Math Geosci (2012) 44:545–581
DOI 10.1007/s11004-012-9396-3

Probability Aggregation Methods in Geoscience

D. Allard · A. Comunian · P. Renard

Received: 3 November 2011 / Accepted: 2 April 2012 / Published online: 26 April 2012

Abstract The need for combining different sources of information in a probabilistic framework is a frequent task in earth sciences. This is a need that can be seen when modeling a reservoir using direct geological observations, geophysics, remote sensing, training images, and more. The probability of occurrence of a certain lithofacies at a certain location for example can easily be computed conditionally on the values observed at each source of information. The problem of aggregating these different conditional probability distributions into a single conditional distribution arises as an approximation to the inaccessible genuine conditional probability given all information. This paper makes a formal review of most aggregation methods proposed so far in the literature with a particular focus on their mathematical properties. Exact relationships relating the different methods is emphasized. The case of events with more than two possible outcomes, never explicitly studied in the literature, is treated in detail. It is shown that in this case, equivalence between different aggregation

The order of the authors is alphabetical.

D. Allard (✉)
UR546 Biostatistique et Processus Spatiaux (BioSP), INRA, Site Agroparc 84914, Avignon, France
e-mail: allard@avignon.inra.fr

A. Comunian · P. Renard
Centre of Hydrogeology and Geothermics, CHYN, Université de Neuchâtel, 11 Rue Emile Argand, 2000 Neuchâtel, Switzerland

A. Comunian
e-mail: alessandro.comunian@gmail.com

P. Renard
e-mail: philippe.renard@unine.ch

Present address:
A. Comunian
National Centre for Groundwater Research and Training, University of New South Wales, Sydney, Australia

formulas is lost. The concepts of calibration, sharpness, and reliability, well known in the weather forecasting community for assessing the goodness-of-fit of the aggregation formulas, and a maximum likelihood estimation of the aggregation parameters are introduced. We then prove that parameters of calibrated log-linear pooling formulas are a solution of the maximum likelihood estimation equations. These results are illustrated on simulations from two common stochastic models for earth science: the truncated Gaussian model and the Boolean. It is found that the log-linear pooling provides the best prediction while the linear pooling provides the worst.

Keywords Data integration · Conditional probability pooling · Calibration · Sharpness · Log-linear pooling

1 Introduction

The problem of aggregating probability assessments coming from different sources of information is probably as old as statistics and stochastic modeling. In geosciences, Tarantola and Valette (1982) and Tarantola (2005) developed the concept of conjunction and disjunction of probabilities in the context of inverse problems. Benediktsson and Swain (1992) adopted consensus theoretic classification methods to aggregate geographical data like satellite images coming from different sources. Journel (2002) proposed the Tau model in a very broad perspective. This model was subsequently used by Strebelle et al. (2003) to map lithofacies using seismic information and multiple-point statistics, and by Comunian et al. (2011) to combine the probability assessments derived from different two-dimensional geostatistical models to simulate three-dimensional geological structures. Okabe and Blunt (2004, 2007) used a linear probability combination method to simulate three-dimensional porous medium from two-dimensional multiple-point statistics extracted from microscope images of a rock sample. Mariethoz et al. (2009) used the probability conjunction method to develop a collocated co-simulation algorithm allowing the modeling of any complex probability relationship between the primary and secondary variable. Ranjan and Gneiting (2010) combined weather forecasts coming from different models with the Beta-transformed Linear opinion Pool (BLP). In the context of risk analysis, Genest and Zidek (1986) and Clemen and Winkler (1999, 2007) provide detailed reviews about probability aggregation methods and their properties.

The diversity of approaches one can find in the literature may be surprising, but this is because aggregating probabilities is usually an ill-posed problem: there is often in practice a lack of information to describe accurately the interactions between the sources of information. In that framework, we are left with making assumptions and select a method without being able to check the accuracy of the estimations. Essentially, there is neither a single method nor a single set of parameters (as several methods are parametric) that can aggregate probabilities accurately under all possible circumstances. Instead, the selection of the most suitable aggregation method depends on the specific problem which is addressed; a clear understanding of the properties characterizing each aggregation method is therefore an important step.

Clemen and Winkler (1999) proposed a classification of the probability aggregation methods into mathematical combination methods and behavioral approaches.

Behavioral approaches are based on the interaction among experts. The aggregation process concludes with an agreement about a common probability term. Note that in the context of behavioral approaches the word interaction has a meaning strictly related to the fact that the experts are human beings who can exchange advice and discuss their assessments. In geosciences, there is no such exchange of information between different sources. We thus restrict ourselves to mathematical aggregation methods which are functions or operators aggregating probability distributions P_i coming from different sources into a global probability distribution P_G.

In this paper, we provide a formal review of most of the available techniques to aggregate probability distributions as well as a few novel methods. We then discuss their properties in the perspective of earth sciences applications. The paper is structured as follows. In Sect. 3, we define the main mathematical properties of the aggregation methods. We then describe and compare formally the different methods (Sect. 4). Section 5 contains an overview of the main statistical measures of performances; because most methods are parametric, we then describe how the parameters can be estimated. That section contains a new result: if a (generalized) log-linear pooling formula is calibrated, its parameters must be those estimated from maximum likelihood. Through a series of numerical examples, Sect. 6 illustrates the different behaviors of the methods. Finally, Sect. 7 provides guidelines for the selection of a suitable aggregation method and discusses the implications of our study.

2 Set-up and Notations

We wish to assess the probability of an event, denoted A, conditional on the occurrence of a set of data events, D_i, $i = 1, \ldots, n$. This means that we wish to approximate the probability $P(A \mid D_1, \ldots, D_n)$ on the basis of the simultaneous knowledge of the n conditional probabilities $P(A \mid D_i)$. The event A can for example be a lithofacies category at a specified location, while the data D_i can represent information provided by core samples at surrounding wells, a seismic survey, lithofacies patterns on training images, or any other source of information. For categorical events or finite discrete data, the formal probabilistic set-up is the following. We need to consider a sample space Ω such that all events A and D_i are subsets of Ω. In the case of categorical data, let $\mathcal{A}$ be the finite set of events in Ω such that the events $A_1, \ldots, A_K$ of $\mathcal{A}$ are mutually exclusive and exhaustive, that is $\mathcal{A}$ forms a finite partition of Ω. For continuous data, the set-up is slightly more technical, but still straightforward in the context of probability measures. For the clarity of exposition, we will focus on the finite discrete set-up above; most if not all results presented in this paper still hold for continuous probability density functions.

The computation of the full conditional probability $P(A \mid D_1, \ldots, D_n)$ necessitates a probabilistic model of the joint distribution of $(A, D_1, \ldots, D_n)$, a task which is rarely achievable. Instead, we will build an approximation of the true conditional probability by the use of an aggregation operator P_G, also called pooling operator or pooling formula, such that

$$P(A \mid D_1, \ldots, D_n) \approx P_G\big(P(A \mid D_1), \ldots, P(A \mid D_n)\big). \tag{1}$$

Aggregating the probabilities is an ill-posed problem because there is not a unique way of constructing the event $D_1 \cap \cdots \cap D_n$ from the knowledge of the conditional probabilities $P(A \mid D_i)$, $i = 1, \ldots, n$. One of the aims of this paper is to discuss the mathematical properties of such operators and, elaborating from a subset of desirable properties, to build and compare some of them, both from a theoretical point of view and on the basis of performances on simulated cases.

In some circumstances, it will be necessary to include a prior probability on the events $A \in \mathcal{A}$, which will be denoted $P_0(A)$. This prior probability is independent on any other probabilities $P(A \mid D_i)$. It can be thought of as arising from an abstract and never specified information D_0 with $P_0(A) = P(A \mid D_0)$. Equation (1) can thus be generalized in the following way

$$P(A \mid D_0, \ldots, D_n) \approx P_G\big(P(A \mid D_0), \ldots, P(A \mid D_n)\big). \tag{2}$$

In geoscience, such a prior probability could be for example a proportion of a lithofacies varying in space and/or imposed by the user. Note that not specifying explicitly a prior distribution is equivalent to specifying an evenly distributed prior. In the following, the more concise notation P_i will sometimes be used to denote $P(A \mid D_i)$ and the RHS of Eq. (2) will often be rewritten as: $P_G(P_0, P_1, \ldots, P_n)(A)$. At the price of a small abuse of notation, we will adopt the more concise notation $P_G(A)$ for $P_G(P_0, P_1, \ldots, P_n)(A)$ when the context permits.

Some probability aggregation methods are formulated in terms of *odd ratios*, denoted O, defined as

$$O(A) = \frac{P(A)}{1 - P(A)}, \quad 0 \leq P(A) < 1, \tag{3}$$

with the convention $O(A) = +\infty$ when $P(A) = 1$. In the simple case of a binary outcome, where $\mathcal{A} = \{A, \bar{A}\}$, it is easy to check that $O(A)O(\bar{A}) = 1$. When there are more than two elements in $\mathcal{A}$, $\prod_{k=1}^{K} O(A_k)$ can be any fixed value, but Eq. (3) will still be used for defining odd ratios.

3 Mathematical Properties

In this section, we first recall and discuss the main properties that can be used for characterizing aggregation methods. Axiomatic approaches (Bordley 1982; Dietrich 2010) use some of these properties as a starting point to derive classes of aggregation operators.

3.1 Dictatorship

Definition 1 (Dictatorship) A method is dictatorial (Genest and Zidek 1986) when the probability P_i provided by the ith source of information is always taken as the group assessment, that is $P_G(P_1, \ldots, P_i, \ldots, P_n)(A) = P_i(A)$, for all $A \in \mathcal{A}$.

Dictatorship is clearly a pathological property. From now on, we will focus on non-dictatorial aggregation operators.

3.2 Convexity

Definition 2 (Convexity) An aggregation operator P_G verifying

$$P_G \in \big[\min\{P_1, \ldots, P_n\}, \max\{P_1, \ldots, P_n\}\big], \tag{4}$$

is convex.

Definition 3 (Unanimity) An aggregation operator P_G verifying $P_G = p$ when $P_i = p$ for $i = 1, \ldots, n$ is said to preserve unanimity.

It is easy to check that when P_G is convex, $P_i = p$ for $i = 1, \ldots, n$ implies $P_G = p$. Thus, any convex operator preserves unanimity, but the converse is not always true. Unanimity, and thus convexity, is not necessarily a desirable property, as we illustrate now in the two following cases. As a first case, consider that all sources of information yield the same probability because they are all induced by the same event of Ω, for example $D_1 = \cdots = D_n$. Then the true conditional probability can be calculated exactly: $P(A \mid D_1 \cap \cdots \cap D_n) = P(A \mid D_1)$. In this first case, unanimity arises because the D_is are all the same.

As a second case, consider that Ω is finite and consider two information $D_1 \neq D_2$ and an event $A \subset (D_1 \cap D_2)$. Then, $P(A \mid D_1) = P(A)/P(D_1)$, and $P(A \mid D_1 \cap D_2) = P(A)/P(D_1 \cap D_2)$. Now, $(D_1 \cap D_2) \subset D_1$ implies that $P(D_1 \cap D_2) < P(D_1)$. Hence $P(A \mid D_1 \cap D_2) > P(A \mid D_1)$. Thus, in this second case, the full conditional probability of A is larger than any partial conditional probability. In this situation, unanimity, and thus convexity are not desirable properties.

These examples show that whether the pieces of information are similar or different, one should expect the aggregation operator to preserve unanimity or not. Quite often in geosciences, unanimity (and convexity) is a limitation because the conditional probabilities we want to aggregate correspond to very different sources of information. In other words, in geoscience, we are essentially in the second case. Therefore, unanimity, and hence convexity, are properties that should not be sought per se.

3.3 Independence Preservation

Consider two events A and B of Ω such that $A \cap B \neq \emptyset$. Note that since $\mathcal{A}$ is a collection of disjoint events, B is not an element of $\mathcal{A}$.

Definition 4 (Independence Preservation) A method preserves the independence if, whenever we choose two events A and B for which $P_i(A \cap B) = P_i(A)P_i(B)$ is valid for every $i = 1, \ldots, n$, the aggregated probability operator P_G preserves independence

$$P_G(P_1, \ldots, P_n)(A \cap B) = P_G(P_1, \ldots, P_n)(A)\, P_G(P_1, \ldots, P_n)(B) \tag{5}$$

holds.

Many authors (Lehrer and Wagner 1983; Genest 1984; Wagner 1984; Genest and Wagner 1987) faced without success the challenge of finding a pooling formula which preserves independence. Independence preservation is of no direct interest in the context described above, since one usually wants to asses the probability of disjoint events A. Together with Genest and Zidek (1986), our conclusion is that independence preservation is not a reasonable requirement to impose on consensus-finding procedures.

3.4 Marginalization

Consider a vector of events $\mathbf{A} = (A_1, A_2)^t$ and $\mathbf{P}(\mathbf{A}) = (P(A_1), P(A_2))^t$. For each component, $k = 1, 2$ of $\mathbf{A}$ one can define the marginalization operator M_k

$$M_k\big\{\mathbf{P}(\mathbf{A})\big\} = P(A_k). \tag{6}$$

Definition 5 (Marginalization) A pooling operator $\mathbf{P}_G$ verifies the marginalization property if, for each component $k = 1, 2$, the operator M_k commutes with the pooling operator

$$P_G\big\{M_k(\mathbf{P}_1), \ldots, M_k(\mathbf{P}_n)\big\} = M_k\big\{\mathbf{P}_G(\mathbf{P}_1, \ldots, \mathbf{P}_n)\big\}. \tag{7}$$

There is only one pooling operator satisfying the marginalization property, namely the linear pooling method. But we will see below that it does not verify other more interesting properties.

3.5 External Bayesianity

The external Bayesianity property is related to the behavior of an aggregation operator when additional information becomes available. Consider that the probabilities can be updated by a likelihood, L, common to all sources of information. We thus consider now the probabilities

$$P_i^L(A) = \frac{L(A)P_i(A)}{\sum_{A\in\mathcal{A}} L(A)P_i(A)}, \quad i = 1, \ldots, n,$$

where $L(A)$ is such that $\sum_{A\in\mathcal{A}} L(A) < \infty$.

Definition 6 (External Bayesianity) An aggregation operator is said to be external Bayesian if the operation of updating the probabilities with the likelihood L commutes with the aggregation operator, that is if

$$P_G\big(P_1^L, \ldots, P_n^L\big)(A) = P_G^L(P_1, \ldots, P_n)(A). \tag{8}$$

Essentially this means that it should not matter whether new information arrives before or after pooling. This property is equivalent to the weak likelihood ratio property in Bordley (1982). External Bayesianity is a very compelling property, both from a theoretical point of view and from an algorithmic point of view. We will see that imposing this property leads to a very specific class of pooling operators.

3.6 Certainty Effect

An interesting feature of an aggregation method is its response to situations where a source of information provides a conditional probability equal to 0 (impossible event) or 1 (certain event). Let us suppose that there exists i such that $P(A \mid D_i) = 0$ and $P(A \mid D_j) \neq 1$ for $j \neq i$.

Definition 7 (0/1 forcing property) An aggregation operator which returns $P_G(A) = 0$ in the above-mentioned case is said to enforce a certainty effect, a property also called the 0/1 forcing property (Allard et al. 2011).

Note that the same is true if $P(A \mid D_i) = 1$, since in this case $P(A' \mid D_i) = 0$, for all $A' \neq A \in \mathcal{A}$. In geoscience, this property is convenient to reproduce depositional sequences or catenary patterns. The drawback is that deadlock situations are possible, when $P(A \mid D_i) = 0$ and $P(A \mid D_j) = 1$ for $j \neq i$. Deadlocks can arise when data are inconsistent with each other. A practical solution can be to consider probabilities in a constrained interval, for example $[0.001, 0.999]$.

4 Aggregation Methods

Aggregation methods can be divided into methods derived from axiomatic approaches and methods derived from model considerations. Genest and Zidek (1986), Bordley (1982) and Dietrich (2010) restricted themselves to the binary case, that is when there are only two possible outcomes, namely A and $\bar{A}$ in $\mathcal{A}$. Bordley (1982) showed that there is only one class of aggregation operator verifying at the same time a set of structural axioms always verified in geoscience (weak ordering of the $O_i(A)$ with respect to A, non-interaction between source of information, continuity) and the weak likelihood ratio condition (or external Bayesianity). The associated pooling formula, hereafter called Bordley formula, combines odds multiplicatively. In the same spirit, Genest and Zidek (1986) show that the unique aggregation operator verifying the same structural axioms and external Bayesianity is the log-linear pooling. These two results turn out to be equivalent in the binary case, but lead to different pooling formulas in the general case of more than two possible outcomes. Still in the binary case, Dietrich (2010) shows that for a very close set of structural axioms, the only pooling formula verifying the property of independent information is a particular case of the log-linear pooling formula.

Following a model-based approach, Journel (2002) proposed the Tau model, which turns out to be equivalent to the Bordley formula (Krishnan 2008). In Polyakova and Journel (2007), the Nu-model is proposed as an alternative to the Tau model. Although no mentions are explicitly made in these papers to any restriction to the binary case, it must be noted that it is, in fact, the case for all considered examples. It turns out that it is equivalent to work with probabilities or with odds in the binary case. This equivalence is lost if there are more than two possible outcomes in $\mathcal{A}$. We will show that there are two quite different routes for generalizing the Nu model to the non-binary case. We will also show how this Nu-model is related to log-linear

pooling methods and that following a maximum entropy principle or equivalently a conditional independence assumption, entails a specific, parameter-free form of the Bordley formula. The resulting pooling formula is similar to the Markovian-type Categorical Prediction (MCP) equations in Allard et al. (2011).

There is yet another enlightening dichotomy. Some methods combine probabilities in an additive way, leading to a linear pooling formula and its generalization, in the spirit of the disjunction operation of probability distributions (Tarantola and Valette 1982; Tarantola 2005). Other methods combine probabilities or odds in a multiplicative way, which corresponds to the conjunction operation of probability distributions (Tarantola and Valette 1982; Tarantola 2005). This last criterion defines two very different groups within which the aggregation methods share many common properties. The next subsections, following and extending the work of Genest and Zidek (1986), Clemen and Winkler (1999), and Clemen and Winkler (2007), provide a summary of some of the most important aggregation methods in earth sciences.

4.1 Additive Methods and Transformed Additive Methods

4.1.1 Linear Pooling

Probably the most intuitive way of aggregating the probabilities $P_1, \ldots, P_n$ is the linear pooling, proposed by Stone (1961) and attributed to Laplace by Bacharach (1979)

$$P_G(A) = \sum_{i=1}^{n} w_i P_i(A), \tag{9}$$

where the w_i are positive weights verifying $\sum_{i=1}^{n} w_i = 1$ in order to have a meaningful global probability. Since the linear pooling is simple to understand and to implement, it is probably the most common aggregation method. However, Ranjan and Gneiting (2010) demonstrated that the linear pooling is intrinsically sub-optimal. This point will be detailed in the next sections.

Linear pooling neither verifies independence preservation, 0/1 forcing properties, nor external Bayesianity unless it is dictatorial (for example $w_i = 1$ for one source D_i and $w_j = 0$, for all $j \neq i$). It is a convex aggregation method, and as a consequence, it does preserve unanimity. As already discussed in Sect. 3.2, this property might be considered as a serious limitation in the context of geoscience modeling. If we provide an equal weight w_i to every probability P_i the method reduces to an arithmetic average; in this case it coincides with the disjunction of probabilities (Tarantola and Valette 1982; Tarantola 2005).

Genest (1984) proved that all pooling operators verifying the marginalization property are of the form

$$P_G(A) = \sum_{i=0}^{n} w_i P_i(A), \tag{10}$$

where P_0 is a prior probability and where the weights $w_0, \ldots, w_n \in [-1, 1]$ add up to one and must satisfy other consistency conditions to ensure that P_G is a probability

measure. The aggregation operator defined by Eq. (10) is called generalized linear pooling. The possibility of negative weights is in theory interesting, but we are faced with the problem of finding weights w_i insuring that P_G is a probability on $\mathcal{A}$. A safe option is to restrict ourselves to weights $w_0, \ldots, w_n \in [0, 1]$ adding to 1. If $w_0 = 0$ we are back to the linear opinion pool.

The resulting probability distribution P_G is very often multi-modal, a not so desired situation. The reasons are profound. From a probabilistic point of view, Eqs. (9) and (10) represent mixture models in which each probability P_i represents a different population; the aggregated probability P_G is then the result of the following hierarchical random experiment: first select a population i with the probability distribution defined by $\mathbf{w} = (w_0, \ldots, w_n)$; then select an event A according to probability distribution P_i. In general, this mixture of population model does not correspond to our geoscience context in which we wish to aggregate partial information on the same object.

4.1.2 Beta-Transformed Linear Pooling

Ranjan and Gneiting (2010) proposed to apply a Beta transformation to linear pooling operators in order to improve their performance, thereby defining the Beta-transformed Linear Pooling (BLP)

$$P_G(A) = H_{\alpha,\beta}\left(\sum_{i=1}^{n} w_i P_i(A)\right), \tag{11}$$

where the weights must be positive and add up to one. The function $H_{\alpha,\beta}$ is the cumulative density function of a beta distribution with shape parameters $\alpha > 0$ and $\beta > 0$

$$\begin{aligned} &H_{\alpha,\beta}(x) = B(\alpha,\beta)^{-1}\int_0^x t^{\alpha-1}(1-t)^{\beta-1}\mathrm{d}t \\ &\text{with} \quad x \in [0,1] \quad \text{and} \quad B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\mathrm{d}t. \end{aligned} \tag{12}$$

BLP includes the linear pooling (LP) when $\alpha = \beta = 1$, since $H_{1,1}(x) = x$, for $0 \le x \le 1$. For other values of the parameters, the marginalization property verified by LP is lost because of the Beta transformation. However, as it is the case for LP, the 0/1 forcing property is not verified unless dictatorship holds. In general, this transformation leads to non-convex aggregation probabilities. In their work, Ranjan and Gneiting (2010) show, on simulations and on real case studies, that the BLP constantly outperforms LP and that it presents very good performances.

4.2 Methods Based on the Multiplication of Probabilities

We have seen in the previous section that additive aggregation methods correspond to mixture models. They are related to union of events and to the logical operator OR. In our context, the information conveyed by the events D_i should rather be aggregated by the logical operator AND, related to the intersection of events. Intuitively,

aggregation operators based on multiplication seem therefore more appropriate than those based on addition. We now present and discuss different aggregation methods based on the multiplication of probabilities.

4.2.1 Log-Linear Pooling

Definition 8 A log-linear pooling operator is a linear operator of the logarithms of the probabilities

$$\ln P_G(A) = \ln Z + \sum_{i=1}^{n} w_i \ln P_i(A), \tag{13}$$

or equivalently

$$P_G(A) \propto \prod_{i=1}^{n} P_i(A)^{w_i}, \tag{14}$$

where Z is a normalizing constant.

Genest and Zidek (1986) showed that all pooling operators verifying external Bayesianity must be of the form Eq. (14) with the additional condition that $\sum_{i=1}^{n} w_i = 1$. This condition also implies that unanimity is preserved. Log-linear pooling does not preserve independence and does not verify the marginalization property. Unlike linear pooling, it is typically unimodal and less dispersed. Since it is based on a product, it verifies the 0/1 forcing property. One particular possibility consists in setting $w_i = 1$ for each $i \neq 0$. This corresponds to the conjunction of probabilities (Tarantola and Valette 1982; Tarantola 2005).

If a prior probability $P_0(A)$ must be included, Eq. (14) becomes $P_G(A) \propto \prod_{i=0}^{n} P_i(A)^{w_i}$ with the restriction $\sum_{i=0}^{n} w_i = 1$ to verify external Bayesianity, yet better written

$$P_G(A) \propto P_0(A)^{1-\sum_{i=1}^{n} w_i} \prod_{i=1}^{n} P_i(A)^{w_i}. \tag{15}$$

In Eq. (15), there is no restriction on the weights $\mathbf{w} = (w_1, \ldots, w_n)$, and $\sum_{i=0}^{n} w_i = 1$ is always verified. Note that if neither external Bayesianity nor unanimity are properties that must be verified, there are no constraints whatsoever on the weights w_i, $i = 0, \ldots, n$.

It is always possible to write the conditional probability $P(A \mid D_1, \ldots, D_n)$ with a log-linear formalism. Let us introduce the following convenient notation. We will denote $D_{<i} = \{D_1 \cap \cdots \cap D_{i-1}\}$, with the convention $D_{<1} = \Omega$. Then

$$\begin{aligned} P(A \mid D_1, \ldots, D_n) &= \frac{P_0(A) P(D_1, \ldots, D_n \mid A)}{\sum_{A \in \mathcal{A}} P_0(A) P(D_1, \ldots, D_n \mid A)} \\ &= \frac{P(A) \prod_{i=1}^{n} P(D_i \mid A, D_{<i})}{\sum_{A \in \mathcal{A}} P(A) \prod_{i=1}^{n} P(D_i \mid A, D_{<i})} \end{aligned} \tag{16}$$

$$= \frac{P(A)^{1-S_{\mathbf{w}}} \prod_{i=1}^{n} P(A \mid D_i)^{w_{A,D_1,\ldots,D_n}}}{\sum_{A \in \mathcal{A}} P(A)^{1-S_{\mathbf{w}}} \prod_{i=1}^{n} P(A \mid D_i)^{w_{A,D_1,\ldots,D_n}}}, \quad (17)$$

with $w_{A,D_1,\ldots,D_n} = \ln P(D_i \mid A, D_{<i}) / \ln P(D_i \mid A)$. This decomposition is exact if there is one weight w per combination $(A, D_1, \ldots, D_n)$. Log-linear pooling, as in Eq. (15), amounts to making the simplifying assumption

$$\ln P(D_i \mid A, D_{<i}) / \ln P(D_i \mid A) = w_i, \quad (18)$$

for all A, all D_i and all $D_{<i}$, which can be verified for some, but not all, probability models.

The sum $S_{\mathbf{w}} = \sum_{i=1}^{n} w_i$ plays an important role in Eq. (15). If $S_{\mathbf{w}} = 1$, the prior probability P_0 is filtered out since $w_0 = 0$ and unanimity is preserved. Otherwise, unanimity is not preserved. Suppose that $P_i = p$ for each $i = 1, \ldots, n$. If $S_{\mathbf{w}} > 1$, the prior probability has a negative weight and P_G will always be further from P_0 than p. This corresponds to the second case illustrating convexity in Sect. 3. Conversely, if $S_{\mathbf{w}} < 1$, P_G is always closer from P_0 than p. And of course, $P_G = p$ if $S_w = 1$. The influence of the prior probability P_0 on the aggregated result P_G can thus be tuned by changing the value of $S_{\mathbf{w}}$.

4.2.2 Generalized Logarithmic Pooling

Genest and Zidek (1986) showed that if we allow the explicit form of P_G to depend upon A, that is if we allow P_G to be of the form

$$P_G(P_1, \ldots, P_n)(A) \propto G\big(A, P_1(A), \ldots, P_n(A)\big),$$

the only pooling operator verifying external Bayesianity is

$$P_G(A) \propto H(A) \prod_{i=1}^{n} P(A \mid D_i)^{w_i}, \quad (19)$$

with $\sum_{i=1}^{n} w_i = 1$ and $H(A)$ being an arbitrary bounded function playing the role of a likelihood on the elements of $\mathcal{A}$. In this case, if all conditional probabilities are equal, the aggregated probability is proportional to p updated by $H(A)$: $P_G(A) \propto H(A)p$.

4.2.3 Maximum Entropy Approach

Instead of establishing a pooling formula from an axiomatic point of view, one can choose to optimize a criterion, for example to minimize the distance between the distribution P and its approximation. The Kullback–Leibler (KL) divergence (Kullback and Leibler 1951) or relative entropy, between a distribution P and another distribution (here its approximation P_G) is

$$D(P_G \,||\, P) = \mathbb{E}_{P_G}\left[\ln \frac{P_G}{P}\right]. \quad (20)$$

Although not a distance in the mathematical sense (it is not symmetrical), the KL divergence is a measure of how much different two probability distributions are. It is always positive and it is equal to zero if, and only if, $P_G = P$. There are strong connections between entropy and KL divergence (Cover and Thomas 2006). In particular, let us assume that some quantities related to P are known, such as moments or conditional probabilities. A natural approach, very common in information theory, computer science, image, and language processing is to find the distribution P_G that shares properties (moments or conditional probabilities) with P and minimizes the KL divergence $D(P_G||P)$. This can be shown equivalent to finding the distribution P_G maximizing its entropy $H(P_G) = \mathbb{E}_{P_G}[P_G]$, subject to the imposed constraints. Allard et al. (2011) developed such an approach for the prediction of spatial categorical variables leading to a Markovian-type categorical prediction (MCP), which was shown to be a very good approximation of the Bayesian maximum entropy (BME) principle (Christakos 1990) with the advantage of being computationally efficient. Following a similar route, we obtain the following result. Here, we need to use the full notation $P_G(P_1, \ldots, P_n)(A)$.

Proposition 1 The pooling formula P_G maximizing the entropy subject to the following univariate and bivariate constraints $P_G(P_0)(A) = P_0(A)$ and $P_G(P_0, P_i)(A) = P(A \mid D_i)$ for $i = 1, \ldots, n$ is

$$P_G(P_1, \ldots, P_n)(A) = \frac{P_0(A)^{1-n} \prod_{i=1}^n P_i(A)}{\sum_{A \in \mathcal{A}} P_0(A)^{1-n} \prod_{i=1}^n P_i(A)}. \tag{21}$$

The proof of this proposition is given in Appendix A. Notice that the maximum entropy approximation Eq. (21) is a special case of the logarithmic pooling formula with $w_i = 1$, for $i = 1, \ldots, n$.

The same formula can also be obtained as a result of the conditional independence assumption. Let us assume that P verifies a conditional independence assumption, that is

$$P(D_0, \ldots, D_n \mid A) = \prod_{i=0}^{n} P(D_i \mid A), \tag{22}$$

for all events $A, D_1, \ldots, D_n$. Conditional independence implies

$$P(D_i \mid A, D_{<i}) = P(D_i \mid A).$$

Hence, Eq. (16) becomes

$$\begin{aligned} P_G(A) &= \frac{P(A) \prod_{i=1}^n P(D_i \mid A)}{\sum_{A \in \mathcal{A}} P(A) \prod_{i=1}^n P(D_i \mid A)} \\ &= \frac{P_0(A)^{1-n} \prod_{i=1}^n P(A \mid D_i) P(D_i)}{\sum_{A \in \mathcal{A}} P_0(A)^{1-n} \prod_{i=1}^n P(A \mid D_i) P(D_i)} \\ &= \frac{P_0(A)^{1-n} \prod_{i=1}^n P_i(A)}{\sum_{A \in \mathcal{A}} P_0(A)^{1-n} \prod_{i=1}^n P_i(A)}. \end{aligned}$$

Put together this last result and Eq. (21) allows us to state the following equivalence.

Proposition 2 Regarding the aggregation of probabilities considered in this work, Maximum Entropy is equivalent to Conditional Independence.

4.2.4 Probability Multiplication Formulas in Summary

Multiplication of the probabilities offers a large class of pooling operators, with interesting subclasses which can be summarized in the following way

$$\begin{aligned}\{\text{Max. Ent.} \equiv \text{Cond. Ind. pooling}\} &\subset \{\text{Ext. Bayes. pooling}\} \\ &\subset \{\text{Log-linear pooling}\}. \end{aligned} \tag{23}$$

The pooling formula corresponding to the maximum entropy principle/conditional independence assumption (21) is particularly easy to implement since it is parameter free. The larger class of pooling formula (15) corresponds to pooling operators verifying the external Bayesianity condition in which the weights are constrained to add up to 1. For this class, the value of $S_{\mathbf{w}}$ is the key factor regarding the behavior with respect to the prior probability P_0. The largest class of pooling operators is of the same form but does not impose any restriction on the weights. This largest class does not verify any mathematical properties presented in Sect. 3, but the 0/1 forcing property.

4.3 Methods Based on the Multiplication of Odds

When using odds $O(A)$, it will be important to distinguish two cases:

1. In the first, more restrictive, case there are only two possible outcomes, such as $\mathcal{A} = \{A, \bar{A}\}$. In this case, $P(A) + P(\bar{A}) = O(A) \cdot O(\bar{A}) = 1$. This case will be called the binary case hereafter.
2. In the second case, there are more than two possible outcomes in $\mathcal{A}$. In this case, there is no general relationships between the odds $O(A)$, and in general $\prod_{A\in\mathcal{A}} O(A) \neq 1$.

We will see that in the binary case, it is completely equivalent to consider operators based on the product of odds and operators based on products of probabilities. In the general case, this equivalence is lost.

4.3.1 Bordley Formula and Tau Model

Binary Case We first restrict ourselves to the binary case. Bordley (1982) showed that in this case, the only pooling operator verifying the weak likelihood ratio axiom (see Definition 6) in addition to other natural axioms is a pooling formula based on the product of the odd ratios

$$O_G(A) = O_0(A)^{w_0} \prod_{i=1}^{n} \left(\frac{O_i(A)}{O_0(A)} \right)^{w_i} = O_0(A)^{w_0 - \sum_{i=1}^{n} w_i} \prod_{i=1}^{n} O_i(A), \tag{24}$$

where the weights w_i can vary in $[0, \infty)$. Now, using $P_i(A) = O_i(A)/(1 + O_i(A))$, and denoting $P_i(A) = P(A \mid D_i)$, Eq. (24) becomes

$$P_G(A) = \frac{P_0(A)\prod_{i=1}^{n}(P_i(A)/P_0(A))^{w_i}}{P_0(A)\prod_{i=1}^{n}(P_i(A)/P_0(A))^{w_i} + (1-P_0(A))\prod_{i=1}^{n}[(1-P_i(A))/(1-P_0(A))]^{w_i}}, \tag{25}$$

or equivalently

$$P_G(A) \propto P_0(A)^{1-\sum_{i=1}^{n} w_i} \prod_{i=1}^{n} P_i(A)^{w_i}, \tag{26}$$

which is nothing but Eq. (15). Hence, we can state the following equivalence in Proposition 3.

Proposition 3 In the binary case, the Bordley formula is equivalent to a log-linear pooling formula verifying external Bayesianity.

Journel (2002) derived a formula for aggregating probabilities that has been later named the Tau model. For presenting this model, we will use our usual notations, which are slightly different than those in Journel (2002), Polyakova and Journel (2007) and Krishnan (2008). In particular, these authors use the inverse of odds-ratio instead of odds-ratio, but since the formulae are purely multiplicative this point is of secondary importance.

In a first step, Journel (2002) sets as an axiom the permanence of ratio principle, which states (using our notations) that "the incremental contribution of data event D_2 to the knowledge of A is the same after or before knowing D_1". Mathematically,

$$\frac{O_G(A \mid D_1, D_2)}{O_G(A \mid D_1)} = \frac{O_G(A \mid D_2)}{O_G(A)}. \tag{27}$$

From this principle, one can easily establish that

$$O_G(A) = O_0(A)^{1-n} \prod_{i=1}^{n} O_i(A),$$

which is a Bordley formula with $w_i = 1$, for $i = 1, \ldots, n$. Replacing $O_i(A)$ by $P_i(A)/(1 + P_i(A))$, one gets $P_G(A) \propto P_0(A)^{1-n} \prod_{i=1}^{n} P_i(A)$, which is nothing but Eq. (21). Hence, we established the following proposition.

Proposition 4 In the case of a binary event, the permanence of ratio principle is equivalent to conditional independence, which is equivalent to a maximum entropy principle.

In a second step, Journel (2002) reintroduced dependence between the source of information by generalizing this formula thus obtaining the general Bordley formula (24). Krishnan (2008) provides the expression of the parameters w_i as a function

of conditional probabilities obtained from the full joint probability, but this exercise is unfortunately only of academic interest since if the full joint model was known, an approximate formula such as the Tau model would not be necessary anymore.

General Case The general case with more than two possible outcomes in $\mathcal{A}$, was not considered in Bordley (1982). In Journel (2002), Polyakova and Journel (2007), and Krishnan (2008), the Tau model is exclusively presented in the case of binary event, either explicitly or implicitly. What happens in the general case with $K > 2$ possible outcomes is rarely addressed explicitly. In this case, the quantities $O(A_1), \ldots, O(A_K)$ in Eq. (24), although computable when the probabilities belong to $[0, 1)$, are not odds in the usual sense. Back-transforming the odds into probabilities using $P_G(\cdot) = O_G(\cdot)/(1 + O_G(\cdot))$ does not lead to quantities adding to one. A normalization step is thus required to obtain a regular probability distribution. A complete formulation of the Tau model in the general case is thus

$$P_G(A) \propto O_G(A)/\big(1 + O_G(A)\big), \quad \text{with}$$
$$O_G(A) = O_0(A)^{1-\sum_{i=1}^n w_i} \prod_{i=1}^n O_i(A)^{w_i}, \quad A \in \mathcal{A}. \tag{28}$$

We thus obtain the following equivalence of Proposition 5.

Proposition 5 The Tau model is equivalent to the Bordley formula; only in the case of a binary event, they both are equivalent to a log-linear pooling.

Note that since $O_G(A) = 0 \Leftrightarrow P_G(A) = 0$, the Tau model (28) verifies the 0/1 forcing property, both in the binary and in the general case.

4.3.2 The Nu Model

The Nu model was proposed in Polyakova and Journel (2007) as an alternative to the Tau model. We first re-derive its expression using our notations before discussing its relationships with the other pooling methods. It will be useful to distinguish the binary case from the general case.

Binary Case Let us first consider the binary case. We start from the exact decomposition of Eq. (16)

$$P(A \mid D_1, \ldots, D_n) = \frac{P(A) \prod_{i=1}^n P(D_i \mid A, D_{<i})}{\sum_{A \in \mathcal{A}} P(A) \prod_{i=1}^n P(D_i \mid A, D_{<i})},$$

and we denote $\nu_i^*(A) = P(D_i \mid A, D_{<i})/P(D_i \mid A)$. Then, defining $\nu^*(A) = \prod_{i=1}^n \nu_i^*(A)$, one can write

$$P(A \mid D_1, \ldots, D_n) = \frac{P(A)\prod_{i=1}^{n} \nu_i^*(A) P(D_i \mid A)}{\sum_{A\in\mathcal{A}} P(A)\prod_{i=1}^{n} \nu_i^*(A) P(D_i \mid A)}$$
$$= \frac{P(A)^{1-n}\nu^*(A)\prod_{i=1}^{n} P(A \mid D_i)}{\sum_{A\in\mathcal{A}} P(A)^{1-n}\nu^*(A)\prod_{i=1}^{n} P(A \mid D_i)}. \quad (29)$$

From this we obtain, the Nu model

$$P_G(A) \propto P_0(A)^{1-n}\nu^*(A)\prod_{i=1}^{n} P(A \mid D_i). \quad (30)$$

In terms of odds, denoting $\nu(A) = \nu^*(A)/(1-\nu^*(A))$

$$O_G(A) = \frac{O_0(A)^{1-n}\nu(A)\prod_{i=1}^{n} O_i(A)}{\sum_{A\in\mathcal{A}} O_0(A)^{1-n}\nu(A)\prod_{i=1}^{n} O_i(A)}, \quad (31)$$

which is the Nu model. Note that in Eq. (30) the factors $\nu^*(A)$ are defined slightly differently than in Polyakova and Journel (2007). After transformation into $\nu(A)$, they lead, however, to the same analytical expression of Eq. (31) the only difference being that our $\nu(A)$ is the inverse of the factor ν_0^{-1} in Polyakova and Journel (2007, Eq. 5). Remember that when applying the Nu model in practice, the quantities $\nu_i(A)$ are not known since $P(D_i \mid A, D_{<i})$ are unknown. They must be considered as parameters to be estimated or set by the user. From Eq. (30), one can see that $\nu^*(A)$ acts as a kind of likelihood which updates the probability $P(A)$ to $P^*(A)^{1-n} = \nu^*(A)P(A)^{1-n}$. The Nu model thus verifies the external Bayesianity condition. Since we are in the binary case, $O_G(\cdot)$ must satisfy $O_G(A).O_G(\bar{A}) = 1$, which implies that $\nu(A).\nu(\bar{A}) = 1$, that is $\nu(A)$ are odds.

Proposition 6 For the binary case $\mathcal{A} = \{A, \bar{A}\}$, the Nu model is equivalent to:

(i) a maximum entropy pooling formula updated by the odds $(\nu(A), 1/\nu(A))$;
(ii) a generalized logarithmic pooling formula with $w_i = 1$, for $i = 1, \ldots, n$.

The maximum entropy formula corresponds to Eq. (30) with $\nu^*(A) = 1$ for all $A \in \mathcal{A}$. Conditional independence in Eq. (22) is a sufficient condition for this, but in theory it is not necessary. If $\nu(A)$ is close to a constant c for all A, the maximum entropy pooling formula Eq. (21) is an excellent approximation of Eq. (30). Note that in Eq. (31) the particular status of $\nu(A)$ as compared to $P_0(A)$ is a little bit unclear.

General Case In the general case with $K > 2$ possible outcomes in $\mathcal{A}$ (Eqs. (30) and (31)) are not equivalent. Two routes are possible for generalizing the Nu model.

1. *The first route* (Nu-1) consists in generalizing the pooling of the probabilities, as in Eq. (30), thus obtaining a generalized or updated maximum entropy formula. Would the full joint probability be accessible, the quantities $\nu^*(A)$ could be ex-

Table 1 Aggregated probability computed according to the two possible generalization of the Nu model

		A_1	A_2	A_3
	P_0	0.6	0.3	0.1
	P_1	1/3	1/3	1/3
	P_2	0.6	0.15	0.25
$(\nu(A_1), \nu(A_2), \nu(A_3))$		P_G		
(1, 1, 1)	Nu-1	0.250	0.125	0.625
(1, 1, 1)	Nu-2	0.302	0.155	0.543
(2, 2, 2)	Nu-1	0.250	0.125	0.625
(2, 2, 2)	Nu-2	0.324	0.189	0.487
(1, 2, 3)	Nu-1	0.105	0.105	0.790
(1, 2, 3)	Nu-2	0.231	0.202	0.567
(0.28, 0.68, 8)	Nu-2	0.105	0.105	0.790

actly computed. This not being the case, $\nu^*(A)$, if not set equal to 1, acts as a kind of likelihood, as already seen in the binary case.

2. *The second route* (Nu-2) considered in Polyakova and Journel (2007) consists in generalizing the pooling of the odds, as in Eq. (31), thus leading to

$$P_G(A) \propto O_G(A)/\big(1 + O_G(A)\big),$$
$$O_G(A) = \frac{O(A)^{1-n}\nu(A)\prod_{i=1}^{n} O_i(A)}{\sum_{A\in\mathcal{A}} O(A)^{1-n}\nu(A)\prod_{i=1}^{n} O_i(A)}. \tag{32}$$

In this second route, $\nu(A)$ acts as an odd updating the product of odds. Increasing $\nu(A)$ leads to an increase of the probability $P_G(A)$.

It is important to stress that, when not in the binary case, these two routes will lead to different values of the aggregated probability $P_G(A)$ for given values of $\nu(A)$. This is illustrated in Table 1, in which $P_G(A)$ is computed according to the Nu-1 or Nu-2 representation for several values of $\nu(A)$. Note that since $w_1 + w_2 = 2 > 1$, the aggregated probability will always be further away from the prior P_0 than the probabilities P_i (see Proposition 6(ii)). Hence, for all considered cases, P_G is the highest for A_3. One can also see that when $\nu(A)$ is evenly distributed, the value of $\nu(A)$ does not play any role when following the first route, which can be seen from Eq. (30), while it does play a role when following the second route. These results illustrate the fact that the first route corresponds to the external Bayesianity condition, with $\nu(A)$ playing the role of an external likelihood. When $\nu(A)$ is uneven, higher values of $\nu(A)$ yield to larger aggregated probabilities. For a given vector for $\nu(A)$, the first route ($\nu(A)$ multiplying probabilities) leads to more extreme probabilities, while the second route ($\nu(A)$ multiplying odds) leads to more equilibrated probabilities. It is, however, possible to find a vector of values along the second route leading to approximately the same aggregated probabilities.

It is also important to understand the profound difference between Bordley/Tau and Nu aggregations. While in the former there is for each source of information a

Table 2 General presentation of non linear aggregation methods

Weights	Likelihood	$K = 2$	$K > 2$	
		Probs ≡ Odds	Probabilities	Odds
When $\sum_{i=1}^{n} w_i = 1$, Ext. Bayesianity and unanimity are verified	$\nu(A) = 1$	Log-linear = Bordley = Tau model	Log-linear	Tau model
	$\nu(A) \neq 1$	Gen. log-linear	Gen. log-linear	–
All $w_i = 1$	$\nu(A) = 1$	Cond. Indep. = Max. Entropy	Cond. Indep. ≡ Max. Entropy	–
	$\nu(A) \neq 1$	Nu model	Nu-1 ≡ updated Max. Ent.	Nu-2 (Polyakova and Journel 2007)

single parameter w_i independent on the event A, in the latter there is a one parameter per event A without any mention to the source of information.

4.4 Multiplication Methods at a Glance

As seen in the previous sections, methods based on the multiplication of probabilities or multiplication of odds are intimately related. Presenting all methods in Table 2 makes it possible to grasp the relationships between the multiplication methods in one glance. At the first level, we make a distinction between the binary case and the general case. We re-emphasize that most of the literature is concerned with the binary case, either explicitly or implicitly, for which methods based on odds are equivalent to methods based on probabilities. On the contrary, it is important to distinguish these two cases when dealing with non-binary events.

A general formulation of all pooling methods is possible

$$T_G(A) = Z + U(A) + \left(1 - \sum_{i=1}^{n} w_i\right) T_0(A) + \sum_{i=1}^{n} w_i T_i(A), \tag{33}$$

in which T is related to probabilities in the following way: $T \equiv P$ for all linear pooling methods; $T \equiv \ln P$ for methods based on the product of probabilities, and $T \equiv \ln O = \ln P - \ln(1 - P)$ for methods based on the product of odds. $U(A)$ is an updating likelihood when considering the general log-linear pooling; it is the logarithm of the Nu parameter for the Nu model. $T_0(A)$ is the prior probability and Z is a normalizing constant. The weight w_0 has been set equal to $1 - \sum_{i=1}^{n} w_i$ in order to respect external Bayesianity. Note that $w_i = 1$ for the Nu model and the maximum entropy. When $T \equiv P$, the Beta-transformed model can also be included by transforming the right-hand side of Eq. (33) with the Beta cumulative probability function $H_{\alpha,\beta}$.

Table 3 Main properties of methods for aggregating n sources of information and a prior term when there are K alternatives

	Lin.	BLP $((\alpha, \beta) \neq (1, 1))$	ME	Nu-1	Nu-2*	Log.-linear	Gen. Log-lin
Convexity	yes	no	no	no	no	no	no
Marginalization	yes	no	no	no	no	no	no
0/1 forcing	no	no	yes	yes	yes	yes	yes
Ext. Bayes.	no	no	yes	yes	no*	cond. yes	cond. yes
# of param.	$n-1$	$n+1^{\dagger}$	0	$K-1$	$K-1$	n	$n+K-1$

Note that some properties not verified in the general case are verified for some very specific values, which either reduce the method to a different method or to dictatorship. The no* are yes when $K = 2$; Nu-2 = Nu-1 when $K = 2$. †Number of parameters in BLP is n if we impose $\alpha = \beta$. Cond. yes means yes when the condition $S_{\mathbf{w}} = 1$ is verified

5 Choosing a Pooling Formula, Estimating the Weights and Assessing the Forecast

5.1 Introduction

Table 3 recapitulates the previous sections about the aggregation methods and their properties. A first dichotomy is between methods based on addition and those based on multiplication. BLP is intermediate. Unlike linear pooling, the BLP is not convex and does not verify marginalization; at the same time, it is different than the multiplicative methods because it does not verify the 0/1 forcing property. This last property is verified by all multiplicative methods. External Bayesianity is verified by the generalized log-linear model, the Nu model and the Bordley formula for binary events. In the more general case, it is always verified by the first route generalizing the Nu model. It is also verified by the log-linear model and the generalized log-linear model, conditional on the sum of the weights being equal to 1.

The role of the prior deserves some discussion. All aggregation formula allow to take into account some form of prior probability, which could for example represent non-stationary proportions. As it can be seen in Eq. (33), in multiplicative methods the role of prior is multiplicative. More precisely, since ratios P_i/P_0 are aggregated, these methods can be very sensitive to the specification of the prior. The influence of the prior depends on the sum $S_{\mathbf{w}} = \sum_{i=1}^{n} w_i$. When $S_{\mathbf{w}} = 1$, the prior is filtered out. When $S_{\mathbf{w}} > 1$, the aggregated probability P_G will be further away from P_0 than the P_is. Contrarily, if $S_{\mathbf{w}} < 1$, P_G will be closer from P_0 than the P_is. Since maximum entropy is a model with $S_{\mathbf{w}} = n$, we can expect this method to greatly amplify the departure to the prior.

At the exception of the maximum entropy approach which is parameter free, all methods presented above have some parameters that need either to be estimated or set by the user. In the Nu model, there are $K-1$ parameters, where K is the cardinality of $\mathcal{A}$, while for the log-linear formula and the Bordley/Tau model there are n parameters. The most general model is the generalized log-linear, with $K+n-1$ parameters if not imposing external Bayesianity. In theory, if the full probability model

was known, expressions for the parameters would be accessible. But in this case, the conditional probability would also be accessible, and a pooling formula would not be sought in the first place.

In the context of aggregating expert opinion, Winkler (1968) suggests four ways of assessing the weights for the linear pool, which could also be applied to the other methods:

(i) equal weights;
(ii) weights proportional to a ranking based on expert's advice;
(iii) weights proportional to a self-rating (each source of information provide a rank for itself) and;
(iv) weights based on some comparison of previously assessed distributions with actual outcomes.

Setting equal weights is sometimes relevant when there is no element which allows to prefer one source of information to another, or when symmetry of information justifies it. But even in this case, the sum $S_{\mathbf{w}}$ needs to be set or estimated. Suggestions (ii) and (iii) might be relevant in the context of human judgments, but of no great use in a geoscience context.

When training data are available (case (iv)) it is possible to estimate the optimum weights according to the optimization of some criterion. Heskes (1998) proposed an algorithm based on the minimization of a Kullback–Leibler distance for selecting weighting factors in logarithmic opinion pools. The optimal weights are found by solving a quadratic programming problem. Ranjan and Gneiting (2010) minimized the likelihood for finding the optimal shape parameters for the Beta-transformed linear opinion pool. Cao et al. (2009) used ordinary kriging to estimate the parameters of the Tau model, but the concept of distance between source of information and that of variogram of probabilities is not at all obvious. We will present the likelihood approach for estimating the parameters for methods based on the multiplication of probabilities in the next sections. A similar derivation for the linear opinion pool and its Beta transform can be found Ranjan and Gneiting (2010).

5.2 Scoring Rules and Divergence

The aggregated probability distribution $P_G(A)$ must be as close as possible to the (unknown) conditional probability distribution $P(A \mid D_1, \ldots, D_n)$, $A \in \mathcal{A}$. Scoring rules (Gneiting and Raftery 2007) provide summary measures for the evaluation of the aggregated probability distributions, by assigning a numerical value, a score, based on P_G and on the event that materializes. Specifically, a scoring rule is a function that associates a value $S(P_G, A_k) \in (-\infty, \infty)$ for each event A_k in $\mathcal{A}$, when the forecasting probability distribution is P_G. $S(P_G, P)$ will denote the expected value of $S(P_G, A_k)$ under the true probability distribution P: $S(P_G, P) = \sum_{A_k \in \mathcal{A}} S(P_G, A_k) P(A_k)$. In the following, we will only consider strictly proper scoring rules, for which $S(P, P) \geq S(Q, P)$ for all probability distribution Q, where equality holds if and only if $Q = P$. Essentially, the highest score is achieved when the aggregated probability distribution is equal to the true distribution. Under mild conditions, if S is a proper scoring rule

$$d(Q, P) = S(P, P) - S(Q, P)$$

is the associated divergence function. It is non-negative and it is equal to 0 if and only if $Q = P$. Note that the order plays an important role in the definition of the divergence, which is thus not necessarily symmetrical. Gneiting and Raftery (2007) review some of the most important scoring rules for categorical variables. We mention two scoring rules which will be important for us in the rest of this work.

Definition 9 (Quadratic or Brier score) The quadratic or Brier score (Brier 1950), is defined by

$$S(P, A_k) = -\sum_{j=1}^{K} (\delta_{jk} - p_j)^2, \tag{34}$$

where $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ otherwise. The associated divergence is the squared Euclidean distance, $d(Q, P) = \sum_{k=1}^{K} (p_k - q_k)^2$. In this particular case, the divergence is symmetrical (and hence is a distance).

Definition 10 (Logarithmic score) The logarithmic score corresponds to

$$S(P, A_k) = \ln p_k. \tag{35}$$

The associated divergence is the Kullback–Leibler divergence, $d(Q, P) = \sum_{k=1}^{K} q_k \ln(p_k/q_k)$. The highest achievable score is $S(P, P) = \sum_{k=1}^{K} p_k \ln(p_k)$, which is nothing but the entropy of the distribution P.

Scoring rules can be used for estimating the parameters of a pooling operator according to the following general approach. Consider a pooling operator $P_{G,\theta}$ depending on some parameters θ and a proper scoring rule, tailored to the problem considered. The estimator $\hat{\theta} = \arg\max_\theta S(\theta)$, where $S(\theta)$ is the empirical score built from the data set, is the optimum score estimator. The logarithmic score is related to the maximum likelihood estimation, while the Brier score is related to calibration and sharpness, presented in the section after next.

5.3 Likelihood for Log-Linear Pooling Formulas

Maximum likelihood estimation is a special case of optimum score estimation, corresponding to maximizing the logarithmic score. We now describe the maximum likelihood approach for estimating the parameters for the pooling formula based on the product of probabilities, which is recalled in its most general form

$$P_G(A_k) = \frac{\nu(A_k) P_0(A_k)^{1-\sum_{i=1}^{n} w_i} \prod_{i=1}^{n} P_i(A_k)^{w_i}}{\sum_{k=1}^{K} \nu(A_k) P_0(A_k)^{1-\sum_{i=1}^{n} w_i} \prod_{i=1}^{n} P_i(A_k)^{w_i}}. \tag{36}$$

This pooling formula includes the log-linear pooling, when all $\nu(A_k) = 1$ and the Nu model (route 1), when all $w_i = 1$. In the binary case it also includes all pooling operators based on the product of odds.

The setting is the following. We denote $\mathbf{w} = (w_1, \ldots, w_n)$ and $\boldsymbol{\nu} = (\nu(A_1), \ldots, \nu(A_K))$ the parameters of the pooling formula and consider M repetitions of a random experiment. For each experiment $m = 1, \ldots, M$, the information $D_i^{(m)}$ is available, allowing to compute the individual conditional probabilities $P_i^{(m)}(A_k)$, and to estimate the aggregated probabilities $P_G^{(m)}(A_k)$ of occurrence of any event A_k. For the sake of lighter notations, we will denote $P_{i,k}^{(m)} = P_i^{(m)}(A_k)$, $P_{G,k}^{(m)} = P_G^{(m)}(A_k)$. In addition to the input information, we also have access to the real occurrence of one of the various possible outcomes. We denote it $Y_k^{(m)}$, $Y_k^{(m)} = 1$ if the outcome is A_k and $Y_k^{(m)} = 0$ otherwise. In the same spirit, we will further denote $\nu_k = \nu(A_k)$. The full log-likelihood is

$$L(\mathbf{w}, \boldsymbol{\nu}) = \ln \prod_{m=1}^{M} \prod_{k=1}^{K} \left(P_{G,k}^{(m)}\right)^{Y_k^{(m)}} = \sum_{m=1}^{M} \sum_{k=1}^{K} Y_k^{(m)} \ln P_{G,k}^{(m)}. \tag{37}$$

Notice that the log-likelihood is nothing but the empirical score of the data-set when applying the logarithmic scoring rule. Replacing $P_{G,k}^{(m)}$ in Eq. (37) by its expression Eq. (36) yields

$$\begin{aligned} L(\mathbf{w}, \boldsymbol{\nu}) = {} & \sum_{m=1}^{M} \sum_{k=1}^{K} Y_k^{(m)} \left\{ \ln \nu_k + \left(1 - \sum_{i=1}^{n} w_i\right) \ln P_{0,k} + \sum_{i=1}^{n} w_i \ln P_{i,k}^{(m)} \right\} \\ & - \sum_{m=1}^{M} \ln \left\{ \sum_{k=1}^{K} \nu_k P_{0,k}^{1-\sum_{i=1}^{n} w_i} \prod_{i=1}^{n} \left(P_{i,k}^{(m)}\right)^{w_i} \right\}. \end{aligned} \tag{38}$$

The parameters $\hat{\mathbf{w}}$ and $\hat{\boldsymbol{\nu}}$ maximizing the log-likelihood in Eq. (38) are the maximum likelihood (ML) estimators of $\mathbf{w}$ and $\boldsymbol{\nu}$. They are found by numerical methods. In theory, it is possible to follow a similar approach for the pooling formulas based on the multiplication of odds, but the expressions are lengthy, without bringing new insight. They are not shown here.

When fitting models, adding parameters leads to increased values of the log-likelihood. But doing so may lead to over-fitting. The Bayesian Information Criterion (BIC) introduced in Schwartz (1978) resolves this problem by adding a penalty term for the number of parameters in the model

$$\mathrm{BIC} = -2L + J \ln M, \tag{39}$$

where L is the log-likelihood, J the total number of parameters of the model considered and M the number of repetitions. Given any two estimated models, the model with the lower value of BIC is the one to be preferred. Lower BIC implies either fewer explanatory variables, better fit, or both. The models being compared need not be nested.

5.4 Calibration and Sharpness

Calibration and sharpness are two particular aspects of the pooling operators which can be used to evaluate their quality. We will follow Ranjan and Gneiting (2010) for

a brief introduction to these notions. We need the following set-up: One considers a random experiment, leading to random information $D_1, \ldots, D_n$ and thus random probabilities P_i. It is convenient to introduce $(Y_1, \ldots, Y_K)$ the random vector corresponding to the outcome, in which $Y_k = 1$ if the outcome is A_k and $Y_k = 0$ otherwise, hence $P(Y_k = 1) = P(A_k) = \mathbb{E}[Y_k]$.

Definition 11 (Calibration) The aggregated probability $P_G(A)$ is said to be calibrated if

$$P\big(Y_k \mid P_G(A_k)\big) = P_G(A_k), \quad k = 1, \ldots, K. \tag{40}$$

This definition is in accordance with economic, meteorological and statistical forecasting literature (Ranjan and Gneiting 2010). Sharpness refers to the concentration of the aggregated distribution. The more concentrated $P_G(\cdot)$ is, the sharper it is. Ranjan and Gneiting (2010) proved that linear opinion pools lack calibration, even though all conditional probabilities $P(A_k \mid D_i)$ are calibrated.

5.5 Calibration of Log-Linear Pooling

In the section presenting log-linear pooling, we showed that it is always possible to write the conditional probability $P(A \mid D_1, \ldots, D_n)$ with a log-linear formalism and that log-linear pooling is exact, thus calibrated, if there is one weight per combination $(A, D_1, \ldots, D_n)$. Log-linear pooling amounts to making the simplifying assumption $\ln P(D_i \mid A, D_{<i}) / \ln P(D_i \mid A) = w_i$ for all A, all D_i and all $D_{<i}$.

We are now ready to state our main result about calibration of log-linear pooling and the relationship between calibrated log-linear pooling (if it exists) and maximum likelihood.

Theorem 1 Suppose there exists a calibrated log-linear pooling. Then, asymptotically, it is the (generalized) log-linear pooling with parameters estimated from maximum likelihood.

Proof Let us first characterize the maximum likelihood solution. At the maximum, the derivatives of the log-likelihood Eq. (38) with respect to the parameters ν_k and w_i are equal to zero. Let us first consider the derivatives with respect to ν_k

$$\sum_{m=1}^{M} Y_k^{(m)} (\nu_k)^{-1} - \sum_{m=1}^{M} P_{0,k}^{1-\sum_{i=1}^{n} w_i} \prod_{i=1}^{n} \big(P_{i,k}^{(m)}\big)^{w_i} \Bigg/ \left(\sum_{l=1}^{K} \nu_l P_{0,l}^{1-\sum_{i=1}^{n} w_i} \big(P_{i,l}^{(m)}\big)^{w_i} \right) = 0. \tag{41}$$

Recognizing in the second term the probability $P_{G,k}^{(m)}$ in Eq. (41) can be better written

$$\sum_{m=1}^{M} Y_k^{(m)} = \sum_{m=1}^{M} P_{G,k}^{(m)}, \quad k = 1, \ldots, K. \tag{42}$$

Likewise, setting the derivatives with respect to w_i to zero leads after some simplifications to

$$\sum_{m=1}^{M}\sum_{k=1}^{K} Y_k^{(m)} \ln P_{i,k}^{(m)} = \sum_{m=1}^{M}\sum_{k=1}^{K} P_{G,k}^{(m)} \ln P_{i,k}^{(m)}, \quad i = 1, \ldots, n. \tag{43}$$

Let us multiply the left- and right-hand sides of Eq. (42) by the maximum likelihood estimates $\hat{v}_k$ and multiply the left- and right-hand sides of Eq. (43) by $\hat{w}_i$. Then the sum of the $K + n$ equations yields

$$\sum_{m=1}^{M}\sum_{k=1}^{K} Y_k^{(m)} \ln P_{\hat{G},k}^{(m)} = \sum_{m=1}^{M}\sum_{k=1}^{K} P_{G,k}^{(m)} \ln P_{\hat{G},k}^{(m)}, \tag{44}$$

where $P_{\hat{G},k}$ denotes the aggregated probabilities with parameters $(\hat{\mathbf{v}}, \hat{\mathbf{w}})$.

Suppose now that the M random experiments are simulated according to P and let us denote $\mathbf{Y} = (Y_1, \ldots, Y_K)$ and $\mathbf{P}_{\hat{G}} = (P_{\hat{G},1}, \ldots, P_{\hat{G},K})$. On the one hand, according to the law of large numbers, Eq. (44) tends in probability to

$$\mathbb{E}\big[\mathbf{Y}^t \ln \mathbf{P}_{\hat{G}}\big] = \mathbb{E}\big[\mathbf{P}_{\hat{G}}^t \ln \mathbf{P}_{\hat{G}}\big] \tag{45}$$

as $M \to \infty$. On the other hand, according to the conditional expectation theorem,

$$\mathbb{E}\big[\mathbf{Y}^t \ln \mathbf{P}_{\hat{G}}\big] = \mathbb{E}\big\{\mathbb{E}\big[\mathbf{Y}^t \ln \mathbf{P}_{\hat{G}} \mid \mathbf{P}_{\hat{G}}\big]\big\} = \mathbb{E}\big\{\mathbb{E}\big[\mathbf{Y}^t \mid \mathbf{P}_{\hat{G}}\big] \ln \mathbf{P}_{\hat{G}}\big\}. \tag{46}$$

If $\mathbf{P}_{\hat{G}}$ is calibrated, that is if $\mathbb{E}[\mathbf{Y}^t \mid \mathbf{P}_{\hat{G}}] = \mathbf{P}_{\hat{G}}$, it is clear that Eq. (45) is verified. Hence, calibration implies that the weights in $P_{\hat{G}}$ are a solution of the maximum likelihood. The theorem is thus proved because the maximum likelihood solution is unique. □

5.6 Empirical Measure of Calibration and Sharpness

Calibration and sharpness of the pooling formulas will be assessed on simulations. They arise naturally considering the Brier score. The empirical mean Brier score is defined as

$$\mathrm{BS} = \frac{1}{M}\left\{\sum_{k=1}^{K}\sum_{m=1}^{M}\big(P_G^{(m)}(A_k) - Y_k^{(m)}\big)^2\right\}, \tag{47}$$

where the superscript refers to the mth random experiment. Suppose that the probability $P_G(A_k)$ takes discrete values $f_k(j)$ (for example from 0 to 1 by step of 0.01), where $j = 1, \ldots, J$. Let $n(j)$ be the number of times $P_G(A_k) = f_k(j)$ and let $q_k(j)$ be the empirical event frequency for A_k when $P_G(A_k) = f_k(j)$. If the pooling formula is calibrated, one must have $q_k(i) = P(A_k \mid P_G(A_k) = f_k(i)) = f_k(i)$. Reliability diagrams plot the empirical event frequency against the aggregated probabilities (Bröcker and Smith 2007). Significant deviation from the diagonal must be interpreted as a lack of calibration.

The Brier score can be decomposed in the following way

$$\mathrm{BS} = \sum_{k=1}^{K}\left\{\frac{1}{M}\sum_{j=1}^{J} n_k(j)\big(f_k(j)-q_k(j)\big)^2\right\} - \sum_{k=1}^{K}\left\{\frac{1}{M}\sum_{j=1}^{J} n_k(j)\big(q_k(j)-\bar{q}_k\big)^2\right\} + \sum_{k=1}^{K}\bar{q}_k(1-\bar{q}_k), \tag{48}$$

where $\bar{q}_k = \frac{1}{M}\sum_{m=1}^{M} Y_k^{(m)}$ is the marginal event frequency.

The first term of the decomposition is the reliability term. It corresponds to the calibration. The lower this term is, the better the pooling formula is calibrated. The second term is a deviation around the re-calibrated probability. For a calibrated pooling formula, it corresponds to the sharpness; in this case, the higher the sharpness, the better. The last term depends on the observation alone; it is independent on the pooling formula. To address the performance of the aggregation methods, Ranjan and Gneiting (2010) proposed diagnostics based on the paradigm of maximizing the sharpness, subject to calibration. With this paradigm, optimal weights can be found using other scoring rules, such as the logarithmic scoring rule.

6 Simulation Study

We now conduct some simulations in order to compare the features of the different aggregation methods. We will first consider three cases with binary outcomes. In these cases, the Bordley/Tau formula is equivalent to a log-linear pooling to which we will refer. In the first case, we consider the aggregation of close to independent information for the prediction of the binary outcome. In this case, maximum entropy (equivalent to conditional independence) should perform reasonably well. In the second case, we will consider a truncated Gaussian model with correlation between three information to be aggregated. In the third case, we will consider a Boolean model with four information. We will then consider a pluri-Gaussian model in which there are three possible categories. For comparing the different aggregation methods we will use the Brier scores (Eq. (48)), BIC (Eq. (39)) and the reliability plots presented in Sect. 5. In some examples, we will have access to the analytical expressions of all conditional probabilities, to which the aggregation formula will be compared.

6.1 First Binary Case: Two Independent Sources of Information

For this first example, we adopt the same analytical setting as in Ranjan and Gneiting (2010), in which the Beta-transformed linear pooling is shown to be superior to linear pooling. The sources of information are two independent $(0, 1)$ Gaussian random variables D_1 and D_2. Let Φ denote the standard normal cumulative distribution function and define $p = \Phi(D_1 + D_2)$. Suppose Y is a Bernoulli random variable with success probability p, and consider the event $A = \{Y = 1\}$. Then

$$P(A \mid p) = P(Y = 1 \mid p) = \mathbb{E}[Y \mid p] = p, \tag{49}$$

and

$$P_1(A) = P(A \mid D_1) = \mathbb{E}[Y \mid D_1] = \mathbb{E}\big[\Phi(D_1 + D_2) \mid D_1\big]$$
$$= \Phi(D_1/\sqrt{3}) = P_2(A). \tag{50}$$

Note that $P(A)$, $P_1(A)$, and $P_2(A)$ are naturally calibrated. A training sample of size $M = 10{,}000$ is generated by simulating D_1, D_2, and Y. The prior is the constant value $p_0 = \mathbb{E}[p] = 1/2$. Table 4 presents the log-likelihood, the BIC and the Brier scores with their reliability and sharpness component for different pooling formula. The log-likelihood is computed according to

$$L = \sum_{m=1}^{M} Y^{(m)} \ln P_G^{(m)}(A) + \big(1 - Y^{(m)}\big) \ln\big(1 - P_G^{(m)}(A)\big).$$

For the sake of comparison, it is also computed for $P_1(A)$ and $P_{12}(A) = P(A \mid D_1, D_2)$. The model with the lowest Brier score, or with the lowest BIC should be preferred. In the case of binary events, remember that the log-linear pooling and Bordley/Tau model are equivalent, and that the Nu model is the generalized log-linear pooling formula with weights $w_i = 1$ for all $i = 1, \ldots, n$. Optimal weights were obtained with the maximum likelihood approach described in the previous section, with the additional constraints of equality $w_1 = w_2$ to account for the symmetry between D_1 and D_2. For the same reason, for the BLP parameters, we imposed $\alpha = \beta$. From Table 4, one can see that although P_1 being calibrated, it lacks sharpness. The exact conditional probability P_{12} is the best achievable prediction: it has the lowest log-likelihood, the lowest Brier score and the highest sharpness. Linear pooling has a lower Brier score than a single information, but at the price of a loss of calibration, and it lacks sharpness. It has the highest BIC among all models considered. As expected from Ranjan and Gneiting (2010), BLP is well calibrated with a high sharpness and the BIC decreases dramatically. Note that the parameter α is quite high, indicating that a strongly unimodal Beta density is necessary to calibrate the linear pooling. Among the multiplicative pooling formula, maximum entropy performs surprisingly well considering that it is parameter free. This is probably due to the fact that D_1 and D_2 are drawn independently. Introducing one parameter in the pooling formula, either for the Nu model or for the log-linear formula decreases the Brier score and the log-likelihood when they are estimated using maximum likelihood, while they can increase when the parameters are away from their optimal values (results not shown here). The log-linear formula leads to the best scores. In particular, it is almost perfectly calibrated. The generalized log-linear formula shows slightly better scores. The lowest BIC is obtained for the log-linear formula, indicating that the extra parameter in the generalized log-linear formula is not significant (note that its value is very close to 1).

6.2 Second Binary Case: Truncated Gaussian Model with Three Data Points

We consider now a truncated model with three data points similar to the construction described in Chugunova and Hu (2008). The prediction point s_0 is located at the

Table 4 First binary case: two sources of close to independent information

	Weight	Param.	−Log-lik	BIC	BS	REL	SH
P_1	–	–	5751.5		0.1973	0.0011	0.0538
P_{12}	–	–	4135.7		0.1352	0.0010	0.1158
Lin.	–	–	5208.7	10417.3	0.1705	0.0346	0.1141
BLP	–	$\alpha = 7.9$	4168.7	8346.5	0.1362	0.0011	0.1148
ME	–	–	5028.7	10057.3	0.1391	0.0045	0.1154
Nu	–	$\nu = 0.99$	4294.8	8598.9	0.1388	0.0043	0.1155
Log-lin.	1.46	–	4139.4	**8289.0**	**0.1353**	0.0010	**0.1156**
Gen. Log-lin.	1.46	$\nu = 0.99$	**4138.8**	8296.1	0.1354	**0.0008**	0.1154

Notes: Maximum likelihood weight and parameter, negative Log-likelihood, BIC, Brier score (BS), reliability term (REL), and sharpness (SH) for different pooling formulas: Linear pooling (Lin.), Beta-transformed Linear Pooling (BLP), Maximum Entropy (ME), Nu-model (Nu), Log-linear pooling, and Generalized Log-linear pooling

origin. The location of the three data points are defined by their distances (d_1, d_2, d_3) and their angles $(\theta_1, \theta_2, \theta_3)$ with the horizontal axis. We consider a random function $X(s)$ with an exponential covariance matrix; the range is set equal to 1 throughout. We define a threshold t and we are interested in the event $A = \{X(s_0) \leq t - a\}$ given the information $D_i = \{X(s_i) \leq t\}$, $i = 1, 2, 3$. Since we know the full model, all conditional probabilities can be numerically computed. A total of 10,000 thresholds t are drawn according to a $(0, 1)$ Gaussian random variable, and we set $a = 1.35$. A Gaussian random vector $(X(s_i))_{i=0,\ldots,3}$ is then simulated conditionally on $X(s_i) \leq t$, for $i = 1, 2, 3$. With this setting, we sample the whole range of probabilities for the event $A = \{X(s_0) \leq t - 1.35\}$, which on average will be close to 0.5. Figure 1 shows the histograms of the marginal and the conditional probabilities of A in one of the cases considered below. Clearly, the whole range of probabilities is sampled, allowing us a good calibration of the different pooling formulas.

6.2.1 Equal Distance: Symmetrical Information

In this experiment, the three data points s_1, s_2, s_3 are located on a circle of radius equal to the range, s_1 being on the horizontal axis, hence $\mathrm{Cor}(X(s_0), X(s_i)) = 0.37$ for all s_i. We thus impose an equal weight to each data. The angles between $s_{2,3}$ and s_1 are set equal to $\pi/3$. Results are presented in the top part of Table 5. The Brier scores, very close to each other, are not a good criterion for assessing the quality of a pooling formula. The log-likelihood shows a more contrasted behavior. Because of the symmetry, the linear pooling is equivalent to a single source of information. It is, by far, the poorest pooling method. A Beta transform improves the performances. The log-likelihood of BLP is intermediate between those obtained with P_1 and P_{12}. BLP has the lowest reliability term and a high sharpness. Among the multiplicative formula, the Maximum entropy is the poorest pooling method; the Log-linear pooling (symmetrical weights equal to 0.75) performs significantly better than the Nu model. It has the lowest Brier score. It shares with BLP the lowest reliability term

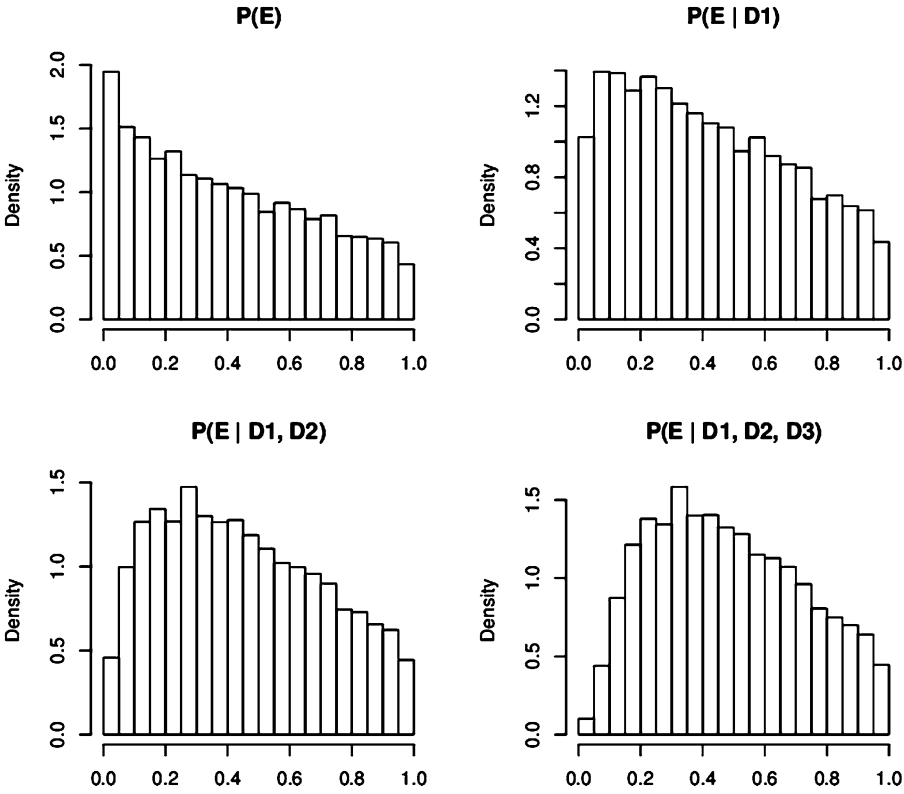

Fig. 1 Histograms of $P(A)$, $P(A \mid D_1)$, $P(A \mid D_1, D_2)$, and $P(A \mid D_1, D_2, D_3)$

and a slightly better sharpness. Its log-likelihood is however significantly higher. The generalized log-linear model achieves a likelihood equal to the true conditional distribution but higher BIC than log-linear pooling and a ν parameter very close to 1 are a strong indication of over-fitting.

6.2.2 Different Distance: Uneven Information

In this situation, the three points are at distances $(d_1, d_2, d_3) = (0.8, 1, 1.2)$. The distances being different, we will consider different weights for the three sources of information. For comparison purposes we will also include equal weight solutions. Results are shown in the bottom part of Table 5. The method with the best indicators related to the Brier score is the BLP. Interestingly, the optimal solution consists in having a 100 % weight for the closest source of information and null weights for all others. It is also the case for the log-linear pooling. When equal weights are imposed for the log-linear pooling formula, the Brier score and the log-likelihood remain almost identical; but because the number of free parameters decreases, the BIC reaches a minimum. In this example, the Brier score and the logarithmic score lead to different selected models. BLP has the lowest Brier score and reliability term and highest sharpness, while the log-linear formula have lower log-likelihood.

Table 5 Second binary case: truncated Gaussian model with three symmetrical sources of information

	Weight	Param.	−Log-lik	BIC	BS	REL	SH
	Same distance						
P_1	–	–	5782.2		0.1943	0.0019	0.0573
P_{12}	–	–	5686.8		0.1939	0.0006	0.0574
P_{123}	–	–	5650.0		0.1935	0.0007	0.0569
Lin.	–	–	5782.2	11564.4	0.1943	0.0019	0.0573
BLP	–	$\alpha = 0.67$	5704.7	11418.7	0.1932	**0.0006**	0.0570
ME	–	–	5720.1	11440.2	0.1974	0.0042	0.0564
Nu	–	$\nu = 0.93$	5695.9	11391.8	0.1952	0.0021	0.0566
Log-Lin.	0.75	–	5651.4	**11312.0**	**0.1931**	**0.0006**	0.0571
Gen. Log-Lin.	0.71	$\nu = 1.03$	**5650.0**	11318.3	0.1937	0.0008	0.0568
	Different distances						
P_1	–	–	5786.6		0.1943	0.0022	0.0575
P_{12}	–	–	5730.8		0.1927	0.0007	0.0577
P_{123}	–	–	5641.4		0.1928	0.0009	0.0579
Lin.eq	(1/3, 1/3, 1/3)	–	5757.2	11514.4	0.1940	0.0018	0.0575
Lin.	(1, 0, 0)	–	5727.2	11482.0	0.1935	0.0015	0.0577
BLP	(1, 0, 0)	$\alpha = 0.66$	5680.5	11397.8	**0.1921**	**0.0004**	**0.0580**
ME	–	–	5727.7	11455.4	0.1972	0.0046	0.0571
Nu	–	$\nu = 0.92$	5791.4	11592.0	0.1950	0.0023	0.0570
Log-Lin.-eq.	(0.72, 0.72, 0.72)	–	5646.1	**11301.4**	0.1928	0.0006	0.0576
Log-Lin.	(1.87, 0, 0)	–	5645.3	11318.3	0.1928	0.0007	0.0576
Gen. Log-Lin.	(1.28, 0.53, 0)	$\nu = 1.04$	5643.1	11323.0	0.1930	0.0010	0.0576

Notes. Same abbreviations as in Table 4. In addition: Lin-eq stands for linear pooling with equal weights; Log-Lin-eq is a log-linear formula with equal weights

6.3 Third Binary Case: Boolean Model with Four Data Points

We simulated in the unit cube a Boolean model of spheres with radius $r = 0.07$. Let us denote $X(s)$ its void indicator function and λ the mean number of spheres per unit volume. Then it is well known (Lantuéjoul 2002) that the void probability is $q = P(X(s) = 1) = \exp\{-\lambda V\}$, with $V = 4\pi r^3/3$. The prediction point s_0 is randomly located in the unit cube and the information points s_i, $i = 1, \ldots, 4$ are randomly located around s_0: two points are in the horizontal plane at a x and y distances uniformly drawn between 0.004 and 0.02, and two other points are similarly located in a vertical plane. The conditional probabilities are easily computed in this model

$$P\big(X(s_0) = 1 \mid X(s_i) = 1\big) = k_r(h),$$

$$P\big(X(s_0) = 1 \mid X(s_i) = 0\big) = \frac{q}{1-q}\big(1 - k_r(h)\big)$$

Table 6 Binary case: Boolean model with four symmetrical data points

	Weights	Param.	−Log-lik	BIC	BS	REL	SH
P_0	–	–	29859.1	59718.2	0.1981	0.0155	0.0479
P_i	–	–	16042.0	32084.0	0.0892	0.0120	0.1532
Lin.	$\simeq 0.25$	–	14443.3	28929.9	0.0774	0.0206	0.1736
BLP	$\simeq 0.25$	(3.64, 4.91)	9690.4	19445.7	0.0575	**0.0008**	0.1737
ME	–	–	7497.3	14994.6	0.0433	0.0019	0.1889
Nu	–	$\nu = 0.96$	7491.3	14993.4	0.0432	0.0018	0.1890
Log-Lin	$\simeq 0.80$	–	7178.0	**14399.3**	**0.0416**	0.0010	0.1897
Gen. Log-Lin.	$\simeq 0.79$	$\nu = 1.04$	**7172.9**	14399.9	0.0417	0.0011	**0.1898**

Notes. Abbreviations as in Table 4. In addition, BLP(2) is the Beta-transformed Linear Pooling with $\alpha \neq \beta$

with $k_r(h) = \exp\{\lambda V[-1.5\|h\|/r + 0.5(\|h\|/r)^3]\}$. The parameter λ is also made random such that q is uniformly distributed in $[0.05, 0.95]$, thus allowing a good measure of the calibration. We performed a total of 50,000 repetitions. Results are presented in Table 6. Since there is a symmetry between the data points, the optimal parameters were equal (up to small statistical fluctuations) for the four data points.

On this example the linear pooling of the four data leads to scores only slightly better than considering only one data point. As usual, BLP is a real improvement. The prediction is calibrated and the Brier score is improved. Perhaps surprisingly, the sharpness is only slightly improved, as compared to the linear pooling. Even more surprising is the quite good performance of maximum entropy. This is perhaps due to the Markovian nature of the Boolean model for which conditional expectation is not a poor approximation. The Nu model performs only marginally better than Maximum entropy. BICs are very close and $\nu \simeq 1$, indicating that this parameter is not really necessary. Log-linear pooling leads to the lowest Brier score. It is almost perfectly calibrated and very sharp. Generalized log-linear pooling has slightly better scores, but at the cost of one additional parameter. Hence, BIC is the lowest for log-linear pooling.

The calibration curve of the four pooling formulas are shown in Fig. 2. On these plots, deviation from the first diagonal indicates a lack of calibration. It is visible for the linear pooling and the maximum entropy. BLP and log-linear pooling have close to calibrated curves. All in all, in this example, log-linear pooling presents the best performances which are significantly better than those of BLP.

6.4 Trinary Events

We keep the same geometrical configuration as in Sect. 6.2. The trinary events are defined by means of two independent $(0, 1)$ Gaussian random functions $U(s)$ and $V(s)$. The category $C(s)$ will depend on $U(s)$ and $V(s)$ according to the following

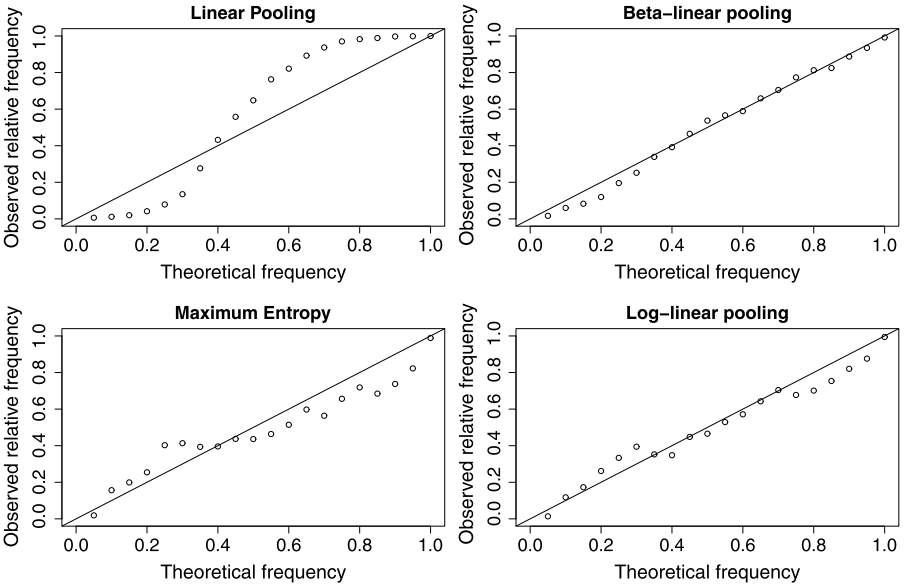

Fig. 2 Calibration curve for the Boolean simulation. BLP and log-linear formulas are computed with their optimal parameters. Deviation from first diagonal indicates lack of calibration

scheme

$$I(s) = 1 \quad \text{if } U(s) \leq t \text{ and } V(s) \leq t, \tag{51}$$

$$I(s) = 2 \quad \text{if } U(s) > t \text{ and } U(s) > V(s), \tag{52}$$

$$I(s) = 3 \quad \text{if } V(s) > t \text{ and } U(s) \leq V(s), \tag{53}$$

where t is a threshold. The marginal probabilities are the following. First, $P(I = k) = P(U \leq t)P(V \leq t) = G^2(t)$ where $G(\cdot)$ denotes the cpf of a (0,1) Gaussian random variable, and $g(t)$ its density. Then symmetry imposes $P(I = 2) = P(I = 3)$, which leads to $P(I = 2) = 0.5[1 - P(I = 1)] = 0.5[1 - G^2(t)]$. The conditional probabilities $P(I(s_0) \mid I(s_i))$ are detailed in Appendix B.

The thresholds t are drawn such that the probability of category 1 is uniformly sampled between 0.1 and 0.8. A total of 20,000 random samples were drawn. It should be remembered that for trinary experiments, the equivalence between methods based on the product of probabilities and those based on the product of odds is lost. It is thus necessary to distinguish between these methods. The Nu-1 route corresponds to a product of probabilities updated by a likelihood on the events, while the Nu-2 route corresponds to a product of odds updated by odds. As expected, linear pooling does not perform well (Table 7). We had some difficulties with BLP, which does not have a straightforward generalization to non-binary events. We chose to pool the three events together. The consequence is that the same parameters were applied to the three categories, which is certainly not optimal. A generalized version, with one parameter per category would probably lead to better performances. ME and $\nu(1)$ do not perform well at all. The reason being that they lead to probabilities very close

Table 7 Trinary event

	−Log-lik	BIC	BS	REL	SH
Lin.	24123.8	24123.9	0.2219	0.0271	0.0262
BLP	21517.9	43045.8	0.2187	0.0218	0.0241
ME	44358.3	88716.6	0.2736	0.0780	0.0254
Nu-1	44278.0	88575.9	0.2770	0.0812	0.0253
Log-Lin.	18744.4	37518.6	0.1890	0.0025	0.0345
Gen. Log-Lin.	**18554.1**	**37157.8**	**0.1868**	**0.0004**	**0.0351**
Bordley/Tau	18846.1	37721.9	0.1904	0.0019	0.0325
Nu-2	21732.6	43494.8	0.2242	0.0300	0.0269
GLO	18733.2	37525.8	0.1896	0.0011	0.0326

Notes. Same abbreviations as in Table 4 and in addition: Nu-1 and Nu-2, two possible routes for generalizing the Nu model for $K > 2$; GLO, generalized log-linear pooling of odds

to 0 or 1, thus strongly penalizing the scores when the prediction is wrong. Methods based on the product of probabilities tend to perform better than the corresponding ones based on the product of odds. The optimal method is the Generalized log-linear pooling formula (Table 7). Unlike the binary case, the extra parameters of this model, as compared to log-linear pooling, offers the flexibility needed to fit to non-binary outcomes. The generalized log-linear pooling of odds is a model, not yet proposed in the literature, that combines w_i, the weights on the sources of information with the parameters $\nu(A)$. It performs slightly better than the Bordley/Tau model, but it is outperformed by the generalized log linear model on probabilities.

7 Discussion and Conclusions

We reviewed a majority of methods proposed in the literature for aggregating probability distributions with a focus on their mathematical properties. By doing so, we were able to better understand the relationships between these methods. We were able to show that conditional independence is equivalent to a particular maximum entropy principle. It is also equivalent to a Nu model with $\nu(A) = 1$ for all $A \in \mathcal{A}$ and to a log-linear formula with $w_i = 1$ for all sources of information. We showed that binary experiments must be distinguished from non-binary ones. In the latter case, the equivalence between Bordley/Tau models (based on odds) and log-linear pooling (based on probabilities) is lost. For this case also, there are two different ways for generalizing the Nu model. The comparison study, illustrated in Table 2, leads us to the definition of one model that has not yet been proposed in the literature: this model would combine weights w_i and $\nu(A)$ on odds. It would be at the same time a generalization of the Tau model and a generalization of the Nu-2 model. This could be called a generalized log-linear combination of odds.

When training are available, maximum likelihood provides an efficient method for estimating the parameters of any chosen model. Our main result is Theorem 1, which states that for (generalized) log-linear poolings, calibration implies parameters

estimated with maximum likelihood. The converse is not true in all generality, but it is verified for some probability models. All simulated examples have shown that log-linear pooling formula with parameters estimated with ML are very close to be calibrated. On one example, log-linear pooling achieved better calibration than the Beta-transformed linear pooling proposed in Ranjan and Gneiting (2010).

On simulations, we were able to show that quadratic and logarithmic scores (the Brier score and the likelihood, or its penalized version BIC) are efficient tools for determining the models leading to the best forecasts. They usually increase or decrease together. However, sometimes they do not lead to the same selected model. Maximum likelihood is related to the logarithmic score and to the Kullback–Leibler (KL) divergence. Maximizing the likelihood, which is equivalent to minimizing the KL divergence to the true unknown conditional probability, does not always lead to the lowest Brier score. But in this case, it is very close to the minimum. In particular, the reliability term REL will always be very close to 0 for the (generalized) log-linear pooling with parameters estimated with maximum likelihood.

A first conclusion of this study is that linear methods should not be used alone for aggregating probability distribution. They can be used if re-calibrated with a Beta transformation whose parameters must be estimated, but methods based on product of probabilities should be preferred. Simulations presented here and other ones not presented here (Comunian 2010) have shown that among methods based on multiplication, the Nu model performs generally worst than any other method. This can be explained from the equations: the parameters $\nu(A)$ act as a likelihood on the events regardless of the information at hand, while other methods provide a transformation of the conditional probabilities which accounts for the redundancy or the interaction between information. This study also indicated that methods based on product of odds (Tau model) are not to be recommended. For binary events, they are equivalent to those based on product of probabilities. For non-binary events they usually perform less well.

The main conclusion of this study is thus the following: for aggregating probability distributions, methods based on product of probabilities (in other words linear combinations of log-probabilities) should be preferred. First, they are easy to implement and to understand. Second, their parameters are easy to estimate using maximum likelihood. According to Theorem 1, if a log-linear pooling formula is calibrated, it is the solution of the maximum likelihood estimation. On all simulations performed so far, we found that log-linear pooling formulas lead to excellent predictions, (slightly) better than or equal to BLP predictions. If no data is available, the parameter free maximum entropy solution is an acceptable approximation. This has profound implications on the practice of spatial prediction and simulation of indicator functions. It implies that the kriging paradigm based on linear combinations of bivariate probabilities and its sequential indicator simulation (SIS) counterpart should probably be replaced by a different paradigm based on the product of probabilities. Allard et al. (2011) arrived at a somehow similar conclusion. We hope that this contribution, together with those cited in this work, will help geoscientists to adopt this new paradigm.

Acknowledgements Funding for A. Comunian and P. Renard was mainly provided by the Swiss National Science foundation (Grants PP002-106557 and PP002-124979) and the Swiss Confederation's In-

novation Promotion Agency (CTI Project No. 8836.1 PFES-ES) A. Comunian was partially supported by the Australian Research Council and the National Water Commission.

Appendix A: Maximum Entropy

Let us define $Q(A, D_0, D_1, \ldots, D_n)$ the joint probability distribution maximizing its entropy $H(Q) = -\sum_{A\in\mathcal{A}} Q(D_0, D_1, \ldots, D_n)(A) \ln Q(D_0, D_1, \ldots, D_n)(A)$ subject to the following constraints.

1. $Q(A, D_0) = Q(A \mid D_0)Q(D_0) \propto P_0(A)$, for all $A \in \mathcal{A}$.
2. $Q(A, D_0, D_i) = Q(A \mid D_i)Q(D_i)Q(D_0) \propto P_i(A)$, for all $A \in \mathcal{A}$ and all $i = 1, \ldots, n$.

We will first show that

$$Q(A, D_0, D_1, \ldots, D_n) \propto P_0(A)^{1-n} \prod_{i=1}^{n} P_i(A),$$

from which the conditional probability

$$\begin{aligned} P_G(P_0, P_1, \ldots, P_n) &= \frac{Q(A, D_0, D_1, \ldots, D_n)}{\sum_A Q(A, D_0, D_1, \ldots, D_n)} \\ &= \frac{P_0(A)^{1-n} \prod_{i=1}^{n} P_i(A)}{\sum_A P_0(A)^{1-n} \prod_{i=1}^{n} P_i(A)} \end{aligned}$$

is immediately derived. For ease of notation, we will use $\sum_A$ as a short notation for $\sum_{A\in\mathcal{A}}$.

Proof The adequate approach is to use the Lagrange multiplier technique on the objective function

$$\begin{aligned} J = &- \sum_A Q(A, D_0, D_1, \ldots, D_n) \ln Q(A, D_0, D_1, \ldots, D_n) \\ &+ \sum_A \mu_A \big\{ Q(A, D_0) - a P_0(A) \big\} \\ &+ \sum_{i=1}^{n} \sum_A \lambda_{A,i} \big\{ Q(A, D_0, D_i) - b_i P_i(A) \big\}, \end{aligned}$$

where μ_A and $\lambda_{A,i}$ are Lagrange multipliers. For finding the solution Q optimizing the constrained problem, we set all partial derivatives to 0. This leads to the system of equations

$$\ln Q(A, D_0, D_1, \ldots, D_n) = -1 + \sum_A \mu_A + \sum_A \sum_{i=1}^{n} \lambda_{A,i}, \tag{54}$$

$$Q(A, D_0) = a P_0(A), \tag{55}$$

$$Q(A, D_0, D_i) = b_i P_i(A), \quad \text{for } i = 1, \ldots, n. \tag{56}$$

From Eqs. (54) and (55), we get

$$Q(A, D_0) = e^{-1} \prod_A e^{\mu_A} \propto P_0(A).$$

Similarly, from Eqs. (54) and (56), we get

$$Q(A, D_0, D_i) = Q(A, D_0) \prod_A e^{\lambda_{A,i}} \propto P_i(A), \quad \text{for } i = 1, \ldots, n,$$

from which we find

$$\prod_A e^{\lambda_{A,i}} \propto P_i(A)/P_0(A), \quad \text{for } i = 1, \ldots, n.$$

Plugging this in Eq. (54) yields

$$Q(A, D_0, D_1, \ldots, D_n) \propto P_0(A) \prod_{i=1}^{n} \frac{P_i(A)}{P_0(A)}.$$

Hence,

$$\begin{aligned} P_G(P_0, P_1, \ldots, P_n)(A) &= \frac{Q(A, D_0, D_1, \ldots, D_n)}{\sum_A Q(A, D_0, D_1, \ldots, D_n)} \\ &= \frac{P_0(A)^{1-n} \prod_{i=1}^n P_i(A)}{\sum_A P_0(F)^{1-n} \prod_{i=1}^n P_i(A)}. \end{aligned}$$

□

Appendix B: Conditional Probabilities for the Trinary Event Example

1. Let us first compute the conditional probability

$$\begin{aligned} P\big(I(s') = 1 \mid I(s) = 1\big) &= P\big(U' \le t, V' \le t \mid U \le t, V \le t\big) \\ &= P\big(U' \le t, V' \le t, U \le t, V \le t\big)/P(U \le t, V \le t) \\ &= P\big(U' \le t, U \le t\big) P\big(V' \le t, V \le t\big)/\big[P(U \le t) P(V \le t)\big] \\ &= G_2^2(t, t; \rho)/G^2(t), \end{aligned}$$

where $G_2^2(t, t; \rho)$ is the bivariate cpf of a $(0, 1)$ bi-Gaussian random vector with correlation ρ. For symmetry reasons, one has $P(I(s') = 2 \mid I(s) = 1) = P(I(s') = 3 \mid I(s) = 1)$, from which it follows immediately

$$P\big(I(s')=2 \mid I(s)=1\big) = P\big(I(s')=3 \mid I(s)=1\big) = 0.5\big[1 - G_2^2(t,t;\rho)/G^2(t)\big].$$

2. We consider now

$$\begin{aligned} P\big(I\big(s'\big)=1 \mid I(s)=2\big) &= P\big(I(s)=2 \mid I\big(s'\big)=1\big)\frac{P(I(s')=1)}{P(I(s)=2)} \\ &= 0.5\left[1 - \frac{G_2^2(t,t;\rho)}{G^2(t)}\right]\frac{G^2(t)}{0.5[1-G^2(t)]} \\ &= \frac{G^2(t) - G_2^2(t,t;\rho)}{1-G^2(t)}. \end{aligned}$$

3. The picture is slightly more complicated for $P(I(s')=2 \mid I(s)=2)$

$$\begin{aligned} &P\big(I\big(s'\big)=2 \mid I(s)=2\big) \\ &\quad = P\big(U'>t, U'>V', U>t, U>V\big)P\big(I(s)=2\big) \\ &\quad = 0.5\big[1-G^2(t)\big]\int_t^{+\infty}\int_t^{+\infty} g_2\big(u,u';\rho\big)\int_{-\infty}^{u'}\int_{-\infty}^{u} g_2\big(v,v';\rho\big)dvdv'dudu' \\ &\quad = 0.5\big[1-G^2(t)\big]\int_t^{+\infty}\int_t^{+\infty} g_2\big(u,u';\rho\big)G_2\big(u,u';\rho\big)dudu'. \end{aligned}$$

There is no closed-form expression for the double integral which must be evaluated numerically. Then $P(I(s')=3 \mid I(s)=2)$ is computed as the complement to 1.

4. The conditional probabilities of $I(s')$ given that $I(s)=3$ are then obtained by symmetry.

References

Allard D, D'Or D, Froidevaux R (2011) An efficient maximum entropy approach for categorical variable prediction. Eur J Soil Sci 62(3):381–393

Bacharach M (1979) Normal Bayesian dialogues. J Am Stat Assoc 74:837–846

Benediktsson J, Swain P (1992) Consensus theoretic classification methods. IEEE Trans Syst Man Cybern 22:688–704

Bordley RF (1982) A multiplicative formula for aggregating probability assessments. Manag Sci 28:1137–1148

Brier G (1950) Verification of forecasts expressed in terms of probability. Mon Weather Rev 78:1–3

Bröcker J, Smith LA (2007) Increasing the reliability of reliability diagrams. Weather Forecast 22:651–661

Cao G, Kyriakidis P, Goodchild M (2009) Prediction and simulation in categorical fields: a transition probability combination approach. In: Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems, GIS'09. ACM, New York, pp 496–499

Christakos G (1990) A Bayesian/maximum-entropy view to the spatial estimation problem. Math Geol 22:763–777

Chugunova T, Hu L (2008) An assessment of the tau model for integrating auxiliary information. In: Ortiz JM, Emery X (eds) VIII international geostatistics congress, Geostats 2008. Gecamin, Santiago, pp 339–348

Clemen RT, Winkler RL (1999) Combining probability distributions from experts in risk analysis. Risk Anal 19:187–203
Clemen RT, Winkler W (2007) Aggregating probability distributions. In: Edwards W, Miles RF, von Winterfeldt D (eds) Advances in decision analysis. Cambridge University Press, Cambridge, pp 154–176
Comunian A (2010) Probability aggregation methods and multiple-point statistics for 3D modeling of aquifer heterogeneity from 2D training images. PhD thesis, University of Neuchâtel, Switzerland
Comunian A, Renard P, Straubhaar J (2011) 3D multiple-point statistics simulation using 2D training images. Comput Geosci 40:49–65
Cover TM, Thomas JA (2006) Elements of information theory, 2nd edn. Wiley, New York
Dietrich F (2010) Bayesian group belief. Soc Choice Welf 35:595–626
Genest C (1984) Pooling operators with the marginalization property. Can J Stat 12:153–165
Genest C, Wagner CG (1987) Further evidence against independence preservation in expert judgement synthesis. Aequ Math 32:74–86
Genest C, Zidek JV (1986) Combining probability distributions: a critique and an annotated bibliography. Stat Sci 1:114–148
Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102:359–378
Heskes T (1998) Selecting weighting factors in logarithmic opinion pools. In: Jordan M, Kearns M, Solla S (eds) Advances in neural information processing systems, vol 10. MIT Press, Cambridge, pp 266–272
Journel A (2002) Combining knowledge from diverse sources: an alternative to traditional data independence hypotheses. Math Geol 34:573–596
Krishnan S (2008) The Tau model for data redundancy and information combination in earth sciences: theory and application. Math Geosci 40:705–727
Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22:76–86
Lantuéjoul C (2002) Geostatistical simulations. Springer, Berlin
Lehrer K, Wagner C (1983) Probability amalgamation and the independence issue: a reply to Laddaga. Synthese 55:339–346
Mariethoz G, Renard P, Froidevaux R (2009) Integrating collocated auxiliary parameters in geostatistical simulations using joint probability distributions and probability aggregation. Water Resour Res 45(W08421):1–13
Okabe H, Blunt MJ (2004) Prediction of permeability for porous media reconstructed using multiple-point statistics. Phys Rev E 70(6):066135
Okabe H, Blunt MJ (2007) Pore space reconstruction of vuggy carbonates using microtomography and multiple-point statistics. Water Resour Res 43(W12S02):1–5
Polyakova EI, Journel AG (2007) The nu expression for probabilistic data integration. Math Geol 39:715–733
Ranjan R, Gneiting T (2010) Combining probability forecasts. J R Stat Soc B 72:71–91
Schwartz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464
Stone M (1961) The opinion pool. Ann Math Stat 32:1339–1348
Strebelle S, Payrazyan K, Caers J (2003) Modeling of a deepwater turbidite reservoir conditional to seismic data using principal component analysis and multiple-point geostatistics. SPE J 8:227–235
Tarantola A (2005) Inverse problem theory. Society for Industrial and Applied Mathematics, Philadelphia
Tarantola A, Valette B (1982) Inverse problems = quest for information. J Geophys 50:159–170
Wagner C (1984) Aggregating subjective probabilities: some limitative theorems. Notre Dame J Form Log 25:233–240
Winkler RL (1968) The consensus of subjective probability distributions. Manag Sci 15:B61–B75