Reading Notes for Large-Scale Inference
Empirical Bayes Methods for Estimation, Testing, and Prediction

Xiang Li
Peking University
lx10077@pku.edu.cn

November 19, 2020

Some High-level Points

1. Empirical Bayes arguments combine frequentist and Bayesian elements in analyzing problems of repeated structure.
2. False discovery rates, Benjamini and Hochberg's seminal contribution, is the great success story of the new methodology and has strong credentials in both the Bayesian and frequentist camps.
3. It is hard to go wrong using maximum likelihood estimation or a t-test on a typical small data set. But it is very easy to go wrong with huge data sets and thousands of questions to answer at once.
4. Typical large-scale applications have been more concerned with testing than estimation. Empirical Bayes blurs the line between testing and estimation as well as between frequentism and Bayesianism.

1 Empirical Bayes and the James–Stein Estimator

The maximum likelihood estimation methods for Gaussian models proposed by Charles Stein are still in use, for good reasons, but Stein-type estimators have pointed the way toward a radically different empirical Bayes approach to high-dimensional statistical inference.

Three Estimators for Gaussian Mean Estimation. Given many independent Gaussian r.v.s,

$$\mu_i \sim \mathcal{N}(0, A) \qquad z_i \mid \mu_i \sim \mathcal{N}(\mu_i, 1) \quad [i = 1, 2, \ldots, N]$$

where (μ_i, z_i) pairs being independent of each other. Using concatenation, it is denoted by $\boldsymbol{\mu} \sim \mathcal{N}_N(\mathbf{0}, AI)$ and $\boldsymbol{z} \mid \boldsymbol{\mu} \sim \mathcal{N}_N(\boldsymbol{\mu}, I)$. where I is the $N \times N$ identity matrix Then Bayes rule gives posterior distribution give

$$\boldsymbol{\mu} \mid \boldsymbol{z} \sim \mathcal{N}_N(B\boldsymbol{z}, BI) \quad [B = A/(A+1)]. \tag{1}$$

We use total squared error loss to measure the error of estimation $L(\boldsymbol{\mu}, \hat{\mu}) = \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 = \sum_{i=1}^{N} (\hat{\mu}_i - \mu_i)^2$ with the corresponding risk function being $R(\boldsymbol{\mu}) = E_{\boldsymbol{\mu}}\{L(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})\} = E_{\boldsymbol{\mu}}\left\{\|t(\boldsymbol{z}) - \boldsymbol{\mu}\|^2\right\}$.

Current statistical practice is quite conservative in protecting individual inferences from the tyranny of the majority, accounting for the continued popularity of standalone methods like MLE.

Empirical Bayes Confidence Intervals. If A were known we could calculate the Bayes posterior distribution for μ_0 according to eqn. (1)

$$\mu_0 \mid z_0 \sim \mathcal{N}(Bz_0, B) \quad [B = A/(A+1)]$$

yielding

$$\mu_0 \in Bz_0 \pm 1.96\sqrt{B}$$

as the obvious 95% posterior interval. A reasonable first try in the empirical Bayes situation is to substitute the unbiased estimate $\hat{B} = 1 - \frac{N-2}{S} \quad \left[S = \|z\|^2\right]$, giving the approximation $\mu_0 \mid z_0, \boldsymbol{z} \dot{\sim} \mathcal{N}\left(\hat{B}z_0, \hat{B}\right)$ and similarly $\hat{B}z_0 \pm 1.96\sqrt{\hat{B}}$.

2 Large-Scale Hypothesis Testing

A new class of "high throughput" biomedical devices, typified by the microarray, routinely produce hypothesis-testing data for thousands of cases at once. A traditional approach to multiple inference uses the Bonferroni bound, reducing the rejection level by a factor of test numbers. However, such classic method will give a too narrow rejection region, while Empirical Bayes methods offer a less conservative approach to multiple testing.

Bayes False Discovery Rate. We suppose that the N cases (the genes for the prostate study) are each either null or non-null with prior probability with z-values having density either $f_0(z)$ or $f_1(z)$, $\pi_0 = \Pr\{\text{ null }\}$ and $\pi_1 = \Pr\{\text{ non-null }\}$. Let F_0 and F_1 denote the probability distributions corresponding to f_0 and f_1 so that, for any subset $\mathcal{Z}$ of the real line,

$$F_0(\mathcal{Z}) = \int_{\mathcal{Z}} f_0(z)dz \quad \text{and} \quad F_1(\mathcal{Z}) = \int_{\mathcal{Z}} f_1(z)dz.$$

The mixture density $f(z) = \pi_0 f_0(z) + \pi_1 f_1(z)$ has the mixture probability distribution $F(\mathcal{Z}) = \pi_0 F_0(\mathcal{Z}) + \pi_1 F_1(\mathcal{Z})$. Suppose we observe $z \in \mathcal{Z}$ and wonder if it corresponds to the null or non-null arm. A direct application of Bayes rule yields

$$\phi(\mathcal{Z}) \equiv \Pr\{\text{null} \mid z \in \mathcal{Z}\} = \pi_0 F_0(\mathcal{Z})/F(\mathcal{Z}) \tag{2}$$

which is the Bayes false discovery rate for $\mathcal{Z}$. Local Bayes false discovery rate is the quantity for a single set $\mathcal{Z}$: $\phi(z_0) \equiv \Pr\{\text{null} \mid z = z_0\} = \pi_0 f_0(z_0)/f(z_0)$. We have $E_{z\sim f(z)}\{\phi(z) \mid z \in \mathcal{Z}\} = \phi(\mathcal{Z})$.

Empirical Bayes Estimates. The empirical distribution of the N z-values is $\bar{F}(\mathcal{Z}) = \#\{z_i \in \mathcal{Z}\}/N$. Substituting into eqn. (2) gives an estimated false discovery rate

$$\overline{\mathrm{Fdr}}(\mathcal{Z}) \equiv \bar{\phi}(\mathcal{Z}) = \pi_0 F_0(\mathcal{Z})/\bar{F}(\mathcal{Z}).$$

Let $N_0(\mathcal{Z})$ be the number of null z_i falling into $\mathcal{Z}$, and likewise $N_1(\mathcal{Z})$ for the non-null z_i in $\mathcal{Z}$. We can only observe the total number $N_+(\mathcal{Z}) = N_0(\mathcal{Z}) + N_1(\mathcal{Z})$. Although $N_0(\mathcal{Z})$ is unobservable, we know its expectation $e_0(\mathcal{Z}) \equiv N\pi_0 F_0(\mathcal{Z})$. So we can express $\overline{\mathrm{Fdr}}(\mathcal{Z})$ as

$$\overline{\mathrm{Fdr}}(\mathcal{Z}) = e_0(\mathcal{Z})/N_+(\mathcal{Z}).$$

$\overline{\mathrm{Fdr}}(\mathcal{Z})$ as a Point Estimate. Consider the following three quantities

$$\overline{\mathrm{Fdr}}(\mathcal{Z}) = \frac{e_0(\mathcal{Z})}{N_+(\mathcal{Z})}, \quad \phi(\mathcal{Z}) = \frac{e_0(\mathcal{Z})}{e_+(\mathcal{Z})}, \quad \text{and} \quad \mathrm{Fdp}(\mathcal{Z}) = \frac{N_0(\mathcal{Z})}{N_+(\mathcal{Z})},$$

we have

$$E\left\{\overline{\mathrm{Fdr}}(\mathcal{Z}) \mid N_1(\mathcal{Z})\right\} \geq \phi_1(\mathcal{Z}) \geq E\left\{\mathrm{Fdp}(\mathcal{Z}) \mid N_1(\mathcal{Z})\right\}$$

where

$$\phi_1(\mathcal{Z}) = \frac{e_0(\mathcal{Z})}{e_0(\mathcal{Z}) + N_1(\mathcal{Z})}.$$

We have the following observations:

1. $E\{\overline{\mathrm{Fdr}}(\mathcal{Z})\} \geq E\{\phi_1(\mathcal{Z})\} \geq E\{\mathrm{Fdp}(\mathcal{Z})\}$: empirical Bayes false discovery rate is a conservatively biased estimate of the actual false discovery proportion (Fdp).
2. $E\{\min(\overline{\mathrm{Fdr}}(\mathcal{Z}), 2)\} \geq \phi(\mathcal{Z})$.
3. Let $\gamma(\mathcal{Z}) = \mathrm{var}\left\{N_+(\mathcal{Z})\right\}/e_+(\mathcal{Z})^2$, standard delta-method calculations yield useful approximations for the mean and variance of $\overline{\mathrm{Fdr}}(\mathcal{Z})$:
$$\frac{\overline{\mathrm{Fdr}}(\mathcal{Z})}{\phi(\mathcal{Z})} \dot{\sim} (1 + \gamma(\mathcal{Z}), \gamma(\mathcal{Z})).$$
4. If $N \sim \mathrm{Poi}(\eta)$ and each z_i is generated independently (Poisson-independence assumptions),
$$E\{\mathrm{Fdp}(\mathcal{Z})\} = \phi(\mathcal{Z}) \cdot [1 - \exp\left(-e_+(\mathcal{Z})\right)]$$
where $e_+(\mathcal{Z}) = E\left\{N_+(\mathcal{Z})\right\} = \eta \cdot F(\mathcal{Z})$.
5. Define $\widetilde{\mathrm{Fdr}}(\mathcal{Z}) = e_0(\mathcal{Z})/\left(N_+(\mathcal{Z}) + 1\right)$, under Poisson-independence assumptions, we have
$$E\{\widetilde{\mathrm{Fdr}}(\mathcal{Z})\} = E\{\mathrm{Fdp}(\mathcal{Z})\} = \phi(\mathcal{Z}) \cdot \left[1 - \exp\left(-e_+(\mathcal{Z})\right)\right].$$

3 Significance Testing Algorithms

p-values. For any value of $\alpha \in [0, 1]$ we construct a rejection region $\mathcal{R}_\alpha$ in the sample space of $\boldsymbol{x}$,

$$\Pr_0\left\{\boldsymbol{x} \in \mathcal{R}_\alpha\right\} = \alpha \quad [\alpha \in (0, 1)].$$

The p -value $p(\boldsymbol{x})$ corresponding to $\boldsymbol{x}$ is defined as the smallest value of α such that $\boldsymbol{x} \in \mathcal{R}_\alpha$,

$$p(\boldsymbol{x}) = \inf_\alpha \left\{\boldsymbol{x} \in \mathcal{R}_\alpha\right\}. \tag{3}$$

For any value of u in $(0, 1)$, the event $p(\boldsymbol{x}) \leq u$ is equivalent to $\boldsymbol{x} \in \mathcal{R}_u$ implying

$$\Pr_0\{p(\boldsymbol{x}) \leq u\} = \Pr_0\left\{\boldsymbol{x} \in \mathcal{R}_u\right\} = u. \tag{4}$$

z-values is defined to be $z(\boldsymbol{x}) = \Phi^{-1}(p(\boldsymbol{x}))$. Under H_0, $P = p(\boldsymbol{x}) \sim \mathcal{U}(0,1)$ is equivalent to $H_0 : z(\boldsymbol{x}) \sim \mathcal{N}(0,1)$.

The Family-Wise Error Rate. FWER is defined as the probability of making at least one false rejection in a family of hypothesis-testing problems. Let $p_1, p_2, \ldots, p_N$ be the p-values obtained in tests of the corresponding family of null hypotheses $H_{01}, H_{02}, \ldots, H_{0N}$, then

$$\text{FWER} = \Pr\{ \text{ Reject any xf } H_{0i}\}.$$

A FWER control procedure is an algorithm that inputs a family of p -values $(p_1, p_2, \ldots, p_N)$ and outputs the list of accepted and rejected null hypotheses, subject to the constraint FWER $\leq \alpha$ for any preselected value of α. Hence, with probability greater than $1 - \alpha$, no true null hypothesis is rejected.

Let FWER $_\alpha(\boldsymbol{x})$ be a FWER level-α test procedure based on $\boldsymbol{x}$. The general definition of adjusted p -value for case i is an analog of eqn. (3)

$$\tilde{p}_i(\boldsymbol{x}) = \inf_\alpha \{H_{0i} \text{ rejected by } \text{FWER}_\alpha(\boldsymbol{x})\}.$$

Holm's Sequentially Rejective Procedure [2]. A *step-down* procedure begins with the ordered p-values. Let the ordered p -values be denoted by

$$p_{(1)} \leq p_{(2)} \leq p_{(3)} \leq \cdots \leq p_{(N)} \tag{5}$$

and reject $H_{0(i)}$, the hypothesis corresponding to $p_{(i)}$, if

$$p_{(j)} \leq \frac{\alpha}{N - j + 1} \quad \text{for } j = 1, 2, \ldots, i. \tag{6}$$

The adjusted p-value for Holm's procedure is

$$\tilde{p}_{(i)} = \max_{j \leq i} \left\{(N - j + 1)p_{(j)}\right\}_1 \tag{7}$$

where $\{x\}_1 \equiv \min(x, 1)$. Holm's method satisfies the FWER control property, which is because

$$\begin{aligned} 1 - \alpha &\overset{\text{Bonferroni}}{\leq} \Pr\left\{p_{(i)} > \frac{\alpha}{N_0} \text{ for all } i \in I_0\right\} = \Pr\left\{p_{(i_0)} > \frac{\alpha}{N_0} = \frac{\alpha}{N + 1 - i_0}\right\} \\ &\leq \Pr\left\{\hat{i} < i_0\right\} \leq \Pr\left\{p_{(\hat{i})} < \frac{\alpha}{N_0}\right\} = \Pr\{\text{none of true null hypotheses rejected}\} \end{aligned}$$

where $I_0 =$ indexes of true null hypotheses, $N_0 = \#I_0$, $i_0 = N - N_0 + 1$ and $\hat{i}$ is the maximum index satisfying eqn. (6). Holm's procedure controls FWER in the *strong sense*, that is, FWER is bounded by α no matter what the pattern of true and false null hypotheses might be.

Simes' Improved Bonferroni Procedure. If we are willing to assume independence among the original p -values $p_1, p_2, \ldots, p_N$, then a better bound, known as Simes' inequality [5, 4], is available: when all the null hypotheses are true, then

$$A_N(\alpha) := \Pr\left\{p_{(i)} \geq \frac{\alpha i}{N} \text{ for } i = 1, 2, \ldots, N\right\} \geq 1 - \alpha$$

with equality if the test statistics are continuous.

Proof. In the independent continuous case, with all $H_{0(i)}$ true, $p_{(1)}, p_{(2)}, \ldots, p_{(N)}$ are the order statistics of N independent $\mathcal{U}(0,1)$ variates, as in eqn. (4). A standard order statistic result then shows that given $p_{(N)}$, the largest p -value, the ratios $(p_{(1)}/p_{(N)}, p_{(2)}/p_{(N)}, \ldots, p_{(N-1)}/p_{(N)})$ are distributed as the order statistics from $(N-1)$ independent $\mathcal{U}(0,1)$ variates, while $p_{(N)}$ has cdf p^N.

Recall the definition of $A_N(\alpha)$, we have

$$A_N(\alpha) = \int_\alpha^1 A_{N-1}\left\{\frac{\alpha(N-1)}{pN}\right\} Np^{N-1}dp$$

If $A_{N-1}(\alpha) = 1-\alpha$ then $A_N(\alpha) = 1-\alpha$ follows. Hence the result is proved by induction. □

Simes discussed another "improved Bonferroni procedure", that is to reject $H_{(i)}$ if $p_{(j)} \leq j\alpha/n$ for some $j \geq i$. This procedure control FWER in the *weak sense* that refers to methods that control FWER only if all the null hypotheses are true (that is $H_0 : H_{01}, \cdots, H_{0N}$ are all true). Simes proved that this procedure has level α under H_0 when the p-values are independent.

Hochberg's Sequentially Acceptive Procedure [1]. Starting from Simes' inequality, Hochberg used the closure principle to improve (i.e., raise) Holm's adjusted p-values eqn. (7) to

$$\tilde{p}_{(i)} = \min_{j \geq i} \left\{(N-j+1)p_{(j)}\right\}_1 \tag{8}$$

with $\{x\}_1 \equiv \min(x, 1)$. This is not a general improvement though, since Simes' inequality depends on independence of the test statistics. Algorithms such as eqn. (8), whose definitions depend on the upper tail of the sequence $p_{(1)} \leq p_{(2)} \leq p_{(3)} \leq \cdots \leq p_{(N)}$ are called "*step-up* procedures."

Remark 3.1.

1. *Holm's scans forward, and stops as soon as a p-value fails to clear its threshold. Holm's step-down procedure eqn. (6) starts with $i = 1$ and keeps rejecting $H_{0(i)}$ until the first time $p_{(i)} > \alpha/(N-i+1)$. This pessimistic approach is called a step-down procedure.*

2. *Hochberg's scans backwards, and stops as soon as a p-value succeeds in passing its threshold. Hochberg's step-back procedure starts with $i = N$ and keeps accepting $H_{0(i)}$ until the first time $p_{(i)} \leq \alpha/(N-i+1)$.*

3. *This shows that Hochberg's procedure is more powerful than Holm's, i.e., rejects more often at the same α level. In general, step-up procedures can be substantially more powerful than step-down procedures.*

The Closure Principle [3]. Let $I \subset \{1, 2, \ldots, N\}$, and $\mathcal{I}$ the statement that all of the null hypotheses in I are true, $\mathcal{I} = \bigcap_I H_{0(i)}$. If I' is a larger subset, $I' \supseteq I$, then logically $\mathcal{I}' \Rightarrow \mathcal{I}$. Suppose that for every subset I we have a level-α non-randomized test function $\phi_I(\boldsymbol{x}) : \phi_I(\boldsymbol{x})$ equals 1 or 0, with 1 indicating rejection of $\mathcal{I}$, satisfying $\Pr_{\mathcal{I}}\{\phi_I(\boldsymbol{x}) = 1\} \leq \alpha$. Consider the simultaneous test function

$$\Phi_I(\boldsymbol{x}) = \min_{I' \supseteq I} \{\phi_{I'}(\boldsymbol{x})\}$$

$\Phi_I(\boldsymbol{x})$ defines a rule that rejects $\mathcal{I}$ if and only if $\mathcal{I}'$ is rejected at level α for every I' containing I. But if $\mathcal{I}$ is true then $I \subseteq I_0$, the set of all true $H_{0(i)}$ and

$$\Pr_{I_0}\{\phi_{I_0}(\boldsymbol{x}) = 1\} \leq \alpha \quad \text{implies} \quad \Pr_I\{\Phi_I(\boldsymbol{x}) = 1\} \leq \alpha$$

So the test Φ_I simultaneously controls the probability of rejecting *any* true subset I at level α.

The closure principle can be used to

1. extend Bonferroni's bound to Holm's procedure by setting

$$\varphi_I(\boldsymbol{x}) = 1 \quad \text{iff} \quad \inf\{p_i(\boldsymbol{x}) : i \in I\} \leq \frac{\alpha}{|I|}$$

at level α . Let $I_i = \{i, i+1, \ldots, N\}$. In terms of the ordered p-values eqn. (5), Bonferroni's rule rejects $\mathcal{I}_i$ if $p_{(i)} \leq \alpha/(N+1-i)$. Note that $I_j \supseteq I_i$ for $j \leq i$.

2. extend Sime's bound to Hochberg's procedure by setting

$$\varphi_I(\boldsymbol{x}) = 1\left(\{p_{(1,I)}(\boldsymbol{x}) \leq \alpha/|I|\}\right) \text{ or } 1\left(\{p_{(2,I)}(\boldsymbol{x}) \leq 2\alpha/|I|\}\right) \text{ or } \ldots \text{ or } 1\left(\{p_{(|I|,I)}(\boldsymbol{x}) \leq \alpha\}\right)$$

where $p_{(j,I)}(\boldsymbol{x})$ denotes the j-th p-value with index in I. Again let $I_i = \{i, i+1, \ldots, N\}$. In terms of the ordered p-values eqn. (5), Sime's rule rejects $\mathcal{I}_i$ if there exist some $j \geq i$ such that $p_{(j)} \leq \alpha/(N+1-i)$. A sufficient condition is there exist some $j \geq i$ such that $p_{(j)} \leq \alpha/(N+1-j)$, which if used is strictly more conservative than the closure of Simes.

The closure of Simes' procedure will be strictly more powerful than Holm's procedure. However, just like Simes, the closure of Simes requires the p-values to be independent.

Permutation Algorithms. Starting with the ordered p-values $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(N)}$, let $r_1, r_2, \ldots, r_N$ indicate the corresponding original indices, $p_{(j)} = p_{r_j}, \quad j = 1, 2, \ldots, N$. Define $R_j = \{r_j, r_{j+1}, \ldots, r_N\}$ and

$$\pi(j) = \Pr_0\left\{\min_{k \in R_j}(P_k) \leq p_{(j)}\right\} \tag{9}$$

Here $(P_1, P_2, \ldots, P_N)$ indicates a hypothetical realization of the unordered p -values $(p_1, p_2, \ldots, p_N)$ obtained under the complete null hypothesis H_0 that all of the H_{0i} are true; eqn. (9) is computed with $p_{(j)}$ fixed at its observed value. The Westfall-Young step-down min-p adjusted p -values are then defined by

$$\tilde{p}_{(i)} = \max_{j \leq i}\{\pi(j)\}$$

The min-p procedure can be difficult to implement. Westfall and Young also proposed a simpler variant, called "max-T" by ordering the values of the original test statistics that gave the observed p-values rather than hard-to-computed p-values.

Other Control Criteria.

1. Per comparison error rate: PCER = $E\{$ Number true null hypotheses rejected $\}/N$;
2. Expected error rate: EER = $E\{$ Number wrong decisions $\}/N$. A wrong decision being rejection of H_{0i} when it should be accepted or vice versa.
3. k-FWER: aims to control the probability of rejecting k or more true null hypotheses; $k = 1$ is the usual FWER, but choosing a larger value of k gives more generous results.

4 False Discovery Rate Control

False Discovery. We wish to test N null hypotheses: $H_{01}, H_{02}, \cdots, H_{0N}$. We will give some decision rule $\mathcal{D}$ that will produce a decision of "null" or "non-null" for each of the N cases. N_0 of the N cases were actually null, of which $\mathcal{D}$ called a non-null (incorrectly) and $N_0 - a$ null (correctly); likewise, N_1 were actually non-null, with $\mathcal{D}$ deciding b of them correctly and $N_1 - b$ incorrectly. The false discovery rate is $\frac{a}{a+b}$ and FWER equals $\Pr\{a > 0\}$.

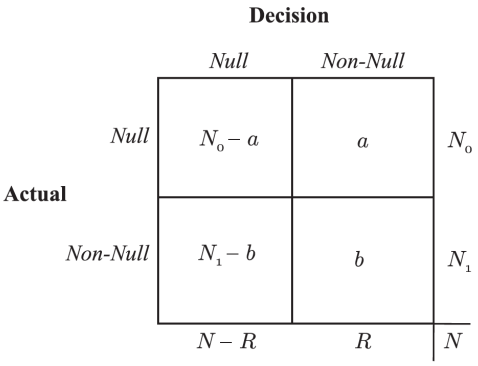

Figure 1: A decision rule $\mathcal{D}$ has rejected R out of N null hypotheses.

Classical Fisherian significance testing calculated *horizontally* in the figure, that is restricting attention to either the null or non-null row, which is to say that they are frequentist calculations. Large-scale testing, with N perhaps in the hundreds or thousands, opens the possibility of calculating *vertically* in the figure, in the Bayesian direction, without requiring Bayesian priors.

Benjamini and Hochberg's FDR Control Algorithm. We assume that our decision rule $\mathcal{D}$ produces a p-value p_i for each case i so that p_i has a uniform distribution if H_{0i} is correct, $H_{0i} : p_i \sim \mathcal{U}(0,1)$. Denote the ordered p -values by $p_{(1)} \le p_{(2)} \le \cdots \le p_{(i)} \le \cdots \le p_{(N)}$. Let $R_{\mathcal{D}}$ be the number of cases rejected, $a_{\mathcal{D}}$ the number of those that are actually null, and $\mathrm{Fdp}_{\mathcal{D}}$ the false discovery proportion

$$\mathrm{Fdp}_{\mathcal{D}} = a_{\mathcal{D}}/R_{\mathcal{D}} \quad [= 0 \text{ if } R_{\mathcal{D}} = 0]$$

The Benjamini-Hochberg (BH) algorithm uses this rule: for a fixed value of q in $(0,1)$, let $i_{\max}$ be the largest index for which

$$p_{(i)} \le \frac{i}{N} q \tag{10}$$

and reject $H_{0(i)}$, the null hypothesis corresponding to $p_{(i)}$, if $i \le i_{\max}$ accepting $H_{0(i)}$ otherwise.

1. **FDR Control.** If the p-values corresponding to the correct null hypotheses are independent of each other, then the rule $\mathrm{BH}(q)$ based on the BH algorithm controls the expected false discovery proportion at q:

$$E\left\{\mathrm{Fdp}_{\mathrm{BH}(q)}\right\} = \pi_0 q \le q \quad \text{where } \pi_0 = N_0/N.$$

2. **From empirical Bayes perspective.** Let $z_{(i)}$ denote the i th ordered value, $z_{(1)} \le z_{(2)} \le \cdots \le z_{(i)} \le \cdots \le z_{(N)}$. Then $p_{(i)} = F_0(z_{(i)})$. Note that the empirical cdf of the z_i values, $\bar{F}(z) = \#\{z_i \le z\}/N$ satisfies $\bar{F}\left(z_{(i)}\right) = i/N$. This means we can write the threshold condition for the BH rule eqn. (10) as

$$F_0\left(z_{(i)}\right)/\bar{F}\left(z_{(i)}\right) \le q, \text{ or } \pi_0 F_0\left(z_{(i)}\right)/\bar{F}\left(z_{(i)}\right) \le \pi_0 q$$

However, $\pi_0 F_0(z)/\bar{F}(z)$ is the empirical Bayes false discovery rate estimate $\overline{\mathrm{Fdr}}(z)$ from eqn. (2) with $\mathcal{Z} = (-\infty, z)$. We can now re-express the Control part. Let $i_{\max}$ be the largest index for which

$$\overline{\mathrm{Fdr}}\left(z_{(i)}\right) \le q$$

and reject $H_{0(i)}$ for all $i \le i_{\max}$, accepting $H_{0(i)}$ otherwise. Then, assuming that the z_i values are independent, the expected false discovery proportion of the rule equals q.

3. **FDR Control is more like an estimate rather than a test statistic.**

   We might follow the rule of rejecting all the null hypotheses H_{0i} corresponding to z_i in $\mathcal{R}$ if $\overline{\text{Fdr}}(\mathcal{R}) \leq q$, and accepting all of them otherwise. Equivalently, we reject all the H_{0i} for z_i in $\mathcal{R}$ if $R \geq e_0(\mathcal{R})/q$.

   It is clear that such rule cannot be a test of the FWER-type null hypothesis that at least one of the R hypotheses is true, $H_0(\text{ union }) = \bigcup_{i:z_i \in \mathcal{R}} H_{0i}$. Rejecting $H_0(\text{ union })$ implies we believe all the H_{0i} for z_i in $\mathcal{R}$ to be incorrect (that is, all should be rejected).

   On the other hand, it is too conservative as a test of $H_0(\text{intersection}) = \bigcap_{i:z_i \in \mathcal{R}} H_{0i}$, which is the hypothesis that all of the R null hypotheses are correct. Rejecting $H_0(\text{intersection})$ says we believe at least one of the R cases is non-null.

4. FDR helps simultaneous tests of correlation, that is, given a case of interest which of the other $N-1$ cases is unusually highly correlated with it? "Unusual" has the meaning here of being in the rejection set of a simultaneous testing procedure.

5 Local False Discovery Rates

Wwe begin with the Bayesian two-groups model, in which each of the N cases is either null or non-null, with prior probability π_0 or π_1

$$\begin{array}{ll} \pi_0 = \Pr\{\text{ null }\} & f_0(z) = \text{ null density} \\ \pi_1 = \Pr\{\text{ non-null }\} & f_1(z) = \text{ non-null density} \end{array}$$

Definition and Estimation. The local false discovery rate is

$$\text{fdr}(z) = \Pr\{\text{null} \mid z\} = \pi_0 f_0(z)/f(z) \tag{11}$$

where $f(z)$ is the mixture density y: $f(z) = \pi_0 f_0(z) + \pi_1 f_1(z)$. $f_0(z)$ is assumed known in this section while we should estimate π_0 and $f(z)$. We can then estimate $\text{fdr}(z)$ by

$$\widehat{\text{fdr}}(z) = \hat{\pi}_0 f_0(z)/\hat{f}(z). \tag{12}$$

A conventional threshold for reporting "interesting" cases is $\widehat{\text{fdr}}(z_i) \leq 0.20$.

Estimation for π_0. Suppose we believe that $f_1(z)$ is zero for a certain subset $\mathcal{A}_0$ of the sample space, perhaps those points near zero, $f_1(z) = 0$, for $z \in \mathcal{A}_0$ that is, all the non-null cases must give z -values outside of $\mathcal{A}_0$ (sometimes called the zero assumption). Then the expected value of $N_+(\mathcal{A}_0)$, the observed number of z_i values in $\mathcal{A}_0$, is $E\{N_+(\mathcal{A}_0)\} = \pi_0 N \cdot F_0(\mathcal{A}_0)$ suggesting the estimators

$$\hat{\pi}_0 = N_+(\mathcal{A}_0) / (N \cdot F_0(\mathcal{A}_0)) \tag{13}$$

and

$$\widehat{\text{Fdr}}(z) = \hat{\pi}_0 F_0(z)/\bar{F}(z)$$

Using $\widehat{\text{Fdr}}(z)$ in place of $\overline{\text{Fdr}}(z) = F_0(z)/\bar{F}(z)$ increases the number of discoveries (i.e., rejections).

We might take $\mathcal{A}_0$ to be the central α_0 proportion of the f_0 distribution on the grounds that all the "interesting" non-null cases should produce z values far from the central region of f_0. If f_0 is $\mathcal{N}(0,1)$ in then $\mathcal{A}_0$ is the interval $\mathcal{A}_0 = \left[\Phi^{-1}(0.5 - \alpha_0/2), \Phi^{-1}(0.5 + \alpha_0/2)\right]$.

Estimation for $f(z)$. suppose f belongs to the J-parameter family:

$$f(z) = \exp\left\{\sum_{j=0}^{J} \beta_j z^j\right\} \tag{14}$$

The difficulty is that we don't assume $z_1, z_2, \cdots$ are independent, so MLE could directly apply. Lindsey's method, an algorithm based on discretizing the z_i values, obtains maximum likelihood estimates $\hat{\boldsymbol{\beta}}$ using standard Poisson regression techniques. In particular,

1. We partition the range $\mathcal{Z}$ of the z_i values into K bins of equal width d, $\mathcal{Z} = \bigcup_{k=1}^{K} \mathcal{Z}_k$
2. Define y_k as the count in the k th bin, $y_k = \#\{z_i \in \mathcal{Z}_k\}$ and let $x_k =$ centerpoint of $\mathcal{Z}_k$. The expected value of y_k is approximately $v_k = Ndf(x_k)$.
3. Lindsey's method assumes that the y_k are independent Poisson counts $y_k \overset{\text{ind}}{\sim} \text{Poi}(\nu_k) \quad k = 1, 2, \ldots, K$ and then fits $\hat{f}$ via a regression model for v_k as a function of x_k. In model eqn. (14), $\log(v_k) = \sum_{j=0}^{J} \beta_j x_k^j$. The above is a standard Poisson generalized linear model (GLM).

Actually, we are fitting a smooth function $f_{\boldsymbol{\beta}}(x_k)$ to the counts y_k. Technically, this is done by minimizing the sum of Poisson deviances,

$$\sum_{k=1}^{K} D\left(y_k, Ndf_{\boldsymbol{\beta}}(x_k)\right), \quad D(y, v) = 2y\left[\log\left(\frac{y}{v}\right) - \left(1 - \frac{y}{v}\right)\right]$$

over the choice of $\boldsymbol{\beta}$. Independence of the z_i or of the y_k is not required.

About Fdr and fdr. Fdr and fdr methods scale in a nice way as the number N of cases changes. Letting N go to infinity shouldn't greatly change $\text{fdr}(z)$, as it converges to $\text{fdr}(z) = \pi_0 f_0(z)/f(z)$ in the usual asymptotic manner. In contrast, inferences from the FWER depend crucially on the number of cases N (e.g., Bonferroni's method would declare case 1 non-null at significance level α if $p_1 \leq \alpha/N$).

It is not necessary to choose between Fdr and fdr; they can be used in tandem. For example, using Fdr and then using fdr gives the investigator both a list of likely prospects and quantitative differentiation within the list.

Local false discovery rates are directly related to "EFP" and "ETP", the expected number of false positives and true positives of a decision procedure. Actually, the decision rule that maximizes ETP for a given value of EFP rejects each null hypothesis at the same threshold value of its local false discovery rate.

$$\text{EFP} = \sum_{i=1}^{N} w_i \pi_{i0} \int_{c_i}^{\infty} f_{i0}(z_i)\, dz_i \quad \text{and} \quad \text{ETP} = \sum_{i=1}^{N} w_i \pi_{i1} \int_{c_i}^{\infty} f_{i1}(z_i)\, dz_i$$

Do simple power diagnostics.

1. The non-null counts y_{1k} can suffer from excess variability due to "histogram noise" in the observed counts y_k. An improved version is $\hat{y}_{1k} = Nd\hat{\pi}_1 \hat{f}_1(z)$ is called the smoothed non-null counts in what follows.

2. The entire non-null cdf of $\text{fdr}(z)$, $\Pr_1\{\widehat{\text{fdr}} \le q\} = \sum_{k:\widehat{\text{dr}_k} \le q} \hat{y}_{1k} / \sum_k \hat{y}_{1k} \quad [q \in (0,1)]$.

3. A simple but useful diagnostic summary statistic for power is $\widehat{E}\text{fdr}_1$, the expectation of $\widehat{\text{fdr}}_k$ under the $\hat{y}_{1k}$ distribution, $\widehat{E\text{fdr}}_1 = \sum_k \widehat{\text{fdr}}_k \hat{y}_{1k} / \sum_k \hat{y}_{1k}$. Low values of $\widehat{E}\text{fdr}_1$ indicate high power − that a typical non-null case will occur where the local false discovery rate is low − and high values indicate low power.

6 Theoretical, Permutation, and Empirical Null Distributions

The case that the theoretical null distribution $\mathcal{N}(0,1)$ provides a reasonably good fit is a less-than-usual occurrence. It is important to choose the null distribution and this section discuss several ways to determine a good null distribution. A key assumption for empirically estimating the null is that π_0, the proportion of null cases, is large.

Use Central Mapping. Let $f_{\pi 0}(z) = \pi_0 f_0(z)$. Central matching estimates $f_0(z)$ and π_0 by assuming that $\log(f(z))$ is quadratic near $z = 0$ (and equal to $f_{\pi 0}(z)$),

$$\log(f(z)) \doteq \beta_0 + \beta_1 z + \beta_2 z^2$$

estimating $(\beta_0, \beta_1, \beta_2)$ from the histogram counts y_k (the count in the k-th bin) around $z = 0$. When we assume that $f_0(z)$ is normal ($f_0(z) \sim \mathcal{N}\left(\delta_0, \sigma_0^2\right)$), $\log\left(f_{\pi 0}(z)\right) = \log\left(\pi_0\right) - \frac{1}{2}\left\{\frac{\delta_0^2}{\sigma_0^2} + \log\left(2\pi\sigma_0^2\right)\right\} + \frac{\delta_0}{\sigma_0^2} z - \frac{1}{2\sigma_0^2} z^2$ is indeed a quadratic function of z.

Uses MLE Method. Given the full set of z -values $z = (z_1, z_2, \ldots, z_N)$, let $\mathcal{I}_0 = \{i : z_i \in \mathcal{A}_0\}$ and $N_0 = \#\mathcal{I}_0$ and define $\boldsymbol{z}_0$ as the corresponding collection of z -values, $\boldsymbol{z}_0 = \{z_i, i \in \mathcal{I}_0\}$. Also, let $\varphi_{\delta_0,\sigma_0}(z)$ be the $\mathcal{N}\left(\delta_0, \sigma_0^2\right)$ density function

$$\varphi_{\delta_0,\sigma_0}(z) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2}\left(\frac{z-\delta_0}{\sigma_0}\right)^2\right\} \quad \text{and} \quad H_0\left(\delta_0, \sigma_0\right) \equiv \int_{\mathcal{A}_0} \varphi_{\delta_0,\sigma_0}(z) dz$$

this being the probability that a $\mathcal{N}\left(\delta_0, \sigma_0^2\right)$ variate falls in $\mathcal{A}_0$. We suppose that the Nz_i values independently follow the two-groups model with $f_0 \sim \mathcal{N}\left(\delta_0, \sigma_0^2\right)$ and $f_1(z) = 0$ for $z \in \mathcal{A}_0$. Then $\boldsymbol{z}_0$ has density and likelihood function

$$f_{\delta_0,\sigma_0,\pi_0}\left(\boldsymbol{z}_0\right) = \left[\binom{N}{N_0} \theta^{N_0} (1-\theta)^{N-N_0}\right] \left[\prod_{z_i \in \mathcal{I}_0} \frac{\varphi_{\delta_0,\sigma_0}\left(z_i\right)}{H_0\left(\delta_0, \sigma_0\right)}\right]$$

when

$$\theta = \pi_0 H_0\left(\delta_0, \sigma_0\right) = \Pr\left\{z_i \in \mathcal{A}_0\right\}$$

Notice that $\boldsymbol{z}_0$ provides $N_0 = \#z_0$, distributed as $\text{Bi}(N, \theta)$, while N is a known constant.

Remark 6.1. *The MLE method yielded smaller standard deviations for estimators. It tend to be less variable than central matching estimates, though more prone to bias.*

Why the Theoretical Null May Fail.

1. Failed mathematical assumptions: the idealized mathematical framework that i.i.d. normal components for a twosample t-statistic is often not the case. But this can be easily cured by permutation calculations.
2. Correlation across sampling units: independence across the n sampling units can fail in practice.
3. Correlation across cases: independence among the z-values is not required for valid false discovery rate inference. Actually, we have following disconcerting fact: even if the theoretical null distribution $z_i \sim \mathcal{N}(0,1)$ is valid for all null cases, correlation among the z_i can make $\mathcal{N}(0,1)$ a misleading choice in likely realizations of $\boldsymbol{z} = (z_1, z_2, \ldots, z_N)$.
4. Unobserved covariates: are ubiquitous in large-scale studies and are perhaps the most common source of trouble for the theoretical null.

7 Estimation Accuracy

Covariance of $\boldsymbol{y}$. Assume that the z_i are divided into a finite number of classes:

$$z_i \sim \mathcal{N}\left(\mu_c, \sigma_c^2\right) \quad \text{for } z_i \in C_c \tag{15}$$

Let N_c be the number of members of C_c, with π_c the proportion, $N_c = \#C_c$ and $\pi_c = N_c/N$ so $\sum_c N_c = N$ and $\sum_c \pi_c = 1$. Under the multi-class model eqn. (15), the covariance of the count vector $\boldsymbol{y}$ has two components

$$\mathbf{cov}(y) = \mathbf{cov}_0 + \mathbf{cov}_1$$

where

$$\mathbf{cov}_0 = N \sum_c \pi_c \left\{ \operatorname{diag}\left(\boldsymbol{P}_c\right) - \boldsymbol{P}_c \boldsymbol{P}_c' \right\}$$

and

$$\mathbf{cov}s_1 = N^2 \sum_c \sum_d \pi_c \pi_d \operatorname{diag}\left(\boldsymbol{P}_c\right) \boldsymbol{\lambda}_{cd} \operatorname{diag}\left(\boldsymbol{P}_d\right) - N \sum_c \operatorname{diag}\left(\boldsymbol{P}_c\right) \boldsymbol{\lambda}_{cc} \operatorname{diag}\left(\boldsymbol{P}_c\right)$$

Summations are over all classes. Here $\operatorname{diag}\left(\boldsymbol{P}_c\right)$ is the $K \times K$ diagonal matrix having diagonal elements $P_{kc} = \Pr_c\left\{z_i \in \mathcal{Z}_k\right\} \doteq d \cdot \varphi\left(x_{kc}\right)/\sigma_c$ similarly $\operatorname{diag}\left(\boldsymbol{P}_d\right)$, while $\boldsymbol{\lambda}_{cd}$ is the $K \times K$ matrix with kl th element $\lambda\left(x_{kc}, x_{ld}\right)$

$$\lambda(u,v) = \int_{-1}^{1} \lambda_\rho(u,v) g(\rho) d\rho,$$

$$\lambda_\rho(u,v) = \frac{\varphi_\rho(u,v)}{\varphi(u)\varphi(v)} - 1 = \left(1-\rho^2\right)^{-\frac{1}{2}} \exp\left\{ \frac{2\rho uv - \rho^2\left(u^2+v^2\right)}{2\left(1-\rho^2\right)} \right\} - 1$$

By Mehler's identity, $\lambda_\rho(u,v) = \sum_{j\geq 1} \frac{\rho^j}{j!} h_j(u) h_j(v)$ where h_j is the j th Hermite polynomial. Denoting the j th moment of the correlation distribution $g(\rho)$ by $\alpha_j = \int_{-1}^{1} \rho^j g(\rho) d\rho$. Then $\lambda(u,v) = \sum_{j\geq 1} \frac{\alpha_j}{j!} h_j(u) h_j(v)$, so $\boldsymbol{\lambda}_{cd}$ in can be expressed in outer product notation as $\boldsymbol{\lambda}_{cd} = \sum_{j\geq 1} \frac{\alpha_j}{j!} h_j\left(\boldsymbol{x}_c\right) h_j\left(\boldsymbol{x}_d\right)'$. Then define $\bar{\phi}^{(j)} = \sum_c \pi_c \varphi_c^{(j)}/\sigma_c$, the correlation penalty becomes

$$\mathbf{cov}_1 = N^2 d^2 \left\{ \sum_{j\geq 1} \frac{\alpha_j}{j!} \bar{\phi}^{(j)} \bar{\phi}^{(j)\prime} - \frac{1}{N} \sum_{j\geq 1} \frac{\alpha_j}{j!} \left(\sum_c \pi_c \varphi_c^{(j)} \varphi_c^{(j)\prime} / \sigma_c^2 \right) \right\}$$

Covariance of $\bar{F}$. Let $\boldsymbol{B}$ be the $K \times K$ matrix

$$\boldsymbol{B}_{kk'} = \begin{cases} 1 & \text{if } k \le k' \\ 0 & \text{if } k > k' \end{cases}$$

so $\overline{\boldsymbol{F}} = \frac{1}{N}\boldsymbol{B}\boldsymbol{y}$ is a K -vector with k th component the proportion of z_i values in bins indexed $\ge k$

$$\bar{F}_k = \#\left\{z_i \ge x_k - d/2\right\}/N \quad (k = 1, 2, \ldots, K).$$

Then $\mathrm{Cov}(\overline{\boldsymbol{F}}) = \mathbf{Cov}_0 + \mathbf{Cov}_1$, where $\mathbf{Cov}_0$ has kl-th entry

$$\frac{1}{N}\sum_c \pi_c \left\{\Phi^+\left(\max\left(x_{kc}, x_{lc}\right)\right) - \Phi^+\left(x_{kc}\right)\Phi^+\left(x_{lc}\right)\right\}$$

and

$$\mathbf{Cov}_1 = \sum_j \frac{\alpha_j}{j!}\bar{\varphi}^{(j-1)}\bar{\varphi}^{(j-1)'} - \frac{1}{N}\sum_j \frac{\alpha_j}{j!}\left\{\sum_c \pi_c \bar{\varphi}_c^{(j-1)}\bar{\varphi}_c^{(j-1)'}\right\}$$

Approximation for practical usage. Convenient approximations to $\mathbf{Cov}_1$ and $\mathbf{cov}_1$ are based on three simplifications:

1. The second terms are of order $1/N$ and can be ignored for N large.
2. By standalization, α_1 is exactly or nearly zero; this leaves α_2 as the lead.
3. With $|\rho| \le 1$, the higher-order moments α_j of $g(\rho), j \ge 3$ decrease quickly to zero if α_2 is not too large.

To summarize, the rms approximations and the homogeneity assumption $\sigma_c = \sigma_0$ together yield these convenient estimates for the covariance matrices of the counts and their cdf values:

$$\widehat{\mathbf{cov}}\{\boldsymbol{y}\} = N\left\{\mathrm{diag}(d\cdot\hat{\boldsymbol{f}}) - d^2\hat{\boldsymbol{f}}\hat{\boldsymbol{f}}'\right\} + \frac{\left(\hat{\alpha}Nd\hat{\sigma}_0^2\right)^2}{2}\hat{\boldsymbol{f}}^{(2)}\hat{\boldsymbol{f}}^{(2)'}$$

and

$$\widehat{\mathrm{Cov}}\{\overline{\boldsymbol{F}}\} = \frac{1}{N}\left\{\left(\bar{F}_{\max(k,l)} - \bar{F}_k\bar{F}_l\right)\right\} + \frac{\left(\hat{\alpha}\hat{\sigma}_0^2\right)^2}{2}\hat{\boldsymbol{f}}^{(1)}\hat{\boldsymbol{f}}^{(1)'}$$

Here $\overline{\boldsymbol{F}} = \left(\bar{F}\left(x_1\right), \bar{F}\left(x_2\right), \ldots, \bar{F}\left(x_K\right)\right)'$.

Under H_0, $z_i \approx \mathcal{N}(0,1)$. Suppose that $x_1, \cdot, x_n$ is such a sample, obtained from cdf F_θ, a member of a one-parameter family $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$ having its moment parameters { mean, standard deviation, skewness, kurtosis }, denoted by $\{\mu_\theta, \sigma_\theta, \gamma_\theta, \delta_\theta\}$ defined differentiably in θ. Under the null hypothesis $H_0 : \theta = 0$, which we can write as $H_0 : x \sim \{\mu_0, \sigma_0, \gamma_0, \delta_0\}$. the standardized variate $Y_0 = \sqrt{n}\left(\frac{\bar{x}-\mu_0}{\sigma_0}\right)$ $[\bar{x} = \sum_{i=1}^n y_i/n]$ satisfies $H_0 : Y_0 \sim \left\{0, 1, \frac{\gamma_0}{\sqrt{n}}, \frac{\delta_0}{n}\right\}$. Normality can be improved to second order by means of a Cornish-Fisher transformation,

$$Z_0 = Y_0 - \frac{\gamma_0}{6\sqrt{n}}\left(Y_0^2 - 1\right) \tag{16}$$

which reduces the skewness in from $O\left(n^{-1/2}\right)$ to $O\left(n^{-1}\right)$, $H_0 : Z_0 \sim \{0, 1, 0, 0\} + O\left(n^{-1}\right)$. We can interpret the last H_0 as saying that Z_0 is a second-order z-value, $H_0 : Z_0 \sim \mathcal{N}(0,1) + O_p\left(n^{-1}\right)$.

Under H_1, Z_0 decays fast. Suppose now that H_0 is false and instead H_1 is true, with $x_1, x_2, \ldots, x_n$ i.i.d. according to $H_1 : x \sim \{\mu_1, \sigma_1, \gamma_1, \delta_1\}$. Setting $Y_1 = \sqrt{n}\left(\frac{\bar{x}-\mu_1}{\sigma_1}\right)$ and $Z_1 = Y_1 - \frac{\gamma_1}{6\sqrt{n}}\left(Y_1^2 - 1\right)$ makes Z_1 second-order normal under H_1, $H_1 : Z_1 \sim \mathcal{N}(0,1) + O_p\left(n^{-1}\right)$ We wish to calculate the distribution of Z_0 eqn. (16) under H_1. Define $c = \sigma_1/\sigma_0$, $d = \sqrt{n}\left(\mu_1 - \mu_0\right)/\sigma_0$ and $g_0 = \gamma_0/(6\sqrt{n})$. Some simple algebra yields the following relationship between Z_0 and Z_1:

$$Z_0 = M + SZ_1 + g_0\left\{\left(\frac{\gamma_1}{\gamma_0}S - c^2\right)\left(Y_1^2 - 1\right) + \left(1 - c^2\right)\right\} \tag{17}$$

where

$$M = d \cdot (1 - dg_0) \quad \text{and} \quad S = c \cdot (1 - 2dg_0)$$

The asymptotic relationships claimed at the start of this section are easily derived from eqn. (17). We consider a sequence of alternatives θ_n approaching the null hypothesis value θ_0 at rate $n^{-1/2}$:

$$\theta_n - \theta_0 = O\left(n^{-1/2}\right). \tag{18}$$

The parameter $d = \sqrt{n}\left(\mu_{\theta_n} - \mu_0\right)/\sigma_0$ is then of order $O(1)$ as is $M = d\left(1 - dg_0\right) = d\left(1 - d\gamma_0/(6\sqrt{n})\right)$. while standard Taylor series calculations give

$$c = 1 + \frac{\dot{\sigma}_0}{\dot{\mu}_0}\frac{d}{\sqrt{n}} + O\left(n^{-1}\right) \quad \text{and} \quad S = 1 + \left(\frac{\dot{\sigma}_0}{\dot{\mu}_0} - \frac{\gamma_0}{3}\right)\frac{d}{\sqrt{n}} + O\left(n^{-1}\right) \tag{19}$$

the dot indicating differentiation with respect to θ. Then under H_1 and eqn. (18),

$$Z_0 \sim \mathcal{N}\left(M, S^2\right) + O_p\left(n^{-1}\right)$$

with M and S as given in eqn. (19). Moreover,

$$\left.\frac{dS}{dM}\right|_{\theta_0} = \frac{1}{\sqrt{n}}\left(\left.\frac{d\sigma}{d\mu}\right|_{\theta_0} - \frac{\gamma_0}{3}\right) + O\left(n^{-1}\right).$$

They show that, under repeated sampling, the non-null distribution of a typical z-value will have standard deviation of order $O\left(n^{-1/2}\right)$ and non-normality $O_p\left(n^{-1}\right)$ as we move away from the null hypothesis: so $z_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$ holds to a good approximation, though we can expect σ_i^2 to differ noticeably from 1 when μ_i is far from 0.

8 Correlation Questions

Row and Column Correlations. We will be interested in both the row-wise and column-wise correlations of the the $N \times n$ matrix $X = (x_{ij})$. It simplifies notation to assume that X has been "demeaned" by the subtraction of row and column means, so that

$$\sum_{i=1}^{N} x_{ij} = \sum_{j=1}^{n} x_{ij} = 0 \quad \text{for } i = 1, 2, \ldots, N \text{ and } j = 1, 2, \ldots, n \tag{20}$$

Sometimes we will go further and assume double standardization: that in addition to eqn. (20),

$$\sum_{i=1}^{N} x_{ij}^2 = N \text{ and } \sum_{j=1}^{n} x_{ij}^2 = n \text{ for } i = 1, 2, \ldots, N \text{ and } j = 1, 2, \ldots, n$$

Double standardization is convenient for our discussion because it makes sample covariances into sample correlations. The $n \times n$ column sample covariance matrix $\hat{\mathbf{\Delta}}$ and the $N \times N$ row sample covariance matrix $\hat{\mathbf{\Sigma}}$ are $\hat{\mathbf{\Delta}} = X'X/N$ and $\hat{\mathbf{\Sigma}} = XX'/n$.

ZOur main result is stated and proved in terms of the singular value decomposition (svd) of X,

$$\underset{N\times n}{X} = \underset{N\times K}{U}\,\underset{K\times K}{\boldsymbol{d}}\,\underset{K\times n}{V'}$$

where K is the rank of X, d is the diagonal matrix of ordered singular values $d_1 \geq d_2 \geq \cdots \geq d_K > 0$, and U and V are orthonormal matrices of sizes $N \times K$ and $n \times K$ $U'U = V'V = I_K$ with I_K the $K \times K$ identity. The squares of the singular values, $e_1 \geq e_2 \geq \cdots \geq e_K > 0 \quad \left[e_k = d_k^2\right]$ are the eigenvalues of $N\hat{\mathbf{\Delta}} = V\boldsymbol{d}^2V'$ and also of $n\hat{\mathbf{\Sigma}} = U\boldsymbol{d}^2U'$.

If X has row and column means equal to zero, then the n^2 elements of $\hat{\Delta}$ and the N^2 elements of $\hat{\mathbf{\Sigma}}$ both have mean zero and empirical variance

$$c_2 = \sum_{k=1}^{K} e_k^2/(Nn)^2. \tag{21}$$

Estimate α. We assume that the columns $\boldsymbol{x}$ of $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ are independent and identically distributed N -vectors, with correlation $\rho_{ii'}$ between entries x_i and $x_{i'}$. The rms correlation is $\alpha = \left[\sum_{i<i'} \rho_{ii'}^2 \Big/ \binom{N}{2}\right]^{\frac{1}{2}}$. Define $\bar{\alpha}$ to be the observed row-wise root mean square correlation, $\bar{\alpha} = \left[\sum_{i<i'} \hat{\rho}_{ii'}^2 \Big/ \binom{N}{2}\right]^{\frac{1}{2}}$ In the doubly standardized case, $\bar{\alpha}^2 = \frac{Nc_2-1}{N-1} \doteq c_2$. Our preferred estimator for the rms correlation is

$$\hat{\alpha} = \left[\frac{n}{n-1}\left(\bar{\alpha}^2 - \frac{1}{n-1}\right)\right]^{\frac{1}{2}} \tag{22}$$

References

[1] Yosef Hochberg. A sharper bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4):800–802, 1988.

[2] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[3] Ruth Marcus, Peritz Eric, and K Ruben Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660, 1976.

[4] Sanat K Sarkar et al. On the simes inequality and its generalization. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, pages 231–242. Institute of Mathematical Statistics, 2008.

[5] R John Simes. An improved bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.