

Consonni, Guido; Leucari, Valentina

Working Paper

Reference priors for discrete graphical models

Quaderni di Dipartimento - EPMQ, No. 181

Provided in Cooperation with:

University of Pavia, Department of Economics and Quantitative Methods (EPMQ)

Suggested Citation: Consonni, Guido; Leucari, Valentina (2005) : Reference priors for discrete graphical models, Quaderni di Dipartimento - EPMQ, No. 181, Università degli Studi di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi (EPMQ), Pavia

This Version is available at:

<https://hdl.handle.net/10419/87114>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

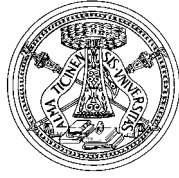
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Quaderni di Dipartimento

Reference priors for discrete graphical models

Guido Consonni
(University of Pavia)

Valentina Leucari
(University of Pavia)

181 (12-05)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Dicembre 2005

Reference priors for discrete graphical models

GUIDO CONSONNI VALENTINA LEUCARI

Dipartimento di Economia Politica e Metodi Quantitativi, University of Pavia,

Via S. Felice 5, 27100 Pavia, Italy

guido.consonni@unipv.it, vl@dimat.unipv.it

SUMMARY

The combination of graphical models and reference analysis represents a powerful tool for Bayesian inference in highly multivariate settings. It is typically difficult to derive reference priors in complex problems. In this paper we present a suitable mixed parameterisation for a discrete decomposable graphical model and derive the corresponding reference prior.

Some key words: Decomposable model; Fisher information matrix; Mixed parameterisation; Natural exponential family.

1 Introduction

This paper brings together two broad statistical methodologies, namely graphical models and reference analysis.

In this paper we deal with undirected graphical models, and assume that each variable, corresponding to a vertex in the graph, is discrete. Our results are therefore relevant also for the analysis of contingency tables under multinomial sampling.

A major breakthrough in the Bayesian analysis of graphical models was achieved with the paper of Dawid & Lauritzen (1993) wherein, among other things, the crucial notion of hyper-Markov laws was introduced. However, subjective elicitation of specified prior distributions for graphical models is still a very hard task, and it is important to develop default, or reference, priors. For informative accounts of reference analysis, see § 5.4 of Bernardo & Smith (1994) and the recent paper by Bernardo (2005).

Finding a reference prior in multiparameter settings is typically highly laborious. Moreover, its derivation and structure may depend on technical aspects that are difficult to justify from a substantive viewpoint. However, under some circumstances, the choice of a suitable parameterisation allows one to derive reference priors uniquely and straightforwardly.

For ease of exposition, in this paper we limit ourselves to the case of decomposable graphical models with binary variables, and focus on a specific parameterisation.

2 Parameterisations for discrete decomposable graphical models

2.1 *Graphical models*

In this section we present the basic prerequisites and notation on graphs and graphical models that are needed. We rely on, and refer to, Lauritzen (1996) for concepts and terminology; see also Cowell et al. (1999).

A graph is a pair $\mathcal{G} = (V, E)$, where V is a set of vertices and E is a set of edges between any two vertices. We only deal with undirected graphs, in which edges do not have a direction. To each vertex v there is associated a random variable X_v . If $A \subseteq V$ we let $X_A = \{X_v : v \in A\}$; for simplicity we set $X_V = X$. A graphical model $\mathcal{M}(\mathcal{G})$ is a family of distributions for X , satisfying the Markov property with respect to \mathcal{G} , namely

that X_u and X_v are conditionally independent given all the remaining variables, written $X_u \perp\!\!\!\perp X_v | X_{V \setminus \{u,v\}}$, whenever u and v are not connected by an edge in \mathcal{G} .

A graph is said to be complete if it possesses all possible edges; that is the corresponding graphical model, also named complete, embodies no conditional independence relationship. A maximal, with respect to inclusion, complete subset of V is a clique. We deal with decomposable graphs, and corresponding models, that are amenable to simpler statistical analysis and interpretation. A notable property of a decomposable graphical model is that the joint density of X , with respect to a product measure, can be factorised in terms of the clique marginal densities; that is

$$f(x) = \frac{\prod_{i=1}^k f_{C_i}(x_{C_i})}{\prod_{i=2}^k f_{S_i}(x_{S_i})}, \quad (1)$$

where the cliques C_1, \dots, C_k are assumed to be arranged in a perfect ordering, and S_2, \dots, S_k denote the corresponding separators, defined as $S_i = \{C_1 \cup \dots \cup C_{i-1}\} \cap C_i$. This will be tacitly understood in the sequel.

2.2 Exponential family representation

Throughout the paper we consider the special case of binary random variables; that is $X_v \in \{0, 1\}$ for each $v \in V$. The extension to the polytomous case is conceptually straightforward, although laborious. We shall offer some comments in § 4. The main theoretical results of this subsection, including notation and proofs, are in the unpublished 2004 University of Pavia Ph. D. thesis by V. Leucari.

We review below briefly a few essential facts about exponential families that shall be needed later on; for a general treatment, see Brown (1986). A recent informative account is provided by Casalis; see Chapter 54 of Kotz et al. (2000).

We regard vectors as column vectors, and denote by y^T the transpose of y . Let ν be a σ -finite positive measure on the Borel sets of \mathbb{R}^d . Suppose ν is not concentrated on an

affine hyperplane of \mathbb{R}^d and consider an exponential family \mathcal{F} whose densities with respect to ν are of the form

$$f(y|\theta) = \exp\{\theta^\top y - M(\theta)\}, \quad \theta \in \Theta, \quad (2)$$

with Θ non-empty. When Θ is the interior of the natural parameter space

$\left\{ \theta \in \mathbb{R}^d : \int \exp\{\theta^\top y\} \nu(dy) < \infty \right\}$, the family \mathcal{F} is said to be a natural exponential family on \mathbb{R}^d , with natural, or canonical, statistic Y and corresponding parameter θ .

The expectation μ of Y is

$$\mu = \mu(\theta) = E(Y|\theta) = \frac{\partial M(\theta)}{\partial \theta};$$

let $\Omega = \mu(\Theta)$ be the mean space. If Ω coincides with the interior of the convex hull of the support of ν , the family \mathcal{F} is said to be steep.

The function

$$V(\mu) = \text{var}(Y|\theta) = \left. \frac{\partial^2 M(\theta)}{\partial \theta^\top \partial \theta} \right|_{\theta=\theta(\mu)}, \quad \mu \in \Omega,$$

is called the variance function of the family \mathcal{F} . Here $\theta(\cdot)$ denotes the inverse of the mapping $\mu(\theta) = \partial M(\theta)/\partial \theta$.

Another useful concept is that of a cut (Barndorff-Nielsen, 1978, p. 50). Consider a general parametric family for the observable X , $f(x|\phi)$, with $\phi \in \Phi$. A statistic $S = s(X)$ is said to be a cut if there exists a parameterisation $\lambda(\phi) = \lambda$, $\lambda \in \Lambda = \lambda(\Phi)$, with $\lambda = (\lambda_1, \lambda_2)$, $\lambda_1 \in \Lambda_1$ and $\lambda_2 \in \Lambda_2$, such that $\Lambda = \Lambda_1 \times \Lambda_2$ and

$$f(x|\lambda) = f_1(s|\lambda_1) f_2(x|\lambda_2, s),$$

where $s = s(x)$. As a consequence λ_1 and λ_2 are likelihood independent.

It is known that undirected graphical models with no hidden variable are exponential families (Geiger et al., 2001). Here we are particularly concerned with alternative parameterisations of graphical models, and we find it useful to treat separately the complete and the general decomposable case.

Consider first the marginal complete model for clique C_i . We denote by \mathcal{C}_i the collection of all non-empty subsets of C_i , namely

$$\mathcal{C}_i = \{D : D \subseteq C_i, D \neq \emptyset\},$$

and similarly $\mathcal{S}_i = \{D : D \subseteq S_i, D \neq \emptyset\}$ for a separator S_i . Define the marginal cell probabilities for clique C_i as

$$p_{\mathcal{C}_i}^* = \{p_{\mathcal{C}_i}^*(D) : D \subseteq C_i, D \neq \emptyset\},$$

where

$$p_{\mathcal{C}_i}^*(D) = \text{pr}(X_D = 1, X_{C_i \setminus D} = 0 | p_{\mathcal{C}_i}^*).$$

In general we shall employ an asterisk to denote quantities, such as parameters, related to marginal models, as opposed to quantities pertaining to the joint distribution. At this stage we introduce a second parameterisation which is inherent in the exponential family representation.

Proposition 1. *Let $\mathcal{M}(\mathcal{G})$ be a discrete complete graphical model for the random variables $X \in \{0, 1\}^{|V|}$. Then the joint distribution of X is a natural exponential family,*

$$f(y|\theta) = \exp\{\theta^T y - M(\theta)\},$$

with $M(\theta) = \log \left[1 + \sum_{D \subseteq V, D \neq \emptyset} \exp \left\{ \sum_{F \subseteq D, F \neq \emptyset} \theta(D) \right\} \right]$. The canonical statistic is

$$Y = \{Y(D) : D \subseteq V, D \neq \emptyset\},$$

where

$$Y(D) = \prod_{v \in D} X_v,$$

and the canonical parameter is

$$\theta = \{\theta(D) : D \subseteq V, D \neq \emptyset\},$$

where

$$\theta(D) = \sum_{F \subseteq D} (-1)^{|D \setminus F|} \log p(F).$$

Proposition 2. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$. Then, for any complete subset $A \subseteq V$, $X_A = \{X_v : v \in A\}$ is a cut.*

Proof. From Theorem 5.4 in Frydenberg (1990), X_A is a cut if and only if $\mathcal{M}(\mathcal{G})$ is collapsible on to A . Asmussen & Edwards (1983) prove that a necessary and sufficient condition for collapsibility is completeness of the boundary of each connected component of $V \setminus A$. If A is complete the boundary of each connected component of $V \setminus A$ is also complete, being a subset of A . \square

Corollary 1. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$. Then, for $i = 1, \dots, k$, the marginal distribution of X_{C_i} is a natural exponential family,*

$$f_{C_i}(y_{C_i} | \theta_{C_i}^*) = \exp \left\{ \theta_{C_i}^{*\top} y_{C_i} - M_{C_i}(\theta_{C_i}^*) \right\}, \quad (3)$$

with $M_{C_i}(\theta_{C_i}^*) = \log \left[1 + \sum_{D \subseteq C_i, D \neq \emptyset} \exp \left\{ \sum_{F \subseteq D, F \neq \emptyset} \theta_{C_i}^*(F) \right\} \right]$. The canonical statistic is

$$Y_{C_i} = \{Y_{C_i}(D) : D \subseteq C_i, D \neq \emptyset\},$$

where

$$Y_{C_i}(D) = \prod_{v \in D} X_v,$$

and the canonical parameter is

$$\theta_{C_i}^* = \{\theta_{C_i}^*(D) : D \subseteq C_i, D \neq \emptyset\},$$

where

$$\theta_{C_i}^*(D) = \sum_{F \subseteq D} (-1)^{|D \setminus F|} \log p_{C_i}^*(F). \quad (4)$$

Proof. From Proposition 2 $\mathcal{M}(\mathcal{G})$ is collapsible on to each clique, and therefore each clique marginal model coincides with the family of Markov distributions over the corresponding subgraph. Such a distribution must then have the structure given in Proposition 1. \square

The mean parameter for the model in (3) is

$$\mu_{C_i} = \{\mu_{C_i}(D) : D \subseteq C_i, D \neq \emptyset\},$$

where

$$\mu_{C_i}(D) = E\{Y_{C_i}(D) | \theta_{C_i}^*\} = \sum_{F \subseteq C_i \setminus D} p_{C_i}^*(D \cup F). \quad (5)$$

Note that there is no need to use an asterisk for the mean parameter in the marginal model since this is just the corresponding subvector of the mean parameter in the joint model. An analogous comment holds for the canonical statistic Y_{C_i} . Clearly Corollary 1 holds more generally for any complete subset of V , and in particular for separators.

If we let $D \subseteq_{\mathcal{G}} V$ denote a complete, with respect to \mathcal{G} , subset of V , a result similar to that in Proposition 1 can be shown for a general, non-complete, decomposable model.

Proposition 3. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$. Then the joint distribution of X is a natural exponential family,*

$$f(y|\theta) = \exp\{\theta^T y - M(\theta)\}, \quad (6)$$

with $M(\theta) = \log \left[1 + \sum_{D \subseteq_{\mathcal{G}} V, D \neq \emptyset} \exp \left\{ \sum_{F \subseteq D, F \neq \emptyset} \theta(F) \right\} \right]$. The canonical statistic is

$$Y = \{Y(D) : D \subseteq_{\mathcal{G}} V, D \neq \emptyset\},$$

where

$$Y(D) = \prod_{v \in D} X_v,$$

and the canonical parameter is

$$\theta = \{\theta(D) : D \subseteq_{\mathcal{G}} V, D \neq \emptyset\},$$

where

$$\theta(D) = \sum_{i=1}^k \theta_{C_i}^*(D) - \sum_{i=2}^k \theta_{S_i}^*(D),$$

with $\theta_{C_i}^*(D) = 0$ whenever D is not contained in C_i , and similarly for $\theta_{S_i}^*(D)$.

The mean parameter for the model in (6) is

$$\mu = \{\mu(D) : D \subseteq_{\mathcal{G}} V, D \neq \emptyset\},$$

where

$$\mu(D) = E\{Y(D)|\theta\}.$$

With reference to the joint model (6), we let $[V(\mu)^{-1}]_{C_i}$ denote the submatrix of $V(\mu)^{-1}$ corresponding to Y_{C_i} , while $[V(\mu)^{-1}]_{C_i}^0$ is the matrix obtained from $[V(\mu)^{-1}]_{C_i}$ by filling up with zero entries to obtain full dimension, and similarly for S_i .

Proposition 4. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$. Then*

$$V(\mu)^{-1} = \sum_{i=1}^k [V(\mu)^{-1}]_{C_i}^0 - \sum_{i=2}^k [V(\mu)^{-1}]_{S_i}^0.$$

Example. Throughout the paper we shall use for illustration an example reported in Edwards & Havranek (1985). A 15-year follow-up study of probable risk factors for coronary heart disease involved 1841 employees of a car factory. Six binary variables are considered: a = smoking, yes or no, b = strenuous mental work, yes or no, c = strenuous physical work, yes or no, d = systolic blood pressure, <140 or ≥ 140 , e = ratio of α and β lipoproteins, < 3

or ≥ 3 , f = family history of coronary heart disease, yes or no. Based on the data collected, Madigan & Raftery (1994) evaluated all possible undirected graphical models for the six variables. The graph in Fig. 1 corresponds to the highest posterior probability model based on 25% of the observations; see Table 6 in Madigan & Raftery (1994). The remaining 75% of the data were allocated for prediction purposes.

FIGURE 1 ABOUT HERE

Cliques and separators are respectively $C_1 = \{b, c\}$, $C_2 = \{b, e\}$, $C_3 = \{a, e\}$, $C_4 = \{d, e\}$, $C_5 = \{f\}$ and $S_2 = \{b\}$, $S_3 = \{e\}$, $S_4 = \{e\}$, $S_5 = \emptyset$.

The natural exponential family representation is based on the following collections of subsets: $\mathcal{C}_1 = \{b, c, bc\}$, $\mathcal{C}_2 = \{b, e, be\}$, $\mathcal{C}_3 = \{a, e, ae\}$, $\mathcal{C}_4 = \{d, e, de\}$, $\mathcal{C}_5 = \{f\}$, $\mathcal{S}_2 = \{b\}$, $\mathcal{S}_3 = \{e\}$, $\mathcal{S}_4 = \{e\}$. To be more precise

$$f(y|\theta) = \frac{f_{\mathcal{C}_1}(y_{\mathcal{C}_1}|\theta_{\mathcal{C}_1}^*) \times \cdots \times f_{\mathcal{C}_5}(y_{\mathcal{C}_5}|\theta_{\mathcal{C}_5}^*)}{f_{\mathcal{S}_2}(y_{\mathcal{S}_2}|\theta_{\mathcal{S}_2}^*) \times \cdots \times f_{\mathcal{S}_4}(y_{\mathcal{S}_4}|\theta_{\mathcal{S}_4}^*)},$$

where, for instance, the marginal density for clique C_1 is

$$f_{\mathcal{C}_1}(y_{\mathcal{C}_1}|\theta_{\mathcal{C}_1}^*) = \exp\{\theta_{\mathcal{C}_1}^*(b)y(b) + \theta_{\mathcal{C}_1}^*(c)y(c) + \theta_{\mathcal{C}_1}^*(bc)y(bc) - M_{\mathcal{C}_1}(\theta_{\mathcal{C}_1}^*)\},$$

with $M_{\mathcal{C}_1}(\theta_{\mathcal{C}_1}^*) = \log\left(1 + e^{\theta_{\mathcal{C}_1}^*(b)} + e^{\theta_{\mathcal{C}_1}^*(c)} + e^{\theta_{\mathcal{C}_1}^*(bc)+\theta_{\mathcal{C}_1}^*(b)+\theta_{\mathcal{C}_1}^*(c)}\right)$. The canonical statistic and parameter are respectively

$$\begin{aligned} Y(b) &= X_b, \\ Y(c) &= X_c, \\ Y(bc) &= X_b X_c, \\ \theta_{\mathcal{C}_1}^*(b) &= \log \frac{p_{\mathcal{C}_1}^*(b)}{p_{\mathcal{C}_1}^*(\emptyset)}, \\ \theta_{\mathcal{C}_1}^*(c) &= \log \frac{p_{\mathcal{C}_1}^*(c)}{p_{\mathcal{C}_1}^*(\emptyset)}, \\ \theta_{\mathcal{C}_1}^*(bc) &= \log \frac{p_{\mathcal{C}_1}^*(bc)p_{\mathcal{C}_1}^*(\emptyset)}{p_{\mathcal{C}_1}^*(b)p_{\mathcal{C}_1}^*(c)}, \end{aligned}$$

where the cell probabilities are those in Table 1.

TABLE 1 ABOUT HERE

An example of the relationship between joint and marginal parameters is the following:

$$\begin{aligned}\theta(c) &= \theta_{\mathcal{C}_1}^*(c), \\ \theta(b) &= \theta_{\mathcal{C}_1}^*(b) + \theta_{\mathcal{C}_2}^*(b) - \theta_{\mathcal{S}_2}^*(b).\end{aligned}$$

2.3 *Mixed parameterisation*

Consider a steep natural exponential family on \mathbb{R}^d with canonical statistic Y . Partition Y as $Y = (Y_1, Y_2)$, and similarly for the canonical parameter $\theta = (\theta_1, \theta_2)$ and the mean parameter $\mu = (\mu_1, \mu_2)$. Suppose now that Y_1 is a cut. From Theorem 3.1 of Barndorff-Nielsen & Koudou (1995) it follows that (μ_1, θ_2) is a reparameterisation of the natural exponential family whose density can be written as

$$f(y_1, y_2 | \mu_1, \theta_2) = f_1(y_1 | \mu_1) f_2(y_2 | y_1, \theta_2).$$

Moreover both f_1 and f_2 are natural exponential families, although the latter may be concentrated on an affine hyperplane of lower dimension for specific values of y_1 .

Gutiérrez Peña & Rueda (2003) provide details about the structure of f_1 and f_2 . Note that the Fisher information matrix for the mixed parameterisation, $H(\mu_1, \theta_2)$, is block-diagonal; that is

$$H(\mu_1, \theta_2) = \begin{pmatrix} H_{11}(\mu_1) & 0 \\ 0 & H_{22}(\mu_1, \theta_2) \end{pmatrix}, \quad (7)$$

with H_{11} only depending on μ_1 . This fact will prove extremely useful in § 3 in the construction of reference priors.

Our objective is to identify a suitable mixed parameterisation for a general discrete decomposable model. Again we shall start by considering first the complete case corresponding to a marginal clique model.

2.4 Mixed parameterisation for a clique marginal model

Before deriving mixed parameterisations for clique marginal models we need some preliminary results. Consider a decomposable model with cliques C_1, \dots, C_k and corresponding separators S_2, \dots, S_k . Define the subset of C_i

$$C_i^1 = \cup_{j \neq i} \{C_i \cap C_j\},$$

and the collection

$$\mathcal{C}_i^1 = \{D : D \subseteq C_i^1, D \neq \emptyset\},$$

containing all subsets of C_i with nonempty intersection with the collection of remaining cliques. Then, if we let $\mathcal{C}_i^2 = C_i \setminus \mathcal{C}_i^1$, we have generated a partition of C_i . Note that \mathcal{C}_i^1 contains subsets of C_i that are in common with other cliques, whereas those of \mathcal{C}_i^2 are specific to clique C_i . Such collections of subsets satisfy several useful properties, namely

Property 1. \mathcal{C}_i^1 is complete for $i = 1, \dots, k$.

Property 2. $C_1^1 \cup \dots \cup C_k^1 = S_2 \cup \dots \cup S_k$.

Property 3. $C_1^1 \cup \dots \cup C_k^1 = \{D : D \subseteq_{\mathcal{G}} (C_1^1 \cup \dots \cup C_k^1), D \neq \emptyset\}$.

Property 4. $Y_{\mathcal{C}_i^1}$ is a cut for $f_{\mathcal{C}_i}(y_{\mathcal{C}_i} | \theta_{\mathcal{C}_i}^*)$ for $i = 1, \dots, k$.

Property 5. $\mathcal{C}_1^2, \dots, \mathcal{C}_k^2$ are pairwise disjoint.

Property 6. $[\theta_{\mathcal{C}_i}^*]_{\mathcal{C}_i^2} = \{\theta_{\mathcal{C}_i}^*(D) : D \in \mathcal{C}_i^2\} = \theta_{\mathcal{C}_i^2} = \{\theta(D) : D \in \mathcal{C}_i^2\}$.

Property 6 means that the subvector, corresponding to the components in \mathcal{C}_i^2 , of the canonical parameter $\theta_{\mathcal{C}_i}^*$ indexing the marginal natural exponential family of clique \mathcal{C}_i , coincides with the corresponding subvector of the canonical parameter θ indexing the joint natural exponential family for Y .

Recall the partition $\mathcal{C}_i = \mathcal{C}_i^1 \cup \mathcal{C}_i^2$ and consider the corresponding partition for $Y_{\mathcal{C}_i}$, namely

$$Y_{\mathcal{C}_i} = (Y_{\mathcal{C}_i^1}, Y_{\mathcal{C}_i^2}), \quad (8)$$

with the allied mixed parameterisation

$$\zeta_{\mathcal{C}_i} = (\mu_{\mathcal{C}_i^1}, \theta_{\mathcal{C}_i^2}), \quad (9)$$

where (9) makes use of Properties 4 and 6.

Proposition 5. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$, with cliques C_1, \dots, C_k . Then the marginal density of $Y_{\mathcal{C}_i}$, parameterised by $\zeta_{\mathcal{C}_i}$, factorises as*

$$f_{\mathcal{C}_i}\{y_{\mathcal{C}_i} | \theta_{\mathcal{C}_i}^*(\zeta_{\mathcal{C}_i})\} = f_{\mathcal{C}_i^1}(y_{\mathcal{C}_i^1} | \mu_{\mathcal{C}_i^1}) f_{\mathcal{C}_i^2 | \mathcal{C}_i^1}(y_{\mathcal{C}_i^2} | y_{\mathcal{C}_i^1}, \theta_{\mathcal{C}_i^2}), \quad (10)$$

with $\mu_{\mathcal{C}_i^1}$ and $\theta_{\mathcal{C}_i^2}$ likelihood independent.

Proof. The marginal density of $Y_{\mathcal{C}_i}$ is obtained in Corollary 1. Since $Y_{\mathcal{C}_i^1}$ is a cut by Property 4, the result follows from Theorem 3.1 of Barndorff-Nielsen & Koudou (1995). \square

Both densities on the right-hand side of (10) have an exponential family structure, so that they can be written as

$$\begin{aligned} f_{\mathcal{C}_i^1}(y_{\mathcal{C}_i^1} | \mu_{\mathcal{C}_i^1}) &\propto \exp \left\{ \theta_{\mathcal{C}_i^1}^*(\mu_{\mathcal{C}_i^1})^T y_{\mathcal{C}_i^1} - M_{\mathcal{C}_i^1}(\theta_{\mathcal{C}_i^1}^*(\mu_{\mathcal{C}_i^1})) \right\}, \\ f_{\mathcal{C}_i^2 | \mathcal{C}_i^1}(y_{\mathcal{C}_i^2} | y_{\mathcal{C}_i^1}, \theta_{\mathcal{C}_i^2}) &\propto \exp \left\{ \theta_{\mathcal{C}_i^2}^T y_{\mathcal{C}_i^2} - M_{\mathcal{C}_i^2}(\theta_{\mathcal{C}_i^2} | y_{\mathcal{C}_i^1}) \right\}, \end{aligned} \quad (11)$$

where $M_{\mathcal{C}_i^2}(\theta_{\mathcal{C}_i^2} | y_{\mathcal{C}_i^1}) = K_{\mathcal{C}_i}(\theta_{\mathcal{C}_i^2}) - y_{\mathcal{C}_i^1}^T G_{\mathcal{C}_i}(\theta_{\mathcal{C}_i^2})$ for some real function $K_{\mathcal{C}_i}$ and vector-valued function $G_{\mathcal{C}_i}$, see also § 3 of Gutiérrez Peña & Rueda (2003). Note that using this partition we do not need to compute marginal parameters, since $\mu_{\mathcal{C}_i^1}$ and $\theta_{\mathcal{C}_i^2}$ are subvectors of μ and θ respectively.

2.5 Mixed parameterisation for a decomposable model

Consider the following partition of Y and the associated mixed parameterisation:

$$\begin{aligned} Y &= (Y_{C_1^1 \cup \dots \cup C_k^1}, Y_{C_1^2 \cup \dots \cup C_k^2}) = (Y_{C^1}, Y_{C^2}), \\ \zeta &= (\mu_{C_1^1 \cup \dots \cup C_k^1}, \theta_{C_1^2 \cup \dots \cup C_k^2}) = (\mu_{C^1}, \theta_{C^2}). \end{aligned} \quad (12)$$

Proposition 6. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$, with cliques C_1, \dots, C_k . Then Y_{C^1} is a cut for $f(y|\theta)$.*

Proof. The result can be shown using Proposition 3 on the induced subgraph $\mathcal{G}_{C_1^1 \cup \dots \cup C_k^1}$. In order to make use of Proposition 3, we first need to prove that $\mathcal{G}_{C_1^1 \cup \dots \cup C_k^1}$ is decomposable and that the marginal density of $\{X_v : v \in C_1^1 \cup \dots \cup C_k^1\}$ is Markov with respect to such a graph.

Decomposability is preserved when considering induced subgraphs, see § 4.2 in Cowell et al. (1999), and hence $\mathcal{G}_{C_1^1 \cup \dots \cup C_k^1}$ is decomposable. In order to prove the Markov property, let p denote the cell probabilities for the joint model, namely $p = \{p(D) : D \subseteq V, D \neq \emptyset\}$, where $p(D) = \text{pr}(X_D = 1, X_{V \setminus D} = 0|p)$. Consider the density obtained via marginalisation,

$$f_{C_1^1 \cup \dots \cup C_k^1}(x_{C_1^1 \cup \dots \cup C_k^1} | p_{C_1^1 \cup \dots \cup C_k^1}^*) = \sum_{x_{V \setminus C_1^1 \cup \dots \cup C_k^1}} f\{y(x)|\theta(p)\}.$$

Using (1) and Property 2 of § 2.4 we can write this as

$$\begin{aligned} & \frac{1}{\prod_{i=2}^k f_{S_i}\{y_{S_i}(x_{S_i})|\theta_{S_i}^*(p)\}} \times \sum_{x_{V \setminus C_1^1 \cup \dots \cup C_k^1}} \prod_{i=1}^k f_{C_i}\{y_{C_i}(x_{C_i})|\theta_{C_i}^*(p)\} \\ &= \frac{1}{\prod_{i=2}^k f_{S_i}\{y_{S_i}(x_{S_i})|\theta_{S_i}^*(p)\}} \times \prod_{i=1}^k h_{C_i^1}\{y_{C_i^1}(x_{C_i^1})|\theta_{C_i^1}^*(p)\} \\ &= \prod_{D \in \mathcal{C}_1^1 \cup \dots \cup \mathcal{C}_k^1} \phi_D\{y(D)\} \\ &= \prod_{D \in \mathcal{C}^1} \phi_D\{y(D)\}, \end{aligned}$$

for some, possibly constant, functions $h_{C_i^1}$ and ϕ_D , where the last equality follows from Property 3. We have thus shown that the density associated with $\mathcal{G}_{C_1^1 \cup \dots \cup C_k^1}$ can be written

as the product of functions defined on the complete subsets of $C_1^1 \cup \dots \cup C_k^1$. This is equivalent to the Markov property with respect to $\mathcal{G}_{C_1^1 \cup \dots \cup C_k^1}$, (Lauritzen, 1996, § 3.2.1). Finally, since the hypotheses of Proposition 3 are satisfied, namely that $\mathcal{M}(\mathcal{G}_{C_1^1 \cup \dots \cup C_k^1})$ is a decomposable graphical model, we can conclude that Y_{C_1} is distributed according to a natural exponential family and is therefore a cut. \square

We can now state a result analogous to that of Proposition 5.

Proposition 7. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$, with cliques C_1, \dots, C_k . Then the following factorisation holds:*

$$f\{y|\theta(\zeta)\} = f_{C^1}(y_{C^1}|\mu_{C^1})f_{C^2|C^1}(y_{C^2}|y_{C^1}, \theta_{C^2}). \quad (13)$$

Proof. By Proposition 6 Y_{C^1} is a cut, and the result follows from Theorem 3.1 of Barndorff-Nielsen & Koudou (1995). \square

Note that, since results in the previous subsection hold for each clique and the Markov property is assumed, the joint density can also be written as

$$\begin{aligned} f(y|\theta) &= \frac{\prod_{i=1}^k f_{C_i}(y_{C_i}|\theta_{C_i}^*)}{\prod_{i=2}^k f_{S_i}(y_{S_i}|\theta_{S_i}^*)} \\ &= \frac{\prod_{i=1}^k f_{C_i^1}(y_{C_i^1}|\mu_{C_i^1})f_{C_i^2|C_i^1}(y_{C_i^2}|y_{C_i^1}, \theta_{C_i^2})}{\prod_{i=2}^k f_{S_i}(y_{S_i}|\theta_{S_i}^*)}. \end{aligned} \quad (14)$$

As a consequence, if we define $\mathcal{V} = \{D : D \subseteq_G V, D \neq \emptyset\}$, the following result holds.

Proposition 8. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$, with cliques C_1, \dots, C_k . Then $Y_{C_i^2} \perp\!\!\!\perp Y_{\mathcal{V} \setminus C_i^2} | Y_{C_i^1}$ for each $i = 1, \dots, k$.*

Proof. Omitting the dependence on parameters, write

$$\begin{aligned} h_{C_i}(y_{C_i^1}, y_{C_i^2}) &= f_{C_i^1}(y_{C_i^1}|\mu_{C_i^1})f_{C_i^2|C_i^1}(y_{C_i^2}|y_{C_i^1}, \theta_{C_i^2}), \\ g_{C_i}(y_{\mathcal{V} \setminus C_i^2}) &= \frac{\prod_{j \neq i} f_{C_j^1}(y_{C_j^1}|\mu_{C_j^1})f_{C_j^2|C_j^1}(y_{C_j^2}|y_{C_j^1}, \theta_{C_j^2})}{\prod_{i=2}^k f_{S_i}\{y_{S_i}|\mu_{S_i}(\theta_{S_i}^*)\}}. \end{aligned}$$

Then, from equation (14), we have

$$f\{y|\theta(\zeta)\} = h_{C_i}(y_{C_i^1}, y_{C_i^2})g_{C_i}(y_{V \setminus C_i^2}),$$

which is the definition of conditional independence of $Y_{C_i^2}$ and $Y_{V \setminus C_i^2}$ given $Y_{C_i^1}$. \square

Since equations (13) and (14) are both representations of the joint density, Proposition 8 yields

$$f_{C^2|C^1}(y_{C^2}|y_{C^1}, \theta_{C^2}) = \prod_{i=1}^k f_{C_i^2|C_i^1}(y_{C_i^2}|y_{C_i^1}, \theta_{C_i^2}), \quad (15)$$

whence

$$f_{C^1}(y_{C^1}|\mu_{C^1}) = \frac{\prod_{i=1}^k f_{C_i^1}(y_{C_i^1}|\mu_{C_i^1})}{\prod_{i=2}^k f_{S_i}\{y_{S_i}|\mu_{S_i}(\theta_{S_i}^*)\}}.$$

Note that μ_{S_i} , $i = 2, \dots, k$, is contained in μ_{C^1} since each S_i is a subset of some clique C_j .

Example (continued). In order to define the mixed parameterisation we first construct the following partitions of C_1, \dots, C_5 :

$$\begin{aligned} C_1^1 &= \{b\} & C_1^2 &= \{c, bc\} \\ C_2^1 &= \{b, e, be\} & C_2^2 &= \emptyset \\ C_3^1 &= \{e\} & C_3^2 &= \{a, ae\} \\ C_4^1 &= \{e\} & C_4^2 &= \{d, de\} \\ C_5^1 &= \emptyset & C_5^2 &= \{f\}. \end{aligned}$$

As a consequence,

$$\begin{aligned} C^1 &= C_1^1 \cup \dots \cup C_5^1 = \{b, e, be\}, \\ C^2 &= C_1^2 \cup \dots \cup C_5^2 = \{a, c, d, f, bc, ae, de\}, \end{aligned}$$

so that the overall partition and associated mixed parameterisation is $Y = (Y_{C^1}, Y_{C^2})$, $\zeta = (\mu_{C^1}, \theta_{C^2})$, with

$$\begin{aligned} \mu_{C^1} &= (\mu(b), \mu(e), \mu(be)), \\ \theta_{C^2} &= (\theta(a), \theta(c), \theta(d), \theta(f), \theta(bc), \theta(ae), \theta(de)), \end{aligned}$$

where for instance $\mu(b) = p_{\mathcal{C}_1}^*(b) + p_{\mathcal{C}_1}^*(bc)$, and similarly for the remaining components of $\mu_{\mathcal{C}^1}$.

3 Reference priors for discrete decomposable models

3.1 Reference priors

Let X be an observable random quantity with density function $f(x|\phi)$, where $\phi \in \Phi \subseteq \mathbb{R}^d$ denotes an unknown parameter. Berger & Bernardo (1992) describe an algorithm for finding reference priors. Such an algorithm is rather elaborate for the general case; it greatly simplifies, however, in the so-called regular case, that is if the posterior distribution of ϕ is asymptotically normal whenever the class of continuous priors for ϕ is strictly positive on Φ .

A reference prior for a given parameter depends on the grouping and ordering of the parameter components. This is why one should strictly speak of ordered-group reference priors. We shall see, however, that in some circumstances order does not matter, so that typically it is the grouping structure that is more important.

We assume that ϕ is decomposed into r groups $(\phi_{(1)}, \dots, \phi_{(r)})$, with $\phi_{(k)}$ having dimension d_k ; we also let $d_k^* = \sum_{j=1}^k d_j$ and define $\phi_{[k]} = (\phi_{(1)}, \dots, \phi_{(k)})$ and $\phi_{[\sim k]} = (\phi_{(k+1)}, \dots, \phi_{(r)})$, for all $k = 1, \dots, r$. The elements of ϕ are usually ordered according to inferential importance; in particular, the parameters of interest should come first. Let $\phi_{(k)} \in \Phi_k$, $k = 1, \dots, r$, and assume that $\Phi = \Phi_1 \times \dots \times \Phi_r$. Let

$$H(\phi) = -E_{X|\phi} \left\{ \frac{\partial^2 \ln f(X|\phi)}{\partial \phi^T \partial \phi} \right\}$$

denote the Fisher information matrix for model $f(x|\phi)$.

For our purposes it is enough to consider the special case for which $H(\phi) = \text{diag}\{H_{11}(\phi), \dots, H_{rr}(\phi)\}$ with $H_{kk}(\phi)$ a $d_k \times d_k$ matrix, with the further condition that there exist positive functions a_k and b_k such that

$$\det\{H_{kk}(\phi)\} = a_k(\phi_{(k)})b_k(\phi_{[k-1]}, \phi_{[\sim k]}), \quad \text{for all } k \in \{1, \dots, r\}. \quad (16)$$

In this case the density with respect to Lebesgue measure of the r -group reference prior on ϕ is given by

$$\pi_{\phi}^{\text{R}}(\phi_{(1)}, \dots, \phi_{(r)}) \propto \prod_{k=1}^r a_k(\phi_{(k)})^{1/2}, \quad \phi_{(k)} \in \Phi_k; \quad (17)$$

see Datta & Ghosh (1995). The Jeffreys prior π_{ϕ}^{J} , which is sometimes used as a noninformative prior, can be recovered as a special case of the reference prior by simply treating the whole parameter ϕ as a single group. As a consequence, if there exists a grouped parameterisation for the model such that the Fisher information matrix is block-diagonal and the factorisation (16) holds, then the reference prior is explicitly available and is provided by (17).

Clearly the mixed parameterisation (μ_1, θ_2) induced through a cut X_1 , as described in § 2.3, leads to a block-diagonal Fisher information matrix. Gutiérrez Peña & Rueda (2003) show in addition that, if the variance function of the underlying natural exponential family is homogeneous or simple quadratic, factorisation (16) holds with $r = 2$. They also identify explicitly the functions $a_1(\mu_1)$ and $a_2(\theta_2)$ and thus obtain the corresponding reference prior for (μ_1, θ_2) .

In the following we shall show that the mixed parameterisation for a discrete decomposable graphical model identified in (12) also affords a factorisation like (16), thus leading explicitly to the appropriate reference prior. We first deal with a clique marginal model and then tackle the general decomposable case.

3.2 Clique marginal model

Proposition 9. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$, with cliques C_1, \dots, C_k . Then the Fisher information matrix relative to the clique marginal model $f_{C_i}(y_{C_i} | \theta_{C_i}^*(\zeta_{C_i}))$ is given by*

$$H_{C_i}(\zeta_{C_i}) = \begin{pmatrix} H_{C_i^1}(\mu_{C_i^1}) & 0 \\ 0 & H_{C_i^2}(\theta_{C_i^2}, \mu_{C_i^1}) \end{pmatrix}, \quad (18)$$

where $H_{C_i^1}(\mu_{C_i^1}) = V_{C_i^1}(\mu_{C_i^1})^{-1}$ is the inverse of the variance function for $f_{C_i^1}\{y_{C_i^1} | \theta_{C_i^1}^*(\mu_{C_i^1}^*)\}$, and

$$\det\{H_{C_i}(\zeta_{C_i})\} = \det\{V_{C_i^1}(\mu_{C_i^1})\}^{-1} a_{C_i^2}(\theta_{C_i^2}) b_{C_i^2}(\mu_{C_i^1}), \quad (19)$$

for some positive functions $a_{C_i^2}$ and $b_{C_i^2}$.

Proof. Equation (18) as well as $H_{C_i^1}(\mu_{C_i^1}) = V_{C_i^1}(\mu_{C_i^1})^{-1}$ follows directly from Lemma 3.3 in Barndorff-Nielsen & Blaesild (1983). Expression (19) follows from the theory in § 4 of Gutiérrez Peña & Rueda (2003) since $f_{C_i}(y_{C_i} | \theta_{C_i}^*)$ is a natural exponential family with a simple quadratic variance function and $Y_{C_i^1}$ is a cut. \square

We are now ready to compute the Jeffreys and reference prior for the mixed parameterisation of clique marginal models.

Proposition 10. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$, with cliques C_1, \dots, C_k . Then the Jeffreys prior and reference prior for the clique marginal model $f_{C_i}\{y_{C_i} | \theta_{C_i}^*(\zeta_{C_i})\}$, relative to the mixed parameterisation*

$\zeta_{C_i} = (\mu_{C_i^1}, \theta_{C_i^2}) \in \Omega_{C_i^1} \times \Theta_{C_i^2}$, are respectively

$$\begin{aligned} \pi_{\zeta_{C_i}}^J &\propto \prod_{D \subseteq C_i^1} \left\{ \sum_{F \subseteq C_i^1 \setminus D} (-1)^{|F|} \mu(D \cup F) \right\}^{-1} \\ &\exp \left(\frac{1}{2} \left[\theta_{C_i^2}^T s_{C_i^2}^0 - \left\{ t_{C_i}^0 K_{C_i}(\theta_{C_i^2}) - s_{C_i^1}^{0T} G_{C_i}(\theta_{C_i^2}) \right\} + \theta_{C_i^1}^* (\mu_{C_i^1})^T s_{C_i^1}^0 - \right. \right. \\ &\quad \left. \left. - t_{C_i}^0 M_{C_i^1} \{ \theta_{C_i^1}^* (\mu_{C_i^1}) \} \right] \right), \\ \pi_{\zeta_{C_i}}^R &\propto \prod_{D \subseteq C_i^1} \left\{ \sum_{F \subseteq C_i^1 \setminus D} (-1)^{|F|} \mu(D \cup F) \right\}^{-1/2} \times \\ &\times \exp \left(\frac{1}{2} \left[\theta_{C_i^2}^T s_{C_i^2}^0 - \left\{ t_{C_i}^0 K_{C_i}(\theta_{C_i^2}) - s_{C_i^1}^{0T} G_{C_i}(\theta_{C_i^2}) \right\} \right] \right), \end{aligned}$$

where

$$K_{C_i}(\theta_{C_i^2}) = \log \left[1 + \sum_{D \subseteq C_i \setminus C_i^1, D \neq \emptyset} \exp \left\{ \sum_{F \subseteq D, F \neq \emptyset} \theta(F) \right\} \right], \quad (20)$$

$$t_{C_i}^0 = (|C_i| + 1), \quad (21)$$

and for each $D \in C_i^1$ the D -component of the vector valued function G_{C_i} is given by

$$\begin{aligned} G_{C_i}(\theta_{C_i^2})(D) &= \sum_{F \subseteq D} (-1)^{|D \setminus F|+1} \log \left[1 + \sum_{G \subseteq C_i \setminus C_i^1, G \neq \emptyset} \exp \left\{ \sum_{H_1 \subseteq G, H_1 \neq \emptyset} \theta(H_1) + \right. \right. \\ &\quad \left. \left. + \sum_{H_2: H_2=H_3 \cup H_4, H_3 \subseteq G, H_3 \neq \emptyset, H_4 \subseteq F, H_4 \neq \emptyset} \theta(H_2) \right\} \right]. \end{aligned} \quad (22)$$

Finally

$$s_{C_i}^0 = (s_{C_i^1}^0, s_{C_i^2}^0),$$

with the individual component of $s_{C_i}^0$ given by

$$s_{C_i}^0(D) = 2^{|C_i \setminus D|}. \quad (23)$$

Proof. As a result of the structure of $\det\{H_{C_i}(\zeta_{C_i})\}$, the Jeffreys and reference priors are computed respectively as

$$\pi_{\zeta_{C_i}}^J \propto \left[\det\{V_{C_i^1}(\mu_{C_i^1})\}^{-1} a_{C_i^2}(\theta_{C_i^2}) b_{C_i^2}(\mu_{C_i^1}) \right]^{1/2}, \quad (24)$$

$$\pi_{\zeta_{C_i}}^R \propto \left[\det\{V_{C_i^1}(\mu_{C_i^1})\}^{-1} a_{C_i^2}(\theta_{C_i^2}) \right]^{1/2}, \quad (25)$$

for $(\mu_{C_i^1}, \theta_{C_i^2}) \in \Omega_{C_i^1} \times \Theta_{C_i^2}$. Consider first $\det\{V_{C_i^1}(\mu_{C_i^1})\}^{-1}$. This is equal to the determinant of the Fisher information matrix, relative to $\mu_{C_i^1}$, for the marginal model $f_{C_i^1}$. To obtain the latter, consider first the determinant of the Fisher information matrix relative to the cell probability parameterisation $p_{C_i^1}^*$. This is easily available from the standard multinomial model with index 1 and is $\left\{\prod_{D \subseteq C_i^1} p_{C_i^1}^*(D)\right\}^{-1}$, where $p_{C_i^1}^*(\emptyset) = 1 - \sum_{D \subseteq C_i^1, D \neq \emptyset} p_{C_i^1}^*(D)$. Now $\mu_{C_i^1}$ is a linear transformation of $p_{C_i^1}^*$, see (5). As a consequence one gets

$$\det\{V_{C_i^1}(\mu_{C_i^1})\}^{-1} = \prod_{D \subseteq C_i^1} \left\{ \sum_{F \subseteq C_i^1 \setminus D} (-1)^{|F|} \mu(D \cup F) \right\}^{-1}, \quad (26)$$

by mere substitution with the understanding that $\mu(\emptyset) = 0$.

We now turn to the expressions for $a_{C_i^2}(\theta_{C_i^2})$ and $b_{C_i^2}(\mu_{C_i^1})$. Using a result in Gutiérrez Peña & Rueda (2003, p. 47) concerning cuts in exponential families having a simple quadratic variance function, we obtain the expressions

$$\begin{aligned} a_{C_i^2}(\theta_{C_i^2}) &= \exp \left[\theta_{C_i^2}^T s_{C_i^2}^0 - \{t_{C_i^2}^0 K_{C_i^2}(\theta_{C_i^2}) - s_{C_i^2}^{0T} G_{C_i^2}(\theta_{C_i^2})\} \right], \\ b_{C_i^2}(\mu_{C_i^1}) &= \det\{V_{C_i^1}(\mu_{C_i^1})\}^{-1} \exp \left[\theta_{C_i^1}^*(\mu_{C_i^1})^T s_{C_i^1}^0 - t_{C_i^1}^0 M_{C_i^1} \{\theta_{C_i^1}^*(\mu_{C_i^1})\} \right], \end{aligned} \quad (27)$$

for some $s_{C_i^2}^0 \in \mathbb{R}^{|C_i^2|}$ and $t_{C_i^2}^0 \in \mathbb{R}$, where

$$\begin{aligned} G_{C_i^2}(\theta_{C_i^2}) &= [\theta_{C_i^1}^*]_{C_i^1}(\mu_{C_i^1}, \theta_{C_i^2}) - \theta_{C_i^1}^*(\mu_{C_i^1}), \\ K_{C_i^2}(\theta_{C_i^2}) &= M_{C_i^1} \{[\theta_{C_i^1}^*]_{C_i^1}(\mu_{C_i^1}, \theta_{C_i^2}), \theta_{C_i^2}\} - M_{C_i^1} \{\theta_{C_i^1}^*(\mu_{C_i^1})\}; \end{aligned}$$

see Efstathiou et al. (1998, p. 82).

The derivation of equations (20) and (22) is highly laborious but relatively straightforward. We omit details. Finally expressions (21) and (23) are derived using Proposition 1 of Consonni et al. (2004). \square

It is apparent from equations (24) and (25) that $\mu_{C_i^1}$ and $\theta_{C_i^2}$ are independent both under the Jeffreys prior and the reference prior.

Proposition 11. *The Jeffreys and reference prior for the clique marginal model*

$f_{C_i}\{y_{C_i}|\theta_{C_i}^*(\zeta_{C_i})\}$, *relative to the mixed parameterisation* $\zeta_{C_i} = (\mu_{C_i^1}, \theta_{C_i^2})$, *are proper.*

Proof. Consider first the Jeffreys prior (24). Note that the clique marginal model, being complete, is a transformation of a standard multinomial model with index 1. Since the Jeffreys prior for the cell probability parameter in the latter model is known to be proper (Bernardo & Smith, 1994, p. 336), and Jeffreys priors are invariant under reparameterisations, it follows that (24) is proper.

Consider now the reference prior (25). To establish the condition of being proper we can work separately on $\mu_{C_i^1}$ and $\theta_{C_i^2}$. Take the marginal on $\mu_{C_i^1}$. Recall from the Proof of Proposition 10 that $\det\{V_{C_i^1}(\mu_{C_i^1})\}^{-1/2} = \prod_{D \subseteq C_i^1} \left\{ \sum_{F \subseteq C_i^1 \setminus D} (-1)^{|F|} \mu(D \cup F) \right\}^{-1/2}$. Consider the one-to-one mapping $\mu_{C_i^1} \rightarrow p_{C_i^1}^*$. As a result of equation (5) and the fact that the Jacobian of the transformation is constant, the induced prior on $p_{C_i^1}^*$ is proportional to $\{\prod_{D \subseteq C_i^1} p_{C_i^1}^*(D)\}^{-1/2}$, which is a Dirichlet distribution with hyperparameters all equal to 1/2, and so the prior on $\mu_{C_i^1}$ is also proper. Take now the component in $\theta_{C_i^2}$ of (25). This coincides with the corresponding marginal under the Jeffreys prior, which has been shown to be proper. \square

3.3 Decomposable model

Results for the Fisher information matrix of clique marginal models can be used to derive the structure of the Fisher information matrix for the joint model.

Proposition 12. *Let* $\mathcal{M}(\mathcal{G})$ *be a discrete decomposable graphical model for the random variables* $X \in \{0, 1\}^{|V|}$, *with cliques* C_1, \dots, C_k . *Then the Fisher information matrix relative to the joint model* $f\{y|\theta(\zeta)\}$ *is given by*

$$H(\zeta) = \text{diag}\{H_{C_1^1}(\mu_{C_1^1}), H_{C_1^2}(\theta_{C_1^2}, \mu_{C_1^1}), \dots, H_{C_k^2}(\theta_{C_k^2}, \mu_{C_k^1})\},$$

where $H_{\mathcal{C}^1}(\mu_{\mathcal{C}^1}) = V_{\mathcal{C}^1}(\mu_{\mathcal{C}^1})^{-1}$ and

$$\det\{H(\zeta)\} = \det\{V_{\mathcal{C}^1}(\mu_{\mathcal{C}^1})\}^{-1} \prod_{i=1}^k \{a_{\mathcal{C}_i^2}(\theta_{\mathcal{C}_i^2}) b_{\mathcal{C}_i^2}(\mu_{\mathcal{C}_i^1})\}, \quad (28)$$

for some positive functions $a_{\mathcal{C}_i^2}$ and $b_{\mathcal{C}_i^2}$, $i = 1, \dots, k$.

Proof. Expression (13) yields $H(\zeta) = \text{diag}\{H_{\mathcal{C}^1}(\mu_{\mathcal{C}^1}), H_{\mathcal{C}^2}(\theta_{\mathcal{C}^2}, \mu_{\mathcal{C}^1})\}$, while Lemma 3.3 in Barndorff-Nielsen & Blaesild (1983) ensures that $H_{\mathcal{C}^1}(\mu_{\mathcal{C}^1}) = V_{\mathcal{C}^1}(\mu_{\mathcal{C}^1})^{-1}$, since $Y_{\mathcal{C}^1}$ is a cut by Proposition 6. Equation (15) implies a block-diagonal structure for $H_{\mathcal{C}^2}(\theta_{\mathcal{C}^2}, \mu_{\mathcal{C}^1})$; on the other hand equation (18), together with (19), provides the expression for the determinant.

□

We are now ready to derive the Jeffreys and reference priors for the mixed parameterisation of the joint model.

Proposition 13. *Let $\mathcal{M}(\mathcal{G})$ be a discrete decomposable graphical model for the random variables $X \in \{0, 1\}^{|V|}$, with cliques C_1, \dots, C_k . Then the Jeffreys prior and reference prior relative to the mixed parameterisation $\zeta = (\mu_{\mathcal{C}^1}, \theta_{\mathcal{C}^2}) \in \Omega_{\mathcal{C}^1} \times \Theta_{\mathcal{C}^2}$ are respectively*

$$\pi_{\zeta}^{\text{J}}(\zeta) \propto \left[\det\{V_{\mathcal{C}^1}(\mu_{\mathcal{C}^1})\}^{-1} \prod_{i=1}^k a_{\mathcal{C}_i^2}(\theta_{\mathcal{C}_i^2}) b_{\mathcal{C}_i^2}(\mu_{\mathcal{C}_i^1}) \right]^{1/2}, \quad (29)$$

$$\pi_{\zeta}^{\text{R}}(\zeta) \propto \left[\det\{V_{\mathcal{C}^1}(\mu_{\mathcal{C}^1})\}^{-1} \prod_{i=1}^k a_{\mathcal{C}_i^2}(\theta_{\mathcal{C}_i^2}) \right]^{1/2}, \quad (30)$$

where $a_{\mathcal{C}_i^2}$ and $b_{\mathcal{C}_i^2}$ are defined in the Proof of Proposition 10.

Proof. Immediate using equation (28) and Proposition 10. □

Under both the Jeffreys and reference priors $\mu_{\mathcal{C}^1}$ and $\theta_{\mathcal{C}^2}$ are independent. Moreover, the marginal prior on $\theta_{\mathcal{C}^2}$ is proper, since it is a product of proper distributions, one for each $\theta_{\mathcal{C}_i^2}$, see Proposition 11; in addition it belongs to the standard conjugate family, as can be gathered from (11) and (27).

Example (continued). We compute equation (30) as follows. Since \mathcal{G} is collapsible on to

the complete subset $\{b, e\}$, the corresponding marginal distribution is a natural exponential family with canonical statistic $Y_{\mathcal{C}_1}$. Using (26) we therefore obtain

$$\det\{V_{\mu_{\mathcal{C}_1}}(\mu_{\mathcal{C}_1})\} = \{\mu(b) - \mu(be)\}\{\mu(e) - \mu(be)\}\mu(be)\{1 - \mu(b) - \mu(e) + \mu(be)\},$$

where $\mu_{\mathcal{C}_1} \in \Omega_{\mathcal{C}_1} = \{\mu(b), \mu(e), \mu(be) : 0 < \mu(b) < 1, 0 < \mu(e) < 1, \mu(be) < \min\{\mu(b), \mu(e)\}, \mu(b) + \mu(e) - \mu(be) < 1\}$. From Proposition 10, we have

$$\begin{aligned} s_{\mathcal{C}_1}^0 &= (s_{\mathcal{C}_1^1}^0, s_{\mathcal{C}_1^2}^0) = (2, (2, 1)), \\ t_{\mathcal{C}_1}^0 &= 4, \\ K_{\mathcal{C}_1}(\theta_{\mathcal{C}_1^2}) &= \log(1 + e^{\theta(c)}), \\ G_{\mathcal{C}_1}(\theta_{\mathcal{C}_1^2}) &= \log\left(\frac{1 + e^{\theta(c)}}{1 + e^{\theta(c) + \theta(bc)}}\right), \end{aligned}$$

and hence

$$a_{\mathcal{C}_1^2}(\theta_{\mathcal{C}_1^2}) = \frac{e^{2\theta(c) + \theta(bc)}}{(1 + e^{\theta(c)})^2(1 + e^{\theta(c) + \theta(bc)})^2},$$

and similarly for the remaining cliques. The overall reference prior for the mixed parameterisation is therefore

$$\begin{aligned} \pi_{\zeta}^{\mathbb{R}}(\zeta) &\propto [\{\mu(b) - \mu(be)\}\{\mu(e) - \mu(be)\}\mu(be)\{1 - \mu(b) - \mu(e) + \mu(be)\}]^{-1/2} \times \\ &\quad \times \frac{e^{\theta(c) + \frac{1}{2}\theta(bc)}}{(1 + e^{\theta(c)})(1 + e^{\theta(c) + \theta(bc)})} \frac{e^{\theta(a) + \frac{1}{2}\theta(ae)}}{(1 + e^{\theta(a)})(1 + e^{\theta(a) + \theta(ae)})} \times \\ &\quad \times \frac{e^{\theta(d) + \frac{1}{2}\theta(de)}}{(1 + e^{\theta(d)})(1 + e^{\theta(d) + \theta(de)})} \frac{e^{\frac{1}{2}\theta(f)}}{(1 + e^{\theta(f)})}, \end{aligned}$$

where $\mu_{\mathcal{C}_1} \in \Omega_{\mathcal{C}_1}$ and $\theta_{\mathcal{C}_2} \in \Theta_{\mathcal{C}_2} = \mathbb{R}^7$. Consider the marginal for $\mu_{\mathcal{C}_1}$ under $\pi_{\zeta}^{\mathbb{R}}$. Since

$$\mu(b) = p_{\mathcal{C}_1^1}^*(b) + p_{\mathcal{C}_1^2}^*(be),$$

$$\mu(e) = p_{\mathcal{C}_1^1}^*(e) + p_{\mathcal{C}_1^2}^*(be),$$

$$\mu(be) = p_{\mathcal{C}_1^2}^*(be),$$

one can see that such a marginal is the prior induced by a Dirichlet distribution on $(p_{\mathcal{C}^1}^*(b), p_{\mathcal{C}^1}^*(e), p_{\mathcal{C}^1}^*(be))$, with hyperparameters all equal to $1/2$. The marginal for $\theta_{\mathcal{C}^2}$ is the same under the reference prior and the Jeffreys prior. Under π_{ζ}^J , the marginal for $\mu_{\mathcal{C}^1}$ is

$$\begin{aligned} \pi_{\mu_{\mathcal{C}^1}}^J(\mu_{\mathcal{C}^1}) &\propto [\{\mu(b) - \mu(be)\}\{\mu(e) - \mu(be)\}\mu(be)\{1 - \mu(b) - \mu(e) + \mu(be)\}]^{-1/2} \times \\ &\quad \times \mu(b)^{1/2}\{1 - \mu(b)\}^{1/2}\mu(e)\{1 - \mu(e)\}, \quad \mu_{\mathcal{C}^1} \in \Omega_{\mathcal{C}^1}, \end{aligned}$$

which differs from the corresponding marginal under the reference prior because of the multiplicative factor $\mu(b)^{1/2}\{1 - \mu(b)\}^{1/2}\mu(e)\{1 - \mu(e)\}$.

For a random sample X_1, \dots, X_n , where $X_i = (X_{ai}, \dots, X_{fi})$, the likelihood function is shown in equation (14). The canonical statistics are $n_a = \sum_{i=1}^n x_{ai}, n_b = \sum_{i=1}^n x_{bi}, n_c = \sum_{i=1}^n x_{ci}, n_d = \sum_{i=1}^n x_{di}, n_e = \sum_{i=1}^n x_{ei}, n_f = \sum_{i=1}^n x_{fi}, n_{ae} = \sum_{i=1}^n x_{ai}x_{ei}, n_{bc} = \sum_{i=1}^n x_{bi}x_{ci}, n_{be} = \sum_{i=1}^n x_{bi}x_{ei}, n_{de} = \sum_{i=1}^n x_{di}x_{ei}$. The posterior distribution is

$$\begin{aligned} \tilde{\pi}_{\zeta}^R(\zeta|x_1, \dots, x_n) &\propto \{\mu(b) - \mu(bc)\}^{n_b - n_{be} - 1/2} \{\mu(e) - \mu(be)\}^{n_e - n_{be} - 1/2} \mu(be)^{n_{be} - 1/2} \times \\ &\quad \times \{1 - \mu(b) - \mu(e) + \mu(be)\}^{n + n_{be} - n_b - n_e - 1/2} \times \\ &\quad \times \frac{e^{(n_c + 1)\theta(c) + (n_{bc} + 1/2)\theta(bc)}}{(1 + e^{\theta(c)})^{n - n_b + 1} (1 + e^{\theta(c) + \theta(bc)})^{n_b + 1}} \times \\ &\quad \times \frac{e^{(n_a + 1)(\theta(a) + (n_{ae} + 1/2)\theta(ae))}}{(1 + e^{\theta(a)})^{n - n_e + 1} (1 + e^{\theta(a) + \theta(ae)})^{n_e + 1}} \times \\ &\quad \times \frac{e^{(n_d + 1)\theta(d) + (n_{de} + 1/2)\theta(de)}}{(1 + e^{\theta(d)})^{n - n_e + 1} (1 + e^{\theta(d) + \theta(de)})^{n_e + 1}} \times \\ &\quad \times \frac{e^{(n_f + 1/2)\theta(f)}}{(1 + e^{\theta(f)})^{n + 1}}, \quad \mu_{\mathcal{C}^1} \in \Omega_{\mathcal{C}^1}, \theta_{\mathcal{C}^2} \in \Theta_{\mathcal{C}^2}. \end{aligned}$$

In particular, the Jeffreys posterior for $\mu_{\mathcal{C}^1}$ is

$$\begin{aligned} \tilde{\pi}_{\mu_{\mathcal{C}^1}}^J(\mu_{\mathcal{C}^1}|x_1, \dots, x_n) &\propto \{\mu(b) - \mu(be)\}^{n_b - n_{be} - 1/2} \{\mu(e) - \mu(be)\}^{n_e - n_{be} - 1/2} \times \\ &\quad \times \mu(be)^{n_{be} - 1/2} \{1 - \mu(b) - \mu(e) + \mu(be)\}^{n + n_{be} - n_b - n_e - 1/2} \times \\ &\quad \times \mu(b)^{1/2} \{1 - \mu(b)\}^{1/2} \mu(e) \{1 - \mu(e)\}, \quad \mu_{\mathcal{C}^1} \in \Omega_{\mathcal{C}^1}. \end{aligned}$$

Even though the reference posterior will be largely dominated by the likelihood, poste-

rior estimates of parameters, such as posterior expectations or posterior modes, differ from maximum likelihood estimates, which can be also regarded as Bayes estimates under a local uniform prior. For example, the expectation of μ_{C_1} under the reference posterior, labelled $\hat{\mu}_{C_1}^R$, is given by

$$\begin{aligned}\hat{\mu}^R(b) &= \frac{n_b + 1}{n + 2}, \\ \hat{\mu}^R(e) &= \frac{n_e + 1}{n + 2}, \\ \hat{\mu}^R(be) &= \frac{n_{be} + 1/2}{n + 2},\end{aligned}$$

while the corresponding maximum likelihood estimates are $\hat{\mu}^{\text{ML}}(b) = n_b/n$, $\hat{\mu}^{\text{ML}}(e) = n_e/n$ and $\hat{\mu}^{\text{ML}}(be) = n_{be}/n$. Note that maximum likelihood estimates will be zero whenever the corresponding counts are zero, whereas this will not be the case under the reference posterior, which creates smoother estimates for the mean component of the parameterisation.

Similar considerations apply for the estimates of the parametric component $\theta_{C_i^2}$. For simplicity we only report those corresponding to clique C_1

$$\begin{aligned}\hat{\theta}^R(c) &= \log \frac{n_{bc} - n_c - 1/2}{n_b + n_c - n - n_{bc} - 1/2} \\ \hat{\theta}^R(bc) &= \log \frac{n_{bc} + 1/2}{n_b - n_{bc} + 1/2} - \log \frac{n_{bc} - n_c - 1/2}{n_b + n_c - n - n_{bc} - 1/2},\end{aligned}$$

while the corresponding maximum likelihood estimates, provided they exist, are $\hat{\theta}^{\text{ML}}(c) = \log\{(n_{bc} - n_c)/(n_b + n_c - n - n_{bc})\}$, $\hat{\theta}^{\text{ML}}(bc) = \log\{n_{bc}/(n_b - n_{bc})\} - \log\{(n_{bc} - n_c)/(n_b + n_c - n - n_{bc})\}$.

4 Discussion

In the paper we have restricted our investigation to the binary case. The extension to the polytomous case is feasible, although tedious. Let X_v take values in the set $\mathcal{X}_v = \{0, \dots, d_v\}$, and set $\mathcal{X}_v^0 = \{1, \dots, d_v\}$. The cell probabilities in the clique C_i marginal model are now defined as $p_{C_i}^* = \{p_{C_i}^*(D), D \subseteq C_i, D \neq \emptyset\}$, where $p_{C_i}^*(D) = \{p_{C_i}^*(D, x_D), x_D \in \mathcal{X}_D^0\}$,

$p_{C_i}^*(D, x_D) = \text{pr}(X_D = x_D, X_{C_i \setminus D} = 0 | p_{C_i}^*)$, and $\mathcal{X}_D^0 = \times_{v \in D} \mathcal{X}_v^0$. Similar representations hold for other quantities like canonical parameters and statistics, as well as mean parameters. The results contained in this paper extends naturally to this setting. We omit details and refer to V. Leucari’s Ph. D. thesis.

Reference priors depend on the grouping and ordering of the parameters. In our case the order of the parameters μ_{C^1} and θ_{C^2} does not matter. On the other hand reference priors are consistent across reparameterisations only for block-lower triangular transformations; see Yang (1995) and Datta & Ghosh (1996). It is not difficult to show that the mapping $(\mu_{C^1}, \theta_{C^2}) \rightarrow (\mu_{C^1}, \mu_{C^2})$ enjoys this property and therefore the corresponding reference prior may be obtained through a standard change of variable.

The exponential family representation of § 2.2 is of course strictly related to the traditional log-linear expansion for hierarchical graphical models, see Whittaker (1990, Ch. 7) and Lauritzen (1996, p. 81), wherein the generating class is the set of all cliques \mathcal{C} . The advantage of using an exponential family representation is that it makes more transparent the connection between the marginal clique models and the overall joint decomposable model through the Markov factorisation (1). In particular, this allows us to relate explicitly marginal and joint parameterisations. Another advantage of working within the exponential family setting relates to the possibility of identifying a useful cut, leading to a mixed parameterisation as detailed in § 2.4 for the clique marginal model and § 2.5 for the joint decomposable model. This in turn affords the explicit derivation of the information matrix for the mixed parameterisation together with the associated reference and Jeffreys priors.

A referee suggested that one might investigate connections between the reference prior for the mixed parameterisation and the hyper-Markov laws of Dawid & Lauritzen (1993). The latter are supported on the set of relevant Markov distributions for a given graph; moreover they satisfy certain conditional independence properties also related to the underlying graphical structure. Dawid & Lauritzen (1993) describe an algorithm for constructing

hyper-Markov priors starting from pairwise hyper-consistent priors on clique parameters.

Our approach is somewhat different. We start by constructing the reference prior on each clique marginal model for a suitable mixed parameterisation. This suggests an appropriate cut for the joint exponential family leading to a mixed parameterisation which is the union of the clique components. Note that the overall reference prior we derive cannot be recovered from the set of clique marginal reference priors; on the other hand it is automatically defined on the relevant parameter space.

Acknowledgement

Research partially supported by Ministero dell'Istruzione dell'Università e della Ricerca, Rome (grant: PRIN 2003138887). We thank Alberto Roverato for helpful discussions relating to § 2.2, two anonymous referees for useful remarks, and the Editor for constructive suggestions that led to a better presentation of the paper.

References

- ASMUSSEN, S. & EDWARDS, D. (1983). Collapsibility and response variables in contingency tables. *Biometrika* **70**, 567-78.
- BARNDORFF-NIELSEN, O. E. (1978). *Information and Exponential Families in Statistical Theory*. Chichester: Wiley.
- BARNDORFF-NIELSEN, O. E. & BLAESILD, P. (1983). Exponential families with affine dual foliations. *Ann. Statist.* **11**, 753-69.
- BARNDORFF-NIELSEN, O. E. & KOUDOU, A. E. (1995). Cuts in natural exponential families. *Theory Prob. Applic.* **40**, 361-72.
- BERGER, J. O. & BERNARDO, J. M. (1992). On the development of the reference prior

- method. In *Bayesian Statistics 4*, Ed. J. M. Bernardo, J. O. Berger, D. V. Lindley and A. F. M. Smith, pp.35-49. London: Oxford University Press.
- BERNARDO, J. M. (2005). Reference analysis. In *Handbook of Statistics 25*, Chapter 2, Ed. D. K. Dey and C. R. Rao. Amsterdam: Elsevier.
- BERNARDO, J. M. & SMITH, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.
- BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families, with Applications in Statistical Decision Theory*. Lectures Notes 9, Hayward, CA: Institute of Mathematical Statistics.
- CONSONNI, G., VERONESE, P. & GUTIÉRREZ PEÑA, E. (2004). Order-invariant group reference priors for natural exponential families having a simple quadratic variance function. *J. Mult. Anal.* **88**, 335-64.
- COWELL, R. G., DAWID, A. P., LAURITZEN, S. L. & SPIEGELHALTER, D. J. (1999). *Probabilistic Networks and Expert Systems*. New York: Springer-Verlag.
- DATTA, G. S. & GHOSH, M. (1995). Some remarks on non-informative priors. *J. Am. Statist. Assoc.* **90**, 1357-63.
- DAWID, A. P. & LAURITZEN, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable models. *Ann. Statist.* **21**, 1272-317.
- EDWARDS, D. & HAVRANEK, T. (1985). A fast procedure for model search in multidimensional contingency tables. *Biometrika* **72**, 339-51.
- EFSTATHIOU, M., GUTIÉRREZ PEÑA, E. & SMITH, A. F. M. (1998). Laplace approximations for natural exponential families with cuts. *Scand. J. Statist.* **25**, 77-92.
- FRYDENBERG, M. (1990). Marginalization and collapsibility in graphical interaction models. *Ann. Statist.*
- GEIGER, D., HECKERMAN, D., KING, H. & MEEK, C. (2001). Stratified exponential families: Graphical models and model selection. *Ann. Statist.* **29**, 505-29.

- GUTIÉRREZ PEÑA, E. & RUEDA, R. (2003). Reference priors for exponential families. *J. Statist. Plan. Infer.* **110**, 35-54.
- KOTZ, S., BALAKRISHNAN, N. & JOHNSON, N. L. (2000). *Continuous Multivariate Distributions. Models and Applications*. Volume 1, 2nd ed. New York: Wiley.
- LAURITZEN, S. L. (1996). *Graphical Models*. Oxford: Oxford University Press.
- MADIGAN, D. & RAFTERY, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Am. Statist. Assoc.* **89**, 1535-46.
- WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.
- YANG, R. (1995). Invariance of the reference prior under reparameterisation. *Test* **4**, 83-94.

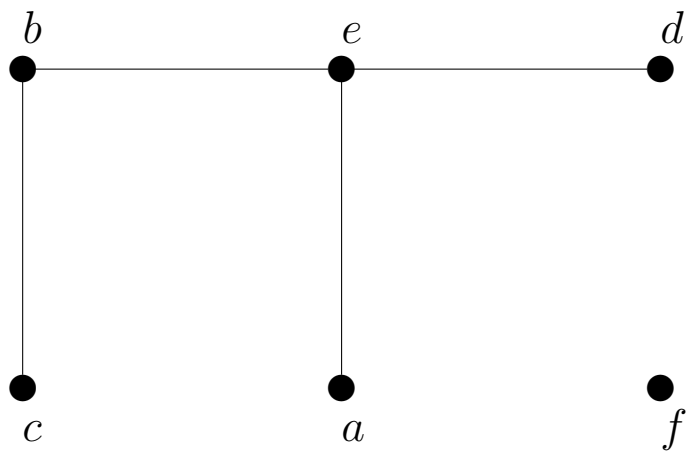


Figure 1: The graph underlying the model in the Example.

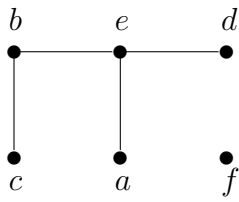


Figure 1: The graph underlying the model in the Example.

	$X_b = 0$	$X_b = 1$
$X_c = 0$	$p_{C_1}^*(\emptyset)$	$p_{C_1}^*(b)$
$X_c = 1$	$p_{C_1}^*(c)$	$p_{C_1}^*(bc)$

Table 1: C_1 -marginal probability table.