Analysis of Local or Asymmetric Dependencies in Contingency Tables using the Imprecise Dirichlet Model

J.-M. BERNARD

Université de Paris 5 & CNRS, France

Abstract

We consider the statistical problem of analyzing the association between two categorical variables from cross-classified data. The focus is put on measures which enable one to study the dependencies at a local level and to assess whether the data support some more or less strong association model. Statistical inference is envisaged using an imprecise Dirichlet model.

Keywords

association, logical model, directional association model, statistical inference, upper and lower probabilities, IDM, prior ignorance, Bayesian inference

1 Introduction

1.1 The problem of association in contingency tables

The problem of measuring association in two-way contingency tables arising from cross-classifications has a long tradition in statistical research (see, *e.g.*, the numerous association measures reviewed by Goodman & Kruskal [6]). Though every one agrees on the meaning of "independence", the opposite notion of "complete association" is felt more ambiguous, because there are several directions in which the data may depart from independence. For the simplest case of 2×2 tables, Kendall & Stuart [9] make the distinction between "complete association" (one empty cell) and "absolute association" (two empty cells on either diagonal of the table). Although such distinctions are occasionally mentioned in the literature, most statistical research appears to have focused on proposing global measures of association.

The motivation behind this article arise from two (apparently) independent goals. The first one is to provide a local and/or asymmetric approach to the analysis of contingency tables and to define well-suited descriptive indices for that purpose. The second one is to build the inferential part of the analysis on a generalization of the Bayesian framework, the *imprecise Dirichlet model (IDM)*. Let us comment on these two aspects.

1.2 Analysis of local/asymmetric dependencies: two examples

The first aim of this article is to address two related types of statistical issues, that we shall illustrate by two psychological examples.

Example 1 (Stages data, Logical model) Jamison [8] studied several cognitive tasks related to the Piaget's stage concept. Table 1 gives the levels attained by a group of children in two tasks, A and B, with three levels each. One model predicts that attaining a given level in task A is a prerequisite for attaining the same level in task B, i.e., predicts that cells a1b2, a1b3 and a2b3 should be empty. This model can also be expressed as the logical expression $\mathcal{M} = [b3 \Longrightarrow a3 \land b2 \Longrightarrow (a2 \lor a3)]$. The issue here is to assess whether a conclusion of quasiagreement of the data with model \mathcal{M} , can be reached or not.

Table 1: "Stages" example. Observed counts **x** for n = 101 children cross-classified according to their performance level in Seriation of lengths (*A*) and Inclusion of lengths (*B*), from [8, p. 248]. For each task, children were classified as "preoperational" (*a*1 and *b*1), "transitional" (*a*2 and *b*2) or "operational" (*a*3 or *b*3). Shaded cells are error cells associated to the logical model $\mathcal{M} = (b3 \Longrightarrow a3 \land b2 \Longrightarrow (a2 \lor a3))$.

	<i>b</i> 1	<i>b</i> 2	<i>b</i> 3
<i>a</i> 1	14	0	0
<i>a</i> 2	15	5	2
<i>a</i> 3	19	20	26

Example 2 (Dyad data, Directional association model) Another type of problem is the study of local dependencies within an $A \times B$ table, which aims at showing that a specified group of cells is over- or under-represented. For example, Danis et al. [5] analyzed data about adult-child verbal interactions in a situation of book reading. Each statement produced by either actor was categorized into one of four levels of increasing complexity. Table 2(left) gives one transition matrix (child statement followed by adult statement) for one dyad. One hypothesis of interest here is that some regions of Table 2(left) should be over- or underrepresented according to the pattern shown in Table 2(right): over-representation of statements of the adult at the same level as the child's (denoted "+"), moderate under-representation of statements at an higher level (denoted "--").

The two types of questions raised by these examples, either asymmetric and expressed in terms of quasi-agreement with a logical model, or local and expressed in terms of over-/under-representation, can be answered using indices of

Table 2: Dyad data. Counts of transitions from the child's statement level (A) to the adult statement level (B) for one dyad (left). Expected pattern of over-representations (+) and under-representations (- or --) (right). Levels correspond to increasing cognitive complexity: "perceptual identification" (a1 and b1), "perceptual relationship" (a2 and b2), "displaced reference" (a3 and b3), and "inferential statement" (a4 and b4); categories a0 and b0 indicate cases in which one of the actors did not speak.

	b0	b1	b2	<i>b</i> 3	b4		b0	b1	b2	<i>b</i> 3	b4
<i>a</i> 0	0	25	2	8	0	<i>a</i> 0					
<i>a</i> 1	6	27	1	3	2	<i>a</i> 1		+	_	_	
<i>a</i> 2	2	0	2	0	0	<i>a</i> 2			+	—	—
<i>a</i> 3	13	0	0	20	2	а3				+	-
<i>a</i> 4	0	2	0	0	0	<i>a</i> 4					+

the same family. Hildebrand *et al.* [7], beside the main trend of research sketched previously, proposed a general index, named *Del*, which measures the degree of agreement of cross-classified data to a specified logical model. The building block of the *Del* index is what [10] call the *association rate* between modalities. Our method will be based on these two indices.

1.3 Inference for local/asymmetric analyses

Several difficulties arise when it comes to making inferences about these indices. The inferential methods that were initially proposed were based on the frequentist framework, and, due to the presence of nuisance parameters, relied on asymptotic arguments (see *e.g.*, [7, Chp. 6]), so that the validity conditions of these methods are satisfied neither for small samples, nor for extreme data sets in which some cells are empty or nearly so. These difficulties come in addition to some fundamental shortcomings of the frequentist methods, and, in particular, the fact that they do not obey the *likelihood principle (LP)*.

The Bayesian approach to inference answers most of these problems. However, it also encounters some difficulties when one wants to make inferences from a prior state of ignorance. None of the various solutions which were proposed for that goal simultaneously satisfies some general desirable principles (see [11]), *i.e.*, the LP, and the *representation invariance principle (RIP)* (invariance with respect to how categories are distinguished).

A generalization of the Bayesian framework, involving *imprecise probabilities*, allows one to overcome most, if not all, of the difficulties of the Bayesian approach, while keeping its attractive features (see [11]). In particular, Walley [12] proposed a new method of inference for categorical data based on the *imprecise Dirichlet model (IDM)*. In the IDM, prior uncertainty about the cells' true

48

frequencies is described by a set of Dirichlet priors, each of which being updated into a Dirichlet posterior using Bayes' theorem. Posterior uncertainty is described by the set of these Dirichlet posteriors. The IDM has several desirable properties as a model for making inferences from a prior state of near ignorance. Firstly, it satisfies both the LP and the RIP. Secondly, the IDM distinguishes between a relative lack of information (high imprecision) and a more substantial state of knowledge (low imprecision). The IDM can also be viewed as a method for making robust inferences.

Our purpose here is to apply the IDM to the problem of studying the association in contingency tables. This article contains relatively few new results about the IDM itself, but we think it is important to face the IDM with several types of applications and data sets, in order to develop more insights about its properties and the scope of its application.

This article is structured as follows. Section 2 defines local or asymmetric association measures. Sections 3 and 4 review the usual Bayesian Dirichlet models and the IDM, respectively. Our main contribution is the study of inferences about association measures from the IDM which is presented in Sections 5 and 6.

2 Descriptive analysis: defining relevant indices

Consider a data set of size *n* categorized in *K* categories, with observed counts $\mathbf{x} = (x_1, \ldots, x_K)$, with $n = \sum_k x_k$. The observed (relative) frequencies are denoted $\mathbf{f} = (f_1, \ldots, f_K)$, with $f_k = x_k/n$. The data \mathbf{x} will be considered as a sample from a larger population, characterized by the parent or true frequencies $\mathbf{\theta} = (\theta_1, \ldots, \theta_K)$, which are the population counterparts of \mathbf{f} . Both \mathbf{f} and $\mathbf{\theta}$ belong to the *K*-dimensional unit simplex S(1, K). Throughout this paper, the generic expression "association model" (or simply "model") denotes some summary statement about a frequency-vector, either \mathbf{f} or $\mathbf{\theta}$, *i.e.*, a statement saying that it belongs to some subset $\mathcal{R} \subset S(1, K)$. The qualifiers "descriptive" and "inductive" are used for models bearing on \mathbf{f} and $\mathbf{\theta}$ respectively. At the descriptive level, a model is either true or false, whereas, at the inductive level, the model's truth can only be assessed with some probability.

In this section, we define various indices in terms of which the association models considered in this paper will be defined. Here, these indices are defined as functions of f, but each one has its inductive counterpart as a function of θ . The problem of making inferences about parameters θ (and indices derived from them) will be envisaged in later sections.

2.1 Notation and preliminary definitions

The *K* categories are obtained here as combinations of modalities of the *A* and *B* variables, so we shall use more specific notations: ab or (a,b) for a cell of the

contingency table, x_{ab} for its observed count, f_{ab} for its observed frequency; we note $f_a = \sum_b f_{ab}$ and $f_b = \sum_a f_{ab}$ the marginal frequencies of categories $a \in A$ or $b \in B$, and $\widehat{f_{ab}} = f_a f_b$ the *product-frequency* of cell ab.¹

Definition 1 (Local independence) *There is* local independence *between modalities a and b, noted a* $\perp \perp b$ *, whenever* $f_{ab} = \widehat{f_{ab}}$.

Definition 2 (Global independence) *There is* global independence *between variables A and B, noted A* $\perp \perp B$, *whenever* $\forall a \in A, b \in B, a \perp \perp b$.

2.2 The association rates as measures of local association

Being interested in the association between variables *A* and *B* amounts to being interested in the departures from global independence, *i.e.*, all departures from local independence. This is done by introducing a measure of local association.

Definition 3 (Association rate, [10]) The association rate between a and b is defined as $t_{ab} = (f_{ab} - \widehat{f_{ab}})/(\widehat{f_{ab}})$.

The sign of t_{ab} indicates whether there is an *attraction* (case $t_{ab} > 0$), a local independence (case $t_{ab} = 0$), or a *repulsion* (case $t_{ab} < 0$) between *a* and *b*. The maximum repulsion is obtained when $t_{ab} = -1$, *i.e.*, when $f_{ab} = 0$, but there is no *a priori* upper limit for t_{ab} . The index t_{ab} can also be interpreted as a *over*- or *under-representation rate* of cell *ab* with respect to the *a* $\perp b$ case: for example, $t_{ab} = +0.50$ (resp. -0.50), indicates that cell *ab* contains 50% *more* (resp. *less*) observations than in the *a* $\perp b$ case.

2.2.1 **Properties of association rates**

As should be clear from properties given below (see also [10, Chp. 7]), the productfrequencies $\hat{f} = (\hat{f}_{ab})_{a \in A, b \in B}$ must be considered as a canonical set of weights for $t = (t_{ab})_{a \in A, b \in B}$. In the following, we denote $Mean_R(t, \hat{f})$ the weighted mean of t (with weights \hat{f}) over $R \subset A \times B$ (R being omitted when $R = A \times B$).

Property 1 The marginal weighted average of \mathbf{t} , for any $a \in A$ or any $b \in B$, is equal to 0, i.e., $\operatorname{Mean}_{\{(a,b),b\in B\}}(\mathbf{t},\widehat{\mathbf{f}}) = 0$ and $\operatorname{Mean}_{\{(a,b),a\in A\}}(\mathbf{t},\widehat{\mathbf{f}}) = 0$.

Corollary 1 If in any row a (resp. column b) some t_{ab} is positive, then some other $t_{ab'}$ (resp. $t_{a'b}$) is negative: over-representation of some cells implies the existence of some under-represented cells. In particular, for a 2×2 table, a $\perp \perp$ b implies $A \perp \perp B$.

¹Throughout this paper, we use *K* to denote both the set of categories and its cardinal, and similarly for *A* and *B*, the distinction being always clear from the context. Unless otherwise stated, all sums over *k* (resp. *a*, *b*) run from 1 to *K* (resp. *A*, *B*).

Property 2 (Pooling) Consider two applications $A \longrightarrow A^*$ and $B \longrightarrow B^*$ and the pooled table, $A^* \times B^*$, then, $\forall a^* \in A^*$, $\forall b^* \in B^*$, $t_{a^*b^*} = \text{Mean}_{\{(ab),a \in a^*, b \in b^*\}}(t, \hat{f})$. In particular, consider cell ab and the pooled table $A^* \times B^*$, where $A^* = \{a, a'\}$ and $B^* = \{b, b'\}$. Then t_{ab} is unchanged, whether it is defined from table $A \times B$ or from $A^* \times B^*$.

Note 1 (Global independence and t) From Definitions 2 and 3, $A \perp B$ occurs if and only if the t_{ab} 's are all equal to 0. Conversely, the departure of any t_{ab} from 0 indicates a departure from independence. What is important here is that the precise pattern of the t_{ab} 's departures from 0 points to the direction of association.

2.2.2 Example: Dyad data (continued)

Table 3 gives the t_{ab} 's for all cells of Table 2(left). Descriptively, (*i*) all diagonal cells but one are over-represented, (*ii*) all cells below the diagonal but one are maximally under-represented, and (*iii*) four of the six cells above the diagonal are under-represented (two maximally). Several of these results go in the direction of the pattern of Table 2(right), but this model, if taken at the cell level, is not descriptively satisfied.

2.3 Mean association rate over a region *R*: index t_R

In order to express the idea that some region $R \subset A \times B$ is over- or under-represented, we shall have recourse to a more global index as in [5].

Definition 4 (Mean association rate) Given a region $R \subset A \times B$, the mean association rate over R is defined as, $t_R = \text{Mean}_R(\mathbf{t}, \widehat{\mathbf{f}})$.

The index t_R varies from -1 (all cells in R are empty), to negative values (under-representation of R), to 0 (independence on average in R), to positive values (over-representation of R) without any a priori upper bound.

Table 3: Dyad data. Observed association rates t_{ab} from data of Table 2.

	b0	<i>b</i> 1	<i>b</i> 2	<i>b</i> 3	<i>b</i> 4
<i>a</i> 0	-1.00	0.52	0.31	-0.15	-1.00
<i>a</i> 1	-0.16	0.47	-0.41	-0.71	0.47
<i>a</i> 2	1.74	-1.00	10.50	-1.00	-1.00
<i>a</i> 3	1.03	-1.00	-1.00	1.12	0.64
<i>a</i> 4	-1.00	1.13	-1.00	-1.00	-1.00

2.3.1 Example: Dyad data (continued)

Consider the Dyad data and the pattern of over-/under-representation of Table 2(right). One possible way to confront the data to this model, at a descriptive level, is to compute the observed mean association rates for the three regions, D for cells on the diagonal, U for cells above and L for cells below it. This yields $t_D = 0.75$, $t_U = -0.50$ and $t_L = -0.91$. A global descriptive summary of the data, which goes in the direction of the expected pattern, is thus: $t_D > 0 > t_U > t_L$.

2.4 The Del index, a measure of agreement with a logical model

2.4.1 Quasi-implication for a 2 × 2 table

Consider a 2 × 2 table, with binary variables $A = \{a, a'\}$ and $B = \{b, b'\}$. We assimilate *a* and *b* to logical propositions, and denote negation by priming, conjunction by concatenation, implication by \Longrightarrow , and the false proposition by \emptyset . Then the statement $a \Longrightarrow b$ (*i.e.*, any observation of type *a* is necessarily of type *b*) is equivalent to $ab' \Longrightarrow \emptyset$, *i.e.*, that cell ab' is empty (cell ab' is an *error cell* for model $a \Longrightarrow b$, see [7]). Bernard [4] weakened the notion of a strict implication $a \Longrightarrow b$ into that of a *quasi-implication*, denoted by $a \longrightarrow b$, by defining the descriptive index $d_{a \Longrightarrow b} = -t_{ab'}$ as a measure of the *degree of agreement* with the logical model $a \Longrightarrow b$. For a given threshold $d_{quasi} > 0$, quasi-implication was defined by: $a \longrightarrow b \iff d_{a \Longrightarrow b} \ge d_{quasi}$.

2.4.2 Generalization to any logical model, the Del index

Definition 5 (Del index, [7]) More generally, consider a logical model \mathcal{M} relative to an $A \times B$ table, and denote by $\mathcal{E}_{\mathcal{M}}$, or \mathcal{E} for short, the set of all error cells that contradict \mathcal{M} , i.e., such that $\mathcal{M} = \bigwedge (ab \Longrightarrow \emptyset)_{(a,b) \in \mathcal{E}}$. Let $t_{\mathcal{E}}$ be the mean association rate over region \mathcal{E} . Then a global measure of the degree of agreement of the data with \mathcal{M} is the Del index, $d_{\mathcal{M}} = -t_{\mathcal{E}}$.

Properties of $d_{\mathcal{M}}$ flow from those of (mean) association rates. The index $d_{\mathcal{M}}$ varies in the range $] -\infty, 1]$; $d_{\mathcal{M}} = 0$ in case of independence on average in region \mathcal{E} and $d_{\mathcal{M}} = 1$ when \mathcal{M} is verified. A value of $d_{\mathcal{M}}$ between 0 and 1 can thus be interpreted as a quasi-agreement of the data with \mathcal{M} at degree $d_{\mathcal{M}}$; the closer to 1 its value is, the better the quasi-agreement is.

Property 3 (Equivalent logical models) Consider two logical models \mathcal{M}_1 , defined on $A \times B$, and \mathcal{M}_2 , defined on a table $A^* \times B^*$ obtained by coarsenings of the A and B classifications, such that \mathcal{M}_1 and \mathcal{M}_2 are logically equivalent. Then, $d_{\mathcal{M}_1} = d_{\mathcal{M}_2}$. This property follows from Property 2.

As seen from Definition 5, $d_{\mathcal{M}}$ and t_R are equivalent indices. In using t_R , we want to stress the over-/under-representation interpretation and the independence

case as a privileged reference ($t_R = 0$), whereas, in using $d_{\mathcal{M}}$, we stress the interpretation in terms of quasi-agreement with a strong/logical model and we point model \mathcal{M} as a privileged reference ($d_{\mathcal{M}} = 1$).

2.4.3 Example: Stages data (continued)

Consider the Stages data in Table 1 and the logical model \mathcal{M} associated with $\mathcal{E} = \{(a1, b2), (a1, b3), (a2, b3)\}$. We see that only two observations fall in region \mathcal{E} and we find $d_{\mathcal{M}} = 0.851$. Descriptively, at threshold, say, $d_{quasi} = 0.80$, we may conclude that the data quasi-agree with model \mathcal{M} .

3 Bayesian inference

We now assume that the data $\mathbf{x} = (x_1, \dots, x_K)$ is a multinomial sample (with $K = A \times B$ categories) of size *n* from a population characterized by the unknown parameters $\mathbf{\theta} = (\theta_1, \dots, \theta_K)$, the true frequencies of the *K* categories: $\mathbf{x} \sim Mn(n, \mathbf{\theta})$. We now want to make inferences about $\mathbf{\theta}$, and, more precisely here, about derived parameters such as τ_{ab} , τ_R and $\delta_{\mathcal{M}}$ which are the population counterparts of the descriptive indices t_{ab} , t_R and $d_{\mathcal{M}}$.

3.1 Dirichlet model for θ

In the usual Bayesian conjugate analysis, prior uncertainty about $\boldsymbol{\theta}$ is described by a Dirichlet prior distribution, $\boldsymbol{\theta} \sim Diri(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ and each hyper-parameter $\alpha_k > 0$. We call the α_k 's the *prior strengths* and $\mathbf{v} = \sum_k \alpha_k$ the *total prior strength*. We shall use an alternative parameterization of the Dirichlet in terms of the *prior frequencies* $\boldsymbol{\varphi} = \boldsymbol{\alpha}/\nu$, where $\boldsymbol{\varphi} \in \mathcal{S}^*(1, K)$ and $\mathcal{S}^*(1, K)$ denotes the interior of simplex $\mathcal{S}(1, K)$.² The prior expectations are simply $E(\boldsymbol{\theta}_k) = \boldsymbol{\varphi}_k$. The posterior distribution on $\boldsymbol{\theta}$ is then an updated Dirichlet distribution, $\boldsymbol{\theta} | \boldsymbol{x} \sim$ $Diri(\boldsymbol{x} + \boldsymbol{\alpha}) = Diri(\boldsymbol{x} + \nu \boldsymbol{\varphi})$, with posterior expectations given by,

$$E(\boldsymbol{\theta}_k | \boldsymbol{x}) = \frac{x_k + \boldsymbol{v} \boldsymbol{\varphi}_k}{n + \boldsymbol{v}}.$$
 (1)

3.2 Objective Bayesian models

For multinomial data, four Dirichlet priors have been proposed as models for prior ignorance about **\Theta**. All are symmetric Dirichlet, that is $\varphi_k = 1/K$ for any k, and they only differ in their respective total prior strength $\nu: \nu \to 0$ (Haldane), $\nu = 1$ (Perks), $\nu = K/2$ (Jeffreys) and $\nu = K$ (Bayes-Laplace's uniform prior).

Haldane's improper prior leads to some undesirable inferences: when $x_k = 0$, it leads to infer that $\theta_k = 0$, even when *n* is small. A major difficulty with the other

²Walley [12] uses symbols *s* and *t_k* in place of v and φ_k respectively.

three objective Bayesian priors is that inferences they produce depend on how the K categories are distinguished, which is partly arbitrary, and thus they do not satisfy the RIP (see [12]). Jeffreys' prior does not satisfy the LP either. Although it is often claimed that inferences from these priors differ in a negligible way when n is not small, large discrepancies can be obtained for statements bearing on unobserved or rare cells, even with large n.

4 Imprecise Dirichlet model

4.1 Presentation of the model

Walley [12] proposed the *imprecise Dirichlet model (IDM)* as a model for prior ignorance in the case of categorical data. The model consists in describing prior uncertainty about $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$ by a set of Dirichlet priors. The prior IDM(v) is defined as the set of all Dirichlet priors on $\boldsymbol{\theta}$ with a fixed total prior strength v > 0, *i.e.*, the set {*Diri*($\boldsymbol{\alpha}$) : $\alpha_k > 0$ for all k, $\sum_k \alpha_k = v$ }, or equivalently

$$\{Diri(\mathbf{v}\boldsymbol{\varphi}): \boldsymbol{\varphi} \in \mathcal{S}^{\star}(1,K)\},\tag{2}$$

where $S^*(1, K)$ is the interior of the simplex S(1, K).

Let $P_{v\varphi}(\cdot)$ and $E_{v\varphi}(\cdot)$ be respectively a prior probability and a prior expectation provided by a particular $Diri(v\varphi)$ in the set (2). The uncertainty about any event Z concerning $\boldsymbol{\theta}$ is described by *prior lower and upper probabilities*, denoted by $\underline{P}(Z)$ and $\overline{P}(Z)$, and calculated by minimizing and maximizing $P_{v\varphi}(Z)$ with respect to $\boldsymbol{\varphi} \in S^*(1, K)$. Similarly, for any real-valued function $\lambda = g(\boldsymbol{\theta})$, *prior lower and upper expectations* $\underline{E}(\lambda)$ and $\overline{E}(\lambda)$ are calculated by minimizing or maximizing the expectation $E_{v\varphi}(\lambda)$ with respect to $\boldsymbol{\varphi}$. Inferences about λ can be summarized by the *prior lower and upper cumulative distribution functions* $(cdf's), \underline{F}_{\lambda}(l) = \underline{P}(\lambda > l)$ and $\overline{F}_{\lambda}(l) = \overline{P}(\lambda > l)$.

Each Dirichlet prior in the prior IDM(v) is updated into a Dirichlet posterior using Bayes' theorem. This updating procedure guarantees coherence of the inferences [11]. Hence the posterior uncertainty about **\theta** from the IDM(v) is expressed by the set

$$\{Diri(\mathbf{x} + \mathbf{v}\mathbf{\phi}) : \mathbf{\phi} \in \mathcal{S}^{\star}(1, K)\}$$
(3)

As for the prior IDM, posterior lower and upper probabilities, expectations and cdf's are obtained by minimization or maximization with respect to $\mathbf{\phi} \in S^*(1, K)$.

The IDM satisfies several desirable principles of inference, and in particular both the LP and the RIP (see [12]). The RIP states that posterior inferences about any derived parameter $\lambda = g(\mathbf{0})$ should not depend on the number of categories *K* used for defining λ . The RIP is satisfied by the IDM in so far as the total prior strength v is specified independently of *K*.

4.2 Prior and posterior inferences about θ_k from the IDM

The posterior lower and upper expectations of θ_k are given by

$$\underline{E}(\boldsymbol{\theta}_k | \boldsymbol{x}) = x_k / (n + \mathbf{v}) \quad \text{and} \quad \overline{E}(\boldsymbol{\theta}_k | \boldsymbol{x}) = (x_k + \mathbf{v}) / (n + \mathbf{v}), \tag{4}$$

55

and are obtained as $\varphi_k \to 0$ and $\varphi_k \to 1$ respectively. The two same limiting values lead to the posterior upper and lower cdf's respectively, $\overline{P}(\theta_k > l | \mathbf{x})$ which is the $Beta(x_k, n - x_k + \mathbf{v})$ cdf, and $\underline{P}(\theta_k > l | \mathbf{x})$ which is the $Beta(x_k + \mathbf{v}, n - x_k)$ cdf.

By setting $n = x_k = 0$ in (4), we see that prior uncertainty about θ_k is maximal. We have $\underline{E}(\theta_k) = 0$ and $\overline{E}(\theta_k) = 1$, and $\underline{P}(\theta_K > l) = 0$ and $\overline{P}(\theta_k > l) = 1$ for any 0 < l < 1, that is *vacuous* lower and upper probabilities.

4.3 Choice of v

The IDM as defined in (2) and (3) depends on the choice of v. The constant v determines how fast the lower and upper probabilities converge one towards the other as *n* increases, and can thus be interpreted as a measure of the caution of the inferences. The larger v is, the more cautious the inferences are. The most important criterion for the choice of v is the requirement that the IDM should be cautious enough to encompass frequentist or objective Bayesian alternatives, while not being too cautious to avoid too weak inferences.

The first researches about the IDM lead to several convincing arguments for choosing $1 \le v \le 2$, but most of these arguments are relative to the binary case (K = 2) only (see [2, 12]. More recent work provides some support for v = 2 in the case of large *K*, for non-parametric inference about a mean [3]. In the following, we shall use v = 2, a value which is also supported by results in Section 5.4.

4.4 Two conjectures about the IDM

Conjecture 1 (Expectation of a derived parameter) Let $\lambda = g(\mathbf{\theta})$ be a real-valued function of $\mathbf{\theta}$, and $E_{v\mathbf{\phi}}(\mathbf{\theta})$ the prior (resp. posterior) expectation of $\mathbf{\theta}$ under the prior Diri $(v\mathbf{\phi})$ (resp. posterior Diri $(\mathbf{x} + v\mathbf{\phi})$). Then the upper and lower expectations of $\mathbf{\theta}$ under the IDM(v) are obtained from the (or one of the) Dirichlet prior which maximizes (resp. minimizes) $g(E_{v\mathbf{\phi}}(\mathbf{\theta}))$ with respect to $\mathbf{\phi}$.

Conjecture 2 (Cdf of a real-valued derived parameter) Let $\lambda = g(\mathbf{0})$ be a realvalued function of $\mathbf{0}$. Let Diri $(\mathbf{v}\mathbf{\phi})$ be a Dirichlet prior which provides the lower (resp. upper) prior or posterior expectation of λ under the IDM(\mathbf{v}), then it also provides the prior or posterior upper (resp. lower) cdf of λ .

The two conjectures hold if g(.) is a linear function of the θ_k 's. We don't expect them to be true in the general case (there are simple counter-examples to Conjecture 1). Nevertheless, we suggest that these conjectures actually provide

reasonable approximations for the lower and upper expectations and cdf's of λ for most functions g(.). In any case, the procedures they induce necessarily lead to an upper (resp. lower) bound for $\underline{E}(\lambda)$ and $\underline{F}_{\lambda}(.)$ (resp. $\overline{E}(\lambda)$ and $\overline{F}_{\lambda}(.)$).

5 Inference about a single association rate τ_{ii}

We first investigate the properties of the inferences about a single association rate τ_{ab} from the IDM. The following lemma shows that inferences about τ_{ab} can be carried out from the analysis of a simple 2 × 2 table.

Lemma 1 Consider the pooled table $A^* \times B^*$, with $A^* = \{a, a'\}$ and $B^* = \{b, b'\}$ and denote τ^*_{ab} the association rate of cell ab from the pooled table. From Property 2, $\tau^*_{ab} = \tau_{ab}$. Further, inferences from the IDM are invariant by such a pooling, since the IDM obeys the RIP. Thus, inferences about any single τ_{ab} only involve the relevant 2×2 table, $A^* \times B^*$.

5.1 Prior upper and lower expectation and cdf

The prior lower and upper expectation of τ_{ab} are given by $\underline{E}(\tau_{ab}) = -1$ and $\overline{E}(\tau_{ab}) \to +\infty$, and are attained respectively by $\varphi_{ab} = \varphi_{a'b'} \to \frac{1}{2}$, and by $\varphi_{ab} = \lambda$, $\varphi_{a'b'} = 1 - \lambda$, with $\lambda \to 0$. The same limiting values of $\boldsymbol{\varphi}$ also lead to the prior upper and lower cdf's respectively, $\underline{P}(\tau_{ab} > t) = 0$ and $\overline{P}(\tau_{ab} > t) = 1$, for any 0 < t < 1. These results show that prior inferences about τ_{ab} are vacuous. The prior IDM thus expresses a state of prior ignorance about parameter τ_{ab} .

5.2 Posterior upper and lower expectation and cdf

As in [4], we have recourse to Conjecture 1 in order to find approximate values for the posterior upper and lower expectations of τ_{ab} . Write $\tau_{ab} = g(\boldsymbol{\theta})$ where g(.)is such that $t_{ab} = g(\boldsymbol{f})$ and g(.) is given by Definition 3. Under a single Dirichlet prior, *Diri*(v $\boldsymbol{\varphi}$), the posterior expectation $E_{v\boldsymbol{\varphi}}(\tau_{ab}|\boldsymbol{x})$ is approximated by replacing each θ_k in g(.) by $E(\theta_k|\boldsymbol{x})$ given in (1), that is

$$E_{\mathbf{v}\boldsymbol{\phi}}^{\star}(\tau_{ab}|\boldsymbol{x}) = \frac{x_{ab} + \mathbf{v}\boldsymbol{\varphi}_{ab}}{(x_a + \mathbf{v}\boldsymbol{\varphi}_a)(x_b + \mathbf{v}\boldsymbol{\varphi}_b)} - 1$$
(5)

where x_a and x_b are the marginal counts of cell ab, and φ_a and φ_b its marginal prior frequencies. Conjecture 1 suggests then to minimize (resp. maximize) $E_{v\varphi}^*(\tau_{ab}|\mathbf{x})$ with respect to $\mathbf{\varphi}$, in order to estimate the posterior lower (resp. upper) expectations of τ_{ab} under the IDM(v). The minimum value is attained by letting $\varphi_{ab'} \rightarrow 1$, $\varphi_{a'b} \rightarrow 1$, or $\varphi_{ab'} = \varphi_{a'b} \rightarrow 1/2$, whether $f_{ab'}$ is lower than, greater than, or equal to $f_{a'b}$ respectively. The maximum value is attained by letting $\varphi_{ab} \rightarrow 1$ or $\varphi_{a'b'} \rightarrow 1$ whether $x_a x_b > x_{ab} (x_a + x_b + v)$ or not. Following Conjecture 2, we use the same values for finding approximate posterior lower and upper cdf's of τ_{ab} .

5.3 Dyad data: Summary of local inferences

Table 4 gives the lower and upper probabilities of a positive association rate from the IDM with v = 2, for each cell *ab* concerned by the prediction given in Table 2(right). Three of the four diagonal cells, (a1,b1), (a2,b2) and (a3,b3), can be assessed to be inductively over-represented with a high guarantee, $\underline{P}(\tau_{ab} > 0)$ being at least 0.99 for any of them. For cell (a4,b4), the probability interval, [0.00; 1.00] is almost vacuous; uncertainty still dominates, even after observing 115 observations. For the regions off the diagonal, only cells (a1,b3) and (a3,b1)are guaranteed to be under-represented, since, in both cases, $\underline{P}(\tau_{ab} < 0) = 1 - \overline{P}(\tau_{ab} > 0) = 1.00$; cells (a2,b1) and (a3,b2) have a probability of at least 0.79 and 0.61 to be under-represented; uncertainty concerning the 8 remaining offdiagonal cells is even larger, since $\underline{P}(\tau_{ab} < 0) < 0.50$ for each cell.

The first overall conclusion that may be drawn from these results is that the model shown in Table 2(right) cannot not be inductively assessed at the cell level.

Of course, any other reference value for τ_{ab} than 0 can be used in a similar way. For instance, the probability intervals for event $\tau_{ab} > 0.50$ for diagonal cells are: [0.30;0.50] for (a1,b1), [0.98;1.00] for (a2,b2), [0.99,1.00] for (a3,b3) and [0.00,0.99] for (a4,b4). Both cells (a2,b2) and (a3,b3) can be assessed to be over-represented by at least 50% with a high lower probability.

Table 4: Dyad data. Lower and upper posterior probabilities for event $\tau_{ab} > 0$, $\underline{P}(\tau_{ab} > 0|\mathbf{x})$ and $\overline{P}(\tau_{ab} > 0|\mathbf{x})$, for cells indexed by $a1, \ldots, a4$ and $b1, \ldots, b4$ only, using the IDM(v = 2).

	b0	<i>b</i> 1	<i>b</i> 2	<i>b</i> 3	<i>b</i> 4
<i>a</i> 0					
<i>a</i> 1		1.00;1.00	0.09;0.65	0.00;0.00	0.45;0.95
<i>a</i> 2		0.00;0.21	0.99;1.00	0.00;0.57	0.00;0.99
<i>a</i> 3		0.00;0.00	0.00;0.39	1.00;1.00	0.53;0.97
<i>a</i> 4		0.56;1.00	0.00;0.99	0.00;0.81	0.00;1.00

5.4 Comparison with frequentist and Bayesian approaches

Let us consider the test of the hypothesis $H_0: \tau_{ab} \le 0$ versus $H_1: \tau_{ab} > 0$. Due to Corollary 1, this test is equivalent to $H_0: \Phi \le 0$ versus $H_1: \Phi > 0$, where Φ is the usual contingency coefficient for a 2 × 2 table.

In the frequentist framework, the usual corresponding test is Fisher's exact test for a 2 × 2 table. The one-sided level p_{inc} of this test is usually computed as the probability of the *observations or more extreme cases* (*inclusive* test) under H_0 . However, as argued by [2], this choice is a matter of convention and one could also envisage the *exclusive* alternative with level p_{exc} involving *more extreme cases*

only. The following lemma shows that both these frequentist tests can actually be reinterpreted in a Bayesian way.

Lemma 2 Let p_{exc} and p_{inc} by the exclusive and the inclusive levels (one-sided) of Fisher's exact test of $H_0: \Phi \leq 0$ versus $H_1: \Phi > 0$ for a 2×2 table with counts \mathbf{x} . Let $P_{v\mathbf{\phi}}(.)$ be a Bayesian probability obtained from the prior Diri $(v\mathbf{\phi})$ on $\mathbf{\theta}$. Then, $p_{exc} = P_{v\mathbf{\phi}}(H_1|\mathbf{x})$ with v = 2 and $\mathbf{\phi} = (0, \frac{1}{2}, \frac{1}{2}, 0)$, and $p_{inc} = P_{v\mathbf{\phi}}(H_1|\mathbf{x})$ with v = 2 and $\mathbf{\phi} = (\frac{1}{2}, 0, 0, \frac{1}{2})$. The former prior allocates non-null strengths evenly to cells (a, b') and (a', b), the latter to cells (a, b) and (a', b').

Lemma 3 Under the same assumptions, the probability $P_{\nu \varphi}(\tau_{ab} > 0|\mathbf{x})$ from any of the four symmetric (φ constant) objective Bayesian priors, i.e., $\nu \rightarrow 0$, $\nu = 1$, $\nu = 2$ and $\nu = 4$, are in the interval $[p_{exc}; p_{inc}]$.

Proof. Lemmas 2 and 3 can be readily deduced from results in [1, Sec. 3].

Theorem 1 For any cell (a,b), the posterior lower and probabilities of event $\tau_{ab} \leq 0$ from the IDM with v = 2 encompass (i) Fisher's exact probabilities for H_0 : $\tau_{ab} \leq 0$ versus $H_1 : \tau_{ab} > 0$ using either the exclusive or the inclusive convention and (ii) the Bayesian posterior probabilities of the same event under the objective priors of Haldane, Perks, Jeffreys and Bayes-Laplace (the latter two being defined on the relevant specific 2×2 table).

Proof. The proof follows from (i) the equivalence between $\tau_{ab} > 0$ and $\Phi > 0$ for the pooled $\{a, a'\} \times \{b, b'\}$ table, (ii) the two Lemmas 2 and 3, and (iii) from the fact that the two Bayesian priors equivalent to p_{exc} and p_{inc} are such that v = 2 and thus belong to the IDM(v = 2).

Note 2 In analyzing a 2 × 2 table, Walley et al. [13, Sec. 5.4] advocate the use of two independent IDM's with same prior strength v_1 , one for each line of the table. They note that the value $v_1 = 1$ leads to $\overline{P}(H_0|\mathbf{x}) = p_{inc}$, a result which is only half of what Lemma 2 says. Here, we propose a more cautious model, a single IDM with $v = 2v_1 = 2$ for the whole table, which encompasses Walley's model. As Theorem 1 implies, our model has the advantage of producing inferences that encompass inferences from alternative objective models for all cells of the table simultaneously. The IDM(v = 2) is the smallest IDM having this property.

5.5 Absent or rare cells

For some cells, posterior uncertainty is still quite large. As an example, consider the unobserved cell (a2, b4) for which the posterior probability interval for $\tau_{ab} > 0$ is almost vacuous, [0.00; 0.986] (see Table 4). Such a wide interval results from the rareness of both a2 and b4 ($f_{a2} = f_{b4} = 4/115$). Even if a2 and b4 were locally independent, the expected number of observations in cell (a2, b4) would

59

be extremely small, $\widehat{x_{a2b4}} = n\widehat{f_{a2b4}} = 16/115$, far less than one observation. Thus, despite the extreme descriptive result $t_{a2b4} = -1$, both the hypotheses $a2 \perp b4$ ($\tau_{a2b4} = 0$) and $a2b4 \implies \emptyset$ ($\tau_{a2b4} = -1$) are compatible with the data. A similar result was found by [4]. This uncertainty is also reflected in the large differences between the alternative objective models: $P(\tau_{a2b4} > 0)$ ranges from 0 (Haldane), 0.350 (Perks), 0.571 (Jeffreys), to 0.802 (Bayes-Laplace), and the corresponding probability from Fisher's exact tests are 0 (exclusive) and 0.866 (inclusive).

6 Inference about a mean association rate τ_R

Without loss of generality (see Property 1), we consider a non-empty region R which does not contain any full row or a full column of the $A \times B$ table. It is easy to find a Dirichlet prior within the IDM for which the prior lower expectation of τ_R is -1 ($\forall (a,b) \in R, \varphi_{ab} \rightarrow 0$ with strengths of cells outside R carefully chosen). This limiting value for $\mathbf{\varphi}$ also provides the prior upper cdf, $\overline{P}(\tau_R > t) = 1$ for 0 < t < 1. We believe that the prior upper expectation and lower cdf of τ_R lead to vacuous inferences about τ_R , but we have no formal proof of that.

6.1 Posterior inferences about a single τ_R

Let $\tau_R = g(\mathbf{\theta})$, with $t_R = g(\mathbf{f})$ as given in Definition 4. We shall assume that allocating v to a single cell suffices to attain the lower or upper expectation or cdf of τ_R . This assumption actually appears to be true in most cases we tested, but is certainly not true in all cases. However, we shall consider that it provides a reasonable approximation for inferences about τ_R from the IDM. As a second level of approximation, we use the same argument as in Section 5.2 using Conjectures 1 and 2. Define $E_{v\mathbf{\theta}}^*(\tau_R | \mathbf{x}) = g(E_{v\mathbf{\theta}}(\mathbf{\theta} | \mathbf{x}))$ and $E_{v\mathbf{\theta}}(\mathbf{\theta} | \mathbf{x})$ is given by (1).

Theorem 2 Denote by r_{ab} the indicator variable of $(a,b) \in R$ and R' the complement of R in $A \times B$. Compute $m_{ab} = \sum_{i=1}^{A} r_{ib}x_i + \sum_{j=1...B} r_{aj}x_j$ for each cell (a,b). Then $\mathbb{E}^*_{v \mathbf{\varphi}}(\mathbf{\tau}_R | \mathbf{x})$ is minimized by letting $\mathbf{\varphi}_{ab} \to 1$ for cell $(a,b) \in R'$ maximizing m_{ab} . (We have no simple formula for maximization of $\mathbb{E}^*_{v \mathbf{\varphi}}(\mathbf{\tau}_R | \mathbf{x})$.) Proof involves tedious but rather simple algebra.

6.2 Stages data: Inference on $\delta_{\mathcal{M}}$

Consider the Stages data (Table 1) and the model \mathcal{M} defined therein. We found $d_{\mathcal{M}} = 0.851$ and we now want to make inferences about parameter $\delta_{\mathcal{M}}$ using the IDM($\nu = 2$). For various statements about $\delta_{\mathcal{M}}$, we find the following probability intervals: [1.00; 1.00] for $\delta_{\mathcal{M}} > 0$, [0.95; 1.00] for $\delta_{\mathcal{M}} > 0.50$, [0.84; 0.98] for $\delta_{\mathcal{M}} > 0.60$ and [0.62; 0.93] for $\delta_{\mathcal{M}} > 0.70$. We thus may assess that the data quasi-agree with \mathcal{M} at threshold $d_{quasi} = 0.50$, with probability at least 0.95.

6.3 Inferences about a complex directional association model

The IDM can be applied to study any kind of complex model expressed as a conjunction of constraints about association rates of specific cells or regions of an $A \times B$ table. Consider the Dyad data in Table 2(left) and the two models $\mathcal{M}_1 = \tau_L < \tau_U < 0 < \tau_D$ and $\mathcal{M}_2 = \tau_L < -0.70 < \tau_U < 0 < 0.50 < \tau_D$. Both try to express the expected pattern shown in Table 2(right), in a more or less strong way. Computing the posterior lower and upper probabilities of \mathcal{M}_1 or \mathcal{M}_2 can be done numerically by minimization/maximization over the set of Dirichlet posteriors. Using the IDM with $\nu = 2$, we find $\underline{P}(\mathcal{M}_1) = 0.98$, $\overline{P}(\mathcal{M}_1) = 1.00$, and $\underline{P}(\mathcal{M}_2) = 0.84$, $\overline{P}(\mathcal{M}_2) = 0.96$. Model \mathcal{M}_1 only is supported by the data with a sufficiently high lower probability.

Of course, models \mathcal{M}_1 and \mathcal{M}_2 are only two candidates amongst the possible inductive summaries of the data. The task of model selection (which is not addressed here) would require taking into account, not only the (lower) probability of each model, but also the degree of specificity or generality of each model.

7 Concluding remarks

This paper proposes a method for analyzing local or asymmetric dependencies in a contingency table, by focusing on previously suggested indices — (mean) association rates [5, 10] and *Del* index [7] —, which, we believe, are simple and natural, and yet provide means to define a wide variety of association models.

We showed how the imprecise Dirichlet model (IDM) can be applied to assess whether the data support such association models or not. Several results provide approximate solutions to the minimizing/maximizing problems required by the IDM. Further research would be needed to develop exact solutions or to measure the accuracy of our approximate procedures.

The exact comparison between the IDM and alternative frequentist or objective Bayesian models, carried out in Section 5.4 (see especially Theorem 1), provides a new argument for choosing v = 2 in the IDM, for a problem involving a possibly large number of categories (see also [4]). The large discrepancies which can be obtained in the inferences from these various alternative models are translated as a high imprecision in the IDM (see an example in Section 5.5). Section 5.4 shows that this phenomenon occurs whenever the frequentist *probability of the observed data* (under some particular null hypothesis) is not negligible.

References

 ALTHAM, P. M. E. Exact Bayesian analysis of a 2 × 2 contingency table and Fisher's exact significance test. J. Roy. Statist. Soc. Ser. B 31, 2 (1968), 261–269.

- [2] BERNARD, J.-M. Bayesian interpretation of frequentist procedures for a Bernoulli process. *The American Statistician* 50, 1 (1996), 7–13.
- [3] BERNARD, J.-M. Non-parametric inference about an unknown mean using the imprecise Dirichlet model. In *Proceedings of the 2nd International Symposium on Imprecise Probabilities and their Applications (ISIPTA'01)* (Maastricht, 2001), G. de Cooman, T. Fine, and T. Seidenfeld, Eds., Shaker Publishing BV, pp. 40–50.
- [4] BERNARD, J.-M. Implicative analysis for multivariate binary data using an imprecise Dirichlet model. J. Statist. Plann. Inference 105 (2002), 83–103.
- [5] DANIS, A., BERNARD, J.-M., AND LEPROUX, C. Shared picture-book reading: A sequential analysis of adult-child verbal interactions. *British Journal of Developmental Psychology 18* (2000), 369–388.
- [6] GOODMAN, L. A., AND KRUSKAL, W. H. Measures of association for cross classifications. II: Further discussion and references. J. Amer. Statist. Assoc. 54 (1959), 123–163.
- [7] HILDEBRAND, D. K., LAING, J. D., AND ROSENTHAL, H. Prediction Analysis of Cross Classifications. John Wiley & sons, 1977.
- [8] JAMISON, W. Developmental inter-relationships among concrete operational tasks: An investigation of Piaget's stage concept. *Journal of Experimental Child Psychology* 24 (1977), 235–253.
- [9] KENDALL, M. G., AND STUART, A. *The Advanced Theory of Statistics, Vol. 2: Inference and Relationship*, 3rd ed. Griffin, 1973.
- [10] ROUANET, H., BERT, M.-P., AND LE ROUX, B. Statistique en Sciences Humaines: Procédures Naturelles. Dunod, Paris, 1987.
- [11] WALLEY, P. Statistical reasoning with imprecise probabilities. In *Mono-graphs on Statistics and Applied Probability*, vol. 42. Chapman & Hall, London, 1991.
- [12] WALLEY, P. Inferences from multinomial data: learning about a bag of marbles. J. Roy. Statist. Soc. Ser. B 58 (1996), 3–57.
- [13] WALLEY, P., GURRIN, L., AND BURTON, P. Analysis of clinical data using imprecise prior probabilities. *The Statistician* 45 (1996), 457–485.

Jean-Marc Bernard is with the Laboratoire de Psychologie Environnementale, Université de Paris 5 & CNRS UMR 8069, 71 avenue Edouard Vaillant, 92774 Boulogne-Billancourt Cedex, France. E-mail: jmbernard@psycho.univ-paris5.fr