# Notes to self

## Setting and notation

Suppose that $K$ random variables are measured on $N$ subjects, with the resulting matrix denoted $\mathbf{Z} \in \mathbb{R}_{N \times K}$. Let $Z_{nk}$ denote, for the $n^{th}$ subject, the $k^{th}$ random variable. Consider a bootstrapping algorithm (which is defined below) that generates, for iterate $j$, a dataset containing the random variables $\left( Z_{n1}^{(j)}, \cdots, Z_{nK}^{(j)} \right)$ for each subject $n$. There are a total of $B$ bootstrapped datasets.

We refer to probability density functions as $f$ and cumulative distribution functions as $F$. Specifically, we refer to the joint distribution of the data as follows. With $\mathcal{B} \in \mathbb{R}_{N \times K}$ denoting a matrix of constants with entries $(b_{11}, \cdots, b_{NK})$, let:

$$F_{\mathbf{Z}}(\mathbf{B}) = P\left( Z_{11} \leq b_{11}, \cdots, Z_{NK} \leq b_{NK} \right)$$

We will usually refer simply to $\mathbf{Z}$ and $F_{\mathbf{Z}}$ for brevity and will denote their counterparts in bootstrapped dataset $j$ as $\mathbf{Z}^{(j)}$ and $F_{\mathbf{Z}^{(j)}}$.

Further suppose that we conduct $W$ hypothesis tests, each with level $\alpha$. Let $c_{w,\alpha}$ be the critical value for the test statistic, $T_w$, of the $w^{th}$ test. We refer to the vector of test statistics as $\mathbf{T}$. Assume that each test concerns a point null hypothesis (**Assumption 1**), that each test statistic is a continuous function of the data with continuous first derivatives, $\mathbf{Z}$ (**Assumption 2**), and that each test is parametric with consistent estimates of all parameters in the likelihood identifiable from the observed data (**Assumption 3**). We define the "strong null" as the case in which all $W$ point null hypotheses hold and use the superscript $^0$ to denote variables generated under the strong null.

Define the statistic corresponding to the total number of rejections in $W$ tests in a sample generated under the strong null as $\widehat{\theta}^0 = \sum_{w=1}^{W} 1\{ T_w^0 > c_{w,\alpha} \}$. Similarly, the total number of rejections in bootstrap sample $j$ is $\widehat{\theta}^{(j)} = \sum_{w=1}^{W} 1\{ T_w^{(j)} > c_{w,\alpha} \}$.
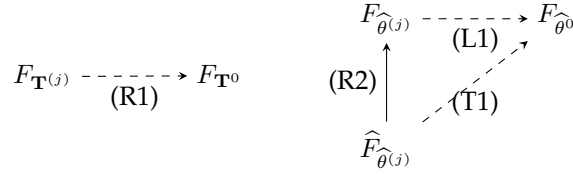
Respectively define the true distribution of the number of rejections under the strong null, its counterpart in the bootstrap samples, and its empirical estimator in the bootstrap samples as:

$$F_{\widehat{\theta}^0}(r) = P\left( \widehat{\theta}^0 \leq r \right)$$

$$F_{\widehat{\theta}^{(j)}}(r) = P\left( \widehat{\theta}^{(j)} \leq r \right)$$

$$\widehat{F}_{\widehat{\theta}^{(j)}}(r) = \frac{1}{B} \sum_{j^*=1}^{B} 1\left\{ \widehat{\theta}^{(j^*)} \leq r \right\}$$

## Results

We now show that under a parametric bootstrapping algorithm defined below, the empirical distribution of the number of rejections in the bootstrap samples converges in distribution to the true distribution of the number of rejections in sample generated under the strong null. Figure 1 depicts the structure of the proofs.

Figure 1: Overview of theory to follow. The dashed arrows denote convergence in probability in either $N$ or both $N$ and $B$. The solid arrow denotes almost sure convergence in $B$. Parenthetical acronyms refer to remarks, a lemma, and a theorem developed below.

$$F_{\widehat{\theta}^{(j)}} \dashrightarrow F_{\widehat{\theta}^0} \qquad (\text{L1})$$

$$F_{\mathbf{T}^{(j)}} \dashrightarrow F_{\mathbf{T}^0} \qquad (\text{R1}) \qquad (\text{R2}) \qquad (\text{T1})$$

$$\widehat{F}_{\widehat{\theta}^{(j)}}$$

**Algorithm 1** (Parametric bootstrap under the strong null). *We will refer throughout to bootstrap samples generated as follows:*

1. *Use the observed data to obtain consistent point estimates of any unknown parameters in the likelihood.*

2. *For any parameters in the likelihood that are constrained under the null hypothesis, set them equal to their null values (which are uniquely defined under Assumption 1 regarding point null hypotheses). For nuisance parameters, set them equal to their consistent estimates.*

3. *For any covariates treated as fixed in the likelihood, set them equal to their observed values.*

4. *Generate outcome data from the estimated likelihood under the strong null as defined above.*

(CITE ME - e.g., Fox, Romano.) Note that unlike some resampling schemes that can induce XXX problems, this algorithm guarantees that the bootstrapped data satisfy all parametric assumptions of the hypothesis tests to be conducted, preserving the nominal $\alpha$ of each test.

**Example 1** (Parametric bootstrap for multiple linear regression). *Suppose that one of the hypothesis tests is conducted using an OLS model regressing the $w^{th}$ variable on the first variable, adjusting for the second variable:*

$$Z_{nw} = \beta_0 + \beta_1 Z_{n1} + \beta_2 Z_{n2} + \epsilon_{nw}$$

*with $H_0 : \beta_1 = 0$ and $\epsilon \sim N(0, \sigma^2)$. Then we can apply Algorithm 1 as follows:*

1. *Fit the model to the observed data to estimate the three nuisance parameters as $\widehat{\beta}_0$, $\widehat{\beta}_2$, and $\widehat{\sigma}^2$.*

2. *For each bootstrap iterate, set the design matrix equal to that of the observed data.*

3. *Set $\beta_1 = 0$ (invoking the strong null), and set nuisance parameters $\beta_0 = \widehat{\beta}_0$, $\beta_2 = \widehat{\beta}_2$, and $\sigma^2 = \widehat{\sigma}^2$.*

4. *Thus, for each bootstrap iterate, generate the bootstrapped outcomes as $Z_{nw}^{(j)} = \widehat{\beta}_0 + \widehat{\beta}_2 Z_{n2} + \epsilon_{nw}^{(j)}$, where $\epsilon_{nw}^{(j)} \sim N(0, \widehat{\sigma}^2)$.*

**Example 2** (Parametric bootstrap for linear mixed model). *The approach in Definition 1 can accommodate parametric models with correlated observations, such as a linear mixed model with random intercepts by subject. One could obtain consistent estimates of the random intercepts using, for example, empirical Bayesian estimation, and these estimates can be plugged into the estimated likelihood under the strong null.*

**Remark 1.** *The distribution of the bootstrapped test statistics is consistent in $N$ for the true distribution of the test*

*statistics under the strong null. By repeatedly invoking the Continuous Mapping Theorem (CMT):*

$$f_{\mathbf{Z}^{(j)}} \xrightarrow[N\to\infty]{P} f_{\mathbf{Z}^0} \qquad\qquad \text{(CMT; Assumption 3)}$$

$$f_{\mathbf{T}^{(j)}} \xrightarrow[N\to\infty]{P} f_{\mathbf{T}^0} \qquad\qquad \text{(CMT; Assumption 2)}$$

$$F_{\mathbf{T}^{(j)}} \xrightarrow[N\to\infty]{P} F_{\mathbf{T}^0} \qquad\qquad \text{(CMT; integration preserves continuity)}$$

$$(1)$$

**Remark 2.** *By the i.i.d construction of the bootstrap samples and the Glivenko-Cantelli theorem, the empirical distribution of the number of rejections in the bootstrap samples converges almost surely in $B$ to its true distribution under the strong null:*

$$\widehat{F}_{\widehat{\theta}^{(j)}} \xrightarrow[B\to\infty]{A.S.} F_{\widehat{\theta}^{(j)}}$$

**Lemma 1.** *The number of rejections in the bootstrap samples converges in distribution in $N$ to the number of rejections in samples from the original population under the strong null:*

$$F_{\widehat{\theta}^{(j)}} \xrightarrow[N\to\infty]{P} F_{\widehat{\theta}^0}$$

*Proof.* Define the $r$-family of "rejection sets" as all possible configurations of the $W$ test statistics that lead to $r$ rejections:

$$\mathcal{A}_r = \left\{ (A_1, \cdots, A_W) \in \mathbb{R}^W : (T_1 \in A_1, \cdots, T_W \in A_W) \Rightarrow \widehat{\theta} = r \right\}$$

(For example, $\mathcal{A}_5$ contains all the 5-vectors of sets such that, if $(T_1 \in A_1, \cdots, T_5 \in A_5)$, then there are 5 total rejections across the $W$ tests. Suppose, for example, that $T_w \in (-\infty, \infty) \ \forall \ w$ and that each null hypothesis is rejected when $T_w > c_\alpha$. Then there are $\binom{W}{r}$ vectors in $\mathcal{A}_5$, representing the number of ways to select $r$ rejections from $W$ tests. One such vector corresponds to rejecting the first 5 tests and failing to reject all remaining tests:
$((c_\alpha, \infty), (c_\alpha, \infty), (c_\alpha, \infty), (c_\alpha, \infty), (c_\alpha, \infty), (-\infty, c_\alpha], \cdots, (-\infty, c_\alpha])$.)

Now consider the distribution of $\widehat{\theta}^{(j)}$:

$$
\begin{aligned}
P\left(\widehat{\theta}^{(j)} = r\right) &= P\left(\sum_{w=1}^{W} 1\left\{T_w^{(j)} > c_{w,\alpha}\right\} = r\right) \\
&= \sum_{(A_1,\cdots,A_W)\in\mathcal{A}} P\left(T_1^{(j)} \in A_1, \cdots, T_W^{(j)} \in A_W\right) \\
&\xrightarrow[N\to\infty]{P} \sum_{(A_1,\cdots,A_W)\in\mathcal{A}} P\left(T_1^0 \in A_1, \cdots, T_W^0 \in A_W\right) \qquad \text{(from Remark 1)} \\
&= P\left(\widehat{\theta}^0 = r\right)
\end{aligned}
$$

$\square$

**Theorem 1.** *The empirical distribution of the number of rejections in the bootstrap samples is consistent in $N$ and $B$ for the true distribution of the number of rejections under the strong null:*

$$\widehat{F}_{\widehat{\theta}^{(j)}} \xrightarrow[B,N\to\infty]{P} F_{\widehat{\theta}^0}$$

*Proof.* Fix an arbitrarily small $\epsilon > 0$ and $\epsilon' > 0$. Remark 2 implies consistency, so there exists a $B^*$ such that, for all $B > B^*$:

$$P\left(|\widehat{F}_{\widehat{\theta}(j)} - F_{\widehat{\theta}(j)}| > \epsilon/2\right) < \epsilon'/2$$

Similarly, from Lemma 1, there exists an $N^*$ such that, for all $N > N^*$:

$$P\left(|F_{\widehat{\theta}(j)} - F_{\widehat{\theta}^0}| > \epsilon/2\right) < \epsilon'/2$$

Thus, for all $B$ and $N$ such that $B > B^*$ and $N > N^*$:

$$\begin{aligned}
P\left(|\widehat{F}_{\widehat{\theta}(j)} - F_{\widehat{\theta}^0}| > \epsilon\right) &= P\left(|\widehat{F}_{\widehat{\theta}(j)} - F_{\widehat{\theta}(j)} + F_{\widehat{\theta}(j)} - F_{\widehat{\theta}^0}| > \epsilon\right) \\
&\leq P\left(|\widehat{F}_{\widehat{\theta}(j)} - F_{\widehat{\theta}(j)}| + |F_{\widehat{\theta}(j)} - F_{\widehat{\theta}^0}| > \epsilon\right) \qquad \text{(triangle inequality)} \\
&\leq P\left(|\widehat{F}_{\widehat{\theta}(j)} - F_{\widehat{\theta}(j)}| > \epsilon/2 \ \text{ or } \ |F_{\widehat{\theta}(j)} - F_{\widehat{\theta}^0}| > \epsilon/2\right) \\
&\leq P\left(|\widehat{F}_{\widehat{\theta}(j)} - F_{\widehat{\theta}(j)}| > \epsilon/2\right) + P\left(|F_{\widehat{\theta}(j)} - F_{\widehat{\theta}^0}| > \epsilon/2\right) \\
&< \epsilon'
\end{aligned}$$

$\square$

# Simulation study

See simulations from 2017-8-1 (written up in "multiple_outcomes (4).pdf"). Maybe use these these scenarios again? Generate at least some data under null to show that power of joint test is <5%, as intended.

## Width of 95% CI generated under null

- Figure 1: Mean number of rejections and average of CI limits to show width

- Figure 2: Average distance of observed rejections from upper CI limit

## Power of joint test

- Figure 3 with panels: Prob of rejecting joint null in samples generated under null (should be <5% for all) and under various alternative scenarios

# Applied example