

ESTIMATING THE NUMBER OF FALSE NULL HYPOTHESES WHEN CONDUCTING MANY TESTS

July 8, 2003

Dan Nettleton

J. T. Gene Hwang

Department of Statistics

Departments of Mathematics and Statistics

Iowa State University

Cornell University

ABSTRACT

Mosig et al. (2001) propose an intuitively appealing method for estimating the number of null hypotheses that are false in a multiple test situation. They present an iterative algorithm that relies on the distribution of observed p -values to obtain their estimator. We characterize the limit of their iterative algorithm and show that their estimator can be computed directly without iteration from the observed distribution of p -values.

1. Introduction. Many genomics experiments involve testing hundreds or thousands of null hypotheses. For example, when searching for associations between genetic loci and a trait of interest, a collection of loci spread throughout the genome at 1 cM increments may be tested, with each null hypothesis being that of no association between the locus and the trait. In microarray experiments it is common to test, for each of thousands of genes, the null hypothesis that the mRNA level of the gene is not altered by a treatment. In such situations it is natural to wish to estimate the number

of loci that are associated with the trait or the number of genes whose mRNA level is altered by treatment.

In addition to being of direct scientific interest, estimates of the number of false null hypotheses can help researchers understand how many additional null hypotheses could be rejected with a more powerful experiment. For example, suppose that 100 of 1000 null hypotheses have been rejected using a method that provides some satisfactory control of an error rate of interest. If it is estimated from the same data set that 500 of the 1000 null hypotheses are false, the researcher has information that perhaps as many as 400 more discoveries could be made with a more powerful experiment.

Estimates of the number of false null hypotheses are also very useful for estimating the proportion of mistakenly rejected null hypotheses among all rejected null hypotheses. Such measures have become very popular in recent years due to the work of Storey (2002, 2003), Storey and Tibshirani (2001), and Storey et al. (2003). As a simple example suppose we conduct 10,000 tests and reject H_{0i} if $p_i \leq 0.01$. Further, suppose 120 p -values are less than 0.01. If we are able to determine that 9,000 null hypotheses are true and 1,000 are false, we would expect around 90 of the 120 rejected null hypotheses to be false positives. Such a performance is likely to be considered unacceptably poor. On the other hand, if we are able to determine that 1,000 null hypotheses are true and 9,000 are false, we would expect only about 10 of the 120 to be false positives, and the list of tests for which the null hypothesis was rejected would be considerably more useful. Distinguishing between the two situations clearly requires knowledge of the number false null hypotheses among all null hypotheses tested.

In this paper we consider the problem of estimating the number of false null hypotheses in a collection of n tests, given the observed p -values p_1, \dots, p_n for the collection of tests. We focus

on the method proposed by Mosig et al. (2001). Section 2 describes this method in detail. Section 3 contains a proof that the iterative algorithm proposed by Mosig et al. (2001) does converge to a limit as the number of iterations goes to infinity. Furthermore, we characterize the limit and show that the estimator can be computed directly from the observed p -values with no iteration.

2. An Intuitive Algorithm for Estimating the Number of False Null Hypotheses. Suppose a collection of n null hypotheses H_{01}, \dots, H_{0n} is to be tested. Suppose that, for all $\alpha \in (0, 1)$, each null hypothesis is tested using a test for which the probability of rejecting a true null hypothesis is exactly α whenever the test is conducted at significance level α . Then the p -values corresponding to tests with true null hypotheses will follow a uniform distribution on the interval $(0, 1)$. Further suppose that the tests used are unbiased, implying that the probability of rejecting a false null hypothesis with a significance-level- α test will be greater than or equal to α . (Note that the rejection probability must be greater than α in some regions of the alternative space so that the rejection probability is not simply equal to a constant α regardless of the data.) When these conditions are satisfied, the distribution of observed p -values from the n tests will be a mixture of a uniform distribution and a distribution that is stochastically smaller than uniform. The mixing proportion associated with the uniform distribution gives the proportion of tests with true null hypotheses while one minus that mixing proportion gives the proportion of tests with false null hypotheses.

Mosig et al. (2001) propose an iterative algorithm that essentially estimates the proportion of observed p -values that follow a uniform distribution. The previously published description of their algorithm is somewhat unclear on a few points. By examining the examples presented by Mosig et al. (2001), we were able to arrive at the following description of their algorithm that can be used to reproduce the results in Table 2 of Mosig et al. (2001).

Informal Description of the Iterative Algorithm:

1. The interval $[0, 1]$ is partitioned into B bins of equal width.
2. Initially assume that all null hypotheses are true.
3. Find the expected number of p -values for each bin given the current estimate of the number of true null hypotheses, bearing in mind that p -values corresponding to true null hypotheses should be uniformly distributed on the interval $(0, 1)$.
4. Beginning with the leftmost bin, sum the number of p -values in excess of expected until a bin with no excess is reached.
5. The excess sum is an estimate of the number of false null hypotheses. The number of tests minus this estimate provides an estimate of the number of true null hypotheses.
6. Return to step 3 and repeat until convergence.

We now give a more formal definition of this algorithm that facilitates the proof in Section 3.

Formal Definition of the Iterative Algorithm:

Suppose the interval $[0, 1]$ is partitioned into B bins numbered from 1 to B such that bin 1 is $[0, 1/B]$ and bin i is $(\frac{i-1}{B}, \frac{i}{B}]$ for $i = 2, \dots, B$. Let n_i denote the number of p -values falling into bin i for $i = 1, \dots, B$. Let $\bar{n}_{i:B} = \sum_{j=i}^B n_j / (B - i + 1)$ for any $i = 1, \dots, B$. Let $N_0 = \sum_{i=1}^B n_i = B\bar{n}_{1:B}$ and define

$$\frac{N_k}{B} = \left(\frac{i_k - 1}{B}\right) \frac{N_{k-1}}{B} + \left(1 - \frac{i_k - 1}{B}\right) \bar{n}_{i_k:B} \quad \text{for all } k \geq 1 \quad (1)$$

where

$$i_k \equiv \min \left\{ i : n_i \leq \frac{N_{k-1}}{B} \right\}.$$

Then N_k is the estimated number of true nulls at iteration k of the algorithm introduced by Mosig et al. (2001).

3. Existence and Characterization of the Limit.

Convergence Theorem: Let $I = \min\{i : n_i \leq \bar{n}_{i:B}\}$. Then $\lim_{k \rightarrow \infty} N_k = B\bar{n}_{I:B}$.

Proof: The proof of the main theorem rests on the following facts which we state here and prove at the end of this section.

A. $\bar{n}_{1:B} > \bar{n}_{2:B} > \cdots > \bar{n}_{I:B}$.

B. If $i_k < I$, then (i) $i_k \leq i_{k+1}$, (ii) $i_{k+1} \leq I$, and (iii) $i_k < i_{k^*}$ for some $k^* > k$.

C. If there exists k^* such that $i_{k^*} = I$, then $i_k = I$ for all $k \geq k^*$.

D. Suppose $\{a_k\}_{k \geq 0}$ is an infinite sequence of real numbers. If there exists $\lambda \in [0, 1)$, an integer

k^* , and a real number a such that $a_k = \lambda a_{k-1} + (1-\lambda)a$ whenever $k \geq k^*$, then $\lim_{k \rightarrow \infty} a_k =$

a .

We begin by showing that the sequence $\{i_k\}$ converges to I in a finite number of iterations.

Note that if $I = 1$, then $n_1 \leq \bar{n}_{1:B} = N_0/B$. Thus $i_1 = 1 = I$ by the definition of i_1 , and fact C guarantees that $i_k = I$ for all $i \geq 1$. If $I > 1$, then $i_1 \leq I$ because $n_I \leq \bar{n}_{I:B} < \bar{n}_{1:B} = N_0/B$ by fact A and the definitions of i_1 and I . Now if $i_1 = I$, fact C guarantees that $i_k = I$ for all $i \geq 1$. If $i_1 < I$, parts (i) through (iii) of fact B imply that i_{k^*} must equal I for some $k^* > 1$. Hence $i_k = I$

for all $k \geq k^*$ by fact C. Thus, in all cases, there is some $k^* \geq 1$ satisfying $i_k = I$ for any $k \geq k^*$.

It follows from (1) that

$$\frac{N_k}{B} = \left(\frac{I-1}{B}\right) \frac{N_{k-1}}{B} + \left(1 - \frac{I-1}{B}\right) \bar{n}_{I:B} \quad \text{for all } k \geq k^*.$$

Now fact D implies that $\lim_{k \rightarrow \infty} \frac{N_k}{B} = \bar{n}_{I:B}$, and the desired result follows. \square

Proof of Fact A: For all $i < I$ we have

$$n_i > \bar{n}_{i:B} = \left(\frac{1}{B-i+1}\right) n_i + \left(1 - \frac{1}{B-i+1}\right) \bar{n}_{i+1:B} \quad (2)$$

Straightforward manipulation of (2) yields $n_i > \bar{n}_{i+1:B}$. Now substituting $\bar{n}_{i+1:B}$ for n_i in the right side of equation (2) yields $\bar{n}_{i:B} > \bar{n}_{i+1:B}$. Thus

$$\bar{n}_{1:B} > \bar{n}_{2:B} > \cdots > \bar{n}_{I:B}. \square$$

Proof of Fact B(i): By the definition of I , $n_{i_k} > \bar{n}_{i_k:B}$ whenever $i_k < I$. By the definition of i_k ,

$n_{i_k} \leq N_{k-1}/B$. Thus $\bar{n}_{i_k:B} < N_{k-1}/B$. Substituting N_{k-1}/B for $\bar{n}_{i_k:B}$ in (1) yields

$$\frac{N_k}{B} < \left(\frac{i_k-1}{B}\right) \frac{N_{k-1}}{B} + \left(1 - \frac{i_k-1}{B}\right) \frac{N_{k-1}}{B} = \frac{N_{k-1}}{B}.$$

Therefore $n_i \leq N_k/B$ implies that $n_i < N_{k-1}/B$. It follows that $\{i : n_i \leq N_k/B\} \subseteq \{i : n_i \leq N_{k-1}/B\}$. Thus $i_k = \min\{i : n_i \leq N_{k-1}/B\} \leq \min\{i : n_i \leq N_k/B\} = i_{k+1}$. \square

Proof of Fact B(ii): By definition of i_k , $n_{i_k} \leq N_{k-1}/B$. Thus, using fact A and the definition of I , we have

$$n_I \leq \bar{n}_{I:B} \leq \bar{n}_{i_k:B} < n_{i_k} \leq N_{k-1}/B$$

whenever $i_k < I$. Thus

$$n_I < \left(\frac{i_k-1}{B}\right) \frac{N_{k-1}}{B} + \left(1 - \frac{i_k-1}{B}\right) \bar{n}_{i_k:B} = \frac{N_k}{B}$$

by (1) which implies that $i_{k+1} \leq I$ by the definition of i_{k+1} . \square

Proof of Fact B(iii): Suppose $i_k < I$ and $i_\ell \leq i_k$ for all $\ell > k$. This would imply that $i_\ell = i_k$ for all $\ell \geq k$ by fact B(i). Thus

$$\frac{N_\ell}{B} = \left(\frac{i_k - 1}{B}\right) \frac{N_{\ell-1}}{B} + \left(1 - \frac{i_k - 1}{B}\right) \bar{n}_{i_k:B} \quad \text{for all } \ell \geq k.$$

Fact D implies that $\lim_{\ell \rightarrow \infty} N_\ell/B = \bar{n}_{i_k:B}$. However, $i_k < I$ implies that $n_{i_k} > \bar{n}_{i_k:B}$. Thus there exists $k^* > k$ such that $n_{i_k} > N_{\ell-1}/B$ whenever $\ell \geq k^*$. This implies that $i_\ell \neq i_k$ when $\ell \geq k^*$.

We have reached a contradiction. Therefore the result follows. \square

Proof of Fact C: If $i_k = I$, then (1) implies that

$$\frac{N_k}{B} = \left(\frac{I - 1}{B}\right) \frac{N_{k-1}}{B} + \left(1 - \frac{I - 1}{B}\right) \bar{n}_{I:B}. \quad (3)$$

By the definition of I , $n_I \leq \bar{n}_{I:B}$. By the definition of i_k , $n_I \leq N_{k-1}/B$ when $i_k = I$. Thus substituting n_I for N_{k-1}/B and n_I for $\bar{n}_{I:B}$ in (3) yields $N_k/B \geq n_I$. Hence $i_{k+1} \leq I$. Now by fact A and the definition of I , $n_i > \bar{n}_{I:B}$ for any $i < I$. Furthermore $i_k = I$ implies $n_i > N_{k-1}/B$ for any $i < I$. Thus substituting n_i for $\bar{n}_{I:B}$ and n_i for N_{k-1}/B in (3) implies $n_i > N_k/B$ for any $i < I$. Thus $i_{k+1} \geq I$. We have shown that $i_{k+1} \leq I$ and $i_{k+1} \geq I$. Thus $i_{k+1} = I$. \square

Proof of Fact D: Let $b_n = a_{k^*+n-1}$ for all $n \geq 0$. Then $b_1 = \lambda b_0 + (1 - \lambda)a$, and an induction argument shows that $b_n = \lambda^n b_0 + a(1 - \lambda) \sum_{i=0}^{n-1} \lambda^i$ for all $n \geq 1$. Now $\lambda \in [0, 1)$ implies that

$$\lim_{n \rightarrow \infty} \lambda^n = 0 \quad \text{and} \quad \sum_{i=0}^{\infty} \lambda^i = (1 - \lambda)^{-1}.$$

Thus $\lim_{n \rightarrow \infty} b_n = a$, and $\lim_{k \rightarrow \infty} a_k = a$ since $\{a_k\}_{k \geq k^*} = \{b_n\}_{n \geq 1}$. \square

4. An Example

Suppose 100 tests are conducted with observed p -values distributed as follows:

Interval	Number of P -Values
[0.0, 0.2]	36
(0.2, 0.4]	22
(0.4, 0.6]	20
0.6, 0.8]	10
(0.8, 1.0]	12

It is straightforward to show that the iterative algorithm of Mosig et al. (20001) produces the sequence $N_0 = 100, N_1 = 82, N_2 = 71.2, N_3 = 64.72, N_4 = 60.832, N_5 = 58.4992, N_6 = 57.09952, N_7 = 56.259712, N_8 = 55.7558272, N_9 = 55.45349632, N_{10} = 55.272097792, N_{11} = 55.1632586752, N_{12} = 55.09795520512, N_{13} = 55.058773123072, \dots$ Due to round-off error the algorithm will converge in a finite number of steps to an estimate of 55 for the number of true null hypotheses. Thus $100 - 55 = 45$ is the estimate of the number of false null hypotheses according to the iterative procedure.

Now note that $n_1 = 36 > 20 = \bar{n}_{1:5}, n_2 = 22 > 16 = \bar{n}_{2:5}, n_3 = 20 > 14 = \bar{n}_{3:5},$ and $n_4 = 10 \leq 11 = \bar{n}_{4:5}.$ Thus $I = 4$ and the limiting value for the iterative algorithm is $B\bar{n}_{I:B} = 5\bar{n}_{4:5} = 5(11) = 55$ according to the Convergence Theorem of Section 3. The estimated number of false null hypothesis is $100 - 55 = 45.$

5. References

Mosig, M. O., Lipkin, E., Galina, K. Tchourzyna, E., Soller, M., and Friedmann, A. (2001). A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. *Genetics* **157**, 1683-1698.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* **64**, 479-498.

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics* in press.

Storey, J. D., Taylor, J. E., and Siegmund, D. (2003). Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B* in press.

Storey, J. D. and Tibshirani, R. (2001). Estimating false discovery rates under dependence, with applications to DNA microarrays. Technical Report 2001-28, Department of Statistics, Stanford University.