SMR: An R Package for Computing the Externally Studentized Normal Midrange Distribution

by Ben Dêivide Oliveira Batista and Daniel Furtado Ferreira

Abstract The main purpose of this paper is to present the main algorithms underlining the construction and implementation of the **SMR** package, whose aim is to compute studentized normal midrange distribution. Details on the externally studentized normal midrange and standardized normal midrange distributions are also given. The package follows the same structure as the probability functions implemented in R. That is: the probability density function (dSMR), the cumulative distribution function (pSMR), the quantile function (qSMR) and the random number generating function (rSMR). Pseudocode and illustrative examples of how to use the package are presented.

Introduction

The SMR package was created to provide an infrastructure for the studentized midrange distribution. This is a new distribution that was inspired by the externally studentized range distribution, which has been largely applied in multiple comparison procedures to identify the best treatment level and has been extensively studied theoretically. Several algorithms to compute the probability density, cumulative distribution, and quantile functions were published by Lund and Lund (1983) and Copenhaver and Holland (1988). Recently, Batista and Ferreira (2014) developed the theory of the externally studentized normal midrange distribution, which is new in the scientific literature to the best knowledge of the authors. The cumulative distribution, the probability density, and the quantile functions were obtained by them analytically.

Computations of the required multidimensional integrations should be done numerically. Therefore, Batista and Ferreira (2014) applied Gaussian quadrature for this task. In particular, they chose the Gauss-Legendre quadrature for solving numerical integrations, because it obtains more accurate results when compared with other Gaussian quadrature methods. The quantile function of the externally studentized normal midrange was computed by the Newton-Raphson method. Based on these numerical methods, the **SMR** package was built and released.

The package name was chosen to identify the Studentized MidRange distribution. The package follows the same structure as the probability functions implemented in R. The following functions were implemented in the package: the probability density function (dSMR), the cumulative distribution function (pSMR), the quantile function (qSMR), and the random number generating function (rSMR).

Therefore, the main purpose of this paper is to present the main algorithms underlining the construction and implementation of the **SMR** package, showing pseudocode of its functions and providing the fundamental ideas for the appropriate use of the package through illustrative examples.

First, details on the externally studentized normal midrange and standardized normal midrange distributions are given. Second, the algorithms for the construction of the package and their respective pseudocodes are presented. Third, details of the **SMR** package functions are showed. Finally, illustrative examples of the package are presented.

The externally studentized normal midrange distribution

Let $X_{1,n}$, $X_{2,n}$, ..., $X_{n,n}$ be the order statistics of a random sample X_1, X_2, ..., X_n of size n from a population with cumulative distribution function (c.d.f.) $F(x)$ and probability density function (p.d.f.) $f(x)$ arranged in order of magnitude of values.

The midrange R for a sample X_1, X_2, ..., X_n is defined as

$$R \quad = \quad (X_{1,n} + X_{n,n})/2,$$

see, e.g., Rider (1957).

The p.d.f. and c.d.f. of R are

$$f_R(r) = \int_{-\infty}^{r} 2n(n-1)f(y)f(2r-y)[F(2r-y) - F(y)]^{n-2}dy \tag{1}$$

and

$$F_R(r) = n \int_{-\infty}^{r} f(y)[F(2r-y) - F(y)]^{n-1} dy, \tag{2}$$

see, e.g., David and Nagaraja (2003).

The studentized midrange Q for a sample $X_1, X_2, \ldots, X_n$ is defined as

$$Q \quad = \quad \frac{R/\sigma}{S/\sigma} = \frac{R}{S} = \frac{W}{X},$$

see, e.g., Batista (2012). When S and R are independent, Q corresponds to the externally studentized midrange, otherwise Q is the internally studentized midrange. The former occurs for example, when R is computed from one normal random sample of size n, and S is obtained from another random sample of size m. It also occurs when R/S is computed from the treatment means of an experimental factor with n levels and $S = \sqrt{(QME)}$, where QME is the associated experimental mean square error with $m-1$ degrees of freedom (Searle, 1987). When considering the externally studentized midrange, it is relevant to compute Q as the ratio of $W = R/\sigma$ the standardized midrange distribution and $X = S/\sigma$.

Considering the particular case of the normal distribution with mean $\mu = 0$, without loss of generality, and variance σ^2 the p.d.f. of $X = S/\sigma$ is given by (Newman, 1939)

$$f(x; \nu) = \frac{\nu^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2-1}} x^{\nu-1} e^{-\nu x^2/2}, x \geq 0. \tag{3}$$

An important distribution for this study, well documented in Batista (2012), Gumbel (1958) and Pillai (1950), is the standardized normal midrange distribution, defined by $W = R/\sigma$. The p.d.f. of W is

$$f_W(w) = \int_{-\infty}^{w} 2n(n-1)\phi(y)\phi(2w-y)[\Phi(2w-y) - \Phi(y)]^{n-2} dy. \tag{4}$$

and the c.d.f. is

$$F_W(w) = \int_{-\infty}^{w} n\phi(y)[\Phi(2w-y) - \Phi(y)]^{n-1} dy, \tag{5}$$

both results found in David and Nagaraja (2003), where $\Phi(.)$ and $\phi(.)$ are the c.d.f. and p.d.f. of the standard normal distribution $N(0,1)$, respectively.

Therefore, according to Batista and Ferreira (2014), the p.d.f. and c.d.f. of Q, in the particular case of the normal distribution with mean $\mu = 0$ and variance σ^2, are respectively

$$f(q; n, \nu) = \int_{0}^{\infty} \int_{-\infty}^{xq} 2n(n-1)x\phi(y)\phi(2xq-y)[\Phi(2xq-y) - \Phi(y)]^{n-2} f(x; \nu) dy dx, \tag{6}$$

and

$$F(q; n, \nu) = \int_{0}^{\infty} \int_{-\infty}^{xq} n\phi(y)[\Phi(2xq-y) - \Phi(y)]^{n-1} f(x; \nu) dy dx, \tag{7}$$

where ν is the number of degrees of freedom.

The p.d.f (4) and c.d.f. (5) are very important to the externally studentized normal midrange algorithms implementation in the **SMR** package.

Algorithms used in the SMR package

The functions implemented in the **SMR** package are dependent on specific functions of R which are: pnorm, to obtain the cumulative distribution function of the standard normal, $\Phi(x)$; dnorm, to obtain the standard normal probability density function, $\phi(x)$; lgamma, to obtain the logarithm of the gamma function. To compute the nodes and weights of the Gauss-Legendre quadrature, an R function based on the method presented by Hildebrand (1974) was implemented. In the following subsections the algorithms and pseudocodes used in the construction of each routine of **SMR** package are presented.

The Gauss-Legendre quadrature and the Newton-Raphson method

The basic idea of Gauss-Legendre quadrature of a function $f(x)$, according to Davis and Rabinowitz (1984), is to write it as follows:

$$\int_{-1}^{1} f(x)dx = \int_{-1}^{1} w(x)g(x)dx \approx \sum_{i=1}^{m} w_i g(x_i), \tag{8}$$

where $f(x) \equiv w(x)g(x)$ and $w(x)$ is the weight function in the Gaussian quadrature, w_i and x_i are the nodes and weights, respectively, in an m-point Gaussian quadrature rule, for $i = 1, 2, \ldots, m$. The weight function is $w(x) = 1$ in the Gauss-Legendre quadrature, thus $f(x) = g(x)$.

Let the symmetric tridiagonal matrix $\mathcal{J}$, with elements $\mathcal{J}_{i,i} = a_{i-1}$, $i = 1, \ldots, m$ and $\mathcal{J}_{i-1,i} = \mathcal{J}_{i,i-1} = \sqrt{\mathcal{J}_{i,i-1}\mathcal{J}_{i-1,i}} = \sqrt{b_{i-1}}$, $i = 2, \ldots, m$, be the Jacobi matrix, given by

$$\mathcal{J} = \begin{pmatrix} a_0 & \sqrt{b_1} & 0 & \ldots & \ldots & \ldots \\ \sqrt{b_1} & a_1 & \sqrt{b_2} & 0 & \ldots & \ldots \\ 0 & \sqrt{b_2} & a_2 & \sqrt{b_3} & 0 & \ldots \\ 0 & \ldots & \ldots & \ldots & \ldots & 0 \\ \ldots & \ldots & 0 & \sqrt{b_{m-2}} & a_{m-2} & \sqrt{b_{m-1}} \\ \ldots & \ldots & \ldots & 0 & \sqrt{b_{m-1}} & a_{m-1} \end{pmatrix},$$

where $a_{i-1} = 0$, for $i = 1, \ldots, m$ and $\sqrt{b_{i-1}} = (i-1)/\sqrt{4(i-1)^2 - 1}$, for $i = 2, \ldots, m$, considering the Gauss-Legendre quadrature.

For computing the quadrature weights and nodes, the eigenvalues and eigenvectors of $\mathcal{J}$ were obtained Hildebrand (1974). The nodes for the Gaussian quadrature were the eigenvalues of this tridiagonal matrix $\mathcal{J}$ and the weights can be computed from the corresponding normalized eigenvectors $\phi^{(i)}$ associated to the eigenvalue x_i. The weight w_i can be computed from the first element $\phi_1^{(i)}$ of the ith corresponding eigenvector by

$$w_i = \mu_0 \left(\phi_1^{(i)} \right)^2,$$

where $\mu_0 = 2$ for the Gauss-Legendre quadrature (Gil et al., 2007).

The set $\{x_i, w_i\}$ is determined such that expression (8) yields an exact result for polynomials of degree $2m - 1$ or less (Chihara, 1978). For non-polynomial functions the Gauss-Legendre quadrature error is defined by

$$\varepsilon_m \approx \left| \sum_{i=1}^{q} w_i g(x_i) - \sum_{i=1}^{m} w_i g(x_i) \right|, \quad q > m. \tag{9}$$

However, the externally studentized normal midrange distribution depends on integrals over infinite intervals. The integral over an infinite range should be changed into an integral over [-1,1] by using the following transformations (Davis and Rabinowitz, 1984)

$$\int_{a}^{\infty} f(y)dy = \int_{-1}^{1} f\left(a + \frac{1+t}{1-t}\right) \frac{2}{(1-t)^2} dt, \tag{10}$$

$$\int_{-\infty}^{b} f(y)dy = \int_{-1}^{1} f\left(b + \frac{1+t}{t-1}\right) \frac{2}{(t-1)^2} dt, \tag{11}$$

$$\int_{a}^{b} f(y)dy = \frac{b-a}{2} \int_{-1}^{1} f\left(\frac{(b-a)}{2}t + \frac{(b+a)}{2}\right) dt. \tag{12}$$

Therefore, the integrals were computed by applying the Gauss-Legendre quadrature rule on these

transformed variables by

$$\int_{a}^{\infty} f(y)dy \approx \sum_{i=1}^{m} w_i \frac{2}{(1-t_i)^2} f\left(a + \frac{1+t_i}{1-t_i}\right), \tag{13}$$

$$\int_{-\infty}^{b} f(y)dy \approx \sum_{i=1}^{m} w_i \frac{2}{(t_i-1)^2} f\left(b + \frac{1+t_i}{t_i-1}\right), \tag{14}$$

$$\int_{a}^{b} f(y)dy \approx \frac{b-a}{2} \sum_{i=1}^{m} w_i f\left(\frac{b-a}{2} t_i + \frac{a+b}{2}\right). \tag{15}$$

For more details, see Olver et al. (2010).

The Newton-Raphson approximation aims to find the roots of a real function, that is,

$$y : f(y) = 0.$$

Thus, iteratively, the Newton-Raphson method follows

$$y_{k+1} = y_k - \frac{f(y_k)}{f'(y_k)}, \quad k = 0, 1, 2, \dots$$

where $f'(y)$ is the first derivative of the function $f(y)$, see, e.g., Miller (2014).

The process starts with an initial arbitrary value y_0, which should be reasonably close to the true root, then the method will usually converge in a few iterations, considering $f'(y_k) \neq 0$.

In this study the main objective was to find the quantile q, given that the cumulative probability is $p, 0 < p < 1$. Hence, considering the cumulative distribution function, $F(q)$, and the probability density, $f(q)$, functions of the random variable Q, the desired solution is

$$q : F(q) - p = 0.$$

The solution is obtained by

$$q_{k+1} = q_k - \frac{F(q_k) - p}{f(q_k)}, \quad k = 0, 1, 2, \dots \tag{16}$$

that should be computed sequentially until a certain convergence criterion is reached.

Standardized normal midrange probability density function

For the standardized normal midrange probability density function (4), the integration interval $(-\infty; q]$ was divided into two subintervals and w was replaced in the algorithm to q, as done before to the cumulative distribution function. Hence, the subintervals $(-\infty, q-8]$ and $[q-8, q]$ were considered. For the first subinterval, the following transformation of variable was used $y = b + (1+t)/(t-1)$, resulting in an integration interval given by $[-1, 1]$, as required to apply the Gauss-Legendre quadrature. With the same purpose, the transformation $z = (a-b)t/2 + (a+b)/2$ was applied to the second subinterval. The algorithm was denoted by dNMR. The approximation of (4) was computed by

$$\begin{aligned} f_Q(q) = &2n(n-1)\left[\int_{-\infty}^{q-8} \phi(y)\phi(2q-y)[\Phi(2q-y) - \Phi(y)]^{n-2}dy \right. \\ &\left. + \int_{q-8}^{q} \phi(y)\phi(2q-y)[\Phi(2q-y) - \Phi(y)]^{n-2}dy \right] \\ \approx &2n(n-1)\left[\sum_{i=1}^{m} \frac{2}{(x_i-1)^2}\phi(y_i)\phi(2q-y_i)[\Phi(2q-y_i) - \Phi(y_i)]^{n-2} \right. \\ &\left. + 4\sum_{i=1}^{m} \phi(z_i)\phi(2q-z_i)[\Phi(2q-z_i) - \Phi(z_i)]^{n-2} \right] \end{aligned} \tag{17}$$

The choice of the value 8 was motivated by the same choice made by Copenhaver and Holland (1988) to the normal range distribution. Since the midrange distribution is more peaked than the range distribution, this value seems to be a good choice for the standardized normal midrange distribution. Further investigations could be undertaken to optimize the choice of these subintervals.

The pseudocode to compute the standardized normal midrange probability density function is

given by:

1. Input q, $n > 1$, $m \geq 2$;

2. Compute m nodes (x_i) and weights (w_i), with $i = 1, 2, \ldots, m$;

3. Transform x_i, from $[-1, 1]$ to $(-\infty, q - 8]$, using:
   $y_i = (q - 8) + (1 - x_i)/(x_i - 1)$, for $i = 1, 2, \ldots, m$;

4. Compute $aux_i = \Phi(2q - y_i) - \Phi(y_i)$, for $i = 1, 2, \ldots, m$;

5. Compute $aux1_i = \phi(y_i)$ and $aux2_i = \phi(2q - y_i)$, for $i = 1, 2, \ldots, m$;

6. If $aux_i \leq 0$, then $aux_i = \epsilon$, for $i = 1, 2, \ldots, m$;

7. Compute $fy_i = \ln(aux1_i) + (n - 1)\ln(aux_i) + \ln(aux2_i) + \ln(2) + \ln(n) + \ln(n - 1)$,
   for $i = 1, 2, \ldots, m$;

8. Apply $fy_i = \exp(fy_i)$, for $i = 1, 2, \ldots, m$;

9. Compute $fy_i = [2/(x_i - 1)^2]fy_i$, for $i = 1, 2, \ldots, m$;

10. Compute $I = \sum\limits_{i=1}^{m} w_i fy_i$;

11. Transform x_i from $[-1, 1]$ to $[q - 8, q]$, using, $a = q - 8$ and $b = q$, where
    $y_i = [(b - a)/2]x_i + (a + b)/2$, for $i = 1, 2, \ldots, m$;

12. Repeat steps 4 to 8, with the y_i from 11, for $i = 1, 2, \ldots, m$;

13. Compute $fy_i = [(b - a)/2]fy_i$, for $i = 1, 2, \ldots, m$;

14. Compute $I = I + \sum\limits_{i=1}^{m} w_i fy_i$;

15. Return I.

The standardized normal midrange cumulative distribution function

The algorithm for computing the cumulative distribution function of standardized normal midrange, expression (5), was developed using the Gauss-Legendre quadrature. This will be essential in the construction of the externally studentized normal midrange. First, the integration interval was divided into two subintervals to achieve higher accuracy as suggested by Quarteroni et al. (2007) and the following division was considered:

$$F_Q(q) = \int_{-\infty}^{q-8} n\phi(y)[\Phi(2q - y) - \Phi(y)]^{n-1}dy + \int_{q-8}^{q} n\phi(y)[\Phi(2q - y) - \Phi(y)]^{n-1}dy. \qquad (18)$$

In these two parts, different variable transformations were considered. In the first part, the transformation $y = (q - 8) + (1 + t)/(t - 1)$ was applied to obtain an integration interval $[-1, 1]$, needed to solve this integral by Gauss-Legendre quadrature. Considering y_i obtained by the above transformation, where $t = x_i$ is the ith Gauss-Legendre quadrature node, $i = 1, 2, \ldots, m$, and m is the number of quadrature points, then the first part of the integral (18) was approximated by using the transformations (11) and (14). Hence,

$$I_1(q) = \int_{-\infty}^{q-8} n\phi(y)[\Phi(2q - y) - \Phi(y)]^{n-1}dy$$

$$\approx n \sum_{i=1}^{m} w_i \frac{2}{(x_i - 1)^2} \phi(x_i)[\Phi(2q - y_i) - \Phi(y_i)]^{n-1}. \qquad (19)$$

For the second part, the transformation $y = (b - a)t/2 + (a + b)/2$ was used, with the same goal presented for the first transformation. These changes were implemented implicitly as a numerical device, as shown above. Therefore, the second part of the integral (18) was approximated by using the transformations (12) and (15). It was given by

$$I_2(q) = \int_{q-8}^{q} n\phi(y)[\Phi(2q - y) - \Phi(y)]^{n-1}dy$$

$$\approx \frac{n(b - a)}{2} \sum_{i=1}^{m} w_i \phi(y_i)[\Phi(2q - y_i) - \Phi(y_i)]^{n-1}, \qquad (20)$$

where $b = q$ and $a = q - 8$.

Considering the expressions (19) and (20) the cumulative distribution function (18) was approximated by

$$F_Q(q) \approx I_1(q) + I_2(q), \tag{21}$$

in the same way as the numerical approximation adopted by McBane (2006) and by Verma and Suarez (2014) for obtaining the probability density and cumulative distribution functions and critical values for the range ratio statistics. The main difference between those works and the present results is the use of the Gauss-Legendre quadrature, besides the distribution functions where the quadrature was applied. All other integrals computed in this work will be based on this type of numerical approximation.

In the algorithm pseudocode, the standardized normal midrange denoted by w, is replaced by q, following the notation of expression (18), avoiding confusion with the quadratures weights. Hence, the pseudocode using the Gauss-Legendre quadrature to obtain the cumulative distribution function of the standardized normal midrange (21), which is denoted by pNMR, is given by:

1. Input $n > 1$, q, and $m \geq 2$;
2. Compute m nodes (x_i) and weights (w_i), with $i = 1, 2, \ldots, m$;
3. Transform x_i, from interval $[-1, 1]$ to $(-\infty, q - 8]$: $y_i = (q - 8) + (1 + x_i)/(x_i - 1)$;
4. Compute $aux1_i = \Phi(y_i)$;
5. Compute $aux2_i = \phi(y_i)$;
6. Compute $aux_i = \Phi(2q - y_i) - aux1_i$;
7. If $aux_i \leq 0$, then $aux_i = \epsilon$;
8. Compute $fy_i = \log(aux2_i) + (n - 1)\log(aux_i)$;
9. Compute $fy_i = \exp(fy_i)$;
10. Compute $fy_i = n(2/(x_i - 1)^2)fy_i$;
11. The quantities in steps 3-10, should be computed for each node x_i, $i = 1, 2, \ldots, m$. Therefore, compute $I = \sum\limits_{i=1}^{m} w_i fy_i$;
12. Set $a = q - 8$ and $b = q$;
13. Transform x_i from interval $[-1, 1]$ to $[q - 8, q]$: $y_i = [(b - a)/2]x_i + (a + b)/2$, for $i = 1, 2, \ldots, m$;
14. Repeat steps 4 to 9, with y_i of step 13, for $i = 1, 2, \ldots, m$;
15. Compute $fy_i = n(b - a)/2 \times fy_i$, for $i = 1, 2, \ldots, m$;
16. Compute $I = I + \sum\limits_{i=1}^{m} w_i fy_i$;
17. Return I.

Standardized normal midrange quantile function

The algorithm qNMR was derived to compute the standardized normal midrange quantile function. The pseudocode makes use of the Newton-Raphson, expression (16), and is given by:

1. Input $n > 1$, $m \geq 2$, and $0 < p < 1$;
2. Set $eps = 1 \times 10^{-13}$, for the error tolerance, and $maxIt = 5000$, for the maximum number of iterations;
3. Get initial estimate of q, called q_0, by:

   3. a. If $p < 0.5$, then $q_0 = -0.5$, else go to step 3b;

   3. b. If $p > 0.5$, then $q_0 = 0.5$, else set $q_0 = 0$;

4. $it = 0$;
5. While ($it \leq maxIt$) do
   $cdf = \text{pNMR}(q_0, n, np)$;
   $pdf = \text{dNMR}(q_0, n, np)$;
   $q_1 = q_0 - (cdf - p)/pdf$;
   go to step 6;
6. If $|q_1 - q_0| \leq eps$, then return q_1 and exit; otherwise, go to step 5, but first update the interactions counter $it = it + 1$ of Newton-Raphson method and do $q_0 = q_1$. In step 5, if this counter exceeds the limit $maxIt$, go to step 7;
7. print the error message: "iterative process did not achieve convergence in $maxIt$ steps."

Externally studentized normal midrange probability density function

For computing the externally studentized normal midrange probability density function, given in (6), note that the innermost integral is the probability density function of the standardized normal midrange, given by (4) and (17). Thus, the dNMR algorithm presented above was reused for computing the probability density function of interest. Also, the variable x (standardized standard deviation) is replaced by s in the implemented algorithm, avoiding confusion with the quadrature nodes. The outermost integral was divided into two parts and a Gauss-Legendre quadrature was also applied in each part. The integration intervals are given by $[0, 1]$ and $[1, \infty)$ for the same reasons mentioned above. Thus,

$$
\begin{aligned}
f(q; n, \nu) &= \int_0^\infty f_Q(sq) f(s; \nu) ds \\
&= \int_0^1 f_Q(sq) f(s; \nu) ds + \int_1^\infty f_Q(sq) f(s; \nu) ds,
\end{aligned}
\tag{22}
$$

where $f(sq)$ was approximated by expression (17), with q substituted by sq, and computed by the algorithm dNMR, described in the previous section.

The innermost integral multiplied by the probability density function of the variable s, $f(s; \nu)$, was computed by

$$
d(q; s, n, \nu) = \mathrm{dNMR}(sq, n, m) \times \frac{\nu^{\nu/2}}{\Gamma(\nu/2) 2^{\nu/2-1}} s^{\nu-1} e^{-\nu s^2/2}
\tag{23}
$$

The auxiliary algorithm dSMR_aux was constructed to compute (23). Its pseudocode is given by:

1. Input s (quadrature node at the ith step), q, $n > 1$ and $m \geq 2$;

2. $Y = \mathrm{dNMR}(s \times q, \ n, \ m)$;

3. Compute

$$
dsln = (nu/2) \ln(nu) - \ln \Gamma(nu/2) - (nu/2 - 1) \ln(2) + (nu - 1) \ln(s) - nu \times s^2/2;
$$

4. Compute $fx = s \times \exp(dsln)$;

5. Compute $d = fx \times Y$;

6. Return d.

Each integration subinterval of (22) was appropriately transformed enabling the Gauss-Legendre quadrature to be applied for approximating the integral (6). For this purpose, in the first integration subinterval, $[0, 1]$, the transformation $y = (a - b)t/2 + (a + b)/2$ was adopted. Hence,

$$
\begin{aligned}
I_1(q; n, \nu) &= \int_0^1 f_Q(sq) f(s; \nu) ds \\
&\approx \sum_{i=1}^m w_i \frac{1}{2} d(q; y_i, n, \nu),
\end{aligned}
\tag{24}
$$

where $d(q; y_i, n, \nu)$ is computed by (23), with corresponding pseudocode given by dSMR_aux.

In the second subinterval of $[1, \infty)$, the transformation (13), with $a = 1$, was applied, being given by $y = 1 + (1 + t)/(1 - t)$. Thus, considering the quadrature nodes x_i's, the integral was approximated by

$$
\begin{aligned}
I_2(q; n, \nu) &= \int_1^\infty f_Q(sq) f(s; \nu) ds \\
&\approx \sum_{i=1}^m w_i \frac{2}{(1 - x_i)^2} d(q; y_i, n, \nu).
\end{aligned}
\tag{25}
$$

The integration (6) was approximated by

$$
f(q; n, \nu) \approx I_1(q; n, \nu) + I_2(q; n, \nu),
\tag{26}
$$

where $I_1(q; n, \nu)$ and $I_2(q; n, \nu)$ were computed by (24) and (25), respectively.

Again, all transformations were implicitly applied as a numerical integration device only. The expression (26) was used to obtain the pseudocode for the computation of the probability density function of the externally studentized normal midrange, denoted dMR. It is given by:

1. Input q, $n > 1$, $nu > 0$ and $m \geq 2$;

2. Compute m nodes (x_i) and weights (w_i), with $i = 1, 2, \ldots, m$;

3. Transform the variable x_i from $[-1, 1]$ to $[0, 1]$, using:
   $y_i = 0,5x_i + 0,5$, for $i = 1, 2, \ldots, m$;

4. Compute $fy_i = \frac{1}{2}\mathtt{dSMR\_aux}(y_i, q, n, m)$, for $i = 1, 2, \ldots, m$;

5. Compute $I = \sum\limits_{i=1}^{m} w_i fy_i$;

6. Transform x_i from $[-1, 1]$ to $[1, \infty]$:
   $y_i = 1 + (1 + x_i)/(1 - x_i)$, for $i = 1, 2, \ldots, m$;

7. Compute $fy_i = \mathtt{dSMR\_aux}(y_i, q, n, m)$, for $i = 1, 2, \ldots, m$;

8. Compute $fy_i = \ln(fy_i) + \ln(2) - 2\ln(1 - x_i)$, for $i = 1, 2, \ldots, m$;

9. Compute $fy_i = \exp(fy_i)$, for $i = 1, 2, \ldots, m$;

10. Compute $I = I + \sum\limits_{i=1}^{m} (fy_i \times w_i)$;

11. Return I.

Quadratures of 64 points are sufficient for most circumstances, although it is possible to increase the number of quadrature points and achieve the desired accuracy. The Newton-Raphson method was used to compute quantiles, since the probability density function, the first derivative, was available. The following subsections describe the algorithms and show their pseudocodes.

The externally studentized normal midrange cumulative distribution function

Note that the innermost integral of the externally studentized normal midrange cumulative distribution function, given in (7), is the cumulative distribution function of standardized normal midrange, given by (5) and (18). Thus, the pNMR algorithm presented above was reused for computing the cumulative distribution function of interest. Also, the variable x (standardized standard deviation) is replaced by s in the implemented algorithm, avoiding confusion with the quadrature nodes. The outermost integral was divided into two parts and a Gauss-Legendre quadrature was also applied in each part. The integration intervals are given by $[0, 1]$ and $[1, \infty)$. The limit 1 was chosen since it is the modal value of the standardized normal standard deviation probability density function $f(x; \nu)$, given in expression (3), following the ideas of Copenhaver and Holland (1988). Thus

$$F(q; n, \nu) = \int_0^\infty F_Q(sq)f(s; \nu)ds$$
$$= \int_0^1 F_Q(sq)f(s; \nu)ds + \int_1^\infty F_Q(sq)f(s; \nu)ds, \qquad (27)$$

where $F_Q(sq)$ was given by expression (18), with q substituted by sq, and computed by the algorithm pNMR, described in the previous section.

The innermost integral multiplied by the probability density function of the variable s, $f(s; \nu)$, was computed by

$$I(q; s, n, \nu) = \mathtt{pNMR}(sq, n, m) \times \frac{\nu^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2-1}} s^{\nu-1}e^{-\nu s^2/2} \qquad (28)$$

Therefore, an auxiliary algorithm, denoted by pNMR_aux, was constructed to compute (28). The pseudocode is given by:

1. Input q, s, $n > 1$, $nu > 0$ and $m \geq 2$;

2. $I = \mathtt{pNMR}(s \times q, n, m)$;

3. Compute

   $fx = (nu/2)\ln(nu) - \ln\Gamma(nu/2) - (nu/2 - 1)\ln(2) + (nu - 1)\ln(s) - s \times nu \times s/2$;

4. Apply $fx = \exp(fx)$;

5. Compute $I = fx \times I$;

6. Return I.

Each subinterval of (27) was appropriately transformed enabling the Gauss-Legendre quadrature to be applied for solving the integral (7). For this purpose, in the first integration subinterval, $[0, 1]$, the transformation $y = (a - b)t/2 + (a + b)/2$, given by (12) and (15), was adopted. Hence, considering all quadrature nodes x_i's, it follows that

$$I_1(q; n, \nu) = \int_0^1 F_Q(sq)f(s; \nu)ds$$

$$\approx \sum_{i=1}^m w_i \frac{1}{2} I(q; y_i, n, \nu), \tag{29}$$

where $I(q; y_i, n, \nu)$ is given by (28).

In the second subinterval of $[1, \infty)$, the transformations (10) and (13), with $a = 1$, was applied and it is given by $y = 1 + (1 + t)/(1 - t)$. Thus, considering the quadrature nodes x_i's, the integral was approximated by

$$I_2(q; n, \nu) = \int_1^\infty F_Q(sq)f(s; \nu)ds$$

$$\approx \sum_{i=1}^m w_i \frac{2}{(1 - x_i)^2} I(q; y_i, n, \nu). \tag{30}$$

The integration (7) was computed by

$$F(q; n, \nu) \approx I_1(q; n, \nu) + I_2(q; n, \nu), \tag{31}$$

using the results of (29) and (30).

All transformations were implicitly applied, i.e., they were only a numerical integration device. The pseudocode for the computation of the cumulative distribution function of the externally studentized normal midrange, denoted pMR, applies the ideas of expression (31) and is given by:

1. Input q, $nu > 0$, $n > 1$, and $m \geq 2$;

2. Compute m nodes (x_i) and weights (w_i), with $i = 1, 2, \ldots, m$;

3. For the first part of the split integral, transform: $y_i = 0, 5x_i + 0, 5$, for $i = 1, 2, \ldots, m$;

4. Compute $fy_i = \frac{1}{2}$pNMR_aux(q, y_i, n, m), for $i = 1, 2, \ldots, m$;

5. Compute $I = \sum_{i=1}^m w_i fy_i$;

6. For the second part of the split integral, transform: $y_i = 1 + (1 + x_i)/(1 - x_i)$, for $i = 1, 2, \ldots, m$;

7. Compute $fy_i = $pNMR_aux$(q, y_i, n, np)$, for $i = 1, 2, \ldots, m$;

8. Compute $fy_i = \ln(fy_i) + \ln(2) - 2\ln(1 - x_i)$, for $i = 1, 2, \ldots, m$;

9. Hence, compute $fy_i = \exp(fy_i)$, for $i = 1, 2, \ldots, m$;

10. Compute $I = I + \sum_{i=1}^m w_i fy_i$;

11. Return I.

Externally studentized normal midrange quantile function

For the externally studentized normal midrange quantile function, the qMR algorithm was constructed. The algorithm applies the Newton-Raphson method and depends on the pMR an dMR methods. Its pseudocode is:

1. Input: $n > 1$, $m \geq 2$, and $0 < p < 1$;

2. Set $eps = 1 \times 10^{-13}$, for the error tolerance, and $maxIt = 5000$, for the maximum number of iterations;

3. Get initial estimate of q, called q_0, by:

   3.a. If $p < 0.5$, then $q_0 = -0.5$, else go to step 3b;

   3.b. If $p > 0.5$, then $q_0 = 0.5$, else set $q_0 = 0$;

4. $it = 0$;

5. While ($it \leq maxIt$) do
   $cdf = \text{pMR}(q_0, n, np)$;
   $pdf = \text{dMR}(q_0, n, np)$;
   $q_1 = q_0 - (cdf - p)/pdf$;
   go to step 6;

6. If $|q_1 - q_0| \leq eps$, then return q_1 and exit; otherwise, go to step 5 but first update the interactions counter $it = it + 1$ of Newton-Raphson method and do $q_0 = q_1$. In step 5, if this counter exceeds the limit $maxIt$, go to step 7;

7. print the error message: "iterative process did not achieve convergence in $maxIt$ steps."

The random number generator

To generate random sample of size N from the normal midrange distribution, the algorithm rMR was constructed, with parameter n and ν. First, a sample from the standard normal distribution of size n, $X_1, X_2, \ldots, X_n$, was simulated, where $X_i \overset{\text{iid}}{\sim} N(0, 1)$. A chi-square variable $U \sim \chi_\nu^2$, with degrees of freedom ν, was also simulated, and the following transformation was computed:

$$Q = \frac{[max(X_i) + min(X_i)]/2}{\sqrt{\frac{U}{\nu}}}.$$

With infinite degrees of freedom, then the following transformation was considered instead

$$W = \frac{max(X_i) + min(X_i)}{2}.$$

The process was repeated N times and the values of Q or W, depending on the considered case, are stored in a vector of size N.

This distribution can be used to obtain cumulative probabilities and quantiles. Thus, for example, given a quantile q or w, the cumulative probability function, ensuring that N is large, is computed in an approximated way by:

$$P(Q \leq q) = \frac{\sum\limits_{i=1}^{N} I(Q_i \leq q)}{N}, \; \nu < \infty$$

$$P(W \leq w) = \frac{\sum\limits_{i=1}^{N} I(W_i \leq w)}{N}, \; \nu = \infty.$$

The random number generator rMR returns the vector of N realizations of Q or W and its pseudocode is presented below. The rMR generator is dependent of the R functions rnorm() and rchisq(), for obtaining a random vector of standard normal independently variables and a chi-square realization, respectively. The matrix(x,p,q) function, presented below, creates a matrix of dimension $p \times q$ filled by columns with the elements given by the vector x.

1. Input: $N > 1$, $n > 1$ and $\nu > 0$;

2. $x = \text{rnorm}(N \times n)$

3. $X = \text{matrix}(x, N, n)$ of dimension $(N \times n)$;

4. if $nu = \infty$, then go to (7), else go to (5);

5. $u = \text{rchisq}(N, \nu)$;

6. $X = X/\sqrt{(u/\nu)}$;

7. for $i = 1$ to N do:
   $\text{rMR}[i] = [max(X[i,]) + min(X[i,])]/2$;

8. return rMR.

Details of the SMR package

The package **SMR** provides the following functions, where np is the number of nodes and weights of the Gauss-Legendre quadrature:

- dSMR(x,size,df,np=32): computes values of the probability density function, given in (6) or (4);
- pSMR(q,size,df,np=32): computes values of the cumulative distribution function, given in (7) or (5);
- qSMR(p,size,df,np=32): computes quantiles of the externally studentized normal midrange;
- rSMR(n,size,df=Inf): drawn a random sample of size *n* from the externally studentized normal midrange.

The value of the argument df can be finite or infinity. If df=Inf, values of the probability density, cumulative distribution and quantile functions of the normal midrange (standardized normal midrange) are computed. If the argument df is not specified in the rSMR function, the default value Inf is used and random samples from the normal midrange distribution are drawn. The other functions presented earlier in the previous section are internal algorithms of the **SMR** package.

As an illustration, consider the following examples:

```
library(SMR)
set.seed(10701)
q    <- 2   # quantile
x    <- 2   # quantile
p    <- 0.9 # probability
n    <- 10  # sample size to be simulated
size <- 5   # normal sample size
df   <- 3   # degrees of freedom
np   <- 32  # number of points of the Gaussian quadrature

dSMR(x, size, df, np) # SMR pdf
[1] 0.01926172

pSMR(q, size, df, np) # SMR cdf
[1] 0.9851739

qSMR(p, size, df, np) # SMR quantile
[1] 0.8350065

rSMR(n, size, df)      # random sample of the SMR distribution
[1]  0.35108979  0.33786356 -0.13753510 -0.58741681 -0.40358907
[6] -0.72528615  0.45845331  0.08906021 -1.64157684  0.07022362
```

In the case $\nu = \infty$, the result of the standardized normal midrange distribution is returned, as illustrated by the following example:

```
library(SMR)
q    <- 2   # quantile
x    <- 2   # quantile
p    <- 0.9 # cumulative probability
n    <- 10  # sample size to be simulated
size <- 5   # normal sample size
df   <- Inf # degrees of freedom
np   <- 32  # number of points of the Gaussian quadrature

dSMR(x, size, df, np) # normal MR pdf
[1] 0.0004487675

pSMR(q, size, df, np) # normal MR cdf
[1] 0.9999408

qSMR(p, size, df, np) # normal MR quantile
[1] 0.6531507

rSMR(n, size, df,)    # random sample of the normal MR distribution
[1] -0.52475079  0.10198842 -0.38647236  0.18939367  0.17756023
[6] -1.03384242  0.35608349  1.00629514  0.06360581  0.70835452
```

A concrete application of the **SMR** package on real dataset is now considered. We use the data on nitrogen contents of red clover plants presented in Steel and Torrie (1980), page 180. A completely randomized design with one factor with 6 levels and $r = 5$ replications were carried out. The mean squared error (MSE) was 11.7887, with $\nu = 24$ degrees of freedom. The factor means is presented in Table 1. A multiple comparison procedure was suggested by Batista (2012). The suggested procedure was similar to the Tukey test (Tukey, 1953). First the least significant difference of the proposed test is computed by

$$\Delta = 2q_{1-\alpha/2;n,\nu}\sqrt{\frac{MSE}{r}},$$

where $q_{1-\alpha/2;n,\nu}$ is the $100\alpha/2\%$ quantile of the externally studentized normal midrange. This value can be computed by using the function qSMR(0.975,6,24,np=32)=1.0049 of the **SMR** package, for the significance level $\alpha = 0.05$.

The Δ value in this case is

$$\Delta = 2 \times 1.0049 \times \sqrt{\frac{11.7887}{5}} = 3.0859.$$

Finally, the test is performed in the same way of the Tukey test. The results of the proposed midrange and Tukey tests are shown in Table 1. The proposed midrange test show no ambiguous results in this example, as happens with the Tukey test (two or more letters per level).

Table 1: The midrange and Tukey multiple comparison procedures applied to the nitrogen contents of 6 red clover plants presented in Steel and Torrie (1980), page 180.

Levels	means	Tukey*	Midrange*
5	13.26	c	d
3	14.64	c	d
6	18.70	cb	c
4	19.92	cb	c
2	23.98	ba	b
1	28.82	a	a

* Means with the same letter are not significantly different.

Concluding remarks

The importance of the externally studentized normal midrange distribution can be enormous in the analysis of experiments, since the midrange estimator is more efficient than the sample mean in platykurtic distributions. Therefore, this distribution could be useful in the proposition of multiple comparison procedures, that could potentially show better results (more robust and powerful) than the traditional tests based on the externally studentized normal range.

This package is easy to use and shows very high accuracy. The accuracy of critical values is estimated through the computation of the difference between two results using different numbers of quadrature points according to the expression (9). The number of quadrature points can be chosen and for most cases, 32 points are enough to achieve high accuracy. Monte Carlo simulations were also used to evaluate the accuracy of the proposed methods and the consistency of the **SMR** package. The algorithms were proposed and implemented using Gauss-Legendre quadrature and the Newton-Raphson method in R software, resulting in the **SMR** package, that is available for download from CRAN. An important aspect is that the **SMR** package is written in R. Code in Fortran or C can make the **SMR** functions faster, a feature that is planned for future releases of the package. Another important aspect is the possibility of using in future releases two-dimensional interpolations when dSMR, pSMR and qSMR are called with long vectors of arguments x, q, and p. In this case, two-dimensional grids for $\{x, q, \text{and } p\}$ and $\{size\}$ can be computed offline and the approx function used to interpolate.

Acknowledgments

Bibliography

B. D. O. Batista. Externally studentized normal midrange exact distribution and development of an R package using gaussian quadrature. Master's thesis, Federal university of Lavras, Brazil, 2012. [p124, 134]

B. D. O. Batista and D. F. Ferreira. *SMR: Externally Studentized Midrange Distribution*, 2014. URL http://CRAN.R-project.org/package=SMR. R package version 2.0.1. [p123, 124]

T. S. Chihara. *An introduction to Orthogonal Polynomials*, volume 13 of *Mathematics and its Applications Series*. Gordon and Breach, New York, 1978. [p125]

M. D. Copenhaver and B. Holland. Computation of the distribution of the maximum studentized range statistic with application to multiple significance testing of simple effects. *Journal of Statistical Computation and Simulation*, 30(1):1–15, 1988. [p123, 126, 130]

H. A. David and H. N. Nagaraja. *Order Statistics*. John Wiley & Sons, Canada, 2003. [p124]

P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration*. Academic Press, New York, 2 edition, 1984. [p125]

A. Gil, J. Segura, and N. Temme. *Numerical Methods for Special Functions*. Society for Industrial and Applied Mathematics, 2007. [p125]

E. J. Gumbel. *Statistics of Extremes*. Columbia University Press, New York, 1958. [p124]

F. B. Hildebrand. *Introduction to Numerical Analysis*. McGraw-Hill, New York, 2 edition, 1974. [p124, 125]

R. E. Lund and J. R. Lund. Algorithm AS 190: Probabilities and upper quantiles for the studentized range. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 32(2):pp. 204–210, 1983. [p123]

G. C. McBane. Programs to compute distribution functions and critical values for extreme value ratios for outlier detection. *Journal of Statistical Software*, 16(3):1–9, 5 2006. [p128]

G. Miller. *Numerical Analysis for Engineers and Scientists*. Cambridge University Press, New York, 2014. [p126]

D. Newman. The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika*, 31(1/2):pp. 20–30, 1939. [p124]

F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010. [p126]

K. C. S. Pillai. On the distribution of midrange and semi-range in samples from a normal population. *Annals of Mathematical Statistics*, 21:100–105, 1950. [p124]

A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*. Texts in Applied Mathematics. Springer, 2007. [p127]

P. R. Rider. The midrange of a sample as an estimator of the population midrange. *J. Amer. Statist. Ass.*, 52(280):537–542, Dec. 1957. [p123]

S. R. Searle. *Linear models for unbalanced data*. J. wiley, New York, 1987. [p124]

R. Steel and J. Torrie. *Principles and Procedures of Statistics: A Biometrical Approach*. McGraw-Hill series in probability and statistics. McGraw-Hill, New York, 2 edition, 1980. [p134]

J. W. Tukey. The problem of multiple comparisons. *Unpublished Dittoed Notes, Princeton University*, 1953. [p134]

M. Verma and M. C. Suarez. Dixontest.criticalvalues: A computer code to calculate critical values for the dixon statistical data treatment approach. *Journal of Statistical Software*, 57(2):1–12, 3 2014. [p128]

Ben Dêivide de Oliveira Batista
Department of Exact Sciences, Federal University of Lavras
37200-000, Lavras, Brazil
URL: http://ufla.br
ben.deivide@gmail.com

Daniel Furtado Ferreira
Department of Exact Sciences, Federal University of Lavras
37200-000, Lavras, Brazil
URL: http://danielfurtado.url.ph
danielff@dex.ufla.br