Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 250 (2014) 25–51

FUZZY sets and systems

www.elsevier.com/locate/fss

Possibilistic and probabilistic likelihood functions and their extensions: Common features and specific characteristics

Giulianella Coletti [a], Davide Petturiti [b], Barbara Vantaggi [b,*]

[a] Dip. Matematica e Informatica, Università di Perugia, 06100 Perugia, Italy
[b] Dip. S.B.A.I., "La Sapienza" Università di Roma, 00185 Roma, Italy

Received 22 December 2012; received in revised form 17 September 2013; accepted 19 September 2013

Available online 25 September 2013

Abstract

We deal with conditional probability in the sense of de Finetti and with T-conditional possibility (with T a triangular norm). We prove that Dubois and Prade conditional possibility is a particular min-conditional possibility and then we compare the two notions of conditioning by an inferential point of view. Moreover, we study T-conditional possibilities as functions of the conditioning event, putting in evidence analogies and differences with conditional probabilities. This allows to characterize likelihood functions (and their aggregations) consistent either with a T-conditional possibility or a conditional probability. This analysis highlights many syntactical coincidences. Nevertheless the main difference is a weak form of monotonicity, which arises only in the possibilistic case.

Keywords: Likelihood function; Coherence; Conditional probability; *DP*-conditional possibility; T-conditional possibility

1. Introduction

In classical Bayesian analysis the inference process is based on a joint probability distribution $f(E, \omega)$, in order to make inferences about an unknown state of nature ω belonging to a set $\mathscr{L}$ of states upon observing the evidence E. The joint probability distribution $f(E, \omega)$ is built from a likelihood function $f(E|\omega)$ and a suitable prior distribution $\varphi(\omega)$. In this framework, a requirement is that both prior probability $\varphi(\cdot)$ and likelihood function $f(E|\cdot)$ are "precise" and completely assessed on the set $\mathscr{L}$ of states of nature, $\mathscr{L} = \{\omega\}$.

This constraint can be too strong to use this model on real problems: in fact sometimes the available information is related to different sets $\mathscr{L}$ and $\mathscr{L}' = \{\psi\}$ (i.e. one has $f(E|\omega)$ and $\varphi(\psi)$) or comes from multiple expert assessors or even from a previous inferential procedure.

For example, often the random vector X related to the states of nature ω is not entirely of interest, so a parameterization $X = (\Theta, \Gamma)$ (i.e. $\omega = (\theta, \gamma)$) is selected for simplifying the analysis on the vector of interest Θ. Then, for example, starting from $f(E|\theta, \gamma)$ and $\varphi(\theta)$ the nuisance parameter γ needs to be "eliminated" from $f(E|\theta, \gamma)$.

* Corresponding author.
E-mail addresses: coletti@dmi.unipg.it (G. Coletti), davide.petturiti@sbai.uniroma1.it (D. Petturiti), barbara.vantaggi@sbai.uniroma1.it (B. Vantaggi).

0165-0114/$ – see front matter
http://dx.doi.org/10.1016/j.fss.2013.09.010

In these situations we could be forced to manage (Bayesian-like) inferential procedures where prior information $\varphi(\cdot)$ and likelihood $f(E|\cdot)$ could be expressed by different uncertainty frameworks. This topic has been already faced, e.g., in [22,25,42,46], in particular the likelihood function has been studied in possibility framework, see [24,25].

In this paper we deal with a theoretical framework for handling inference processes in which the available information is expressed in probabilistic and possibilistic terms (for analogous issues see [4,16,17,23,25,37,39,45,48,49]).

This need can arise from the following problems:

- The likelihood f and the prior φ are given on the same partition $\mathscr{L} = \{(\theta, \gamma)\}$. If γ is a nuisance parameter, then the aim is to "eliminate the nuisance parameter γ", that means to find an aggregated likelihood function on the coarser partition $\mathscr{L}' = \{(\theta)\}$ (see, e.g., [5]).
- The available information consists of a probabilistic likelihood $f(E|H_i)$ with H_i ranging in a partition $\mathscr{L}$ and a possibility distribution φ on $\mathscr{L}$, which can be obtained as upper envelope of a class of probabilities (see, e.g., [3,13,16,21,28,38]) or as translation of a human perception or a fuzzy information (see for instance [12,45,47]): we are interested in evaluating the uncertainty on the events of the form $H_i|E$.
- Given a likelihood $f(E|\cdot)$ assessed on a partition $\mathscr{L}_1$ and a probability or possibility distribution φ on a coarser partition $\mathscr{L}_2$, in order to evaluate the uncertainty on the events $K_h|E$, with K_h ranging in $\mathscr{L}_2$, we need to compute the aggregated likelihood $g(E|K_h)$ by considering all (and only) the given information.
- Given a possibility distribution on a partition $\mathscr{L}_1$ and a class of probabilistic likelihood functions $f_j(E_j|H_i)$ with $H_i \in \mathscr{L}_1$ and E_j ranging in another partition $\mathscr{L}_2$, we are interested in computing a generalized information measure [33] of the partition $\mathscr{L}$ obtained as conjunction of the two partitions $\mathscr{L}_1, \mathscr{L}_2$. The possibility distribution on $\mathscr{L}_1$ can be derived, for instance, through the maximum specificity principle, as a transformation of a univariate probability distribution (see, e.g., [30,35,36]), that gives rise to the more informative possibility dominating the probability measure. This is relevant not just for mathematical ease but also to recognize systematic and random errors in measurements.

The question is: how can these situations be managed by maintaining coherence inside either probability or possibility theory?

For this aim we focus on the *likelihood function*, regarded as a function or assessment on a class of conditional events $\{E|H_i\}$, with E an arbitrary event and $\{H_i\}$ a finite partition $\mathscr{L}$ of the sure event Ω. Then, we need to study the *aggregated likelihood function*, which is a function defined on the conditional events $\{E|K\}$, with K belonging to the set $\mathscr{H}$ of finite disjunctions of the H_i's. Concerning this point, in [24] it is shown that the aggregated likelihood obtained according to the maximum likelihood principle is maxitive as possibility measures.

Now first of all we are interested in studying when a probabilistic likelihood has the same properties of a possibilistic one from a syntactical point of view. Then, by referring to coherent T-conditional possibility assessments (introduced in [19,29]) and their particular extensions, we study in which way we can extend both a possibilistic and a probabilistic likelihood function to the events of a partition less fine than that in which it is defined.

Since conditioning in possibility theory is deeply debated (see, e.g., [20,26], see also [23] for a state of the art of possibility theory), we compare Dubois and Prade's approach [25,26] (in the following called *DP*-conditioning) with the approach of T-conditional possibility. In particular, we prove that *DP*-conditional possibilities are particular min-conditional possibilities and we introduce the notion of coherence also for *DP*-conditioning. Then we show that coherent *DP*-conditional possibility assessments, as well as coherent T-conditional possibility assessments (with T a continuous t-norm), are extendable to any new conditional event. However, the set of possible extensions in the case of min-conditional possibility is a closed interval, while in the case of *DP*-conditioning the set is not necessarily convex (it could be the union of a possibly degenerate interval and the point 1), see Example 2.

Then we consider T-conditional possibilities (where T is again a continuous triangular norm) and we show the role of Goodman–Nguyen's relation [32], which generalizes the implication between events to conditional events: T-conditional possibility, as well as conditional probability, is monotone with respect to Goodman–Nguyen's implication.

Our aim is to give a thorough comparison of probabilistic and possibilistic (aggregated) likelihood functions. Concerning the likelihood function on a partition $\mathscr{L}$, we find that, from a syntactical point of view, any possibilistic likelihood is also a probabilistic likelihood, and vice versa. Similar results are obtained by checking coherence of a likelihood together with either a probabilistic or a possibilistic prior. This property continues to hold when we consider