Journal of Statistical Planning and Inference 141 (2011) 2987–3004

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On a class of distributions on the simplex

Stefano Favaro, Georgia Hadjicharalambous, Igor Prünster [*,1]

Department of Statistics and Applied Mathematics and Collegio Carlo Alberto, University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy

A R T I C L E I N F O

Article history:
Received 2 September 2010
Received in revised form
23 February 2011
Accepted 7 March 2011
Available online 16 March 2011

Keywords:
Bayes estimator
Dirichlet distribution
Distribution on the simplex
Generalized Bessel function distribution
Infinitely divisible distribution
Inverse-Gaussian distribution
Lévy measure
Positive stable distribution

A B S T R A C T

In the present paper we define and investigate a novel class of distributions on the simplex, termed normalized infinitely divisible distributions, which includes the Dirichlet distribution. Distributional properties and general moment formulae are derived. Particular attention is devoted to special cases of normalized infinitely divisible distributions which lead to explicit expressions. As a by-product new distributions over the unit interval and a generalization of the Bessel function distribution are obtained.

1. Introduction

The Dirichlet distribution on the simplex is motivated by the well-known Dirichlet integral and arises in a variety of probabilistic and statistical contexts. These include Bayesian analysis, order statistics, limit laws, statistical genetics, Pearson systems of curves, nonparametric inference, distribution-free tolerance intervals, reliability theory, probability inequalities, multivariate analysis, characterization problems, stochastic processes and other areas. In order to define the Dirichlet distribution, let us consider a collection of random variables (r.v.) $X_1,\dots,X_n$ which are assumed to be independent and distributed according to a gamma distribution with parameters $(\alpha_1,1),\dots,(\alpha_n,1)$, respectively, being $\alpha_i>0$ the shape parameters for $i=1,\dots,n$. Then, by defining the r.v. $W:=X_1+\cdots+X_n$, the Dirichlet distribution with parameter $(\alpha_1,\dots,\alpha_n)$ is defined as the distribution of the random vector $(P_1,\dots,P_n):=(X_1/W,\dots,X_n/W)$ on the $(n-1)$-dimensional simplex $\Delta^{(n-1)}:=\{(p_1,\dots,p_{n-1}):p_i>0, i=1,\dots,n-1,\sum_{i=1}^{n-1}p_i\le 1\}$. In other terms, the Dirichlet distribution is definable via a "normalization" operation as the joint distribution of a set of independent r.v. distributed according to a gamma distribution divided by their sum.

The Dirichlet distribution represents one of the most widely studied distributions over the last century. The reason for its popularity can be mainly traced back to its mathematical tractability; however, this also implies some constraints in terms of flexibility, which constitute a limitation in some of its numerous applications. Therefore, various approaches for constructing distributions on the simplex have been undertaken with the aim of overcoming some of the drawbacks of the

* Corresponding author.
E-mail addresses: stefano.favaro@unito.it (S. Favaro), georgia.hadjicharalambous@carloalberto.org (G. Hadjicharalambous), igor@econ.unito.it (I. Prünster).

[1] Also affiliated to ICER.

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2011.03.015

Dirichlet distribution. For instance, in the context of the analysis of compositional data, which are subject to nonnegativity and constant-sum constraints, and in Bayesian analysis we mention the scaled Dirichlet distribution (Aitchison, 1982), the generalized logistic-normal distribution (Aitchison, 1985), the Liouville distribution and its various generalizations (see Rayens and Srinivasan, 1994 and references therein), the generalized Dirichlet distribution (Connnor and Mosimann, 1969; James, 1972, 1975), the grouped Dirichlet distribution (Ng et al., 2008), the nested Dirichlet distribution (Ng et al., 2009; Tian et al., 2010) and the Dirichlet tree distribution usually also known as the hyper-Dirichlet distribution (Dennis, 1991, 1992). In the present paper, we propose an alternative approach to the problem of defining distributions on the simplex by resorting to the above mentioned "normalization" approach. Such an approach has been fruitfully introduced in Regazzini et al. (2003) for the definition of random probability measures for Bayesian nonparametric inference (see, e.g., Nieto-Barajas et al., 2004; James et al., 2006; Lijoi et al., 2005; Griffin, 2007; James et al., 2009; Griffin and Walker, 2011, for following developments).

Let us consider a r.v. X distributed according to a gamma distribution with parameter $(\alpha,1)$; it is well known that the gamma distribution is infinitely divisible (ID), i.e for any $n \in \mathbb{N}$ there exists a collection of independent and identically distributed r.v. $Y_1,\ldots,Y_n$ such that $X \overset{d}{=} Y_1+\cdots+Y_n$ or, alternatively, the r.v. X can be "separated" into the sum of an arbitrary number of independent and identically distributed r.v. In particular, according to the Lévy–Khintchine representation theorem for ID distributions, the Laplace transform of the gamma distribution with parameter $(\alpha,1)$ coincides with

$$\mathbb{E}[e^{-uX}] = \exp\left\{-\alpha\int_0^\infty (1-e^{-ux})\frac{e^{-x}}{x}\,dx\right\} = (1+u)^{-\alpha}, \quad u \ge 0,$$

where $\alpha x^{-1}e^{-x}\,dx$ is the so-called Lévy measure and it completely characterizes the gamma distribution. See the comprehensive and stimulating monograph (Steutel and Van Harn, 2004) and references therein. At this stage we are in a position to describe the main purpose of the present paper, that is, the generalization of the "normalization" approach applied for defining the Dirichlet distribution to cases in which the gamma r.v. are replaced by arbitrary positive ID r.v. In other terms, we consider a collection of positive r.v. $X_1,\ldots,X_n$ which are assumed to be independent and distributed according to, not necessarily coinciding, ID distributions. The Laplace transform of each X_i can be represented canonically via the Lévy–Khintchine representation as

$$\mathbb{E}[e^{-uX_i}] = \exp\{-\Psi_i(u)\} = \exp\left\{-\int_0^\infty (1-e^{-ux})\nu_i(dx)\right\} \quad u \ge 0,$$

where Ψ_i is typically referred to as Laplace exponent and the Lévy measure ν_i is any nonnegative Borel measure satisfying the condition:

$$\int_0^\infty \min(1,x)\nu_i(dx) < \infty,$$

which completely characterizes the distribution of the r.v. X_i, for each $i=1,\ldots,n$. Thus, by defining the r.v. $W := X_1+\cdots+X_n$, the "normalization" approach yields a wide class of distributions over $\Delta^{(n-1)}$ for the random vector $(P_1,\ldots,P_n) := (X_1/W,\ldots,X_n/W)$. In particular, each of these distributions is completely characterized by the corresponding collection of Lévy measures $\{\nu_1,\ldots,\nu_n\}$. We term this class of distributions the normalized ID (NID) distributions.

The class of NID distributions represents a natural extension of the Dirichlet distribution, which can be recovered as special case of NID distributions by assuming the collection of Lévy measures to be $\nu_i(dx) = \alpha_i x^{-1}e^{-x}dx$ for $i=1,\ldots,n$. Further important examples of NID distributions, which have found important applications in Bayesian theory, are the normalized inverse-Gaussian distribution (Lijoi et al., 2005), the normalized 1/2-stable distributions (Carlton, 2002) and the normalized tempered stable distribution (Kolossiatis et al., 2011; Lijoi et al., 2007). Our main aim is to investigate the class of NID distributions in terms of their distributional properties and to provide general moment formulae. Specifically, the key tools exploited for deriving the results are given by Gurland's inversion formula for characteristic functions (Gurland, 1948) and the Faà di Bruno formula which allows to obtain expressions in terms of partial Bell polynomials (Comtet, 1974). As an application of the general results, some interesting examples of NID distributions are considered by combining three well-known ID distributions: the gamma distribution, the positive 1/2-stable distribution and the inverse-Gaussian distribution. These lead to relatively simple and tractable expressions, which are readily exploitable in applications. Following these guidelines, in Section 2 we state the definition of NID distribution, investigate properties of the corresponding distribution and provide general moment formulae. In Section 3 interesting special cases of NID distributions are analyzed in detail. As a by-product new distributions over the unit interval and a generalization of the Bessel function distribution are obtained.

2. Normalized infinitely divisible distributions

Relying on the "normalization" concept briefly described in Section 1, we start by stating the formal definition of NID distribution on the simplex.