A NOTE ON QUASI - INDEPENDENCY

Johan Van Horebeek and Jef L. Teugels

Comunicación Técnica No I-02-03/25-02-2002
(CC/CIMAT)

A Note on Quasi-independency

Johan Van Horebeek & Jef L. Teugels

Abstract: *Quasi-independency is usually defined in a technical way by means of restrictions on the parameters of a particular probability model. In this contribution we discuss an alternative, intuitively appealing and generally applicable construction and make a comparison with the classical approach.*

Key words: *quasi-independency, log-linear model, categorical data.*

AMS Subject classification: 62H17, 62H05, 62H20

1 Introduction

Generally, quasi-independency refers to a situation where two variables in a contingency table are independent *discarding* some cells that correspond to combinations that are special in a certain sense.

The concept of quasi-independence can be traced back to the sixties (e.g., Caussinus (1965), Goodman (1961)), with roots in the work of K. Pearson and J.A. Harris in the late twenties. Despite the long tradition and the broad scope of applications covering social mobility, economics and genetics - to mention just a few - little attention has been paid to explore alternative ways to formalize the concept.

For a contingency table for the variables (X_1, X_2), quasi-independency on a set S of cells of the table is traditionally defined in terms of a factorization property of the underlying probability distribution:

$$P(X_1 = x_1, X_2 = x_2) = f(x_1)g(x_2) \text{ for } (x_1, x_2) \in S. \tag{1}$$

In log-linear models this condition can be expressed immediately by equating certain λ-parameters of the model to zero.

As a motivation for (1), many authors *suggest* implicitly or explicitly (cp. Agresti (1990) pag. 355) that X_1 is quasi-independent of X_2 on S if and only if

$$X_1 \perp X_2 | (X_1, X_2) \in S \tag{2}$$

i.e., they relate quasi-independency to the classical definition of independence. Although (2) is a formally correct statement, it is shown in Appendix I, that the condition generally defines a family of degenerated distributions. Hence, it should not be considered as the foundation of (1).

In the literature, two special cases are well known for which a correct formal definition (with a clear intuitive meaning) is available.

(i) The first one is when *all* (excluded) cells not belonging to S have *more* observations than what one expects under the classical independence hypothesis. As a classical example, look at the vote intentions with panel data, known for their inertia of the political affiliation of the respondents. Typically, a latent variable is introduced dividing the population in two groups: stayers who firmly belong to a certain political party and movers whose political preference might change. The frequencies of the movers are modelled by a classical independence model (Goodman (1965)).

(*ii*) A second special case is when the set of cells S has a particular configuration such that quasi-independence is equivalent to the simultaneous classical independency assumption for particular subtables of the original contingency table (Bishop et al. (1995)).

In the current paper we aim to present a *generally applicable* definition of quasi-independence with an intuitive interpretation. In many cases it will coincide with the classical definition. Nevertheless, there exist non-pathological configurations for S for which significant differences arise. In section 2, the definition is given while in section 3, a comparison is made with the classical definition. In the last sections we discuss some examples and the corresponding estimation problem.

In the sequel, we will suppose that $X_i \in \{0, \cdots, r_i - 1\}, i = 1,2$ and that all cell probabilities on S are nonzero. Outside S, cell probabilities are completely arbitrary.

2 Quasi-independency and Probability Ratios

Independency of two stochastic discrete variables is defined either by a factorization property of the joint probability distribution or by means of restrictions on conditional probabilities:

$$X_1 \perp X_2 \Leftrightarrow \forall x_1, x_2, y_1, y_2 : \begin{array}{l} P(X_1 = x_1 | X_2 = x_2) = P(X_1 = x_1 | X_2 = y_2) \\ P(X_2 = x_2 | X_1 = x_1) = P(X_2 = x_2 | X_1 = y_1). \end{array} \tag{3}$$

This first approach leads to (1) as a definition for quasi-independency, with the advantages and disadvantages already mentioned in the previous section.

An alternative starting point to introduce quasi-independency is to restrict the values of x_1, x_2, y_1, y_2 in (3) to those $(x_1, x_2), (x_1, y_2), (y_1, x_2)$ belonging to S. Nevertheless, as shown in Appendix II, the corresponding definition is not immediately useful for practical purposes. In general, due to the existence of *excluded* cells, typically defined in terms of the values of both variables, one should avoid defining quasi-independency using marginal distributions (as one does with conditional probabilities).

Instead of (3), we take as a starting point the characterization of independence of two stochastic variables X_1, X_2, which is expressed by means of the following relationships between ratios of probabilities:

$$X_1 \perp X_2 \quad \text{iff} \quad \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = y_1, X_2 = x_2)} \text{ does not depend of } x_2. \tag{4}$$

The restriction of (4) to cells belonging to S leads to the following definition.

Definition 2.1 *For a given set of cells S of a contingency table of (X_1, X_2), X_1 and X_2 are quasi-independent on S iff*

$$\frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = y_1, X_2 = x_2)} \tag{5}$$

does not depend on x_2 for every $(x_1, x_2), (y_1, x_2)$ belonging to S.

An equivalent requirement to (5) is

$$\frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1, X_2 = y_2)} \tag{6}$$

does not depend on x_1 for every $(x_1, x_2), (x_1, y_2)$ belonging to S.
We use the fact that the mean of a geometric distribution with parameter θ equals $1/\theta$ to give an intuitively appealing reformulation of the above definition.

Definition 2.1 *(alternative) For a given set of cells S of a contingency table of (X_1, X_2), X_1 and X_2 are on S iff under independent sampling of (X_1, X_2), the fraction of the mean waiting time to observe the first time (x_1, x_2) to the mean waiting time to observe the first time (y_1, x_2) does not depend on x_2 for every $(x_1, x_2), (y_1, x_2) \in S$.*

To get a flavour of the similarities and differences with the usual definition of quasi-independency, consider the contingency table of Fig. 1 where the excluded cells are marked in black while the cell probabilities are given in terms of the parameters f_i, g_i and μ, up to a normalization constant.

		$f_0 g_2$	$f_0 g_3$	$f_0 g_4$	$f_0 g_5$
		$f_1 g_2$	$f_1 g_3$	$f_1 g_4$	$f_1 g_5$
$\mu f_2 g_0$	$\mu f_2 g_1$			$f_2 g_4$	$f_2 g_5$
$\mu f_3 g_0$	$\mu f_3 g_1$			$f_3 g_4$	$f_3 g_5$
$\mu f_4 g_0$	$\mu f_4 g_1$	$\mu f_4 g_2$	$\mu f_4 g_3$		
$\mu f_5 g_0$	$\mu f_5 g_1$	$\mu f_5 g_2$	$\mu f_5 g_3$		

Figure 1.

Clearly, this distribution satisfies Def. 2.1. Because of the presence of the additional free parameter μ, there is in general no quasi-independence according to (1). Clearly, equation (1) is more restrictive. For the table of Fig. 1, taking $\mu = 1$, contrary to Def. 2.1, relationships like

$$\frac{P(X_1 = 0, X_2 = 2)}{P(X_1 = 5, X_2 = 2)} \frac{P(X_1 = 5, X_2 = 0)}{P(X_1 = 3, X_2 = 0)} = \frac{P(X_1 = 0, X_2 = 4)}{P(X_1 = 3, X_2 = 4)}$$

are implied as a by-product.

3 A Comparison with the Classical Definition

The configuration of the cells that belong to S determines the possible differences between the two approaches. We therefore use the following classifications.

Definition 3.1 *(Bishop et al. (1995)) Two cells in S are associated if they are in the same row or column of the contingency table.*

Definition 3.2 *(Bishop et al. (1995)) The set of cells S is connected if every pair of cells in S can be linked by a chain of cells, any two consecutive members of which must be associated.*

Here is a well known result: if S is not connected, there exists a permutation of the rows and columns such that the corresponding contingency table has a block-diagonal structure and each block can be treated in its own isolated way. For this reason we can and do assume that S is connected. If not, we apply the analysis to each of its connected components.

Definition 3.3 *A subset A of cells in S is united if every pair of two cells in S can be linked by a chain of cells of S, any two consecutive members of which must be neighbours (i.e., of the form $\{(x_1, x_2), (x_1, x_2 + 1)\}$, $\{(x_1, x_2), (x_1, x_2 - 1)\}$, $\{(x_1, x_2), (x_1 + 1, x_2)\}$ or $\{(x_1, x_2), (x_1 - 1, x_2)\}$).*

For example, in Fig. 2, the set $\{c_1, c_2, c_4\}$ is united. The set $\{c_1, c_3\}$ is not.

Definition 3.4 *Two disjoint subsets of cells A and B are correlated if there exist values x_1, x_2, y_1, y_2 such that of the four elements of the set $\{(x_1, x_2), (x_1, y_2), (y_1, y_2), (y_1, x_2)\}$, three belong to one set while the remaining one belongs to the other set.*

In Fig. 2, the cells $\mathcal{U}_1$ and $\mathcal{U}_2$ are correlated because of the set of cells $\{c_1, c_2, c_3, c_4\}$. The sets $\mathcal{U}_1$ and $\mathcal{U}_5$ are uncorrelated.

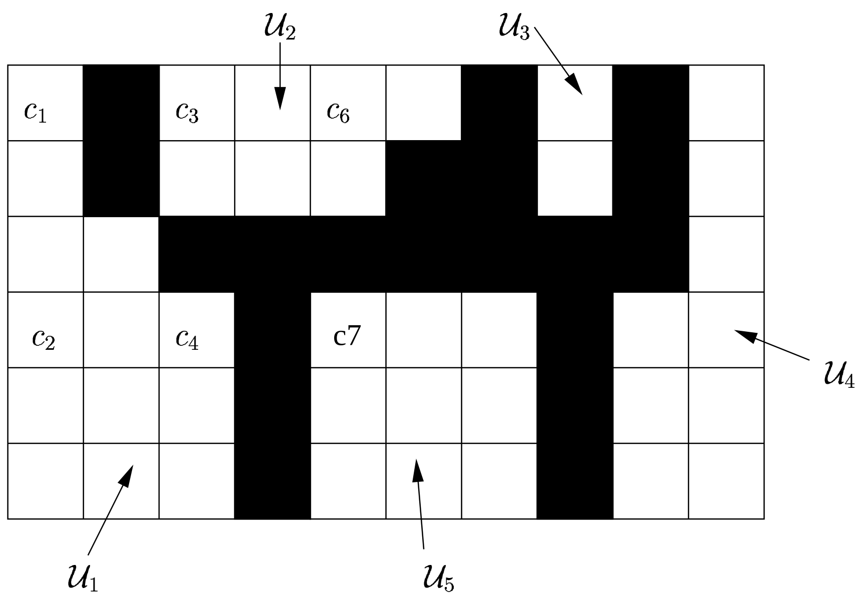

Figure 2.

Given a connected set S, denote by $\{\mathcal{U}_i\}$ the partition of S into its united subsets. Taking repeatedly the union of those $\{\mathcal{U}_i\}$ that are correlated, leads to an *uncorrelated partition* of S:

$$S = \cup_i \mathcal{V}_i \ ,$$

such that each $\mathcal{V}_i$ is uncorrelated with $S \setminus \mathcal{V}_i$.

For the table of Fig. 2, the two sets $\{c_1, c_2, c_3, c_4\}$ and $\{c_3, c_4, c_6, c_7\}$ lead to the uncorrelated partition consisting of $\mathcal{V}_1 = \mathcal{U}_1 \cup \mathcal{U}_2 \cup \mathcal{U}_5$, $\mathcal{V}_2 = \mathcal{U}_3$ and $\mathcal{V}_3 = \mathcal{U}_4$.

Finally, we define *border columns* and *border rows.*

Definition 3.5 *(i) Given S, we call the row x_1, $0 \leq x_1 \leq r_1 - 2$, a border row on S iff there does not exist a column x_2, $0 \leq x_2 \leq r_2 - 1$ such that (x_1, x_2) and $(x_1 + 1, x_2)$ both belong to S.*
(ii) Given S, we call the column x_2, $0 \leq x_2 \leq r_2 - 2$, a border column on S iff there does not exist a row x_1, $0 \leq x_1 \leq r_1 - 1$ such that (x_1, x_2) and $(x_1, x_2 + 1)$ both belong to S.

In Figure 2, column 7 is a border column. Observe that a border row or column will always separate at least two uncorrelated sets.

We now reformulate our concept of quasi-independence in terms of a representation for the elements in the contingency table.

Property 3.1 *For a given connected set of cells S and $P(\cdot)$, a positive distribution on S, $P(\cdot)$ is quasi-independent on S according to Def. 2.1 iff there exist functions $f(\cdot)$, $g(\cdot)$ and constants μ_i such that:*

$$\log P(X_1 = x_1, X_2 = x_2) = f(x_1) + g(x_2) + \sum_i \mu_i \mathbb{I}((x_1, x_2) \in \mathcal{V}_i) \quad , \quad \forall (x_1, x_2) \in S \quad (7)$$

where $\{\mathcal{V}_i\}$ is an uncorrelated partition of S and $\mathbb{I}(\cdot)$ denotes the indicator function.

Proof

Assume that quasi-independency holds. Then we define for $0 < x_1 < r_1 - 1$:

$$f_{x_1} = \begin{cases} \frac{P(X_1=x_1+1,X_2=x_2)}{P(X_1=x_1,X_2=x_2)} & \text{if } x_1 \text{ is not a border row} \\ 1 & \text{in the other case} \end{cases}$$

and similarly for $0 < x_2 < r_2 - 1$:

$$g_{x_2} = \begin{cases} \frac{P(X_1=x_1,X_2=x_2+1)}{P(X_1=x_1,X_2=x_2)} & \text{if } x_2 \text{ is not a border column} \\ 1 & \text{in the other case .} \end{cases}$$

The essentials of the proof will be explained for the particular case of the table of Fig. 2. Consider the set $\mathcal{U}_1$.

Use the fact that the cells are united and choose as a *reference cell* $P(X_1 = 0, X_2 = 0)$. We obtain

$$\forall (x_1, x_2) \in \mathcal{U}_1 : P(X_1 = x_1, X_2 = x_2) = P(X_1 = 0, X_2 = 0) \prod_{i=0}^{x_1-1} f_i \prod_{j=0}^{x_2-1} g_j, \tag{8}$$

where we take $\prod_{i=a}^{b} f_i = \prod_{j=a}^{b} g_j = 1$ if $a > b$.

Defining:

$$f(x_1) = \log \prod_{i=0}^{x_1-1} f_i, \quad g(x_2) = \log \prod_{j=0}^{x_2-1} g_j$$

and $\mu_1 = \log P(X_1 = 0, X_2 = 0)$, we obtain (7) for all cells in $\mathcal{U}_1$.

Next, consider the cells of $\mathcal{U}_2$. Repeat the above scheme but take $P(X_1 = 0, X_2 = 2)$ as reference cell. We obtain a factorization similar to (8) as

$$\forall (x_1, x_2) \in \mathcal{U}_2 : P(X_1 = x_1, X_2 = x_2) = P(X_1 = 0, X_2 = 2) \prod_{i=0}^{x_1-1} f_i \prod_{j=2}^{x_2-1} g_j. \tag{9}$$

Now, $\mathcal{U}_2$ and $\mathcal{U}_1$ are correlated. Therefore by quasi-independency, $P(X_1 = 0, X_2 = 2)$ can be entirely written in terms of cell probabilities of $\mathcal{U}_1$ in that

$$P(X_1 = 0, X_2 = 2) = \frac{P(X_1 = 0, X_2 = 0)P(X_1 = 2, X_2 = 2)}{P(X_1 = 2, X_2 = 0)}. \tag{10}$$

Note that all terms on the right hand side are of the form (7). If we substitute (10) into (9), we obtain a factorization of the left hand side in (9) which is of the requested form (7).

In general, it will not always be possible to take as a reference cell the one with the lowest index value for x_1 and lowest index value for x_2. Nevertheless, the fact that the

set is united, still guarantees an expansion of the form (7).
We repeat the above procedure for all united sets of $\mathcal{V}_1$.

For the cells of $\mathcal{U}_3$ in $\mathcal{V}_2$, the following expansion can be obtained:

$$\forall (x_1, x_2) \in \mathcal{U}_3 : P(X_1 = x_1, X_2 = x_2) = \frac{P(X_1 = 0, X_2 = 7)}{\prod_{j=0}^{7-1} g_j} \prod_{i=0}^{x_1-1} f_i \prod_{j=0}^{x_2-1} g_j. \qquad (11)$$

Taking $\mu_2 = \frac{P(X_1=0,X_2=7)}{\prod_{j=0}^{7-1} g_j}$ we again end up with an expansion of the form (7). The above argument can be repeated for each correlated set.

Corollary 3.1 *For a given connected set S, the difference in the number of parameters between the quasi-independence model of Def. 2.1 and (1) is equal to the number of uncorrelated subsets of S, decreased by 1 plus the number of border rows and border columns.*

One important class of tables where there is a difference between the number of parameters according to the classical and our definition of quasi-independency is formed by all $n \times n$ tables ($n \geq 3$), where S is a subset of all cells with a distance to the main diagonal of the table, that is greater than or equal to $\lfloor n/3 \rfloor$. The *smallest* member of this class is a 3×3 table with all diagonal cells excluded.

4 Estimation

First note that the model (7) is a Generalized Linear Model. Therefore, standard optimization procedures can be used to obtain the Maximum Likelihood Estimators (MLE) in case they exist.

Only for very particular configurations of the sampling zeros, a MLE might not exist. Adapting slightly the results of Fienberg (1970), one obtains the following sufficient condition.

Property 4.1 *Given an $r_1 \times r_2$ table and a connected set S, a unique nonzero MLE exists if the table restricted to the nonzero sampling cells, has the same uncorrelated partition and the same border rows and columns as the original one under quasi-independency on S.*

Situations where the restriction of the table to the nonzero sampling cells, *does* change the parameterization are, e.g., when some marginal (row or column) frequencies are zero, or when two correlated sets of the original table become uncorrelated.

Proof
The structure of the proof is basically the same as the one used in Fienberg (1970) to derive sufficient conditions to guarantee unique nonzero maximum likelihood estimators under classical quasi-independency.

For the sake of compatibility with Fienberg (1970), we will assume Poisson sampling. If we denote by m_{x_1,x_2} the expected number of observations in cell (x_1, x_2) and by n_{x_1,x_2} the number of observations in cell (x_1, x_2), the log-likelihood, L, equals, up to a constant:

$$L = \sum_{(x_1,x_2)\in S} (n_{x_1,x_2} \log m_{x_1,x_2} - m_{x_1,x_2}).$$

Observe that it is sufficient to prove that L satisfies the following conditions: (C1) it obtains its supremum for finite nonzero values of the parameters; (C2) it is a strictly concave function.

- In order to prove (C1), as shown in Freeman (1987), it is sufficient to show that for all cells (x_1, x_2) with sampling zeros, the maximum likelihood estimator for m_{x_1,x_2} under quasi-independency is nonzero (this is the only case where $n_{x_1,x_2} \log m_{x_1,x_2} - m_{x_1,x_2}$ does not tend to $-\infty$ when $\log m_{x_1,x_2} \to \pm\infty$).

  Suppose the opposite is true. In this case the likelihood function is equal to the one corresponding to the contingency table restricted to those cells without sampling zeros. To these the well-known result (Birch (1963)) applies that, for a contingency without zero cells, the likelihood is bounded from above and obtains its supremum on the inside of the parameter space.

  As we suppose that the set of parameters of the restricted table are the same as for the original table, $m_{x_1,x_2} > 0$ for the original table, leading to the required contradiction.

- To prove (C2): as $\log(\cdot)$ is strictly concave, it is sufficient to show that the parameterization of m_{x_1,x_2} in(7) is a 1-1 mapping on all cells with no sampling zeros.

  As we suppose that the sampling zeros do not change the parameterization (7), and as $\{m_{x_1,x_2}\}$ determines completely the parameter values (see the proof of (7)), we obtain the required 1-1 correspondence.

5 Examples

In this section we analyze two data sets, each one focusing on a different aspect of the above introduced quasi-independency concept. The first example illustrates the utility of having an intuitively clear definition of the quasi-independence concept; the second example illustrates that the new parameterization leads to conclusions, different from those resulting from the classical approach.

5.1 Example 1

The following data set was originally published in Jekel et al. (1978) and is extensively analyzed in Freeman (1987). It concerns a study of the occurrence of soft-tissue sarcomas. The data are summarized in Table 1; the variable X_1 denotes the type of sarcoma and X_2 the time period when the case was diagnosed.

		X_1		
		Fibroid (0)	Lipoid (1)	Mixed/Others (2)
	1935-44 (0)	40	12	21
X_2	1945-54 (1)	70	11	31
	1955-64 (2)	93	38	47
	1965-74 (3)	43	51	67

Table 1

The independency hypothesis is not acceptable due to cell $(3, 0)$ which has a standardized residual of -4.08175 under (full) independency. In Freeman (1987) (pag. 91 eq. 3.34), the author proposes the model

$$H_o : p_{x_1,x_2} = p_{x_1,+}p_{+,x_2} \text{ except for } x_1 = 3,\ x_2 = 0 \tag{12}$$

First of all, observe that in the above $p_{x_1,+}$ and p_{+,x_2} refer to the marginal probabilities defining $p_{3,0}$ equal to zero, i.e., not to the marginal distribution of X_1, X_2. Indeed, the latter would make no sense as $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$ for all but one cell, together with $\sum P(X_1 = x_1, X_2 = x_2) = 1$ automatically imply that the factorization holds for the complete table.

Technically, model (12) is equivalent to (7) and is acceptable with a p-value of 0.1618. Although it seems a minor detail, the use of Def. 2.1 instead of (12) offers a clear interpretation of what one is really testing. Assume we apply independent sampling. Then the ratio of (i) the mean waiting time to observe for the first time a person from period x_2 and with cell tissue x_1 and (ii) the mean waiting time to observe for the first time a person from period x_2 and with cell tissue x_1' does not depend on x_2, in case neither (x_1, x_2) nor (x_1', x_2) is the excluded cell.
Observe that the above interpretation is conceptually close to the basic research question of the authors in that they were *interested in studying changing patterns in sarcomas over time* (Freeman (1987)).

5.2 Example 2

The second example is taken from a 1976 Danish Welfare Survey where the structure of the classification in social rank groups was studied for married women with age between 40 and 59, versus their husband's social rank. The data are summarized in Table 2 (a). We do not pretend to present a complete statistical analysis of these data, but only point out how the choice of the definition of the quasi-independency concept influences the modeling stage.

As usual with this type of data, we suppose that the interest is in the off-diagonal cells. In Table 2 (b) the standardized residuals are given under quasi-independency excluding the diagonal cells (in this case, the two definitions coincide). One observes large residuals in two cells. We first exclude the cell $(1, 0)$ with the largest residual and adjust a quasi-independency model for the remaining cells (once again, in this case, the two definitions coincide). The resulting standardized residuals are given in Table 2 (c).

If one wishes to repeat the above procedure, one should exclude the cell $(0, 3)$. But then it will make a difference which definition of quasi-independency one is using. As Figure 3 shows, the deletion of $(0, 3)$ leads to two uncorrelated sets. Consequently, there will a difference of one parameter between the new and old definition. More specifically, under Definition 2.1 one obtains that the loglikelihood ratio statistic, G^2, equals 0.46 with 2 degrees of freedom while under (1), $G^2 = 4.43$ with 3 degrees of freedom.

		X_1 Woman's social rank			
		I-II (0)	III (1)	IV (2)	V (3)
X_2	I-II(0)	20	35	42	22
Husband's	III (1)	4	44	122	71
social	IV (2)	6	12	49	71
rank	V (3)	0	6	32	146

(a)

		X_1 Woman's social rank			
		I-II (0)	III (1)	IV (2)	V (3)
X_2	I-II(0)	-	-0.46	1.03	-1.54
Husband's	III (1)	4.10	-	-3.44	-1.01
social	IV (2)	-1.26	0.43	-	0.74
rank	V (3)	-1.92	-0.43	1.78	-

(b)

		X_1 Woman's social rank			
		I-II (0)	III (1)	IV (2)	V (3)
X_2	I-II(0)	-	-0.19	0.71	-1.54
Husband's	III (1)	-	-	-0.60	1.03
social	IV (2)	0.21	-0.03	-	-0.17
rank	V (3)	-0.28	0.09	0.07	-

(c)

Table 2

Of course, as with classical quasi-independency, the exclusion of cells in a statistical analysis should be done with a lot of caution.

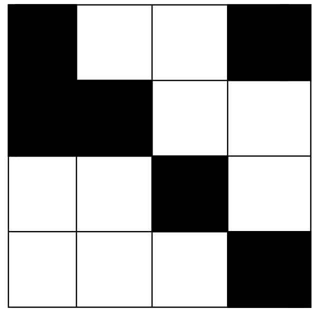

Figure 3.

6 Conclusions

In this paper we showed how the concept of quasi-independency can be formalized in intuitively appealing way independent of the log-linear parametrization. More specifically, our definition of quasi-independency is based on probabilistic concepts but can be translated into a technical representation. This approach turns out to provide insight in the exact nature of quasi-independency. Without doubt, it should be possible to apply this notion to other concepts from categorical data analysis.

Appendix I

Consider a squared contingency table and S the set of all off-diagonal cells. Using (2) as the definition of quasi-independency, we obtain

$$X_1 \perp X_2 | X_1 \neq X_2$$

or, equivalently,

$$\forall x_1 : P(X_1 = x_1, X_2 = x_2 | X_1 \neq X_2) = P(X_1 = x_1 | X_1 \neq X_2) P(X_2 = x_2 | X_1 \neq X_2).$$

Taking the summation over all x_1 such that $x_1 \neq x_2$, we get :

$$P(X_2 = x_2 | X_1 \neq X_2) = (1 - P(X_1 = x_2 | X_1 \neq X_2)) P(X_2 = x_2 | X_1 \neq X_2),$$

This implies $P(X_1 = x_2 | X_1 \neq X_2) = 0$, which is too strong a restriction for practical purposes.

Appendix II

Suppose one defines quasi-independency by the requirements

$$\forall (x_1, x_2), (x_1, y_2), (y_1, x_2) \in S : \quad \begin{array}{l} P(X_1 = x_1 | X_2 = x_2) = P(X_1 = x_1 | X_2 = y_2) \\ P(X_2 = x_2 | X_1 = x_1) = P(X_2 = x_2 | X_1 = y_1) \; . \end{array} \tag{13}$$

Let us take the particular case where S contains all but one specific cell of the table (x_k, x_l). Now consider all cells in row x_1, that do not contain (x_k, x_l), i.e., $x_1 \neq x_k$. Relation (13) implies

$$\frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1, X_2 = y_2)} = \frac{P(X_2 = x_2)}{P(X_2 = y_2)}.$$

As this is true for every y_2, repeated application of the basic property of real numbers

$$\frac{a}{b_1} = \frac{c}{d_1} \text{ and } \frac{a}{b_2} = \frac{c}{d_2} \quad \rightarrow \quad \frac{a}{b_1 + b_2} = \frac{c}{d_1 + d_1},$$

yields that

$$\frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)} = \frac{P(X_2 = x_2)}{1},$$

or that

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2) \ . \tag{14}$$

The same result can be obtained for all cells in a column that do not contain (x_k, x_l). As every cell different from (x_k, x_l), is in a row (or column) not containing (x_k, x_l), factorization (14) is valid for all cells belonging to S.

Use the identity $\sum_{x_1,x_2} P(X_1 = x_1, X_2 = x_2) = 1$ and some algebra to see that

$$P(X_1 = x_k, X_2 = x_l) = P(X_1 = x_k)P(X_2 = x_l) \ .$$

But then the independency also holds for the excluded cell (x_k, x_l). Hence (13) can not be used to define quasi-independency on all but one cell.

References

Agresti, A. (1990). *Categorical Data Analysis.* John Wiley & Sons.

Birch, M.W. (1963). Maximum likelihood in three-way contingency tables, *J. Royal Statist. Society B* 25, 220–233.

Bishop, Y., Fienberg, S., Holland, P. (1995). *Discrete Multivariate Analysis.* MIT Press.

Caussinus, H.(1965). Contribution á l'analyse statistique des tableaux de correlation, *Ann. Fac. Sci., Univ. Toulouse* 29, 77–182.

Fienberg, S.E. (1970). Quasi-independence and maximum likelihood estimation in incomplete contingency tables, *J. Amer. Statist. Assoc.* 65, 1610–1615.

Freeman, D.H. (1987). *Applied Categorical Data Analysis*, Marcel Dekker.

Goodman, L.A.(1961). Statistical methods for the mover-stayer model, *J. Amer. Statist. Assoc.* 56, 841–868.

Goodman, L.A.(1965). On the statistical analysis of mobility tables, *Amer. J. Sociol.* 70, 564–585.

Jekel, J.F., Freeman, D.H., and Meigs, J.W.(1978). A study of trends in upper arms soft tissue sarcomas in the State of Connecticut following the introduction of alum-adsorbed allergenic extract, *Annals of Allergy* 40 , 28–31.