# Bayesian inference for marginal models under equality and inequality constraints

Luisa Scaccia
Dip. di Istituzioni Economiche e Finanziarie
Università di Macerata
Via Crescimbeni, 20 - 62100 Macerata
scaccia@unimc.it

Francesco Bartolucci
Dip. di Economia, Finanza e Statistica
Università di Perugia
Via A. Pascoli, 20 - 06123 Perugia
bart@stat.unipg.it

**Abstract:** We develop a Bayesian framework for making inference on a class of marginal models for categorical variables, which is formulated through equality and/or inequality constraints on generalized logits, generalized log-odds ratios and similar higher-order interactions. A Markov chain Monte Carlo (MCMC) algorithm is used for parameter estimation and for computing the Bayes factor between competing models. The approach is illustrated through the application to a well-known dataset on social mobility.

**Keywords:** Encompassing priors, Generalised logits, MCMC, Positive association.

## 1 Introduction

An interesting class of models for the analysis of contingency tables is based on the multivariate logistic transform of Glonek and McCullagh (1995). These models parametrise the joint distribution of the variables of interest on the basis of marginal logits, marginal log-odds ratios and similar higher-order interactions. This parametrisation is then substantially different from a log-linear parametrisation. The class of models of Glonek and McCullagh (1995) can be extended in several ways. In particular, we deal here with models in which: (i) the parameters of the saturated model are given by *generalised* logits for each univariate marginal distribution, generalised log-odds ratios for each bivariate marginal distribution and similar interactions for each higher-order marginal distribution; (ii) linear equality and inequality constraints can be formulated on such parameters. Thus, several hypotheses of positive association can be expressed (Bartolucci *et al.*, 2001; Colombi and Forcina, 2001).

The literature on Bayesian inference for models such as those described above is rather scarce, especially when the parameters are subject to inequality constraints. In order to bridge this gap, we propose a Bayesian framework that is based on the Bayes factor (BF); see Kass and Raftery (1995). The BF has been recently used in categorical data analysis by several authors among which Dellaportas and Forster (1999) who propose a general framework for selecting a log-linear model under a multivariate Normal prior distribution on the parameters. Also in our framework we assume a multivariate Normal prior. However, our approach is more general in that it also allows for inequality constraints on the parameters. Furthermore, we use different numerical methods for computing the posterior distribution of the parameters and the BF between competing models.

The paper is organised as follows. In Section 2 we describe the class of models of interest and the prior distribution on the parameters. In Section 3 we deal with Bayesian estimation and model selection. Finally, in Section 4 we propose an application.

## 2 A class of marginal models for categorical variables

Let $\mathbf{A} = (A_1, \ldots, A_q)'$ be a vector of $q$ categorical variables, the $i$-th of which has support $\{1, \ldots, m_i\}$, and $\boldsymbol{\pi}$ a vector of length $r = \prod_i m_i$ with elements $p(\mathbf{A} = \mathbf{a})$ for every possible configuration $\mathbf{a}$ of $\mathbf{A}$. We describe a saturated parametrisation of $\boldsymbol{\pi}$ based on marginal logits, log-odds ratios and similar higher-order interactions, which generalises the one of Glonek and McCullagh (1995). The marginal logits may be of the following types:

- *local*: $\eta_i(a_i; l) = \log p(A_i = a_i + 1)/p(A_i = a_i)$;
- *global*: $\eta_i(a_i; g) = \log p(A_i \geq a_i + 1)/p(A_i \leq a_i)$;
- *continuation*: $\eta_i(a_i; c) = \log p(A_i \geq a_i + 1)/p(A_i = a_i)$;
- *reverse continuation*: $\eta_i(a_i; r) = \log p(A_i = a_i + 1)/p(A_i \leq a_i)$.

for $a_i = 1, \ldots, m_i - 1$. Local logits are appropriate for variables with non-ordered categories, whereas global and continuation logits are suitable for ordinal variables. Marginal log-odds ratios are defined as contrasts between conditional logits. For example, when logits of type $l$ and $g$ are used for $A_i$ and $A_j$, respectively, we obtain the *local-global* log-odds ratios $\eta_{ij}(a_i, a_j; l, g) = \eta_j(a_j; g | A_i = a_i + 1) - \eta_j(a_j; g | A_i = a_i)$, with $a_i = 1, \ldots, m_i - 1$ and $a_j = 1, \ldots, m_j - 1$. Similarly, three-way interactions are defined as contrasts between conditional log-odds ratios and so on for higher-order interactions. Once the type of logit has been chosen for each variable, all the marginal parameters are collected in a $(r-1)$-dimensional vector $\boldsymbol{\eta}$ which may be expressed as

$$\boldsymbol{\eta} = \mathbf{C} \log(\mathbf{M}\boldsymbol{\pi}), \tag{1}$$

where $\mathbf{C}$ and $\mathbf{M}$ are matrices whose construction is described in Colombi and Forcina (2001). Note that parametrisation (1) defines a saturated parametrisation of $\boldsymbol{\pi}$ which may be inverted by a simple iterative algorithm.

Several models may be formulated by equality and inequality constraints on $\boldsymbol{\eta}$ of type $\mathbf{E}\boldsymbol{\eta} = \mathbf{0}$, $\mathbf{U}\boldsymbol{\eta} \geq \mathbf{0}$. For instance, we can formulate the hypothesis of positive association between two variables, $A_1$ and $A_2$ say, by constraining all log-odds ratios to be non-negative. The type of association depends on the logits used: local logits for both variables lead to *Total Positivity of Order 2* ($TP_2$), whereas global logits determine the less stringent hypothesis of *Positive Quadrant Dependence* (PQD). Regardless the type of log-odds ratios, independence requires all of them to be 0. Moreover, if $A_1$ and $A_2$ have the same categories, we may formulate the hypothesis that $A_2$ is stochastically greater than $A_1$ or that of marginal homogeneity; see also Bartolucci *et al.* (2001).

For what concerns the prior distribution, we assume $\boldsymbol{\eta} \sim N(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta)$ when the model is saturated, i.e. neither equality nor inequality constraints are formulated on $\boldsymbol{\eta}$. The hyperparameters $\boldsymbol{\mu}_\eta$ and $\boldsymbol{\Sigma}_\eta$ are chosen as, respectively, the expected value and variance of $\boldsymbol{\eta} = \mathbf{C}\log(\mathbf{M}\boldsymbol{\pi})$, under a Dirichlet distribution with parameter $\mathbf{1}_r$ on $\boldsymbol{\pi}$. For a constrained model, we employ the *encompassing prior* strategy of Klugkist *et al.* (2005); see also Consonni and Veronese (2006). The rule underlying this strategy is that, once a prior distribution has been formulated on the parameters of a certain model, the prior for any nested model derives by conditioning on the corresponding parameter space. Consider first the case in which we only assume the equality constraints of type $\mathbf{E}\boldsymbol{\eta} = \mathbf{0}$. We can equivalently express this constraint as $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, with $\mathbf{X}$ being a suitable design matrix. The prior for $\boldsymbol{\beta}$ is automatically determined from that for $\boldsymbol{\eta}$ as $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, with $\boldsymbol{\mu}_\beta = \mathbf{X}'\boldsymbol{\mu}_\eta - \mathbf{X}'\boldsymbol{\Sigma}_\eta\mathbf{E}'(\mathbf{E}\boldsymbol{\Sigma}_\eta\mathbf{E}')^{-1}\mathbf{E}\boldsymbol{\mu}_\eta$ and $\boldsymbol{\Sigma}_\beta = \mathbf{X}'\boldsymbol{\Sigma}_\eta\mathbf{X} - \mathbf{X}'\boldsymbol{\Sigma}_\eta\mathbf{E}'(\mathbf{E}\boldsymbol{\Sigma}_\eta\mathbf{E}')^{-1}\mathbf{E}\boldsymbol{\Sigma}_\eta\mathbf{X}$. If also inequality constraints are used, the distribution is further conditioned in order to satisfy $\mathbf{U}\boldsymbol{\eta} = \mathbf{U}\mathbf{X}\boldsymbol{\beta} \geq \mathbf{0}$.

## 3 Bayesian estimation of the parameters and model selection

Let $\mathbf{y}$ be the vector of frequencies for the observed contingency table. The posterior distribution of the parameters under a certain model is obtained by an MCMC algorithm which is briefly illustrated in the following.

Consider first the saturated model. Under this model, the algorithm draws realizations from the posterior distribution of $\boldsymbol{\eta}$, $p(\boldsymbol{\eta}|\mathbf{y})$, by using the acceptance-rejection rule of Metropolis-Hastings (Metropolis *et al.*, 1953; Hastings, 1970) based on the proposal distribution $N(\boldsymbol{\nu}_\eta, \boldsymbol{\Omega}_\eta)$, with parameters $\boldsymbol{\nu}_\eta$ and $\boldsymbol{\Omega}_\eta$ chosen by a pilot chain. For a model formulated by the equality constraints $\mathbf{E}\boldsymbol{\eta} = \mathbf{0}$, which are equivalent to $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, we use a similar algorithm to sample from the posterior $p(\boldsymbol{\beta}|\mathbf{y})$. In this case, we use $N(\boldsymbol{\nu}_\beta, \boldsymbol{\Omega}_\beta)$ as proposal distribution, where $\boldsymbol{\nu}_\beta = \mathbf{X}'\boldsymbol{\nu}_\eta - \mathbf{X}'\boldsymbol{\Omega}_\eta\mathbf{E}'(\mathbf{E}\boldsymbol{\Omega}_\eta\mathbf{E}')^{-1}\mathbf{E}\boldsymbol{\nu}_\eta$ and $\boldsymbol{\Omega}_\beta = \mathbf{X}'\boldsymbol{\Omega}_\eta\mathbf{X} - \mathbf{X}'\boldsymbol{\Omega}_\eta\mathbf{E}'(\mathbf{E}\boldsymbol{\Omega}_\eta\mathbf{E}')^{-1}\mathbf{E}\boldsymbol{\Omega}_\eta\mathbf{X}$. When also inequality constraints are used, the proposal has to be properly conditioned.

An important issue is typically that of choosing a model in the set $\mathscr{M} = \{M_1, \ldots, M_K\}$ of competing models defined by different choices of the type of logits or of the equality and/or inequality constraints on the saturated parameter vector. For this aim, we rely on the BF that for two models, say $M_l$ and $M_k$, is defined as $B_{lk} = p(\mathbf{y}|l)/p(\mathbf{y}|k)$ where $p(\mathbf{y}|k)$ is the *marginal likelihood* of $M_k$. The larger $B_{lk}$, the greater the evidence in favour of $M_l$ with respect to $M_k$ (Kass and Raftery, 1995). In practice, when $B_{lk} > 1$, or equivalently $\log(B_{lk}) > 0$, model $M_l$ has to be preferred to $M_k$. To compare more than two models, it is convenient to single out a "reference" model, $M_1$ say, and compute the BF between any other model and this one. Note that the BF allows us to easily compare models parametrised through different types of logits, an otherwise cumbersome task in a likelihood-ratio approach. The latter may also be difficult to apply in the presence of nuisance parameters which, however, do not limit the use of the BF.

The BF cannot be computed analytically for the models at issue. Following the approach of Chib and Jeliazkov (2001), we obtain an estimate of the marginal likelihood of each model as a by-product of the Metropolis-Hastings algorithm, previously described.

## 4 An application

As an example we analyze the data in Table 1, which concern a sample of British males cross-classified according to their occupational status ($A_2$) and that of their father ($A_1$).

**Table 1**: *Father ($A_1$) and son ($A_2$) occupational status for a sample of British males.*

| | $A_2$ | | | | | |
|---|---|---|---|---|---|---|
| $A_1$ | I | II | III | IV | V | VI |
| I | 125 | 60 | 26 | 49 | 14 | 5 |
| II | 47 | 65 | 66 | 123 | 23 | 21 |
| III | 31 | 58 | 110 | 223 | 64 | 32 |
| IV | 50 | 114 | 185 | 715 | 258 | 189 |
| V | 6 | 19 | 40 | 179 | 143 | 71 |
| VI | 3 | 14 | 32 | 141 | 91 | 106 |

The data were already considered by Dardanoni and Forcina (1998) who, on the basis of a likelihood-ratio approach, concluded that the data conform to some forms of positive

association. However, they did not reach a conclusion about $TP_2$, due to presence of nuisance parameters, given by marginal column probabilities.
For these data we first compared the saturated model ($M_1$), the independence one ($M_2$) and those incorporating PQD ($M_3$) and $TP_2$ ($M_4$), obtaining:

$$\log(\hat{B}_{21}) = -357.49 \qquad \log(\hat{B}_{31}) = 4.79 \qquad \log(\hat{B}_{41}) = 33.88.$$

The hypothesis of independence must be definitely rejected, whereas that of positive association may not be rejected. The model incorporating $TP_2$ has to be preferred to that incorporating PQD. Thus, sons coming from a better family seem to have a better chance of success also conditional on remaining within any given subset of neighbouring classes. The hypothesis of uniform association has instead to be rejected: comparing model $M_5$, incorporating this constraint in addition to $TP_2$, with $M_4$, we obtained $\log(\hat{B}_{54}) = -31.86$. We also considered constraints on the marginal distributions: $M_6$ incorporates in $M_4$ the constraint that the marginal distributions of $A_1$ and $A_2$ are equal, while $M_7$ incorporates in $M_4$ the constraint that $A_2$ is stochastically greater than $A_1$. We obtained:

$$\log(\hat{B}_{64}) = -1.27 \qquad \log(\hat{B}_{74}) = 1.63.$$

The data seem to support $M_7$. Thus we can observe not only *pure mobility*, i.e. positive association between family's origin and the son's status, but also *structural mobility*, which refers to how far apart the two marginals are and is related to socioeconomic growth.

## References


Bartolucci F., Forcina A. and Dardanoni V. (2001) Positive quadrant dependence and marginal modelling in two-way tables with ordered margins, *Journal of the American Statistical Association*, 96, 1497–1505.

Chib S. and Jeliazkov I. (2001) Marginal likelihood from the metropolis-hastings output, *Journal of the American Statistical Association*, 96, 270–281.

Colombi R. and Forcina A. (2001) Marginal regression models for the analysis of positive association of ordinal response variables, *Biometrika*, 88, 1007–1019.

Consonni G. and Veronese P. (2006) Compatibility of prior specifications across linear models, Technical report, Università di Pavia.

Dardanoni V. and Forcina A. (1998) A unified approach to likelihood inference on stochastic orderings in a nonparametric context, *Journal of the American Statistical Association*, 93, 1112–1123.

Dellaportas P. and Forster J.J. (1999) Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models, *Biometrika*, 86, 615–633.

Glonek G. and McCullagh P. (1995) Multivariate logistic models, *Journal of the Royal Statistical Society, B*, 57, 533–546.

Hastings W.K. (1970) Monte Carlo sampling methods using markov chains and their applications, *Biometrika*, 57, 97–109.

Kass R.E. and Raftery A.E. (1995) Bayes factors, *Journal of the American Statistical Association*, 90, 773–795.

Klugkist I., Bernet K. and Herbert H. (2005) Bayesian model selection using encompassing priors, *Statistica Neerlandica*, 59, 57–69.

Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. and Teller E. (1953) Equations of state calculations by fast computing machines, *Journal of Chemistry Physics*, 21, 1087–1091.