



## Calculation of Critical Values for Somerville's FDR Procedures

Paul N. Somerville  
University of Central Florida

---

### Abstract

A Fortran 95 program has been written to calculate critical values for the step-up and step-down FDR procedures developed by Somerville (2004). The program allows for arbitrary selection of number of hypotheses, FDR rate, one- or two-sided hypotheses, common correlation coefficient of the test statistics and degrees of freedom. An MCV (minimum critical value) may be specified, or the program will calculate a specified number of critical values or steps in an FDR procedure.

The program can also be used to efficiently ascertain an upper bound to the number of hypotheses which the procedure will reject, given either the values of the test statistics, or their  $p$  values. Limiting the number of steps in an FDR procedure can be used to control the number or proportion of false discoveries (Somerville and Hemmelmann 2007). Using the program to calculate the largest critical values makes possible efficient use of the FDR procedures for very large numbers of hypotheses.

*Keywords:* FDR, critical values, Fortran, Somerville's procedure.

---

## 1. Introduction

Recent developments in diverse fields, including genomics, have resulted in very large data sets where it may be desired to simultaneously test many thousands of null hypotheses. Traditional procedure to control the probability of rejecting true hypotheses in multiple hypotheses testing has been to control the family-wise error rate (FWER). For a large number of hypotheses, however, extremely low power for testing single hypotheses may result.

Recently, procedures have been proposed which control the false discovery rate (FDR). The false discovery rate is defined to be the expected value of the proportion of rejected hypotheses which are actually true, with the false discovery rate defined to be zero when no hypotheses are rejected. Using an FDR procedure, the hypotheses  $H_1, H_2, \dots, H_m$ , are simultaneously tested using the corresponding test statistics  $t_1, t_2, \dots, t_m$ . Let  $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(m)}$  be

the ordered test statistic values, and denote by  $H_{(i)}$  the hypotheses corresponding to  $T_{(i)}$ . Then critical constants  $d_1 \leq d_2 \leq \dots \leq d_m$ , are used to compare  $T_{(i)}$  with  $d_i$ .

For a step-down FDR procedure, for  $i = m, m - 1, \dots$ , compare  $T_{(i)}$  with  $d_i$  until for the first time  $T_{(i)} < d_i$ , rejecting in turn all the hypotheses for which  $T_{(i)} \geq d_i$ . For a step-up FDR procedure, for  $i = 1, 2, \dots$ , compare  $T_{(i)}$  with  $d_i$  until for the first time  $T_{(i)} \geq d_i$ , rejecting all the hypotheses, except those for which the comparisons already made have shown  $T_{(i)} \leq d_i$ . The critical values  $d_i$  are chosen such that the expected value of the FDR is always less than some value  $q$  (or  $\alpha$ ). FDR procedures may also be defined in terms of critical  $p$  values. Benjamini and Hochberg (1995) proposed an FDR step-up procedure valid for independent test statistics. Benjamini and Lui (1999) presented a step-down FDR procedure valid under the same conditions. Sarkar (2002) showed that both procedures were valid under positive dependency. Troendle (2000) developed both step-up and step-down FDR procedures which asymptotically control the FDR when the test statistics have a multivariate  $t$  distribution. Somerville (2004), assuming a multivariate  $t$  distribution and common correlation of the test statistics used least favorable configurations to develop step-down and step-up FDR procedures. His step-down procedure yielded the same critical values as those of Troendle. It should be stated that, although extensive calculations strongly suggested the validity of the assumed location of the least favorable locations, a rigorous proof was not realized. Also, instead of setting negative critical values equal to zero, Somerville used a minimum critical value (MCV). The step-down FDR procedures of Troendle and Somerville have been shown by Horn and Dunnett (2004) and Somerville (2003) to have superior power, under the assumption that the test statistics have a multivariate  $t$  distribution.

While FDR procedures control the expected value of the proportion of true hypotheses which are rejected, it is possible that a very large number of true hypotheses could be rejected. To control the number of true hypotheses, which are falsely rejected, van der Laan, Dudoit, and Pollard (2004) proposed the generalized family-wise error rate  $\text{gFWER}(k)$ . It is defined as the probability of at least  $(k + 1)$  Type 1 errors ( $k = 0$  for the usual FWER), i.e.  $\text{P}[U \leq k] \geq 1 - \alpha$  where  $U$  is the number of true hypotheses which are rejected. Augmentation procedures proposed by van der Laan *et al.* (2004) and Dudoit, van der Laan, and Birkner (2004) can then be used to control PFP (proportion of false positives), where  $\text{P}[\text{PFP} \leq \gamma] \geq 1 - \alpha$  for some pre-specified  $\gamma$  and  $\alpha$ . More recently Lehmann and Romano (2005) suggested new methods of controlling  $\text{gFWER}(k)$  (called by them  $k$ -FWER) and PFP (called by them FDP - False Discovery Proportion). Both single-step and step-down procedures were derived which control the  $k$ -FWER. The procedures make no assumptions concerning the dependence structure or the  $p$  values of the individual tests. The step-down procedure is simple to apply, and cannot be improved without violation of control of the  $k$ -FWER. Lehmann and Romano (2005) also proposed two methods for controlling the PFP. The first holds under "mild conditions on the dependence structure of  $p$  values" while the second requires no dependence assumptions.

Somerville (2004) noticed that by limiting the number of steps in his procedures, the FDR could be reduced with a small subsequent loss of power. (An  $s$  step procedure is equivalent to an  $m$  step procedure where the  $m - s$  smallest critical values are replaced with the value  $d_{m-s+1}$ , also called MCV. Thus, only the largest  $s$  critical values are used in the comparisons). Somerville and Hemmelmann (2007), by limiting the maximum number of steps in step-down and step-up procedures, developed new procedures to control  $k$ -FWER. The procedures require only an arbitrary set of critical constants  $d_1 \leq d_2 \leq \dots \leq d_m$ , and need not be related to any step-down or step-up procedure. However, by starting with a set of critical values

which satisfy an FDR requirement, the procedures can be used to simultaneously control FDR and  $k$ -FWER requirements. Using data from the literature, the method of [Somerville and Hemmelmann \(2007\)](#) was compared with those of [Lehmann and Romano \(2005\)](#), and under the assumption of multivariate normality and a common correlation of the test statistics, considerable improvement in the reduction of false positives, in control of PFP, and in power, were accomplished while still controlling the FDR. The Fortran 95 program `seq.f95` can be used in four different modes to calculate the critical values for the procedure of [Somerville \(2004\)](#). In mode 1, the program calculates the critical values for step-down or step-up FDR procedures for one- or two-sided hypotheses, arbitrary degrees of freedom, arbitrary FDR levels and arbitrary common correlation coefficients. Using mode 3, the program calculates the largest critical values (number of critical values is specified by the user). Using mode 4, the Fortran 95 program, given the test statistic values, efficiently calculates an upper bound to the number of hypotheses which will be rejected. Given  $p$  values for the test statistics, the program, using mode 5, converts the  $p$  values to test statistic values and also obtains an upper bound to the number of rejected hypotheses. This paper has a number of objectives:

1. Describe the ways in which the Fortran program `seq` can be used in hypotheses testing, when the number of hypotheses is large.
2. Describe how the program obtains the constants used in Somerville's FDR procedures.
3. Present instructions and examples, which make it relatively easy for a user, using the Fortran program `seq` to solve many hypotheses testing problems involving a large number of hypotheses.

## 2. Calculation of critical values

The procedure for calculating the critical values is described in [Somerville \(2004\)](#). Let  $H_1, H_2, \dots, H_m$  be  $m$  hypotheses to be simultaneously tested using the test statistics  $T_1, T_2, \dots, T_m$ . We require that the false discovery rate (FDR), be less than a pre-specified value  $q$  (or  $\alpha$ ). If  $Q$  is the proportion of rejected hypotheses which are true, then  $\text{FDR} = \text{E}(Q)$ . We wish to calculate the  $m$  critical values  $d_1 \leq d_2 \leq \dots \leq d_m$ , such that  $\text{E}(Q) \leq q$ . We outline here the procedure for the step-down case.

Let  $A_i$  be the probability that exactly  $i$  hypotheses are rejected. Then

$$\begin{aligned} A_0 &= \text{P}[T_{(m)} < d_m] \\ A_1 &= \text{P}[T_{(m)} \geq d_m, T_{(m-1)} < d_{m-1}] \\ &\dots \\ A_{m-1} &= \text{P}[T_{(m)} \geq d_m, \dots, T_{(2)} \geq d_2, T_{(1)} < d_1] \\ A_m &= \text{P}[T_{(m)} \geq d_m, \dots, T_{(2)} \geq d_2, T_{(1)} \geq d_1] \end{aligned}$$

where  $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(m)}$ . To obtain the critical values, use is made of  $m$  least favorable configurations of the location parameters of the test statistics. Define  $LFC_i$  as the configuration where  $i$  of the location parameters are zero and the remainder are infinite. (The case where all  $m$  hypotheses are true corresponds to  $LFC_m$ ). To obtain  $d_1$ , use  $LFC_1$ . Then, with probability 1,  $T_{(m)}, \dots, T_{(2)}$  are infinite, and  $A_0, \dots, A_{m-2}$  are zero, and  $\text{E}(Q) = A_{m-1}(0/(m-1)) + A_m(1/m) \leq q$  or  $\text{P}[T_{(1)} \geq d_1] \leq mq$ .

To maximize power considerations, we choose the smallest value  $d_i$  which satisfies the equation. To obtain  $d_i$ , ( $1 < i \leq m$ ), given the values of  $d_1, \dots, d_{i-1}$ , use  $LFC_i$ . With probability 1,  $T_{(m)}, T_{(m-1)}, \dots, T_{(i+1)}$  are infinite and  $A_0, \dots, A_{m-i-1}$  are zero. Under  $LFC_i$ , when exactly  $r(\geq m-i)$  hypotheses are rejected, the proportion of true hypotheses rejected equals  $(r - (m-i)/r)$  with probability 1. We may then write, using the basic expectation algorithm,

$$E(Q) = A_{m-i}(0/(m-i)) + A_{m-i+1}(1/(m-i+1)) + \dots + A_m(i/m)$$

where

$$\begin{aligned} A_{(m-i+1)} &= \mathbb{P}[T_{(i)} \geq d_i, T_{(i-1)} < d_{(i-1)}]^1 \\ &\dots \\ A_{m-1} &= \mathbb{P}[T_{(i)} \geq d_i, \dots, T_{(2)} \geq d_2, T_{(1)} < d_1] \\ A_m &= \mathbb{P}[T_{(i)} \geq d_i, \dots, T_{(2)} \geq d_2, T_{(1)} \geq d_1]. \end{aligned}$$

Since  $A_{m-i+1}, \dots, A_m$  are each decreasing functions of  $d_i$ , we choose  $d_i$  as the smallest value such that  $E(Q) \leq q$ . The value for  $d_m$  for the FDR step-down procedure may also be calculated using the equation  $\mathbb{P}[T_{(m)} < d_m] = 1 - q$ . For the FDR step-up procedure, the formula slightly underestimates  $d_m$ .

### 3. Fortran 95 implementation

#### 3.1. Mode 1

The program first calculates  $d_1$  and  $d_m$ . The problem is to find the smallest value for  $d_i$  given previously obtained values  $d_1, \dots, d_{i-1}$ , subject to  $FDR = E(Q) \leq q$ . Since  $d_1 \leq d_2 \leq \dots \leq d_m$ , starting lower and upper bounds, `boundXL` and `boundXU`, for  $d_i$  are  $d_{i-1}$  and  $d_m$ . (It is well known that  $d_m$  for the step-up case is never less than  $d_m$  for the step-down case for the same parameter values. Extensive calculations suggest that the ratio is never greater than approximately 1.01, with the values close to 1.0 for large values of  $m$ . For step-up FDR, the upper bound for  $d_m$  is "conservatively" replaced with 1.1 times the values of  $d_m$  for the step-down case.) Starting "attained" lower and upper bounds, `boundFL` and `boundFU`, for FDR are set at 0 and 1. Two initial estimates of  $d_i$ ,  $x_1$  and  $x_2$  are chosen to calculate the corresponding values of FDR, namely  $F_1$  and  $F_2$ . (The subroutines `getFDR` and `getFDRup` are used to obtain the corresponding FDR values.) Using those calculations, the subroutine `bound(X,F)` is used to improve both sets of bounds. Linear interpolation is then used to obtain a third estimate  $x_3$  (modified, if necessary, to keep it within the existing bounds), and the corresponding FDR value  $F_3$  calculated. If  $F_1$  is closer to  $q$  than  $F_2$ , then the values  $x_2$  and  $F_2$  are replaced with the values  $x_3$  and  $F_3$ , and likewise the values  $x_3$  and  $F_3$  replace the values  $x_1$  and  $F_1$  if  $F_2$  is closer. Again linear interpolation using the new pair,  $x_1$  and  $x_2$  is used to obtain a new  $x_3$  and the iteration continues until the process is defined to converge. The program attempts to obtain an estimate for  $d_i$  with an FDR within .0001 of  $q$ .

#### Subroutine `getFDR`

The subroutine `getFDR` (`getFDRup` for the step-up case) uses Monte Carlo to get estimates for  $A_{m-i+1}$  to  $A_m$  (step-down case) given the previously obtained values  $d_1, \dots, d_{i-1}$  and an

<sup>1</sup>Somerville (2004), page 102 line 18, gives an erroneous formula for  $A_{m-i+1}$ .

estimate for  $d_i$ . The estimate of  $A_j$  is the proportion of the  $n$  vectors, whose elements are the randomly obtained values for the  $m$  test statistics, for which the FDR procedure rejects exactly  $j$  hypotheses under  $LFC_i$ . The FDR is then obtained as

$$E(Q) = A_{m-i}(0/(m-i)) + A_{m-i+1}(1/(m-i+1)) + \dots + A_m(i/m).$$

*Subroutine* bound(X,L)

The subroutine `bound(X,L)` replaces `boundXU` with  $x_i$  and `boundFL` with  $FDR_i$  if `boundFL`  $< F_i < q$ , and replaces `boundXL` with  $x_i$ , and `boundFU` with  $FDR_i$  if  $q < F_i < \text{boundFU}$ .

### 3.2. Mode 3

In mode 3, the program calculates the pre-specified “ $ic$ ” “unique” values (number of steps  $s$ ). The values  $d_1, \dots, d_{m-ic+1}$  are equal and trial values  $x_1$  and  $x_2$  are chosen to calculate the corresponding FDR values  $F_1$  and  $F_2$ . As in mode 1, linear interpolation is used to obtain  $x_3$  and iteration is employed to obtain the common value of the  $m-ic+1$  smallest critical values. `seq.f95` uses the procedures of mode 1 to calculate the remaining  $ic-1$  critical values.

### 3.3. Mode 4

Calculating all of the critical values for Somerville’s FDR procedure would present an almost insurmountable task when the number of hypotheses to be simultaneously tested is very large. The object of mode 4 is, given the values of all of the test statistics (or their  $p$  values), is to estimate an upper bound to the number of hypotheses that the procedure would reject. Having this upper bound, a user could use an  $s$ -step version of the procedure. (The value of  $s$  would be equal to the upper bound). This would require the calculation of only  $s$  critical values, which should then present a manageable task. The program is designed so that using mode 4 estimates both an upper bound to the number of test statistics to be rejected, and also calculates the needed  $s$  critical values for a corresponding  $s$ -step procedure.

Starting with  $i = 1$ , the program assumes that MCV equals the  $i^{\text{th}}$  largest test statistic and estimates the probability that at least  $i$  test statistics will each be greater than or equal to MCV. The program finds the largest value of  $i$  ( $ub$ ) such that the probability is less than or equal to  $q$ . Mode 3 can be used to calculate the largest  $s$  ( $= ub$ ) critical values. The values of `nCV` and `nCVend` in line 2 are not used in the program. For a quick estimate, the value of  $n$  need not be large, with  $n = 1000$  usually adequate.

### 3.4. Mode 5

`seq.f95` converts the  $p$  values to test statistic values and mode 4 is used.

## 4. Errors in the estimates

The iterative procedure is complicated by the fact that the FDR calculations are Monte Carlo based. To reduce the effect of such errors, the same random number seeds are used for each FDR calculation. For fixed critical values, the error in the calculated value of FDR can be approximated by  $\sqrt{(FDR(1-FDR)/n)}$ . Numerous calculation to estimate the partial

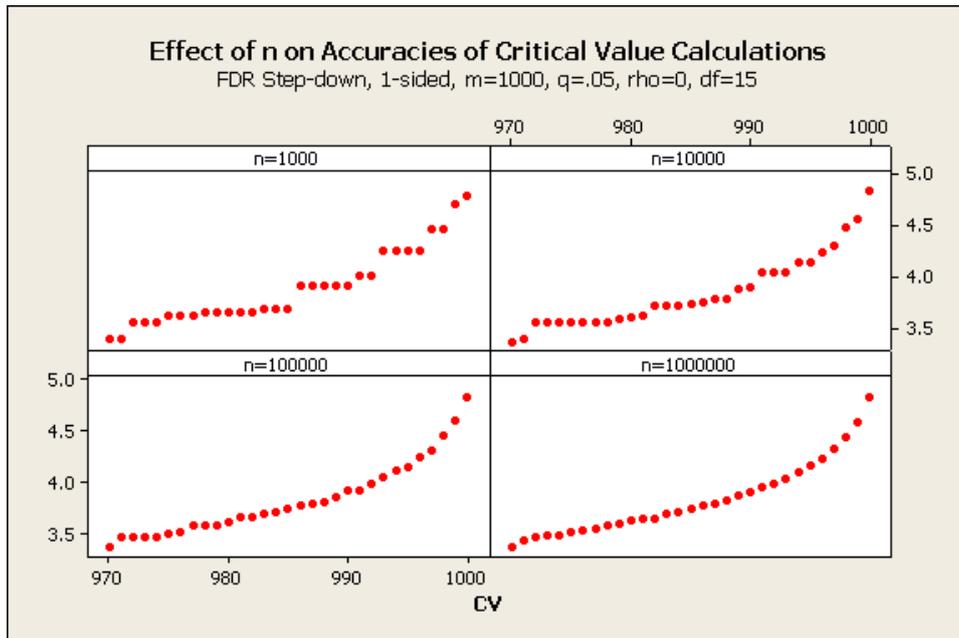


Figure 1: Effect of  $n$  on accuracies of critical value calculations

derivative of  $d_j$  with respect to FDR indicate that for a change of  $\epsilon$  in FDR, changes in the value of  $d_j$  from  $7\epsilon$  to  $40\epsilon$  may be required and the change in  $d_j$  is a function of  $m$ ,  $q$ , and the value of  $j$ . The fortunate aspect is that modest errors in the estimation of a critical value may result in much smaller errors in the achieved False Discovery Rate. Since calculation of each critical value is a function of all previously calculated critical values, there is also a self-correcting effect, which may be enhanced by smoothing the final set of critical values. Values of  $n$  greater than 10,000 are recommended when calculating critical values. Figure 1 illustrates the effect on the accuracy of the critical values as  $n$  increases. Values used are step-down FDR, 1-sided hypotheses with  $m = 1000$ ,  $\rho = 0$ ,  $q = .05$ , and four values for  $n$ , namely  $10^3$ ,  $10^4$ ,  $10^5$  and  $10^6$ , and using exactly 31 “unique” critical values.

## 5. Examples

The program was used in mode 4 to obtain an upper bound to the number of rejected hypotheses for the data of Hedenfalk *et al.* (2001). Using  $n = 10,000$  or  $100,000$ ,  $q = .05$ ,  $\rho = 0$ ,  $df = \infty$ , 2-sided hypotheses, `seq.f95` has given an upper bound to the number of hypotheses to be rejected as 153.

An example for each of the four modes is given. The input file `crit.in` is given for each, as are the two output files `crit.out` and `seq.out`.

## 6. Summary and conclusions

Using the program `seq.f95` to obtain a specified limited number of critical values for the FDR procedure is simple and efficient and the results can be used to control the number

and proportion of false positives. The FDR procedure requires that the test statistics have a multivariate  $t$  distribution with a common correlation. Limited studies suggest that the procedure has robust qualities with regard to the assumption of the multivariate  $t$  distribution, and is conservative when correlations are underestimated. It is not difficult to show that  $d_1$  and  $d_m$ , the smallest and largest critical values are decreasing functions of  $\rho$ , the common correlation coefficient.

## Acknowledgments

The author is indebted to the referees for helpful comments and suggestions.

## References

- Benjamini Y, Hochberg Y (1995). “Controlling the False Discovery Rate, a Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society B*, **57**, 289–300.
- Benjamini Y, Lui W (1999). “A Step-down Multiple Hypotheses Testing Procedure that Controls the False Discovery Rate under Independence.” *Journal of Statistical Planning and Inference*, **82**, 163–170.
- Dudoit S, van der Laan MJ, Birkner MD (2004). “Multiple Testing Procedures for Controlling Tail Probability Error Rates.” *Working Paper 166*, University of California, Berkeley, Division of Biostatistics, Working Paper Series. URL <http://www.bepress.com/ucbbiostat/paper166/>.
- Hedenfalk I, *et al.* (2001). “Gene Expression Profiles in Hereditary Breast Cancer.” *The New England Journal of Medicine*, **344**, 539–548.
- Horn M, Dunnett CW (2004). “Power and Sample Size Comparisons of Stepwise FWE and FDR Controlling Test Procedures in the Normal Many-One Case.” In “Lecture Notes Monograph Series,” volume 47, pp. 48–64. Institute of Mathematical Statistics.
- Lehmann EL, Romano JP (2005). “Generalizations of the Familywise Error Rate.” *The Annals of Statistics*, **29**, 1138–1154.
- Sarkar SK (2002). “Some Results on False Discovery Rate in Stepwise Multiple Testing.” *The Annals of Statistics*, **30**, 239–257.
- Somerville PN (2003). “Optimum FDR Procedures and MCV Values.” *Technical Report TR-03-01*, Department of Statistics, University of Central Florida.
- Somerville PN (2004). “FDR Step-down and Step-up Procedures for the Correlated Case.” In “Lecture Notes Monograph Series,” volume 47, pp. 100–118. Institute of Mathematical Statistics.
- Somerville PN, Hemmelmann C (2007). “Step-up and Step-down Procedures Controlling the Number and Proportion of False Positives.” *Computational Statistics & Data Analysis*, **51**.

Troendle JF (2000). "Stepwise Normal Theory Multiple Test Procedures Controlling the False Discovery Rate." *Journal of Statistical Planning and Inference*, **84**, 139–158.

van der Laan MJ, Dudoit S, Pollard KS (2004). "Augmentation Procedures for Control of the Generalized Familywise Error Rate and Tail Probabilities for the Proportion of False Positives." *Statistical Applications in Genetics and Molecular Biology*, **3**, 1–25.

## A. User manual

### A.1. Capabilities of the program

Some of the capabilities are:

- The program can be used to calculate all or a specified number of the largest critical values needed to simultaneously test  $m$  null hypotheses. The hypotheses may all be one-sided or all two-sided.
- If all of the values of the test statistics (or all of the  $p$  values) are available, the program can be used to estimate an upper bound to the number of hypotheses which will be rejected, and a corresponding MCV will be given. The program can then be used to calculate that number of critical values or the user can specify the MCV. This is an especially useful capability of the program, and eliminates the practically insurmountable task of calculating all of the  $m$  critical values when  $m$  is very large.
- A minimum critical value (MCV) can be specified. That is, the user decides that a hypothesis will never be rejected if the test statistic value is less than MCV. The program will then calculate all of the necessary critical values.
- Enables user to calculate critical values for an implementation of the procedure of [Somerville and Hemmelmann \(2007\)](#) (controlling the number of false positives, given a set of critical values).

In all cases, the user will need to specify all of the following parameters:

- `m` the number of hypotheses,
- `df` the common number of degrees of freedom for all of the test statistics (or  $p$  values),
- `q` the false discovery rate (also often denoted by  $\alpha$ ),
- `upordown` for step-down, a negative integer, for step-up, a positive integer or zero,
- `nbrsided` 1 for one-sided hypotheses, 2 for two-sided hypotheses,
- `rho` common correlation between the test statistics.

In addition, the user will need to specify values for `iseed`, `n`, `nCV` and `nCVend`:

- `iseed` a positive integer for the random number generator, e.g., 757.
- `n` the number of random  $m$ -dimensional vector to be generated. The size of  $n$  determines the accuracy of the critical values generated. A value less than 10,000 is almost never recommended. A value of 1,000,000 will usually be sufficient to produce near 3 decimal accuracy.
- `nCV`, `nCVend`  $d_{nCV}$  is the first value and  $d_{nCV}$  is the last value to be calculated. It is assumed that the values  $d_1$  to  $d_{nCV-1}$  have already been calculated. Note: if the mode is not 1, arbitrary values for `nCV` and `nCVend` may be input.

Three files must exist prior to the running of the program. `crit.in` consists of three lines and contains the necessary input for the program. The two files `crit.out` and `seq.out` will contain the output of the program at the completion of running the program `seq.f95`.

The program has four modes: 1, 3, 4, and 5.

#### *Mode 1*

In this mode, the program calculates all of the critical values from  $d_{nCV}$  to  $d_{nCVend}$ . It is assumed that the values from  $d_1$  to  $d_{nCV-1}$  are known. If all of the critical values are to be calculated, set `nCV` = 1, and set `nCVend` =  $m$ .

This mode can also be used when an MCV is given. Set `nCV` = 2 and `nCVend` =  $m$ , and in line 3, make both the first and second values equal to MCV. This could be very inefficient if  $m$  is large unless a small value of  $n$  can be used. If one is able to make a close underestimate of the number critical values which will be equal to MCV (say  $ub$ ), then setting `nCV` equal to  $ub$  and `nCVend` equal to  $m$ , and inserting  $ub$  consecutive values of MCV in line 3 of `crit.in` will be much more efficient. A method to make an estimate of  $ub$  might be to make several runs using guessed values for  $ub$ , using a smaller value for  $n$  (say 1000). See also the note in Mode 3.

#### *Mode 3*

This mode is used when only the largest critical values need to be calculated. (See modes 4 and 5.) When using mode 3, the values for `nCV` and `nCVend` are ignored. If  $ub$  is the number of critical values to be calculated, line 3 should contain only the value  $ub$ . Note: The smallest critical value for an  $s$ -step procedure is also the MCV for the procedure.

#### *Mode 4*

If the set of  $m$  test statistic values are given, the program efficiently calculates an upper bound (say  $ub$ ) to the number of hypotheses that will be rejected, and a corresponding MCV. Mode 3 can then be used to calculate only the largest  $ub$  critical values. The remaining critical values can then be set equal to the smallest of the calculated critical values, constituting the set of critical values needed for an  $s$ -step FDR procedure.

#### *Mode 5*

Given the  $p$  values corresponding to the  $m$  hypotheses, the  $p$  values are converted to test statistic values, and the methodology used in mode 4 is used to determine an upper bound to the number of hypotheses which will be rejected.

#### *Contents of the input file crit.in*

`crit.in` contains three lines. The first line contains the digit 1, 3, 4 or 5, corresponding to the desired mode. The second line contains ten elements: `n`, `iseed`, `m`, `nCV`, `nCVend`, `df`, `q`, `nbrsided`, `upordown`, `rho`. Note that unless the mode is 1, arbitrary integers may be used for `nCV` and `nCVend`.

If the mode is 1, the third line contains the previously obtained values for  $d_1$  to  $d_{nCV-1}$ , followed by an estimate for  $d_{nCV}$ . If no values have been previously obtained, the line contains

an estimate for  $d_1$ . If the mode is 3, the line contains only the integer giving the number of critical values to be obtained. If the mode is 4 or 5, the line contains exactly  $m$  values. If the mode is 4, the elements are the values of the test statistics (in non-decreasing order). If the mode is 5, the elements are the  $p$  values corresponding to the hypotheses (in non-increasing order).

#### *Contents of the output file seq.out*

If the mode was 1 or 3, `seq.out` contains the requested calculated critical values and their critical value number `nCV`. If the mode is 4 or 5, `seq.out` gives an upper bound to the number of hypotheses that may be rejected (say  $ub$ ). It gives also the  $ub$ -th largest test statistic which is also a corresponding MCV.

#### *Contents of the output file crit.out*

This file contains some of the results of intermediate calculations. It may be of interest to some users.

### **A.2. Calculating all of the critical values**

If 20 hypotheses to be simultaneously tested with the expected value of the FDR to be controlled at the level .05, the hypotheses are all two-sided, the degrees of freedom for each of the test statistics is 19, and the test statistics have a common correlation of zero. For step-down FDR, the following 3 lines in `crit.in` can be used:

```
1
10000 757 20 1 20 19 .05 2 -3 0.0
```

1.024 (This number is your estimate of the value of  $d_1$  and could be set equal to zero.)

See also Example 4.1 in `seq.f95`.

### **A.3. Determining an upper bound for the number of rejected hypotheses**

Suppose the values of the test statistics are as follows (arranged in non-descending order of magnitude): 0.74 1.01 1.30 1.42 1.42 1.50 1.60 1.73 1.82 1.82 1.95 2.04 2.42 2.61 3.02 3.15 4.02 4.32 5.12 5.23

With the expected value of FDR to be controlled at the level .05, two-sided hypotheses, 19 degrees of freedom and a common correlation of zero and step-down FDR, the following 3 lines in `crit.in` can be used:

```
4
10000 757 20 17 20 19 .05 2 -3 0.0

0.74 1.01 1.30 1.42 1.42 1.50 1.60 1.73 1.82 1.82 1.95 2.04 2.42 2.61 3.02
3.15 4.02 4.32 5.12 5.23
```

The program will then give 8 hypotheses as the upper bound (corresponding to an MCV of 2.42). See also Example 4.3 in `seq.f95`.

Using the  $p$  values corresponding to the above test statistic values, the following 3 lines can be used in `crit.in`:

```
5
10000 757 20 17 20 19 .05 2 -3 0.0
0.468346 0.325186 0.209152 0.171809 0.171809 0.150049 0.126095
0.099842 0.084551 0.084551 0.066089 0.055499 0.025711 0.017214
0.007043 0.005273 0.000732 0.000369 0.000061 0.000048
```

#### A.4. Calculating the $s$ largest critical values

For  $m = 1000$ , I would like to calculate the 10 largest critical value (use a 10-step procedure). I wish to use 2 sided FDR, assume a common correlation of 0.0. Each test statistic has 20 `df`. Arbitrary values may be given for `nCV` and `nCVend`. The following may be used as input for `crit.in`.

```
3
10000 757 1000 1 44 20 .05 2 -3 0.0
10
```

See also Example 4.2 in `seq.f95`.

#### A.5. Specifying a minimum critical value

For any set of parameters,  $m$ ,  $\rho$ , `upordown`, `df`, `nbrsided`,  $q$  there is a relationship between an MCV and the number of "unique" critical values (number of steps in the corresponding  $s$ -step procedure). Knowing the approximate relationship may assist in determining whether to use a "unique" number of critical values or choose an MCV. A quick way to explore this is to choose several numbers of "unique" critical values and find the corresponding MCVs. If  $m = 5000$ , `df` = 30,  $q = .05$ ,  $\rho = 0$ , and 1-sided step-down, choosing 20 "unique" values (a small value for  $n$  will be sufficient here, and values for `nCV` and `nCVend` are arbitrary), `crit.in` could be the following:

```
3
1000 757 5000 4 7 30 .05 1 -3 0.0
20
```

Note that after the program calculates the first critical value, the following statement occurs: `MCV = 3.887 for exactly 20 "unique" critical values`. Abort the program (using `CTRL BREAK`) and replace 20 in line 3 with another number, and repeat the process. Using

successively the values 10, 20, 40, 60, 80, 100, 500, one obtains corresponding MCV's of 4.111, 3.887, 3.620, 3.501, 3.398, 3.316, 2.656.

## A.6. Implementing the procedure of Somerville and Hemmelmann

Somerville and Hemmelmann (2007) have shown that the number of false discoveries can be controlled by limiting the number of steps in a step-down procedure.

Suppose there are 5000 test statistics each with 22 *df*. A step-down procedure is desired such that the probability of more than 10 false discoveries is less than .05 (approximately). Using the FDR step-down procedure of Somerville (2004) what is the maximum number of steps that can be used?

Assuming the test statistics have a multivariate *t* distribution with a common correlation of zero, the formulas for the number of steps given by Somerville and Hemmelmann (2007) and interpolation can be used. If *k* is the number of false discoveries:

$$\text{for } df = 15: \text{ Max Steps} = -0.24923 + 3.85682k - .000425084k^2 \text{ (use for } k \leq 33).$$

$$\text{for } df = \infty : \text{ Max Steps} = -5.0138 + 10.1836k + 0.3642844k^2 \text{ (use for } k \leq 19).$$

Using *k* = 10, the maximum number of steps would be 38.276 and 133.25 for *df* = 15 and *df* =  $\infty$ , respectively. Using linear interpolation in  $1/df$ , the number of steps for *df* = 22 would be 62.02. Similarly, for *k* = 19, the number of steps would be 134.65. That is, if *U* is the number of false discoveries:

$$P[U \leq 10] \leq .05 \text{ if no more than 62 steps can be used.}$$

$$P[U \leq 19] \leq .05 \text{ if no more than 134 steps can be used.}$$

The program `seq` can be used to calculate the necessary critical values. For higher common correlations between the test statistics, the Fortran 95 program `FDRPWR` could be used to find the maximum number of steps that could be used and still meet the stated requirements. See Somerville and Hemmelmann (2007).

## A.7. Special conditions for step-up FDR

Somerville (2004) noted that using his step-up FDR procedure, an MCV was sometimes necessary. Using `seq` for small values of *m*, even for large values for *n*, the following message may occur if an attempt is made to obtain all of the critical values: `Unable to get large enough d value. Try larger values for n. For step-up FDR, a larger value of MCV may be required.` A practical solution, and an efficient procedure, is to use mode 3, and obtain only the largest values (e.g. 20, 50 or 100).

## A.8. An additional capability when using modes 4 or 5

Examining the file `seq.out`, where the output for Example 4 (Hedenfalk data) is given, the following statement appears: `As many as 153 hypotheses may be rejected (FDR level .05).` It should be noted that, examining the FDR column, additional statements may be made: `As many as 54 hypotheses may be rejected (FDR level .01).` `As many as 28 hypotheses may be rejected (FDR level .001).`

**Affiliation:**

Paul N. Somerville  
Department of Statistics and Actuarial Science  
University of Central Florida  
Orlando, Florida, United States of America  
E-mail: [somervil@pegasus.cc.ucf.edu](mailto:somervil@pegasus.cc.ucf.edu)  
URL: <http://pegasus.cc.ucf.edu/~somervil/>