International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Composition operator for densities of continuous variables

Vladislav Bína

Faculty of Management, University of Economics in Prague, Jarošovská 1117/II, 37701 Jindřichův Hradec, Czechia

A R T I C L E I N F O

Article history:
Received 31 March 2016
Received in revised form 30 December 2016
Accepted 23 January 2017
Available online xxxx

Keywords:
Operator of composition
Continuous random variable
Dirac delta function
Conditioning
Exponential family

A B S T R A C T

The paper presents an operator of composition for densities of continuous random variables and analyzes its properties, particularly the assertions useful for marginalization, lemmata concerning the conditional independence, theorems describing entropy of the composition and assertions for reordering of composed densities in the model. The generalized function of Dirac delta is proposed as a degenerated distribution allowing the expression of operation of conditioning using the composition of Dirac delta with a continuous density or with a compositional model. A special case of composition within the exponential family of distributions is considered.

1. Introduction

The paper presents an approach to the composition of multidimensional densities of continuous random variables. In a way, it is an alternative to the copula approach based on Sklar's famous theorem (see [13]) elaborated later into the form of a vine approach, e.g., [1]. For a basic discussion of the properties, advantages and disadvantages of the compositional model approach and copula based approaches, see [4] and [3].

The paper brings a definition of an operator of composition (already introduced in [3]) and analyzes its basic properties in a manner analogous to [6]. The basic aim is to show that the well-known generalized function of Dirac delta may serve as a (degenerated) distribution providing the possibility of defining the operation of conditioning (for a discrete case see [5]) when computing with distributions composed from continuous densities.

2. Preliminaries and basic notions

Throughout the presented paper we consider a finite index set $N = \{1, \ldots, n\}$ together with a set of random variables $\{X_i\}_{i \in N}$ with values, or vectors of values, denoted by the corresponding lowercase letters. The domain of variables will be denoted by the corresponding bold uppercase letter $\mathbf{X}_i$. Variables with a finite or countable set of possible *states* are called *discrete*; other variables are called *continuous*. Both discrete (with real domain) and continuous variables can be described using (multi-dimensional) *cumulative distribution function* (CDF). However, to define an operator of composition, in a discrete case we use *probability mass function* which, in the theory of compositional models, is usually denoted by small Greek letters $(\kappa, \lambda, \mu, \nu, \pi, \ldots)$. The theory of discrete compositional models and all related notions can be found in [6].

In order to define the operator of composition for continuous variables, we require the cumulative distribution function F to have a corresponding density f that coincides with the derivative of F whenever it exists (but almost everywhere in the sense of Lebesgue measure). For the sake of simplicity, in all opposite cases we require that the density f vanishes

E-mail address: bina@fm.vse.cz.

http://dx.doi.org/10.1016/j.ijar.2017.01.008
0888-613X/

to zero which makes the following considerations easier (the formulae hold everywhere instead of almost everywhere). Similarly, in the case of multivariate (joint) probability density function we require CDF to be sufficiently differentiable and corresponding partial derivatives to exist.

As previously mentioned, the probability density functions (of continuous random variables) will be denoted by lowercase letters of the Latin alphabet ($f, g, h, \dots$), e.g., the abbreviated notation $f(x_K)$ denotes a multidimensional density of variables having indices from set $K \subseteq N$.

For a probability density function $f(x_K)$ and any set of variable indices $L \subset K$, a marginal probability density $f(x_L)$ of $f(x_K)$ can be computed for each x_L as follows

$$f(x_L) = \int_{\mathbf{X}_{K \setminus L}} f(x_K) dx_{K \setminus L}$$

where obviously the integration runs over the domains of all variables in $K \setminus L$. We will also employ an equivalent way to denote the marginal $f(x_L)$, namely $f^{\downarrow\{L\}}$ which was introduced by Prakash Shenoy and Glenn Shafer (see, e.g., [12]).

Let us recall a notion of *support* (of a function), i.e. the set of points where the function f has non-zero values

$$S_f = \{x \mid f(x) \neq 0\}.$$

Having probability density $f(x_K)$ and two disjoint subsets $L, M \subseteq K$ we define the *conditional probability density* of X_L given a value $x_M = \mathbf{x_M}$ for every $x_{L \cup M}$ as

$$f(x_L \mid x_M = \mathbf{x_M}) f(x_M = \mathbf{x_M}) = f(x_L, x_M = \mathbf{x_M}).$$

Let us note that for $f(x_M = \mathbf{x_M}) = 0$ the definition is ambiguous, but we do not need to exclude such cases.

3. Conditional independence

Continuous random variables x_K having a joint density $f(x_K)$ are all *independent* from each other if and only if for each x_K the joint density function can be expressed as a product $f(x_K) = f_1(x_1) \cdots f_k(x_k)$ ($K = \{1, \dots, k\}$). An important notion is so-called conditional independence.

Definition 1 *(Conditional independence).* Let us have probability density function $f(x_K)$ and three disjoint subsets $L_1, L_2, L_3 \subseteq K$ where L_1 and L_2 are non-empty. We say that groups of variables X_{L_1} and X_{L_2} are *conditionally independent* given group X_{L_3} if for each $x_{L_1 \cup L_2 \cup L_3}$ it holds

$$f(x_{L_1 \cup L_2 \cup L_3}) f(x_{L_3}) = f(x_{L_1 \cup L_3}) f(x_{L_2 \cup L_3}). \tag{1}$$

We write $X_{L_1} \perp\!\!\!\perp X_{L_2} \mid X_{L_3}[f]$.

Let us clarify what we will call the *principle of marginal zeroing*: if we have probability density $f(x_K)$ and $M \subseteq L \subseteq K$ then for each x_K it holds that $f(x_M) = 0$ implies $f(x_L) = 0$ and this implies $f(x_K) = 0$.

Now it is apparent that in Formula (1) the equality holds for x_K such that $f(x_{L_3}) = 0$. For each x_K such that $f(x_{L_3}) > 0$ we can, according to the above definition, introduce conditional probability density and divide both sides of Formula (1) by $f(x_{L_3})$ obtaining

$$f(x_{L_1 \cup L_2 \cup L_3}) = f(x_{L_1 \cup L_3}) f(x_{L_2} | x_{L_3}). \tag{2}$$

4. Operator of composition

Firstly, we can refer to an important binary relation of absolute continuity or dominance that can be found in any basic textbook on functional analysis, e.g., in [10]. Namely:

Definition 2. Let us have two probability density functions $f(x_K)$ and $g(x_K)$ with the same set of variables X_K. Then f is said to be *absolutely continuous with respect to* g, or *dominated by* g (denoted by $f \ll g$) if for each $x_K \in X_K$ it holds

$$(g(x_K) = 0 \Rightarrow f(x_K) = 0).$$

The definition of the composition operator and its properties are in a way analogous to the situation in discrete probability distributions, see [6].

Definition 3. Consider two sets of continuous variables X_L and X_M, a probability density $f(x_L)$, and a probability density $g(x_M)$ such that $f(x_{L \cap M}) \ll g(x_{L \cap M})$. The *right composition* is given by

$$f(x_L) \triangleright g(x_M) = \frac{f(x_L) g(x_M)}{g(x_{L \cap M})}.$$