Journal of Mathematical Psychology 74 (2016) 11–33

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Context–content systems of random variables: The Contextuality-by-Default theory

Ehtibar N. Dzhafarov [a,*], Janne V. Kujala [b]

[a] Purdue University, USA
[b] University of Jyväskylä, Finland

HIGHLIGHTS

- Contextuality is about random variables classified by content and by context.
- Same-context variables possess joint distributions, with observed probabilities.
- Same-content variables can be joined to be equal with maximal probabilities.
- Are these probabilities compatible with the observed probabilities?
- If they are not, the system is contextual.

ARTICLE INFO

Article history:
Available online 3 June 2016

Keywords:
Contextuality
Couplings
Connectedness
Random variables

ABSTRACT

This paper provides a systematic yet accessible presentation of the Contextuality-by-Default theory. The consideration is confined to finite systems of categorical random variables, which allows us to focus on the basics of the theory without using full-scale measure-theoretic language. Contextuality-by-Default is a theory of random variables identified by their contents and their contexts, so that two variables have a joint distribution if and only if they share a context. Intuitively, the content of a random variable is the entity the random variable measures or responds to, while the context is formed by the conditions under which these measurements or responses are obtained. A system of random variables consists of stochastically unrelated "bunches," each of which is a set of jointly distributed random variables sharing a context. The variables that have the same content in different contexts form "connections" between the bunches. A probabilistic coupling of this system is a set of random variables obtained by imposing a joint distribution on the stochastically unrelated bunches. A system is considered noncontextual or contextual according to whether it can or cannot be coupled so that the joint distributions imposed on its connections possess a certain property (in the present version of the theory, "maximality"). We present a criterion of contextuality for a special class of systems of random variables, called cyclic systems. We also introduce a general measure of contextuality that makes use of (quasi-)couplings whose distributions may involve negative numbers or numbers greater than 1 in place of probabilities.

1. Introduction

Contextuality-by-Default (CbD) is an approach to probability theory, specifically, to the theory of random variables. CbD is not a model of empirical phenomena, and it cannot be corroborated or falsified by empirical data. However, it provides a sophisticated conceptual framework in which one can describe empirical data and formulate models that involve random variables.

In Kolmogorovian Probability Theory (KPT) random variables are understood as measurable functions mapping from one (domain) probability space into another (codomain) probability space. CbD can be viewed as a theory within the framework of KPT if the latter is understood as allowing for multiple domain probability spaces, freely introducible and unrelated to each other. However, CbD can also be (in fact, is better) formulated with no reference to domain probability spaces, with random variables understood as entities identified by their probability distributions and their unique labels within what can be called sets of random variables "in existence" or "in play".

Although one cannot deal with probability distributions without the full-fledged measure-theoretic language, we avoid

* Corresponding author.
E-mail addresses: ehtibar@purdue.edu (E.N. Dzhafarov), jvk@iki.fi (J.V. Kujala).

http://dx.doi.org/10.1016/j.jmp.2016.04.010
0022-2496/

technicalities some readers could find inhibitive by focusing in this paper on *finite* systems of *categorical random variables* (those with finite numbers of possible values). Virtually all of the content of this paper, however, is generalizable *mutatis mutandis* to arbitrary systems of arbitrary random entities.

1.1. A convention

In the following we introduce sets of random variables classified in two ways, by their *contexts* and by their *contents*, and we continue to speak of contexts and contents throughout the paper. The two terms combine nicely, but they are also easily confused in reading. For this reason, in this paper we do violence to English grammar and write "conteXt" and "conteNt" when we use these words as special terms.

1.2. Two conteNts in two conteXts

We begin with a simple example. A person randomly chosen from some population is asked two questions, q and q'. Say, $q =$ "Do you like bees?" and $q' =$ "Do you like to smell flowers?". The answer to the first question (Yes or No) is a random variable whose *identity* (that which allows one to uniquely identify it within the class of all random variables being considered) clearly includes q, so it can be denoted R_q. We will refer to the question q as the *conteNt* of the random variable R_q. The second random variable then can be denoted $R_{q'}$, and its conteNt is q'. The set of all random variables being considered here consists of R_q and $R_{q'}$, and we do not confuse them because they have distinct conteNts: we know which of the two responses answers which question.

The two random variables have a joint distribution that can be presented, because they are binary, by values of the three probabilities

$$\Pr\left[R_q = \text{Yes}\right], \qquad \Pr\left[R_{q'} = \text{Yes}\right],$$
$$\Pr\left[R_q = \text{Yes and } R_{q'} = \text{Yes}\right].$$

The joint distribution exists because the two responses, R_q and $R_{q'}$, occur together in a well-defined empirical sense: the empirical sense of "togetherness" of the responses here is "to be given by one and the same person". In other situations the empirical meaning can be different, e.g., "to be recorded in the same trial".

Our example is too simple for our purposes. Let us assume therefore that the two questions q, q' are asked under varying controlled conditions, e.g., one randomly chosen person can be asked these questions after having watched a movie about the killer bees spreading northwards (let us call this condition c), another after watching a movie about deciphering the waggle dances of the honey bees (c'). Most people would consider q as one and the same question whether posed under the condition c or the condition c'; and the same applies to the question q'. In other words, the conteNts q and q' of the two respective random variables would normally be considered unchanged by the conditions c and c'.

However, the random variables themselves (the responses) are clearly affected by these conditions. In particular, nothing guarantees that the joint distribution of $\left(R_q, R_{q'}\right)$ will be the same under the two conditions. It is necessary therefore to include c and c' in the description of the random variables representing the responses. We will call c and c' *conteXts* of (or for) the corresponding random variables and present them as $R_q^c, R_{q'}^c, R_q^{c'}, R_{q'}^{c'}$. There are now four random variables in play, and we do not confuse them because each of them is uniquely identified by its conteNt and its conteXt.

1.3. Jointly distributed versus stochastically unrelated random variables

In each of the two conteXts, the two random variables are jointly distributed, i.e., we have well-defined probabilities

$$\left.\begin{array}{r} \Pr\left[R_q^c = \text{Yes}\right], \\ \Pr\left[R_{q'}^c = \text{Yes}\right], \\ \Pr\left[R_q^c = \text{Yes and } R_{q'}^c = \text{Yes}\right] \end{array}\right\} \quad \text{in conteXt } c,$$

and

$$\left.\begin{array}{r} \Pr\left[R_q^{c'} = \text{Yes}\right], \\ \Pr\left[R_{q'}^{c'} = \text{Yes}\right], \\ \Pr\left[R_q^{c'} = \text{Yes and } R_{q'}^{c'} = \text{Yes}\right] \end{array}\right\} \quad \text{in conteXt } c'.$$

No joint probabilities, however, are defined between the random variables picked from different conteXts. We cannot determine such probabilities as

$$\Pr\left[R_q^c = \text{Yes and } R_{q'}^{c'} = \text{Yes}\right],$$
$$\Pr\left[R_q^c = \text{Yes and } R_q^{c'} = \text{Yes}\right],$$
$$\Pr\left[R_q^c = \text{Yes and } R_q^{c'} = \text{Yes and } R_{q'}^{c'} = \text{Yes}\right],$$

etc.

We express this important fact by saying that any two variables recorded in different conteXts are *stochastically unrelated*. The reason for stochastic unrelatedness is simple: no random variable in conteXt c can co-occur with any random variable in conteXt c' in the same empirical sense in which two responses co-occur within either of these conteXts, because c and c' are mutually exclusive conditions. The empirical sense of co-occurrence in our example is "to be given by the same person", and we have assumed that a randomly chosen person is either shown one movie or another. If some respondents were allowed to watch both movies before responding, we would have to redefine the classification of our random variables by introducing a third conteXt, $c'' = \left(c, c'\right)$. We would then have three pairwise mutually exclusive conteXts, c, c', c'', and six random variables, $R_q^c, R_{q'}^c, R_q^{c'}, R_{q'}^{c'}, R_q^{c''}, R_{q'}^{c''}$, such that, e.g., $R_q^{c''}$ is jointly distributed with $R_{q'}^{c''}$ but not with R_q^c.

In case one is tempted to consider joint probabilities involving R_q^c and $R_q^{c'}$ simply equal to zero (because these two responses never co-occur), this thought should be dismissed. Indeed, then all four joint probabilities,

$$\Pr\left[R_q^c = \text{Yes and } R_{q'}^{c'} = \text{Yes}\right],$$
$$\Pr\left[R_q^c = \text{Yes and } R_{q'}^{c'} = \text{No}\right],$$
$$\Pr\left[R_q^c = \text{No and } R_{q'}^{c'} = \text{Yes}\right],$$
$$\Pr\left[R_q^c = \text{No and } R_{q'}^{c'} = \text{No}\right],$$

would have to be equal to zero, which is not possible as they should sum to 1. These probabilities are not zero, they are *undefined*.

1.4. Bunches and connections in conteXt–conteNt matrices

The picture of the system consisting of our four random variables is now complete. Let us call this system $\mathcal{A}$. It is an example of a *conteXt–conteNt (c–c) system* of random variables, and it can be schematically presented in the form of the *conteXt–conteNt (c–c)*