Discussiones Mathematicae
Probability and Statistics 30 (2010) 53–86

DMPS Page

NEAR-EXACT DISTRIBUTIONS FOR THE GENERALIZED WILKS LAMBDA STATISTIC

Luís M. Grilo

Polytechnic Institute of Tomar, Mathematics Department
Tomar, Portugal

e-mail: lgrilo@ipt.pt

and

Carlos A. Coelho

Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia, Departamento de Matemática, Portugal

e-mail: cmac@fct.unl.pt

Dedicated to Professor J. Tiago Mexia
on his Jubilee

Abstract

Two near-exact distributions for the generalized Wilks Lambda statistic, used to test the independence of several sets of variables with a multivariate normal distribution, are developed for the case where two or more of these sets have an odd number of variables. Using the concept of near-exact distribution and based on a factorization of the exact characteristic function we obtain two approximations, which are very close to the exact distribution but far more manageable. These near-exact distributions equate, by construction, some of the first exact moments and correspond to cumulative distribution functions which are practical to use, allowing for an easy computation of quantiles. We also develop three asymptotic distributions which also equate some of the first exact moments. We assess the proximity of the asymptotic and near-exact distributions obtained to the exact distribution using two measures based on the Berry-Esseen bounds. In our comparative numerical study we consider different numbers of sets of variables, different numbers of variables per set and different sample sizes.

Keywords: independent Beta random variables, characteristic function, sum of Gamma random variables, likelihood ratio test statistic, proximity measures.

2000 Mathematics Subject Classification: 62E20, 62E15, 62E10.

1. Introduction

The generalized Wilks Lambda statistic (Wilks, 1932, 1935) is used in multivariate analysis to test the independence among m sets ($m \geq 2$) of random variables (r.v.'s), under the normality assumption. For the case when there is at most one set with an odd number of variables among the m sets, we have the exact distribution in the form of a Generalized Integer Gamma (GIG) distribution obtained by Coelho (1998), but for the case where at least two sets, among the m sets, have an odd number of variables, we do not have yet an exact distribution in a manageable form, adequate for further manipulation. Although we have, for this general case, some asymptotic distributions (see for example Box (1949) and Anderson (2003)) and some near-exact distributions (Coelho, 2003, 2004), in this paper we develop three asymptotic distributions and two new near-exact distributions, these later ones obtained in a concise and manageable form but nonetheless extremely close to the exact distribution in terms of characteristic function (c.f.), probability density function (p.d.f.), cumulative distribution function (c.d.f.), moments and quantiles.

In order to develop the near-exact distributions we first factor the exact c.f. and then we replace a suitably chosen part of the exact c.f., which corresponds to the c.f. of a Logbeta distribution, by an adequate asymptotic approximation. Depending on the asymptotic result used, one may obtain different near-exact approximations. In one case we replace the c.f. of a Logbeta r.v. by the c.f. of the sum of two Gamma r.v.'s and, in the other case, by the c.f. of a mixture of two Gamma r.v.'s. These distributions match the first three and four exact moments, respectively. By joining this small part with the remaining unchanged part of the original c.f., we get what we call a near-exact c.f. In the first case this c.f. corresponds to a particular Generalized Near-Integer Gamma (GNIG) distribution, while in the second case it corresponds to a mixture of two GNIG distributions. The corresponding near-exact c.d.f.'s are obtained in a concise and manageable form, perfectly handled by a number of available software programs, allowing for the computation of near-exact quantiles.

The concept of near-exact distribution has already been introduced in a number of papers (Coelho, 2003, 2004; Grilo and Coelho, 2007, 2010) and also a similar derivation procedure has already been applied to obtain near-exact distributions for the product of an odd number of particular independent Beta r.v.'s (Grilo and Coelho, 2007). Now, based on a factorization of the exact c.f. of the logarithm of the generalized Wilks Λ statistic, we develop near-exact distributions for this well-known statistic.

Our paper is organized as follows: in Section 2 we present some useful distributions for our work; in Section 3 we develop two near-exact distributions, based on factorizations of the exact c.f., and also three asymptotic distributions for the generalized Wilks Λ statistic. In Section 4, we use two measures based on the Berry-Esseen bounds to assess the behavior of the near-exact and asymptotic distributions proposed and also to compare them with a rather well-known asymptotic distribution (Box, 1949; Anderson, 2003) and with another near-exact distribution (Coelho, 2004). In Section 5, we provide some conclusions and final remarks.

2. Some distributions used in the paper

Since some of our near-exact and asymptotic distributions are GNIG distributions or finite mixtures of GNIG distributions we now introduce this distribution along with the useful Logbeta distribution.

Let Z be a r.v. with a GIG (Generalized Integer Gamma) distribution of depth g (Coelho, 1998), with shape parameters $r_1, \ldots, r_g \in \mathbb{N}$ (where $\mathbb{N}$ is the set of positive integers) and all different rate parameters $\lambda_1, \ldots, \lambda_g \in \mathbb{R}^+$(being $\mathbb{R}^+$ the set of positive reals). We will denote this fact by

$$Z \sim GIG(r_1, \ldots, r_g; \lambda_1, \ldots, \lambda_g)\,.$$

The p.d.f. of Z is given by

$$f_Z(z) = K \sum_{i=1}^{g} P_i(z) \mathrm{e}^{-\lambda_i z}\,, \qquad (z > 0),$$

where

$$K = \prod_{i=1}^{g} \lambda_i^{r_i} \tag{1}$$

and $P_i(z)$ is a polynomial of degree $r_i - 1$ in z, which may be written as

$$P_i(z) = \sum_{k=1}^{r_i} c_{i,k} z^{k-1},$$

where

$$c_{i,r_i} = \frac{1}{(r_i-1)!} \prod_{\substack{j=1 \\ j\neq i}}^{g} (\lambda_j - \lambda_i)^{-r_j} \tag{2}$$

and, for $k = 1, \ldots, r_i - 1$,

$$c_{i,r_i-k} = \frac{1}{k} \sum_{j=1}^{k} \frac{(r_i - k + j - 1)!}{(r_i - k - 1)!} R(j-1, i) c_{i,r_i-(k-j)}\,, \tag{3}$$

where

$$R(n,j) = \sum_{\substack{i=1 \\ i\neq j}}^{g} r_i (\lambda_j - \lambda_i)^{-n-1}, \quad (n = 0, \ldots, r_i - 1)\,. \tag{4}$$

The c.d.f. of Z is given by

$$F_Z(z) = K \sum_{i=1}^{g} P_i^*(z)\,, \qquad (z > 0)$$

with K given by (**??**) and where

$$P_i^*(z) = \sum_{k=1}^{r_i} c_{i,k} \frac{(k-1)!}{\lambda_i^k} \left\{ 1 - \left(\sum_{j=0}^{k-1} \frac{\lambda_i^j z^j}{j!} \right) \mathrm{e}^{-\lambda_i z} \right\}$$

with $c_{i,k}$ $(i = 1, \ldots, g; k = 1, \ldots, r_i)$ given by (**??**) through (**??**).

Now, let us consider $Z \sim GIG(r_1, \ldots, r_g; \lambda_1, \ldots, \lambda_g)$ and $X \sim G(r, \lambda)$, two independent r.v.'s with $r \in \mathbb{R}^+ \backslash \mathbb{N}$ and $\lambda \neq \lambda_j$, $\forall j \in \{j = 1, \ldots, g\}$. Then the r.v. $W = Z + X$ has a GNIG (Generalized Near-Integer Gamma) distribution with depth $g + 1$ (Coelho, 2004). Symbolically,

$$W \sim GNIG(r_1, \ldots, r_g, r; \lambda_1, \ldots, \lambda_g, \lambda)\,. \tag{5}$$

The p.d.f. of W is given by
(6)

$$f_W(w) = K\lambda^r \sum_{j=1}^{g} \mathrm{e}^{-\lambda_j w} \sum_{k=1}^{r_j} \left\{ c_{j,k} \frac{\Gamma(k)}{\Gamma(k+r)} w^{k+r-1} {}_1F_1(r, k+r, -(\lambda-\lambda_j)w) \right\}, \quad (w>0)$$

and the c.d.f. by

$$\begin{aligned} F_W(w) &= \lambda^r \frac{w^r}{\Gamma(r+1)} {}_1F_1(r, r+1, -\lambda w) \\ &\quad -K\lambda^r \sum_{j=1}^{g} \mathrm{e}^{-\lambda_j w} \sum_{k=1}^{r_j} c^*_{j,k} \sum_{i=0}^{k-1} \frac{w^{r+i}\lambda_j^i}{\Gamma(r+1+i)} \\ &\qquad {}_1F_1(r, r+1+i, -(\lambda-\lambda_j)w), (w>0), \end{aligned} \tag{7}$$

where

$$K = \prod_{j=1}^{g} \lambda_j^{r_j} \qquad \text{and} \qquad c^*_{jk} = \frac{c_{jk}}{\lambda_j^k}\Gamma(k)$$

with $c_{j,k}$ given by (**??**) through (**??**). In the above expressions

$$\begin{aligned} {}_1F_1(a,b,z) &= \frac{\Gamma(b)}{\Gamma(a)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)}{\Gamma(b+j)} \frac{z^j}{j!} \\ &= \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 \mathrm{e}^{zt} t^{a-1} (1-t)^{b-a-1} dt, \qquad (a \neq b), \end{aligned}$$

is the Kummer confluent hypergeometric function (Abramowitz and Stegun, 1974) which has good convergence properties and nowadays it can be found in a number of software packages, such as Mathematica.

The c.f. of W in (**??**) is given by

$$\phi_W(t) = \lambda^r (\lambda - \mathrm{i}t)^{-r} \prod_{j=1}^{g} \lambda_j^{r_j} (\lambda_j - \mathrm{i}t)^{-r_j}, \tag{8}$$

where $r \in \mathbb{R}^+ \backslash \mathbb{N}$, $\lambda \in \mathbb{R}^+$, $r_j \in \mathbb{N}$ and $\lambda \neq \lambda_j$, $\forall j \in \{1, \ldots, g\}$. If $r \in \mathbb{N}$ then the GNIG distribution of depth $g+1$ reduces to a GIG distribution of depth $g+1$. That is, the GIG distribution is a particular case of the GNIG distribution.

If the r.v. W has a distribution that is a mixture, with k components, of GNIG distributions, the j-th component with weight π_j and depth g_j, we will denote this fact by

$$W \sim MkGNIG$$
$$(\pi_1; r_{11}, \ldots, r_{g_1 1}; \lambda_{11}, \ldots, \lambda_{g_1 1} | \ldots | \pi_k; r_{1k}, \ldots, r_{g_k k}; \lambda_{1k}, \ldots, \lambda_{g_k k})\,.$$

If X is a r.v. with a Beta distribution, with parameters $\alpha > 0$ and $\beta > 0$, symbolically

$$X \sim Beta(\alpha, \beta),$$

then the h-th moment of X is given by

$$E(X^h) = \frac{B(\alpha+h, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+h)}{\Gamma(\alpha+\beta+h)}\,, \qquad (h > -\alpha)\,. \tag{9}$$

If $Y = -\ln X$ then Y is a r.v. with a Logbeta distribution with parameters α and β (Johnson *et al.*, 1995), denoted by

$$Y \sim Logbeta(\alpha, \beta)\,.$$

The p.d.f. of Y is

$$f_Y(y) = \frac{1}{B(\alpha, \beta)} \mathrm{e}^{-\alpha y} (1 - \mathrm{e}^{-y})^{\beta-1}\,, \qquad (y > 0)\,.$$

Since the Gamma functions in (**??**) are still defined for h complex (in strict sense), the c.f. of Y is given by

$$\phi_Y(t) = E(\mathrm{e}^{\mathrm{i}tY}) = E(\mathrm{e}^{-\mathrm{i}t \ln X}) = E(X^{-\mathrm{i}t}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha-\mathrm{i}t)}{\Gamma(\alpha+\beta-\mathrm{i}t)}, \tag{10}$$

where $\mathrm{i} = (-1)^{1/2}$ and $t \in \mathbb{R}$ (being $\mathbb{R}$ the set of real numbers). Through (**??**) we know that, if $E(|Y^h|) < \infty$ then

$$E(Y^h) = \frac{1}{\mathrm{i}^h} \frac{d^h}{dt^h} \phi_Y(t) \bigg|_{t=0}\,, \quad (h \in \mathbb{N}),$$

and thus we can get expressions for some of the first moments, μ'_h, for the r.v. Y.

For example, the expressions of the first four moments are given by

$$
\begin{aligned}
\mu_1' &= E(Y) = \psi(\alpha+\beta) - \psi(\alpha) \\
\mu_2' &= E(Y^2) = \psi'(\alpha) - \psi'(\alpha+\beta) + [\psi(\alpha+\beta) - \psi(\alpha)]^2, \\
\mu_3' &= E(Y^3) = \psi''(\alpha+\beta) - \psi''(\alpha) + [\psi(\alpha+\beta) - \psi(\alpha)]^3 \\
&\qquad\qquad +3\,[\psi(\alpha+\beta) - \psi(\alpha)]\left[\psi'(\alpha) - \psi'(\alpha+\beta)\right], \\
\mu_4' &= E(Y^4) = \psi'''(\alpha) - \psi'''(\alpha+\beta) + [\psi(\alpha) - \psi(\alpha+\beta)]^4 \\
&\qquad\qquad +6\,[\psi(\alpha) - \psi(\alpha+\beta)]^2\left[\psi'(\alpha) - \psi'(\alpha+\beta)\right] \\
&\qquad\qquad +3\left[\psi'(\alpha) - \psi'(\alpha+\beta)\right]^2 + 4\,[\psi(\alpha) - \psi(\alpha+\beta)] \\
&\qquad\qquad\qquad\qquad \left[\psi''(\alpha) - \psi''(\alpha+\beta)\right],
\end{aligned}
\tag{11}
$$

where $\psi(x) = \frac{d}{dx}\ln\Gamma(x)$ is the digamma function, $\psi'(x) = \frac{d^2}{dx^2}\ln\Gamma(x) = \frac{d}{dx}\psi(x)$ is the trigamma function, $\psi''(x) = \frac{d}{dx}\psi'(x)$ is the quadrigamma function, and so on.

3. Near-exact and asymptotic distributions for the generalized Wilks Λ statistic

Let $\underline{X}$ be a random vector with dimension p, where the r.v.'s have a joint p-multivariate Normal distribution $N_p(\underline{\mu}, \Sigma)$. Let us consider $\underline{X}$ split into m subvectors, where the k-th subvector has p_k variables, being $p = \sum_{k=1}^{m} p_k$ the overall number of variables. Then, each subvector $\underline{X}_k (k = 1, \ldots, m)$ will have a p_k-multivariate Normal distribution $N_{p_k}(\underline{\mu}_k, \Sigma_{kk})$. Symbolically,

$$\underline{X} = [\underline{X}_1', \ldots, \underline{X}_k', \ldots, \underline{X}_m']' \sim N_p\left(\underline{\mu}, \Sigma\right)$$

where

$$\underline{\mu} = [\underline{\mu}_1', \ldots, \underline{\mu}_k', \ldots, \underline{\mu}_m']', \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1k} & \cdots & \Sigma_{1m} \\ \vdots & \ddots & \vdots & & \vdots \\ \Sigma_{k1} & \cdots & \Sigma_{kk} & \cdots & \Sigma_{km} \\ \vdots & & \vdots & \ddots & \vdots \\ \Sigma_{m1} & \cdots & \Sigma_{mk} & \cdots & \Sigma_{mm} \end{bmatrix}.$$

For a sample of size $n+1$, the $2/(n+1)$-th power of likelihood ratio test statistic, used to test the null hypothesis of independence of the m subvectors $\underline{X}_k$,

$$H_0 : \Sigma = diag(\Sigma_{11}, \ldots, \Sigma_{kk}, \ldots, \Sigma_{mm}) \,, \tag{12}$$

is the generalized Wilks Λ statistic

$$\Lambda = \frac{|V|}{\prod_{k=1}^{m} |V_{kk}|} \,, \tag{13}$$

where $|\,.\,|$ stands for the determinant and V is either the Maximum Likelihood Estimator (MLE) of Σ or the sample variance-covariance matrix of $\underline{X}$, and V_{kk} is either the MLE of Σ_{kk} or the sample variance-covariance matrix of $\underline{X}_k$.

The generalized Wilks Λ statistic may be written as (Anderson, 2003, Theorem 9.3.2)

$$\Lambda = \prod_{k=1}^{m-1} \Lambda_{k(k+1,\ldots,m)} \,, \tag{14}$$

where $\Lambda_{k(k+1,\ldots,m)}$ denotes the Wilks Λ statistic used to test the independence between the k-th subvector and the vector formed by joining subvectors $k+1$ through m. In other words, for $k = 1, \ldots, m-1$, $\Lambda_{k(k+1,\ldots,m)}$ is the Wilks Λ statistic used to test the null hypothesis,

$$H_0^{(k)} : [\Sigma_{k,k+1} \ldots\ \Sigma_{km}] = 0_{p_k \times (p_{k+1}+\ldots+p_m)}, \qquad k = 1, \ldots, m-1. \tag{15}$$

Using the result in Theorem 9.3.2 in Anderson (2003) and considering that the k-th subvector has p_k variables ($k = 1, \ldots, m$), the distribution of $\Lambda_{k(k+1,\ldots,m)}$ in (**??**), under the null hypothesis $H_0^{(k)}$, is the same as the distribution of $\prod_{j=1}^{p_k} Y_j$, where, for a sample of size $n+1$ (with $n \geq p_1 + \cdots + p_m$), Y_j are p_k independent r.v.'s with Beta distributions,

$$Y_j \sim Beta\left(\frac{n+1-q_k-j}{2}, \frac{q_k}{2}\right), \qquad j = 1, \ldots, p_k \,,$$

where $q_k = p_{k+1} + \cdots + p_m$. This way, based on expression (**??**) we may write

$$E(Y_j^h) = \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}+h\right)}{\Gamma\left(\frac{n+1-j}{2}+h\right)}, \qquad \left(h > -\frac{n+1-q_k-j}{2}\right)$$

and, given the independence of the p_k r.v.'s Y_j, under the null hypothesis $H_0^{(k)}$ in (**??**),

$$E\left(\Lambda^h_{k(k+1,\dots,m)}\right) = \prod_{j=1}^{p_k} E(Y_j^h) = \prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}+h\right)}{\Gamma\left(\frac{n+1-j}{2}+h\right)},$$

$$\left(h > -\frac{n+1-q_k-p_k}{2}\right).$$

Given the independence of the $m-1$ statistics $\Lambda_{k(k+1,\dots,m)}$ in (**??**), under the null hypothesis of independence of the m sets of variables in (**??**), we obtain the h-th moment of the generalized Wilks Λ statistic in (**??**), for a sample of size $n+1$, as

$$\begin{aligned} E(\Lambda^h) &= \prod_{k=1}^{m-1} E[\Lambda^h_{k(k+1,\dots,m)}] = \prod_{k=1}^{m-1}\prod_{j=1}^{p_k} E(Y_j^h) \\ &= \prod_{k=1}^{m-1}\prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}+h\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}+h\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)}. \end{aligned} \tag{16}$$

Since the Gamma functions in (**??**) are still valid for any strictly complex h, for a sample of size $n+1$, the c.f. of the r.v. $W = -\ln\Lambda$ is given by

$$\begin{aligned} \phi_W(t) &= E(\mathrm{e}^{\mathrm{i}tW}) = E(\mathrm{e}^{-\mathrm{i}t\ln\Lambda}) = E(\Lambda^{-\mathrm{i}t}) \\ &= \prod_{k=1}^{m-1}\prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)}, \end{aligned} \tag{17}$$

where $\mathrm{i} = (-1)^{1/2}$ and $t \in \mathbb{R}$. Taking this c.f. as a basis, we will develop in the next subsections two near-exact and three asymptotic distributions for W.

3.1. Two near-exact distributions for the generalized Wilks Λ statistic

In Theorem 1 we present two near-exact distributions for the generalized Wilks Λ statistic, in the case where at least two sets have an odd number of variables. One of these distributions is a GNIG distribution that matches the first three exact moments and the other is a M2GNIG distribution which matches the first four exact moments. These distributions emerge as the direct application of the procedure used by Grilo (2005) and Grilo and Coelho (2007) to obtain two near-exact distributions for the product of particular independent Beta r.v.'s.

Theorem 1. *When, among the m sets of variables there are l sets with an even number of variables, i.e., there are $m-l$ sets that have an odd number of variables, then let $m-l=2k^*$, if $m-l$ is even or $m-l=2k^*+1$, if $m-l$ is odd (where $k^*=\left\lfloor\frac{m-l}{2}\right\rfloor$ is the integer part of $\frac{m-l}{2}$). Then, under* (**??**) *and for a sample of size $n+1$, we may obtain two different near-exact distributions for the r.v. $W=-\ln\Lambda$. A first near-exact distribution may be obtained in the form of a GNIG distribution of depth $p=p_1+p_2+\ldots+p_m$,*

$$W \overset{ne}{\sim} GNIG(r_1^*,\ldots,r_{p-2}^*,r_{p-1}^*,r_p^*;\lambda_1,\ldots,\lambda_{p-2},\lambda_{p-1},\lambda_p)$$

with rate parameters

$$\lambda_j=\frac{n-p+j}{2}, \qquad j=1,\ldots,p-2\,, \tag{18}$$

and shape parameters

$$r_j^*=\sum_{k=1}^{m-2k^*-1} r_{k,j-p_k^*}+\sum_{\substack{k=m-2k^*\\ step\,2}}^{m-2} r_{k,j-p_k^*}+\sum_{\substack{k=m-2k^*+1\\ step\,2}}^{m-1} r^*_{k,j-p_k^*},\quad j=1,\ldots,p-2 \tag{19}$$

with $p_k^*=\sum_{l=1}^{k-1}p_l$, and

$$r_{k,j-p_k^*}=0 \quad if \quad p_k^*\geq j,$$

$$r^*_{k,j-p_k^*}=0 \quad if \quad p_k^*\geq j \quad or \quad j=p-2,$$

where, for $k=1,\ldots,m-2k^*-1$ (step 1) and $k=m-2k^*,\ldots,m-2$ (step 2),

$$r_{kj}=\begin{cases} h_{kj} & j=1,\ 2,\\ r_{k,j-2}+h_{kj} & j=3,\ldots,\ p_k+q_k-2\end{cases} \tag{20}$$

with

$$h_{kj} = (\textit{number of elements of } \{p_k, q_k\} \geq j) - 1 \tag{21}$$

and for $k = m - 2k^* + 1, \ldots, m - 1$ *(with step* 2*)*

$$r^*_{kj} = \begin{cases} r'_{kj} & j = 1, \ldots, p_k - 1, \\ & j = p_k + 2n + 1;\ \ n = 0, \ldots, \frac{q_k - 5}{2}, \\ r'_{kj} + 1 & j = p_k + 2n;\ \ n = 0, \ldots, \frac{q_k - 5}{2}, \end{cases} \tag{22}$$

where

$$r'_{kj} = \begin{cases} h'_{kj} & j = 1,\ 2, \\ r'_{k,j-2} + h'_{kj} & j = 3, \ldots, p_k + q_k - 3 \end{cases} \tag{23}$$

with

$$h'_{kj} = (\textit{number of elements of } \{p_k - 1, q_k\} \geq j) - 1 \tag{24}$$

and, yet with $r^*_{p-1} = 1$*, and* r^*_p*,* λ_{p-1} *and* λ_p *obtained by numeric solution of the system of equations*

$$\begin{cases} \mu'_1 = \dfrac{1}{\lambda_{p-1}} + \dfrac{r^*_p}{\lambda_p}, \\[2ex] \mu'_2 = \dfrac{2\lambda_p^2 + 2\lambda_{p-1}\lambda_p r^*_p + \lambda^2_{p-1} r^*_p(1+r^*_p)}{\lambda^2_{p-1}\lambda^2_p}, \\[2ex] \mu'_3 = \dfrac{6\lambda_p^3 + 6\lambda_{p-1}\lambda_p^2 r^*_p + 3\lambda^2_{p-1}\lambda_p r^*_p(1+r^*_p) + \lambda^3_{p-1} r^*_p(2+3r^*_p+r^{*2}_p)}{\lambda^3_{p-1}\lambda^3_p}, \end{cases} \tag{25}$$

where, on the first member of (**??**)*,* μ'_1, μ'_2 *and* μ'_3 *are the first three moments of a Logbeta r.v. with parameters* $\alpha = \frac{n}{2} - \frac{3}{2}$ *and* $\beta = \frac{3}{2}$*, obtained from* (**??**) *by replacing* α *and* β *by the appropriate values, and on the second member we have the expressions of the first three moments of the sum of two independent Gamma r.v.'s, the first one with shape parameter* $r^*_{p-1} = 1$ *and rate parameter* λ_{p-1} *and the second one with shape parameter* r^*_p *and rate parameter* λ_p*.*

The second near-exact distribution for the r.v. $W = -\ln\Lambda$ is a M2GNIG distribution, where both components have depth $p-1$,

$$W \overset{ne}{\sim} M2GNIG(\pi; r_1^*, \ldots, r_{p-2}^*, r_{p-1}; \lambda_1, \ldots, \lambda_{p-2}, \lambda_{p-1} |$$
$$1-\pi; r_1^*, \ldots, r_{p-2}^*, r_{p-1}; \lambda_1, \ldots, \lambda_{p-2}, \lambda'_{p-1})$$

where the shape parameters $r_1^*, \ldots, r_{p-2}^*$ are given by (??) through (??) and the rate parameters $\lambda_1, \ldots, \lambda_{p-2}$ by (??). Considering the same shape parameter r_{p-1} for both GNIG distributions in the mixture, we obtain π, r_{p-1}, λ_{p-1} and λ'_{p-1} by numeric solution of the system of equations

$$\begin{cases} \mu'_1 = \pi \dfrac{\Gamma(r_{p-1}+1)}{\Gamma(r_{p-1})} \dfrac{1}{\lambda_{p-1}} + (1-\pi) \dfrac{\Gamma(r_{p-1}+1)}{\Gamma(r_{p-1})} \dfrac{1}{\lambda'_{p-1}}, \\[2ex] \mu'_2 = \pi \dfrac{\Gamma(r_{p-1}+2)}{\Gamma(r_{p-1})} \dfrac{1}{\lambda^2_{p-1}} + (1-\pi) \dfrac{\Gamma(r_{p-1}+2)}{\Gamma(r_{p-1})} \dfrac{1}{\lambda'^2_{p-1}}, \\[2ex] \mu'_3 = \pi \dfrac{\Gamma(r_{p-1}+3)}{\Gamma(r_{p-1})} \dfrac{1}{\lambda^3_{p-1}} + (1-\pi) \dfrac{\Gamma(r_{p-1}+3)}{\Gamma(r_{p-1})} \dfrac{1}{\lambda'^3_{p-1}}, \\[2ex] \mu'_4 = \pi \dfrac{\Gamma(r_{p-1}+4)}{\Gamma(r_{p-1})} \dfrac{1}{\lambda^4_{p-1}} + (1-\pi) \dfrac{\Gamma(r_{p-1}+4)}{\Gamma(r_{p-1})} \dfrac{1}{\lambda'^4_{p-1}}, \end{cases} \tag{26}$$

where, on the first member of (??), μ'_1, μ'_2, μ'_3 and μ'_4 represent the first four moments of the sum of k^* independent and identically distributed (i.i.d.) Logbeta r.v.'s with parameters $\alpha = \frac{n}{2} - \frac{3}{2}$ and $\beta = \frac{3}{2}$, and in the second member we have the first four moments of a mixture of two Gamma distributions (M2G) with weights π and $1-\pi$, the first one with shape parameter r_{p-1} and rate parameter λ_{p-1} and the second one with shape parameter r_{p-1} and rate parameter λ'_{p-1}.

Proof. We will consider that, without any loss of generality, the sets of variables with an odd number of variables are, among the m sets, the last $m-l$ sets of variables, that is, the sets $1, \ldots, l$ have an even number of variables and the remaining, $l+1, \ldots, m$, have an odd number of variables. Take $k^* = \lfloor \frac{m-l}{2} \rfloor$ with $k^* \in \mathbb{N}_0$. Then, we may write

$$\phi_W(t) = \prod_{k=1}^{m-(2k^*+1)} \underbrace{\prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)}}_{p_k \text{ even}}$$

$$\times \prod_{\substack{k=m-2k^*\\ \text{step } 2}}^{m-2} \underbrace{\prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)}}_{q_k \text{ even}}$$

$$\times \prod_{\substack{k=m-(2k^*-1)\\ \text{step } 2}}^{m-1} \underbrace{\prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)}}_{p_k \text{ and } q_k \text{ odd}},$$

where for the first two factors (with p_k or q_k even), we use the identity

$$\prod_{j=1}^{p} \frac{\Gamma\left(c+\frac{p}{2}-\frac{j}{2}+\frac{b}{2}\right)}{\Gamma\left(c+\frac{p}{2}-\frac{j}{2}\right)} = \prod_{j=1}^{p+b-2} \left(c+\frac{j}{2}-\frac{1}{2}\right)^{r_j}$$

with $c \in \mathbb{R}^+$ and $\frac{b}{2} \in \mathbb{N}$ or $\frac{p}{2} \in \mathbb{N}$ (Coelho, 1998), to rewrite the c.f. of W in the form

$$\phi_W(t) = \prod_{k=1}^{m-2k^*-1} \underbrace{\prod_{j=1}^{p_k+q_k-2} \left(\frac{n-p_k-q_k+j}{2}\right)^{r_{kj}} \left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r_{kj}}}_{p_k \text{ even}}$$

$$\times \prod_{\substack{k=m-2k^*\\ \text{step } 2}}^{m-2} \underbrace{\prod_{j=1}^{p_k+q_k-2} \left(\frac{n-p_k-q_k+j}{2}\right)^{r_{kj}} \left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r_{kj}}}_{q_k \text{ even}}$$

$$\times \prod_{\substack{k=m-2k^*+1\\ \text{step } 2}}^{m-1} \underbrace{\prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)}}_{p_k \text{ and } q_k \text{ odd}}$$

with r_{kj} given by (**??**) and (**??**). For the last factor, where p_k and q_k are both odd, we may write

$$\begin{aligned}
&\prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)} \\
&\qquad = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n-q_k}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-q_k}{2}\right)} \prod_{j=2}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)} \\
&\qquad = \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{3}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n}{2}-\frac{3}{2}\right)\Gamma\left(\frac{n}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n}{2}-\frac{3}{2}\right)\Gamma\left(\frac{n-q_k}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-q_k}{2}\right)\Gamma\left(\frac{n}{2}-\frac{3}{2}-\mathrm{i}t\right)} \\
&\qquad\qquad \times \prod_{j=1}^{p_k-1} \frac{\Gamma\left(\frac{n+1-(j+1)}{2}\right)}{\Gamma\left(\frac{n+1-(j+1)}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_k-(j+1)}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-(j+1)}{2}\right)} \\
&\qquad = \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{3}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n}{2}-\frac{3}{2}\right)\Gamma\left(\frac{n}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n-q_k}{2}+\frac{q_k-3}{2}\right)\Gamma\left(\frac{n-q_k}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-q_k}{2}\right)\Gamma\left(\frac{n-q_k}{2}+\frac{q_k-3}{2}-\mathrm{i}t\right)} \\
&\qquad\qquad \times \prod_{j=1}^{p_k-1} \frac{\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-q_k-j}{2}\right)} \,.
\end{aligned}$$

Since q_k is a positive odd integer and thus $\frac{q_k-3}{2}$ is a positive integer, we may use the identity,

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} = \prod_{j=0}^{\beta-1}(\alpha+j)\,,$$

which is valid for $\beta \in \mathbb{N}$ and α real or complex, with $\alpha = \frac{n-q_k}{2}$ and $\beta = \frac{q_k-3}{2}$, and write

$$\prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)}$$

$$= \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{3}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n}{2}-\frac{3}{2}\right)\Gamma\left(\frac{n}{2}-\mathrm{i}t\right)} \prod_{j=0}^{\frac{q_k-3}{2}-1} \left(\frac{n-q_k}{2}+j\right)\left(\frac{n-q_k}{2}+j-\mathrm{i}t\right)^{-1}$$

$$\times \prod_{j=1}^{p_k-1} \frac{\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-q_k-j}{2}\right)},$$

where, given that p_k is odd, we have p_k-1 even, so that we may write

$$\prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)}$$

$$= \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{3}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n}{2}-\frac{3}{2}\right)\Gamma\left(\frac{n}{2}-\mathrm{i}t\right)} \prod_{j=0}^{\frac{q_k-3}{2}-1} \left(\frac{n-q_k}{2}+j\right)\left(\frac{n-q_k}{2}+j-\mathrm{i}t\right)^{-1}$$

$$\times \prod_{j=1}^{p_k+q_k-3} \left(\frac{n-q_k-p_k}{2}+\frac{j}{2}\right)^{r_{kj}} \left(\frac{n-q_k-p_k}{2}+\frac{j}{2}-\mathrm{i}t\right)^{-r_{kj}}$$

$$= \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{3}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n}{2}-\frac{3}{2}\right)\Gamma\left(\frac{n}{2}-\mathrm{i}t\right)} \prod_{j=1}^{p_k+q_k-3} \left(\frac{n-q_k-p_k}{2}+\frac{j}{2}\right)^{r^*_{kj}}$$

$$\left(\frac{n-q_k-p_k}{2}+\frac{j}{2}-\mathrm{i}t\right)^{-r^*_{kj}}$$

with r_{kj} $(k=m-2k^*+1, m-2k^*+3, \ldots, m-1; j=1,\ldots,p_k+q_k-3)$ and r^*_{kj} given by (**??**) through (**??**). We may thus rewrite the c.f. of W, as

(27)

$$
\begin{aligned}
\phi_W(t) \\
&= \prod_{k=1}^{m-2k^*-1} \prod_{j=1}^{p_k+q_k-2} \left(\frac{n-p_k-q_k+j}{2}\right)^{r_{kj}} \left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r_{kj}} \\
&\quad \times \prod_{\substack{k=m-2k^* \\ \text{step } 2}}^{m-2} \prod_{j=1}^{p_k+q_k-2} \left(\frac{n-p_k-q_k+j}{2}\right)^{r_{kj}} \left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r_{kj}} \\
&\quad \times \prod_{\substack{k=m-2k^*+1 \\ \text{step } 2}}^{m-1} \left\{ \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{3}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n}{2}-\frac{3}{2}\right)\Gamma\left(\frac{n}{2}-\mathrm{i}t\right)} \prod_{j=1}^{p_k+q_k-3} \left(\frac{n-p_k-q_k+j}{2}\right)^{r^*_{kj}} \right. \\
&\qquad\qquad \left. \left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r^*_{kj}} \right\} \\
&= \prod_{k=1}^{m-2k^*-1} \prod_{j=1}^{p_k+q_k-2} \left(\frac{n-p_k-q_k+j}{2}\right)^{r_{kj}} \left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r_{kj}} \\
&\quad \times \prod_{\substack{k=m-2k^* \\ \text{step } 2}}^{m-2} \prod_{j=1}^{p_k+q_k-2} \left(\frac{n-p_k-q_k+j}{2}\right)^{r_{kj}} \left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r_{kj}} \\
&\quad \times \prod_{\substack{k=m-2k^*+1 \\ \text{step } 2}}^{m-1} \prod_{j=1}^{p_k+q_k-3} \left(\frac{n-p_k-q_k+j}{2}\right)^{r^*_{kj}} \left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r^*_{kj}} \\
&\quad \times \left\{ \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{3}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n}{2}-\frac{3}{2}\right)\Gamma\left(\frac{n}{2}-\mathrm{i}t\right)} \right\}^{k^*} \\
&= \left\{ \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}-\frac{3}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n}{2}-\frac{3}{2}\right)\Gamma\left(\frac{n}{2}-\mathrm{i}t\right)} \right\}^{k^*} \prod_{j=1}^{p-2} \left(\frac{n-p+j}{2}\right)^{r^*_j} \left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r^*_j},
\end{aligned}
$$

where r_j^* are given by (**??**). In (**??**), we will replace the c.f. of a Logbeta r.v. with parameters $\frac{n}{2}-\frac{3}{2}$ and $\frac{3}{2}$, by the c.f. of the sum of two Gamma r.v.'s,

$$\lambda_{p-1}(\lambda_{p-1}-\mathrm{i}t)^{-1}\lambda_p^{r_p^*}(\lambda_p-\mathrm{i}t)^{-r_p^*}\,,$$

where the parameters r_p^*, λ_{p-1} and λ_p are obtained in such a way that the first three derivatives of both c.f.'s with respect to t, at $t=0$, are equal. This means that the distributions to which they correspond will have the same first three moments. This leads us to obtain such parameters as the solutions of the system of equations (**??**).

The expression of the near-exact c.f. of W obtained in this way is of the type in (**??**), more precisely, it is given by

$$\begin{aligned}
&\left\{\lambda_{p-1}(\lambda_{p-1}-\mathrm{i}t)^{-1}\,\lambda_p^{r_p^*}(\lambda_p-\mathrm{i}t)^{-r_p^*}\right\}^{k^*}\\
&\qquad\qquad\times\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_j^*}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_j^*}\\
&=\lambda_{p-1}^{k^*}(\lambda_{p-1}-\mathrm{i}t)^{-k^*}\,\lambda_p^{k^*r_p^*}(\lambda_p-\mathrm{i}t)^{-k^*r_p^*}\\
&\qquad\qquad\times\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_j^*}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_j^*}\,,
\end{aligned}\tag{28}$$

that is the c.f. of a r.v. with a GNIG distribution of depth p, whose first three moments will match the first three moments of the exact distribution. More precisely, (**??**) is the product of the c.f. of the sum of $p-2$ independent r.v.'s with Gamma distribution, which corresponds to a GIG distribution of depth $p-2$, with shape parameters r_j^* given by (**??**) and rate parameters λ_j given by (**??**), by the c.f. of the sum of two independent r.v.'s with Gamma distribution, with shape parameters $k^*\in\mathbb{N}$ and $k^*r_p^*$ and rate parameters λ_{p-1} and λ_p. Thus, the c.f. in (**??**) is the c.f. of the sum of a r.v. with a GIG distribution of depth $p-2$ with a r.v. with a GNIG distribution of depth 2, yielding a GNIG distribution of depth p.

We may obtain another near-exact c.f. if, in (**??**), we replace the part that corresponds to the sum of k^* i.i.d. r.v.'s with a Logbeta distribution with parameters $\frac{n}{2}-\frac{3}{2}$ and $\frac{3}{2}$ by the c.f. of a M2G distribution with equal shape parameters, r_{p-1}, and rate parameters λ_{p-1} and λ'_{p-1}, i.e.,

$$\pi \frac{\lambda_{p-1}^{r_{p-1}}}{(\lambda_{p-1}-\mathrm{i}t)^{r_{p-1}}} + (1-\pi)\frac{\lambda_{p-1}'^{\,r_{p-1}}}{(\lambda_{p-1}'-\mathrm{i}t)^{r_{p-1}}},$$

where the parameters π, r_{p-1}, λ_{p-1} and λ'_{p-1} are obtained in such a way that the first four derivatives of both functions with respect to t, at $t=0$, are equal. That is, the first four moments of the exact and near-exact distributions of W will be the same. Such parameters are obtained as the solution of the system of equations in (**??**).

The expression of the near-exact c.f. of W is then given by

$$\left\{\pi \frac{\lambda_{p-1}^{r_{p-1}}}{(\lambda_{p-1}-\mathrm{i}t)^{r_{p-1}}} + (1-\pi)\frac{\lambda_{p-1}'^{\,r_{p-1}}}{(\lambda_{p-1}'-\mathrm{i}t)^{r_{p-1}}}\right\} \times \prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_j^*}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_j^*}, \tag{29}$$

that is the product of the c.f. of the sum of $p-2$ independents r.v.'s with Gamma distributions, which corresponds to a GIG distribution of depth $p-2$ (with shape parameters r_j^* given by (**??**) and rate parameters λ_j given by (**??**)), by the c.f. of a M2G distribution with both shape parameters equal to r_{p-1} and rate parameters λ_{p-1} and λ'_{p-1}, and weights π and $1-\pi$. In other words, (**??**) is thus the c.f. of the sum of a r.v. with a GIG distribution of depth $p-2$ with a r.v. with a M2G distribution, or yet, the c.f. of a r.v. with a M2GNIG distribution of depth $p-1$, which, by construction, matches the first four moments of the exact distribution.

The expressions for the near-exact density and cumulative distribution functions of $W=-\ln\Lambda$ may be obtained from (**??**) and (**??**), respectively, by making the appropriate replacement of parameters. From these we may easily derive, by simple transformation, the corresponding near-exact density and cumulative distribution functions for the generalized Wilks Λ statistic. This way we obtain, for the first near-exact distribution in Theorem 1

$$f_\Lambda(u) \approx K\lambda_p^{r_p^*}\sum_{j=1}^{p-1} u^{\lambda_j}\sum_{k=1}^{r_j^*} c_{j,k}\frac{\Gamma(k)}{\Gamma(k+r_p^*)}(-\ln u)^{k+r_p^*-1} \times {}_1F_1(r_p^*, k+r_p^*, (\lambda_p-\lambda_j)\ln u), \quad (u>0),$$

as near-exact p.d.f. for Λ, and

$$F_\Lambda(u) \approx 1 - \lambda_p^{r_p^*} \frac{(-\ln u)^{r_p^*}}{\Gamma(r_p^*+1)} \, {}_1F_1(r_p^*, r_p^*+1, \lambda_p \ln u)$$

$$+K\lambda_p^{r_p^*} \sum_{j=1}^{p-1} u^{\lambda_j} \sum_{k=1}^{r_j^*} c_{j,k}^* \sum_{i=0}^{k-1} \frac{(-\ln u)^{r_p^*+i} \lambda_j^i}{\Gamma(r_p^*+1+i)}$$

$$\times \, {}_1F_1(r_p^*, r_p^*+1+i, (\lambda_p-\lambda_j)\ln u), \quad (u>0)\,,$$

as near-exact c.d.f., with

$$K = \prod_{j=1}^{p-1} \lambda_j^{r_j^*} \quad \text{and} \quad c_{j,k}^* = \frac{c_{j,k}}{\lambda_j^k} \Gamma(k),$$

while for the second near-exact distribution in Theorem 1, we have

$$f_\Lambda(u) \approx \pi K \lambda_{p-1}^{r_{p-1}} \sum_{j=1}^{p-2} u^{\lambda_j} \sum_{k=1}^{r_j^*} c_{j,k} \frac{\Gamma(k)}{\Gamma(k+r_{p-1})} (-\ln u)^{k+r_{p-1}-1}$$

$$\times \, {}_1F_1(r_{p-1}, k+r_{p-1}, (\lambda_{p-1}-\lambda_j)\ln u)$$

$$+ (1-\pi) K {\lambda'}_{p-1}^{r_{p-1}} \sum_{j=1}^{p-2} u^{\lambda_j} \sum_{k=1}^{r_j^*} c_{j,k} \frac{\Gamma(k)}{\Gamma(k+r_{p-1})} (-\ln u)^{k+r_{p-1}-1}$$

$$\times \, {}_1F_1(r_{p-1}, k+r_{p-1}, (\lambda'_{p-1}-\lambda_j)\ln u), \quad (u>0),$$

as the near-exact p.d.f. for Λ, and

$$F_\Lambda(u) \approx 1 - \pi \, \lambda_{p-1}^{r_{p-1}} \frac{(-\ln u)^{r_{p-1}}}{\Gamma(r_{p-1}+1)} \, {}_1F_1(r_{p-1}, r_{p-1}+1, \lambda_{p-1} \ln u)$$

$$+K\lambda_{p-1}^{r_{p-1}} \sum_{j=1}^{p-2} u^{\lambda_j} \sum_{k=1}^{r_j^*} c_{j,k}^* \sum_{i=0}^{k-1} \frac{(-\ln u)^{r_{p-1}+i} \lambda_j^i}{\Gamma(r_{p-1}+1+i)}$$

$$\times \, {}_1F_1(r_{p-1}, r_{p-1}+1+i, (\lambda_{p-1}-\lambda_j)\ln u)$$

$$- (1-\pi)\lambda'^{r_{p-1}}_{p-1} \frac{(-\ln u)^{r_{p-1}}}{\Gamma(r_{p-1}+1)} {}_1F_1(r_{p-1}, r_{p-1}+1, \lambda'_{p-1}\ln u)$$

$$+ (1-\pi)K\lambda'^{r_{p-1}}_{p-1} \sum_{j=1}^{p-2} u^{\lambda_j} \sum_{k=1}^{r_j^*} c^*_{j,k} \sum_{i=0}^{k-1} \frac{(-\ln u)^{r_{p-1}+i}\lambda_j^i}{\Gamma(r_{p-1}+1+i)}$$

$$\times\, {}_1F_1(r_{p-1}, r_{p-1}+1+i, (\lambda'_{p-1}-\lambda_j)\ln u), \quad (u>0),$$

as the near-exact c.d.f. of Λ, with

$$K = \prod_{j=1}^{p-2} \lambda_j^{r_j^*} \quad \text{and} \quad c^*_{j,k} = \frac{c_{j,k}}{\lambda_j^k}\Gamma(k).$$

Based on the c.d.f.'s presented it is quite easy to compute near-exact quantiles.

3.2. Asymptotic distributions for the generalized Wilks Λ statistic

As approximations for the generalized Wilks Λ statistic we also consider the asymptotic distribution proposed by Box (1949) and Anderson (2003) and three asymptotic distributions developed by us, which match some of the first exact moments.

3.2.1. Box-Anderson asymptotic distribution for the statistic $W = -\ln\Lambda$

Box (1949) and Anderson (2003, Section 9.4 of Chapter 9) developed two well-known asymptotic distributions for linear transformations of the logarithm of the Wilks Λ statistic, under the null hypotheses of independence of the m sets of variables. These are based on series expansions which use Chi-square distributions. As we can see in Appendix A, the two asymptotic distributions proposed by the two authors agree to terms of order η^{-2}, with η given by (**??**).

Based on the results obtained by those two authors we will use, as asymptotic approximation for the distribution of the r.v. $V_2 = \eta W$, a mixture of two Chi-square distributions, i.e., we will use (see Appendix A)

$$\phi_{V_2}(t) \cong \left(1 - \frac{\gamma_2}{\eta^2}\right)\phi_{\chi^2_f}(t) + \frac{\gamma_2}{\eta^2}\phi_{\chi^2_{f+4}}(t), \tag{30}$$

where

$$\gamma_2 = \frac{S_4}{48} - \frac{5}{96}S_2 - \frac{(S_3)^2}{72S_2} = \frac{p^4 - \sum_{k=1}^{m} p_k^4}{48} - \frac{5\left(p^2 - \sum_{k=1}^{m} p_k^2\right)}{96} - \frac{\left(p^3 - \sum_{k=1}^{m} p_k^3\right)^2}{72\left(p^2 - \sum_{k=1}^{m} p_k^2\right)},$$

(31) $$\eta = n + 1 - \frac{9S_2 + 2S_3}{6S_2}$$

and

$$\phi_{\chi^2_f}(t) = \left(\frac{1}{2}\right)^{\frac{f}{2}} \left(\frac{1}{2} - \mathrm{i}t\right)^{-\frac{f}{2}}$$

is the c.f. of a r.v. with a Chi-square distribution with f degrees of freedom. Since we have

$$\phi_W(t) = E(\mathrm{e}^{\mathrm{i}tW}) = E(\mathrm{e}^{\mathrm{i}(t/\eta)V}),$$

the use of (**??**), is equivalent to the use, for the c.f. of the r.v. $W = -\ln\Lambda$, of the approximation

(32) $$\phi_W(t) \cong \left(1 - \frac{\gamma_2}{\eta^2}\right) \phi_{\chi^2_f}\left(\tfrac{t}{\eta}\right) + \frac{\gamma_2}{\eta^2}\phi_{\chi^2_{f+4}}\left(\tfrac{t}{\eta}\right) .$$

We will call the asymptotic distribution derived from (**??**) the Box-Anderson distribution.

3.2.2. Asymptotic distributions for the statistic $W = -\ln\Lambda$ which equate moments

We will also approximate the whole c.f., $\phi_W(t)$ in (**??**), by the c.f. of a Gamma r.v., by the c.f. of a GNIG r.v. with depth 2 with c.f.

$$\lambda_{p-1}(\lambda_{p-1} - \mathrm{i}t)^{-1}\lambda_p^{r_p^*}(\lambda_p - \mathrm{i}t)^{-r_p^*}$$

or by the c.f. of a M2G distribution (with both components with the same shape parameters). The approximation is done in such a way that if these approximating c.f.'s have d parameters, their first d derivatives with respect to t, at $t=0$, will match the corresponding first d derivatives of $\phi_W(t)$ with respect to t, at $t=0$. The asymptotic distributions obtained in this way are: a Gamma, a GNIG and a M2G distribution, which match the first two, three and four exact moments, respectively.

4. Comparative numerical studies

To assess the performance of the asymptotic and near-exact distributions proposed we use two proximity measures, based on the difference between the exact and asymptotic or near-exact c.f.'s. These measures were used by Grilo and Coelho (2007) and they are directly derived from the inversion formulas respectively for the p.d.f. and the c.d.f.. Their expressions are

$$\Delta_1 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left|\phi_W(t) - \phi(t)\right| \, dt$$

and

$$\Delta_2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left|\frac{\phi_W(t) - \phi(t)}{t}\right| \, dt, \tag{33}$$

where $\phi_W(t)$ represents the exact c.f. of the r.v. W and $\phi(t)$ the approximate (asymptotic or near-exact) c.f., corresponding to the distribution under study. The measure Δ_2 in (**??**) may be seen as directly derived from the Berry-Esseen bound and the use of the measures Δ_1 and Δ_2 enables us to obtain upper bounds on the absolute value of the differences of the density and the cumulative distribution functions, respectively. More precisely,

$$\max_{w>0} |f_W(w) - f(w)| \leq \Delta_1 \qquad \text{and} \qquad \max_{w>0} |F_W(w) - F(w)| \leq \Delta_2\,,$$

where $f_W(w)$ and $F_W(w)$ are, respectively, the exact p.d.f. and c.d.f. of W, evaluated at $w>0$, and $f(w)$ and $F(w)$ are, respectively, the asymptotic or near-exact p.d.f. and c.d.f. of W. The proposed measures are an important tool to assess the proximity between asymptotic or near-exact distributions and exact distributions, mainly in cases where the expressions for the exact p.d.f. or c.d.f. are not known, or being known they are so complicated that they are not manageable. This way, smaller values of the measures are associated with better closeness of the distributions (in terms of moments, quantiles and c.f., and as such also in terms of density and cumulative distribution functions). The measures Δ_1 and Δ_2 are accurate to evaluate the proximity of quantiles, with smaller values of these measures being associated with smaller differences among quantiles (see Grilo and Coelho, 2007, 2010).

In this stage we perform a comparative numerical study among the approximations proposed. We consider four asymptotic distributions: the Box-Anderson which does not equate any moments (Box, 1949; Anderson, 2003), a Gamma, a GNIG and a M2G, which equate the first two, three and four exact moments, respectively (developed according to Subsection 3.2.2); and three near-exact distributions: a GNIG which equates two exact moments (Coelho, 2004), a GNIG and a M2GNIG which equate the first three and four exact moments, respectively (developed in Subsection 3.1). These approximations and the number of exact moments they match are shown in Table 1.

Table 1. Asymptotic and near-exact distributions and the number of exact moments equated.

Distributions		**No. of moments equated**
Asymptotic	Box-Anderson	0
	Gamma	2
	GNIG	3
	M2G	4
Near-exact	GNIG	2
	GNIG	3
	M2GNIG	4

We will use the measures Δ_1 and Δ_2 to assess the proximity of the different distributions, for variations in the number of sets (m), in the number of variables per set (p_k) and in the sample size (n). In Table 2 is displayed a summary of the cases considered in the comparative study.

Table 2. Number of sets, number of variables per set and sample size.

No. of sets	**No. of variables per set**	**Total no. of variables**	**Sample size**
$m = 3$	$p_1 = 5,\ p_2 = 7,\ p_3 = 3$	$p = 15$	$n = 25$ and $n = 100$
	$p_1 = 5,\ p_2 = 7,\ p_3 = 9$	$p = 21$	
$m = 4$	$p_1 = 5,\ p_2 = 7,\ p_3 = 3,\ p_4 = 6$	$p = 21$	

In Tables 3 through 5 we may see how, opposite to the asymptotic distributions, the near-exact distributions show an asymptotic behavior also for an increasing number of variables, not only in terms of increasing values of p_k, when keeping m unchanged, but also for increasing m, keeping $p = p_1 + p_2 + \ldots + p_m$ unchanged.

As expected, the values of the proximity measures decline with increasing values of the sample size both for the asymptotic and near-exact distributions. Also, systematically, distributions that equate a larger number of exact moments have lower values of the proximity measures. Both for the asymptotic and near-exact distributions we have with lower values of measures the two approximations based on mixtures: the M2G in the case of asymptotic distributions and the M2GNIG in the case of near-exact distributions. We may note that both distributions match four exact moments, but the near-exact distribution has always lower values of the proximity measures. The asymptotic distribution Box-Anderson, which does not equate any moment, has almost always the highest values for the proximity measures, mainly for smaller sample sizes.

In a more detailed comparative analysis between asymptotic and near-exact distributions, we may see that the best asymptotic distribution (the M2G distribution, which equates four exact moments) is always worse than the least performant near-exact distribution (the GNIG distribution, which equates two moments). The difference is more visible for smaller samples, what therefore enhances the advantage of the near-exact distributions over the asymptotic, with regard to smaller samples. For large samples the asymptotic distributions have a relative improvement in the quality of approximation which is however not enough to beat the near-exact distributions. In addition, when the difference $n - p$ decreases, the near-exact distributions are still much closer to the exact distribution, even when the number of sets of variables increases (compare the values of proximity measures between distributions in Tables 3 and 5).

For the same sample size, an increase in the total number of variables leads to an increase in the values of the proximity measures for the asymptotic distributions. This instability of asymptotic distributions contrasts with the behavior of near-exact distributions, whose values of proximity measures decrease in this case (compare, for example, Tables 3 and 4). The near-exact distributions always have a better performance than the asymptotic ones. They lay closer to the exact distribution than the asymptotic ones, namely for smaller sample sizes.

Some quantiles, for the distributions and cases in Tables 3 through 5, are presented in Appendix B, where we consider the first fifteen decimal places of quantiles to assess the precision and performance of the approximations proposed. Note that smaller values of the proximity measures are generally associated with smaller differences between the exact and approximate quantiles. Thus, although we do not have the exact quantiles for the examples presented, we can compare the quantiles of different approximations with the quantiles of the near-exact distribution M2GNIG (for $n = 25$ or $n = 100$), since this approximation has lower values of Δ_1 and Δ_2.

Table 3. Values of measures Δ_1 and Δ_2 for asymptotic and near-exact distributions. Case $m = 3$ with $p_1 = 5$, $p_2 = 7$, $p_3 = 3$; $n = 25$ and $n = 100$.

Distributions		**Proximity measures**			
		$n = 25$		$n = 100$	
		Δ_1	Δ_2	Δ_1	Δ_2
Asymptotic	Box-Anderson(0 m.)	8.815E-02	1.063E-02	1.104E-03	2.844E-05
	Gamma(2 m.)	1.371E-02	9.355E-04	2.112E-03	2.620E-05
	GNIG(3 m.)	1.914E-03	1.122E-04	5.029E-04	5.225E-06
	M2G(4 m.)	3.370E-04	1.896E-05	2.053E-06	1.909E-08
Near-exact	GNIG(2 m.)	8.356E-07	5.566E-08	5.581E-07	6.898E-09
	GNIG(3 m.)	2.244E-08	1.262E-09	3.168E-09	3.320E-11
	M2GNIG(4 m.)	6.369E-11	3.135E-12	3.163E-12	7.082E-15

Table 4. Values of measures Δ_1 and Δ_2 for asymptotic and near-exact distributions. Case $m = 3$ with $p_1 = 5$, $p_2 = 7$, $p_3 = 9$; $n = 25$ and $n = 100$.

Distributions		**Proximity measures**			
		$n = 25$		$n = 100$	
		Δ_1	Δ_2	Δ_1	Δ_2
Asymptotic	Box-Anderson(0 m.)	7.795E-01	1.151E-01	4.538E-03	1.597E-04
	Gamma(2 m.)	2.435E-02	3.214E-03	2.114E-03	3.905E-05
	GNIG(3 m.)	4.797E-03	5.451E-04	1.126E-04	1.772E-06
	M2G(4 m.)	1.965E-03	1.944E-04	4.096E-06	5.674E-08
Near-exact	GNIG(2 m.)	6.385E-08	8.140E-09	1.182E-07	2.178E-09
	GNIG(3 m.)	9.273E-10	9.942E-11	4.631E-10	7.235E-12
	M2GNIG(4 m.)	1.416E-12	1.328E-13	3.200E-13	1.299E-14

Table 5. Values of measures Δ_1 and Δ_2 for asymptotic and near-exact distributions. Case $m = 4$ with $p_1 = 5$, $p_2 = 7$, $p_3 = 3$, $p_4 = 6$; $n = 25$ and $n = 100$.

Distributions		**Proximity measures**			
		$n = 25$		$n = 100$	
		Δ_1	Δ_2	Δ_1	Δ_2
Asymptotic	Box-Anderson(0 m.)	8.331E-01	1.673E-01	5.865E-03	2.224E-04
	Gamma(2 m.)	2.352E-02	3.190E-03	1.956E-03	3.819E-05
	GNIG(3 m.)	4.663E-03	5.444E-04	1.044E-04	1.736E-06
	M2G(4 m.)	1.907E-03	1.937E-04	3.872E-06	5.669E-08
Near-exact	GNIG(2 m.)	5.712E-08	7.500E-09	9.509E-08	1.852E-09
	GNIG(3 m.)	8.052E-10	8.900E-11	3.532E-10	5.834E-12
	M2GNIG(4 m.)	1.192E-12	2.077E-14	2.310E-13	9.125E-15

5. Conclusions and final remarks

The near-exact distributions developed are very close to the exact distribution and although some of the general expressions obtained for the c.d.f.'s may seem complicated, they are, in fact, very manageable and easily allow for the calculation of near-exact quantiles and p-values through the use of a symbolic software. Note that even when we have the expressions for the exact p.d.f.'s and c.d.f.'s available from the literature, these are usually only available for specific numbers of variables per set and the expressions are highly complex, since they make use of unsolved integrals and/or series, which render the computation of exact quantiles impossible.

The comparative analysis conducted allowed us to confirm and reinforce the importance of near-exact distributions over the asymptotic ones. Even when we compare asymptotic and near-exact distributions that equate the same number of exact moments we confirm that the near-exact distributions are always closer to the exact distribution. The near-exact distributions are still very close to the exact distribution when the difference between the sample size and the total number of variables, $n-p$, is very small, which is the usual situation where asymptotic distributions work less well. The near-exact distributions developed also display an asymptotic behavior for increasing number of variables.

Among the near-exact distributions considered for the Wilks Λ statistic, for the general case of several sets of variables, the near-exact M2GNIG distribution is the one that allows for the computation of near-exact quantiles closer to the exact ones. So if we want more accuracy, the near-exact distributions, expressed under the form of mixtures, are the best option, because they lie closer to the exact distribution.

The procedure used in this paper may also be applied to obtain near-exact distributions for other likelihood ratio test statistics, used in several multivariate tests, as well as other statistic tests whose exact distributions are usually seen as hard to obtain in a manageable form.

Appendix

A. Box and Anderson asymptotic distributions for the generalized Wilks Λ statistic

In this appendix we present the asymptotic distributions of Box (1949) and Anderson (2003) for the statistic $W=-\ln\Lambda$ and the fact that these two approaches match by terms of certain order.

A.1. Asymptotic distribution of Box for the statistic $W = -\ln\Lambda$

Box (1949) obtained an asymptotic distribution for the statistic $V_1 = \mu W$ ($\mu > 0$), for the general case of m sets of variables, for a sample size of $n+1$, based on a series expansion until terms of order μ^{-2}. After some simple manipulation we get an approximation to the c.d.f. of the r.v. V_1 in the form of

$$P(V_1 \le v) \cong \left(1 - \frac{\alpha_2}{\mu^2}\right) P(\chi^2_f \le v) + \frac{\alpha_2}{\mu^2} P(\chi^2_{f+4} \le v), \tag{34}$$

where χ^2_f is a r.v. with Chi-square distribution with f degrees of freedom and where

$$\alpha_2 = \alpha_2' - \alpha_1'\beta + \frac{f}{4}\beta^2$$

with

$$\left\{\begin{array}{l} \alpha_1' = \frac{1}{24}(2S_3 + 3S_2) \\[2mm] \alpha_2' = \frac{1}{48}(S_4 + 2S_3 - S_2) \end{array}\right. , \tag{35}$$

where

$$S_i = \left(\sum_{k=1}^{m} p_k\right)^i - \sum_{k=1}^{m} p_k^i = p^i - \sum_{k=1}^{m} p_k^i\,,$$

where p_k represents the number of variables in the k-th set, and

$$f = \frac{1}{2} S_2 \tag{36}$$

and where, according to Box (1949), the best choice for β is

$$\beta = \frac{2S_3 + 3S_2}{6S_2}\,. \tag{37}$$

Under these circumstances, μ is given by $\mu = n - \beta = n - \frac{2S_3+3S_2}{6S_2}$.

A.2. Asymptotic distribution of Anderson for the statistic $W = -\ln\Lambda$

Anderson (2003) obtained an asymptotic distribution for the statistic $V_2 = \eta W$, also for the general case of m sets of variables, for a sample size $n+1$, which gives as c.d.f. for the r.v. V_2,

$$P(V_2 \le v) = \left(1 - \frac{\gamma_2}{\eta^2}\right) P(\chi^2_f \le v) + \frac{\gamma_2}{\eta^2} P(\chi^2_{f+4} \le v) + O(\eta^{-3}),$$

where χ^2_f is a r.v. with Chi-square distribution with f degrees of freedom, with f given by (**??**), and where

$$\eta = n + 1 - \frac{9S_2 + 2S_3}{6S_2}$$

and

$$\gamma_2 = \frac{S_4}{48} - \frac{5}{96}S_2 - \frac{(S_3)^2}{72S_2} = \frac{p^4 - \sum_{k=1}^{m} p_k^4}{48} - \frac{5\left(p^2 - \sum_{k=1}^{m} p_k^2\right)}{96} - \frac{\left(p^3 - \sum_{k=1}^{m} p_k^3\right)^2}{72\left(p^2 - \sum_{k=1}^{m} p_k^2\right)}$$

with

$$p = \sum_{k=1}^{m} p_k\,.$$

This distribution agrees, until terms of order η^{-2}, with the distribution in (**??**). We just have to prove that $\eta = \mu$ and $\gamma_2 = \alpha_2$.

In fact,

$$\eta = n + 1 - \frac{9S_2 + 2S_3}{6S_2} = n - \frac{2S_3 + 3S_2}{6S_2} = \mu$$

while, given the definition of α_1' and α_2' in (**??**) and taking into account (**??**) and (**??**), we have

$$\begin{aligned}
\alpha_2 &= \frac{S_4 + 2S_3 - S_2}{48} - \frac{2S_3 + 3S_2}{24}\left(\frac{2S_3 + 3S_2}{6S_2}\right) + \frac{S_2}{8}\left(\frac{2S_3 + 3S_2}{6S_2}\right)^2 \\
&= \frac{S_4 + 2S_3 - S_2}{48} - \frac{4(S_3)^2 + 12S_2S_3 + 9(S_2)^2}{288S_2} \\
&= \frac{S_4}{48} - \frac{5}{96}S_2 - \frac{(S_3)^2}{72S_2} = \gamma_2.
\end{aligned}$$

Appendix

B. Some quantiles of asymptotic and near-exact distributions

In this appendix we have some quantiles for the asymptotic and near-exact distributions presented in Table 1, and for the cases considered in Table 2.

Table B.1. Some quantiles of asymptotic and near-exact distributions, for $m = 3$ with $p_1 = 5$, $p_2 =$ $p_3 = 3$ and $n = 25$.

Distributions		Quantile		
		0.90	0.95	0.99
Asymptotic	Box-Anderson(0 m.)	5.031785461796158	5.323031958069611	5.898005586512672
	Gamma(2 m.)	5.070377562043812	5.370357926673817	5.963786660003066
	GNIG(3 m.)	5.070276333237788	5.372126647829524	5.971982498923960
	M2G(4 m.)	5.070609220349255	5.372523848045243	5.971819900646903
Near-exact	GNIG(2 m.)	5.070602168477183	5.372467807060278	5.971703926691035
	GNIG(3 m.)	5.070602124092140	5.372467665422931	5.971703537687081
	M2GNIG(4 m.)	5.070602126798732	5.372467667053351	5.971703532349906

Table B.2. Some quantiles of asymptotic and near-exact distributions, for $m = 3$ with $p_1 = 5$, $p_2 =$ $p_3 = 3$ and $n = 100$.

Distributions		Quantile		
		0.90	0.95	0.99
Asymptotic	Box-Anderson(0 m.)	0.935323168711130	0.989715419238025	1.097231449665216
	Gamma(2 m.)	0.935339192611802	0.989726082542498	1.097223554713795
	GNIG(3 m.)	0.935342850877693	0.989738547373362	1.097259383804981
	M2G(4 m.)	0.935340711254230	0.989737150205773	1.097263392208844
Near-exact	GNIG(2 m.)	0.935340709285214	0.989737142448385	1.097263384191019
	GNIG(3 m.)	0.935340708748366	0.989737139450099	1.097263374024131
	M2GNIG(4 m.)	0.935340708764024	0.989737139462930	1.097263374001226

Table B.3. Some quantiles of asymptotic and near-exact distributions, for $m = 3$ with $p_1 = 5$, $p_2 =$
$p_3 = 9$ and $n = 25$.

Distributions		**Quantile**		
		0.90	0.95	0.99
Asymptotic	Box-Anderson(0 m.)	11.591586686879699	12.032350722190023	12.89841425069464
	Gamma(2 m.)	12.345918745169811	12.896339241630456	13.97171235654050
	GNIG(3 m.)	12.348171464976944	12.912166618988656	14.02700835429970
	M2G(4 m.)	12.348171464976944	12.912166618988656	14.02700835429970
Near-exact	GNIG(2 m.)	12.348022879983701	12.910964952132374	14.02458359609938
	GNIG(3 m.)	12.348022863197591	12.910964910723551	14.02458349791948
	M2GNIG(4 m.)	12.348022863501334	12.910964910801998	14.02458349704613

Table B.4. Some quantiles of asymptotic and near-exact distributions, for $m = 3$ with $p_1 = 5$, $p_2 =$
$p_3 = 9$ and $n = 100$.

Distributions		**Quantile**		
		0.90	0.95	0.99
Asymptotic	Box-Anderson(0 m.)	1.836423568798852	1.912728356791496	2.061422164719888
	Gamma(2 m.)	1.836561835231905	1.912879136056215	2.061589621633479
	GNIG(3 m.)	1.836566487919830	1.912904876661781	2.061678211637670
	M2G(4 m.)	1.836567768617353	1.912905919850552	2.061676226936147
Near-exact	GNIG(2 m.)	1.836567748918097	1.912905868635871	2.061676170614511
	GNIG(3 m.)	1.836567748527033	1.912905867113034	2.061676165947668
	M2GNIG(4 m.)	1.836567748531739	1.912905867116187	2.061676165938647

Table B.5. Some quantiles of asymptotic and near-exact distributions, for $m = 4$ with $p_1 = 5$, $p_2 =$
$p_3 = 3$, $p_4 = 6$ and $n = 25$.

Distributions		Quantile		
		0.90	0.95	0.99
Asymptotic	Box-Anderson(0 m.)	12.460941411465933	12.907514159294704	13.78390061885836
	Gamma(2 m.)	13.296058071513843	13.856824747820833	14.95011724014824
	GNIG(3 m.)	13.295201884762036	13.868568563988037	15.00677256704869
	M2G(4 m.)	13.298577869482085	13.873174216085569	15.00638700746535
Near-exact	GNIG(2 m.)	13.298396069376416	13.871917301706467	15.00395005153875
	GNIG(3 m.)	13.298396053550118	13.871917262268838	15.00394995723122
	M2GNIG(4 m.)	13.298396053835702	13.871917262346372	15.00394995641449

Table B.6. Some quantiles of asymptotic and near-exact distributions, for $m = 4$ with $p_1 = 5$, $p_2 =$
$p_3 = 3$, $p_4 = 6$ and $n = 100$.

Distributions		Quantile		
		0.90	0.95	0.99
Asymptotic	Box-Anderson(0 m.)	2.039136431292056	2.119137075841269	2.274715205452883
	Gamma(2 m.)	2.039331328087639	2.119352060184476	2.274962859788668
	GNIG(3 m.)	2.039336486121450	2.119378976623912	2.275054213119008
	M2G(4 m.)	2.039337803157476	2.119380026799260	2.275052099885275
Near-exact	GNIG(2 m.)	2.039337781112693	2.119379971964562	2.275052042593969
	GNIG(3 m.)	2.039337780749925	2.119379970585481	2.275052038400520
	M2GNIG(4 m.)	2.039337780753908	2.119379970588093	2.275052038392689

Acknowledgments

The authors wish to thank the financial support of the Portuguese Foundation for Science and Technology (FCT), through CMA/FCT, the Center for Mathematics and Its Applications of Faculdade de Ciências e Tecnologia of Universidade Nova de Lisboa ('Financiamento Base' 2009 ISFL-1-297 from FCT/MCTES/PT).

REFERENCES

[1] M.Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions,* (eds.), 9-th printing, Dover, New York 1974.

[2] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis,* 3.rd ed., J. Wiley & Sons, New York 2003.

[3] G.E.P. Box, *A general distribution theory for a class of likelihood criteria*, Biometrika **36** (1949), 317–346.

[4] C.A. Coelho, *The generalized integer Gamma distribution – a basis for distributions in Multivariate Statistics*, Journal of Multivariate Analysis **64** (1998), 86–102.

[5] C.A. Coelho, *A generalized Integer Gamma distribution as an asymptotic replacement for a Logbeta distribution & applications*, American Journal of Mathematical Management Sciences **23** (2003), 383–399.

[6] C.A. Coelho, *The generalized near-integer Gamma distribution: a basis for 'near-exact' approximations to the distribution of statistics which are the product of an odd number of independent Beta random variables*, Journal of Multivariate Analysis **89** (2004), 191–218.

[7] L.M. Grilo, *Development of near-exact distributions for different scenarios of application of the Wilks Lambda statistic* (in Portuguese), Ph.D. thesis, Technical University of Lisbon, Portugal 2005.

[8] L.M. Grilo and C.A. Coelho, *Development and Comparative Study of two Near-exact Approximations to the Distribution of the Product of an Odd Number of Independent Beta Random Variables,* Journal of Statistical Planning and Inference **137** (2007), 1560–1575.

[9] L.M. Grilo and C.A. Coelho, *The exact and near-exact distribution for the Wilks Lambda statistic used in the test of independence of two sets of variables*, American Journal of Mathematical and Management Sciences (2010), (in print).

[10] N.L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 2, J. Wiley & Sons, New York 1995.

[11] S.S. Wilks, *Certain generalizations in the analysis of variance*, Biometrika **24** (1932), 471–494.

[12] S.S. Wilks, *On the independence of k sets of normally distributed statistical variables*, Econometrika **3** (1935), 309–326.

Received 15 March 2010