SOME RESULTS ON NON-CENTRAL BETA DISTRIBUTIONS

Andrea Ongaro

Dipartimento di Economia, Metodi Quantitativi e Strategie di Impresa, Università degli Studi di Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italia

Carlo Orsi [1]

Dipartimento di Economia, Metodi Quantitativi e Strategie di Impresa, Università degli Studi di Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italia

1. INTRODUCTION

The beta distribution plays a central role in modeling a wide variety of random phenomena which can take on values with lower and upper bounds. For example, such distribution has been employed to model the time to perform an activity in the project management area (Malcolm *et al.*, 1959) and has been used to fit the sample distribution of soil parameters (Haskett *et al.*, 1995). In particular, the beta distribution is useful for modeling proportions. Such distribution, in fact, has been applied to fit the monthly frequency distribution of the daily sunshine duration measurements collected at various locations in Algeria (Ettoumi *et al.*, 2002) and has been employed to model heterogeneity in the probability of HIV transmission per sexual contact (Wiley *et al.*, 1989).

Therefore, the beta distribution is versatile enough to model different types of data. Despite this, several generalizations of such distribution have been developed in order to achieve further flexibility. In this regard, see Johnson *et al.* (1995) and McDonald and Xu (1995). However, neither the beta distribution nor the major part of its generalizations offer flexibility to the limits at 0 and 1 of their probability density functions. On the contrary, the class of the so-called "non-central" beta distributions display such behavior which, to our knowledge, has never been studied. Such class, however, is affected by strong limitations in terms of tractability and interpretability.

That said, in this paper we shall introduce a new distribution on the real interval $(0, 1)$ which can be viewed as a more easily tractable and interpretable analogue of the doubly non-central beta distribution (Johnson *et al.*, 1995). More specifically, the present paper is organized as follows. In Section 2 some notations are introduced and some useful facts are recalled. Particular attention is focused on the non-central chi-squared distribution, which plays a crucial role in the study of the family of the generalizations of the beta distribution we are interested in. In Section 3 we shall provide a novel analysis of the standard non-central beta distribution. In particular, the perturbation representation of such distribution highlights its uneasy tractability and interpretability. In Section 4 we shall perform a preliminary investigation of the properties of a new non-central beta distribution. The analysis of such distribution is based on the same techniques employed in the study of the standard one and will allow a better understanding of the similarities and differences between the two distributions.

---

[1] Corresponding Author. E-mail: c.orsi@campus.unimib.it

2. Preliminaries

It is well known that if Y_i, $i = 1, 2$, are independent χ^2 random variables with $2\alpha_i > 0$ degrees of freedom, denoted by $\chi^2_{2\alpha_i}$, then the random variable:

$$X = \frac{Y_1}{Y_1 + Y_2} \tag{1}$$

has a beta distribution with shape parameters α_i, denoted by Beta(α_1, α_2). Its density is:

$$\text{Beta}(x; \alpha_1, \alpha_2) = \frac{x^{\alpha_1 - 1} (1 - x)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)}, \quad 0 < x < 1. \tag{2}$$

The beta distribution can also be obtained as conditional distribution of X given the sum $Y^+ = Y_1 + Y_2$, as a consequence of the independence of X and Y^+. The latter is a characterizing property of (independent) gamma random variables.

Let us now recall the definition and some useful properties of the non-central chi-squared distribution, which plays a crucial role in the study of the family of the generalizations of the beta distribution we are interested in.

Let W_k, $k = 1, \ldots, g$, be independent and normally distributed random variables with means μ_k and unitary variances. Then, a random variable is said to have a non-central chi-squared distribution with $g > 0$ degrees of freedom and non-centrality parameter $\lambda = \sum_{k=1}^{g} \mu_k^2 \geq 0$, denoted by $\chi'^2_g(\lambda)$, if it is distributed as:

$$Y' = \sum_{k=1}^{g} W_k^2$$

(Johnson et $al.$, 1995). The case $\lambda = 0$ corresponds to the χ^2_g distribution.

The density function $f_{Y'}$ of $Y' \sim \chi'^2_g(\lambda)$ can be expressed as:

$$f_{Y'}(y; g, \lambda) = \sum_{i=0}^{+\infty} \frac{e^{-\frac{\lambda}{2}} \left(\frac{\lambda}{2}\right)^i}{i!} \frac{y^{\frac{g+2i}{2} - 1} e^{-\frac{y}{2}}}{\Gamma\left(\frac{g+2i}{2}\right) 2^{\frac{g+2i}{2}}}, \quad y > 0, \tag{3}$$

i.e. as series of the χ^2_{g+2i} densities, $\forall i \in \mathbb{N} \cup \{0\}$, weighted by the probabilities of a Poisson random variable with mean $\lambda/2$, $\lambda \geq 0$ (the case $\lambda = 0$ corresponding to a random variable degenerate at zero).

In view of Eq. (3), the $\chi'^2_g(\lambda)$ distribution admits the following mixture representation.

PROPOSITION 1 (MIXTURE REPRESENTATION). *Let Y' have a $\chi'^2_g(\lambda)$ distribution and M be a Poisson random variable with mean $\lambda/2$. Then, Y' admits the following representation:*

$$Y' \,|\, (M = i) \sim \chi^2_{g+2i}. \tag{4}$$

The $\chi'^2_g(\lambda)$ density can assume a larger variety of shapes than the central one. In particular, the limit at 0 of such density is decreasing in λ when $g = 2$.

PROPOSITION 2 (LIMIT AT 0 OF THE $\chi'^2_g(\lambda)$ DENSITY WHEN $g = 2$). *Let Y' be a $\chi'^2_g(\lambda)$ random variable with $g = 2$ and $\lambda \geq 0$. Let $f_{Y'}$ be the density of Y'. Then:*

$$\lim_{y \to 0^+} f_{Y'}(y; 2, \lambda) = \frac{1}{2} e^{-\frac{\lambda}{2}}. \tag{5}$$

In this regard, see Figure 1.

Finally, we recall that the non-central chi-squared distribution is reproductive with respect to both degrees of freedom and non-centrality parameter.
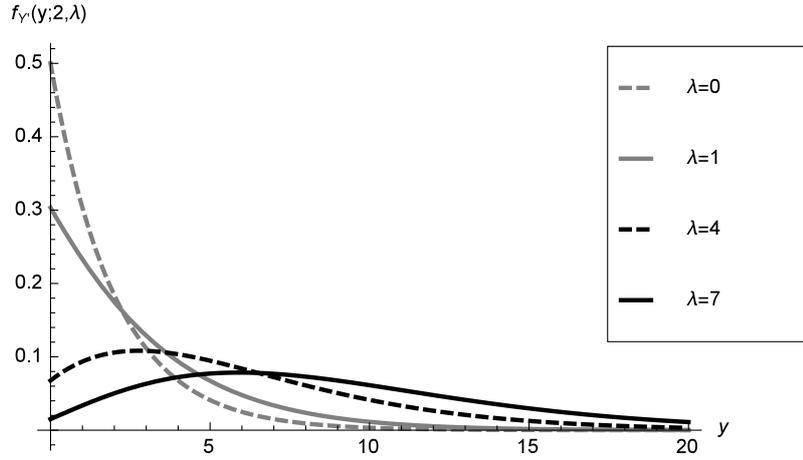
Figure 1 – Plots of the density of $Y' \sim \chi_g'^2(\lambda)$ for $g = 2$ and selected values of λ.

PROPOSITION 3 (REPRODUCTIVE PROPERTY). *If Y_j', $j = 1, \ldots, m$, are independent with $\chi_{g_j}'^2(\lambda_j)$ distributions, then:*

$$Y'^+ = \sum_{j=1}^{m} Y_j' \sim \chi_g'^2(\lambda),$$

where $g = \sum_{j=1}^{m} g_j$ and $\lambda = \sum_{j=1}^{m} \lambda_j$.

Such property can be easily derived from its characteristic function (Johnson *et al.*, 1995).

3. THE STANDARD NON-CENTRAL BETA DISTRIBUTION

Let us now recall the definition and some useful representations of the standard non-central beta distribution.

Let Y_i', $i = 1, 2$, be independent $\chi_{2\alpha_i}'^2(\lambda_i)$ random variables. Then, a random variable is said to have a doubly non-central beta distribution with shape parameters α_1, α_2 and non-centrality parameters λ_1, λ_2, denoted by $\text{DNCB}(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$, if it is distributed as

$$X' = \frac{Y_1'}{Y_1' + Y_2'} \tag{6}$$

(Johnson *et al.*, 1995).

The DNCB density can be easily derived by using the mixture representation of the non-central χ^2 distribution. Specifically, let M_i, $i = 1, 2$, be independent Poisson random variables with means $\lambda_i/2$. Conditionally on (M_1, M_2), X' has a $\text{Beta}(\alpha_1 + M_1, \alpha_2 + M_2)$ distribution, $Y_i'|(M_1, M_2)$ being independent with distributions $\chi_{2\alpha_i + 2M_i}^2$, $i = 1, 2$. Therefore, the density function $f_{X'}$ of $X' \sim \text{DNCB}(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$ can be expressed as:

$$f_{X'}(x; \alpha_1, \alpha_2, \lambda_1, \lambda_2) = \sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} \frac{e^{-\frac{\lambda_1}{2}}\left(\frac{\lambda_1}{2}\right)^j}{j!} \frac{e^{-\frac{\lambda_2}{2}}\left(\frac{\lambda_2}{2}\right)^k}{k!} \frac{x^{\alpha_1+j-1}(1-x)^{\alpha_2+k-1}}{B(\alpha_1 + j, \alpha_2 + k)}, \quad 0 < x < 1,$$
$$\tag{7}$$

i.e. as double series of the $\text{Beta}(\alpha_1 + j, \alpha_2 + k)$ densities, $\forall j, k \in \mathbb{N} \cup \{0\}$, weighted by the joint probabilities of the bivariate random variable (M_1, M_2), where M_i, $\forall i = 1, 2$, are independent with $\text{Poisson}(\lambda_i/2)$ distributions.

The above discussion directly leads to the following mixture representation.

PROPOSITION 4 (MIXTURE REPRESENTATION). *Let X' have a* DNCB $(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$ *distribution. Then, X' admits the following representation:*

$$X' \,|\, (M_1 = j,\, M_2 = k) \sim \text{Beta}\,(\alpha_1 + j, \alpha_2 + k)\ , \tag{8}$$

where M_i, $i = 1, 2$, are independent Poisson random variables with means $\lambda_i/2$.

The perturbation representation of the doubly non-central beta distribution highlights the uneasy tractability and interpretability of such distribution.

PROPOSITION 5 (PERTURBATION REPRESENTATION). *Let $X' \sim$ DNCB $(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$. Then, the density $f_{X'}$ of X' can be written as:*

$$f_{X'}\,(x; \alpha_1, \alpha_2, \lambda_1, \lambda_2) = \text{Beta}\,(x; \alpha_1, \alpha_2) \cdot e^{-\frac{\lambda^+}{2}} \, \Psi_2 \left[\alpha^+; \alpha_1, \alpha_2; \frac{\lambda_1}{2} x, \frac{\lambda_2}{2}\,(1 - x) \right]\ , \tag{9}$$

where $\alpha^+ = \alpha_1 + \alpha_2$, $\lambda^+ = \lambda_1 + \lambda_2$,

$$\Psi_2\,[\alpha; \gamma, \gamma'; x, y] = \sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} \frac{(\alpha)_{j+k}}{(\gamma)_j\,(\gamma')_k} \frac{x^j}{j!} \frac{y^k}{k!}, \quad x, y \ge 0 \tag{10}$$

is the Humbert's confluent hypergeometric function (Srivastava and Karlsson, 1985) and

$$(a)_0 = 1, \quad (a)_i = a\,(a+1)\ldots(a+i-1), \qquad \forall i \in \mathbb{N} \tag{11}$$

is the Pochhammer's symbol or ascendent factorial.

The perturbing factor of the beta density in the doubly non-central beta distribution is a function in two variables given by the sum of a double power series (Eq. (10)). Such function has not a simple behavior and, to our knowledge, is not reducible to a more tractable analytical form. Therefore, it's not easy to understand the effect of such factor.

Despite this, the DNCB density can assume a large variety of shapes far beyond the beta model ones. In this regard, Figures 2, 3, 4, 5, 6 display the DNCB density for selected values of the parameters. In particular, the effect of the perturbing factor can be clearly seen when $\alpha_1 = \alpha_2 = 1$, because in this case the beta density reduces to the uniform one (see Figures 4, 5, 6).
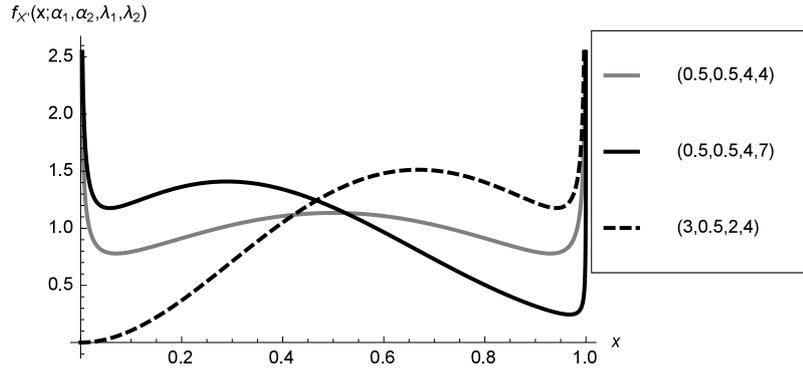
Figure 2 – Plots of the density of $X' \sim$ DNCB $(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$ for selected values of $(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$.

Moreover, it is well known that the limits at 0 and 1 of the beta density are equal to 0 or $+\infty$ if $\alpha_i \ne 1$ and are equal to 1 if $\alpha_i = 1$, $i = 1, 2$. When $\alpha_i \ne 1$ the DNCB density shows the same limiting characteristics as the beta model ones. On the contrary, when $\alpha_i = 1$ the doubly non-central beta density shows the attractive feature of taking on arbitrary finite and positive limits at 0 and 1. To our knowledge, such characteristic of the DNCB density has never been studied. More specifically, such limits have the following expressions.
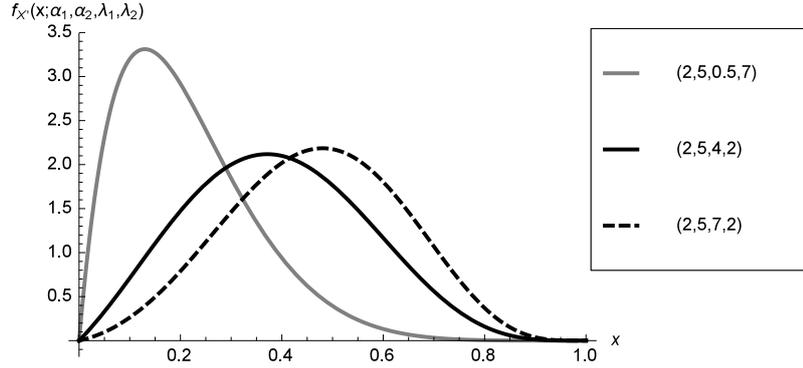
Figure 3 – Plots of the density of $X' \sim \mathrm{DNCB}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right)$ for selected values of $(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$.
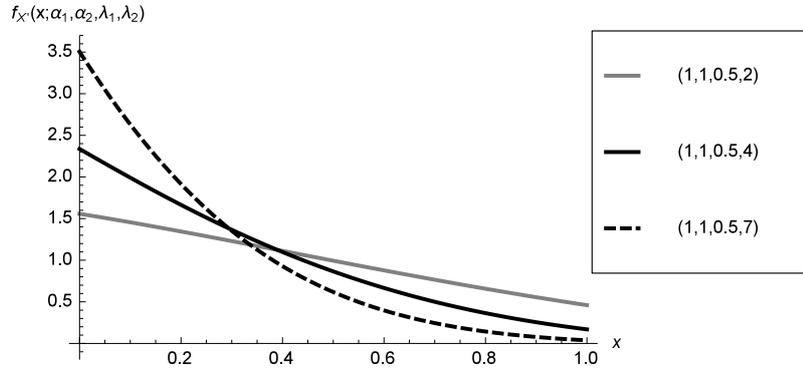
Figure 4 – Plots of the density of $X' \sim \mathrm{DNCB}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right)$ for $\alpha_1 = \alpha_2 = 1$ and selected values of λ_1, λ_2.

REMARK 6 (LIMITS AT 0 AND 1 OF THE DNCB DENSITY WHEN $\alpha_1 = \alpha_2 = 1$). *Let X' $\sim \mathrm{DNCB}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right)$. Then, the limits at 0 and 1 of the density $f_{X'}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right)$ of X' when $\alpha_1 = \alpha_2 = 1$ are:*

$$\lim_{x \to 0^+} f_{X'}\left(x; 1, 1, \lambda_1, \lambda_2\right) = e^{-\frac{\lambda_1}{2}} \left(\frac{\lambda_2}{2} + 1\right), \tag{12}$$

$$\lim_{x \to 1^-} f_{X'}\left(x; 1, 1, \lambda_1, \lambda_2\right) = e^{-\frac{\lambda_2}{2}} \left(\frac{\lambda_1}{2} + 1\right). \tag{13}$$

Notice that, under the case of unitary shape parameters, the limit at 0 of the DNCB density is increasing in λ_2 for any fixed $\lambda_1 > 0$ and decreasing in λ_1 for any fixed $\lambda_2 > 0$, while the limit at 1 is decreasing in λ_2 for any fixed $\lambda_1 > 0$ and increasing in λ_1 for any fixed $\lambda_2 > 0$.

Finally, by exploiting the mixture representation in Eq. (8), it is possible to derive a new general formula for the moments of the standard non-central beta distribution. Specifically, the moments of the DNCB distribution can be written and represented as follows.

PROPOSITION 7 (MOMENTS). *Let $X' \sim \mathrm{DNCB}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right)$. Let M_j, $j = 1, 2$, be independent Poisson random variables with means $\lambda_j/2$ and $M^+ = M_1 + M_2$. Then, $\forall r \in \mathbb{N}$, the r-th moment of X' can be written as:*

$$\mathbb{E}\left[\left(X'\right)^r \mid M^+\right] = \frac{1}{\left(\alpha^+ + M^+\right)_r} \sum_{i=0}^{M^+} (\alpha_1 + i)_r \binom{M^+}{i} \left(\frac{\lambda_1}{\lambda^+}\right)^i \left(1 - \frac{\lambda_1}{\lambda^+}\right)^{M^+ - i} \tag{14}$$

$f_{X'}(x; \alpha_1, \alpha_2, \lambda_1, \lambda_2)$
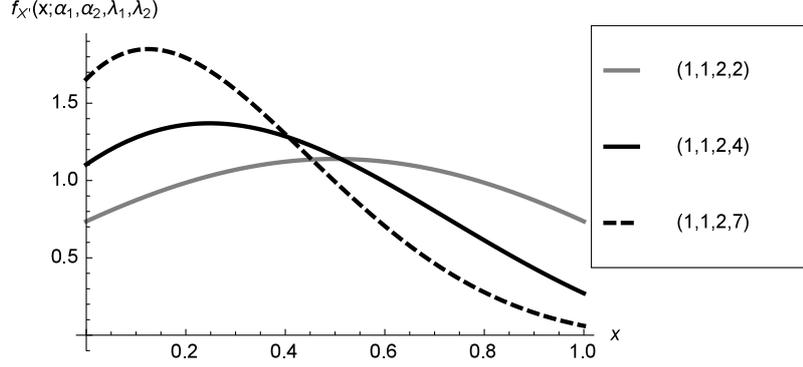
Figure 5 – Plots of the density of $X' \sim \mathrm{DNCB}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right)$ for $\alpha_1 = \alpha_2 = 1$ and selected values of λ_1, λ_2.

$f_{X'}(x; \alpha_1, \alpha_2, \lambda_1, \lambda_2)$
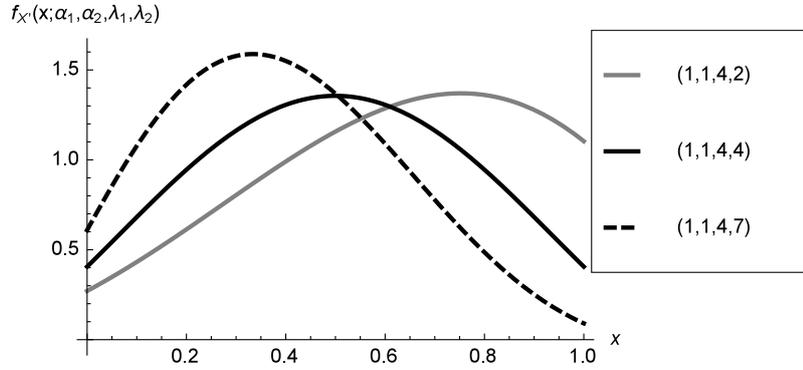
Figure 6 – Plots of the density of $X' \sim \mathrm{DNCB}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right)$ for $\alpha_1 = \alpha_2 = 1$ and selected values of λ_1, λ_2.

and admits the following representation:

$$\mathbb{E}\left[(X')^r\right] = \mathbb{E}_{(M_1, M_2)}\left[\frac{(\alpha_1 + M_1)_r}{(\alpha^+ + M^+)_r}\right]. \tag{15}$$

PROOF. Observe first that, in the notation of Eq. (6), $Y'^+ = Y'_1 + Y'_2$ and $X' = Y'_1/Y'^+$ are independent conditionally on M^+. In this regard, notice that Eq. (8) can be rewritten as follows:

$$X'|\left(M_1, M^+\right) \sim \mathrm{Beta}\left(\alpha_1 + M_1, \alpha_2 + M^+ - M_1\right); \tag{16}$$

similarly, since $Y'^+ \sim \chi'^2_{2\alpha^+}\left(\lambda^+\right)$ in view of Proposition 3, Eq. (4) can be rewritten as follows:

$$Y'^+\big|\left(M_1, M^+\right) \stackrel{d}{=} Y'^+\big|\, M^+ \sim \chi^2_{2\alpha^+ + 2M^+}. \tag{17}$$

Therefore, X' and Y'^+ are conditionally independent given (M_1, M^+), the latter being a characterizing property of independent gamma random variables. Hence, conditionally on (M_1, M^+), the joint distribution of (X', Y'^+) factorizes into the marginal distributions of X' and Y'^+. Finally, the joint density function of $(X', Y'^+)|\, M^+$ turns out to factorize into the marginal density functions of $X'|M^+$ and $Y'^+|M^+$. Indeed, by recalling that

$M_1 | M^+ \sim \text{Binomial}\left(M^+, \frac{\lambda_1}{\lambda^+}\right)$, one can obtain:

$$
\begin{aligned}
f_{(X', Y'+)|M^+}(x, y) &= \sum_{i=0}^{M^+} \Pr\left(M_1 = i \,\middle|\, M^+\right) \cdot f_{(X', Y'+)|(M_1, M^+)}(x, y) = \\
&= f_{Y'+|M^+}(y) \cdot f_{X'|M^+}(x),
\end{aligned}
$$

where:

$$
f_{X'|M^+}(x) = \sum_{i=0}^{M^+} \text{Binomial}\left(i; M^+, \frac{\lambda_1}{\lambda^+}\right) \cdot \text{Beta}\left(x; \alpha_1 + i, \alpha_2 + M^+ - i\right),
$$

with:

$$
\text{Binomial}\left(i; M^+, \frac{\lambda_1}{\lambda^+}\right) = \binom{M^+}{i} \left(\frac{\lambda_1}{\lambda^+}\right)^i \left(1 - \frac{\lambda_1}{\lambda^+}\right)^{M^+ - i}, \qquad i = 0, \ldots, M^+.
$$

That said, because of the foregoing arguments, the following holds:

$$
\mathbb{E}\left[(X')^r \,\middle|\, M^+\right] = \frac{\mathbb{E}\left[(Y'_1)^r \,\middle|\, M^+\right]}{\mathbb{E}\left[(Y'+)^r \,\middle|\, M^+\right]}, \tag{18}
$$

where, in view of the general formula for the moments of the gamma distribution (Johnson *et al.*, 1994):

$$
\mathbb{E}\left[(Y'_1)^r \,\middle|\, M^+\right] = \mathbb{E}\left[2^r\,(\alpha_1 + M_1)_r \,\middle|\, M^+\right] = 2^r \sum_{i=0}^{M^+} (\alpha_1 + i)_r \binom{M^+}{i} \left(\frac{\lambda_1}{\lambda^+}\right)^i \left(1 - \frac{\lambda_1}{\lambda^+}\right)^{M^+ - i} \tag{19}
$$

and:

$$
\mathbb{E}\left[(Y'+)^r \,\middle|\, M^+\right] = 2^r \left(\alpha^+ + M^+\right)_r. \tag{20}
$$

Hence, Eq. (14) is established.

Finally, by letting $L \sim \text{Binomial}(M^+, \lambda_1 / \lambda^+)$ in Eq. (14), one can obtain that:

$$
\mathbb{E}\left[(X')^r \,\middle|\, M^+\right] = \frac{\mathbb{E}\left[(\alpha_1 + L)_r\right]}{(\alpha^+ + M^+)_r}. \tag{21}
$$

Thus, in view of the fact that $L \overset{d}{=} M_1 | M^+$, Eq. (15) is established. $\qquad\square$

The formula in Eq. (14) reduces the computation of moments of the DNCB distribution from a double series to a single one. In particular, by setting $r = 1$ and $r = 2$ and after some manipulations, the first two moments of the DNCB distribution can be written in terms of perturbations of the corresponding moments of the beta distribution as follows:

$$
\mathbb{E}(X') = \frac{\alpha_1}{\alpha^+} e^{-\frac{\lambda^+}{2}} \left[{}_1F_1\left(\alpha^+; \alpha^+ + 1; \frac{\lambda^+}{2}\right) + \frac{\alpha^+ \frac{\lambda_1}{2}}{\alpha_1\,(\alpha^+ + 1)}\,{}_1F_1\left(\alpha^+ + 1; \alpha^+ + 2; \frac{\lambda^+}{2}\right)\right], \tag{22}
$$

$$
\begin{aligned}
\mathbb{E}\left[(X')^2\right] &= \frac{(\alpha_1)_2}{(\alpha^+)_2} e^{-\frac{\lambda^+}{2}} \left[{}_1F_1\left(\alpha^+; \alpha^+ + 2; \frac{\lambda^+}{2}\right) + \frac{\alpha^+ \lambda_1}{\alpha_1\,(\alpha^+ + 2)} \cdot \right. \\
&\left. \cdot\ {}_1F_1\left(\alpha^+ + 1; \alpha^+ + 3; \frac{\lambda^+}{2}\right) + \frac{(\alpha^+)_2 \left(\frac{\lambda_1}{2}\right)^2}{(\alpha_1)_2\,(\alpha^+ + 2)_2}\,{}_1F_1\left(\alpha^+ + 2; \alpha^+ + 4; \frac{\lambda^+}{2}\right)\right], \tag{23}
\end{aligned}
$$

where:

$$
{}_1F_1\left(a; b; x\right) = \sum_{i=0}^{+\infty} \frac{(a)_i}{(b)_i} \frac{x^i}{i!}, \qquad a, b > 0,\ x \geq 0
$$

is the generalized hypergeometric function ${}_pF_q$ with $p = 1$ and $q = 1$ coefficients respectively at numerator and denominator, otherwise known as Kummer's confluent hypergeometric function (Srivastava and Karlsson, 1985).

4. A NEW NON-CENTRAL BETA DISTRIBUTION

Our basic idea to introduce a new family of non-central beta distributions is to consider the conditional distribution of a DNCB random variable X' given $Y_1' + Y_2'$, Y_i' being independent with $\chi_{2\alpha_i}'^2(\lambda_i)$ distributions, $i = 1, 2$. Because of the above mentioned characterizing property of independent gamma random variables, such conditional distribution must depend on $Y_1' + Y_2'$ and therefore must be different from the DNCB distribution, unless $\lambda_1 = \lambda_2 = 0$. The latter case corresponds to the beta distribution.

The density of the new non-central beta distribution can be derived by using the mixture representation of the non-central χ^2 distribution. Specifically, let M_i, $i = 1, 2$, be independent Poisson random variables with means $\lambda_i/2$. Conditionally on (M_1, M_2), $X' = Y_1'/(Y_1' + Y_2')$ and $Y_1' + Y_2'$ are independent; therefore, X' given $(M_1, M_2, Y_1' + Y_2')$ has a Beta$(\alpha_1 + M_1, \alpha_2 + M_2)$ distribution. A direct application of Bayes' theorem then shows that, conditionally on $Y_1' + Y_2' = y$, for any fixed $y > 0$, (M_1, M_2) has the following joint probability function:

$$\Pr\left(M_1 = j, M_2 = k \mid Y_1' + Y_2' = y\right) = \frac{1}{{}_0F_1\left(\alpha^+, \frac{y\lambda^+}{4}\right)} \frac{\left(\frac{y\lambda_1}{4}\right)^j \left(\frac{y\lambda_2}{4}\right)^k}{j!\,k!\,(\alpha^+)_{j+k}}, \quad j, k \in \mathbb{N} \cup \{0\}, \tag{24}$$

where

$$_0F_1\left(a; x\right) = \sum_{i=0}^{+\infty} \frac{1}{(a)_i} \frac{x^i}{i!}, \qquad a > 0, \; x \geq 0 \tag{25}$$

is the generalized hypergeometric function $_pF_q$ with $p = 0$ and $q = 1$ coefficients respectively at numerator and denominator (Srivastava and Karlsson, 1985).

As the conditioning value y in Eq. (24) can be incorporated in the non-centrality parameters λ_i's, we are then led to the following definition of the above mentioned new distribution.

DEFINITION 8. *A random variable is said to have a conditional doubly non-central beta distribution with shape parameters α_1, α_2 and non-centrality parameters λ_1, λ_2, denoted by* CDNCB$(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$, *if it is distributed as*

$$X'' = \frac{Y_1'}{Y_1' + Y_2'} \bigg| Y_1' + Y_2' = 1 \tag{26}$$

where Y_i', $i = 1, 2$, are independent $\chi_{2\alpha_i}'^2(\lambda_i)$ random variables. The random variable in Eq. (26) is meant to be distributed according to the regular version of the conditional density of $Y_1' | Y_1' + Y_2' = 1$ which is proportional to the product of the densities of Y_1' and Y_2' respectively evaluated in x and $1 - x$.

By mixturing the density function of X' given $(M_1, M_2, Y_1' + Y_2' = 1)$ with respect to the joint probability function of (M_1, M_2) given $Y_1' + Y_2' = 1$, one can obtain the density function $f_{X''}$ of $X'' \sim$ CDNCB$(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$ which turns out to have the following expression:

$$f_{X''}\left(x; \alpha_1, \alpha_2, \lambda_1, \lambda_2\right) = \sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} \frac{\frac{\left(\frac{\lambda_1}{4}\right)^j \left(\frac{\lambda_2}{4}\right)^k}{j!\,k!\,(\alpha^+)_{j+k}}}{{}_0F_1\left(\alpha^+, \frac{\lambda^+}{4}\right)} \frac{x^{\alpha_1+j-1}\left(1-x\right)^{\alpha_2+k-1}}{B\left(\alpha_1+j, \alpha_2+k\right)}, \quad 0 < x < 1, \tag{27}$$

given by the double series of the Beta$(\alpha_1 + j, \alpha_2 + k)$ densities, $\forall j, k \in \mathbb{N} \cup \{0\}$, weighted by the joint probabilities of the bivariate random variable (N_1, N_2), whose probability function is given by Eq. (24) with $y = 1$. The above discussion directly leads to the following mixture representation.

PROPOSITION 9 (MIXTURE REPRESENTATION). *Let X'' have a CDNCB $(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$ distribution. Then, X'' admits the following representation:*

$$X'' \mid (N_1 = j, \ N_2 = k) \sim \text{Beta}\left(\alpha_1 + j, \alpha_2 + k\right) , \tag{28}$$

where the joint probability function of (N_1, N_2) is given by Eq. (24) with $y = 1$.

Although the distribution of (N_1, N_2) is slightly more complicated than the one of (M_1, M_2), it can be easily handled and understood in terms of the random variables (N_1, N^+), where $N^+ = N_1 + N_2$ has the following probability function:

$$\Pr\left(N^+ = i\right) = \frac{1}{{}_0F_1\left(\alpha^+; \frac{\lambda^+}{4}\right)} \frac{1}{(\alpha^+)_i} \frac{\left(\frac{\lambda^+}{4}\right)^i}{i!}, \quad i \in \mathbb{N} \cup \{0\} . \tag{29}$$

Interestingly, $N_1 \mid N^+$ shares the same distribution of $M_1 \mid M^+$, with $M^+ = M_1 + M_2$. More specifically, the following holds true.

REMARK 10. *Let (N_1, N_2) be jointly distributed as in Eq. (24) with $y = 1$ and $N^+ = N_1 + N_2$ be distributed as in Eq. (29). Then:*

$$N_i \mid N^+ \sim \text{Binomial}\left(N^+, \frac{\lambda_i}{\lambda^+}\right), \qquad i = 1, 2. \tag{30}$$

This allows us to derive the expression for the covariance between N_1 and N_2 as follows:

$$\begin{aligned}
\text{Cov}\left(N_1, N_2\right) &= \ \mathbb{E}_{N^+}\left\{\text{Cov}\left[(N_1, N_2) \mid N^+\right]\right\} + \text{Cov}\left[\mathbb{E}\left(N_1 \mid N^+\right), \mathbb{E}\left(N_2 \mid N^+\right)\right] = \\
&= \ \frac{\lambda_1 \lambda_2}{(\lambda^+)^2}\left[\text{Var}\left(N^+\right) - \mathbb{E}\left(N^+\right)\right] ,
\end{aligned} \tag{31}$$

which can be easily expressed in terms of the hypergeometric function ${}_0F_1$. An extensive numerical investigation shows that such covariance is negative.

It is also simple to check that the CDNCB admits an unconditional "ratio" type representation.

REMARK 11 (RATIO REPRESENTATION). *Let (N_1, N_2) be jointly distributed as in Eq. (24) with $y = 1$. Let (Z_1, Z_2) be conditionally independent random variables given (N_1, N_2), with $Z_i \mid (N_1, N_2) \sim \chi^2_{2\alpha_i + 2N_i}$, $i = 1, 2$. Then:*

$$\frac{Z_1}{Z_1 + Z_2} \sim \text{CDNCB}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right). \tag{32}$$

Notice that (Z_1, Z_2) are negatively correlated as $\text{Cov}\left(Z_1, Z_2\right) = 4\,\text{Cov}\left(N_1, N_2\right)$.

The CDNCB model can be easily simulated by using either its mixture or ratio representations. In both cases it's necessary to simulate the joint distribution of (N_1, N_2). Such issue can be addressed by first generating the random variable N^+ in Eq. (29) and then $N_1 \mid N^+$ in Eq. (30). As the former distribution is similar to the Poisson one, it can be simulated by using, for example, the inverse-transform method (see e.g. Law and Kelton, 2000).

To fully understand the behavior of the new non-central beta distribution it is instructive to realize how it perturbates the beta one.

PROPOSITION 12 (PERTURBATION REPRESENTATION). *Let $X'' \sim \text{CDNCB}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right)$. Then the density $f_{X''}$ of X'' can be written as:*

$$f_{X''}\left(x; \alpha_1, \alpha_2, \lambda_1, \lambda_2\right) = \text{Beta}(x; \alpha_1, \alpha_2) \cdot \frac{{}_0F_1\left(\alpha_1; \frac{\lambda_1 x}{4}\right) {}_0F_1\left[\alpha_2; \frac{\lambda_2(1-x)}{4}\right]}{{}_0F_1\left(\alpha^+; \frac{\lambda^+}{4}\right)} . \tag{33}$$

PROOF. Observe that Eq. (11) is tantamount to:

$$(a)_i = \frac{\Gamma(a+i)}{\Gamma(a)}, \qquad \forall i \in \mathbb{N} \cup \{0\}, \tag{34}$$

where $\Gamma(\cdot)$ is the gamma function.
In the light of Eq. (34) and in the notation of Eq. (27), one can obtain:

$$B(\alpha_1 + j, \alpha_2 + k) = \frac{\Gamma(\alpha_1)(\alpha_1)_j \, \Gamma(\alpha_2)(\alpha_2)_k}{\Gamma(\alpha^+)(\alpha^+)_{j+k}} = B(\alpha_1, \alpha_2) \frac{(\alpha_1)_j (\alpha_2)_k}{(\alpha^+)_{j+k}}.$$

Hence, by means of easy manipulations, Eq. (27) can be rewritten as follows:

$$f_{X''}(x; \alpha_1, \alpha_2, \lambda_1, \lambda_2) = \text{Beta}(x; \alpha_1, \alpha_2) \cdot \frac{\sum_{j=0}^{+\infty} \frac{\left(\frac{\lambda_1}{4}x\right)^j}{(\alpha_1)_j \, j!} \sum_{k=0}^{+\infty} \frac{\left[\frac{\lambda_2}{4}(1-x)\right]^k}{(\alpha_2)_k \, k!}}{{}_0F_1\left(\alpha^+; \frac{\lambda^+}{4}\right)}.$$

Finally, by virtue of Eq. (25), Eq. (33) is established.   □

Such representation highlights the greater interpretability and tractability of the CDNCB distribution with respect to the DNCB. The perturbing factor of the beta density in the CDNCB distribution, unlike the DNCB one, has a product form involving only the ${}_0F_1$ function (and therefore a single series) in a completely symmetric fashion. Furthermore, ${}_0F_1(a; x)$ has a very simple behavior and is implemented in common statistical packages. For any fixed $a > 0$, it is increasing, with values in $[1, +\infty)$ and convex. For fixed $x \geq 0$, ${}_0F_1(a; x)$ and its first derivative with respect to x are decreasing in a. See plots in Figure 7.
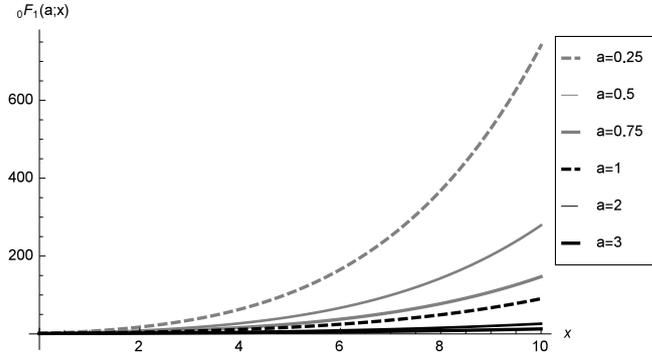
Figure 7 – Plots of ${}_0F_1(a; x)$ for $x \geq 0$ and selected values of $a > 0$.

It follows that ${}_0F_1(\alpha_1; \lambda_1 x/4)$ in Eq. (33) has the effect of giving more weight to the right tail of the beta density through the scale parameter λ_1 and the shape parameter α_1 determining the rate of increase of the function. Perfectly symmetric considerations hold for the other component ${}_0F_1[\alpha_2; \lambda_2(1-x)/4]$ of the perturbing factor.

That said, Figures 8, 9, 10, 11, 12 display the CDNCB density for selected values of the parameters. They show that such density assumes shapes similar to the DNCB ones.

However, in order to appreciate such a variety of shapes, the non-centrality parameters λ_1, λ_2 have to be assigned higher values than the ones used to represent the DNCB density in Figures 2, 3, 4, 5 and 6. This discrepancy can be explained as follows.

The mixture representations given in Propositions 4 and 9 show that differences between the DNCB and the CDNCB can be entirely explained by differences between the distributions of (M_1, M_2) and (N_1, N_2). The latter distributions, by Remark 10, share the same conditional distributions $N_i | N^+$ and $M_i | M^+$. It follows that differences between the DNCB
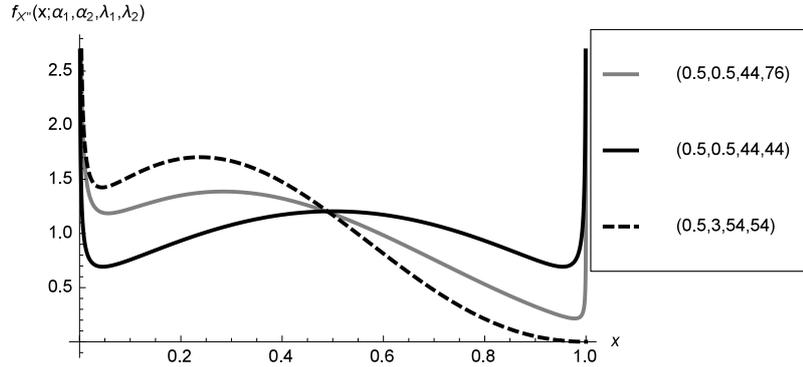
Figure 8 – Plots of the density of $X'' \sim \mathrm{CDNCB}\,(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$ for selected values of $(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$.
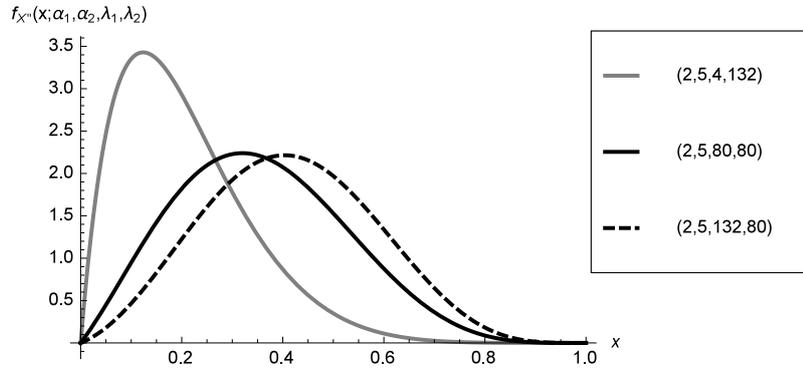
Figure 9 – Plots of the density of $X'' \sim \mathrm{CDNCB}\,(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$ for selected values of $(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$.

and the CDNCB are entirely due to differences between M^+, which has a Poisson$(\lambda^+/2)$ distribution, and N^+, with probability function given by Eq. (29). Some theoretical comparisons between M^+ and N^+ can be made by applying the following well known monotone likelihood ratio result. Suppose $p_1(i)$ and $p_2(i)$ are two probability functions corresponding to the random variables X_1 and X_2 respectively. Then a sufficient condition for X_1 to be stochastically larger than X_2 is that the likelihood ratio $\mathrm{LR}(i) = p_1(i)/p_2(i)$ is increasing (see, for example, Lehmann and Romano (2005), Lemma 3.4.2). A direct application of this result proves that N^+ is stochastically smaller than M^+, for any given $\alpha^+ > 1/2$. Furthermore, N^+ is stochastically decreasing in α^+, for any given λ^+, and stochastically increasing in λ^+, for given α^+. This explains why higher values of λ^+ for the CDNCB are needed in order that its density resembles the DNCB one. It also indicates that this is even more so when α^+ is large.

To appreciate the extent of this phenomenon, a graphical comparison of the probability functions of M^+ and N^+ for selected values of λ^+ and of α^+ is presented in Figures 13, 14 and 15. In particular, one can see that the discrepancy between the distributions of N^+ and M^+ increases when λ^+ gets larger. A rough idea of the link between the λ^+ values of the two distributions can be obtained by comparing their expected values. Figure 16 reports the plots of $\mathbb{E}(M^+) = \lambda^+/2$ and of

$$\mathbb{E}(N^+) = \frac{\frac{\lambda^+}{4}\, {}_0F_1\left(\alpha^+ + 1; \frac{\lambda^+}{4}\right)}{\alpha^+\, {}_0F_1\left(\alpha^+; \frac{\lambda^+}{4}\right)}$$
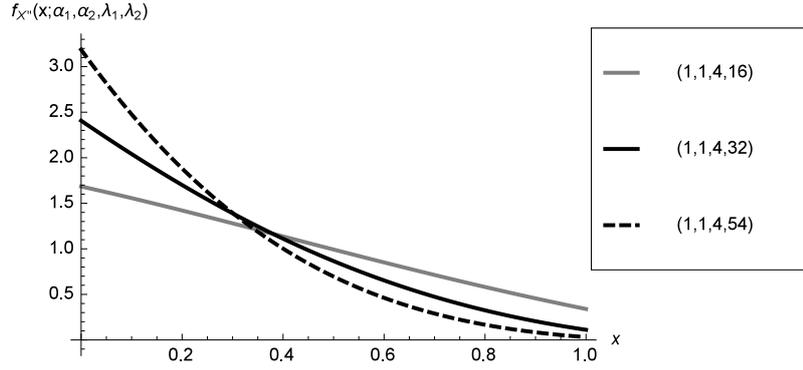
Figure 10 – Plots of the density of $X'' \sim \text{CDNCB}(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$ for $\alpha_1 = \alpha_2 = 1$ and selected values of λ_1, λ_2.
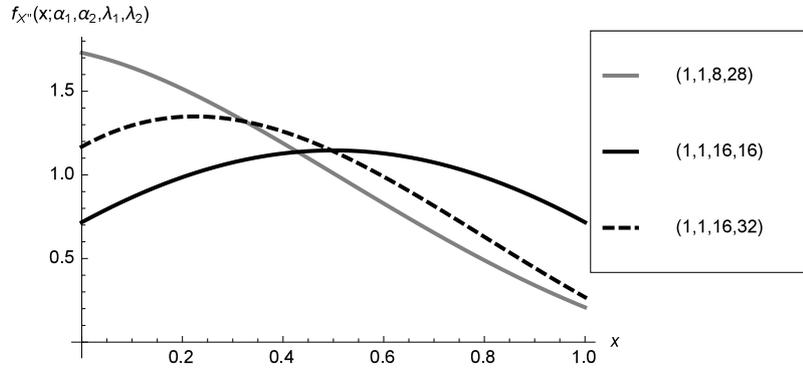
Figure 11 – Plots of the density of $X'' \sim \text{CDNCB}(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$ for $\alpha_1 = \alpha_2 = 1$ and selected values of λ_1, λ_2.

as a function of λ^+. From this plot, for a given value of $\mathbb{E}(M^+)$ one can find the value λ^+ relative to N^+ which produces similar expectation. As expected, this λ^+ value gets larger when α^+ increases. It is also noticeable the strong increment of the difference between the two expectations for growing λ^+.

As a further relevant property, the CDNCB density can take on arbitrary finite and positive limits at 0 and 1, as well as the DNCB (see Figures 10, 11 and 12). More specifically, such limits have the following expressions.

REMARK 13 (LIMITS AT 0 AND 1 OF THE CDNCB DENSITY WHEN $\alpha_1 = \alpha_2 = 1$). *Let* $X'' \sim \text{CDNCB}(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$. *Then, the limits of the density $f_{X''}(\alpha_1, \alpha_2, \lambda_1, \lambda_2)$ of X'' at 0 and 1 when $\alpha_1 = \alpha_2 = 1$ are:*

$$\lim_{x \to 0^+} f_{X''}(x; 1, 1, \lambda_1, \lambda_2) = \frac{{}_0F_1\left(1; \frac{\lambda_2}{4}\right)}{{}_0F_1\left(2; \frac{\lambda^+}{4}\right)}, \tag{35}$$

$$\lim_{x \to 1^-} f_{X''}(x; 1, 1, \lambda_1, \lambda_2) = \frac{{}_0F_1\left(1; \frac{\lambda_1}{4}\right)}{{}_0F_1\left(2; \frac{\lambda^+}{4}\right)}. \tag{36}$$

Finally, notice that the general formulas for the moments of the DNCB distribution in Eq. (14) and (15) hold for the CDNCB distribution as well, provided we simply replace
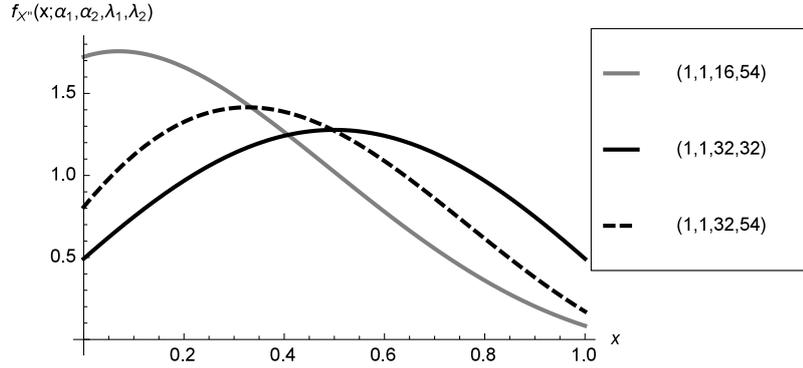
Figure 12 – Plots of the density of $X'' \sim \mathrm{CDNCB}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right)$ for $\alpha_1 = \alpha_2 = 1$ and selected values of λ_1, λ_2.
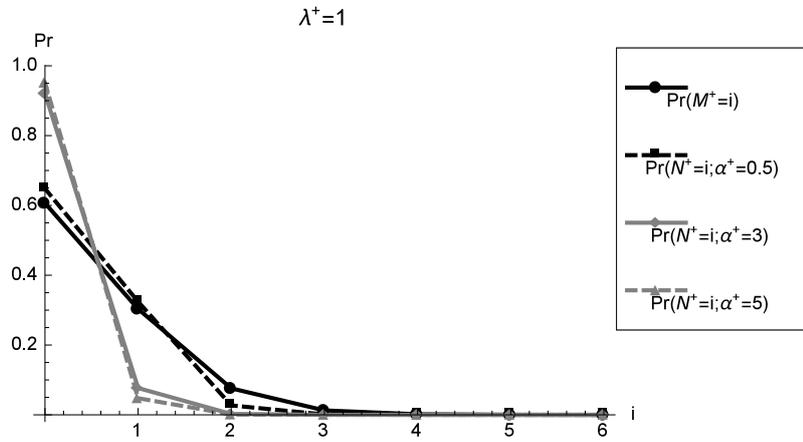
Figure 13 – Probability functions of M^+ and N^+ for $\lambda^+ = 1$ and $\alpha^+ \in \{0.5, 3, 5\}$.

M^+ by N^+. Specifically, the moments of the CDNCB distribution can be written and represented as follows.

PROPOSITION 14 (MOMENTS). *Let X'' have a $\mathrm{CDNCB}\left(\alpha_1, \alpha_2, \lambda_1, \lambda_2\right)$ distribution. Let (N_1, N_2) be jointly distributed as in Eq. (24) with $y = 1$ and $N^+ = N_1 + N_2$. Then, $\forall r \in \mathbb{N}$, the r-th moment of X'' can be written as:*

$$\mathbb{E}\left[(X'')^r \mid N^+\right] = \frac{1}{(\alpha^+ + N^+)_r} \sum_{i=0}^{N^+} (\alpha_1 + i)_r \binom{N^+}{i} \left(\frac{\lambda_1}{\lambda^+}\right)^i \left(1 - \frac{\lambda_1}{\lambda^+}\right)^{N^+ - i} \tag{37}$$

and admits the following representation:

$$\mathbb{E}\left[(X'')^r\right] = \mathbb{E}_{(N_1, N_2)}\left[\frac{(\alpha_1 + N_1)_r}{(\alpha^+ + N^+)_r}\right]. \tag{38}$$

PROOF. The proof follows from arguments similar to those of Proposition 7 by making use of the CDNCB ratio representation in Remark 11. □

By setting $r = 1$ and $r = 2$ in Eq. (37) and after some manipulations, the first two moments of the CDNCB distribution can be written as perturbations of the corresponding

$\lambda^+ = 7$
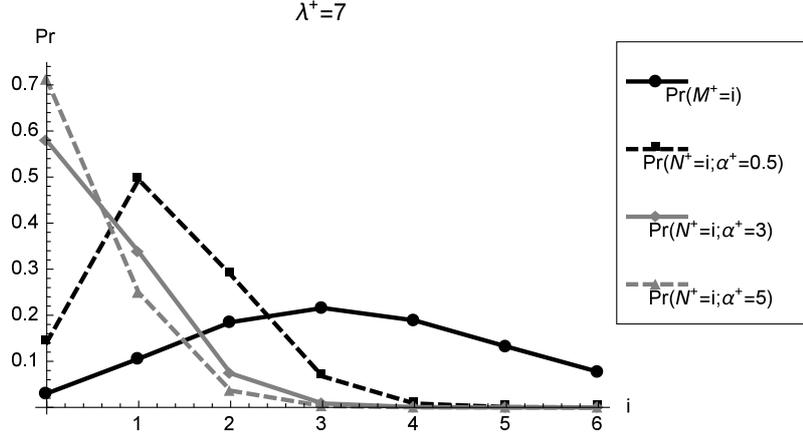
Figure 14 – Probability functions of M^+ and N^+ for $\lambda^+ = 7$ and $\alpha^+ \in \{0.5, 3, 5\}$.

$\lambda^+ = 11$
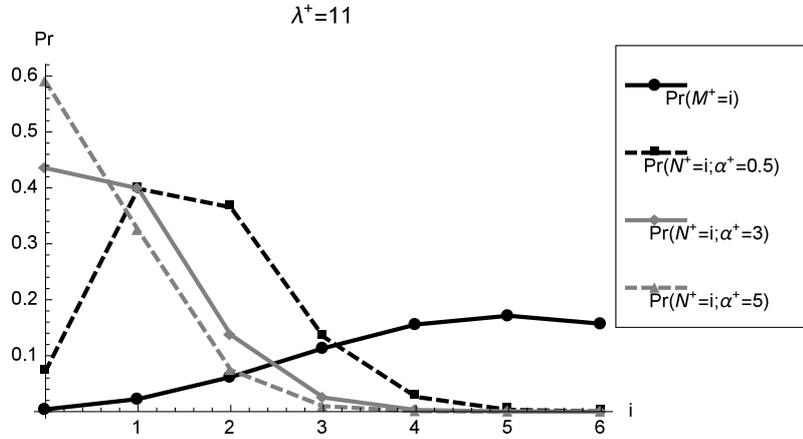
Figure 15 – Probability functions of M^+ and N^+ for $\lambda^+ = 11$ and $\alpha^+ \in \{0.5, 3, 5\}$.

moments of the beta distribution in terms of the easily tractable $_0F_1$ function as follows:

$$\mathbb{E}\left(X''\right) = \frac{\alpha_1}{\alpha^+ \, _0F_1\left(\alpha^+; \frac{\lambda^+}{4}\right)} \left[_0F_1\left(\alpha^+ + 1; \frac{\lambda^+}{4}\right) + \frac{\frac{\lambda_1}{4}}{\alpha_1\left(\alpha^+ + 1\right)} \, _0F_1\left(\alpha^+ + 2; \frac{\lambda^+}{4}\right) \right], \quad (39)$$

$$\mathrm{E}\left[\left(X''\right)^2\right] = \frac{(\alpha_1)_2}{(\alpha^+)_2 \, _0F_1\left(\alpha^+; \frac{\lambda^+}{4}\right)} \left[_0F_1\left(\alpha^+ + 2; \frac{\lambda^+}{4}\right) + \frac{2\frac{\lambda_1}{4}}{\alpha_1\left(\alpha^+ + 2\right)} \cdot \right.$$

$$\left. \cdot \, _0F_1\left(\alpha^+ + 3; \frac{\lambda^+}{4}\right) + \frac{\left(\frac{\lambda_1}{4}\right)^2}{(\alpha_1)_2\left(\alpha^+ + 2\right)_2} \, _0F_1\left(\alpha^+ + 4; \frac{\lambda^+}{4}\right) \right]. \quad (40)$$

5. Conclusions

In this paper a new distribution on the real interval $(0, 1)$ was introduced. It can be viewed as a more easily tractable and interpretable analogue of the doubly non-central beta distribution, the latter corresponding to the distribution of the compositional ratio of two independent non-central chi-squared random variables.
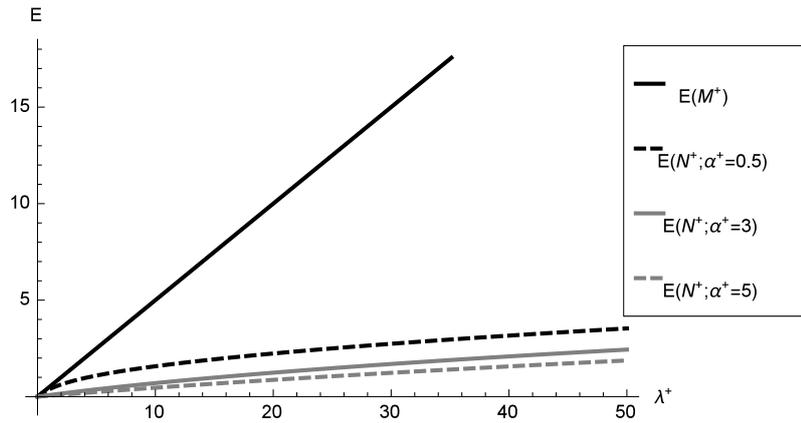
Figure 16 – Plot of the expected values ($\mathbb{E}$) of M^+ and N^+ versus λ^+ for $\alpha^+ \in \{0.5, 3, 5\}$.

The doubly non-central beta density can assume a large variety of shapes far beyond the beta model ones. Furthermore, it can take on arbitrary finite and positive limits at 0 and 1 when the shape parameters are unitary; on the contrary, the beta density cannot display such behavior. However, the perturbing factor of the beta density in the perturbation representation of the doubly non-central beta density is given by the sum of a double power series which has not a simple behavior and is not reducible in an easily tractable analytical form. Moreover, a new general formula for the moments of the standard non-central beta distribution was derived. Such formula reduces the computation of moments from a double series to a single one.

The analysis of the new non-central beta distribution has been based on the same techniques employed in the study of the standard one and has allowed a better understanding of the similarities and differences between the two distributions. Several representations of such distribution have been derived. In particular, the perturbation representation highlights the greater interpretability and tractability of the new distribution with respect to the standard one. Indeed, the perturbing factor of the beta density has a product form involving single series characterized by a simple behavior in a completely symmetric fashion.

Further analysis of such distributions is needed, in particular inferential aspects have to be addressed. We plan to tackle these issues, by starting from parameters estimation, in future work.

References

F. Y. Ettoumi, A. Mefti, A. Adane, M. Y. Bouroubi (2002). *Statistical analysis of solar measurements in Algeria using beta distributions.* Reneweable Energy, 26, no. 1, pp. 47–67.

J. D. Haskett, Y. A. Pachepsky, B. Acock (1995). *Use of the beta distribution for parameterizing variability of soil properties at the regional level for crop yield estimation.* Agricultural Systems, 48, no. 1, pp. 73–86.

N. L. Johnson, S. Kotz, N. Balakrishnan (1994). *Continuous Univariate Distributions Vol. 1, 2nd edition.* John Wiley & Sons, New York.

N. L. JOHNSON, S. KOTZ, N. BALAKRISHNAN (1995). *Continuous Univariate Distributions Vol. 2, 2nd edition.* John Wiley & Sons, New York.

A. M. LAW, W. D. KELTON (2000). *Simulation Modeling and Analysis, 3rd edition.* McGraw-Hill, Boston.

E. L. LEHMANN, J. P. ROMANO (2005). *Testing Statistical Hypotheses, 3rd edition.* Springer, New York.

D. G. MALCOLM, J. H. ROSEBOOM, C. E. CLARK, W. FAZAR (1959). *Application of a technique for research and development program evaluation.* Operations Research, 7, no. 5, pp. 646–669.

J. B. MCDONALD, Y. J. XU (1995). *A generalization of the beta distribution with applications.* Journal of Econometrics, 66, no. 1-2, pp. 133–152.

H. M. SRIVASTAVA, W. KARLSSON (1985). *Multiple Gaussian Hypergeometric Series.* Ellis Horwood, Chichester.

J. A. WILEY, S. J. HERSCHKORN, N. S. PADIAN (1989). *Heterogeneity in the probability of HIV transmission per sexual contact: the case of male-to-female transmission in penile-vaginal intercourse.* Statistics in Medicine, 8, no. 1, pp. 93–102.

SUMMARY

In this paper a new non-central beta distribution is defined. Several properties are derived (including various representations and moments expressions) both for the new and the standard non-central beta distribution, showing a greater tractability and interpretability of the former.

Keywords: beta distribution, non-centrality, moments, mixture representations, hypergeometric functions