0(P)$, proposed in Dudoit et al. (2004b) and van der Laan et al. (2004a), is defined as the asymptotic distribution of the M-vector $Z_n = (Z_n(m) : m = 1, \ldots, M)$ of null shift and scale-transformed test statistics. That is,

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

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.

In contrast, the new null distribution $\breve{Q}_0 = \breve{Q}_0(P)$ of van der Laan and Hubbard (2006) is defined as the asymptotic distribution of the M-vector $\breve{Z}_n = (\breve{Z}_n(m) : m = 1, \ldots, M)$ of null quantile-transformed test statistics. That is,

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

where $q_{0,m}$ are user-supplied marginal test statistics null distributions.

1. **Main ingredients: Null shift and scale values and null quantiles.** While our first proposal requires M-vectors of null values $\lambda_0 \in I\!R^M$ and $\tau_0 \in I\!R^{+M}$, so that $\limsup_n \mathrm{E}[T_n(m)] \leq \lambda_0(m)$ and $\limsup_n \mathrm{Var}[T_n(m)] \leq \tau_0(m)$ for $m \in \mathcal{H}_0$, the new proposal of van der Laan and Hubbard (2006) relies on marginal null distributions $q_{0,m}$ that dominate the true marginal distributions $Q_{n,m}$, i.e., satisfy marginal null domination Assumption mgNDT.
2. **$\mathcal{H}_0$-specific joint null distributions.** If the true marginal distributions $Q_{n,m}$, of the test statistics $T_n(m)$ for the true null hypotheses $m \in \mathcal{H}_0$, converge weakly (up to a location shift) to the corresponding user-supplied marginal null distributions $q_{0,m}$, then the two $\mathcal{H}_0$-specific joint null distributions $Q_{0,\mathcal{H}_0}$ and $\breve{Q}_{0,\mathcal{H}_0}$ coincide.

3. **$\mathcal{H}_1$-specific joint null distributions.** In general, for the false null hypotheses $\mathcal{H}_1$, the null-transformed test statistics $Z_n(m)$ and $\breve{Z}_n(m)$ can have very different finite sample and asymptotic marginal distributions. In particular, whereas the marginal distributions of $\breve{Q}_0$ coincide with the user-supplied marginal null distributions (i.e., $\breve{Q}_{0,m} = q_{0,m}$), the marginal distributions of Q_0 do not necessarily have this property. Hence, the $\mathcal{H}_1$-specific joint null distributions $Q_{0,\mathcal{H}_1}$ and $\breve{Q}_{0,\mathcal{H}_1}$ could in principle be very different and thus lead to procedures with different power properties.
4. **Estimation of the test statistics null distributions.** In practice, both test statistics null distributions $Q_0 = Q_0(P)$ and $\breve{Q}_0 = \breve{Q}_0(P)$ are unknown, as they depend on the unknown data generating distribution P. Similar bootstrap procedures may be used to obtain consistent estimators Q_{0n} and $\breve{Q}_{0n}$, of Q_0 and $\breve{Q}_0$, respectively (Procedures 2.3 and 2.4). However, bootstrap estimators $\breve{Q}_{0n}$ of the null quantile-transformed null distribution $\breve{Q}_0$ are expected to be more efficient than bootstrap estimators Q_{0n} of the null shift and scale-transformed null distribution Q_0. To see this, suppose that the two $\mathcal{H}_0$-specific joint null distributions $Q_{0,\mathcal{H}_0}$ and $\breve{Q}_{0,\mathcal{H}_0}$ coincide. The bootstrap estimator of $\breve{Q}_0$ is based on a model where all marginal distributions are given, whereas the bootstrap estimator of Q_0 ignores this information and considers a larger model with unspecified marginal distributions. As a result, the bootstrap marginal distributions $Q_{0n,m}$ are subject to finite sample variability and typically differ from the user-supplied marginal distributions $q_{0,m}$.
5. **Known optimal marginal null distributions.** The new null quantile-transformed null distribution is particularly appealing when one has available optimal marginal null distributions $q_{0,m}$ for single hypothesis testing. For example, consider a data structure $X = (X(m) : m = 1, \ldots, M+1)$, where $(X(m) : m = 1, \ldots, M)$ is an M-dimensional covariate/genotype vector and $Y = X(M+1)$ is a univariate outcome/phenotype. The covariates/genotypes could correspond to M microarray gene expression measures and the outcome/phenotype to a (censored) survival time or a tumor class. Suppose one wishes to test the M null hypotheses $H_0(m)$ of independence between the covariates $X(m)$ and the outcome $Y = X(M+1)$, $m = 1, \ldots, M$, based on an M-vector of arbitrary test statistics $T_n = (T_n(m) : m = 1, \ldots, M)$. Then, one can set the marginal null distributions $q_{0,m}$ equal to the permutation distributions of the corresponding test statistics $T_n(m)$. One knows from single hypothesis testing that if the null hypothesis $H_0(m)$ is true, then the permutation distribution of $T_n(m)$ is (exactly) equal to the true conditional distribution of $T_n(m)$, given the marginal empirical distributions of $X(m)$ and Y. In the special case of the test of single-parameter null hypotheses based on t-statistics, one could use standard normal marginal null distributions, that is, set $q_{0,m} = \Phi$, where Φ is the $\mathrm{N}(0,1)$ CDF.

van der Laan and Hubbard (2006) argue in their Section 3 that Type I error control results proved in our earlier articles for the original null shift and scale-transformed test statistics null distribution Q_0 and its bootstrap estimators Q_{0n} also hold for the new null quantile-transformed test statistics null distribution. Specifically, the null quantile-transformed test statistics null distribution $\breve{Q}_0$ and its bootstrap estimators $\breve{Q}_{0n}$ provide Type I error control for: $\Theta(F_{V_n})$-controlling joint single-step common-cut-off and common-quantile procedures (Chapter 4); FWER-controlling joint step-down common-cut-off (maxT) and common-quantile (minP) procedures (Chapter 5); gTP-controlling (marginal/joint single-step/stepwise) augmentation multiple testing procedures (Chapter 6); gTP-controlling joint resampling-based empirical Bayes procedures (Chapter 7). The main point is that both test statistics null distributions satisfy joint null domination Assumption jtNDT for the $\mathcal{H}_0$-specific test statistics.

Section 4 in van der Laan and Hubbard (2006) is analogous to Sections 2.6 and 2.7, below, in that it examines properties of the null quantile-transformed null distribution for two types of testing problems: the test of single-parameter null hypotheses using t-statistics (e.g., tests of means, correlation coefficients, regression coefficients) and the test of multiple-parameter null hypotheses using χ^2-statistics.

In summary, either test statistics null distribution Q_0 or $\breve{Q}_0$ (or consistent estimators thereof) may be used in any of the multiple testing procedures proposed in Chapters 3–7 of this book, as they both satisfy the key property of joint null domination for the $\mathcal{H}_0$-specific test statistics (Assumption jtNDT). In particular, 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 (defined in terms of submodels for the data generating distribution), and test statistics (e.g., t-statistics, χ^2-statistics, F-statistics). The newly proposed null quantile-transformed null distribution has the additional advantage that it allows the user to select optimal marginal null distributions and hence tends to outperform the original null shift and scale-transformed null distribution. Unless stated otherwise, the simpler notation Q_0 and Q_{0n} refers to either null distribution.

2.5 Null distribution for transformations of the test statistics

2.5.1 Null distribution for transformed test statistics

Suppose one is interested in deriving rejection regions for an M-vector of test statistics $T_n^\ell = (T_n^\ell(m) : m = 1, \ldots, M)$, defined as *transformations* of the original test statistics $T_n = (T_n(m) : m = 1, \ldots, M)$, by $T_n^\ell(m) \equiv \ell_m(T_n(m))$, in terms of a collection of M functions $\ell_m : I\!R \rightarrow I\!R$.

The special case of the absolute value function ($\ell(z) = |z|$) is discussed in general terms in Section 2.5.2 and also in Section 4.5, in the context of single-step common-cut-off and common-quantile procedures.

As in Equation (2.2), given a random M-vector $Z = (Z(m) : m = 1, \ldots, M)$, with joint distribution Q, and a collection of M rejection regions $\mathcal{C} = \{\mathcal{C}(m) : m = 1, \ldots, M\}$, denote the numbers of rejected hypotheses and Type I errors for the transformed test statistics $\ell_m(Z(m))$ by

$$R^\ell(\mathcal{C}|Q) \equiv \sum_{m=1}^{M} \mathrm{I}\left(\ell_m(Z(m)) \in \mathcal{C}(m)\right) \tag{2.29}$$

and

$$V^\ell(\mathcal{C}|Q) \equiv \sum_{m\in\mathcal{H}_0} \mathrm{I}\left(\ell_m(Z(m)) \in \mathcal{C}(m)\right),$$

respectively. Also adopt the shorthand notation of Equation (2.3), for the special cases where Q corresponds to the true distribution Q_n and null distribution Q_0 for the original test statistics T_n,

$$\begin{aligned} R_n^\ell &\equiv R^\ell(\mathcal{C}|Q_n), & R_0^\ell &\equiv R^\ell(\mathcal{C}|Q_0), \\ V_n^\ell &\equiv V^\ell(\mathcal{C}|Q_n), & V_0^\ell &\equiv V^\ell(\mathcal{C}|Q_0). \end{aligned} \tag{2.30}$$

The proposition below specifies conditions under which deriving rejection regions for the transformed test statistics T_n^ℓ, based on a null distribution Q_0 for the original test statistics T_n, leads to proper Type I error control.

Proposition 2.5. [Null distribution for transformed test statistics] *Consider the simultaneous test of M null hypotheses $H_0(m)$, $m = 1, \ldots, M$, based on an M-vector of test statistics $T_n^\ell = (T_n^\ell(m) : m = 1, \ldots, M)$, defined as transformations of the original test statistics $T_n = (T_n(m) : m = 1, \ldots, M)$, by $T_n^\ell(m) = \ell_m(T_n(m))$, in terms of a collection of M functions $\ell_m : I\!R \rightarrow I\!R$. Let $Q_n = Q_n(P)$ and Q_0 denote, respectively, the true finite sample joint distribution of T_n and a null distribution that satisfies joint null domination Assumption jtNDT for the $\mathcal{H}_0$-specific subvector of test statistics $(T_n(m) : m \in \mathcal{H}_0)$.*

Scenario 1. *If the functions ℓ_m are continuous and non-decreasing, then joint null domination Assumption jtNDT for the original test statistics $(T_n(m) : m \in \mathcal{H}_0)$ implies joint null domination Assumption jtNDT for the transformed test statistics $(T_n^\ell(m) : m \in \mathcal{H}_0)$. Hence, for one-sided rejection regions of the form $\mathcal{C}(m) = (c(m), +\infty)$ for the transformed test statistics $T_n^\ell(m)$, null domination Assumption NDV is satisfied by the numbers of Type I errors V_n^ℓ and V_0^ℓ. If one further assumes that the Type I error rate mapping Θ meets monotonicity Assumption MΘ and continuity Assumption CΘ at $F_{V_0^\ell}$, then null domination Assumption NDΘ is* *satisfied by the Type I error rates* $\Theta(F_{V_0^\ell})$ *and* $\Theta(F_{V_n^\ell})$*. This means that one-sided rejection regions for the transformed test statistics* T_n^ℓ *may be derived based on the null distribution* Q_0 *for the original test statistics* T_n*.*

Scenario 2. *If joint null domination Assumption* jtNDT *holds with equality for the original test statistics* $(T_n(m) : m \in \mathcal{H}_0)$*, then it also holds with equality for the transformed test statistics* $(T_n^\ell(m) : m \in \mathcal{H}_0)$*, for any continuous functions* ℓ_m*. Hence, for any type of rejection regions* $\mathcal{C}(m)$ *for the transformed test statistics* $T_n^\ell(m)$*, null domination Assumption* NDV *is satisfied with equality by the numbers of Type I errors* V_n^ℓ *and* V_0^ℓ*. If one further assumes that the Type I error rate mapping* Θ *meets monotonicity Assumption* MΘ *and continuity Assumption* CΘ *at* $F_{V_0^\ell}$*, then null domination Assumption* NDΘ *is satisfied with equality by the Type I error rates* $\Theta(F_{V_0^\ell})$ *and* $\Theta(F_{V_n^\ell})$*. This means that any type of rejection regions for the transformed test statistics* T_n^ℓ *may be derived based on the null distribution* Q_0 *for the original test statistics* T_n*.*

The proof of this proposition is straightforward and is therefore omitted.

An alternative and more general approach for obtaining rejection regions for transformed test statistics T_n^ℓ would be to derive a null distribution Q_0^ℓ directly for T_n^ℓ, using the general constructions of Sections 2.3 and 2.4.

There is, however, a trade-off between generality and simplicity. For instance, consider the test of single-parameter null hypotheses using t-statistics T_n (Section 2.6). For the null shift and scale-transformed approach of Section 2.3, the null values are $\lambda_0(m) = 0$ and $\tau_0(m) = 1$ and the null distribution Q_0 for T_n is an M-variate Gaussian distribution, with mean vector zero and covariance matrix $\sigma^* = \Sigma^*(P)$ equal to the correlation matrix of the vector influence curve. For the transformed test statistics T_n^ℓ, the null shift and scale values are no longer 0 and 1 and the null distribution Q_0^ℓ is no longer Gaussian.

2.5.2 Example: Absolute value transformation

A special case of interest is the *absolute value* function, $\ell(z) = |z|$, which corresponds to *symmetric two-sided rejection regions* for the original test statistics $T_n(m)$: $\mathcal{C}_n(m;\alpha) = (-\infty, -c_n(m;\alpha)) \cup (c_n(m;\alpha), +\infty)$, for an M-vector of non-negative cut-offs $c_n(\alpha) = (c_n(m;\alpha) : m = 1, \ldots, M) \in I\!R^{+M}$. That is, for a MTP with nominal Type I error level α, the set of rejected null hypotheses is given by

$$\begin{aligned}\mathcal{R}^{||}(T_n, Q_0, \alpha) &= \{m : T_n(m) < -c_n(m;\alpha) \text{ or } T_n(m) > c_n(m;\alpha)\} \\ &= \{m : |T_n(m)| > c_n(m;\alpha)\}\,.\end{aligned} \tag{2.31}$$

Specifically, consider the two-sided test of single-parameter null hypotheses $H_0(m) = \mathrm{I}(\psi(m) = \psi_0(m))$ against alternative hypotheses $H_1(m) =$

$\mathrm{I}(\psi(m) \neq \psi_0(m))$, based on an M-vector of t-statistics, defined as in Section 2.6 by

$$T_n(m) = \sqrt{n}\frac{\psi_n(m) - \psi_0(m)}{\sigma_n(m)},$$

where ψ_n is an asymptotically linear estimator of the parameter ψ.

A similar argument as in the proof of Theorem 2.6 shows that

$$(T_n(m) : m \in \mathcal{H}_0) \overset{\mathcal{L}}{\Rightarrow} Q_{0,\mathcal{H}_0} = \mathrm{N}(0, \sigma^*_{\mathcal{H}_0}).$$

Hence, asymptotic joint null domination Assumption jtNDT is satisfied with equality for the $\mathcal{H}_0$-specific absolute t-statistics $(|T_n(m)| : m \in \mathcal{H}_0)$ and the null distribution $Q_0 = \mathrm{N}(0, \sigma^*)$ of Theorem 2.6. It follows from Proposition 2.5, Scenario 2, that asymptotic null domination Assumptions NDV and $\mathsf{ND}\boldsymbol{\Theta}$, for the number of Type I errors and Type I error rate, are also satisfied with equality for any type of rejection regions for $|T_n(m)|$. Hence, as dictated by the three-step road map of Procedure 2.1, one has

$$\lim_{n\to\infty} \Theta(F_{V_n^{||}}) = \Theta(F_{V_0^{||}}) \leq \Theta(F_{R_0^{||}}) \leq \alpha. \tag{2.32}$$

Thus, multiple testing procedures based on absolute t-statistics $|T_n(m)|$ and any type of rejection regions $\mathcal{C}_n(m) = \mathcal{C}(m; T_n, Q_0, \alpha)$, derived under the Gaussian null distribution $Q_0 = \mathrm{N}(0, \sigma^*)$ of Theorem 2.6, do indeed provide the desired Type I error control. The special cases of single-step common-quantile Procedure 4.1 and common-cut-off Procedure 4.2 are discussed in detail in Section 4.5.

Note that, for the absolute value function and two-sided rejection regions, the stronger requirement of asymptotic *equality* of the test statistics true distribution $Q_{n,\mathcal{H}_0}$ and null distribution $Q_{0,\mathcal{H}_0}$ is essential, as the weaker domination property would only guarantee Type I error control for one of the tails.

2.5.3 Example: Null shift and scale and null quantile transformations

The random M-vectors of *null-transformed test statistics* Z_n and $\breve{Z}_n$ (Equations (2.15) and (2.25)), defining the null distributions proposed in Sections 2.3 and 2.4, correspond, respectively, to the following transformations,

$$\ell_{0,m}(z) \equiv \nu_{0,n}(m)\,(z - \mathrm{E}[T_n(m)]) + \lambda_0(m) \tag{2.33}$$

and

$$\breve{\ell}_{0,m}(z) \equiv q_{0,m}^{-1} Q_{n,m}^{\Delta}(z).$$

The null shift and scale functions $\ell_{0,m}$ are continuous and non-decreasing. For continuous marginal distributions $Q_{n,m}$ and $q_{0,m}$, the null quantile functions $\breve{\ell}_{0,m}$ are also continuous and non-decreasing. Thus, Scenario 1 in Proposition 2.5 applies to a broad range of testing problems.

2.5.4 Bootstrap estimation of the null distribution for transformed test statistics

Regarding the bootstrap estimation of rejection regions and adjusted p-values for MTPs based on transformed test statistics T_n^ℓ, one could first use general Procedure 2.3 or 2.4 (or a related procedure from Section 2.6 or 2.7) to derive a matrix $\mathbf{Z}_n^B = \left(Z_n^B(m,b)\right)$, of null-transformed bootstrap test statistics $Z_n^B(m,b)$, based on the original test statistics T_n. The null distribution Q_0, for the original test statistics T_n, is estimated by the empirical distribution Q_{0n} of the columns of matrix $\mathbf{Z}_n^B$. An estimated null distribution Q_{0n}^ℓ, for the transformed test statistics T_n^ℓ, is given by the empirical distribution of the columns of the transformed matrix $\ell(\mathbf{Z}_n^B) = \left(\ell_m(Z_n^B(m,b))\right)$.

Using Q_{0n}^ℓ to obtain rejection regions for the transformed test statistics T_n^ℓ leads to procedures that control the Type I error rate $\Theta(F_{V_n^\ell})$ under the two scenarios considered in Proposition 2.5.

For instance, bootstrap versions of single-step common-quantile Procedure 4.1 and common-cut-off Procedure 4.2 may be implemented as in Procedures 4.20 and 4.21, respectively, using transformed test statistics T_n^ℓ and the estimated null distribution Q_{0n}^ℓ.

2.6 Testing single-parameter null hypotheses based on t-statistics

2.6.1 Set-up and assumptions

In this section, we consider the one-sided test of M single-parameter null hypotheses $H_0(m) = \mathrm{I}\left(\psi(m) \leq \psi_0(m)\right)$ against alternative hypotheses $H_1(m) = \mathrm{I}\left(\psi(m) > \psi_0(m)\right)$, where $\Psi(P) = \psi = (\psi(m) : m = 1, \ldots, M)$ is an M-vector of real-valued parameters $\Psi(P)(m) = \psi(m)$.

The null hypotheses can be tested using *t-statistics*, defined as in Section 1.2.5 by

$$T_n(m) \equiv \sqrt{n}\frac{\psi_n(m) - \psi_0(m)}{\sigma_n(m)}, \tag{2.34}$$

where $\hat{\Psi}(P_n) = \psi_n = (\psi_n(m) : m = 1, \ldots, M)$ is an *asymptotically linear estimator* of the parameter M-vector $\Psi(P) = \psi$, with M-dimensional vector *influence curve* (IC) $IC(X|P) = (IC(X|P)(m) : m = 1, \ldots, M)$, such that

$$\psi_n(m) - \psi(m) = \frac{1}{n}\sum_{i=1}^{n} IC(X_i|P)(m) + o_P(1/\sqrt{n}), \tag{2.35}$$

and $\sigma_n^2(m)$ are consistent estimators of the variances $\sigma^2(m) = \sigma(m,m) = \mathrm{E}[IC^2(X|P)(m)]$, $m = 1, \ldots, M$. Let $Q_n = Q_n(P)$ denote the finite sample joint distribution of T_n, under the true, unknown data generating distribution P. Large values of the t-statistic $T_n(m)$ are assumed to provide evidence against the corresponding null hypothesis $H_0(m) = \mathrm{I}(\psi(m) \leq \psi_0(m))$, that is, tests are based on one-sided rejection regions of the form $\mathcal{C}_n(m) = (c_n(m), +\infty)$.

Next, we propose a *t-statistic-specific null distribution* Q_0^t that leads to asymptotic control of Type I error rates $\Theta(F_{V_n})$, defined as arbitrary parameters of the distribution of the number of Type I errors V_n.

2.6.2 Test statistics null distribution

Theorem 2.6. [**t-statistic-specific null distribution**] *Consider t-statistics $T_n = (T_n(m) : m = 1, \ldots, M)$, defined as in Equation (2.34), and a test statistics null distribution $Q_0^t = Q_0^t(P) \equiv N(0, \Sigma^*(P))$, defined as the M-variate Gaussian distribution with covariance matrix $\sigma^* = \Sigma^*(P)$ equal to the correlation matrix of the vector influence curve $IC(X|P)$ of Equation (2.35). Then, asymptotic null domination Assumption* NDV, *for the number of Type I errors, is satisfied by the t-statistics T_n and the null distribution Q_0^t. That is, for all $c = (c(m) : m = 1, \ldots, M) \in I\!R^M$ and $x \in \{0, \ldots, M\}$,*

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

Thus, according to the three-step road map of Procedure 2.1, multiple testing procedures based on t-statistics T_n and the t-statistic-specific null distribution Q_0^t provide asymptotic control of general Type I error rates $\Theta(F_{V_n})$, for the one-sided test of single-parameter null hypotheses $H_0(m) = \mathrm{I}(\psi(m) \leq \psi_0(m))$ against alternative hypotheses $H_1(m) = \mathrm{I}(\psi(m) > \psi_0(m))$.

Proof of Theorem 2.6. Let us verify asymptotic null domination Assumption NDV for the t-statistics T_n of Equation (2.34) and the null distribution $Q_0^t = \mathrm{N}(0, \sigma^*)$. Firstly, note that the t-statistics $T_n(m)$ can be rewritten as

$$\begin{aligned} T_n(m) &= \sqrt{n}\frac{\psi_n(m) - \psi(m)}{\sigma_n(m)} + \frac{\sigma(m)}{\sigma_n(m)}\sqrt{n}\frac{\psi(m) - \psi_0(m)}{\sigma(m)} \qquad (2.36) \\ &= Z_n^t(m) + \frac{\sigma(m)}{\sigma_n(m)} d_n(m), \end{aligned}$$

in terms of deterministic shifts, $d_n(m) \equiv \sqrt{n}(\psi(m) - \psi_0(m))/\sigma(m)$, and standardized statistics, $Z_n^t(m) \equiv \sqrt{n}(\psi_n(m) - \psi(m))/\sigma_n(m)$. By Equation (2.35) and the Central Limit Theorem (Theorem B.4), one has

$$Z_n^t \stackrel{\mathcal{L}}{\Rightarrow} Z^t \sim Q_0^t(P) = \mathrm{N}(0, \Sigma^*(P)), \tag{2.37}$$

where $\sigma^* = \Sigma^*(P) = \mathrm{Cov}[Z^t]$ is the correlation matrix of the M-vector influence curve $IC(X|P)$. For $m \in \mathcal{H}_0$, $d_n(m) \leq 0$, so that $T_n(m) \leq Z_n^t(m)$. Thus, from the Continuous Mapping Theorem (Theorem B.3) and Proposition B.2,

$$\begin{aligned}
&\liminf_{n\to\infty} \Pr\left(\sum_{m\in\mathcal{H}_0} \mathrm{I}\left(T_n(m) > c(m)\right) \leq x\right) \\
&\geq \liminf_{n\to\infty} \Pr\left(\sum_{m\in\mathcal{H}_0} \mathrm{I}\left(Z_n^t(m) > c(m)\right) \leq x\right) \\
&= \Pr\left(\sum_{m\in\mathcal{H}_0} \mathrm{I}\left(Z^t(m) > c(m)\right) \leq x\right),
\end{aligned}$$

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

□

The above theorem proposes a test statistics null distribution Q_0^t derived specifically in terms of the t-statistics T_n of Equations (2.34) and (2.35). As described below, it turns out that this null distribution Q_0^t corresponds to the general proposals Q_0 and $\breve{Q}_0$ of Sections 2.3 and 2.4, respectively.

Comparison to null shift and scale-transformed null distribution

One can show, under mild regularity conditions, that the t-statistic-specific null distribution $Q_0^t = \mathrm{N}(0, \sigma^*)$ of Theorem 2.6 corresponds to the general null shift and scale-transformed null distribution Q_0 of Theorem 2.2, with null values $\lambda_0(m) = 0$ and $\tau_0(m) = 1$.

To see this, consider the simple known variance case, where $\sigma_n(m) = \sigma(m)$. Then, $\mathrm{E}[T_n] = d_n$ and $\mathrm{Cov}[T_n] = \mathrm{Cor}[T_n] = \sigma^*$. Hence, $T_n(m) = Z_n^t(m) + \mathrm{E}[T_n(m)]$. In addition, for null values $\lambda_0(m) = 0$ and $\tau_0(m) = 1$, the M-vector Z_n, defining the general null distribution Q_0 in Theorem 2.2, reduces to Z_n^t. Hence, $Q_0 = Q_0^t = \mathrm{N}(0, \sigma^*)$.

Comparison to null quantile-transformed null distribution

A similar equivalence result is provided for the null quantile-transformed null distribution in Section 4.1 of van der Laan and Hubbard (2006). Specifically, for standard normal marginal null distributions $q_{0,m} = \Phi$, it is argued that the asymptotic null quantile-transformed null distribution $\breve{Q}_0$ is equal to the t-statistic-specific null distribution $Q_0^t = \mathrm{N}(0, \sigma^*)$ of Theorem 2.6. That is,

$$\breve{Z}_n = (\Phi^{-1} Q_{n,m}^{\Delta}(T_n(m)) : m = 1, \ldots, M) \stackrel{\mathcal{L}}{\Rightarrow} Q_0^t.$$

Theorem 2 of van der Laan and Hubbard (2006) further shows that the bootstrap estimator $\breve{Q}_{0n}$ of Procedure 2.4 converges weakly to Q_0^t.

2.6.3 Estimation of the test statistics null distribution

One can exploit the specific form of the t-statistics defined in Equations (2.34) and (2.35), to derive consistent estimators of the null distribution $Q_0^t = \mathrm{N}(0, \sigma^*)$ of Theorem 2.6.

First, consider the case where one knows the form of the M-vector influence curve, $IC(X|P) = (IC(X|P)(m) : m = 1, \ldots, M)$, for the estimator $\psi_n = \hat{\Psi}(P_n)$ (e.g., tests for means, correlation coefficients, and regression coefficients, treated in Sections 2.6.4–2.6.6, below). Given an estimator $IC_n(X) = (IC_n(X)(m) : m = 1, \ldots, M)$ of $IC(X|P)$, one can obtain the following estimator of the $M \times M$ influence curve covariance matrix $\sigma = \Sigma(P)$,

$$\sigma_n = \hat{\Sigma}(P_n) = \frac{1}{n} \sum_{i=1}^{n} IC_n(X_i) IC_n^\top(X_i). \tag{2.38}$$

An estimator of Q_0^t is then given by the M-variate Gaussian distribution $Q_{0n}^t = \mathrm{N}(0, \sigma_n^*)$, where $\sigma_n^* = \hat{\Sigma}^*(P_n)$ is the correlation matrix corresponding to the estimated covariance matrix σ_n.

When the influence curve is not readily available, $\sigma^* = \Sigma^*(P)$ can be estimated with the bootstrap as follows. Given an estimator $P_n^\star$ of the true data generating distribution P, let $\mathcal{X}_n^\# = \{X_i^\# : i = 1, \ldots, n\}$ denote a random sample of n IID copies of a random variable $X^\# \sim P_n^\star$. For each bootstrap sample $\mathcal{X}_n^\#$, with empirical distribution $P_n^\#$, compute the estimator $\psi_n^\# = \hat{\Psi}(P_n^\#)$. A bootstrap estimator of the covariance (and correlation) matrix $\sigma^* = \mathrm{Cov}_P[Z^t]$ is given by the covariance (and correlation) matrix $\sigma_n^* = \mathrm{Cov}_{P_n^\star}[Z_n^{t,\#}]$, of standardized bootstrap test statistics $Z_n^{t,\#}$ defined as either

$$Z_n^{t,\#}(m) = \frac{(\psi_n^\#(m) - \mathrm{E}_{P_n^\star}[\psi_n{}^\#(m)])}{\sqrt{\mathrm{Var}_{P_n^\star}[\psi_n{}^\#(m)]}} \tag{2.39}$$

or

$$Z_n^{t,\#}(m) = \frac{(\psi_n^\#(m) - \psi_n(m))}{\sqrt{\mathrm{Var}_{P_n^\star}[\psi_n{}^\#(m)]}}, \qquad m = 1, \ldots, M.$$

A parametric bootstrap estimator of the null distribution Q_0^t is then given by $Q_{0n}^t = \mathrm{N}(0, \sigma_n^*)$; a non-parametric bootstrap estimator is also provided by the joint distribution of the M-vector of standardized statistics $Z_n^{t,\#}$.

Note that, when an estimator of the influence curve is available, using the bootstrap to estimate σ^* does not necessarily pay off over direct estimation based on the original sample $\mathcal{X}_n$. When the correlation matrix is sparse, shrinkage estimation methods may be beneficial.

Alternately, a consistent estimator of the null distribution Q_0^t can be obtained using general bootstrap Procedure 2.3, for the null shift and scale-transformed null distribution, with null values $\lambda_0(m) = 0$ and $\tau_0(m) = 1$. Likewise, one could apply general bootstrap Procedure 2.4, for the null quantile-transformed null distribution, with standard normal marginal null distributions $q_{0,m} = \Phi$.

As mentioned in Section 2.3.2, above, one of the main advantages of a parametric estimator $Q_{0n}^t = \mathrm{N}(0, \sigma_n^*)$ is that it is continuous and hence does not suffer from the discreteness of non-parametric bootstrap estimators. Similar issues arise for F-statistics, as discussed in Section 2.7, below.

2.6.4 Example: Tests for means

A familiar testing problem, that falls within our single-parameter hypothesis testing framework, is that where $X \sim P$ is a random J-vector and the parameter of interest is the *mean* vector of X, $\Psi(P) = \psi = (\psi(j) : j = 1, \ldots, J) = \mathrm{E}[X]$, with elements $\psi(j) = \Psi(P)(j) = \mathrm{E}[X(j)]$. The $M = J$ null hypotheses, $H_0(m) = \mathrm{I}(\psi(m) \leq \psi_0(m))$, then refer to individual elements of the mean vector ψ.

Given a random sample $\mathcal{X}_n = \{X_i : i = 1, \ldots, n\}$, from the data generating distribution P, the test statistics $T_n(m)$ of Equation (2.34) are the usual *one-sample t-statistics*, where $\psi_n(m) = \hat{\Psi}(P_n)(m) = \bar{X}_n(m) = \sum_i X_i(m)/n$ and $\sigma_n^2(m) = \sum_i (X_i(m) - \bar{X}_n(m))^2/n$ are the empirical means and variances of the M elements of X, respectively.

In this simple case, the elements of the M-vector influence curve are $IC(X|P)(m) = X(m) - \psi(m)$ and can be estimated by $IC_n(X)(m) = X(m) - \bar{X}_n(m)$. Thus, a consistent estimator of the test statistics null distribution Q_0^t of Theorem 2.6 is the M-variate Gaussian distribution $Q_{0n}^t = \mathrm{N}(0, \sigma_n^*)$, where $\sigma_n^* = \hat{\Sigma}^*(P_n)$ is the $M \times M$ empirical correlation matrix $\mathrm{Cor}_{P_n}[X]$.

2.6.5 Example: Tests for correlation coefficients

Another common testing problem covered by Theorem 2.6 is that where the parameter of interest is the $J \times J$ *correlation* matrix for a random J-vector $X \sim P$, that is, $\Psi(P) = \psi = (\psi(j, j') : j,\ j' = 1, \ldots, J) = \mathrm{Cor}[X]$, with elements $\psi(j, j') = \Psi(P)(j, j') = \mathrm{Cor}[X(j), X(j')]$. Suppose one is interested in testing the $M = J(J-1)/2$ null hypotheses that the J elements of X are uncorrelated, that is, null hypotheses $H_0(j, j') = \mathrm{I}(\psi(j, j') = 0)$, $j = 1, \ldots, J-1$, $j' = j+1, \ldots, J$.

Commonly-used test statistics for this problem are $T_n(j, j') = \sqrt{n}\psi_n(j, j')$, where $\psi_n(j, j') = \hat{\Psi}(P_n)(j, j')$ are the *empirical correlation coefficients*. As discussed in Westfall and Young (1993, Example 2.2, p. 43), subset pivotality fails for this testing problem. To see this, consider the simple case where $J = 3$ (and $M = 3$) and assume that $H_0(1, 2)$ and $H_0(1, 3)$ are true, so that $\psi(1, 2) = \psi(1, 3) = 0$. Then, the joint distribution of $(T_n(1, 2), T_n(1, 3))$ is asymptotically Gaussian, with mean vector zero, unit variances, and correlation of $\psi(2, 3)$, and thus depends on the truth or falsity of the third hypothesis $H_0(2, 3)$. In other words, the covariance matrix of the vector influence curve for the empirical correlation coefficients differs under the true data generating distribution P and under a data generating null distribution P_0 for which $\psi(j,j') = 0, \forall\, j \neq j'$. Tests for correlation coefficients thus provide an example where standard procedures based on subset pivotality fail, whereas procedures based on the t-statistic-specific null distribution of Theorem 2.6 or the general null distributions of Sections 2.3 and 2.4 achieve the desired Type I error control (Pollard et al., 2005a; Pollard and van der Laan, 2004).

The influence curves for the empirical correlation coefficients $\psi_n(j,j')$ can be obtained by applying the Delta-method with the function

$$f(\xi(j,j')) = \psi(j,j') = \frac{\gamma(j,j') - \gamma(j)\gamma(j')}{\sqrt{\gamma(j,j) - \gamma^2(j)}\sqrt{\gamma(j',j') - \gamma^2(j')}}, \tag{2.40}$$

defined in terms of a 5×1 parameter column vector $\xi(j,j') = \Xi(P)(j,j') = [\gamma(j), \gamma(j'), \gamma(j,j), \gamma(j',j'), \gamma(j,j')]^\top$, with elements $\gamma(j) = \Gamma(P)(j) = \mathrm{E}[X(j)]$ and $\gamma(j,j') = \Gamma(P)(j,j') = \mathrm{E}[X(j)X(j')]$, $j, j' = 1, \ldots, J$. Let $f'(\xi)$ denote the 1×5 gradient row vector of $f(\xi)$. Then,

$$\psi_n(j,j') - \psi(j,j') = f'(\xi(j,j'))\,(\xi_n(j,j') - \xi(j,j')) + o_P(1/\sqrt{n}), \tag{2.41}$$

where $\xi_n(j,j') = \hat{\Xi}(P_n)(j,j') = [\gamma_n(j), \gamma_n(j'), \gamma_n(j,j), \gamma_n(j',j'), \gamma_n(j,j')]^\top$ is a 5×1 estimator column vector for $\xi(j,j')$, based on the empirical moments. Hence, the influence curve for the estimator $\psi_n(j,j')$ is

$$\begin{aligned} IC(X|P)(j,j') &= f'(\xi(j,j'))(\xi_1(j,j') - \xi(j,j')) \qquad (2.42)\\ &= \frac{1}{\sqrt{\sigma(j,j)}\sqrt{\sigma(j',j')}} \begin{bmatrix} \gamma(j)\frac{\sigma(j,j')}{\sigma(j,j)} - \gamma(j') \\ \gamma(j')\frac{\sigma(j,j')}{\sigma(j',j')} - \gamma(j) \\ -\frac{1}{2}\frac{\sigma(j,j')}{\sigma(j,j)} \\ -\frac{1}{2}\frac{\sigma(j,j')}{\sigma(j',j')} \\ 1 \end{bmatrix}^\top \begin{bmatrix} X(j) - \gamma(j) \\ X(j') - \gamma(j') \\ X^2(j) - \gamma(j,j) \\ X^2(j') - \gamma(j',j') \\ X(j)X(j') - \gamma(j,j') \end{bmatrix}, \end{aligned}$$

where covariances are denoted by $\sigma(j,j') = \gamma(j,j') - \gamma(j)\gamma(j')$.

Section 8.4 examines the choice of a test statistics null distribution in testing problems concerning correlation coefficients. Section 9.3 considers the identification of co-expressed miRNAs based on tests for correlations coefficients.

2.6.6 Example: Tests for regression coefficients

Consider a random $J = (M+1)$-vector $X \sim P$, from a data generating distribution P, where $(X(m) : m = 1, \ldots, M)$ is an M-dimensional covariate/genotype vector and $Y = X(M+1)$ is a univariate outcome/phenotype. For instance, the covariates/genotypes could correspond to M microarray gene expression measures and the outcome/phenotype to a (censored) survival time or a tumor class.

Assume the following model for the conditional expected value of the outcome Y given individual covariates $X(m)$,

$$\mathrm{E}[Y|X(m)] = g(X(m);\gamma_m) = h(\gamma_m(1) + \gamma_m(2)X(m)), \qquad m = 1,\ldots,M, \tag{2.43}$$

where $\Gamma_m(P) = \gamma_m = (\gamma_m(1), \gamma_m(2))$ are *regression coefficients* for the mth covariate $X(m)$. The parameter of interest is the M-vector of *slope parameters*, $\Psi(P) = \psi = (\psi(m) = \gamma_m(2) : m = 1,\ldots,M)$.

Given a random sample $\mathcal{X}_n = \{X_i : i = 1,\ldots,n\}$, from the data generating distribution P, one can estimate the regression parameters γ_m for each covariate $X(m)$ using the method of least squares, that is, by seeking γ_m that minimizes the sum of squared residuals, $\sum_i (Y_i - g(X_i(m);\gamma_m))^2$. The *least squares estimator*, $\hat{\Gamma}_m(P_n) = \gamma_{m,n} = (\gamma_{m,n}(1), \gamma_{m,n}(2))$, is obtained by solving the following equation for γ,

$$0 = \frac{\partial}{\partial\gamma} \sum_{i=1}^{n} (Y_i - g(X_i(m);\gamma))^2 ,$$

that is,

$$0 = \sum_{i=1}^{n} \left(\frac{\partial}{\partial\gamma} g(X_i(m);\gamma)\right) (Y_i - g(X_i(m);\gamma)) .$$

Let $IC_m(X|P) = (IC_m(X|P)(1), IC_m(X|P)(2))$ denote the two-dimensional vector influence curve for the least squares estimator $\gamma_{m,n}$ of the regression parameters γ_m corresponding to covariate $X(m)$. Under mild regularity conditions (Lemma 2.1, p. 105, van der Laan and Robins (2003)), one can show that

$$\begin{aligned}\gamma_{m,n} - \gamma_m = {} & \frac{1}{n}\sum_{i=1}^{n} c_m^{-1}(\gamma_m) \left.\left(\frac{\partial}{\partial\gamma} g(X_i(m);\gamma)\right)\right|_{\gamma=\gamma_m} (Y_i - g(X_i(m);\gamma_m)) \\ & + o_P(1/\sqrt{n}),\end{aligned} \tag{2.44}$$

where, for a given $\gamma \in I\!R^2$,

$$c_m(\gamma) = E \begin{bmatrix} \left(\frac{\partial}{\partial\gamma(1)} g(X(m);\gamma)\right)^2 & \left(\frac{\partial}{\partial\gamma(1)} g(X(m);\gamma)\right) \times \left(\frac{\partial}{\partial\gamma(2)} g(X(m);\gamma)\right) \\ \left(\frac{\partial}{\partial\gamma(1)} g(X(m);\gamma)\right) \times \left(\frac{\partial}{\partial\gamma(2)} g(X(m);\gamma)\right) & \left(\frac{\partial}{\partial\gamma(2)} g(X(m);\gamma)\right)^2 \end{bmatrix} .$$

From the above expression, the influence curves are

$$IC_m(X|P) = c_m^{-1}(\gamma_m) \left(\frac{\partial}{\partial \gamma} g(X(m);\gamma) \right)\Bigg|_{\gamma=\gamma_m} (Y - g(X(m);\gamma_m)) . \quad (2.45)$$

The M-dimensional vector influence curve for the least squares estimators $\hat{\Psi}(P_n) = \psi_n = (\psi_n(m) = \gamma_{m,n}(2) : m = 1, \ldots, M)$, of the M slope parameters ψ, is

$$IC(X|P) = (IC_m(X|P)(2) : m = 1, \ldots, M).$$

The covariance matrix of the vector influence curve $IC(X|P)$ is

$$\sigma = \Sigma(P) = \mathrm{E}\left[IC(X|P)IC^\top(X|P)\right],$$

and can be estimated as in Equation (2.38), using the empirical covariance matrix for an estimator $IC_n(X)$ of the vector influence curve.

Linear regression

A common model for a continuous outcome $Y \in I\!R$ is the *linear model*, corresponding to the identity function $h(z) = z$. That is,

$$\mathrm{E}[Y|X(m)] = g(X(m);\gamma_m) = \gamma_m(1) + \gamma_m(2)X(m). \quad (2.46)$$

In this case, the influence curves for the least squares estimators $\gamma_{m,n}$ of the regression coefficients γ_m are given by

$$\begin{aligned} IC_m(X|P) = & \frac{1}{\mathrm{Var}[X(m)]} \begin{bmatrix} \mathrm{E}[X^2(m)] & -\mathrm{E}[X(m)] \\ -\mathrm{E}[X(m)] & 1 \end{bmatrix} \begin{bmatrix} 1 \\ X(m) \end{bmatrix} \\ & \times (Y - \gamma_m(1) - \gamma_m(2)X(m)) . \end{aligned} \quad (2.47)$$

Logistic regression

A common model for a binary outcome $Y \in \{0, 1\}$ is the *logistic model*, corresponding to the *softmax* or *inverse logit* function $h(z) = \exp(z)/(1+\exp(z))$. That is,

$$\Pr(Y = 1|X(m)) = g(X(m);\gamma_m) = \frac{\exp(\gamma_m(1) + \gamma_m(2)X(m))}{1 + \exp(\gamma_m(1) + \gamma_m(2)X(m))}. \quad (2.48)$$

Here,

$$\left(\frac{\partial}{\partial \gamma} g(X(m);\gamma) \right)\Bigg|_{\gamma=\gamma_m} = \frac{\exp(\gamma_m(1) + \gamma_m(2)X(m))}{(1 + \exp(\gamma_m(1) + \gamma_m(2)X(m)))^2} \begin{bmatrix} 1 \\ X(m) \end{bmatrix}, \quad (2.49)$$

and the influence curves for the least squares estimators $\gamma_{m,n}$ of the regression coefficients γ_m can be derived by substituting for $\partial g(X(m);\gamma)/\partial\gamma$ in Equation (2.45), above.

Section 8.3 examines the choice of a test statistics null distribution in testing problems concerning regression coefficients in linear models where the covariates and error terms are allowed to be dependent. Section 9.3 considers tests for regression coefficients in logistic models relating cancer status to miRNA expression measures and tissue type (Pollard et al., 2005a).

2.7 Testing multiple-parameter null hypotheses based on *F*-statistics

2.7.1 Set-up and assumptions

Consider random M-vectors $X_k = (X_k(m) : m = 1, \ldots, M) \sim P_k$, from K different populations, with respective data generating distributions P_k, $k = 1, \ldots, K$. Let $\psi_k = \Psi(P_k) = \mathrm{E}[X_k]$ and $\sigma_k = \Sigma(P_k) = \mathrm{Cov}[X_k]$ denote, respectively, the mean vector and covariance matrix for Population k. Denote the elements of the covariance matrix σ_k by $\sigma_k(m, m') = \mathrm{Cov}[X_k(m), X_k(m')]$ and adopt the shorter notation $\sigma_k^2(m) = \sigma_k(m, m)$ for the diagonal elements of σ_k, i.e., the variances. Consider testing the M null hypotheses $H_0(m) = \mathrm{I}\,(\psi_1(m) = \psi_2(m) = \cdots = \psi_K(m))$, that the elements of the mean vectors are constant across the K populations.

Suppose one observes a random sample $\mathcal{X}_{k,n_k} = \{X_{k,i} : i = 1, \ldots, n_k\}$, of size n_k, from Population k, $k = 1, \ldots, K$[4]. Let $n = \sum_k n_k$ denote the total sample size and $\eta_{k,n} = n_k/n$ the empirical frequency for Population k. Assume that $\lim_n \eta_{k,n} = \eta_k > 0$, $\forall\ k = 1, \ldots, K$. The null hypotheses can be tested using *F-statistics*,

$$T_n(m) \equiv \frac{\frac{1}{K-1}\sum_{k=1}^K n_k(\bar{X}_{k,n_k}(m) - \bar{X}_n(m))^2}{\frac{1}{n-K}\sum_{k=1}^K \sum_{i=1}^{n_k}(X_{k,i}(m) - \bar{X}_{k,n_k}(m))^2}, \qquad m = 1, \ldots, M, \tag{2.50}$$

where $\bar{X}_{k,n_k} = \sum_i X_{k,i}/n_k$ denotes the empirical mean vector for the sample $\mathcal{X}_{k,n_k}$ from Population k and $\bar{X}_n = \sum_k \eta_{k,n}\bar{X}_{k,n_k} = \sum_k \sum_i X_{k,i}/n$ denotes the empirical mean vector for the pooled sample of size n. Large values of the F-statistic $T_n(m)$ are assumed to provide evidence against the corresponding null hypothesis $H_0(m) = \mathrm{I}\,(\psi_1(m) = \psi_2(m) = \cdots = \psi_K(m))$, that is, tests are based on one-sided rejection regions of the form $\mathcal{C}_n(m) = (c_n(m), +\infty)$.

Next, we propose an *F-statistic-specific null distribution* Q_0^F that leads to asymptotic control of Type I error rates $\Theta(F_{V_n})$, defined as arbitrary parameters of the distribution of the number of Type I errors V_n.

[4] N.B. With proper care, one could allow random sample sizes n_k.

2.7.2 Test statistics null distribution

Theorem 2.7. [**F-statistic-specific null distribution**] *Consider F-statistics $T_n = (T_n(m) : m = 1, \ldots, M)$, defined as in Equation (2.50), and a test statistics null distribution $Q_0^F = Q_0^F(P_1, \ldots, P_K)$, defined as the joint distribution of a random M-vector $Z^F = (Z^F(m) : m = 1, \ldots, M)$ of quadratic forms*

$$Z^F(m) \equiv \frac{1}{(K-1)\sum_{k=1}^K \eta_k \sigma_k^2(m)} \tag{2.51}$$

$$\times \left(\sum_{k=1}^K (1-\eta_k) Y_k^2(m) - \sum_{\substack{k=1 \\ k \neq k'}}^K \sum_{k'=1}^K \sqrt{\eta_k \eta_{k'}} Y_k(m) Y_{k'}(m) \right),$$

based on K independent Gaussian M-vectors $Y_k = (Y_k(m) : m = 1, \ldots, M) \sim N(0, \sigma_k)$. In matrix notation, the quadratic forms are defined by

$$Z^F(m) \equiv \widetilde{Y}_m^\top A_m \widetilde{Y}_m, \tag{2.52}$$

based on M dependent Gaussian K-vectors $\widetilde{Y}_m = (Y_k(m) : k = 1, \ldots, K) \sim N(0, \widetilde{\sigma}_m)$, with diagonal covariance matrices $\widetilde{\sigma}_m$ such that $\widetilde{\sigma}_m(k,k) = \sigma_k^2(m)$, and M symmetric $K \times K$ matrices A_m with elements

$$A_m(k,k') \equiv \frac{1}{(K-1)\sum_{k=1}^K \eta_k \sigma_k^2(m)} \begin{cases} (1-\eta_k), & \text{if } k = k' \\ -\sqrt{\eta_k \eta_{k'}}, & \text{if } k \neq k' \end{cases}. \tag{2.53}$$

Then, the F-statistics $(T_n(m) : m \in \mathcal{H}_0)$ for the true null hypotheses converge weakly to the $\mathcal{H}_0$-specific quadratic forms $(Z^F(m) : m \in \mathcal{H}_0)$, that is,

$$(T_n(m) : m \in \mathcal{H}_0) \overset{\mathcal{L}}{\Rightarrow} (Z^F(m) : m \in \mathcal{H}_0) \sim Q_{0,\mathcal{H}_0}^F.$$

It follows that asymptotic null domination Assumption NDV*, for the number of Type I errors, is satisfied with* equality *by the F-statistics T_n and the null distribution Q_0^F. That is, for all $c = (c(m) : m = 1, \ldots, M) \in I\!R^{+M}$ and $x \in \{0, \ldots, M\}$,*

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

$$= \Pr_{Q_0^F} \left(\sum_{m \in \mathcal{H}_0} \mathrm{I}\left(Z^F(m) > c(m)\right) \leq x \right).$$

Thus, according to the three-step road map of Procedure 2.1, multiple testing procedures based on F-statistics T_n and the F-statistic-specific null distribution Q_0^F provide asymptotic control of general Type I error rates $\Theta(F_{V_n})$, *for the test of multiple-parameter null hypotheses of the form* $H_0(m) = \mathrm{I}\left(\psi_1(m) = \psi_2(m) = \cdots = \psi_K(m)\right)$.

Furthermore, the quadratic forms $Z^F(m)$ *have means and variances given, respectively, by*

$$\mathrm{E}[Z^F(m)] = \frac{1}{(K-1)\sum_{k=1}^K \eta_k \sigma_k^2(m)} \sum_{k=1}^K (1-\eta_k)\sigma_k^2(m) \tag{2.54}$$

and

$$\mathrm{Var}[Z^F(m)] = \frac{2}{(K-1)^2(\sum_{k=1}^K \eta_k \sigma_k^2(m))^2} \times \left(\left(\sum_{k=1}^K (1-2\eta_k)\sigma_k^4(m) \right) + \left(\sum_{k=1}^K \eta_k \sigma_k^2(m) \right)^2 \right).$$

In the special case of constant variances across populations, i.e., $\sigma_k^2(m) = \sigma^2(m)$, *then* $\mathrm{E}[Z^F(m)] = 1$, $\mathrm{Var}[Z^F(m)] = 2/(K-1)$, *and the quadratic forms have marginal* χ^2*-distributions with* $(K-1)$ *degrees of freedom, that is,*

$$(K-1)Z^F(m) \sim \chi^2(K-1). \tag{2.55}$$

Proof of Theorem 2.7. Firstly, note that the denominators of the F-statistics can be rewritten as

$$D_n(m) \equiv \frac{n}{n-K} \sum_{k=1}^K \eta_{k,n} \sigma_{k,n_k}^2(m), \tag{2.56}$$

where the empirical frequencies $\eta_{k,n} = n_k/n$ converge to the population frequencies η_k and the empirical variances $\sigma_{k,n_k}^2(m) = \sum_i (X_{k,i}(m) - \bar{X}_{k,n_k}(m))^2/n_k$ are consistent estimators of the population variances $\sigma_k^2(m)$, i.e., $\eta_{k,n} \to \eta_k > 0$ and $\sigma_{k,n_k}^2(m) \xrightarrow{P} \sigma_k^2(m)$, $k = 1, \ldots, K$. Thus, as $n \to \infty$,

$$D_n(m) \xrightarrow{P} D(m) \equiv \sum_{k=1}^K \eta_k \sigma_k^2(m). \tag{2.57}$$

The numerators of the F-statistics can be rewritten as quadratic forms

$$N_n(m) \equiv \frac{1}{K-1} \sum_{k=1}^{K} \left(Y_{k,n_k}(m) - \sqrt{\eta_{k,n}} \sum_{k'=1}^{K} \sqrt{\eta_{k',n}} Y_{k',n_{k'}}(m) \right)^2 \tag{2.58}$$
$$= \frac{1}{K-1} \left(\sum_{k=1}^{K} Y_{k,n_k}^2(m) - 2 \left(\sum_{k=1}^{K} \sqrt{\eta_{k,n}} Y_{k,n_k}(m) \right) \left(\sum_{k'=1}^{K} \sqrt{\eta_{k',n}} Y_{k',n_{k'}}(m) \right) + \sum_{k=1}^{K} \eta_{k,n} \left(\sum_{k'=1}^{K} \sqrt{\eta_{k',n}} Y_{k',n_{k'}}(m) \right)^2 \right)$$
$$= \frac{1}{K-1} \left(\sum_{k=1}^{K} Y_{k,n_k}^2(m) - \left(\sum_{k=1}^{K} \sqrt{\eta_{k,n}} Y_{k,n_k}(m) \right)^2 \right)$$
$$= \frac{1}{K-1} \left(\sum_{k=1}^{K} (1-\eta_{k,n}) Y_{k,n_k}^2(m) - \sum_{\substack{k=1 \\ k \neq k'}}^{K} \sum_{k'=1}^{K} \sqrt{\eta_{k,n}\eta_{k',n}} Y_{k,n_k}(m) Y_{k',n_{k'}}(m) \right),$$

where $Y_{k,n_k} = (Y_{k,n_k}(m) : m = 1, \ldots, M)$ are K independent M-vectors defined by $Y_{k,n_k}(m) = \sqrt{n_k}(\bar{X}_{k,n_k}(m) - \bar{\psi}(m))$ and $\bar{\psi}(m) = \sum_k \eta_k \psi_k(m)$, $k = 1, \ldots, K$.

Thus, asymptotically, one can approximate the F-statistics $T_n = (T_n(m) : m = 1, \ldots, M)$ by a random M-vector $Z_n^F = (Z_n^F(m) : m = 1, \ldots, M)$ of quadratic forms, as follows,

$$T_n(m) \approxeq \frac{N_n(m)}{D(m)} \tag{2.59}$$
$$\approxeq \frac{1}{(K-1)\sum_{k=1}^{K} \eta_k \sigma_k^2(m)} \times \left(\sum_{k=1}^{K} (1-\eta_k) Y_{k,n_k}^2(m) - \sum_{\substack{k=1 \\ k \neq k'}}^{K} \sum_{k'=1}^{K} \sqrt{\eta_k \eta_{k'}} Y_{k,n_k}(m) Y_{k',n_{k'}}(m) \right)$$
$$\equiv Z_n^F(m).$$

That is, the mth element $Z_n^F(m)$ of the random M-vector Z_n^F is a simple quadratic function $f_m(Y_{1,n_1}, \ldots, Y_{K,n_K})$ of the mth elements $Y_{k,n_k}(m)$ of the

K random M-vectors Y_{k,n_k}, $k = 1, \ldots, K$. The M-vector Z_n^F may be expressed as $Z_n^F = f(Y_{1,n_1}, \ldots, Y_{K,n_K}) = (f_m(Y_{1,n_1}, \ldots, Y_{K,n_K}) : m = 1, \ldots, M)$.

By the Central Limit Theorem (Theorem B.4),

$$(Y_{k,n_k}(m) : m \in \mathcal{H}_0) \stackrel{\mathcal{L}}{\Rightarrow} (Y_k(m) : m \in \mathcal{H}_0),$$

for independent Gaussian M-vectors $Y_k = (Y_k(m) : m = 1, \ldots, M) \sim \mathrm{N}(0, \sigma_k)$, $k = 1, \ldots, K$. By the Continuous Mapping Theorem (Theorem B.3), it then follows that

$$(T_n(m) : m \in \mathcal{H}_0) \stackrel{\mathcal{L}}{\Rightarrow} (Z^F(m) : m \in \mathcal{H}_0) \sim Q_{0,\mathcal{H}_0}^F,$$

where $Z^F = f(Y_1, \ldots, Y_K)$ is the random M-vector of quadratic forms with joint distribution Q_0^F, defined as in Equations (2.51)–(2.53).

Hence, asymptotic null domination Assumption NDV, for the number of Type I errors, is satisfied with equality by the F-statistics T_n and the null distribution Q_0^F.

Note that the F-statistics $(T_n(m) : m \in \mathcal{H}_1)$ for the false null hypotheses have infinite limits. Indeed, for $m \in \mathcal{H}_1$, $Y_{k,n_k}(m) = \sqrt{n_k}(\bar{X}_{k,n_k}(m) - \psi_k(m)) + \sqrt{n_k}(\psi_k(m) - \bar{\psi}(m))$ converges to either $+\infty$ or $-\infty$ for some k, hence $\lim_n T_n(m) = +\infty$.

The moments of $Z^F(m)$ are obtained from standard results on quadratic forms (Theorem 1, p. 55, and Corollary 1.3, p. 57, Searle (1971)). In the special case of constant variances across populations, i.e., $Diag(\sigma_k) = (\sigma^2(m) : m = 1, \ldots, M)$, the matrices $(K-1)A_m \operatorname{Cov}[\widetilde{Y}_m]$ are idempotent; hence, the quadratic forms $(K-1)Z^F(m)$ have marginal $\chi^2(K-1)$-distributions (Theorem 2, p. 57, Searle (1971)).

□

The above theorem proposes a test statistics null distribution Q_0^F derived specifically in terms of the F-statistics T_n of Equation (2.50). This null distribution is the joint distribution of an M-vector of quadratic forms of Gaussian random variables and is entirely specified by the population covariance matrices σ_k and frequencies η_k (via the matrices A_m and the random M-vectors $Y_k \sim \mathrm{N}(0, \sigma_k)$, defining the quadratic forms Z^F in Equations (2.51)–(2.53)). Although properties of the marginal distributions of the F-statistics follow from standard univariate results on quadratic forms, Theorem 2.7 provides as a main contribution a joint null distribution Q_0^F that takes into account the dependence structure of these test statistics. Specifically, the dependence structure of the null distribution Q_0^F is implied by the dependence structure of the data generating distributions P_k, as indicated by the presence of the covariance matrices σ_k in the definition of the quadratic forms Z^F.

Note that the F-statistics $(T_n(m) : m \in \mathcal{H}_0)$ for the true null hypotheses converge weakly to the $\mathcal{H}_0$-specific joint null distribution $Q_{0,\mathcal{H}_0}^F$. Asymptotic joint null domination Assumption jtNDT for the test statistics $(T_n(m) : m \in \mathcal{H}_0)$ is therefore satisfied with equality. In contrast, the F-statistics $(T_n(m) :$

$m \in \mathcal{H}_1$) for the false null hypotheses have infinite limits, i.e., $\lim_n T_n(m) = +\infty$, for $m \in \mathcal{H}_1$. Key Assumption NDV of asymptotic null domination for the number of Type I errors is nonetheless satisfied, as it only concerns the test statistics $(T_n(m) : m \in \mathcal{H}_0)$ corresponding to the true null hypotheses. In other words, neither convergence to nor the weaker domination by Q_0^F is needed for the false null hypotheses.

Gaussian data generating distributions with constant variances across populations

In the special case of Gaussian data generating distributions $P_k = \mathrm{N}(\psi_k, \sigma_k)$, with constant variances across populations, i.e., $Diag(\sigma_k) = (\sigma^2(m) : m = 1, \ldots, M)$, the test statistics T_n have marginal *non-central F-distributions* (Section 2.4, Searle (1971)). Specifically, $T_n(m) \sim \mathrm{F}(\nu_1, \nu_2, \upsilon_n(m))$, where the degrees of freedom are $\nu_1 = (K-1)$ and $\nu_2 = (n-K)$ and the non-centrality parameter is

$$\upsilon_n(m) = \frac{1}{\sigma^2(m)} \sum_{k=1}^{K} n_k(\psi_k(m) - \bar{\psi}(m))^2, \qquad \bar{\psi}(m) = \sum_{k=1}^{K} \eta_k \psi_k(m). \quad (2.60)$$

For the true null hypotheses (i.e., for $m \in \mathcal{H}_0$), $\upsilon_n(m) = 0$. For the false null hypotheses (i.e., for $m \in \mathcal{H}_1$) and non-local alternative mean parameters $\psi_k(m)$, $\lim_n \upsilon_n(m) = +\infty$. In addition, $\lim_n \nu_2 = +\infty$.

The means and variances of the F-statistics are given by, respectively,

$$\mathrm{E}[T_n(m)] = \frac{(\nu_1 + \upsilon_n(m))\nu_2}{\nu_1(\nu_2 - 2)} \to \begin{cases} 1, & \text{if } m \in \mathcal{H}_0 \\ +\infty, & \text{if } m \in \mathcal{H}_1 \end{cases} \quad (2.61)$$

and

$$\begin{aligned} \mathrm{Var}[T_n(m)] &= \frac{2\nu_2^2\left(\nu_1^2 + (2\upsilon_n(m) + \nu_2 - 2)\nu_1 + \upsilon_n(m)(\upsilon_n(m) + 2\nu_2 - 4)\right)}{\nu_1^2(\nu_2 - 4)(\nu_2 - 2)^2} \\ &\to \begin{cases} 2/(K-1), & \text{if } m \in \mathcal{H}_0 \\ +\infty, & \text{if } m \in \mathcal{H}_1 \end{cases}. \end{aligned} \quad (2.62)$$

Furthermore, the F-statistics $T_n(m)$ have asymptotic marginal *non-central χ^2-distributions*, with $(K-1)$ degrees of freedom and non-centrality parameter $\upsilon_n(m)$. That is,

$$(K-1)T_n(m) \overset{\mathcal{L}}{\Rightarrow} \chi^2(K-1, \upsilon_n(m)). \quad (2.63)$$

Comparison to null shift and scale-transformed null distribution

Instead of the F-statistic-specific null distribution Q_0^F proposed in Theorem 2.7, one could apply the general construction of Theorem 2.2, whereby the null distribution Q_0 is defined as the asymptotic distribution of the M-vector $Z_n = (Z_n(m) : m = 1, \ldots, M)$ of null shift and scale-transformed test statistics,

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

For F-statistics, the null values $\lambda_0(m)$ and $\tau_0(m)$ are based on, respectively, the means and variances of the quadratic forms Z^F (Equation (2.54)). In the special case of constant variances across populations, i.e., $Diag(\sigma_k) = (\sigma^2(m) : m = 1, \ldots, M)$, the null values do not depend on the unknown data generating distributions P_k and are given by $\lambda_0(m) = 1$ and $\tau_0(m) = 2/(K-1)$. Otherwise, one needs to estimate the population frequencies η_k and variances $\sigma_k^2(m)$ in order to use Equation (2.54).

Note that, in the construction of $Z_n(m)$, it is important to scale the test statistics $T_n(m)$ by $\nu_{0,n}(m) = \sqrt{\min\{1, \tau_0(m)/\,\mathrm{Var}[T_n(m)]\}}$, as these F-statistics converge to infinity for non-local alternative hypotheses. Without this scaling, one could have asymptotically infinite test statistic cut-offs and hence no power against the alternative hypotheses.

The F-statistic-specific null distribution Q_0^F of Theorem 2.7 and the general null distribution Q_0 of Theorem 2.2 are the same for the true null hypotheses ($m \in \mathcal{H}_0$), but may differ for the false null hypotheses ($m \in \mathcal{H}_1$). Thus, in choosing between Q_0^F and Q_0, the main issue is power.

Comparison to null quantile-transformed null distribution

Section 4.2 of van der Laan and Hubbard (2006) addresses a similar testing problem using the new null quantile-transformed null distribution introduced in Section 2.4. Specifically, for χ^2-statistics T_n and marginal null distributions $q_{0,m} = \chi^2(K-1)$, Theorem 3 proves that the null quantile-transformed test statistics $(\breve{Z}_n(m) : m \in \mathcal{H}_0)$ for the true null hypotheses converge weakly to the $\mathcal{H}_0$-specific subdistribution $Q^{\chi}_{0,\mathcal{H}_0}$, of a joint null distribution Q_0^{χ} with marginal $\chi^2(K-1)$-distributions. Theorem 3 further provides conditions under which estimators of Q_0^{χ} lead to proper Type I error control.

As previously discussed, the ability to control marginal null distributions should confer greater power to this new approach.

2.7.3 Estimation of the test statistics null distribution

A consistent estimator Q_{0n}, of the general null shift and scale-transformed null distribution Q_0 of Theorem 2.2, can be obtained using bootstrap Procedure 2.3, with null values $\lambda_0(m)$ and $\tau_0(m)$ defined as in Equation (2.54). In the special case of constant variances across populations, the null values are $\lambda_0(m) = 1$ and $\tau_0(m) = 2/(K-1)$. Otherwise, one needs to estimate the null values, as they depend on the unknown population frequencies η_k and variances $\sigma_k^2(m)$.

Estimation approaches for the general null quantile-transformed null distribution of Section 2.4 are discussed in Section 4.2 of van der Laan and Hubbard (2006).

Alternately, one can exploit properties of F-statistics to derive a consistent estimator Q_{0n}^F of the F-statistic-specific null distribution Q_0^F of Theorem 2.7. Recall that this null distribution is the joint distribution of an M-vector of quadratic forms of Gaussian random variables and is entirely specified by the population covariance matrices σ_k and frequencies η_k (Equations (2.51)–(2.53)). The main task is therefore to derive estimators $\sigma_{k,n}$ and $\eta_{k,n}$ of these population covariance matrices and frequencies, based on the K random samples $\mathcal{X}_{k,n_k} = \{X_{k,i} : i = 1, \ldots, n_k\}$, $k = 1, \ldots, K$. The null distribution Q_0^F may then simply be estimated by the joint distribution Q_{0n}^F of an M-vector of quadratic forms, defined using the empirical analogues of Equations (2.51)–(2.53), in terms of independent Gaussian M-vectors $Y_k \sim \mathrm{N}(0, \sigma_{k,n})$. Unlike the general non-parametric bootstrap estimator of Procedure 2.3, for the null distribution Q_0 of Theorem 2.2, this F-statistic-specific estimator has the advantage of being continuous.

Finally, another F-statistic-specific approach involves bootstrapping the centered observations $X_{k,i} - \bar{X}_{k,n_k}$ and estimating the null distribution Q_0^F by the bootstrap distribution of the corresponding F-statistics. In this method, the estimated null distribution of the test statistics is based on a data generating null distribution.

The last two approaches both provide consistent estimators of the F-statistic-specific null distribution Q_0^F of Theorem 2.7.

2.8 Weak and strong Type I error control and subset pivotality

As mentioned in Section 2.2.4, the multiple testing methodology developed in this book differs in a number of fundamental aspects from existing approaches to Type I error control and the choice of a test statistics null distribution. Our proposed multiple testing procedures are: (i) only concerned with controlling the Type I error rate under the *true data generating distribution* P, i.e., under the joint distribution $Q_n = Q_n(P)$ of the test statistics T_n implied by P; (ii) based on a *test statistics null distribution* rather than a data generating null distribution.

In this regard, one of our main contributions is the general characterization (Section 2.2.3) and explicit construction (Sections 2.3 and 2.4) of proper null distributions Q_0 (and estimators thereof, Q_{0n}) for the test statistics T_n. Procedures based on the proposed null distributions provide Type I error control for general data generating distributions (with arbitrary dependence structures among variables), null hypotheses, and test statistics.

In our framework, the notions of weak and strong control of a Type I error rate become irrelevant and Type I error control does not involve associated restrictive assumptions such as subset pivotality. The present section attempts nonetheless to formalize these concepts and discusses how they relate to the approach introduced in Section 2.2.

2.8.1 Weak and strong control of a Type I error rate

Usual definitions of weak and strong Type I error control

As discussed in Hochberg and Tamhane (1987, p. 3) and Westfall and Young (1993, p. 9–10), the multiple testing literature commonly distinguishes between weak and strong control of a Type I error rate.

Weak control refers to control of the Type I error rate under a data generating distribution P_0 that satisfies the *complete null hypothesis*, $H_0^C = \prod_{m=1}^M H_0(m) = \prod_{m=1}^M \mathrm{I}(P \in \mathcal{M}(m)) = \mathrm{I}\left(P \in \cap_{m=1}^M \mathcal{M}(m)\right)$, that all M null hypotheses are true, i.e., under a distribution P_0 that belongs to the intersection $\cap_{m=1}^M \mathcal{M}(m)$ of all M submodels.

In contrast, *strong control*, as defined in Westfall and Young (1993), considers *all* 2^M *possible subsets of null hypotheses*, $\mathcal{J}_0 \subseteq \{1, \ldots, M\}$, and refers to control of the Type I error rate under each of 2^M distributions $P_{\mathcal{J}_0} \in \cap_{m \in \mathcal{J}_0} \mathcal{M}(m)$ that satisfy subsets of null hypotheses $\mathcal{J}_0$. In particular, strong control implies weak control for $\mathcal{J}_0 = \{1, \ldots, M\}$.

As detailed below, the definitions of weak and strong control implicitly assume the existence of a *mapping* $\mathcal{J}_0 \to P_{\mathcal{J}_0}$, from subsets $\mathcal{J}_0$ of null hypotheses to data generating distributions $P_{\mathcal{J}_0} \in \cap_{m \in \mathcal{J}_0} \mathcal{M}(m)$ that satisfy each of the null hypotheses in $\mathcal{J}_0$.

It is important to recognize that, although strong control does consider the subset $\mathcal{H}_0 = \mathcal{H}_0(P)$ of true null hypotheses corresponding to the true data generating distribution P, Type I error control under P is not guaranteed by strong control, unless the mapping $\mathcal{J}_0 \to P_{\mathcal{J}_0}$ results in $P_{\mathcal{H}_0} = P$.

Defining a data generating distribution that satisfies a given subset of null hypotheses

In much of the multiple testing literature, Type I error rates are defined loosely in terms of probabilities *given subsets of null hypotheses*, rather than probabilities *under distributions that satisfy subsets of null hypotheses*, i.e., *under distributions that belong to intersections of submodels*. For example, Westfall and Young (1993, p. 9) refer to the FWEP as the family-wise error rate "computed under the *partial null hypothesis* (meaning that some subcollection of nulls, say $H_{j_1}, \ldots, H_{j_t}$, is true)" and provide the following definition in their Equation (1.2),

$$FWEP = \Pr(\text{Reject at least one } H_i,\ i = j_1, \ldots, j_t | H_{j_1}, \ldots, H_{j_t} \text{ are true}).$$

As discussed in Dudoit et al. (2004b) and Pollard and van der Laan (2004), such a quantity is not well-defined, because Type I error rates are parameters of a distribution for the number of Type I errors (and possibly the number of rejected hypotheses, as for the FDR) and can only be defined meaningfully with respect to such a distribution (Section 1.2.9). A more precise definition would be that FWEP is the family-wise error rate *under a data generating distribution* $P_{\mathcal{J}_0}$ that satisfies a certain subset $\mathcal{J}_0 = \{j_1, \ldots, j_t\}$ of null hypotheses, i.e., defined such that $P_{\mathcal{J}_0} \in \cap_{m \in \mathcal{J}_0} \mathcal{M}(m)$.

This immediately raises the issue of how to map from a subset $\mathcal{J}_0$ of null hypotheses to a well-defined data generating distribution $P_{\mathcal{J}_0} \in \cap_{m \in \mathcal{J}_0} \mathcal{M}(m)$. Except in very simple situations (e.g., null hypotheses concerning the mean vector of a multivariate Gaussian data generating distribution), each subset $\mathcal{J}_0$ of null hypotheses corresponds to a family of possible distributions. One approach is to define the distribution $P_{\mathcal{J}_0}$ as a projection of the true data generating distribution P onto the submodel $\cap_{m \in \mathcal{J}_0} \mathcal{M}(m)$, selecting, for example, the distribution $P_{\mathcal{J}_0} \in \cap_{m \in \mathcal{J}_0} \mathcal{M}(m)$ with the smallest Kullback-Leibler divergence with P. That is,

$$
\begin{aligned}
P_{\mathcal{J}_0} &= \Pi_{KL}(P | \cap_{m \in \mathcal{J}_0} \mathcal{M}(m)) \\
&\equiv \arg\max_{P' \in \cap_{m \in \mathcal{J}_0} \mathcal{M}(m)} \int \log\left(\frac{dP'(x)}{d\mu(x)}\right) dP(x),
\end{aligned}
\tag{2.64}
$$

for a dominating measure μ. Another possibility is to select the distribution $P_{\mathcal{J}_0}$ on the conservative boundary of the submodel $\cap_{m \in \mathcal{J}_0} \mathcal{M}(m)$. The reader is referred to Pollard and van der Laan (2004) for a discussion of multivariate data generating null distributions and proposals for specifying such joint distributions based on projections of the true data generating distribution P onto submodels satisfying subsets of null hypotheses.

However, as discussed by these authors, in many testing problems of interest, one simply cannot identify a data generating null distribution $P_0 \in \cap_{m=1}^{M} \mathcal{M}(m)$ that provides proper control of the Type I error rate under the true data generating distribution P. That is, in many cases, the assumed *null distribution* $Q_{n,\mathcal{H}_0}(P_0)$ and the *true distribution* $Q_{n,\mathcal{H}_0}(P)$ of the $\mathcal{H}_0$-specific subvector $(T_n(m) : m \in \mathcal{H}_0)$ of test statistics have different limits and thus violate null domination Assumption $\mathsf{ND}\Theta$ for the Type I error rate, i.e., $\lim_n \Theta(F_{V_n}) > \Theta(F_{V_0}) = \alpha$. Instead, for the test of single-parameter null hypotheses using t-statistics (Section 2.6), Pollard and van der Laan (2004) recommend using a test statistics null distribution such as the Kullback-Leibler projection of $Q_n = Q_n(P)$ onto the space of multivariate Gaussian distributions with mean vector zero. The projection null distribution corresponds to the null distribution $Q_0^t(P) = \mathrm{N}(0, \Sigma^*(P))$ proposed in Theorem 2.6.

Revised definitions of weak and strong Type I error control

As usual, consider the simultaneous test of M null hypotheses $H_0(m)$, $m = 1, \ldots, M$, based on test statistics $T_n = (T_n(m) : m = 1, \ldots, M)$, with true finite sample joint distribution $Q_n = Q_n(P)$ and null distribution Q_0. Let $\mathcal{C}_n(m) = \mathcal{C}(m; T_n, Q_0, \alpha)$, $m = 1, \ldots, M$, and $\mathcal{R}_n = \mathcal{R}(T_n, Q_0, \alpha)$ denote, respectively, the M rejection regions and corresponding set of rejected null hypotheses, for a MTP with nominal Type I error level α. That is,

$$\mathcal{R}(T_n, Q_0, \alpha) = \{m : T_n(m) \in \mathcal{C}(m; T_n, Q_0, \alpha)\}.$$

Given a subset of null hypotheses $\mathcal{J}_0 \subseteq \{1, \ldots, M\}$, define a data generating distribution $P_{\mathcal{J}_0} \in \cap_{m \in \mathcal{J}_0} \mathcal{M}(m)$ and let $Q_n(P_{\mathcal{J}_0})$ denote the corresponding joint distribution for the test statistics T_n. Following the notation introduced in Equations (2.2) and (2.3), denote the numbers of rejected hypotheses and Type I errors by

$$R_n(\mathcal{J}_0) \equiv R(\mathcal{C}_n | Q_n(P_{\mathcal{J}_0})) = \sum_{m=1}^{M} \mathrm{I}\left(T_n(m) \in \mathcal{C}_n(m)\right) \tag{2.65}$$

and

$$V_n(\mathcal{J}_0) \equiv V(\mathcal{C}_n | Q_n(P_{\mathcal{J}_0})) = \sum_{m \in \mathcal{J}_0} \mathrm{I}\left(T_n(m) \in \mathcal{C}_n(m)\right),$$

respectively, under the assumption that $T_n \sim Q_n(P_{\mathcal{J}_0})$.

Strong control of a Type I error rate at level α requires that

$$\max_{\mathcal{J}_0 \subseteq \{1,\ldots,M\}} \Theta(F_{V_n(\mathcal{J}_0), R_n(\mathcal{J}_0)}) \leq \alpha \qquad \text{[finite sample strong control]}$$

$$\limsup_{n \to \infty} \max_{\mathcal{J}_0 \subseteq \{1,\ldots,M\}} \Theta(F_{V_n(\mathcal{J}_0), R_n(\mathcal{J}_0)}) \leq \alpha \qquad \text{[asymptotic strong control]}. \tag{2.66}$$

Thus, strong control involves considering 2^M distributions $P_{\mathcal{J}_0}$, each corresponding to a subset $\mathcal{J}_0$ of null hypotheses. Note also that this definition of strong control is completely dependent upon the choice of mapping $\mathcal{J}_0 \to P_{\mathcal{J}_0}$.

Weak control corresponds to $\mathcal{J}_0 = \{1, \ldots, M\}$ and $P_0 = P_{\{1,\ldots,M\}}$.

Type I error control under the true data generating distribution P does not necessarily follow from strong control, unless the mapping $\mathcal{J}_0 \to P_{\mathcal{J}_0}$ results in $P_{\mathcal{H}_0} = P$ for $\mathcal{J}_0 = \mathcal{H}_0$. In other words, control under the true P could fail under strong control when an improper mapping is used to define $P_{\mathcal{H}_0}$.

In contrast, as discussed in Section 2.2, the methodology proposed in this book is only concerned with Type I error control under the true data generating distribution P. That is, we only require that Equation (2.66) hold in the special case where $\mathcal{J}_0 = \mathcal{H}_0$ and $P_{\mathcal{H}_0} = P$.

2.8.2 Subset pivotality

In practice, it is not feasible to consider all 2^M possible subsets of null hypotheses and commonly-used single-step and stepwise multiple testing procedures are typically based on cut-offs derived under a data generating distribution P_0 that satisfies the complete null hypothesis $H_0^C = \prod_{m=1}^{M} H_0(m)$, i.e.,

$P_0 \in \cap_{m=1}^{M} \mathcal{M}(m)$. Strong control of a Type I error rate, and in particular control under the true data generating distribution P, are then claimed to follow from weak control under conditions such as subset pivotality.

As stated in Condition 2.1, p. 42, in Westfall and Young (1993), "The distribution of $\mathbf{P}$ has the *subset pivotality* property if the joint distribution of the subvector $\{P_i : i \in K\}$ is identical under the restrictions $\cap_{i \in K} H_{0i}$ and H_0^C, for all subsets $K = \{i_1, \ldots, i_j\}$ of true null hypotheses". In our notation, K is a subset $\mathcal{J}_0 \subseteq \{1, \ldots, M\}$ of null hypotheses and $\mathbf{P}$ refers to the vector $(P_{0n}(m) : m = 1, \ldots, M)$ of unadjusted p-values (Section 1.2.12).

As for the definitions of weak and strong control, subset pivotality implicitly assumes the existence of a mapping $\mathcal{J}_0 \rightarrow P_{\mathcal{J}_0}$, from subsets $\mathcal{J}_0$ of null hypotheses to data generating distributions $P_{\mathcal{J}_0} \in \cap_{m \in \mathcal{J}_0} \mathcal{M}(m)$ that satisfy each of the null hypotheses in $\mathcal{J}_0$. A (finite sample) subset pivotality condition for test statistics can then be stated as

$$Q_{n,\mathcal{J}_0}(P_{\mathcal{J}_0}) = Q_{n,\mathcal{J}_0}(P_0), \qquad \forall\ \mathcal{J}_0 \subseteq \{1, \ldots, M\}, \tag{2.67}$$

in terms of distributions $P_{\mathcal{J}_0}$ corresponding to subsets $\mathcal{J}_0$ of null hypotheses and where $P_0 = P_{\{1,\ldots,M\}}$.

Note that the subset pivotality condition considers all 2^M possible subsets of null hypotheses, and not simply the subset $\mathcal{J}_0 = \mathcal{H}_0(P)$ corresponding to the true data generating distribution P. In this sense, and provided $P_{\mathcal{H}_0} = P$, the assumption is stronger than needed, because it is only of interest to control Type I error rates under the true P, that is, the only relevant condition is $Q_{n,\mathcal{H}_0}(P) = Q_{n,\mathcal{H}_0}(P_0)$ for $\mathcal{J}_0 = \mathcal{H}_0$. In general, however, subset pivotality does not guarantee control under the true P, if an improper mapping $\mathcal{J}_0 \rightarrow P_{\mathcal{J}_0}$ is used and $P_{\mathcal{H}_0} \neq P$.

Finally, as discussed in Section 2.2.4, the subset pivotality assumption in Equation (2.67) differs from our (finite sample) joint null domination Assumption jtNDT which: (i) only considers the subset $\mathcal{J}_0 = \mathcal{H}_0$; (ii) does not require the test statistics null distribution $Q_{0,n}$ or Q_0 to be defined in terms of a data generating null distribution P_0, i.e., $Q_{0,n} = Q_n(P_0)$; (iii) does not require equality of the true and null test statistics distributions, but the weaker null domination, i.e., $Q_{n,\mathcal{H}_0}(P) \geq Q_{n,\mathcal{H}_0}(P_0)$.

2.9 Test statistics null distributions based on bootstrap and permutation data generating distributions

Permutation procedures are widely-used in multiple testing to obtain data generating null distributions P_0 and corresponding test statistics null distributions $Q_n(P_0)$ (Westfall and Young, 1993). This section builds on Pollard and van der Laan (2004) and compares bootstrap- and permutation-based test statistics null distributions.

2.9.1 The two-sample test of means problem

Consider a two-sample test of means problem, with data structure $(X, Y) \sim P \in \mathcal{M}$, where $X = (X(m) : m = 1, \ldots, M)$ is a random M-vector and $Y \in \{1, 2\}$ a binary population label. For Population k, $k = 1, 2$, let $\eta_k = \Pr(Y = k)$ denote the population frequency, let $P_{X|k}$ denote the conditional data generating distribution of X given $Y = k$ (i.e., $X|Y = k \sim P_{X|k}$), and let $\psi_k = (\psi_k(m) : m = 1, \ldots, M) = \mathrm{E}[X|Y = k]$ and $\sigma_k = \mathrm{Cov}[X|Y = k]$ denote, respectively, the conditional M-dimensional mean vector and $M \times M$ covariance matrix of X. Consider testing the following M null hypotheses concerning the differences $\psi(m) = \psi_1(m) - \psi_2(m)$ in conditional means,

$$H_0(m) = \mathrm{I}\,(\psi(m) = 0)\,, \qquad m = 1, \ldots, M. \tag{2.68}$$

Suppose one has a random sample $\mathcal{XY}_n = \{(X_i, Y_i) : i = 1, \ldots, n\}$, of n IID copies of the pair $(X, Y) \sim P \in \mathcal{M}$. Denote the (random) sample size for Population k by $n_k = \sum_i \mathrm{I}\,(Y_i = k)$ and estimate the conditional mean vector ψ_k by the corresponding empirical mean vector $\psi_{k,n_k} = \bar{X}_{k,n_k}$, with elements $\psi_{k,n_k}(m) = \bar{X}_{k,n_k}(m) = \sum_i \mathrm{I}\,(Y_i = k)\, X_i(m)/n_k$. The null hypotheses can be tested using (unstandardized) difference statistics,

$$\begin{aligned} D_n(m) &\equiv \sqrt{n}(\psi_{2,n_2}(m) - \psi_{1,n_1}(m)) \qquad (2.69)\\ &= \sqrt{n} \sum_{i=1}^{n} \left(\frac{\mathrm{I}\,(Y_i = 2)\, X_i(m)}{n_2} - \frac{\mathrm{I}\,(Y_i = 1)\, X_i(m)}{n_1} \right), \qquad m = 1, \ldots, M. \end{aligned}$$

Consider the following two models, $\mathcal{M}$ and $\mathcal{M}_= \subseteq \mathcal{M}$, corresponding, respectively, to general non-parametric and location-shifted conditional data generating distributions $P_{X|1}$ and $P_{X|2}$.

Non-parametric model, $\mathcal{M}$. For the *non-parametric model* $(X, Y) \sim P \in \mathcal{M}$, $X|Y = 1 \sim P_{X|1}$ and $X|Y = 2 \sim P_{X|2}$, where $P_{X|1}$ and $P_{X|2}$ are arbitrary conditional data generating distributions for Populations 1 and 2, respectively.

Location shift model, $\mathcal{M}_=$. For the *location shift model* $(X, Y) \sim P_= \in \mathcal{M}_=$, $X|Y = 1 \sim P_{X|1} = P_X(\cdot - \psi_1)$ and $X|Y = 2 \sim P_{X|2} = P_X(\cdot - \psi_2)$, where P_X is a common M-dimensional distribution with mean vector zero. That is, $P_{X|1}$ and $P_{X|2}$ are identical except for a location shift.

The implications of each model are investigated in terms of the choice of an appropriate null distribution for the test statistics D_n. Model $\mathcal{M}_= \subseteq \mathcal{M}$ makes the strong assumption that, under the complete null hypothesis $H_0^C = \prod_{m=1}^{M} H_0(m) = \mathrm{I}\,(\psi_1 = \psi_2)$, the random vector X has the same conditional distribution in the two populations ($P_{X|1} = P_{X|2}$), that is, X and Y are independent. If one were testing the null hypothesis $\mathrm{I}\left(P_{X|1} = P_{X|2}\right)$ that the conditional data generating distributions are identical for the two populations, then $\mathcal{M}_=$ would clearly be a good choice of model from which to select a data generating null distribution. However, model $\mathcal{M}_=$ may be a poor choice for testing the null hypotheses in Equation (2.68), which only concern differences in means between the two populations and allow, in particular, different covariance structures σ_k in each population.

2.9.2 Distribution of the test statistics under two different data generating distributions

By the Central Limit Theorem (Theorem B.4), the difference statistics D_n have a Gaussian asymptotic distribution. This distribution is fully specified by its mean vector (with elements equal to zero for the true null hypotheses) and its covariance matrix. In what follows, we therefore focus on properties and estimation of the covariance matrix of the test statistics.

For simplicity, and without loss of generality, consider only $M = 2$ null hypotheses, i.e., a bivariate random vector X.

Proposition 2.8, below, provides asymptotic variances and covariances for the difference statistics D_n under two different data generating distributions for (X, Y).

Proposition 2.8. [Asymptotic variances and covariances of difference statistics for two-sample test of means, under two different data generating distributions] *Consider a data structure $(X, Y) \sim P \in \mathcal{M}$, where $X = (X(1), X(2))$ is a bivariate random vector and $Y \in \{1, 2\}$ is a binary population label, with $\eta_k = \Pr(Y = k)$, $k = 1, 2$. Let $P_{X|k}$ denote a bivariate distribution, with mean vector $\psi_k = [\psi_k(1), \psi_k(2)]^\top$ and covariance matrix $\sigma_k = (\sigma_k(m, m') : m, m' = 1, 2)$, $k = 1, 2$. Specifically, consider the following two data generating distributions for (X, Y).*

Non-parametric data generating distribution, $\boldsymbol{P}$. *For $(X, Y) \sim P$, the conditional distribution of X given $Y = k$ is $P_{X|k}$, that is, $X|Y = k \sim P_{X|k}$, $k = 1, 2$.*

Independence data generating distribution, $\boldsymbol{P_\perp}$. *For $(X, Y) \sim P_\perp$, X and Y are independent and X has the mixture distribution $X \sim \eta_1 P_{X|1} + \eta_2 P_{X|2}$.*

Then, for a random sample $\mathcal{XY}_n = \{(X_i, Y_i) : i = 1, \ldots, n\}$, of n IID copies of the pair $(X, Y) \sim P$, the asymptotic covariance matrix of the difference statistics $D_n = (D_n(1), D_n(2))$ of Equation (2.69) is given by

$$\varsigma \equiv \lim_{n\to\infty} \mathrm{Cov}_P[D_n] = \begin{bmatrix} \frac{\sigma_1(1,1)}{\eta_1} + \frac{\sigma_2(1,1)}{\eta_2} & \frac{\sigma_1(1,2)}{\eta_1} + \frac{\sigma_2(1,2)}{\eta_2} \\ \frac{\sigma_1(1,2)}{\eta_1} + \frac{\sigma_2(1,2)}{\eta_2} & \frac{\sigma_1(2,2)}{\eta_1} + \frac{\sigma_2(2,2)}{\eta_2} \end{bmatrix}. \tag{2.70}$$

For $(X, Y) \sim P_\perp$,

$$\varsigma_\perp \equiv \lim_{n\to\infty} \mathrm{Cov}_{P_\perp}[D_n] = \begin{bmatrix} \frac{\sigma_1(1,1)}{\eta_2} + \frac{\sigma_2(1,1)}{\eta_1} & \frac{\sigma_1(1,2)}{\eta_2} + \frac{\sigma_2(1,2)}{\eta_1} \\ \frac{\sigma_1(1,2)}{\eta_2} + \frac{\sigma_2(1,2)}{\eta_1} & \frac{\sigma_1(2,2)}{\eta_2} + \frac{\sigma_2(2,2)}{\eta_1} \end{bmatrix}. \tag{2.71}$$

It is interesting to note that the asymptotic covariance matrices ς and $\varsigma_\perp$ of the difference statistics D_n are identical, except for the roles of population frequencies η_1 and η_2 being reversed.

The expressions for ς and $\varsigma_\perp$ illustrate that, for most values of the parameters σ_k and η_k, $k = 1, 2$, the difference statistics have different asymptotic distributions for data generating distributions P and $P_\perp$. If, however, either (i) $\eta_1 = \eta_2$ or (ii) $\sigma_1 = \sigma_2$, then the asymptotic distributions are the same for both scenarios, i.e., $\varsigma = \varsigma_\perp$.

As discussed below, the bootstrap estimator of the distribution of the test statistics D_n converges to the asymptotic distribution of D_n under P, while the permutation estimator of the distribution of D_n converges to the asymptotic distribution of D_n under $P_\perp$. Thus, under the reduced location shift model $\mathcal{M}_=$, for which $\sigma_1 = \sigma_2$, a permutation data generating distribution yields a sensible test statistics null distribution.

It is somewhat surprising that, even when the data generating distribution P is not an element of the reduced model $\mathcal{M}_=$ (e.g., $\sigma_1 \neq \sigma_2$), one still has $\varsigma = \varsigma_\perp$ when $\eta_1 = \eta_2$. Thus, in the case of equal population frequencies (i.e., $\eta_1 = \eta_2$), permutation distributions, corresponding to the independence data generating distribution $P_\perp$, yield valid test statistics null distributions.

In summary, Proposition 2.8 suggests that, unless either (i) $\eta_1 = \eta_2$ or (ii) $\sigma_1 = \sigma_2$, one should use the bootstrap (rather than permutation) to estimate the null distribution of the test statistics D_n, since the bootstrap preserves the covariance structure ς of these test statistics. However, for equal population frequencies (i.e., $\eta_1 = \eta_2$) or covariance structures (i.e., $\sigma_1 = \sigma_2$, as in model $\mathcal{M}_=$), one could use permutation estimators, because the asymptotic covariance matrix of the test statistics D_n is the same for both data generating distributions P and $P_\perp$ (i.e., $\varsigma = \varsigma_\perp$). Furthermore, permutation estimators of the covariance matrix tend to be more efficient than non-parametric bootstrap estimators, because they correspond to a smaller model and make use of all n observations (Pollard and van der Laan, 2004).

Similar conclusions apply to the usual (standardized) two-sample Welch t-statistics,

$$T_n(m) \equiv \frac{\psi_{2,n_2}(m) - \psi_{1,n_1}(m)}{\sqrt{\frac{\sigma^2_{1,n_1}(m)}{n_1} + \frac{\sigma^2_{2,n_2}(m)}{n_2}}}, \tag{2.72}$$

where n_k, $\psi_{k,n_k}(m)$, and $\sigma^2_{k,n_k}(m)$ denote, respectively, the sample size, empirical means, and empirical variances, for Population k, $k = 1, 2$.

Proof of Proposition 2.8. The derivations of variances and covariances for the difference statistics $D_n = (D_n(1), D_n(2))$ are similar for the two data generating distributions P and $P_\perp$ and make use of the Double Expectation Theorem. For simplicity, and without loss of generality, assume that both null hypotheses $H_0(1)$ and $H_0(2)$ are true and that the mean vectors for $P_{X|1}$ and $P_{X|2}$ are zero, i.e., $\psi_1 = \psi_2 = [0, 0]^\top$. Then,

$\mathrm{E}[D_n] = [0,0]^\top$ and $\mathrm{Cov}[D_n] = \mathrm{E}[D_n D_n^\top]$ under both distributions P and $P_\perp$. Let $\mathcal{Y}_n = \{Y_i : i = 1, \ldots, n\}$.

Variances. First, derive the asymptotic variances of the difference statistics $D_n(m)$.

$$
\begin{aligned}
\lim_{n\to\infty} \mathrm{Var}[D_n(m)] &= \lim_{n\to\infty} \mathrm{E}[D_n^2(m)] \\
&= \lim_{n\to\infty} n\,\mathrm{E}\left[\mathrm{E}\left[\left(\sum_{i=1}^n \left(\frac{\mathrm{I}\,(Y_i = 2)\,X_i(m)}{n_2} - \frac{\mathrm{I}\,(Y_i = 1)\,X_i(m)}{n_1}\right)\right)^2 \middle| \mathcal{Y}_n\right]\right] \\
&= \lim_{n\to\infty} n\,\mathrm{E}\left[\mathrm{E}\left[\sum_{i=1}^n \left(\frac{\mathrm{I}\,(Y_i = 2)\,X_i(m)}{n_2} - \frac{\mathrm{I}\,(Y_i = 1)\,X_i(m)}{n_1}\right)^2 \middle| \mathcal{Y}_n\right]\right] \\
&= \lim_{n\to\infty} n\,\mathrm{E}\left[\mathrm{E}\left[\sum_{i=1}^n \left(\frac{\mathrm{I}\,(Y_i = 2)\,X_i^2(m)}{n_2^2} + \frac{\mathrm{I}\,(Y_i = 1)\,X_i^2(m)}{n_1^2}\right) \middle| \mathcal{Y}_n\right]\right] \\
&= \lim_{n\to\infty} n\,\mathrm{E}\left[\sum_{i=1}^n \left(\frac{\mathrm{I}\,(Y_i = 2)\,\mathrm{E}[X_i^2(m)|Y_i = 2]}{n_2^2}\right.\right. \\
&\qquad\qquad\qquad \left.\left. + \frac{\mathrm{I}\,(Y_i = 1)\,\mathrm{E}[X_i^2(m)|Y_i = 1]}{n_1^2}\right)\right] \\
&= \mathrm{E}[X^2(m)|Y = 2] \lim_{n\to\infty} \mathrm{E}\left[\frac{n}{n_2}\right] + \mathrm{E}[X^2(m)|Y = 1] \lim_{n\to\infty} \mathrm{E}\left[\frac{n}{n_1}\right] \\
&= \frac{\mathrm{E}[X^2(m)|Y = 2]}{\eta_2} + \frac{\mathrm{E}[X^2(m)|Y = 1]}{\eta_1}.
\end{aligned}
$$

The third equality follows by noting that the (X_i, Y_i) are independent, with $\mathrm{E}[X_i(m)|Y_i = k] = 0$, $m = 1, 2$, $k = 1, 2$; the fourth from $\mathrm{I}\,(Y_i = 1)\,\mathrm{I}\,(Y_i = 2) = 0$; the sixth from the fact that the (X_i, Y_i) are identically distributed; and the seventh from $\lim_n n_k/n = \eta_k$ a.s., $k = 1, 2$.

When $(X, Y) \sim P$,

$$
\begin{aligned}
\lim_{n\to\infty} \mathrm{Var}[D_n(m)] &= \frac{\mathrm{E}[X^2(m)|Y = 1]}{\eta_1} + \frac{\mathrm{E}[X^2(m)|Y = 2]}{\eta_2} \\
&= \frac{\sigma_1(m,m)}{\eta_1} + \frac{\sigma_2(m,m)}{\eta_2}, \qquad m = 1, 2.
\end{aligned}
$$

Similarly, when $(X, Y) \sim P_\perp$, the asymptotic variance of the difference statistic $D_n(m)$ is as follows.

$$\begin{aligned}\lim_{n\to\infty}\operatorname{Var}[D_n(m)] &= \frac{\mathrm{E}[X^2(m)|Y=1]}{\eta_1}+\frac{\mathrm{E}[X^2(m)|Y=2]}{\eta_2}\\ &= \frac{\mathrm{E}[X^2(m)]}{\eta_1}+\frac{\mathrm{E}[X^2(m)]}{\eta_2}\\ &= \left(\frac{1}{\eta_1}+\frac{1}{\eta_2}\right)(\eta_1\sigma_1(m,m)+\eta_2\sigma_2(m,m))\\ &= \frac{\sigma_1(m,m)}{\eta_2}+\frac{\sigma_2(m,m)}{\eta_1}, \qquad m=1,\,2.\end{aligned}$$

The second and third equalities follow by noting that X and Y are independent, with X having the mixture distribution $X \sim \eta_1 P_{X|1} + \eta_2 P_{X|2}$, so that $\operatorname{Var}[X(m)] = \mathrm{E}[X^2(m)] = \mathrm{E}[X^2(m)|Y=1] = \mathrm{E}[X^2(m)|Y=2] = \eta_1\sigma_1(m,m)+\eta_2\sigma_2(m,m)$, $m=1,\,2$.

Covariances. Now consider the asymptotic covariance between the difference statistics $D_n(1)$ and $D_n(2)$.

$$\begin{aligned}\lim_{n\to\infty}\operatorname{Cov}[D_n(1),D_n(2)] &= \lim_{n\to\infty}\mathrm{E}[D_n(1)D_n(2)]\\
&= \lim_{n\to\infty} n\,\mathrm{E}\Bigg[\mathrm{E}\Bigg[\sum_{i=1}^n\left(\frac{\mathrm{I}\,(Y_i=2)\,X_i(1)}{n_2}-\frac{\mathrm{I}\,(Y_i=1)\,X_i(1)}{n_1}\right)\\
&\qquad\times\sum_{i=1}^n\left(\frac{\mathrm{I}\,(Y_i=2)\,X_i(2)}{n_2}-\frac{\mathrm{I}\,(Y_i=1)\,X_i(2)}{n_1}\right)\Bigg|\,\mathcal{Y}_n\Bigg]\Bigg]\\
&= \lim_{n\to\infty} n\,\mathrm{E}\Bigg[\mathrm{E}\Bigg[\sum_{i=1}^n\left(\frac{\mathrm{I}\,(Y_i=2)\,X_i(1)}{n_2}-\frac{\mathrm{I}\,(Y_i=1)\,X_i(1)}{n_1}\right)\\
&\qquad\times\left(\frac{\mathrm{I}\,(Y_i=2)\,X_i(2)}{n_2}-\frac{\mathrm{I}\,(Y_i=1)\,X_i(2)}{n_1}\right)\Bigg|\,\mathcal{Y}_n\Bigg]\Bigg]\\
&= \lim_{n\to\infty} n\,\mathrm{E}\Bigg[\mathrm{E}\Bigg[\sum_{i=1}^n\Bigg(\frac{\mathrm{I}\,(Y_i=2)\,X_i(1)X_i(2)}{n_2^2}\\
&\qquad+\frac{\mathrm{I}\,(Y_i=1)\,X_i(1)X_i(2)}{n_1^2}\Bigg)\Bigg|\,\mathcal{Y}_n\Bigg]\Bigg]\\
&= \lim_{n\to\infty} n\,\mathrm{E}\Bigg[\sum_{i=1}^n\Bigg(\frac{\mathrm{I}\,(Y_i=2)\,\mathrm{E}[X_i(1)X_i(2)|Y_i=2]}{n_2^2}\\
&\qquad+\frac{\mathrm{I}\,(Y_i=1)\,\mathrm{E}[X_i(1)X_i(2)|Y_i=1]}{n_1^2}\Bigg)\Bigg]\\
&= \mathrm{E}[X(1)X(2)|Y=2]\lim_{n\to\infty}\mathrm{E}\left[\frac{n}{n_2}\right]+\mathrm{E}[X(1)X(2)|Y=1]\lim_{n\to\infty}\mathrm{E}\left[\frac{n}{n_1}\right]\\
&= \frac{\mathrm{E}[X(1)X(2)|Y=2]}{\eta_2}+\frac{\mathrm{E}[X(1)X(2)|Y=1]}{\eta_1}.\end{aligned}$$

The third equality follows by noting that the (X_i, Y_i) are independent, with $\mathrm{E}[X_i(m)|Y_i = k] = 0$, $m = 1, 2$, $k = 1, 2$; the fourth from $\mathrm{I}(Y_i = 1)\,\mathrm{I}(Y_i = 2) = 0$; the sixth from the fact that the (X_i, Y_i) are identically distributed; and the seventh from $\lim_n n_k/n = \eta_k$ a.s., $k = 1, 2$.

When $(X, Y) \sim P$,

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Cov}[D_n(1), D_n(2)] &= \frac{\mathrm{E}[X(1)X(2)|Y=1]}{\eta_1} + \frac{\mathrm{E}[X(1)X(2)|Y=2]}{\eta_2} \\ &= \frac{\sigma_1(1,2)}{\eta_1} + \frac{\sigma_2(1,2)}{\eta_2}.\end{aligned}$$

Similarly, when $(X, Y) \sim P_\perp$, the asymptotic covariance of the difference statistics $D_n(1)$ and $D_n(2)$ is as follows.

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Cov}[D_n(1), D_n(2)] &= \frac{\mathrm{E}[X(1)X(2)|Y=1]}{\eta_1} + \frac{\mathrm{E}[X(1)X(2)|Y=2]}{\eta_2} \\ &= \frac{\mathrm{E}[X(1)X(2)]}{\eta_1} + \frac{\mathrm{E}[X(1)X(2)]}{\eta_2} \\ &= \left(\frac{1}{\eta_1} + \frac{1}{\eta_2}\right)(\eta_1\sigma_1(1,2) + \eta_2\sigma_2(1,2)) \\ &= \frac{\sigma_1(1,2)}{\eta_2} + \frac{\sigma_2(1,2)}{\eta_1}.\end{aligned}$$

The second and third equalities follow by noting that X and Y are independent, with X having the mixture distribution $X \sim \eta_1 P_{X|1} + \eta_2 P_{X|2}$, so that $\mathrm{Cov}[X(1), X(2)] = \mathrm{E}[X(1)X(2)] = \mathrm{E}[X(1)X(2)|Y = 1] = \mathrm{E}[X(1)X(2)|Y = 2] = \eta_1\sigma_1(1, 2) + \eta_2\sigma_2(1, 2)$.

□

2.9.3 Bootstrap and permutation test statistics null distributions

As suggested by Proposition 2.8, the non-parametric model $\mathcal{M}$ and the smaller location shift model $\mathcal{M}_=$ imply different bootstrap sampling distributions for estimating the distribution of the test statistics $D_n = (D_n(m) : m = 1, \ldots, M)$. In particular, each model implies different data generating and test statistics null distributions. For non-parametric model $\mathcal{M}$, one samples from the joint empirical distribution of the pair (X, Y), whereas for reduced model $\mathcal{M}_=$, one samples from a model-based estimator of the data generating distribution.

Bootstrap test statistics null distribution for model $\mathcal{M}$

For the non-parametric model $\mathcal{M}$, the bootstrap estimator of the *joint data generating distribution* P of the pair (X, Y) is the *joint empirical distribution*

P_n of the $n = n_1 + n_2$ pairs of (X, Y)-observations, $\{(X_i, Y_i) : i = 1, \ldots, n\}$. One resamples n pairs of (X, Y)-observations at random, with replacement from P_n, to form a bootstrap sample $\{(X_i^\#, Y_i^\#) : i = 1, \ldots, n\}$. The bootstrap test statistics null distribution Q_{0n} is the empirical distribution of the M-vectors of centered difference statistics, $Z_n^\# = \sqrt{n}((\psi_{2,n_2}^\# - \psi_{1,n_1}^\#) - (\psi_{2,n_2} - \psi_{1,n_1}))$, where $\psi_{k,n_k}^\#$ denotes the bootstrap empirical mean vector for Population k, that is, $\psi_{k,n_k}^\# = \sum_i \mathrm{I}\left(Y_i^\# = k\right) X_i^\# / \sum_i \mathrm{I}\left(Y_i^\# = k\right)$, $k = 1, 2$.

Note that an asymptotically equivalent estimator could be obtained by sampling n_1 observations at random, with replacement from the Population 1 sample, $\{(X_i, Y_i) : Y_i = 1, i = 1, \ldots, n\}$, and n_2 observations at random, with replacement from the Population 2 sample, $\{(X_i, Y_i) : Y_i = 2, i = 1, \ldots, n\}$.

Bootstrap test statistics null distribution for model $\mathcal{M}_=$

For the reduced location shift model $\mathcal{M}_=$, the bootstrap estimator of the *common mean-zero marginal data generating distribution* P_X of X is the *centered marginal empirical distribution* $P_{X,n}$ of the $n = n_1 + n_2$ centered X-observations, $\{X_{=,i} : i = 1, \ldots, n\}$, where $X_{=,i} = X_i - \mathrm{I}(Y_i = 1)\, \psi_{1,n_1} - \mathrm{I}(Y_i = 2)\, \psi_{2,n_2}$. One resamples n centered X-observations, $\{X_{=,i}^\# : i = 1, \ldots, n\}$, at random, with replacement from $P_{X,n}$, and sets $X_i^\# = X_{=,i}^\# + \psi_{k,n_k}$ and $Y_i^\# = k$ for a random subset of n_k such observations, $k = 1, 2$, to form a bootstrap sample $\{(X_i^\#, Y_i^\#) : i = 1, \ldots, n\}$. Again, the bootstrap test statistics null distribution Q_{0n} is the empirical distribution of the M-vectors of centered difference statistics, $Z_n^\# = \sqrt{n}((\psi_{2,n_2}^\# - \psi_{1,n_1}^\#) - (\psi_{2,n_2} - \psi_{1,n_1}))$.

Note that the above bootstrap procedure for model $\mathcal{M}_=$ is equivalent to the following approach: form the *mixture marginal empirical distribution* $P_{X,n}$ of the $n = n_1 + n_2$ (uncentered) X-observations, $\{X_i : i = 1, \ldots, n\}$; resample n X-observations, $\{X_i^\# : i = 1, \ldots, n\}$, at random, with replacement from $P_{X,n}$; set $Y_i^\# = k$ for a random subset of n_k such observations, $k = 1, 2$; and define Q_{0n} as the empirical distribution of the M-vectors of (uncentered) difference statistics, $D_n^\# = \sqrt{n}(\psi_{2,n_2}^\# - \psi_{1,n_1}^\#)$. This yields the non-parametric bootstrap (sampling *with* replacement) analogue of the commonly-used permutation (sampling *without* replacement) test, corresponding to the independence data generating distribution $P_\perp$ in Proposition 2.8.

Permutation test statistics null distribution

Permutation tests are known to be exact (up to the discreteness of the permutation distribution) under the location shift model $\mathcal{M}_=$ and the complete null hypothesis (Theorem 6, p. 231, Lehmann (1986); Puri and Sen (1971)). Indeed, if X and Y are independent, then the permutation distribution is equal to the conditional joint distribution of the pair (X, Y), given the marginal empirical distributions of X and Y. In contrast, bootstrap procedures corresponding to non-parametric model $\mathcal{M}$ are only approximate. In other words, model $\mathcal{M}_{=} \subseteq \mathcal{M}$ implies a stronger null model restriction than $\mathcal{M}$, as needed for an exact test.

As remarked in Section 2.9.2, above, when the permutation approach is appropriate, it tends to provide less variable estimators of the test statistics null distribution than the non-parametric bootstrap. Indeed, it should come as no surprise that, for small sample sizes, one typically obtains more accurate test results using a model-based (permutation or other suitable) estimator of the null distribution than a non-parametric estimator.

In general, estimation of the test statistics null distribution involves a bias-variance trade-off and raises the interesting open question of model selection.

The reader is referred to Pollard et al. (2005a) and Pollard and van der Laan (2004) for a more detailed discussion of the relative merits of bootstrap- and permutation-based multiple testing procedures.

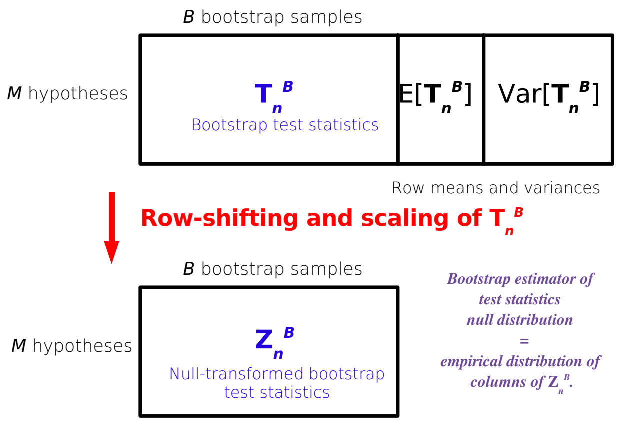

Figure 2.1. *Bootstrap estimation of the null shift and scale-transformed test statistics null distribution Q_0 (Procedure 2.3).* The bootstrap test statistics are stored in the $M \times B$ matrix $\mathbf{T}_n^B = \big(T_n^B(m,b)\big)$, with rows corresponding to the M null hypotheses and columns to the B bootstrap samples. Expected values, $\mathrm{E}[T_n(m)]$, and variances, $\mathrm{Var}[T_n(m)]$, of the test statistics are estimated by taking, respectively, row means and variances of $\mathbf{T}_n^B$. The matrix of test statistics $\mathbf{T}_n^B$ can then be row-shifted and scaled using the null values $\lambda_0(m)$ and $\tau_0(m)$, to produce an $M \times B$ matrix $\mathbf{Z}_n^B = \big(Z_n^B(m,b)\big)$. The null distribution Q_0 is estimated by the empirical distribution Q_{0n} of the columns of $\mathbf{Z}_n^B$.

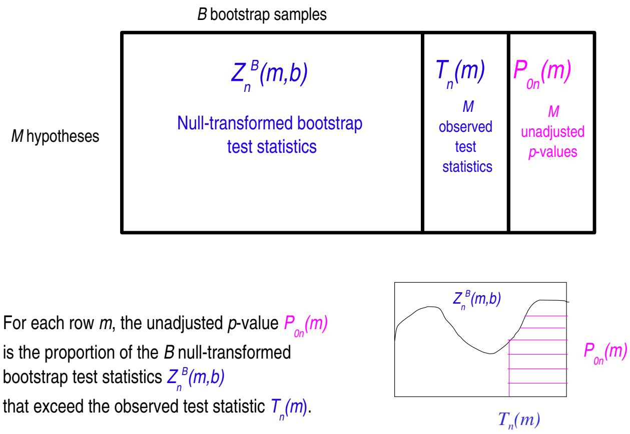

Figure 2.2. *Bootstrap estimation of the unadjusted p-values $P_{0n}(m)$.* Bootstrap estimators of the unadjusted p-values $P_{0n}(m)$ are obtained from the matrix $\mathbf{Z}_n^B = \left(Z_n^B(m,b)\right)$, of null-transformed bootstrap test statistics, by recording, for each row m, the proportion of $Z_n^B(m,b)$ that are greater than or equal to the observed test statistic $T_n(m)$.

http://www.springer.com/978-0-387-49316-9

Multiple Testing Procedures with Applications to Genomics
Dudoit, S.; van der Laan, M.J.
2008, XXXIII, 590 p., Hardcover
ISBN: 978-0-387-49316-9