A family of near-exact distributions based on truncations of the exact distribution for the generalized Wilks Lambda statistic[♦]

L. M. Grilo, *Mathematics Department of Instituto Politécnico de Tomar ‘Centro de Matemática e Aplicações’ of Faculdade de Ciências e Tecnologia,Universidade Nova de Lisboa, Portugal.*
Email: lgrilo@ipt.pt

C. A. Coelho[1], *Mathematics Department and ‘Centro de Matemática e Aplicações’ Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal.*
Email: cmac@fct.unl.pt

Abstract

Since for the case where at least two sets have an odd number of variables we do not have the exact distribution of the generalized Wilks Lambda statistic in a manageable form, adequate for manipulation, we develop in this paper a family of very accurate near-exact distributions for this statistic for the case where two or three sets have an odd number of variables. We first express the exact characteristic function of the logarithm of the statistic in the form of the characteristic function of an infinite mixture of Generalized Integer Gamma distributions. Then, based on truncations of this exact characteristic function, we obtain a family of near-exact distributions, which, by construction, match the first two exact moments. These near-exact distributions display an asymptotic behaviour for increasing number of variables involved. The corresponding cumulative distribution functions are obtained in a concise and manageable form, relatively easy to implement computationally, allowing for the computation of virtually exact quantiles. We undertake a comparative study for small sample sizes, using two proximity measures based on the Berry-Esseen bounds, to assess the performance of the near-exact distributions for different numbers of sets of variables and different numbers of variables in each set.

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

Key Words: Independent random variables, characteristic function, sum of Gamma, mixtures, proximity measures, small sample sizes.

1. Introduction

When two or more sets, among the *m* sets, have an odd number of variables, we do not have the exact distribution of the generalized Wilks Lambda statistic (Wilks, 1932, 1935), used in

---

[♦]This research received the financial support of the Portuguese Foundation for Science and Technology through grant 2010 ISFL-1-297 from FCT/MCTES/PT.

[1] Corresponding author: Dep. Matemática, FCT/UNL, Campus de Caparica, 2829-516, Caparica, Portugal

multivariate analysis to test the independence among m ($m \geq 2$) sets of variables, under the assumption of normality (and indeed also under the assumption of elliptically contoured or left orthogonal-invariant distributions), in a manageable form, adequate for further manipulation. The aim of this paper is thus to develop a family of very accurate near-exact distributions, which, by construction, will match the first two exact moments. The concept of near-exact distribution has already been introduced in a number of papers (Coelho, 2003, 2004; Grilo, 2005; Grilo and Coelho, 2007, 2010). These are distributions which lay very 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 and which may be developed either based on factorizations or based on truncations of the exact distribution. In addition, these distributions are relatively easy to implement computationally, allowing for an easy calculation of near-exact quantiles.

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

$$\underline{X} = [\underline{X}'_1, \ldots, \underline{X}'_k, \ldots, \underline{X}'_m]' \sim N_p\left(\underline{\mu}, \Sigma\right), \tag{1}$$

where the population mean vector and population variance-covariance matrix are given by

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

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

$$H_0 : \Sigma = diag(\Sigma_{11}, \ldots, \Sigma_{kk}, \ldots, \Sigma_{mm}), \tag{2}$$

is the generalized Wilks Λ statistic

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

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 $\underline{X}_k$ $(k = 1, \ldots, m)$.

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

$$\Lambda = \prod_{k=1}^{m-1} \Lambda_{k(k+1,\ldots,m)} \tag{3}$$

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 subvectores $k+1$ 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,

$$H_0^{(k)}:[\Sigma_{k,k+1}\ldots\Sigma_{km}] = 0_{p_k \times (p_{k+1}+\ldots+p_m)}, \quad k=1,\ \ldots,\ m\text{-}1. \tag{4}$$

Then, using the result in Theorem 9.3.3 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 (3), under the null hypothesis $H_0^{(k)}$ in (4), is the same as $\prod_{j=1}^{p_k} X_j$, where X_j are p_k independents r.v.'s with

$$X_j \sim Beta\left(\frac{n+1-q_k-j}{2}, \frac{q_k}{2}\right), \qquad j=1,\ldots,p_k \tag{5}$$

for a sample of size *n* + 1 (with $n \ge p_1+\ldots+p_m$) and where $q_k = p_{k+1}+\ldots+p_m$.

Concerning the distribution of the generalized Wilks Lambda statistic we have to consider two cases: when there is at most one set with an odd number of variables among the *m* sets we have the exact distribution obtained by Coelho (1998) in the form of a Generalized Integer Gamma (GIG) distribution; and, when there are at least two sets with an odd number of variables among the *m* sets, situation in which we do not have the exact distribution of the generalized Wilks Lambda statistic in a manageable form, adequate for further manipulation. For this second case there are a few asymptotic distributions (Box, 1949; Coelho, 2000; Anderson, 2003; Grilo and Coelho, 2010) and some near-exact distributions based on factorizations of the exact c.f. (Coelho, 2003, 2004; Coelho *et al*, 2006; Grilo, 2005; Grilo and Coelho, 2010). In this paper we study the development of near-exact distributions for the generalized Wilks Λ statistic specifically adapted for the case where there are two or three sets with an odd number of variables. We first express the exact c.f. of the logarithm of the generalized Wilks Lambda statistic under the form of an infinite mixture of GIG distributions, and then, based on truncations of this exact c.f., we obtain a family of near-exact distributions, as a finite mixture of a number of GIG distributions and a Generalized Near-Integer Gamma (GNIG) distribution, which, by construction, will match the two first exact moments.

In Section 2 we introduce some distributions used in our work (GIG, GNIG and Logbeta distribution). In Section 3 we express the exact distribution of the generalized Wilks Lambda statistic under the form of an infinite mixture of GIG distributions and then, based on truncations of this exact c.f., we obtain in Section 4, a family of near-exact distributions for this statistic. In Section 5, we use two measures of proximity to assess the performance of the near-exact distributions developed. In Section 6, we compare a member of the family of near-exact distributions, based on truncations, with the exact distribution and also with a near-exact distribution based on factorizations, developed in Grilo and Coelho (2010). Finally, we provide some concluding comments in Section 7.

2. Some distributions used

We introduce here the GIG, GNIG and Logbeta distributions, which will be used in the sections ahead.

Let Z be a r.v. with a GIG distribution of depth g, 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 $\lambda_1,\ldots,\lambda_g \in \mathbb{R}^+$ (Coelho, 1998, 2003), denoted by

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

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

$$f_Z(z) = K\sum_{i=1}^{g} P_i(z)\mathrm{e}^{-\lambda_i z}, \qquad (z>0) \tag{6}$$

where

$$K = \prod_{i=1}^{g}\lambda_i^{r_i} \tag{7}$$

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

$$P_i(z) = \sum_{k=1}^{r_i} c_{i,k}\, z^{k-1} \tag{8}$$

where

$$c_{i,r_i} = \frac{1}{(r_i-1)!}\prod_{\substack{j=1\\ j\neq i}}^{g}(\lambda_j-\lambda_i)^{-r_j} \tag{9}$$

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

$$c_{i,r_i-k} = \frac{1}{k}\sum_{j=1}^{k}\frac{(r_i-k+j-1)!}{(r_i-k-1)!}R(j-1,i)c_{i,r_i-(k-j)}\,, \tag{10}$$

where

$$R(n,j) = \sum_{\substack{i=1\\ i\neq j}}^{g} r_i(\lambda_j-\lambda_i)^{-n-1}, \quad (n=0,\ldots,r_i-1)\,. \tag{11}$$

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

$$F_Z(z) = K\sum_{i=1}^{g}P_i^*(z), \qquad (z>0) \tag{12}$$

with K given by (7) and where

$$P_i^*(z) = \sum_{k=1}^{r_i} c_{i,k}\,\frac{(k-1)!}{\lambda_i^k}\left[1-\left(\sum_{j=0}^{k-1}\frac{\lambda_i^j z^j}{j!}\right)\mathrm{e}^{-\lambda_i z}\right] \tag{13}$$

with $c_{i,k} \quad (i=1,\ldots,g; k=1,\ldots,r_i)$ given by (9) through (11).

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

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

Let us consider, now, $Z \sim GIG(r_1,...,r_g;\lambda_1,...,\lambda_g)$ and $X \sim Gamma(r,\lambda)$ two independent r.v.'s with $r \in \mathbb{R}^+ \setminus \mathbb{N}$ and $\lambda \neq \lambda_j$, $\forall_j \in \{j=1,...,g\}$. Then the r.v. $W = Z + X$ has a GNIG distribution with depth *g* + 1 (Coelho, 2004). Symbolically,

$$W \sim GNIG(r_1,...,r_g,r;\lambda_1,...,\lambda_g,\lambda)\,. \tag{14}$$

The p.d.f. of *W* is given by

$$f_W(w) = K\lambda^r \sum_{j=1}^{g} \mathrm{e}^{-\lambda_j w} \sum_{k=1}^{r_j} \left\{ c_{j,k} \frac{\Gamma(k)}{\Gamma(k+r)} w^{k+r-1} {}_1F_1(r,k+r,-(\lambda-\lambda_j)w) \right\}, \qquad (w>0) \tag{15}$$

and the c.d.f. by

$$\begin{aligned} F_W(w) = {} & \lambda^r \frac{w^r}{\Gamma(r+1)} {}_1F_1(r,r+1,-\lambda w) \\ & - K\lambda^r \sum_{j=1}^{g} \mathrm{e}^{-\lambda_j w} \sum_{k=1}^{r_j} c^*_{j,k} \sum_{i=0}^{k-1} \frac{w^{r+i}\lambda_j^i}{\Gamma(r+1+i)} {}_1F_1(r,r+1+i,-(\lambda-\lambda_j)w), \qquad (w>0) \end{aligned} \tag{16}$$

Where

$$K = \prod_{j=1}^{g} \lambda_j^{r_j} \quad \text{and} \quad c^*_{j,k} = \frac{c_{jk}}{\lambda_j^k}\Gamma(k) \tag{17}$$

with $c_{j,k}$ given by (9) through (11). 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). Nowadays we may find these functions, which have good convergence properties, in several software packages, like Mathematica™.

The c.f. of *W* is given by

$$\varphi_W(t) = \lambda^r (\lambda - \mathrm{i}t)^{-r} \prod_{j=1}^{g} \lambda_j^{r_j} (\lambda_j - \mathrm{i}t)^{-r_j}\,, \tag{18}$$

where $r \in \mathbb{R}^+ \setminus \mathbb{N}$, $\lambda \in \mathbb{R}^+$, $r_j \in \mathbb{N}$ and $\lambda \neq \lambda_j$, $\forall j \in \{1,...,g\}$. If $r \in \mathbb{N}$ then the GNIG distribution of depth $g+1$ reduce into a GIG distribution of depth $g+1$. This way we may look at the GNIG distribution as a generalization of the GIG distribution.

Let *X* be a r.v. with Beta distribution, with parameters $\alpha > 0$ and $\beta > 0$, what we denote by

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

The *h*-th moment of *X* is

$$E(X^h) = \frac{B(\alpha+h,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+h)}{\Gamma(\alpha+\beta+h)} \qquad (h > -\alpha)\,. \tag{19}$$

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

$$Y \sim Logbeta(\alpha,\beta)\,. \tag{20}$$

The p.d.f. of Y is

$$f_Y(y) = \frac{1}{B(\alpha,\beta)}\,\mathrm{e}^{-\alpha y}(1-\mathrm{e}^{-y})^{\beta-1}, \qquad (y>0)\,. \tag{21}$$

Since the Gamma functions in (19) are still defined for any complex h (in strict sense), the c.f. of Y is given by

$$\varphi_Y(t) = E(\mathrm{e}^{\mathrm{i}tY}) = E(\mathrm{e}^{-\mathrm{i}t\ln X}) = E(X^{-\mathrm{i}t}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\frac{\Gamma(\alpha-\mathrm{i}t)}{\Gamma(\alpha+\beta-\mathrm{i}t)} \tag{22}$$

where $\mathrm{i}=(-1)^{1/2}$ and $t\in\mathbb{R}$.

We may also write the c.f. of Y in (20) as

$$\varphi_Y(t) = \frac{1}{B(\alpha,\beta)}\frac{1}{\Gamma(1-\beta)}\sum_{j=0}^{\infty}\frac{\Gamma(1-\beta+j)}{j!\ (\alpha+j)}(\alpha+j)(\alpha+j-\mathrm{i}t)^{-1}\,, \tag{23}$$

which is an infinite mixture of the c.f.'s of a $Gamma(1,\alpha+j)\equiv Exponential(\alpha+j)$ distribution with weights given by (Coelho et al., 2006),

$$\pi_j = \frac{1}{B\left(\alpha,\beta\right)}\frac{\Gamma\left(1-\beta+j\right)}{\Gamma\left(1-\beta\right)j!(\alpha+j)} \quad (j=0,\ldots).$$

3. The exact distribution of generalized Wilks Lambda statistic when a maximum of three sets have an odd number of variables

In Theorem 1 we present the exact distribution of the generalized Wilks Lambda statistic for the case where two or three of the sets have an odd number of variables, expressed in the form of an infinite mixture of GIG distributions.

Theorem 1 *When there are two or at most three sets of variables with an odd number of variables among the m sets, then under (2) and for a sample of size n+1, the exact distribution of W = - ln Λ is an infinite mixture of GIG distributions,* $\frac{q_{m-2}}{2}-2$ *of them with depth* $p-2$ *and the remaining with depth* $p-1$*, symbolically,*

$$W \sim MGIG\left(\frac{1}{B\left(c,\frac{p_m}{2}\right)\Gamma\left(1-\frac{p_m}{2}\right)}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)};r^{**}_{\nu 1},\ldots,r^{**}_{\nu,p-2};\lambda_1,\ldots,\lambda_{p-2}\left(\nu=1,\ldots,\frac{q_{m-2}}{2}-2\right);\right.$$
$$\left. r^*_1,\ldots,r^*_{p-1};\lambda_1,\ldots,\lambda_{p-1}\left(\nu=0,\nu\geq\frac{q_{m-2}}{2}-1\right)\right)$$

where $c=\frac{n+1-q_{m-2}}{2}$ *and* $q_{m-2}=p_{m-1}+p_m$*. The shape parameters* $r^{**}_{\nu j}$ *are given by*

$$r_{\nu j}^{**}=\sum_{k=1}^{m-2} r_{k,j-p_k^*}+r_{m-1,j-p_k^*}^{*}\ ,\quad j=1,\ldots,p-2\,, \tag{24}$$

where

$$r_j^{*}=\sum_{k=1}^{m-2} r_{k,j-p_k^*}+r_{m-1,j-p_k^*}\ ,\quad j=1,\ldots,p-2 \tag{25}$$

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

$$\begin{aligned} &r_{k,j-p_k^*}=0 && \text{if}\quad p_k^*\ge j\\ &r_{m-1,j-p_k^*}^{*}=r_{m-1,j-p_k^*}=0 && \text{if}\quad p_k^*\ge j\quad \text{or}\quad j=p-2, \end{aligned} \tag{26}$$

where, for $k=1,\ldots,m-2$ *and* $q_k=p_{k+1}+\ldots+p_m$,

$$r_{kj}=\begin{cases} h_{kj} & j=1,2\\ r_{k,j-2}+h_{kj} & j=3,\ldots,p_k+q_k-2 \end{cases} \tag{27}$$

with

$$h_{kj}=(\textit{number of elements of }\{p_k,q_k\}\ge j)-1\,. \tag{28}$$

For $\nu=0$ *or* $\nu\ge \frac{q_{m-2}}{2}-1$

$$r_{m-1,j}=\begin{cases} h_j & j=1,\ 2\\ r_{j-2}+h_j & j=3,\ldots,q_{m-2}-3 \end{cases} \tag{29}$$

with

$$h_j=(\textit{number of elements of }\{p_{m-1}-1,p_m\}\ge j)-1\,, \tag{30}$$

and the rate parameters λ_j *are given by*

$$\lambda_j=\frac{n-p+j}{2}\qquad (j=1,\ldots,p-2) \tag{31}$$

and $\lambda_{p-1}=c+\nu$ *; while for* $\nu=1,\ldots,\frac{q_{m-2}}{2}-2$

$$r_{m-1,j}^{*}=\begin{cases} r_{m-1,j}+1 & j=2\nu\\ r_{m-1,j} & j\in\{1,\ldots,q_{m-2}-3\}\setminus\{2\nu\}, \end{cases} \tag{32}$$

with $r_{m-1,j}$ *given by (29) and (30), since the rates parameters* λ_j $(j=1,\ldots,p-2)$ *are given by (31).*

Proof. From (19) and the independence of the p_k r.v.'s X_j in (5), under the null hypothesis $H_0^{(k)}$ in (4), we may write

$$E[\Lambda_{k(k+1,\ldots,m)}^h]=\prod_{j=1}^{p_k}E(X_j^h)=\prod_{j=1}^{p_k}\frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)}\frac{\Gamma\left(\frac{n+1-q_k-j}{2}+h\right)}{\Gamma\left(\frac{n+1-j}{2}+h\right)},\quad \left(h>-\frac{n+1-q_k-p_k}{2}\right).$$

Then, given the independence of $m-1$ statistics $\Lambda_{k(k+1,\ldots,m)}$ in (3), under the null hypothesis of independence of the *m* sets of variables in (2) (Anderson, 2003, Theorem 9.3.2), we obtain the *h*-th moment of generalized Wilks Λ statistic, for a sample size *n* + 1 , as

$$E(\Lambda^h)=\prod_{k=1}^{m-1}E[\Lambda_{k(k+1,\dots,m)}^h]=\prod_{k=1}^{m-1}\prod_{j=1}^{p_k}E(X_j^h)=\prod_{k=1}^{m-1}\prod_{j=1}^{p_k}\frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}+h\right)}\frac{\Gamma\left(\frac{n+1-q_k-j}{2}+h\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)} \tag{33}$$

and, since the Gamma functions in (33) 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

$$\varphi_W(t)=E(\mathrm{e}^{\mathrm{i}tW})=E(\mathrm{e}^{-\mathrm{i}t\ln\Lambda})=E(\Lambda^{-\mathrm{i}t})=\prod_{k=1}^{m-1}\prod_{j=1}^{p_k}\frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)}\frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)} \tag{34}$$

where $\mathrm{i}=(-1)^{1/2}$ and $t\in\mathbb{R}$ (being $\mathbb{R}$ the set of reals). To factorize the c.f. of *W* in (34) we consider, without any loss of generality, that among the *m* sets of variables the last two or three sets are the ones that have an odd number of variables. Then, we may write the c.f. of *W* as

$$\begin{aligned}\varphi_W(t)=&\underbrace{\prod_{k=1}^{m-3}\prod_{j=1}^{p_k}\frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)}\frac{\Gamma\left(\frac{n+1-q_k-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_k-j}{2}\right)}}_{p_k\ \text{even}}\\
&\times\underbrace{\prod_{j=1}^{p_{m-2}}\frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)}\frac{\Gamma\left(\frac{n+1-q_{m-2}-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_{m-2}-j}{2}\right)}}_{q_{m-2}\ \text{even}}\\
&\times\underbrace{\prod_{j=1}^{p_{m-1}}\frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)}\frac{\Gamma\left(\frac{n+1-q_{m-1}-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_{m-1}-j}{2}\right)}}_{p_{m-1}\ \text{and}\ q_{m-1}\ \text{odd}}.\end{aligned}$$

Using the equality

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

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

$$\begin{aligned}\varphi_W(t)=&\underbrace{\prod_{k=1}^{m-3}\prod_{j=1}^{p_k+q_k-2}\left(\frac{n-p_k-q_k+j}{2}\right)^{r_{kj}}\left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r_{kj}}}_{p_k\ \text{even}}\\
&\times\underbrace{\prod_{j=1}^{p_{m-2}+q_{m-2}-2}\left(\frac{n-p_{m-2}-q_{m-2}+j}{2}\right)^{r_{m-2,j}}\left(\frac{n-p_{m-2}-q_{m-2}+j}{2}-\mathrm{i}t\right)^{-r_{m-2,j}}}_{q_{m-2}\ \text{even}}\\
&\times\underbrace{\prod_{j=1}^{p_{m-1}}\frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)}\frac{\Gamma\left(\frac{n+1-q_{m-1}-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_{m-1}-j}{2}\right)}}_{p_{m-1}\ \text{and}\ q_{m-1}\ \text{odd}}\end{aligned} \tag{35}$$

where the parameters r_{kj} $(k=1,\dots,m-2;\, j=1,\dots,p_k+q_k-2)$ are given by (27) and (28) and $q_{m-1}=p_m$. We may write the part of $\varphi_W(t)$ corresponding to the third factor in (35), where p_{m-1} and q_{m-1} are both odd, as

$$
\begin{aligned}
&\prod_{j=1}^{p_{m-1}} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_{m-1}-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_{m-1}-j}{2}\right)} \\
&= \frac{\Gamma\left(\frac{n+1-p_{m-1}}{2}\right)}{\Gamma\left(\frac{n+1-p_{m-1}}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_{m-1}-p_{m-1}}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_{m-1}-p_{m-1}}{2}\right)} \prod_{j=1}^{p_{m-1}-1} \frac{\Gamma\left(\frac{n+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_{m-1}-j}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_{m-1}-j}{2}\right)} \\
&= \frac{\Gamma\left(\frac{n+1-p_{m-1}}{2}\right)}{\Gamma\left(\frac{n+1-p_{m-1}}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_{m-2}}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_{m-2}}{2}\right)} \prod_{j=1}^{q_{m-2}-3} \left(c+\frac{j}{2}\right)^{r_{m-1,j}} \left(c+\frac{j}{2}-\mathrm{i}t\right)^{-r_{m-1,j}} \\
&= \frac{\Gamma\left(\frac{n+1-p_{m-1}}{2}\right)}{\Gamma\left(\frac{n+1-p_{m-1}}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{n+1-q_{m-2}}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n+1-q_{m-2}}{2}\right)} \prod_{j=1}^{q_{m-2}-3} \left(\frac{n+1-q_{m-2}+j}{2}\right)^{r_{m-1,j}} \left(\frac{n+1-q_{m-2}+j}{2}-\mathrm{i}t\right)^{-r_{m-1,j}}
\end{aligned}
\tag{36}
$$

where the parameters $r_{m-1,j}$ are given by (29) and (30). Then replacing, in this c.f. expression, the part corresponding to the c.f. of a Logbeta distribution with parameters $\frac{n+1-q_{m-2}}{2}$ and $\frac{q_{m-1}}{2}$ by its representation in the form of an infinite mixture of Exponential distributions in (22), we may write the c.f. of W as

$$
\begin{aligned}
\varphi_W(t) = &\prod_{k=1}^{m-3} \underbrace{\prod_{j=1}^{p_k+q_k-2} \left(\frac{n-p_k-q_k+j}{2}\right)^{r_{kj}} \left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r_{kj}}}_{p_k \text{ even}} \\
&\times \underbrace{\prod_{j=1}^{p_{m-2}+q_{m-2}-2} \left(\frac{n-p_{m-2}-q_{m-2}+j}{2}\right)^{r_{m-2,j}} \left(\frac{n-p_{m-2}-q_{m-2}+j}{2}-\mathrm{i}t\right)^{-r_{m-2,j}}}_{q_{m-2} \text{ even}} \\
&\times \prod_{j=1}^{q_{m-2}-3} \left(\frac{n-p_{m-1}-q_{m-1}+j}{2}\right)^{r_{m-1,j}} \left(\frac{n-p_{m-1}-q_{m-1}+j}{2}-\mathrm{i}t\right)^{-r_{m-1,j}} \\
&\times \left[\frac{1}{B\left(c,\frac{p_m}{2}\right)} \frac{1}{\Gamma\left(1-\frac{p_m}{2}\right)} \sum_{\nu=0}^{\infty} \frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)} (c+\nu)(c+\nu-\mathrm{i}t)^{-1}\right] \\
= &\left[\frac{1}{B\left(c,\frac{p_m}{2}\right)} \frac{1}{\Gamma\left(1-\frac{p_m}{2}\right)} \sum_{\nu=0}^{\infty} \frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)} (c+\nu)(c+\nu-\mathrm{i}t)^{-1}\right] \\
&\times \prod_{j=1}^{p-2} \left(\frac{n-p+j}{2}\right)^{r_j^*} \left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_j^*}
\end{aligned}
$$

where the parameters r_j^* are given by (25). This way, we may finally write

$$\varphi_W(t)=\frac{1}{B\left(c,\frac{p_m}{2}\right)}\frac{1}{\Gamma\left(1-\frac{p_m}{2}\right)}\left[\sum_{\nu=1}^{\frac{q_{m-2}}{2}-2}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)}\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_{\nu j}^{**}}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_{\nu j}^{**}}\right.$$
$$\left.+\sum_{\nu=0,\nu\geq\frac{q_{m-2}}{2}-1}^{\infty}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)}(c+\nu)(c+\nu-\mathrm{i}t)^{-1}\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_j^*}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_j^*}\right] \qquad (37)$$

which is the c.f. of an infinite mixture of GIG distributions, $\frac{q_{m-2}}{2}-2$ of them with depth $p-2$, with shape parameters $r_{\nu j}^{**}$ given by (24) and rate parameters λ_j given by (31) $(\nu=1,\ldots,\frac{q_{m-2}}{2}-2;\ j=1,\ldots,p-2)$, while the remaining GIG distributions have depth $p-1$, with shape parameters r_j^* given by (25) and $r_{p-1}^*=1$, and the rate parameters λ_j given by (31) and $\lambda_{p-1}=c+\nu$ $(\nu=0,\nu\geq\frac{q_{m-2}}{2}-1;\ j=1,\ldots,p-2)$. ■

We may note that, if $m=2$ then $q_{m-2}=q_0=p_1+p_2=p$ and we consider that any product with an upper limit smaller than its lower limit evaluates to one. In this case we have in (24) and (25), respectively, $r_{\nu j}^{**}=r^*_{m-1,j-p_k^*}$ and $r_j^*=r_{m-1,j-p_k^*}$, while in (35) the first product vanishes, or rather, evaluates to 1, as well as the second product, where we take $r_{0,j}=0$ for any *j*; we are then left with only the third product in the case where the only two sets have an odd number of variables.

From (37) and (6)-(13), the exact p.d.f. of Λ is, for $c=\frac{n+1-q_{m-2}}{2}$,

$$f_\Lambda(u)=\frac{1}{B\left(c,\frac{p_m}{2}\right)\Gamma\left(1-\frac{p_m}{2}\right)}\left[\sum_{\nu=1}^{\frac{q_{m-2}}{2}-2}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)}K_\nu\sum_{j=1}^{p-2}P_{\nu j}(-\ln u)u^{\lambda_j}\right.$$
$$\left.+\sum_{\nu=0,\nu\geq\frac{q_{m-2}}{2}-1}^{\infty}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)}K\sum_{j=1}^{p-1}P_j(-\ln u)u^{\lambda_j}\right] \qquad (0<u<1)$$

where,

$$K_\nu=\prod_{j=1}^{p-2}\lambda_j^{r_{\nu j}^{**}}\quad,\qquad P_{\nu j}(-\ln u)=\sum_{i=1}^{r_{\nu j}^{**}}c_{\nu j,i}(-\ln u)^{i-1} \qquad (38)$$

and

$$K=\prod_{j=1}^{p-1}\lambda_j^{r_j^*}\quad,\qquad P_j(-\ln u)=\sum_{m=1}^{r_j^*}c_{j,m}(-\ln u)^{m-1}, \qquad (39)$$

while the exact c.d.f. is

$$F_\Lambda(u)=\frac{1}{B\left(c,\frac{p_m}{2}\right)\Gamma\left(1-\frac{p_m}{2}\right)}\left[\sum_{\nu=1}^{\frac{q_{m-2}}{2}-2}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)}K_\nu\sum_{j=1}^{p-2}P^*_{\nu j}(-\ln u)\right.$$

$$\left.+\sum_{\nu=0,\nu\geq\frac{q_{m-2}}{2}-1}^{\infty}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)}K\sum_{j=1}^{p-1}P^*_j(-\ln u)\right]\qquad(0<u<1)$$

where,

$$P^*_{\nu j}(-\ln u)=\sum_{i=1}^{r^{**}_{\nu j}}c_{\nu j,i}\,\frac{(i-1)!}{\lambda_j^i}\left[1-\left(\sum_{i^*=0}^{i-1}\frac{\lambda_j^{i^*}(-\ln u)^{i^*}}{i^*!}\right)u^{\lambda_j}\right],\qquad(40)$$

and

$$P^*_j(-\ln u)=\sum_{m=1}^{r^*_j}c_{j,m}\,\frac{(m-1)!}{\lambda_j^m}\left[1-\left(\sum_{i=0}^{m-1}\frac{\lambda_j^{i}(-\ln u)^{i}}{i!}\right)u^{\lambda_j}\right],\qquad(41)$$

and where $c_{\nu\, j,i}$ and $c_{j,m}$ are given by (9)-(11).

4. A family of near-exact distributions for the generalized Wilks Lambda statistic when two or three sets have an odd number of variables

The near-exact distributions developed in this section are based on truncations of the exact distribution of $W=-\ln\Lambda$. They are manageable and expressed as finite mixtures, which equate the two first exact moments and which allow for an easy computation of near-exact quantiles. These distributions are obtained in Theorem 2.

Theorem 2. *When among m sets of variables there are two or at most three sets with an odd number of variables then, under (2) and for a sample of size n+1, a family of near-exact distributions for W = - ln Λ is given by a finite mixture of* n^*+2 *distributions, where* n^*+1 *components are GIG distributions and the last component is a GNIG distribution, with weights*

$$\pi_\nu=\frac{1}{B\left(c,\frac{p_m}{2}\right)}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\Gamma\left(1-\frac{p_m}{2}\right)\nu!(c+\nu)}\quad(\nu=0,1,...n^*)\qquad(42)$$

and

$$\pi_{n^*+1}=1-\sum_{\nu=0}^{n^*}\pi_\nu\ .$$

Symbolically,

$$W\overset{ne}{\sim}M(n^*+1)GIG+GNIG(\pi_\nu;r^*_{\nu1},...,r^*_{\nu,p-1};\lambda_1,...,\lambda_{p-1};\nu=0,...,n^*+1),$$

where the shape and rate parameters are as follows:

- *for* $n^* \le \frac{q_{m-2}}{2} - 2$ *we have a finite mixture of* $n^* + 2$ *distributions, where* $n^* + 1$ *components are GIG distributions (* n^* *with depth* $p-2$ *and one with depth* $p-1$*) and the last component is a GNIG distribution of depth* $p-1$*; for* $\nu = 1, \ldots, n^* < \frac{q_{m-2}}{2} - 2$ *the shape parameters* $r^{**}_{\nu j}$ $(j = 1, \ldots, p-2)$ *are given by (24) through (28) and the rate parameters* λ_j $(j = 1, \ldots, p-2)$ *are given by (31); for* $\nu = 0$ *we have the* $r^{**}_{\nu j} = r^*_j$ $(j = 1, \ldots, p-2)$ *given by (25) through (28) and* $r^*_{p-1} = 1$*, with the rates* λ_j $(j = 1, \ldots, p-2)$ *given by (31) and* $\lambda_{p-1} = c$ *;*

- *for* $n^* > \frac{q_{m-2}}{2} - 2$ *we have a finite mixture of* $n^* + 2$ *distributions, where* $n^* + 1$ *components are GIG distributions,* $\frac{q_{m-2}}{2} - 2$ *with depth* $p-2$ *and the remaining* $n^* + 1 - (\frac{q_{m-2}}{2} - 2)$ *with depth* $p-1$*, and the last component is a GNIG distribution of depth* $p-1$*; for* $\nu = 1, \ldots, \frac{q_{m-2}}{2} - 2$ *the shape parameters* $r^{**}_{\nu j}$ $(j = 1, \ldots, p-2)$ *are given by (24) through (28) and the rates* λ_j $(j = 1, \ldots, p-2)$ *are given by (31), while for* $\nu = 0$ *and* $\nu \ge \frac{q_{m-2}}{2} - 1$ *we have the* $r^{**}_{\nu j} = r^*_j$ $(j = 1, \ldots, p-2)$ *given by (25) through (28) and* $r^*_{p-1} = 1$*, since the rates* λ_j $(j = 1, \ldots, p-2)$ *are given by (31) with* $\lambda_{p-1} = c + \nu$ *;*

- *both for* $n^* \le \frac{q_{m-2}}{2} - 2$ *and* $n^* > \frac{q_{m-2}}{2} - 2$ *we have for the GNIG distribution of depth* $p-1$*, the* $r^{**}_{\nu j} = r^*_j$ $(j = 1, \ldots, p-2)$ *given by (25) through (28) and the rates* λ_j $(j = 1, \ldots, p-2)$ *given by (31), with* $r^*_{\nu, p-1} = r$ *and* $\lambda_{p-1} = \lambda$*, obtained in such a way that the two first moments of the near-exact distributions match the two first exact moments.*

Proof. Considering the c.f. of W in (37), and truncating the infinite mixture of Exponential distributions corresponding to the p_{m-1}-th Logbeta we may write,

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

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

$$= \underbrace{\frac{1}{B\left(c, \frac{p_m}{2}\right)} \frac{1}{\Gamma\left(1 - \frac{p_m}{2}\right)} \sum_{\nu=0}^{\infty} \frac{\Gamma\left(1 - \frac{p_m}{2} + \nu\right)}{\nu!(c+\nu)} (c+\nu)(c+\nu-\mathrm{i}t)^{-1}}_{\varphi^*(t)} \underbrace{\prod_{j=1}^{q_{m-2}-3} \left(c + \frac{j}{2}\right)^{r_j} \left(c + \frac{j}{2} - \mathrm{i}t\right)^{-r_j}}_{\varphi^{**}(t)},$$

where $c=\frac{n+1-q_{m-2}}{2}$, so that taking

$$\varphi^*(t)=\sum_{\nu=0}^{\infty}\pi_\nu\varphi_{X\nu}(t)=\underbrace{\sum_{\nu=0}^{n^*}\pi_\nu\varphi_{X\nu}(t)}_{T_{n^*}(t)}+\underbrace{\sum_{\nu=n^*+1}^{\infty}\pi_\nu\varphi_{X\nu}(t)}_{R_{n^*}(t)}$$

where $\varphi_{X_\nu}(t)$ is the c.f. of an $Exponential(c+\nu)$ distribution $(\nu=0,1,\ldots)$ and the weights π_ν are given by

$$\pi_\nu=\frac{1}{B\left(c,\frac{p_m}{2}\right)}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\Gamma\left(1-\frac{p_m}{2}\right)\nu!(c+\nu)}\quad(\nu=0,1,\ldots),$$

then we approximate $R_{n^*}(t)$ by $\theta\varphi_1(t)$, where $\varphi_1(t)=\lambda^r(\lambda-\mathrm{i}t)^{-r}$ is the c.f. of a $Gamma(r,\lambda)$ distribution and the weight θ is given by $\theta=\sum_{\nu=n^*+1}^{\infty}\pi_\nu=1-\sum_{\nu=0}^{n^*}\pi_\nu$. The parameters r and λ of the Gamma distribution are obtained in such a way that the two first derivatives of $R_{n^*}(t)$ and $\theta\varphi_1(t)$ with respect to t, at $t=0$, are equal, i.e., in such a way that

$$\left.\frac{d^h}{dt^h}R_{n^*}(t)\right|_{t=0}=\theta\left.\frac{d^h}{dt^h}\varphi_1(t)\right|_{t=0}\iff\left.\frac{d^h}{dt^h}\varphi^*(t)\right|_{t=0}-\left.\frac{d^h}{dt^h}T_{n^*}(t)\right|_{t=0}=\theta\left.\frac{d^h}{dt^h}\varphi_1(t)\right|_{t=0}\qquad(h=1,2),$$

where

$$\left.\frac{d^h}{dt^h}\varphi^*(t)\right|_{t=0}=\frac{d^h}{dt^h}\left[\sum_{k=0}^{n^*}\pi_k\varphi_{X_k}(t)+\theta\varphi_1(t)\right]\Bigg|_{t=0}\qquad(h=1,2),$$

that is, in such a way that the two first moments of the exact and near-exact distributions are equal. Thus, we approximate $\varphi^*(t)$ by

$$T_{n^*}(t)+\theta\varphi_1(t)=\left(\frac{1}{B\left(c,\frac{p_m}{2}\right)}\frac{1}{\Gamma\left(1-\frac{p_m}{2}\right)}\sum_{\nu=0}^{n^*}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)}(c+\nu)(c+\nu-\mathrm{i}t)^{-1}\right)+\theta\lambda^r(\lambda-\mathrm{i}t)^{-r}.$$

Since we may write

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

$$
= \left[\left(\frac{1}{B\left(c,\frac{p_m}{2}\right)} \frac{1}{\Gamma\left(1-\frac{p_m}{2}\right)} \sum_{\nu=0}^{n^*} \frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)} (c+\nu)(c+\nu-\mathrm{i}t)^{-1}\right) + \theta\lambda^r(\lambda-\mathrm{i}t)^{-r}\right]
$$

$$
\times \prod_{j=1}^{q_{m-2}-3} \left(c+\frac{j}{2}\right)^{r_j} \left(c+\frac{j}{2}-\mathrm{i}t\right)^{-r_j}
$$

$$
= \frac{1}{B\left(c,\frac{p_m}{2}\right)} \frac{1}{\Gamma\left(1-\frac{p_m}{2}\right)} \left[\sum_{\nu=0}^{n^*} \frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)} (c+\nu)(c+\nu-\mathrm{i}t)^{-1} \prod_{j=1}^{q_{m-2}-3} \left(c+\frac{j}{2}\right)^{r_j} \left(c+\frac{j}{2}-\mathrm{i}t\right)^{-r_j}\right]
$$

$$
+ \theta\lambda^r(\lambda-\mathrm{i}t)^{-r} \prod_{j=1}^{q_{m-2}-3} \left(c+\frac{j}{2}\right)^{r_j} \left(c+\frac{j}{2}-\mathrm{i}t\right)^{-r_j},
$$

the near-exact c.f. of W may be given by

$$
\left[\left(\frac{1}{B\left(c,\frac{p_m}{2}\right)} \frac{1}{\Gamma\left(1-\frac{p_m}{2}\right)} \sum_{\nu=0}^{n^*} \frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)} (c+\nu)(c+\nu-\mathrm{i}t)^{-1}\right) \right.
$$

$$
\left. \times \prod_{j=1}^{q_{m-2}-3} \left(c+\frac{j}{2}\right)^{r_{m-1,j}} \left(c+\frac{j}{2}-\mathrm{i}t\right)^{-r_{m-1,j}} + \theta\lambda^r(\lambda-\mathrm{i}t)^{-r} \prod_{j=1}^{q_{m-2}-3} \left(c+\frac{j}{2}\right)^{r_{m-1,j}} \left(c+\frac{j}{2}-\mathrm{i}t\right)^{-r_{m-1,j}}\right] \tag{43}
$$

$$
\times \prod_{j=1}^{p-2} \left(\frac{n-p+j}{2}\right)^{r_j} \left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_j},
$$

which is the c.f. of a finite mixture of n^*+2 distributions, where n^*+1 components are GIG distributions and the last one is a GNIG distribution.

Then, we have to consider two situations:

- $n^* \le \frac{q_{m-2}}{2}-2$, case where (43) may be written as

$$\frac{1}{B\left(c,\frac{p_m}{2}\right)}\frac{1}{\Gamma\left(1-\frac{p_m}{2}\right)}\left[\sum_{\nu=1}^{n^*}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)}\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_{\nu j}^{**}}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_{\nu j}^{**}}\right.$$
$$\left.+\frac{\Gamma\left(1-\frac{p_m}{2}\right)}{c}c(c-\mathrm{i}t)^{-1}\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_j^*}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_j^*}\right] \quad (44)$$
$$+\theta\lambda^r(\lambda-\mathrm{i}t)^{-r}\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_j^*}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_j^*}$$

which is the c.f. of a finite mixture of n^*+2 distributions, where n^*+1 components are GIG distributions (n^* of them with depth $p-2$ and one with depth $p-1$) and the last one is a GNIG distribution of depth $p-1$, with shape and rate parameters mentioned in the body of Theorem 2;

- $n^*>\frac{q_{m-2}}{2}-2$, case where (43) may be written as

$$\frac{1}{B\left(c,\frac{p_m}{2}\right)}\frac{1}{\Gamma\left(1-\frac{p_m}{2}\right)}\left[\sum_{\nu=1}^{\frac{q_{m-2}}{2}-2}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)}\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_{\nu j}^{**}}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_{\nu j}^{**}}\right.$$
$$\left.+\sum_{\nu=0,\nu\geq\frac{q_{m-2}}{2}-1}^{n^*}\frac{\Gamma\left(1-\frac{p_m}{2}+\nu\right)}{\nu!(c+\nu)}(c+\nu)(c+\nu-\mathrm{i}t)^{-1}\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_j^*}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_j^*}\right] \quad (45)$$
$$+\theta\lambda^r(\lambda-\mathrm{i}t)^{-r}\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r_j^*}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r_j^*}$$

which is the c.f. of a finite mixture of n^*+2 distributions, where n^*+1 of them are GIG distributions, $\frac{q_{m-2}}{2}-2$ with depth $p-2$ and $n^*+1-(\frac{q_{m-2}}{2}-2)$ with depth $p-1$ and the last one is a GNIG distribution of depth $p-1$, with shape and rate parameters mentioned in the body of Theorem 2. ▪

The near-exact c.f.'s obtained in this way are asymptotic for increasing values of n^*, in the sense that they converge to the exact c.f., $\varphi_W(t)$, when $n^*\to+\infty$. The corresponding p.d.f.s and c.d.f.s may be obtained in a concise and manageable form, suitable for an expeditious computation of quantiles.

Taking into account that, from (42), we have

$$\pi_0=\frac{1}{B\left(c,\frac{p_m}{2}\right)}\frac{1}{c},$$

then, from (44) and (6)-(17) in Section 2, we have, for $n^*\leq\frac{q_{m-2}}{2}-2$, the near-exact p.d.f. and c.d.f. of Λ given by

$$f_\Lambda(u)=\frac{1}{B\left(c,\frac{p_m}{2}\right)c}\;K\sum_{j=1}^{p-1}P_j(-\ln u)u^{\lambda_j}\;+\sum_{\nu=1}^{n^*}\pi_\nu\,K_\nu\sum_{j=1}^{p-2}P_{\nu j}(-\ln u)u^{\lambda_j}+\pi_{n^*+1}K\,\lambda^r\sum_{j=1}^{p-2}P_j^{**}(-\ln u)u^{\lambda_j}$$

$$(0<u<1)$$

and

$$F_\Lambda(u)=\frac{1}{B\left(c,\frac{p_m}{2}\right)c}\;K\sum_{j=1}^{p-1}P_j^*(-\ln u)u^{\lambda_j}\;+\sum_{\nu=1}^{n^*}\pi_\nu\,K_\nu\sum_{j=1}^{p-2}P_{\nu j}^*(-\ln u)u^{\lambda_j}$$

$$+\,\pi_{n^*+1}\left(\lambda^r\frac{(-\ln u)^r}{\Gamma(r+1)}\,{}_1F_1(r,r+1,\lambda\ln u)-K\,\lambda^r\sum_{j=1}^{p-2}P_j^{***}(-\ln u)u^{\lambda_j}\right)$$

$$(0<u<1),$$

while for $n^*>\frac{q_{m-2}}{2}-2$ they are respectively given by

$$f_\Lambda(u)=\frac{1}{B\left(c,\frac{p_m}{2}\right)c}\;K\sum_{j=1}^{p-1}P_j(-\ln u)u^{\lambda_j}\;+\sum_{\nu=1}^{\frac{q_{m-2}}{2}-2}\pi_\nu\,K_\nu\sum_{j=1}^{p-2}P_{\nu j}(-\ln u)u^{\lambda_j}+\sum_{\nu=\frac{q_{m-2}}{2}-1}^{n^*}\pi_\nu\,K\sum_{j=1}^{p-1}P_j(-\ln u)u^{\lambda_j}$$

$$+\,\pi_{n^*+1}K\,\lambda^r\sum_{j=1}^{p-2}P_j^{**}(-\ln u)u^{\lambda_j}$$

$$(0<u<1)$$

and

$$F_\Lambda(u)=\frac{1}{B\left(c,\frac{p_m}{2}\right)c}\;K\sum_{j=1}^{p-1}P_j^*(-\ln u)u^{\lambda_j}\;+\sum_{\nu=1}^{\frac{q_{m-2}}{2}-2}\pi_\nu\,K_\nu\sum_{j=1}^{p-2}P_{\nu j}^*(-\ln u)u^{\lambda_j}+\sum_{\nu=\frac{q_{m-2}}{2}-1}^{n^*}\pi_\nu\,K\sum_{j=1}^{p-1}P_j^*(-\ln u)u^{\lambda_j}$$

$$+\,\pi_{n^*+1}\left(\lambda^r\frac{(-\ln u)^r}{\Gamma(r+1)}\,{}_1F_1(r,r+1,\lambda\ln u)-K\,\lambda^r\sum_{j=1}^{p-2}P_j^{***}(-\ln u)u^{\lambda_j}\right)$$

$$(0<u<1),$$

for K_ν, $P_{\nu j}(-\ln u)$, K and $P_j(-\ln u)$ given by (38) and (39), $P_{\nu j}^*(-\ln u)$ and $P_j^*(-\ln u)$ given by (40) and (41), π_ν given by (42) and

$$P_j^{**}(-\ln u)=\sum_{k=1}^{r_j}\left\{c_{j,k}\frac{\Gamma(k)}{\Gamma(k+r)}(-\ln u)^{k+r-1}\,{}_1F_1\big(r,k+r,(\lambda-\lambda_j)\ln u\big)\right\}$$

and

$$P_j^{***}(-\ln u)=\sum_{k=1}^{r_j}\left\{c_{j,k}^*\sum_{i=0}^{k-1}\frac{(-\ln u)^{r+i}\,\lambda_j^i}{\Gamma(r+1+i)}\,{}_1F_1\big(r,r+1+i,(\lambda-\lambda_j)\ln u\big)\right\}.$$

5. Proximity measures to assess the quality of near-exact distributions

We assess the proximity of the near-exact distributions developed using two measures, Δ_1 and Δ_2, based on c.f.'s. These two proximity measures were already used by Grilo and Coelho (2007, 2010) and Coelho and Mexia (2010), in situations where the expressions for the exact p.d.f. and c.d.f. are not known, and they are directly derived from the inversion formulas. Their expressions are

$$\Delta_1 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left|\varphi_W(t) - \varphi(t)\right| dt \tag{46}$$

and

$$\Delta_2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left|\frac{\varphi_W(t) - \varphi(t)}{t}\right| dt \tag{47}$$

where $\varphi_W(t)$ represents the exact c.f. of r.v. W and $\varphi(t)$ the near-exact c.f. under study. The measure Δ_2 in (47) may be seen as directly derived from Berry-Esseen bound. The use of these measures enables us to obtain an upper bound on the absolute value of the difference of the densities or the cumulative distribution functions, since

$$\max_{w>0}\left| f_W(w) - f(w)\right| \le \Delta_1 \quad \text{and} \quad \max_{w>0}\left|F_W(w) - F(w)\right| \le \Delta_2 ,$$

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 near-exact p.d.f. and c.d.f. of W, corresponding to $\varphi(t)$.

Smaller values of the measures Δ_1 and Δ_2 correspond to better approximations. This way, these measures are an useful tool for evaluating and comparing the performance of the near-exact distributions proposed.

6. Comparative numerical study

In order to assess the quality of the family of near-exact distributions M(n^*+1)GIG+GNIG, based on truncations of the exact c.f. and which equate the first 2 exact moments, we use the proximity measures Δ_1 and Δ_2 in (46) and (47). We analyze the proximity of the near-exact distributions developed to the exact distribution and also compare them with the near-exact M2GNIG distribution, used in , presented in Grilo and Coelho (2007, 2010), based on a factorization of the exact c.f. and which equates the first 4 exact moments. We consider here the near-exact M2GNIG distribution, since this was the best performing approximation among the asymptotic and near-exact distributions, based on factorizations, presented in Grilo and Coelho (2007, 2010).

We analyse the behaviour of these near-exact distributions for different values of m, that is, for different numbers of sets of variables and also for different numbers of variables in each set. In Table 1 we have a summary of the particular cases studied, for a sample size of n = 25: the case of 3 sets (all with an odd number of variables) and the case of 4 sets (in which only one has an even number of variables).

In Tables 2 through 7 we have, in alternative sequence, values for proximity measures Δ_1 and Δ_2 and some quantiles for the near-exact distributions, for the cases presented in Table 1.

Table 1: Number of sets, number of variables in each set and sample size.

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

We consider in our study a member of the family with five hundred terms. This is, for the truncation $n^*+1 = 499$ we obtain, as member, the near-exact M499GIG+GNIG distribution. In the calculations of the values of both proximity measures we use the exact c.f. in (34) and the near-exact c.f. in (43). In Table 2 we have m = 3 sets, all with an odd number of variables, where the global number of variables is 15 and the sample size is 25, making the difference between them equal to 10 ($n-p=10$). The values of the measures show that the near-exact M499GIG+GNIG distribution beats the near-exact M2GNIG distribution, although this last one equated more moments. In Table 4 we also have m = 3, but the global number of variables is 21 and the sample size is again 25, reducing the difference between them, which is now equal to $n-p=4$. When we increase the number of variables in one of the sets, the performance of both near-exact distributions becomes even better, with a larger improvement in the near-exact distribution based on the truncation of the exact distribution, developed in this paper. In Table 6 we consider m = 4 sets, 3 of which with an odd number of variables, being the global number of variables again equal to 21. Once again we may see the better performance of the near-exact distribution M499GIG+GNIG based on truncations, with lower values for both proximity measures. However, when the number of sets of variables increases the quality of the approximation given by the near-exact distribution M499GIG+GNIG becomes a bit worse. Anyway, we may always surpass this minor drawback increasing the number of terms considered in the truncations.

In Tables 3, 5 and 7 we have some quantiles of the M499GIG+GNIG and M2GNIG near-exact distributions. As we may see in Grilo and Coelho (2010) and in Coelho and Mexia (2010), smaller values of measures Δ_1 and Δ_2 are associated with smaller differences among quantiles.

Table 2: Values of measures Δ_1 and Δ_2 for near-exact distributions.
Case $m=3$ with $p_1=5, p_2=7, p_3=3$ and $n=25$.

Near-exact distributions	**Proximity measures**	
	Δ_1	Δ_2
M499GIG+GNIG (2 moments)	4.585E-17	5.612E-18
M2GNIG (4 moments)	6.369E-11	3.135E-12

Table 3: Some quantiles of near-exact distributions, for $m=3$ with $p_1=5, p_2=7, p_3=3$ and $n=25$.

Near-exact distributions	**Quantile**		
	0.90	0.95	0.99
M499GIG+GNIG (2m.)	5.070602126803278	5.372467667062496	5.971703532356292
M2GNIG (4 m.)	5.070602126798732	5.372467667053351	5.971703532349906

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

Near-exact distributions	**Proximity measures**	
	Δ_1	Δ_2
M499GIG+GNIG (2 moments)	2.512E-20	3.072E-21
M2GNIG (4 moments)	1.416E-12	1.328E-13

Table 5: Some quantiles of near-exact distributions, for $m = 3$ with $p_1 = 5$, $p_2 = 7$, $p_3 = 9$ and $n = 25$.

Near-exact distributions	**Quantile**		
	0.90	0.95	0.99
M499GIG+GNIG (2m.)	12.348022863501802	12.910964910802633	14.024583497046132
M2GNIG (4 m.)	12.348022863501334	12.910964910801998	14.024583497046136

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

Near-exact distributions	**Proximity measures**	
	Δ_1	Δ_2
M499GIG+GNIG (2 moments)	4.935E-18	6.340E-19
M2GNIG (4 moments)	1.192E-12	2.077E-14

Table 7: Some quantiles of near-exact distributions, for $m = 4$ with $p_1 = 5$, $p_2 = 7$, $p_3 = 3$, $p_4 = 6$ and $n = 25$.

Near-exact distributions	**Quantile**		
	0.90	0.95	0.99
M499GIG+GNIG (2m.)	13.298396053836127	13.871917262346954	15.003949956414504
M2GNIG (4 m.)	13.298396053835702	13.871917262346372	15.003949956414496

7. Conclusions and final remarks

The expression obtained for the exact distribution of the generalized Wilks Λ statistic needs a very large number of terms in the series of the exact distribution GIG to obtain accurate enough approximations to the exact p.d.f., c.d.f., moments and quantiles. For the same accuracy, this number is much larger than the number of terms needed in the finite mixture of the near-exact M(n^*+1)GIG+GNIG distributions, for a similar accuracy. This way, this family of near-exact distributions, which equates the first two exact moments and lays very close to the exact distribution, is very useful for practical purposes, mainly for small sample sizes.

We have to point out the excellent performance of the family of near-exact distributions M(n^*+1)GIG+GNIG for small values of n and for small values of $n - p$. For higher values of n or $n - p$, we may consider the near-exact distribution M2GNIG, based on factorizations of the exact c.f., as an alternative.

We should stress here the wide range of application of the near-exact distributions obtained, since they apply not only to the case of underlying multivariate Normal distributions but also to the cases of underlying elliptically contoured and left orthogonal-invariant distributions. Such very precise approximate distributions may indeed be necessary and adequate for cases where, for some reason, one may need extra precision that the usual asymptotic distributions available are not able to provide.

Acknowledgments

The authors want to thank the financial support of FCT/MCTES/PT (the Portuguese Foundation for Science and Technology) through grant 2010 ISFL-1-297.

References

Abramowitz, M. and Stegun, I. A. (1974). *Handbook of Mathematical Functions.* (eds.). 9.ª ed., Dover, New York.

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

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

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

Coelho, C. A. (2000). On two asymptotic normal distributions for the generalized Wilks lambda statistic. *Comm. Statist. – Theory and Methods*, **29**, 7, 1465-1486.

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

Coelho, C. A. (2004). 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**, 191-218.

Coelho, C. A., Alberto, R. P. and Grilo, L. M. (2006). A mixture of Generalized Integer Gamma distribution as the exact distribution of the product of an odd number of independent Beta random variables with first parameter evolving by 1/2. Applications. *Journal of Interdisciplinary Mathematics*, **9**, 2, 228-248.

Coelho, C. A. and Mexia, J. T. (2010). *Product and Ratio of Generalized Gamma-Ratio Random Variables – Exact and Near-exact distributions – Applications*, 145+v pp. Lambert Academic Publishing AG & Co. KG, Saarbrüken, Germany.

Grilo, L. M. (2005). *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.

Grilo, L. M. and Coelho, C. A. (2007). 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**, 1560-1575.

Grilo, L. M. and Coelho, C. A. (2010). Near-exact distributions for the generalized Wilks Lambda statistic. *Discussiones Mathematicae Probability and Statistics*, **30** (1) 53-86.

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

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471-494.

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