

# Chapter 3

## Testing Joint Conditional Independence of Categorical Random Variables with a Standard Log-Likelihood Ratio Test



Helmut Schaeben

**Abstract** While tests for pairwise conditional independence of random variables have been devised, testing joint conditional independence of several random variables seems to be a challenge in general. Restriction to categorical random variables implies in particular that their common distribution may initially be thought of as contingency table, and then in terms of a log-linear model. Thus, Hammersley–Clifford theorem applies, and provides insight in the factorization of the log-linear model corresponding to assumptions of independence or conditional independence. Such assumptions simplify the full joint log-linear model, and in turn any conditional distribution. If the joint log-linear model corresponding to the assumption of joint conditional independence given the conditioning variable is not sufficiently large to explain some data according to a standard log-likelihood test, its null-hypothesis of joint conditional independence may be rejected with respect to some significance level. Enlarging the log-linear model by some product terms of variables and running the log-likelihood test on different models may provide insight which variables are lacking conditional independence. Since the joint distribution determines any conditional distribution, the series of tests eventually provides insight which variables and product terms a proper logistic regression model should comprise.

### 3.1 Introduction

Conditional independence is a probabilistic approach to causality (Suppes 1970; Dawid 1979, 2004, 2007; Spohn 1980, 1994; Pearl 2009; Chalak and White 2012) while for instance correlation is obviously not as it is a symmetric relationship. Features of conditional independence are

---

H. Schaeben (✉)

Geophysics and Geoinformatics, TU Bergakademie Freiberg, Freiberg, Germany  
e-mail: [schaeben@tu-freiberg.de](mailto:schaeben@tu-freiberg.de)

© The Author(s) 2018

B. S. Daya Sagar et al. (eds.), *Handbook of Mathematical Geosciences*,  
[https://doi.org/10.1007/978-3-319-78999-6\\_3](https://doi.org/10.1007/978-3-319-78999-6_3)

- Conditionally independent random variables are conditionally uncorrelated.
- Conditionally independent random variables may be significantly correlated or not.
- Independence does not imply conditional independence and vice versa.
- Pairwise conditional independence does not imply joint conditional independence.

Statistical tests for pairwise conditional independence of random variables have been devised, e.g., Bergsma (2004), Su and White (2007), Su and White (2008), Song (2009), Bergsma (2010), Huang (2010), Zhang et al. (2011), Bouezmarni et al. (2012), Györfi and Walk (2012), Doran et al. (2014), Ramsey (2014), Huang et al. (2016), testing joint conditional independence of several random variables seems to be a challenge in general. For the special case of dichotomous variables, the “omnibus test” (Bonham-Carter 1994) and the “new omnibus test” (Agterberg and Cheng 2002) have been suggested.

Weak conditional independence of random variables was introduced in Wong and Butz (1999), and elaborated on in Butz and Sanscartier (2002). Extended conditional independence has recently been introduced in Constantinou and Dawid (2015). The definition of weak conditional independence given in Cheng (2015) refers to conditional independent random events, and rephrases conditional independence in terms of ratios of conditional probabilities rather than conditional probabilities to avoid the distinction of conditional independence given a conditioning event or its complement. This definition becomes irrelevant when proceeding from elementary probability of events to probability of random variables, and to the general definition of conditionally independent random variables.

Conditional independence is an issue in a Bayesian approach to estimate posterior (conditional) probabilities of a dichotomous random target variable in terms of weights-of-evidence (Good 1950, 1960, 1985). In turn, conditional independence is the major mathematical assumption of potential modeling with weights of evidence, cf. (Bonham-Carter et al. 1989; Agterberg and Cheng 2002; Schaeben 2014b), e.g., applied to prospectivity modeling of mineral deposits. The method requires a training dataset laid out in regular cells (pixels, voxels) of equal physical size representing the support of probabilities. The sum of posterior probabilities over all cells equals the sum of the target variable over all cells. Deviations indicate a violation of the assumption of conditional independence, and are used as statistic of a test (Agterberg and Cheng 2002) which involves a normality assumption. Funny enough, ArcSDM calculates so-called normalized probabilities, i.e., posterior probabilities rescaled so that the overall measure of conditional independence is satisfied (ESRI 2018); of course, the trick does not fix any problem. Violation of the assumption of conditional independence does not only corrupt the posterior (conditional) probabilities estimated with weights of evidence, but also their ranks, cf. (Schaeben 2014b), which is worse. Thus, the method of weights-of-evidence requires the mathematical modeling assumption of conditional independence to yield reasonable predictions. However, conditional independence is an issue with respect to logistic regression, too.

## 3.2 From Contingency Tables to Log-Linear Models

A comprehensive exposure of log-linear models is Christensen (1997). Let  $\mathbf{Z}$  be a random vector of categorical random variables  $Z_\ell, \ell = 0, \dots, m$ , i.e.,  $\mathbf{Z} = (Z_0, Z_1, \dots, Z_m)^\top$ . It is completely characterized by its distribution

$$p_\kappa = P_{\mathbf{Z}}(s_\kappa) = P(\mathbf{Z} = s_\kappa) = P((Z_0, \dots, Z_m) = (s_{k_0}, \dots, s_{k_m}))$$

with the multi-index  $\kappa = (k_0, \dots, k_m)$ , where  $s_{k_\ell}$  with  $k_\ell = 1, \dots, K_\ell$  denotes all possible categories of the categorical random variable  $Z_\ell, \ell = 0, \dots, m$ . Since it is assumed that there is a total of  $K_\ell$  different categories with  $P_{Z_\ell}(s_{k_\ell}) > 0$ , there is a total of  $\prod_{\ell=0}^m K_\ell$  different categorical states for  $\mathbf{Z} = \bigotimes_{\ell=0}^m Z_\ell$ .

The distribution of a categorical random vector may initially be thought of as being provided by contingency tables. More conveniently, the distribution of a categorical random vector  $\mathbf{Z}$  can generally be written in terms of a log-linear model as

$$\log p_\kappa = \sum_{\kappa} w_\kappa f_{\mathbf{Z}}^\kappa(\mathbf{z})$$

with

$$w_\kappa = \log p_\kappa, \\ f_{\mathbf{Z}}^\kappa(\mathbf{z}) = \mathbf{I}_{\{s_\kappa\}}(\mathbf{z}) = \mathbf{I}_{\{s_{k_0}, \dots, s_{k_m}\}}(z_0, \dots, z_m).$$

## 3.3 Independence, Conditional Independence of Random Variables

If the random variables  $Z_\ell, \ell = 1, \dots, m$ , are independent, then the joint probability of any subset of random variables  $Z_\ell$  can be factorized into the product of the individual probabilities, i.e.,

$$P_{\bigotimes_{\ell \in M} Z_\ell} = \bigotimes_{\ell \in M} P_{Z_\ell}.$$

where  $M$  denotes any non-empty subset of the set  $\{1, \dots, m\}$ . In particular

$$P_{\mathbf{Z}} = P_{\bigotimes_{\ell=1}^m Z_\ell} = \bigotimes_{\ell=1}^m P_{Z_\ell}.$$

If the random variables  $Z_\ell, \ell = 1, \dots, m$ , are conditionally independent given  $Z_0$ , then the joint conditional probability of any subset of random variables  $Z_\ell$  given  $Z_0$  can be factorized into the product of the individual conditional probabilities, i.e.,

$$P_{\otimes_{\ell \in M} Z_\ell | Z_0} = \bigotimes_{\ell \in M} P_{Z_\ell | Z_0}, \quad (3.1)$$

and in particular

$$P_{\otimes_{\ell=1}^m Z_\ell | Z_0} = \bigotimes_{\ell=1}^m P_{Z_\ell | Z_0}.$$

### 3.4 Logistic Regression, and Its Special Case of Weights-of-Evidence

Conditional expectation of a dichotomous random target variable  $Z_0$  given a  $m$ -variate random predictor vector  $\mathbf{Z} = (Z_1, \dots, Z_m)^\top$  is equal to a conditional probability, i.e.,

$$E(Z_0 | \mathbf{Z}) = P(Z_0 = 1 | \mathbf{Z}).$$

Then the ordinary logistic regression model (without interaction terms) neglecting the error term yields

$$\text{logit}P(Z_0 = 1 | \mathbf{Z}) = \beta_0 + \boldsymbol{\beta}^\top \mathbf{Z}, \quad \beta_0 \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^m.$$

Omitting the error term it can be rewritten in terms of a probability as

$$P(Z_0 = 1 | \mathbf{Z}) = \Lambda(\beta_0 + \boldsymbol{\beta}^\top \mathbf{Z}),$$

where  $\Lambda$  denotes the logistic function. The logistic regression model with interaction terms reads in terms of a logit transformed probability

$$\text{logit}P(Z_0 = 1 | \mathbf{Z}) = \beta_0 + \sum_{\ell} \beta_{\ell} Z_{\ell} + \sum_{\ell_i, \dots, \ell_j} \beta_{\ell_i, \dots, \ell_j} Z_{\ell_i} \dots Z_{\ell_j}, \quad (3.2)$$

and in terms of a probability

$$P(Z_0 = 1 | \mathbf{Z}) = \Lambda \left( \beta_0 + \sum_{\ell} \beta_{\ell} Z_{\ell} + \sum_{\ell_i, \dots, \ell_j} \beta_{\ell_i, \dots, \ell_j} Z_{\ell_i} \dots Z_{\ell_j} \right).$$

If all predictor variables are dichotomous variables and conditionally independent given the target variable then the parameters of the ordinary logistic regression model simplify to

$$\beta_0 = \text{logit}P(Z_0 = 1) + W^{(0)}, \quad \beta_{\ell} = C_{\ell}, \quad \ell = 1, \dots, m,$$

with contrasts

$$C_\ell = W_\ell^{(1)} - W_\ell^{(0)}, \quad \ell = 1, \dots, m,$$

defined as differences of weights of evidence

$$W_\ell^{(1)} = \ln \frac{P(Z_\ell = 1 \mid Z_0 = 1)}{P(Z_\ell = 1 \mid Z_0 = 0)}, \quad W_\ell^{(0)} = \ln \frac{P(Z_\ell = 0 \mid Z_0 = 1)}{P(Z_\ell = 0 \mid Z_0 = 0)},$$

and with  $W^{(0)} = \sum_{\ell=1}^m W_\ell^{(0)}$  provided all conditional probabilities are different from 0 (Schaeben 2014b). Obviously the model parameters become independent of one another, and can be estimated by mere counting. This special case of a logistic regression model is usually referred to as the method of “weights-of-evidence”. In turn, the canonical generalization of Bayesian weights-of-evidence is logistic regression.

That weights of evidence  $W_\ell$  agree with the logistic regression parameters  $\beta_\ell$  in case of joint conditional independence becomes obvious when recalling

$$\begin{aligned} C_\ell &= W_\ell^{(1)} - W_\ell^{(0)} \\ &= \ln \frac{P(Z_\ell = 1 \mid Z_0 = 1)}{P(Z_\ell = 1 \mid Z_0 = 0)} - \ln \frac{P(Z_\ell = 0 \mid Z_0 = 1)}{P(Z_\ell = 0 \mid Z_0 = 0)} \\ &= \ln \left( \frac{O(Z_0 = 1 \mid Z_\ell = 1)}{O(Z_0 = 1 \mid Z_\ell = 0)} \right) = \beta_\ell, \end{aligned}$$

which is the log odds ratio, the usual interpretation of  $\beta_\ell$  (Hosmer and Lemeshow 2000).

If  $\mathbf{Z}$  comprises  $m$  dichotomous predictor variables  $Z_\ell, \ell = 1, \dots, m$ , there are  $2^m$  possible different realizations  $\mathbf{z}_k, k = 1, \dots, 2^m$ , of  $\mathbf{Z}$ . Then

$$\begin{aligned} \sum_{i=1}^n \hat{P}(Z_0 = 1 \mid \mathbf{Z} = \mathbf{z}(i)) &= \sum_{k=1}^{2^m} \hat{P}(Z_0 = 1 \mid \mathbf{Z} = \mathbf{z}_k) H(\mathbf{Z} = \mathbf{z}_k) \\ &= \sum_{k=1}^{2^m} \hat{P}(Z_0 = 1 \mid \mathbf{Z} = \mathbf{z}_k) n \hat{P}(\mathbf{Z} = \mathbf{z}_k) \\ &= n \hat{P}(Z_0 = 1) = \sum_{i=1}^n z_0(i), \end{aligned}$$

where the last equation is an application of the formula of total probability. It is a constitutive equation to estimate the parameters of a logistic regression model and holds always for fitted logistic regression models. With respect to weights-of-evidence, the test statistic of the so-called “new omnibus test” of conditional independence (Agterberg and Cheng 2002) is

$$t = \sum_{i=1}^n \left( \widehat{P}(Z_0 = 1 \mid \mathbf{Z} = \mathbf{z}(i)) - z_0(i) \right)$$

and should not be too large for conditional independence to be reasonably assumed.

### 3.5 Hammersley–Clifford Theorem

Rephrasing the proper statement (Lauritzen 1996) casually, the Hammersley–Clifford Theorem states that a probability distribution with a positive density satisfies one of the Markov properties with respect to an undirected graph  $G$  if and only if its density can be factorized over the cliques of the graph. Since the distribution of a categorical random vector can be represented in terms of a log-linear model, Hammersley–Clifford theorem applies. Given  $(m + 1)$  random variables  $Z_0, \dots, Z_m$ , there is a total of  $\binom{m+1}{\ell+1}$  different product terms each involving  $(\ell + 1)$  variables,  $\ell = 0, \dots, m$ , summing to a total of  $\sum_{\ell=0}^m \binom{m+1}{\ell+1} = 2^{m+1} - 1$  different terms. Thus there is a total of  $(m + 1)$  single variable terms, and a total of  $2^{m+1} - (m + 2)$  multi variable terms.

The full log-linear model encompasses all terms and reads

$$\log p_{\kappa} = \sum_{\ell=0}^m \sum_{\alpha \in C_{\ell+1}^{m+1}} \sum_{\kappa(\alpha)} \phi_{\kappa(\alpha)} \mathbf{I}_{s_{\kappa(\alpha)}}(\mathbf{z}_{\kappa(\alpha)}) \quad (3.3)$$

where  $\alpha \in C_{\ell+1}^{m+1}$  denotes an  $(\ell + 1)$ -combination of the set  $\{1, \dots, m + 1\} \subset \mathbb{N}$ , and  $\kappa(\alpha) = (k_{i_1}, \dots, k_{i_{\ell+1}})$  denotes a multi-index with  $(\ell + 1)$  entries  $k_{i_\ell} = 1, \dots, K_{i_\ell}$ , for  $\ell = 0, \dots, m$ . The random vector  $\mathbf{Z}_{\kappa(\alpha)}$  is the product of any tuple of  $(\ell + 1)$  components of  $\mathbf{Z}$ , the total number of which is  $\binom{m+1}{\ell+1}$ .

Assumptions of independence or conditional independence simplify the distribution of  $\mathbf{Z}$ , i.e., its full log-linear model, considerably. Assuming independence for all its components  $Z_\ell, \ell = 0, \dots, m$ , the log-linear model simplifies according to Eq. (3.1) to

$$\log p_{\kappa} = \sum_{\ell=0}^m \log p_{k_\ell} = \sum_{\ell=0}^m \sum_{k_\ell=1}^{K_\ell} \phi_{k_\ell} \mathbf{I}_{\{s_{k_\ell}\}}(z_\ell), \quad (3.4)$$

where  $\phi_{k_\ell} = \log p_{k_\ell}$ .

Assuming joint conditional independence of all components  $Z_\ell, \ell = 1, \dots, m$ , given  $Z_0$ , the log-linear model, Eq. (3.3), simplifies according to Eq. (3.1) to

$$\log p_{\kappa} = \sum_{\ell=0}^m \sum_{k_\ell=1}^{K_\ell} \phi_{k_\ell} \mathbf{I}_{\{s_{k_\ell}\}}(z_\ell) + \sum_{\ell=1}^m \sum_{\alpha \in \{0, \ell\}} \sum_{\kappa(\alpha)} \phi_{\kappa(\alpha)} \mathbf{I}_{\{s_{\kappa(\alpha)}\}}(\mathbf{z}_{\kappa(\alpha)}). \quad (3.5)$$

Thus the latter model, Eq. (3.5), assuming conditional independence differs from the model for independence, Eq. (3.4), in the additional product terms  $Z_0 \otimes Z_\ell$ ,  $\ell = 1, \dots, m$ .

Any violation of joint conditional independence given  $Z_0$  results in additional cliques of the graph and in additional product terms. Assuming that conditional independence given  $Z_0$  does not hold for a particular subset  $Z_{\ell_1}, \dots, Z_{\ell_k}$  of variables  $Z_\ell$  results in an enlarging of the log-linear model of Eq. (3.5) by additional terms referring to  $Z_0 \otimes \bigotimes_{\ell=1}^k \bigotimes_{\ell_i \in C_\ell^k} Z_{\ell_i}$  and  $\bigotimes_{\ell=1}^k \bigotimes_{\ell_i \in C_\ell^k} Z_{\ell_i}$ , respectively.

### 3.6 Testing Joint Conditional Independence of Categorical Random Variables

The statistic of the likelihood ratio test (Neyman and Pearson 1933; Casella and Berger 2001) is the ratio of the maximized likelihood of a restricted model and the maximized likelihood of the full model. The assumption of the likelihood ratio test concerns the choice of the model family of distributions.

The null-hypothesis is that a given log-linear model is sufficiently large to represent the joint distribution. If the random variables are categorical, the full log-linear model is always sufficiently large as was explicitly shown above. More interesting are tests whether a smaller log-linear model is sufficiently large. Testing the null-hypothesis whether a log-linear model encompassing one-variable and two-variable terms, all of which involve  $Z_0$ , is sufficiently large provides a test of conditional independence of all  $Z_\ell$ ,  $\ell = 1, \dots, m$ , given  $Z_0$  because this log-linear model is sufficiently large in case of conditional independence given  $Z_0$ . Thus, a reasonable rejection of the initial null-hypothesis implies a reasonable rejection of the assumption of conditional independence given  $Z_0$ .

### 3.7 Conditional Distribution, Logistic Regression

Since the joint distribution implies all marginal and conditional distribution, respectively, the conditional distribution

$$P_{Z_0 | \bigotimes_{\ell=1}^m Z_\ell} = \frac{P_{\bigotimes_{\ell=0}^m Z_\ell}}{P_{\bigotimes_{\ell=1}^m Z_\ell}} \quad (3.6)$$

is explicitly given here by

$$\frac{P_{\bigotimes_{\ell=0}^m Z_\ell}(s_{k_0}, \dots, s_{k_\ell})}{P_{\bigotimes_{\ell=1}^m Z_\ell}(s_{k_1}, \dots, s_{k_\ell})} = \frac{P_{\bigotimes_{\ell=0}^m Z_\ell}(s_{k_0}, \dots, s_{k_\ell})}{\sum_{k_0=1}^{K_0} P_{\bigotimes_{\ell=0}^m Z_\ell}(s_{k_0}, s_{k_1}, \dots, s_{k_\ell})}.$$

Assuming independence, Eq. (3.6) immediately reveals

$$P_{Z_0|\otimes_{\ell=1}^m Z_\ell} = P_{Z_0}.$$

Assuming conditional independence of all  $Z_\ell, \ell = 1, \dots, m$ , given  $Z_0$  and further that  $Z_0$  is dichotomous, then

$$P_{Z_0|\otimes_{\ell=1}^m Z_\ell}(1 | s_{k_1}, \dots, s_{k_\ell}) = \frac{P_{\otimes_{\ell=0}^m Z_\ell}(1, s_{k_1}, \dots, s_{k_\ell})}{\sum_{i=0}^1 P_{\otimes_{\ell=0}^m Z_\ell}(i, s_{k_1}, \dots, s_{k_\ell})} \quad (3.7)$$

with

$$P_{\otimes_{\ell=0}^m Z_\ell}(1, s_{k_1}, \dots, s_{k_\ell}) = \exp\left(\phi_1 + \sum_{\ell=1}^m \phi_{k_\ell} + \sum_{\ell=1}^m \sum_{k_\ell=1}^{K_\ell} \phi_{1,k_\ell}\right)$$

and

$$\sum_{i=0}^1 P_{\otimes_{\ell=0}^m Z_\ell}(i, s_{k_1}, \dots, s_{k_\ell}) = \sum_{i=0}^1 \exp\left(\phi_i + \sum_{\ell=1}^m \phi_{k_\ell} + \sum_{\ell=1}^m \sum_{k_\ell=1}^{K_\ell} \phi_{i,k_\ell}\right)$$

Thus,

$$\begin{aligned} & \frac{P_{\otimes_{\ell=0}^m Z_\ell}(1, s_{k_1}, \dots, s_{k_\ell})}{\sum_{s=0}^1 P_{\otimes_{\ell=0}^m Z_\ell}(s, s_{k_1}, \dots, s_{k_\ell})} \\ &= \frac{\exp\left(\phi_1 \mathbf{I}_{\{1\}}(1) + \sum_{\ell=1}^m \phi_{1,k_\ell} \mathbf{I}_{\{1, s_{k_\ell}\}}(1, Z_\ell)\right)}{\sum_{s=0}^1 \exp\left(\phi_s \mathbf{I}_{\{s\}}(1) + \sum_{\ell=1}^m \phi_{k_\ell} \mathbf{I}_{\{s, s_{k_\ell}\}}(1, Z_\ell)\right)} \\ &= \frac{\exp\left(\phi_1 + \sum_{\ell=1}^m \phi_{1, s_{k_\ell}} \mathbf{I}_{\{s_{k_\ell}\}}(Z_\ell)\right)}{1 + \exp\left(\phi_1 + \sum_{\ell=1}^m \phi_{1, k_\ell} \mathbf{I}_{\{s_{k_\ell}\}}(Z_\ell)\right)} \\ &= \Lambda\left(\phi_1 + \sum_{\ell=1}^m \phi_{1, k_\ell} \mathbf{I}_{\{s_{k_\ell}\}}(Z_\ell)\right). \end{aligned}$$

Finally,

$$P_{Z_0|\otimes_{\ell=1}^m Z_\ell} = \Lambda\left(\beta_0 + \sum_{\ell=1}^m \beta_\ell Z_\ell\right),$$

which is obviously logistic regression

$$\text{logit} P_{Z_0|\otimes_{\ell=1}^m Z_\ell} = \beta_0 + \sum_{\ell=1}^m \beta_\ell Z_\ell. \quad (3.8)$$

It should be noted that additional product terms in the joint probability  $P_{\otimes_{\ell=0}^m Z_{\ell}}$  on the right hand side of Eq. (3.7) of the form  $\otimes_{\ell=1}^k \otimes_{\ell_i \in C_{\ell}^k} Z_{\ell_i}$  including  $Z_{\ell}$ ,  $\ell = 1, \dots, m$ , only, i.e., not including  $Z_0$ , would not effect the form of the conditional probability, Eq. (3.8). Additional product terms of the form  $Z_0 \otimes \otimes_{\ell=1}^k \otimes_{\ell_i \in C_{\ell}^k} Z_{\ell_i}$ , i.e., including  $Z_0$ , result in a logistic regression model with interaction terms, Eq. (3.2).

Ordinary logistic regression is optimum, if the joint probability of the (dichotomous) target variable and the predictor variables is of log-linear form and all predictor variables are jointly conditionally independent given the target variable; in particular, it is optimum if the predictor variables are categorical and jointly conditionally independent given the target variable (Schaeben 2014a). Logistic regression with interaction terms is optimum, if the joint probability of the (dichotomous) target variable and the predictor variables is of log-linear form and the interaction terms correspond to lacking conditionally independence given the target variable; for categorical predictor variables, interaction terms can compensate for any lack of conditional independence exactly. Logistic regression with interaction terms is optimum in case of lacking conditional independence (Schaeben 2014a).

### 3.8 Practical Applications

The practical application of the log-likelihood ratio test of joint conditional independence generally includes the following steps

- test the null-hypothesis that the full log-linear model is sufficiently large to represent the joint probability of all predictor variables and the target variables;
- if the first null-hypothesis is not reasonably rejected, test the null-hypotheses that smaller log-linear models are sufficiently large; in particular;
- test the null hypothesis that the log-linear model without any interaction term is sufficiently large;
- if the final null-hypothesis is rejected, then the predictor variables must not be assumed to be jointly conditionally independent given the target variable.

#### 3.8.1 Practical Application with Fabricated Indicator Data

##### 3.8.1.1 The Data Set BRY

The data set BRY is derived from the [https://en.wikipedia.org/wiki/Conditional\\_independence](https://en.wikipedia.org/wiki/Conditional_independence). Initially it comprises three random events  $B, R, Y$ , denoting the subsets of the set of all 49 pixels which are blue, red or yellow with given probabilities  $P(B) = \frac{18}{49} = 0.367$ ,  $P(R) = \frac{16}{49} = 0.326$ ,  $P(Y) = \frac{12}{49} = 0.244$ . The random events

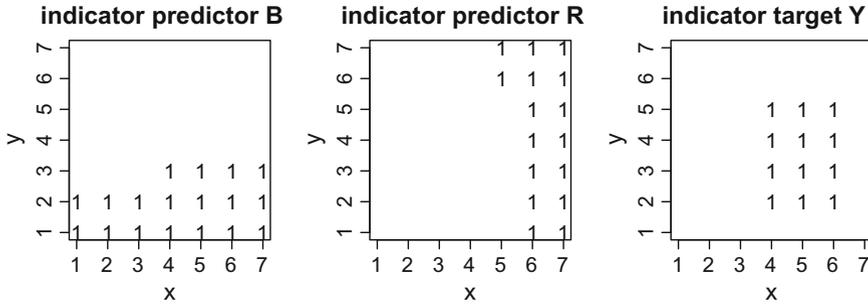


Fig. 3.1 Map images of random events  $B, R, Y$ .

$B, R, Y$  are distinguished from their corresponding random indicator variables  $\mathbf{B}, \mathbf{R}, \mathbf{Y}$  defined as usually, e.g.,

$$\mathbf{B}(\omega) = \mathbf{I}_B(\omega), \omega \in \Omega,$$

where  $\mathbf{I}$  denotes the indicator variable. They are assigned to pixels of a  $7 \times 7$  digital map image, Fig. 3.1.

It should be noted that in this example any spatial references are solely owed to the purpose of visualization as map images, and that the test itself does not take any spatial references or spatially induced dependences into account.

Checking independence according to its definition in reference to random events, the figures

$$P(B \cap R) = 0.122, \quad P(B) P(R) = 0.119$$

indicate that the random events  $B$  and  $R$  are not independent. However, the deviation is small.

Next, conditional independence is checked in terms of its definition referring to random events. Since conditional independence of the random events  $B$  and  $R$  given  $Y$  does not imply conditional independence of the random events  $B$  and  $R$  given the complement  $\mathcal{C}Y$ , two checks are required. The results are

$$P(B \cap R | Y) = \frac{1}{6} = P(B | Y) P(R | Y)$$

$$P(B \cap R | \mathcal{C}Y) = \frac{4}{37} \neq \left(\frac{12}{37}\right)^2 = P(B | \mathcal{C}Y) P(R | \mathcal{C}Y),$$

and indicate that the random events  $B$  and  $R$  are conditionally independent given the random event  $Y$ , but that they are not conditionally independent given the complement  $\mathcal{C}Y$ . It should be noted that the deviation of the joint conditional probability and the product of the two individual conditional probabilities in terms of their ratio is 1.027. In fact, the events  $B$  and  $R$  are conditionally independent given either  $Y$  or  $\mathcal{C}Y$  if one white pixel, e.g. pixel (1,7) with  $\mathbf{B} = \mathbf{R} = \mathbf{Y} = 0$ , is omitted.

Generalizing the view to random variables  $B, R, Y$  and their unique joint realization as shown in Fig. 3.1, Pearson's  $\chi^2$  test with Yates' continuity correction of the null-hypothesis of independence of the random variables  $B$  and  $R$  given the data returns a  $p$ -value of 1 indicating that the null-hypothesis cannot reasonably be rejected.

The likelihood ratio test is applied with respect to the log-linear distribution corresponding to the null-hypothesis of conditional independence and results in a  $p$ -value of 0.996 indicating that the null-hypothesis cannot reasonably be rejected.

Thus, given the data the tests suggest to infer that the random variables  $B$  and  $R$  are independent and conditionally independent given the random variable  $Y$ .

### 3.8.1.2 The Data Set SCCI

The next data set SCCI comprises three random events  $B_1, B_2, T$  with given probabilities  $P(B_1) = P(B_2) = P(T) = \frac{7}{49} = \frac{7}{49} = 0.142$ . They are assigned to pixels of a  $7 \times 7$  digital map image, Fig. 3.2.

Checking independence according to its definition for random events, the figures

$$P(B_1 \cap B_2) = 0.102, \quad P(B_1) P(B_2) = 0.020$$

indicate that the random events  $B_1$  and  $B_2$  are not independent.

Next, conditional independence is checked in terms of its definition referring to random events. Since conditional independence of the random events  $B_1$  and  $B_2$  given  $T$  does not imply conditional independence of the random events  $B_1$  and  $B_2$  given  $\complement T$ , two checks are required. The results are

$$P(B_1 \cap B_2 | T) = 0.714 \neq 0.734 = P(B_1 | T) P(B_2 | T)$$

$$P(B_1 \cap B_2 | \complement T) = 0 \neq 0.0005 = P(B_1 | \complement T) P(B_2 | \complement T),$$

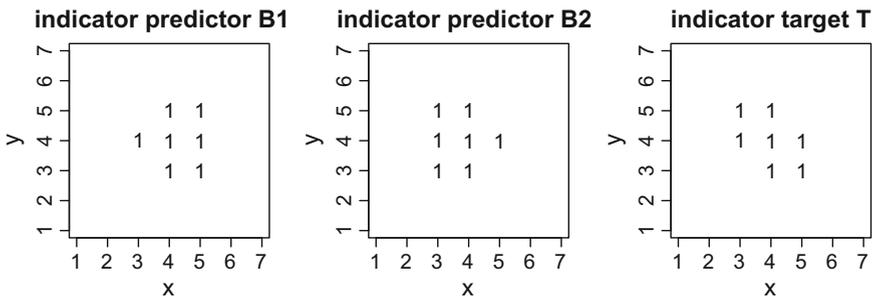


Fig. 3.2 Map images of random events  $B_1, B_2, T$  with  $P(B_1) = P(B_2) = P(T) = \frac{7}{49} = \frac{7}{49} = 0.142$ .

and indicate that the random events  $B_1$  and  $B_2$  are neither conditionally independent given the random event  $T$  nor given the complement  $\bar{T}$ .

Testing the null-hypothesis of independence of the random variables  $B_1$  and  $B_2$  with Pearson's  $\chi^2$  test with Yates' continuity correction given the data returns a  $p$ -value of practically equal to 0 indicating that the null-hypothesis should be rejected. The likelihood ratio test is applied with respect to the log-linear distribution corresponding to the null-hypothesis of conditional independence and results in a  $p$ -value of 0.825 indicating that the null-hypothesis cannot reasonably be rejected.

Thus, given the data the tests imply that the random variables  $B_1$  and  $B_2$  are not independent but conditionally independent given the random variable  $T$ .

### 3.9 Discussion and Conclusions

Since pairwise conditional independence does not imply joint conditional independence, the  $\chi^2$ -test (Bonham-Carter 1994) of independence given  $Z_0 = 1$  does not apply to checking the modeling assumption of weights-of-evidence. The disadvantage of both the "omnibus" test (Bonham-Carter 1994) and the "new omnibus" test (Agterberg and Cheng 2002) is twofold. First, it involves an assumption of normal distribution which itself should be subject to a test. Second, weights-of-evidence has to be applied to calculate the test statistic which is the sum of all predicted conditional probabilities within the training data set. If the test actually suggests rejection of the null-hypothesis of conditional independence, the user learns that the application of weights-of-evidence was not mathematically authorized to predict the conditional probabilities. The standard likelihood ratio test suggested here resolves both shortcomings.

**Acknowledgements** The author would like to thank Prof. Juanjo Egozcue, UPC Barcelona, Spain, and Prof. K. Gerald van den Boogaart, HIF, Germany, for emphatic and instructive discussions of conditional independence.

### References

- Agterberg FP, Cheng Q (2002) Conditional independence test for weights-of-evidence modeling. *Nat Resour Res* 11:249–255
- Bergsma, WP (2004) Testing conditional independence for continuous random variables. *Eurandom Report Vol. 2004048*, Eindhoven
- Bergsma WP (2010) Nonparametric testing of conditional independence by means of the partial copula, SSRN: <https://ssrn.com/abstract=1702981>, <https://doi.org/10.2139/ssrn.1702981>
- Bonham-Carter GF (1994) *Geographic information systems for geoscientists: modeling with GIS*. Elsevier, Pergamon
- Bonham-Carter GF, Agterberg FP, Wright DF (1989) Weights of evidence modeling: a new approach to mapping mineral potential. In: Agterberg FP, Bonham-Carter GF (eds.) *Statistical Applications in the Earth Sciences*, Geological Survey of Canada, Paper 89-9, pp. 171–183

- Bouezmarni T, Rombouts JVK, Taamouti A (2012) Nonparametric copula-based test for conditional independence with applications to granger causality. *J Bus Econ Stat* 30:275–287
- Butz CJ, Sanscartier MJ (2002) Properties of weak conditional independence. In: Alpigini JJ, Peters JF, Skowron A, Zhong N (eds) *Rough Sets and Current Trends in Computing: third international conference, RSCTC 2002 Malvern, PA, USA, 14–16 October, 2002*, vol 2475. *Proceedings Lecture Notes in Computer Science*. Springer, Berlin, pp 349–356
- Casella G, Berger RL (2001) *Statistical inference*, 2nd edn. Duxbury Thomson Learning
- Chalak K, White H (2012) Causality, conditional independence, and graphical separation in settable systems. *Neural Comput* 24:1611–1668
- Cheng Q (2015) BoostWofE: A new sequential weights of evidence model reducing the effect of conditional dependency. *Math Geosci* 47:591–621
- Christensen R (1997) *Log-Linear models and logistic regression*, 2nd edn. Springer, Berlin
- Constantinou P, Dawid AP (2015) Extended conditional independence and applications in causal inference. [arXiv:1512.00245](https://arxiv.org/abs/1512.00245)
- Dawid AP (1979) Conditional independence in statistical theory. *J R Stat Soc B* 41:1–31
- Dawid AP (2004) Probability, causality and the empirical world: A Bayes-de Finetti-Popper-Borel synthesis. *Stat Sci* 19:44–57
- Dawid AP (2007) *Fundamentals of statistical causality*. Research Report 279, Department of Statistical Science, University College London
- Doran G, Muandet K, Zhang K, Schölkopf B (2014) A permutation-based kernel conditional independence test. *Proceedings of UAI*
- ESRI: ArcSDM3.1 User Guide Spatial Data Modeller 3 Extension for ArcMap 9.1 file: [http://C:/arcgis/ArcSDM/Documentation/sdmrpn.htm\(9of16\)2/22/200612:05:17PM](http://C:/arcgis/ArcSDM/Documentation/sdmrpn.htm(9of16)2/22/200612:05:17PM)
- Good IJ (1950) *Probability and the weighing of evidence*. Griffin, London
- Good IJ (1960) Weight of evidence, corroboration, explanatory power, information and the utility of experiments. *J R Stat Soc B* 22:319–331
- Good IJ (1985) Weight of evidence: A brief survey. *Bayesian Stat* 2:249–270
- Györfi L, Walk H (2012) Strongly consistent nonparametric tests of conditional independence. *Stat Probab Lett* 82:1145–1150
- Hosmer DW, Lemeshow S (2000) *Applied logistic regression*, 2nd edn. Wiley, New Jersey
- Huang T-M (2010) Testing conditional independence using maximal nonlinear conditional correlation. *Ann Stat* 38:2047–2091
- Huang M, Sun Y, White H (2016) A flexible nonparametric test for conditional independence. *Econom. Theory* 32:1434–1482
- Lauritzen SL (1996) *Graphical models*. Clarendon Press, Oxford
- Neyman J, Pearson ES (1933) On the problem of the most efficient tests of statistical hypotheses. *Philos Trans R Soc Lond A* 231:289–337
- Pearl J (2009) *Causality: models, reasoning, and inference*, 2nd edn. Cambridge University Press, New York
- Ramsey JD (2014) A scalable conditional independence test for nonlinear, non Gaussian data. [arXiv:1401.5031](https://arxiv.org/abs/1401.5031)
- Schaeben H (2014a) A mathematical view of weights-of-evidence, conditional independence, and logistic regression in terms of Markov random fields. *Math Geosci* 46:691–709
- Schaeben H (2014b) Potential modeling: conditional independence matters. *Int J Geomath* 5:99–116
- Song K (2009) Testing conditional independence via Rosenblatt transforms. *Ann Statist* 37:4011–4045
- Spohn W (1980) Stochastic independence, causal independence, and shieldability. *J Philos Log* 9:73–99
- Spohn W (1994) On the properties of conditional independence. In: Suppes P, Humphreys P (eds) *Scientific Philosopher*, vol 1. *Probability and Probabilistic Causality*. Kluwer, Dordrecht, pp 173–194

- Su L, White H (2007) A consistent characteristic function-based test for conditional independence. *J Econom* 141:807–834
- Su L, White H (2008) A nonparametric Hellinger metric test for conditional independence. *Econo Theory* 24:829–864
- Suppes P (1970) *A probabilistic theory of causality*. North-Holland, Amsterdam
- Wong SKM, Butz CJ (1999) Contextual weak independence in Bayesian networks. In: Fifteenth conference on uncertainty in artificial intelligence, pp. 670–679
- Zhang K, Peters J, Janzing D, Schölkopf B (2011) Kernel-based conditional independence test and application in causal discovery. In: Cozman FG, Pfeffer A (eds) *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI 2011)*, Barcelona, Spain, July 14–17, 2011. AUAI Press, Corvallis, OR, USA, pp 804–813

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

