A Mixture of Generalized Integer Gamma distributions as the exact distribution of the product of an odd number of independent Beta random variables. Applications

Carlos A. Coelho [a,1,*] Rui P. Alberto [b,1] Luís M. Grilo [c,1]

[a] *The New University of Lisbon – Faculty of Sciences and Technology – Mathematics Department (cmac@fct.unl.pt)*

[b] *Lisbon Institute of Agriculture Technology – Lisbon University of Technology, Research Unit in Applied Mathematics (rui.pesado@sapo.pt)*

[c] *Polytechnic Institute of Tomar, Dep. of Mathematics (lgrilo@ipt.pt)*

Abstract

In this paper we show first how the distribution of the logarithm of a random variable with a Beta distribution may be expressed either as a mixture of Gamma distributions or as a mixture of Generalized Integer Gamma (GIG) distributions and then how the exact distribution of the product of an odd number of independent Beta random variables whose first parameter evolves by 1/2 and whose second parameter is the half of an odd integer may be expressed as a mixture of GIG distributions. Some particularities of these mixtures are analysed. The results are then used to obtain the exact distribution of the logarithm of the Wilks Λ statistic to test the independence of two sets of variables, both with an odd number of variables, and the exact distribution of the logarithm of the generalized Wilks Λ statistic to test the independence of several sets of variables, in the case where two or three of them have an odd number of variables. A discussion of relative advantages and disadvantages of the use of the exact versus near-exact distributions is carried out.

Key words: sum independent Gamma random variables, mixtures, mixture Exponential distributions, mixture Generalized Integer Gamma distributions, improper weights, generalized Wilks Lambda, likelihood ratio statistic.

[*] Corresponding author: Mathematics Department, Faculty of Sciences and Technology, The New University of Lisbon, Quinta da Torre, 2829-516 Caparica, Portugal; Tel:(351) 21 294 8388; Fax:(351) 21 294 8391.

[1] This research was financially supported by the Portuguese Foundation for Science and Technology (FCT).

Accepted for publication in J. of Interdisciplinary Mathematics

1 Introduction

Obtaining the exact distribution of the product of an odd number of independent random variables with Beta distributions which have the first parameter evolving by 1/2 and the same second parameter is an interesting and long-standing problem. Besides, its solution would be useful since it would allow us to obtain the exact distribution of a number of statistics, namely some of the likelihood ratio statistics used in Multivariate Statistics (under the normality assumptions). In this paper we will show how, starting from the expression of the p.d.f. (probability density function) of a Logbeta random variable, that is, a random variable whose exponential has a Beta distribution, we may express both the distribution of such random variable and the distribution of the product of independent random variables with Beta distributions as mixtures of GIG (Generalized Integer Gamma) distributions. The GIG distribution was introduced by Coelho (1998).

In order to achieve this purpose we need to establish some notation and a preliminary result.

2 A Preliminary Result

Let X have a Beta distribution with parameters $a, b \in I\!R^+$ and let

$$Y = -\log X\,.$$

We will say that the random variable Y has a Logbeta distribution with parameters a and b. The p.d.f. of Y is

$$f_Y(y) \;=\; \frac{1}{B(a,b)}\; \mathrm{e}^{-ay} \left(1-\mathrm{e}^{-y}\right)^{b-1} \;.$$

We will next show how for non-integer b the distribution of Y may be expressed as a mixture of Exponential or GIG distributions, which for $b > 1$ has improper weights, that is, some of which are negative and/or greater than 1 in absolute value, although always adding up to 1.

Hereon we will use the notation

$$X \;\sim\; G(r, \lambda)$$

to denote the fact that the random variable X has a Gamma distribution with shape parameter r and rate parameter λ, that is, that the p.d.f. of X is

$$f_X(x) \;=\; \frac{\lambda^r}{\Gamma(r)}\; \mathrm{e}^{-\lambda x}\; x^{r-1}\,,$$

and the notation

$$Z \sim MG\,(p_i; r_i, \lambda_i; i = lb, \ldots, ub)$$

to denote the fact that the distribution of the random variable Z is a mixture with weights p_i of Gamma distributions with shape parameters r_i and rate parameters λ_i for $i = lb, \ldots, ub$, where if ub is ommitted it means that the mixture is infinite, being then usually used the notation $i = lb, lb + s, \ldots$, where s is the step in i. Accordingly, we will also use the notation

$$W \sim MGIG\,(p_i;\, r_{1i}, \ldots, r_{gi};\, \lambda_{1i}, \ldots, \lambda_{gi};\, i = lb, \ldots, ub)$$

to denote the fact that the distribution of the random variable W is a mixture with weights p_i of GIG distributions of depth g, the i-th one of which has shape parameters $r_{1i}, \ldots, r_{gi}$ and rate parameters $\lambda_{1i}, \ldots, \lambda_{gi}$, for $i = lb, \ldots, ub$, where once again, if ub is ommitted means that the mixture is infinite.

Theorem 1 *Let*

$$Y \sim Logbeta(a, b)\,.$$

Then if $b \in I\!N$, the distribution of Y is a particular Generalized Integer Gamma distribution of depth b with shape parameters all equal to 1 and rate parameters $a, a+1, \ldots, a+b-1$, that is

$$Y \sim GIG\Big(\underbrace{1, 1, \ldots, 1}_{b};\ \underbrace{a, a+1, \ldots, a+b-1}_{b}\Big)\,,$$

and if $b \in I\!R^+ \backslash I\!N$, with $r = \lfloor b \rfloor$ and $r^* = b - r$ (where $\lfloor\cdot\rfloor$ denotes the floor of the argument, that is, the largest integer that does not exceed the argument), the distribution of Y is either an infinite mixture of Exponential distributions or an infinite mixture of Generalized Integer Gamma distributions of depth $r+1$, more precisely, in this case we may write,

$$\begin{aligned}
Y \sim & MG\Big(\frac{\Gamma(1-b+j)}{B(a,b)\,\Gamma(1-b)\,j!\,(a+j)};\ 1, a+j;\ j=0,1,\ldots\Big)\\
\equiv & MGIG\Big(\frac{\Gamma(1-r^*+j)}{B(a,r^*)\,\Gamma(1-r^*)\,j!\,(a+j)};\underbrace{1,\ldots,1}_{r+1};\underbrace{a+r^*,\ldots,a+b-1,a+j}_{r+1};\\
& \qquad\qquad j=0,1,\ldots\Big)\\
\equiv & MGIG\Big(\frac{\Gamma(1-r^*+j)}{B(a+r,r^*)\,\Gamma(1-r^*)\,j!\,(a+r+j)};\underbrace{1,\ldots,1}_{r+1};\underbrace{a,\ldots,a+r-1,a+r+j}_{r+1};\\
& \qquad\qquad j=0,1,\ldots\Big).
\end{aligned}$$

For $b > 1$ the first of these mixtures has improper weights, some of which are negative and/or have absolute value greater than one, although always adding up to 1, and some of which may have enormously large absolute values when a is large.

Proof: Since for $a, b \in I\!R^+$, if

$$X \sim Beta(a,b)$$

then

$$E\left(X^h\right) = \frac{\Gamma(a+b)\,\Gamma(a+h)}{\Gamma(a)\,\Gamma(a+b+h)}\,,$$

the characteristic function of $Y = -\log X$, for $b \in I\!N$, using $\mathrm{i} = (-1)^{1/2}$ and the fact that for any complex a and integer n,

$$\frac{\Gamma(a+n)}{\Gamma(a)} = \prod_{j=0}^{n-1}(a+j)\,, \tag{1}$$

may be written as

$$\begin{aligned}\Phi_Y(t) &= E\left[\mathrm{e}^{\mathrm{i}tY}\right] = E\left[\mathrm{e}^{-\mathrm{i}t\log X}\right]\\ &= E\left[X^{-\mathrm{i}t}\right] = \frac{\Gamma(a+b)\,\Gamma(a-\mathrm{i}t)}{\Gamma(a)\,\Gamma(a+b-\mathrm{i}t)}\\ &= \prod_{j=0}^{b-1}(a+j)\,(a+j-\mathrm{i}t)^{-1}\end{aligned}$$

that is the characteristic function of a sum of b independent random variables with Exponential distributions with parameters $a+j$ $(j = 0, \ldots, b-1)$. This distribution is a GIG distribution of depth b with shape parameters all equal to 1 and rate parameters $a, a+1, \ldots, a+b-1$. This result had already been obtained in Coelho (1998) but it is shown here for completeness so that we may have in one only Theorem the results for $b \in I\!N$ and $b \in I\!R^+\backslash I\!N$.

For $b \in I\!R^+\backslash I\!N$, since we know that for $|x| < 1$,

$${}_1F_0(c;x) = \sum_{j=0}^{\infty}\frac{\Gamma(c+j)}{\Gamma(c)}\,\frac{x^j}{j!} = (1-x)^{-c}\,,$$

we may write, for $y > 0$,

$$\left(1-\mathrm{e}^{-y}\right)^{b-1} = {}_1F_0\left(1-b;\mathrm{e}^{-y}\right) = \sum_{j=0}^{\infty}\frac{\Gamma(1-b+j)}{\Gamma(1-b)}\,\frac{(\mathrm{e}^{-y})^j}{j!}$$

so that the p.d.f. of Y may be written as

$$
\begin{aligned}
f_Y(y) &= \frac{1}{B(a,b)}\ \mathrm{e}^{-ay} \sum_{j=0}^{\infty} \frac{\Gamma(1-b+j)}{\Gamma(1-b)}\ \frac{\mathrm{e}^{-jy}}{j!} \\
&= \frac{1}{B(a,b)}\ \frac{1}{\Gamma(1-b)} \sum_{j=0}^{\infty} \frac{\Gamma(1-b+j)}{j!(a+j)}\ \underbrace{(a+j)\ \mathrm{e}^{-(a+j)y}}_{\text{p.d.f. of } G(1,a+j)} = \sum_{j=0}^{\infty} p_j\ (a+j)\ \mathrm{e}^{-(a+j)y}
\end{aligned}
$$

that is the p.d.f. of a mixture of $G(1, a+j)$ $(j = 0, 1, \ldots)$ distributions, with weights

$$
p_j \;=\; \frac{1}{B(a,b)}\ \frac{\Gamma(1-b+j)}{\Gamma(1-b)\, j!\, (a+j)} \qquad (j = 0, 1, \ldots)\,. \tag{2}
$$

That these weights add up to 1 it may be shown through the known relation for the Gauss hypergeometric function

$$
{}_2F_1(a,b;c;1) \;=\; \sum_{i=0}^{\infty} \frac{\Gamma(a+i)\,\Gamma(b+i)\,\Gamma(c)}{\Gamma(a)\,\Gamma(b)\,\Gamma(c+i)}\ \frac{1}{i!} \;=\; \frac{\Gamma(c)\,\Gamma(c-a-b)}{\Gamma(c-a)\,\Gamma(c-b)}
$$

so that

$$
\begin{aligned}
\sum_{i=0}^{\infty} \frac{\Gamma(1-b+i)}{\Gamma(1-b)}\ \frac{1}{i!\,(a+i)} &= \sum_{i=0}^{\infty} \frac{\Gamma(1-b+i)}{\Gamma(1-b)}\ \frac{\Gamma(a+i)}{\Gamma(a+i+1)}\ \frac{1}{i!} \\
&= \frac{1}{a} \sum_{i=0}^{\infty} \frac{\Gamma(1-b+i)}{\Gamma(1-b)}\ \frac{\Gamma(a+i)}{\Gamma(a)}\ \frac{\Gamma(a+1)}{\Gamma(a+1+i)}\ \frac{1}{i!} \\
&= \frac{1}{a}\ {}_2F_1(1-b, a; a+1; 1) \;=\; \frac{1}{a}\ \frac{\Gamma(a+1)\,\Gamma(b)}{\Gamma(a+b)} \\
&= B(a,b)
\end{aligned}
$$

and thus

$$
\sum_{j=0}^{\infty} \frac{1}{B(a,b)}\ \frac{\Gamma(1-b+j)}{\Gamma(1-b)\, j!\, (a+j)} \;=\; 1\,.
$$

We should note that for $0 < b < 1$ all the weights are proper weights, that is, they have values between 0 and 1, while for $b > 1$ some of them are negative and/or greater than one in absolute value. In fact since

$$
\Gamma(b)\,\Gamma(1-b) \;=\; \frac{\pi}{\sin(\pi b)}
$$

the weights p_j in (2) may be written as

$$
p_j \;=\; \frac{\sin(\pi b)}{\pi}\ \frac{\Gamma(a+b)}{\Gamma(a)}\ \frac{\Gamma(1-b+j)}{j!\,(a+j)}
$$

so that since for $0 < b < 1$ we have $1-b+j > 0$ ($j = 0, 1, \ldots$) and $\sin(\pi b) > 0$, all the weights are positive and since they add up to 1, in this case

$$0 < p_j < 1 \;, \qquad \forall j \in \{0, 1, \ldots\} \,,$$

while for $b > 1$, $1 - b + j$ has some values negative so that both $\Gamma(1 - b + j)$ and $\sin(\pi b)$ may either have positive or negative values, making some of the weights p_j positive and other negative, and since they add up to 1, some of them will also have absolute values larger than 1.

Indeed even in the cases where $b > 1$ we may work through the characteristic function of Y in order to obtain its distribution as a mixture of GIG distributions with proper weights. In fact, let us first, for $b \in I\!R^+ \backslash I\!N$, as a consequence of the result above write the characteristic function of Y as a mixture of characteristic functions of Gamma random variables, that is

$$\begin{aligned}\Phi_Y(t) = E\left[\mathrm{e}^{\mathrm{i}tY}\right] &= \frac{\Gamma(a+b)}{\Gamma(a)}\,\frac{\Gamma(a-\mathrm{i}t)}{\Gamma(a+b-\mathrm{i}t)} \\ &= \frac{1}{B(a,b)}\,\frac{1}{\Gamma(1-b)}\sum_{j=0}^{\infty}\frac{\Gamma(1-b+j)}{j!(a+j)}\,(a+j)\,(a+j-\mathrm{i}t)^{-1} \,. \end{aligned} \tag{3}$$

Let then Y be a random variable with a Logbeta distribution with parameters a and b, where $b > 1$ and let $r = \lfloor b \rfloor$. Let further $r^* = b - r$, so that $b = r + r^*$. Then, using (1) and (3), we may write the characteristic function of Y as

$$\begin{aligned}\Phi_Y(t) = E\left[\mathrm{e}^{\mathrm{i}tY}\right] &= \frac{\Gamma(a+b)\,\Gamma(a-\mathrm{i}t)}{\Gamma(a)\,\Gamma(a+b-\mathrm{i}t)} \\ &= \frac{\Gamma(a+r+r^*)}{\Gamma(a+r^*)}\,\frac{\Gamma(a+r^*)}{\Gamma(a)}\,\frac{\Gamma(a-\mathrm{i}t)}{\Gamma(a+r^*-\mathrm{i}t)}\,\frac{\Gamma(a+r^*-\mathrm{i}t)}{\Gamma(a+r+r^*-\mathrm{i}t)} \\ &= \frac{\Gamma(a+r^*)}{\Gamma(a)}\,\frac{\Gamma(a-\mathrm{i}t)}{\Gamma(a+r^*-\mathrm{i}t)}\prod_{k=0}^{r-1}(a+r^*+k)\,(a+r^*+k-\mathrm{i}t)^{-1} \\ &= \frac{1}{B(a,r^*)}\,\frac{1}{\Gamma(1-r^*)}\sum_{j=0}^{\infty}\frac{\Gamma(1-r^*+j)}{j!\,(a+j)}(a+j)(a+j-\mathrm{i}t)^{-1} \\ &\qquad\qquad \prod_{k=0}^{r-1}(a+r^*+k)\,(a+r^*+k-\mathrm{i}t)^{-1} \end{aligned}$$

that is the characteristic function of an infinite mixture of GIG distributions of depth $r+1$ with all shape parameters equal to 1 and rate parameters $a+r^*+k$ ($k = 0, \ldots, r-1$) and the $(r+1)$-th equal to $a+j$ ($j = 0, 1, \ldots$), with weights $\frac{1}{B(a,r^*)}\,\frac{1}{\Gamma(1-r^*)}\,\frac{\Gamma(1-r^*+j)}{j!\,(a+j)}$ ($j = 0, 1, \ldots$).

Yet another way of decomposing the characteristic function of Y would be

$$\begin{aligned}\Phi_Y(t)=E\left[\mathrm{e}^{\mathrm{i}tY}\right] &= \frac{\Gamma(a+b)\,\Gamma(a-\mathrm{i}t)}{\Gamma(a)\,\Gamma(a+b-\mathrm{i}t)} \\ &= \frac{\Gamma(a+r+r^*)}{\Gamma(a+r)}\,\frac{\Gamma(a+r)}{\Gamma(a)}\,\frac{\Gamma(a-\mathrm{i}t)}{\Gamma(a+r-\mathrm{i}t)}\,\frac{\Gamma(a+r-\mathrm{i}t)}{\Gamma(a+r+r^*-\mathrm{i}t)} \\ &= \frac{\Gamma(a+r+r^*)}{\Gamma(a+r)}\,\frac{\Gamma(a+r-\mathrm{i}t)}{\Gamma(a+r+r^*-\mathrm{i}t)}\prod_{k=0}^{r-1}(a+k)\,(a+k-\mathrm{i}t)^{-1} \\ &= \frac{1}{B(a+r,r^*)}\,\frac{1}{\Gamma(1-r^*)}\sum_{j=0}^{\infty}\frac{\Gamma(1-r^*+j)}{j!\,(a+r+j)}(a+r+j)(a+r+j-\mathrm{i}t)^{-1} \\ &\qquad\qquad \prod_{k=0}^{r-1}(a+k)\,(a+k-\mathrm{i}t)^{-1}\end{aligned}$$

yielding the characteristic function of an infinite mixture of GIG distributions of *depth* $r+1$ with all shape parameters equal to 1 and rate parameters $a+k$ ($k=0,\ldots,r-1$) and the $(r+1)$-th equal to $a+r+j$ ($j=0,1,\ldots$), with weights $\frac{1}{B(a+r,r^*)}\,\frac{1}{\Gamma(1-r^*)}\,\frac{\Gamma(1-r^*+j)}{j!\,(a+r+j)}$ $(j=0,1,\ldots)$. □

Actually for the mixtures with proper weights, the sum of the first weights declines as the value of the first parameter of the Logbeta whose characteristic function is expressed as a sum grows large, while for the mixtures with improper weights the absolute value of some of the weights grows enormously as the value of that parameter also grows large. For both reasons it seems we should try to express as a mixture Logbeta random variables with the first parameter as small as possible and the second parameter with a value between 0 and 1.

3 The exact distribution of the product of an odd number of independent Beta random variables

We will use the notation

$$W \sim MGIG\,(p_i; r_{1i},\ldots,r_{gi}|r^*_{1i},\ldots,r^*_{hi};\lambda_{1i},\ldots,\lambda_{gi}|\lambda^*_{1i},\ldots,\lambda^*_{hi};$$
$$i=lb_1,\ldots,ub_1|i=lb_2,\ldots,ub_2)$$

to denote that the distribution of W is a mixture, with weights p_i, of GI Gamma distributions of depth g, for $i=lb_1,\ldots,ub_1$, and depth h, for $i=lb_2,\ldots,ub_2$, where if ub_2 is missing it is taken as infinity.

Theorem 2 *Let p be an odd integer and let*

$$Y_j \sim Beta\left(a_j, \frac{f}{2}\right) \qquad j = 1, \ldots, p$$

be independent random variables with Beta distributions, where f is also an odd positive integer and $a_j = c - j/2$ $(j = 1, \ldots, p)$, with $c = a + p/2$, $a \in I\!R^+$. Further let

$$W_1' = \prod_{j=1}^{p} Y_j \qquad \text{and} \qquad W_1 = -\log W_1' = -\sum_{j=1}^{p} \log Y_j\,.$$

Let also

$$Y_j^* \sim Beta\left(a_j^*, \frac{p}{2}\right) \qquad j = 1, \ldots, f$$

be independent random variables with Beta distributions, where $a_j^* = c^* - j/2$ $(j = 1, \ldots, f)$, with $c^* = a + f/2$, $a \in I\!R^+$. Let then

$$W_2' = \prod_{j=1}^{f} Y_j^* \qquad \text{and} \qquad W_2 = -\log W_2' = -\sum_{j=1}^{f} \log Y_j^*\,.$$

Then the exact distribution of W_1 is a mixture of GIG distributions, the first $\frac{p+f}{2} - 2$ of which are of depth $p + f - 3$ and the remaining of depth $p + f - 2$, what we will denote as

$$W_1 \sim MGIG\Bigg(\frac{1}{B\left(a, \frac{1}{2}\right)\sqrt{\pi}} \frac{\Gamma\left(k + \frac{1}{2}\right)}{k!\,(a+k)};\ r_1^{**}, \ldots, r_{p+f-3}^{**} \,\Big|\, 1, r_1^*, \ldots, r_{p+f-3}^*;$$
$$\underbrace{a + \tfrac{1}{2}, \ldots, a + \tfrac{p+f-3}{2}}_{p+f-3} \,\Big|\, \underbrace{a + k, a + \tfrac{1}{2}, \ldots, a + \tfrac{p+f-3}{2}}_{p+f-2};$$
$$k = 1, \ldots, \tfrac{p+f}{2} - 2 \,\Big|\, k = 0, k \geq \tfrac{p+f}{2} - 1 \Bigg)$$

where for $j = 1, \ldots, p + f - 3$, (for $k = 1, 2, \ldots, \frac{p+f}{2} - 2$)

$$r_j^{**} = \begin{cases} r_j & j = 2n \ \left(n = 1, \ldots, \frac{f-3}{2}\right),\ j \neq 2k \\ & j = f - 1, \ldots, f + p - 3 \\ r_j + 1 & j = 1 + 2n \ \left(n = 0, \ldots, \frac{f-3}{2}\right) \\ r_j + 2 & j = 2k \end{cases} \tag{4}$$

and (for $k=0$ and $k \geq \frac{p+f}{2}-1$)

$$r_j^* = \begin{cases} r_j & j=2n \ \left(n=1,\ldots,\frac{f-3}{2}\right) \\ & j=f-1,\ldots,f+p-3 \\ r_j+1 & j=1+2n \ \left(n=0,\ldots,\frac{f-3}{2}\right) \end{cases} \tag{5}$$

with

$$r_j = \begin{cases} h_j & j=1,2 \\ r_{j-2}+h_j & j=3,\ldots,p+f-3 \end{cases} \tag{6}$$

where

$$h_j = \begin{cases} 1 & j=1,\ldots,\min(p-1,f) \\ 0 & j=1+\min(p-1,f),\ldots,\max(p-1,f) \\ -1 & j=1+\max(p-1,f),\ldots,p+f-3 \end{cases} \tag{7}$$

or equivalently,

$$h_j = (\textit{number of elements of } \{p-1,f\} \textit{ greater or equal to } j)-1\,. \tag{8}$$

The distribution of W_2 is the same as the distribution of W_1 and also the distribution of W_2' is the same as the distribution of W_1'.

Proof: In order to obtain a mixture with all proper weights vanishing as fast as possible we will leave for decomposition, under the form of a series, the characteristic function of a Logbeta random variable with the second parameter equal to 1/2 and first parameter as small as possible.

Indeed, given the independence of the p random variables Y_j and the fact that the h-th moment of Y_j is given by

$$E\left(Y_j^h\right) = \frac{\Gamma\left(a_j+\frac{f}{2}\right)\,\Gamma(a_j+h)}{\Gamma(a_j)\,\Gamma\left(a_j+\frac{f}{2}+h\right)}\,,$$

using (3) and a by-product of the proof of Theorem 2 in Coelho (1998), which for even p states that for $a_j = a+\frac{p}{2}-\frac{j}{2}$,

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

where

$$r_j = \begin{cases} h_j & j=1,2 \\ r_{j-2}+h_j & j=3,\ldots,p+f-2 \end{cases}$$

with

$$h_j = (\text{number of elements of } \{p,f\} \text{ greater or equal to } j)-1\,,$$

what for odd p yields

$$\prod_{j=1}^{p-1} \frac{\Gamma\left(a_j+\frac{f}{2}\right)}{\Gamma(a_j)} = \prod_{j=1}^{p-1} \frac{\Gamma\left(a+\frac{1}{2}+\frac{p-1}{2}-\frac{j}{2}+\frac{f}{2}\right)}{\Gamma\left(a+\frac{1}{2}+\frac{p-1}{2}-\frac{j}{2}\right)} = \prod_{j=1}^{p+f-3} \left(a+\frac{j}{2}\right)^{r_j}\,,$$

with r_j defined as above, with p replaced by $p-1$, using $r=\lfloor f/2 \rfloor = \frac{f-1}{2}$, the characteristic function of W_1 may be written as

$$\begin{aligned}
\Phi_{W_1}(t) &= E\left(e^{\mathrm{i}tW_1}\right) = E\left(e^{-\mathrm{i}t\log W_1'}\right) = E\left(W_1'^{-\mathrm{i}t}\right) = E\left(\prod_{j=1}^{p} Y_j^{-\mathrm{i}t}\right) \\
&= \prod_{j=1}^{p} E\left(Y_j^{-\mathrm{i}t}\right) = \prod_{j=1}^{p} \frac{\Gamma\left(a_j+\frac{f}{2}\right)\Gamma(a_j-\mathrm{i}t)}{\Gamma(a_j)\,\Gamma\left(a_j+\frac{f}{2}-\mathrm{i}t\right)} \\
&= \frac{\Gamma\left(a_p+\frac{f}{2}\right)\Gamma(a_p-\mathrm{i}t)}{\Gamma(a_p)\,\Gamma\left(a_p+\frac{f}{2}-\mathrm{i}t\right)} \prod_{j=1}^{p-1} \frac{\Gamma\left(a_j+\frac{f}{2}\right)\Gamma(a_j-\mathrm{i}t)}{\Gamma(a_j)\,\Gamma\left(a_j+\frac{f}{2}-\mathrm{i}t\right)} \\
&= \frac{1}{B\left(a_p,\frac{1}{2}\right)\Gamma(1/2)} \sum_{k=0}^{\infty} \frac{\Gamma\left(k+\frac{1}{2}\right)}{k!\,(a_p+k)} (a_p+k)\,(a_p+k-\mathrm{i}t)^{-1} \\
&\qquad \prod_{l=0}^{r-1}\left(a_p+\frac{1}{2}+l\right)\left(a_p+\frac{1}{2}+l-\mathrm{i}t\right)^{-1} \prod_{j=1}^{p+f-3}\left(a+\frac{j}{2}\right)^{r_j}\left(a+\frac{j}{2}-\mathrm{i}t\right)^{-r_j} \\
&= \frac{1}{B\left(a,\frac{1}{2}\right)\sqrt{\pi}} \left[\sum_{k=1}^{\frac{p+f}{2}-2} \frac{\Gamma\left(k+\frac{1}{2}\right)}{k!\,(a+k)} \prod_{j=1}^{p+f-3}\left(a+\frac{j}{2}\right)^{r_j^{**}}\left(a+\frac{j}{2}-\mathrm{i}t\right)^{-r_j^{**}}\right. \\
&\qquad + \sum_{k=0,\frac{p+f}{2}-1}^{\infty} \frac{\Gamma\left(k+\frac{1}{2}\right)}{k!\,(a+k)} (a+k)\,(a+k-\mathrm{i}t)^{-1} \\
&\qquad\qquad \left.\prod_{j=1}^{p+f-3}\left(a+\frac{j}{2}\right)^{r_j^{*}}\left(a+\frac{j}{2}-\mathrm{i}t\right)^{-r_j^{*}}\right]
\end{aligned} \tag{9}$$

with r_j given by (6) through (8), r_j^* given by (5) through (8) and r_j^{**} $(j = 1, \ldots, p+f-3)$ given by (4) and (6) through (8).

That the characteristic function of W_2 is the same as the characteristic function of W_1 is most easily shown using, for positive integer p and f, the relation

$$\prod_{j=1}^{p} \frac{\Gamma\left(a+\frac{p}{2}-\frac{j}{2}+\frac{f}{2}\right)}{\Gamma\left(a+\frac{p}{2}-\frac{j}{2}\right)} = \prod_{j=1}^{f} \frac{\Gamma\left(a+\frac{f}{2}-\frac{j}{2}+\frac{p}{2}\right)}{\Gamma\left(a+\frac{f}{2}-\frac{j}{2}\right)} \tag{10}$$

that comes out of the proof of Theorem 2 in Coelho (1998) and the results in Coelho (1999). Then, a direct application of (10) to the characteristic function of W_2 leads us immediately to the characteristic function of W_1, since

$$\begin{aligned} E\left(\mathrm{e}^{\mathrm{i}tW_2}\right) &= \prod_{j=1}^{f} \frac{\Gamma(a_j^*-\mathrm{i}t)\,\Gamma(a_j^*+p/2)}{\Gamma(a_j^*+p/2-\mathrm{i}t)\,\Gamma(a_j^*)} \\ &= \prod_{j=1}^{p} \frac{\Gamma\left(a+\frac{p}{2}-\frac{j}{2}-\mathrm{i}t\right)\,\Gamma\left(a+\frac{f}{2}+\frac{p}{2}-\frac{j}{2}\right)}{\Gamma\left(a+\frac{f}{2}+\frac{p}{2}-\frac{j}{2}-\mathrm{i}t\right)\,\Gamma\left(a+\frac{p}{2}-\frac{j}{2}\right)}, \end{aligned}$$

so that the distributions of W_2 and W_1 are the same and so are the distributions of W_2' and W_1'. □

The application of the result in this Theorem and the results in Coelho (1998) on the GIG distribution will enable us to easily obtain the p.d.f. and c.d.f. (cumulative distribution function) for both W_1 and W_2 and thus also for W_1' and W_2'.

4 An application of the results obtained

The generalized Wilks Λ statistic is the well known statistic used to test the fit of the Generalized Canonical Analysis model (Coelho, 1992) or the independence among m sets of jointly normally distributed variables. Let us assume that

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

where

$$\underline{\mu} = \left[\underline{\mu}_1', \ldots, \underline{\mu}_k', \ldots, \underline{\mu}_m'\right]', \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \ldots & \Sigma_{1k} & \ldots & \Sigma_{1m} \\ \Sigma_{21} & \ldots & \Sigma_{2k} & \ldots & \Sigma_{2m} \\ \vdots & & \vdots & & \vdots \\ \Sigma_{k1} & \ldots & \Sigma_{kk} & \ldots & \Sigma_{km} \\ \vdots & & \vdots & & \vdots \\ \Sigma_{m1} & \ldots & \Sigma_{mk} & \ldots & \Sigma_{mm} \end{bmatrix} \tag{11}$$

and

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

is the overall number of variables, being p_k the number of variables in $\underline{X}_k$, the k-th set.

The generalized Wilks Λ statistic (Wilks, 1932, 1935, 1946; Anderson, 1984; Coelho, 1992)

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

where $|\cdot|$ stands for the determinant and V is either the MLE of Σ or the sample variance-covariance matrix of $\underline{X}$, split in a manner similar to Σ in (11) (being thus V_{kk} either the MLE of Σ_{kk} or the sample variance-covariance matrix of $\underline{X}_k$) is, for a sample of size $n+1$, the $(2/(n+1))$th power of the likelihood ratio test statistic to test the null hypothesis of independence of the m sets of variables,

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

It may be shown that the generalized Wilks Λ statistic may be written as

$$\Lambda = \prod_{k=1}^{m-1} \Lambda_{k(k+1,\ldots,m)} \tag{13}$$

where $\Lambda_{k(k+1,\ldots,m)}$ denotes the Wilks Λ statistic used to test the independence between the k-th set and the set formed by joining sets $k+1$ 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

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

Under the null hypothesis (12) the $m-1$ Λ statistics on the right hand side of (13) are independent (Anderson, 1984; Coelho, 1992), where $\Lambda_{k(k+1,\ldots,m)}$ has the same distribution as $\prod_{j=1}^{p_k} Y_j$, where, for a sample of size $n+1$, with $n \geq p$, Y_j are p_k independent Beta random variables with parameters $(n+1-q_k-j)/2$ and $q_k/2$, where $q_k = p_{k+1} + \ldots + p_m$ (Anderson, 1984, Theorem. 9.3.2). Then, from this fact and from (13) above we may write the h-th moment of Λ under (12), for a sample of size $n+1$, as

$$E\left(\Lambda^h\right) = \prod_{k=1}^{m-1} \prod_{j=1}^{p_k} \frac{\Gamma\left(\frac{n+1-j}{2}\right) \Gamma\left(\frac{n+1-q_k-j}{2} + h\right)}{\Gamma\left(\frac{n+1-j}{2} + h\right) \Gamma\left(\frac{n+1-q_k-j}{2}\right)} \tag{14}$$

where p_k is the number of variables in the k-th set and $q_k = p_{k+1} + \ldots + p_m$.

In this section we will apply the results obtained so far to the distribution of the generalized Wilks Lambda statistic, when there are two or three sets of variables with an odd number of variables.

Theorem 3 *When among the* m (≥ 2) *sets of variables there are two or three that have an odd number of variables, then, under* (12) *and for a sample of size* $n+1$,

$$W \sim MGIG\left(\frac{1}{B\left(a,\frac{1}{2}\right)\sqrt{\pi}}\,\frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\nu!\,(a+\nu)};\, r_1^{**},\ldots,r_{p-2}^{**}\,\middle|\, r_1^{*},\ldots,r_{p-2}^{*},1;\right.$$
$$\underbrace{\frac{n+1-p}{2},\ldots,\frac{n-2}{2}}_{p-2}\,\middle|\,\underbrace{\frac{n+1-p}{2},\ldots,\frac{n-2}{2},a+\nu}_{p-1};$$
$$\left.\nu=1,\ldots,\frac{q_{m-2}}{2}-2\,\middle|\,\nu=0,\nu\geq\frac{q_{m-2}}{2}-1\right)$$

where $W=-\log\Lambda$, $a=\frac{n+1-q_{m-2}}{2}$, $q_k=p_{k+1}+\ldots+p_m$ *and the shape parameters* r_j^* *and* r_j^{**} $(j=1,\ldots,p-2)$ *are given by*

$$r_j^* = \sum_{k=1}^{m-2} r_{k,j-p_k^*} + r^*_{m-1,j-p_k^*} \qquad (j=1,\ldots,p-2) \tag{15}$$

and

$$r_j^{**} = \sum_{k=1}^{m-2} r_{k,j-p_k^*} + r^{**}_{m-1,j-p_k^*} \qquad (j=1,\ldots,p-2) \tag{16}$$

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

$$\begin{aligned} &r_{k,j-p_k^*}=0 && if \quad p_k^*\geq j\\ &r^{**}_{m-1,j-p_k^*}=r^{*}_{m-1,j-p_k^*}=0 && if \quad p_k^*\geq j \quad or \quad j=p-2\,, \end{aligned} \tag{17}$$

and where for $k{=}1,\ldots,m{-}2$,

$$r_{kj} = \begin{cases} h_{kj} & j=1,2\\ r_{k,j-2}+h_{kj} & j=3,\ldots,p_k+q_k-2 \end{cases} \tag{18}$$

with

$$h_{kj} = (number\ of\ elements\ in\ \{p_k,q_k\}\ greater\ or\ equal\ to\ j)-1 \tag{19}$$

and, for $\nu = 1, \ldots, \frac{q_{m-2}}{2} - 2$

$$r^{**}_{m-1,j} = \begin{cases} r_j & j = 2n \ \left(n = 1, \ldots, \frac{p_m-3}{2}\right), \ j \neq 2\nu \\ & j = p_m - 1, \ldots, q_{m-2} - 3 \\ r_j + 1 & j = 1 + 2n \ \left(n = 0, \ldots, \frac{p_m-3}{2}\right) \\ r_j + 2 & j = 2\nu \end{cases} \tag{20}$$

and for $\nu = 0$ *and* $\nu \geq \frac{q_{m-2}}{2} - 1$,

$$r^{*}_{m-1,j} = \begin{cases} r_j & j = 2n \ \left(n = 1, \ldots, \frac{p_m-3}{2}\right) \\ & j = p_m - 1, \ldots, q_{m-2} - 3 \\ r_j + 1 & j = 1 + 2n \ \left(n = 0, \ldots, \frac{p_m-3}{2}\right) \end{cases} \tag{21}$$

with

$$r_j = \begin{cases} h^*_j & j = 1, 2 \\ r^*_{j-2} + h^*_j & j = 3, \ldots, q_{m-2} - 3 \end{cases} \tag{22}$$

where

$$h^*_j = (\textit{number of elements of } \{p_{m-1} - 1, p_m\} \textit{ greater or equal to } j) - 1\,. \tag{23}$$

Proof: From (14) we know that the characteristic function of $W = -\log \Lambda$, where Λ represents the generalized Wilks Λ statistic, may be written, for a sample of size $n + 1$, as

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

where $\mathrm{i} = (-1)^{1/2}$, p_k is the number of variables in the k-th set ($k = 1, \ldots, m$) and $q_k = p_{k+1} + \ldots + p_m$.

Without any loss of generality, let the two or three sets with an odd number of variables be the last ones among the m sets of variables.

Then, it will be for $k = m-1$ that the product

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

will have p_k and q_k both odd. This product has to be handled in the same way the characteristic function of W_1 was in the proof of Theorem 2 in section 3.

We may write the characteristic function of W as

$$\begin{aligned} E\left(\mathrm{e}^{\mathrm{i}tW}\right) = \prod_{k=1}^{m-3} & \underbrace{\prod_{j=1}^{p_k+q_k-2} \left(\frac{n-p_k-q_k+j}{2}\right)^{r_{kj}} \left(\frac{n-p_k-q_k+j}{2} - \mathrm{i}t\right)^{-r_{kj}}}_{p_k \text{ even}} \\ & \underbrace{\prod_{j=1}^{p_{m-2}+q_{m-2}-2} \left(\frac{n-p_{m-2}-q_{m-2}+j}{2}\right)^{r_{m-2,j}} \left(\frac{n-p_{m-2}-q_{m-2}+j}{2} - \mathrm{i}t\right)^{-r_{m-2,j}}}_{q_{m-2} \text{ even}} \\ & \underbrace{\prod_{j=1}^{p_{m-1}} \frac{\Gamma\left(\frac{n+1-j}{2}\right)\,\Gamma\left(\frac{n+1-q_{m-1}-j}{2} - \mathrm{i}t\right)}{\Gamma\left(\frac{n+1-j}{2} - \mathrm{i}t\right)\,\Gamma\left(\frac{n+1-q_{m-1}-j}{2}\right)}}_{\text{both } p_{m-1} \text{ and } q_{m-1} \text{ odd}} \end{aligned} \tag{24}$$

where r_{kj} $(k = 1, \ldots, m-2;\ j = 1, \ldots, p_k+q_k-2)$ is given by (18) and (19) and $q_{m-1} = p_m$. Then, using the same procedure used to handle the characteristic function of W_1 in the proof of Theorem 2 we may write

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

as in the result in the last row of (9) with

$$\begin{aligned} p &= p_{m-1} \\ f &= q_{m-1} = p_m \\ p+f &= p_{m-1} + p_m = q_{m-2} \\ a &= \frac{n+1-q_{m-1}-p_{m-1}}{2} = \frac{n+1-q_{m-2}}{2}\,, \end{aligned}$$

and then we have for $a = \frac{n+1-q_{m-2}}{2}$

$$
\begin{aligned}
E\left(\mathrm{e}^{\mathrm{i}tW}\right) &= \\
&\frac{1}{B\left(a,\frac{1}{2}\right)\sqrt{\pi}}\left[\sum_{\nu=1}^{\frac{q_{m-2}}{2}-2}\frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\nu!\,(a+\nu)}\prod_{j=1}^{q_{m-2}-3}\left(a+\frac{j}{2}\right)^{r^{**}_{m-1,j}}\left(a+\frac{j}{2}-\mathrm{i}t\right)^{-r^{**}_{m-1,j}}\right.\\
&\qquad+\sum_{\nu=0,\frac{q_{m-2}}{2}-1}^{\infty}\frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\nu!\,(a+\nu)}(a+\nu)\,(a+\nu-\mathrm{i}t)^{-1}\\
&\qquad\left.\prod_{j=1}^{q_{m-2}-3}\left(a+\frac{j}{2}\right)^{r^{*}_{m-1,j}}\left(a+\frac{j}{2}-\mathrm{i}t\right)^{-r^{*}_{m-1,j}}\right]\\
&\qquad\prod_{k=1}^{m-2}\prod_{j=1}^{p_k+q_k-2}\left(\frac{n-p_k-q_k+j}{2}\right)^{r_{kj}}\left(\frac{n-p_k-q_k+j}{2}-\mathrm{i}t\right)^{-r_{kj}}
\end{aligned}
$$

so that we may finally write

$$
\begin{aligned}
E\left(\mathrm{e}^{\mathrm{i}tW}\right) &= \\
&\frac{1}{B\left(a,\frac{1}{2}\right)\sqrt{\pi}}\left[\sum_{\nu=1}^{\frac{q_{m-2}}{2}-2}\frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\nu!\,(a+\nu)}\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r^{**}_{j}}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r^{**}_{j}}\right.\\
&\qquad+\sum_{\nu=0,\frac{q_{m-2}}{2}-1}^{\infty}\frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\nu!\,(a+\nu)}(a+\nu)\,(a+\nu-\mathrm{i}t)^{-1}\\
&\qquad\left.\prod_{j=1}^{p-2}\left(\frac{n-p+j}{2}\right)^{r^{*}_{j}}\left(\frac{n-p+j}{2}-\mathrm{i}t\right)^{-r^{*}_{j}}\right]
\end{aligned}
$$

with r^{**}_j and r^*_j given by (15) through (23). □

If one wants the exact distribution of the Wilks Λ statistic used to test the independence of two sets with an odd number of variables then one only has to set $m = 2$ in the Theorem above. Although some of the expressions may then seem a bit strange they all work out well, since for $m = 2$ we will have

$$q_{m-2} = q_0 = p_1 + p_2 = p\,,$$

and given that we use the rule that any summation with an upper bound smaller than its lower bound evaluates to zero and any product with an upper bound smaller than its lower bound evaluates to one, thus, for $m = 2$, (15) and (16) will respectivelly yield

$$r^*_j = r^*_{1,j-p^*_k}$$

and

$$r_j^{**} = r_{1,j-p_k^*}^{**}$$

while in (24) only the third product is active in this case since the first product vanishes, or rather, evaluates to one, since its upper bound is smaller than its lower bounder while the second product also evaluates to one given that we take $r_{0,j} = 0$ for any j.

This way if $m = 2$ and both sets of variables have an odd number of variables, Theorem 3 still gives the exact distribution, being enough to use in its statement $q_{m-2} = p$.

5 Discussion and final remarks

Kabe (1962) already obtained the exact distribution of the Wilks Λ statistic for the case of only two groups of variables under the form of hypergeometric series related with the distribution of the sum of independent Gamma random variables. However, the expressions for the p.d.f. of such distributions involved representations with several infinite summations.

Our aim was to obtain representations for both the p.d.f. and the c.d.f. of the product of an odd number of particular independent Beta random variables and at least for some cases of the generalized Wilks Λ statistic, under the null hypothesis, which would involve only one infinite summation, that is, which would have the form of an infinite mixture of distributions whose p.d.f.'s and c.d.f.'s would not involve any series representations.

Our goal was attained by obtaining such distributions under the form of infinite mixtures of GIG distributions. The representation of both the p.d.f. and c.d.f. of such distributions does not involve any infinite sums or unsolved integrals. This way, the esults obtained seem to be much adequate for the computation of quantiles, mainly for very small values of the parameter a in Theorem 2 or of the sample size in the distributin of the generalized Wilks Λ statistic, since in these cases the weights in the mixture will vanish rather quickly so that just a few terms in the mixture may be enough to obtain a very accurate result.

Anyway, further numerical studies concerning the comparison of the near-exact approximations in Grilo and Coelho (2003) and Alberto and Coelho (2003) and the exact distribution in this paper need to be carried out, in order to be able to better assess the situations where each of these distributions is the most useful. These calculations, falling out of the scope of this paper, were left for further work.

Also left for further work is the application of the exact distributions studied in this paper in building new near-exact distributions for both the product of an odd number of independent Beta random variables and the generalized Wilks Λ statistic.

References

Alberto, R. P. and Coelho, C. A., 2003. Study of the Quality of Several Asymptotic and Near-Exact Approximations Based on Moments for the Distributions of the Wilks Lambda. (submitted to the *Journal of Statistical Planning and Inference*).

Anderson, T. W., 1984. *An Introduction to Multivariate Statistical Analysis*, 2nd ed., J. Wiley & Sons, New York.

Coelho, C. A., 1992. *Generalized Canonical Analysis*, Ph.D. thesis. The University of Michigan, Ann Arbor, MI.

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., 1999. Addendum to the paper "The Generalized Integer Gamma Distribution – A Basis for Distributions in Multivariate Statistics". *Journal of Multivariate Analysis*, **69**, 281–285.

Grilo, L. M. and Coelho, C. A., 2003. Development and Comparative Study of two Near-Exact Approximations to the Distribution of the Product of an Odd Number of Independent Beta Random Variables. (submitted to the *Journal of Statistical Planning and Inference*).

Kabe, D. G., 1962. On the exact distribution of a class of multivariate test criteria. *Ann. Math. Statist.*, **33**, 1197–1200.

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. *Econometrica* **3**, 309–326.

Wilks, S. S., 1946. Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Ann. Math. Stat.*, **17**, 257–281.