The Annals of Statistics
2001, Vol. 29, No. 3, 715–747

PRIOR INDUCTION IN LOG-LINEAR MODELS FOR GENERAL CONTINGENCY TABLE ANALYSIS[1]

BY R. KING AND S. P. BROOKS

University of Cambridge

Log-linear modelling plays an important role in many statistical applications, particularly in the analysis of contingency table data. With the advent of powerful new computational techniques such as reversible jump MCMC, Bayesian analyses of these models, and in particular model selection and averaging, have become feasible. Coupled with this is the desire to construct and use suitably flexible prior structures which allow efficient computation while facilitating prior elicitation. The latter is greatly improved in the case where priors can be specified on interpretable parameters about which relevant experts can express their beliefs.

In this paper, we show how the specification of a general multivariate normal prior on the log-linear parameters induces a multivariate log-normal prior on the corresponding cell counts of a contingency table. We derive the parameters of this distribution in an explicit practical form and state the corresponding mean and covariances of the cell counts. We discuss the importance of these results in terms of applying both uninformative and informative priors to the model parameters and provide an illustration in the context of the analysis of a 2^3 contingency table.

1. Introduction. The analysis of general k-way contingency table data is of interest in a wide variety of areas of statistical application. Model selection is notoriously difficult in such situations, since the number of models rises doubly exponentially with dimension. Obviously, exhaustive comparisons are impossible, but various computational techniques have been proposed for obtaining posterior model probabilities for problems of this sort, depending upon the parameters upon which the analyst wishes to express prior opinions. With the introduction of powerful new computational techniques these forms of analysis have been greatly simplified and are becoming increasingly common in the applied literature.

In undertaking a Bayesian analysis of contingency table data, it is necessary to specify priors either for the cell counts (which could alternatively be expressed in terms of the cell probabilities and total cell count) or, equivalently, the log-linear parameters. Madigan and York (1997) choose to choose to place hyper-Dirichlet priors [Dawid and Lauritzen (1993)] on the cell probabilities which has the advantage that priors of this form allow a factorization of the likelihood through the identification of cliques within the corresponding model graph. Giudici, Green and Tarantola (1999) illustrate how this decomposition

Received November 1999; revised February 2001.

[1]Supported in part by UK Engineering and Physical Sciences Research Council.

AMS 2000 *subject classifications*. Primary 62F15, 62H99.

Key words and phrases. Bayesian analysis, contingency table, multivariate normal, prior elicitation.

leads to local computations for the cell probabilities, improving the computational efficiency of individual parameter updates.

However, there are also several disadvantages. The first is that by adopting a prior of this form, we restrict ourselves to decomposable models, which may or may not be appropriate for any particular problem. For example, Dellaportas and Forster (1999) provide a simple 2^6 contingency table example where the posterior model probabilities of hierarchical log-linear models, graphical models and decomposable models are compared. The corresponding results show that the most probable log-linear model is approximately 700 times more likely than the most probable decomposable model. Second, the hyper-Dirichlet prior requires a very large number of hyper-parameters to be specified, which must be hyperconsistent and ideally compatible across all models. This is often achieved by assuming that the distributions on any clique are obtained by marginalization from a unique distribution on a complete graph, but it certainly adds an additional level of complexity to the prior specification problem. Finally, though each parameter update involves only local computation, there may be a large number of them, so that a large number of local updates are required.

An alternative approach is to specify a prior distribution on the log-linear parameters [Knuiman and Speed (1998), Evans, Gilula and Guttman (1993), Dellaportas and Forster (1999)]. This is the approach adopted by King and Brooks (2001) who place independent normal priors on the log-linear parameters for a 2^k contingency table analysis. The advantage of this form of prior is both the conceptual and computational simplicity as well as the fact that we are no longer restricted to decomposable models. The disadvantages are that there is no longer a decomposition of the likelihood, and log-linear parameter updates involve global rather than local computations. However, this drawback is mitigated by the fact that there are typically far fewer log-linear parameters than there are cells, so correspondingly fewer updates are required.

In practice, the choice of prior should have more to do with accurately expressing prior beliefs than mathematical convenience and typically prior information may be available on both log-linear parameters and cell counts [King and Brooks (2001)]. For example, we may be interested in the total cell count as well as the presence of some dependence between two or more sources. However, the prior must be placed on either the cell probabilities or the log-linear parameters. In practice, prior information may be available in terms of both parameterizations and so it is vital that if a prior is placed on one set of parameters, then the prior induced on the other must also be sensible. To check this condition an explicit form for the induced prior must be available. The injective nature of the log-linear modelling structure means that this is only possible if priors are placed on the log-linear parameters and in this paper we provide a practical form (and discuss the properties) of the corresponding prior induced on the cell counts or probabilities. We therefore provide the results necessary to ensure that any prior specified is consistent with prior beliefs under both parameterizations of the model.

Some results along these lines already exist [Knuiman and Speed (1988), Forster (1992)]. However, these results concern the modelling of the cell probabilities and impose the usual "sum to zero" constraints through the prior, restricting the range of priors to those of a certain form. When eliciting priors, this is an additional complication which can be avoided by imposing the constraints directly through the design matrix linking the cell counts to the log-linear parameters. We return to the idea of modelling cell probabilities in Section 6. In this context, our approach has the additional advantage that a general correlation structure can be directly imposed on the log-linear parameters without having to limit ourselves to those prior structures that impose the necessary constraints.

It is fairly obvious that under a general multivariate normal prior on the log-linear parameters, the corresponding cell counts have a multivariate log-normal distribution but in this paper we derive the associated parameter values in a usable form. From this distribution, we are able to calculate the induced prior mean and covariances of the cell counts, which can then be used to check that the prior chosen for the log-linear parameters, as well as reflecting our prior about those parameters, is also consistent with our prior beliefs about the associated cell counts.

We begin, in Section 2, by introducing the notation to be used throughout the paper. We then derive the form of the corresponding design matrix in Section 3 and discuss some of its properties. In Section 4, we present our main results. We derive the form of the multivariate normal distribution induced on the log of the cell counts as well as some associated moments of the cell counts and the total cell count which can be used to check consistency with prior beliefs. In Section 5 we investigate the effect of both noninformative and informative priors placed on the log-linear parameters. We begin by showing that a diffuse prior placed on the log-linear parameters is similarly diffuse in terms of the cell counts allowing for consistency in the presence of weak prior information. We then discuss the use of informative priors in the context of the analysis of an incomplete 2^3 contingency table. Finally, in Section 6, we provide some general discussion of the usefulness and general context of these results.

2. The log-linear model. We assume that data is obtained from a set of factors or sources, S. We let $|S|$ denote the number of sources and label each source such that $S = \{S_\gamma : \gamma = 1, \ldots, |S|\}$. We denote the set of levels for source S_γ by K_γ, for $\gamma = 1, \ldots, |S|$. Further, we let the levels in source S_γ be $\{1, \ldots, |K_\gamma|\}$, so that there are $|K_\gamma| \geq 1$ levels for $\gamma = 1, \ldots, |S|$. Then we can produce a contingency table with each cell representing a possible combination of responses from the sources. The cells can be expressed as the set $K = K_1 \times \cdots \times K_{|S|}$, so that each of the $|K|$ cells can be indexed by $\mathbf{k} \in K$ with corresponding cell probability $p_{\mathbf{k}}$. Similarly, for $\mathbf{k} \in K$, we let $n_{\mathbf{k}}$ denote the cell count for the corresponding cell. Finally we define $\mathscr{P}(S)$ to be the set of subsets, (or power set) of S, so that $\mathscr{P}(S) = \{s : s \subseteq S\}$. Then, to represent a log-linear model, we use the index $m \subseteq \mathscr{P}(S)$, where m lists the log-linear

terms present in the model, with $|m|$ denoting the number of terms within model m. We allow for a constant log-linear term, by the inclusion of the empty set in $\mathscr{P}(S)$, that is, $\varnothing \in \mathscr{P}(S)$.

For element $c \in \mathscr{P}(S)$, we let $\mathbf{M}^c$ be the set of all possible combinations of levels for the sources contained within c, that do not include the last level (generally the highest level). Then, in general, we set $\mathbf{M}^c = \{\mathbf{m}_1^c, \ldots, \mathbf{m}_{|\mathbf{M}^c|}^c\}$. For example, suppose that $S = \{S_1, S_2\}$, where $K_1 = K_2 = \{1, 2, 3\}$. Then $\mathscr{P}(S) = \{\varnothing, \{S_1\}, \{S_2\}, \{S_1, S_2\}\}$. For $c = \{S_1, S_2\}$, we have that $\mathbf{M}^c = \{\mathbf{m}_1^c, \mathbf{m}_2^c, \mathbf{m}_3^c, \mathbf{m}_4^c\} = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$, where (i, j) corresponds to the ith level of source S_1 and the jth level of source S_2. The ordering of the levels is generally unimportant so long as it is consistent, but in general a natural ordering will nearly always be apparent. For each possible element, $c \in \mathscr{P}(S)$, we denote the associated log-linear vector, by $\boldsymbol{\theta}^c = (\theta_{\mathbf{m}_1}^c, \ldots, \theta_{\mathbf{m}_{|\mathbf{M}^c|}}^c)^T$, dropping the dependence on c of $\{\mathbf{m}_1, \ldots, \mathbf{m}_{|\mathbf{M}^c|}\}$, since it is implicit from the context. Under the usual "sum to zero" constraints the log-linear parameters $\theta_{\mathbf{j}}^c$, $\mathbf{j} \in \mathbf{M}^c$, are linearly independent. Hence, given a model, the log-linear parameters are uniquely identifiable.

If we let m be an ordered subset of $\mathscr{P}(S)$, then we denote the associated log-linear parameter vector for model m by $\boldsymbol{\theta}_m = ((\boldsymbol{\theta}^{c_1})^T, \ldots, (\boldsymbol{\theta}^{c_{|m|}})^T)$ where $\boldsymbol{\theta}^{c_i}$ denotes the parameter vector associated with the ith element. We can describe the relationship between the expected cell counts and log-linear parameters as

$$\ln n_{\mathbf{k}} = \sum_{c \in m} (\mathbf{I}^c(\mathbf{k}))^T \boldsymbol{\theta}^c, \qquad \mathbf{k} \in K, \tag{1}$$

where $(\mathbf{I}^c(\mathbf{k}))^T = (I_{\mathbf{m}_1}^c(\mathbf{k}), \ldots, I_{\mathbf{m}_{|\mathbf{M}^c|}}^c(\mathbf{k}))$, and $I_{\mathbf{j}}^c(\mathbf{k}) = 0, \pm 1, \forall\, \mathbf{j} \in \mathbf{M}^c$ and $c \in \mathscr{P}(S)$, are functions ensuring that the usual conditions for identifiability are observed. We define $\mathbf{I}^{\varnothing}(\mathbf{k}) = \mathbf{1}_{|K|}$, $\forall\, \mathbf{k} \in K$, that is, the vector of length $|K|$, such that each element is equal to unity. We denote by $L_l^{\mathbf{e}}$ the set $\{\mathbf{k}: k(\gamma) = e(\gamma);\ \forall\, \gamma \neq l; k(l) \in K_l\}$, where $k(\gamma)$ and $e(\gamma)$ represent the elements corresponding to source S_γ of the vectors $\mathbf{k}$ and $\mathbf{e}$, respectively. We can then specify the identifiability conditions as

$$\sum_{\mathbf{k} \in L_l^{\mathbf{e}}} \mathbf{I}^c(\mathbf{k}) = \mathbf{0} \qquad \forall\, l\colon S_l \in c, \forall\, \mathbf{e} \in K \quad \text{and } \forall\, c \in \mathscr{P}(S)\backslash\varnothing \tag{2}$$

with $I_{\mathbf{0}}^c(\mathbf{0}) = \mathbf{1}$ for uniqueness. We assume that each source S_γ, $\gamma = 1, \ldots, |S|$ has levels denoted by $\{1, \ldots, |K_\gamma|\}$, where $|K_\gamma|$ is the total number of levels of the source. We can express this function explicitly as follows.

LEMMA 2.1. *Denote by* $I_{\mathbf{j}}^c(\mathbf{k})$, *the element of* $\mathbf{I}^c(\mathbf{k})$ *corresponding to* $\mathbf{j} \in \mathbf{M}^c$, *for* $c \in \mathscr{P}(S)\backslash\varnothing$ *and* $\mathbf{k} \in K$. *Then a solution to* (2) *which satisfies the constraint that* $I_{\mathbf{1}}^c(\mathbf{1}) = 1\, \forall\, c$ *is given by*

$$I_{\mathbf{j}}^c(\mathbf{k}) = \prod_{\gamma\colon S_\gamma \in c} [\mathscr{I}(k(\gamma) = j(\gamma)) - \mathscr{I}(k(\gamma) = |K_\gamma|)], \tag{3}$$

where $\mathcal{I}(\cdot)$ *is the indicator function and* $j(\gamma)$ *is the* γ*th element of* $\mathbf{j}$, *that is, the element corresponding to source* S_γ.

The proof is given in the Appendix.

Having established a notation for the models and their associated parameters, we now consider how we might place priors on the parameters in the two distinct cases where we first have little or no prior information and, second, where prior information is available and we wish to incorporate this within the analysis.

The most convenient manner to express the relationship between the expected cell counts and log-linear parameters, under any model m, is in terms of a design matrix, so that

$$\ln \mathbf{n} = \sum_{c \in m} X^c \boldsymbol{\theta}^c, \tag{4}$$

where $\ln \mathbf{n}$ is the vector consisting of elements $\ln n_{\mathbf{k}}, \mathbf{k} \in K$ and X^c denotes the design matrix associated with parameter $c \in \mathcal{P}(S)$. We can then denote the complete design matrix by X_m, where,

$$X_m = (X^{c_1} \mid X^{c_2} \mid \cdots \mid X^{c_{|m|}}),$$

so that,

$$\sum_{c \in m} X^c \boldsymbol{\theta}^c = X_m \boldsymbol{\theta}_m.$$

From this representation it is clear that if we place a multivariate normal prior on $\boldsymbol{\theta}_m$, then we induce a multivariate normal prior on $\log \mathbf{n}$. Consider the model $m \subseteq \mathcal{P}(S)$, with corresponding design matrix X_m. If we take a multivariate normal $N(\boldsymbol{\mu}_m, \Sigma_m)$ for the log-linear parameters, then the log cell counts have a $N(\boldsymbol{\mu}, \Sigma)$ distribution where

$$\boldsymbol{\mu} = X_m \boldsymbol{\mu}_m$$

and

$$\Sigma = X_m \Sigma_m X_m^T.$$

In this paper, we are concerned with deriving the exact form of Σ. In order to do this, we must first establish an ordering for the rows and columns of the design matrix X_m and this we do in Section 3.1. Once an order has been constructed, it is easy to show that if we let $\mathbf{x}_{\mathbf{i}}^c$ and $\mathbf{x}_{\mathbf{j}}^d$ be the columns of the design matrix X_m corresponding to level $\mathbf{i} \in \mathbf{M}^c$ and $\mathbf{j} \in \mathbf{M}^d$, respectively, then,

$$\begin{aligned}
\Sigma &= X_m \Sigma_m X_m^T \\
&= \sum_{d \in m} \sum_{\mathbf{j} \in \mathbf{M}^d} \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \mathbf{x}_{\mathbf{j}}^d \sigma_{\mathbf{ji}}^{dc} (\mathbf{x}_{\mathbf{i}}^c)^T \quad \text{by definitions of } X_m \text{ and } \Sigma_m \\
&= \sum_{d \in m} \sum_{\mathbf{j} \in \mathbf{M}^d} \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \sigma_{\mathbf{ij}}^{cd} \mathbf{x}_{\mathbf{j}}^d (\mathbf{x}_{\mathbf{i}}^c)^T \quad \text{since } \sigma_{\mathbf{ji}}^{dc} = \sigma_{\mathbf{ij}}^{cd}, \qquad (5)
\end{aligned}$$

where $\sigma_{\mathbf{ij}}^{cd}$ denotes the element of Σ_m corresponding to the covariance between sources c and d at levels $\mathbf{i}$ and $\mathbf{j}$.

Thus, the elements of Σ can be expressed as a sum over products of elements of the design matrix and the corresponding prior covariance. In Section 3 we first derive a form for the elements of the design matrices X^c, $c \in m$ and then a form for the product term in (5).

3. The design matrix. In this section, we provide an explicit expression for the design matrices and for the products of columns with transposed columns of the design matrix, for all possible parameters. However, we first begin by establishing an ordering for the log-linear parameters and the cell counts which will be used to relate specific entries in the design matrix to the corresponding cell count and log-linear parameter.

3.1. *Ordering the parameters.* For any parameter vector $\boldsymbol{\theta}^c$, $c \in m \subseteq \mathscr{P}(S)$, we define a corresponding matrix X^c of dimension

$$\left(\prod_{\gamma=1}^{|S|} |K_\gamma| \times \prod_{\gamma:\, S_\gamma \in c} (|K_\gamma| - 1)\right).$$

This is then the design matrix corresponding to parameter $c \in m$ with the null product (occurring when $m = \varnothing$ in which case $\{\gamma\colon S_\gamma \in c\} = \varnothing$) being defined to be 1. Then, within each $\boldsymbol{\theta}^c$ we adopt the natural lexicographic ordering for the log-linear parameters in $\boldsymbol{\theta}^c$. Now, the log-linear parameter vector $\boldsymbol{\theta}^c$ contains the log-linear parameters $\theta_{\mathbf{j}}^c$ such that $\mathbf{j} \in \mathbf{M}^c$ with element $C(c, \mathbf{j})$ of $\boldsymbol{\theta}^c$ corresponding to parameter $\theta_{\mathbf{j}}^c$. This ordering function C is defined as follows. If we have sources $S = \{S_1, S_2, \ldots\}$ where source S_γ has levels $\{1, \ldots, |K_\gamma|\}$, for $\gamma = 1, \ldots, |S|$, then

$$C(c, \mathbf{j}) = 1 + \sum_{\gamma:\, S_\gamma \in c} \left(\frac{j(\gamma) - 1}{|K_\gamma| - 1} \prod_{l \le \gamma} (|K_l| - 1) \right),$$

where $j(\gamma)$ is the element of $\mathbf{j}$, corresponding to source S_γ.

For example, suppose that $c = \{S_1, S_2, S_3\}$, with $|K_1| = 2$, $|K_2| = 3$ and $|K_3| = 4$. We then order the corresponding log-linear parameter vector,

$$\boldsymbol{\theta}^c = (\theta_{(111)}^c, \theta_{(121)}^c, \theta_{(112)}^c, \theta_{(122)}^c, \theta_{(113)}^c, \theta_{(123)}^c)^T.$$

Intuitively, we order the parameters by translating $\mathbf{j}$ into a decimal integer by treating its elements as if they represented a number in base $b = \max_{\gamma:\, S_\gamma \in c} |K_\gamma|$. Here element $j(i)$ corresponds to coefficient corresponding to b^{i-1}. The parameters are then ordered from smallest to largest in decimal value and renumbered so that no gaps in the numbering occur.

We order the cell counts similarly, so that for the vector $\mathbf{n}$, we allocate element $R(\mathbf{k})$ to be the value corresponding to cell $\mathbf{k} \in K$, by ordering the

sources in an analogous way. We set

$$R(\mathbf{k}) = 1 + \sum_{\gamma=1}^{|S|} \left(\frac{k(\gamma) - 1}{|K_\gamma|} \prod_{l \le \gamma} |K_l| \right),$$

where $k(\gamma)$ is the element of $\mathbf{k}$ corresponding to source S_γ.

These orderings thus provide a structure for the design matrix X^c, so that row $R(\mathbf{k})$ corresponds to cell $n_{\mathbf{k}}$ (and remains unchanged for all $c \in m$) and column $C(c, \mathbf{j})$ corresponds to parameter $\theta^c_{\mathbf{j}}$. We next provide an explicit form for the elements of the design matrix.

3.2. *Elements of the design matrix.* The following lemma provides the definition of a general design matrix X^c in terms of a Kronecker product of indicator functions.

LEMMA 3.1. *For $c \in \mathscr{P}(S)$ then,*

$$X^c = \bigotimes_{\gamma=1}^{|S|} \Big[\mathscr{I}(S_\gamma \in c) L_{|K_\gamma|} + \mathscr{I}(S_\gamma \notin c) \mathbf{1}_{|K_\gamma|} \Big], \tag{6}$$

where $\mathscr{I}(\cdot)$ is the indicator function, $\mathbf{1}_{|K_\gamma|}$ is the vector of length $|K_\gamma|$ with each element equal to unity and

$$L_{|K_\gamma|} = \begin{pmatrix} I_{|K_\gamma|-1} \\ -\mathbf{1}^T_{|K_\gamma|-1} \end{pmatrix},$$

where $I_{|K_\gamma|-1}$ is the identity matrix of dimension $|K_\gamma| - 1$.

Here, for a matrix function f, the multiple Kronecker product is defined so that

$$\bigotimes_{\gamma=1}^{|S|} f(S_\gamma) = f(S_{|S|}) \otimes f(S_{|S|-1}) \otimes \cdots \otimes f(S_1).$$

The proof is given in the Appendix.

Thus, we obtain an explicit form for the design matrix. We now derive a natural form for the product of any column of the design matrix with any transposed column so that we may calculate the value of Σ in (5). This will be used in the next section to derive our main result relating the priors induced upon the expected cell counts by a multivariate normal prior placed on the log-linear parameters.

We begin by introducing some additional notation. We first define the matrix $G^\gamma_{\mathbf{i}}$ whose (m, n)th element is given by

$$g^\gamma_{\mathbf{i}}(m, n) = \begin{cases} 1, & \text{if } (m, n) = (i(\gamma), i(\gamma)); \text{ or } (|K_\gamma|, |K_\gamma|), \\ -1, & \text{if } (m, n) = (i(\gamma), |K_\gamma|); \text{ or } (|K_\gamma|, i(\gamma)), \\ 0, & \text{otherwise.} \end{cases}$$

We also define a matrix $A_{\mathbf{ij}}^{cd}(\gamma)$ whose (m,n)th element can be defined under three separate cases as follows.

Case I. $S_\gamma \in c$; $S_\gamma \in d$,

$$a_{\mathbf{ij}}^{cd}(m,n) = \begin{cases} 1, & \text{if } (m,n) = (i(\gamma), j(\gamma)); (j(\gamma), i(\gamma)); \text{ or } (|K_\gamma|, |K_\gamma|), \\ 0, & \text{otherwise.} \end{cases}$$

Case II. $S_\gamma \in c$; $S_\gamma \notin d$,

$$a_{\mathbf{ij}}^{cd}(m,n) = \begin{cases} 1, & \text{if } (m,n) = (r, i(\gamma)), \text{ for } r = 1, \ldots, |K_\gamma| - 1, \\ -1, & \text{if } (m,n) = (|K_\gamma|, |K_\gamma|), \\ 0, & \text{otherwise.} \end{cases}$$

Case III. $S_\gamma \notin c$; $S_\gamma \in d$,

$$a_{\mathbf{ij}}^{cd}(m,n) = \begin{cases} 1, & \text{if } (m,n) = (j(\gamma), j(\gamma)), \\ -1, & \text{if } (m,n) = (|K_\gamma|, |K_\gamma|), \\ 0, & \text{otherwise.} \end{cases}$$

The matrix is left undefined if $c, d \notin S_\gamma$.

Finally, we define two Kronecker product matrices,

$$W_{\mathbf{ij}}^{cd} = \bigotimes_{\gamma=1}^{|S|} \Big[\mathcal{I}(S_\gamma \in c \cup d) A_{\mathbf{ij}}^{cd}(\gamma) + (1 - \mathcal{I}(S_\gamma \in c \cup d)) I_{|K_\gamma|} \Big] \tag{7}$$

and

$$V_{\mathbf{i}}^{c} = \bigotimes_{\gamma=1}^{|S|} \Big[\mathcal{I}(S_\gamma \in c) G_{\mathbf{i}}^{\gamma} + (1 - \mathcal{I}(S_\gamma \in c)) J_{|K_\gamma|} \Big], \tag{8}$$

where $I_{|K_\gamma|}$ denotes the identity matrix of dimension $|K_\gamma|$ and $J_{|K_\gamma|}$ denotes the $|K_\gamma| \times |K_\gamma|$ matrix with each entry set to 1.

The following result states that the product of any column of the design matrix with any transposed column is equal to the product of the corresponding W and V matrices.

LEMMA 3.2. *If we let* $\mathbf{x}_{\mathbf{j}}^{d}$ *be the column of the design matrix,* X^d, *corresponding to level* $\mathbf{j} \in \mathbf{M}^d$, *with the analogous definition for* $\mathbf{x}_{\mathbf{i}}^{c}$, *for* $c, d \in \mathcal{P}(S)$, *then*

$$\mathbf{x}_{\mathbf{j}}^{d} (\mathbf{x}_{\mathbf{i}}^{c})^T = W_{\mathbf{ij}}^{cd} V_{\mathbf{i}}^{c}, \tag{9}$$

where $W_{\mathbf{ij}}^{cd}$ *and* $V_{\mathbf{i}}^{c}$ *are given in* (7) *and* (8).

The proof is given in the Appendix.

We now have a form for the matrix product in (5) that can be used to obtain the matrix Σ. In the next section we provide our main results, deriving an analytic form for the mean vector and covariance matrix of the induced prior on $\log \mathbf{n}$ and examining some of the properties of this induced distribution.

4. The induced prior on the cell counts. As discussed earlier, it is simple to demonstrate that when we place a multivariate normal prior distribution on the log-linear parameters the corresponding prior for the log of the expected cell count is also multivariate normal. The following theorem provides an exact analytic form for the mean vector and covariance matrix of the induced prior on $\log \mathbf{n}$.

THEOREM 4.1. *Suppose that for model $m \subseteq \mathscr{P}(S)$, we take a multivariate normal prior for the model parameters*,

$$\boldsymbol{\theta}_m \sim \mathscr{N}(\boldsymbol{\mu}_m, \Sigma_m),$$

where the elements of $\boldsymbol{\mu}_m$ are given by $\mu_{\mathbf{i}}^c$ representing the mean of $\theta_{\mathbf{i}}^c$, and the elements of Σ_m are given by $\sigma_{\mathbf{ij}}^{cd}$, representing the covariance between log-linear parameters $\theta_{\mathbf{i}}^c$ and $\theta_{\mathbf{j}}^d$, such that $c, d \in m$ and $\mathbf{i} \in \mathbf{M}^c$, $\mathbf{j} \in \mathbf{M}^d$.

Then, under the model described in (1), *the mean vector $\boldsymbol{\mu}$ of $\ln \mathbf{n}$ has elements*

$$\mu_{\mathbf{k}} = \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \mu_{\mathbf{i}}^c I_{\mathbf{i}}^c(\mathbf{k}),$$

and the corresponding covariance matrix of the log of the cell counts can be expressed as

$$\Sigma = \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} B_{\mathbf{i}}^c(m) V_{\mathbf{i}}^c,$$

where $V_{\mathbf{i}}^c$ is given in (8) *and*

$$B_{\mathbf{i}}^c(m) = \sum_{d \in m} \sum_{\mathbf{j} \in \mathbf{M}^d} \sigma_{\mathbf{ij}}^{cd} W_{\mathbf{ij}}^{cd},$$

with $W_{\mathbf{ij}}^{cd}$ given in (7).

PROOF. Clearly, comparing (1) and (4), the row of X^c corresponding to cell $\mathbf{k} \in K$ is equal to $I^c(\mathbf{k})^T$. Therefore if $x_{\mathbf{j}}^c(\mathbf{k})$ denotes the element of the matrix X^c, corresponding to cell $\mathbf{k} \in K$ and sources c at level $\mathbf{j} \in \mathbf{M}^c$, we have,

$$x_{\mathbf{j}}^c(\mathbf{k}) = I_{\mathbf{j}}^c(\mathbf{k}). \tag{10}$$

Then, by definition we have that $\boldsymbol{\mu} = X_m \boldsymbol{\mu}_m$, therefore,

$$\mu_{\mathbf{k}} = \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \mu_{\mathbf{i}}^c x_{\mathbf{i}}^c(\mathbf{k}) = \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \mu_{\mathbf{i}}^c I_{\mathbf{i}}^c(\mathbf{k}),$$

by (10).

Similarly,

$$\begin{aligned}\Sigma &= \sum_{d\in m}\sum_{\mathbf{j}\in\mathbf{M}^d}\sum_{c\in m}\sum_{\mathbf{i}\in\mathbf{M}^c}\sigma_{\mathbf{ij}}^{cd}\mathbf{x}_{\mathbf{j}}^{d}(\mathbf{x}_{\mathbf{i}}^{c})^T \quad \text{by (5)} \\ &= \sum_{d\in m}\sum_{\mathbf{j}\in\mathbf{M}^d}\sum_{c\in m}\sum_{\mathbf{i}\in\mathbf{M}^c}\sigma_{\mathbf{ij}}^{cd}W_{\mathbf{ij}}^{cd}V_{\mathbf{i}}^{c} \quad \text{using Lemma 3.2} \\ &= \sum_{c\in m}\sum_{\mathbf{i}\in\mathbf{M}^c}V_{\mathbf{i}}^{c}\sum_{d\in m}\sum_{\mathbf{j}\in\mathbf{M}^d}\sigma_{\mathbf{ij}}^{cd}W_{\mathbf{ij}}^{cd} \\ &= \sum_{c\in m}\sum_{\mathbf{i}\in\mathbf{M}^c}V_{\mathbf{i}}^{c}B_{\mathbf{i}}^{c}(m) \quad \text{by definition of } B_{\mathbf{i}}^{c}(m). \end{aligned} \tag{11}$$

□

As a corollary to this theorem, we also obtain a form for the density of the corresponding cell counts.

COROLLARY 4.1. *With a multivariate normal prior placed on the log-linear parameters, the corresponding density induced on the cell counts,* $\mathbf{n}$*, is given by the multivariate log-normal distribution with density function*

$$f_{\mathbf{n}}(\mathbf{n}) = \frac{1}{|\Sigma|^{1/2}(2\pi)^{|K|/2}}\prod_{\mathbf{k}\in K}\left(\frac{1}{n_{\mathbf{k}}}\right)\exp\left(-\frac{1}{2}(\ln\mathbf{n}-\boldsymbol{\mu})^T(\Sigma)^{-1}(\ln\mathbf{n}-\boldsymbol{\mu})\right). \tag{12}$$

The proof of this result follows directly from Theorem 4.1 using a simple transformation of variables.

Next we obtain some useful distributional results for the cell counts, and in particular the mean vector and covariance matrix associated with the density for the cell counts given in Corollary 4.1.

LEMMA 4.1. *Given that* $\mathbf{n}$ *has a multivariate log-normal* $(\boldsymbol{\mu},\Sigma)$ *distribution, then for* $\mathbf{k}_1,\mathbf{k}_2\in K$,

$$\mathbb{E}(n_{\mathbf{k}_1}) = \exp\left(\mu_{\mathbf{k}_1}+\frac{\sigma_{\mathbf{k}_1}^2}{2}\right); \tag{13}$$

$$\text{Var}\,(n_{\mathbf{k}_1}) = \exp\left(2\mu_{\mathbf{k}_1}+\sigma_{\mathbf{k}_1}^2\right)(\exp(\sigma_{\mathbf{k}_1}^2)-1); \quad \textit{and} \tag{14}$$

$$\text{Cov}(n_{\mathbf{k}_1},n_{\mathbf{k}_2}) = \exp\left(\mu_{\mathbf{k}_1}+\mu_{\mathbf{k}_2}+\tfrac{1}{2}(\sigma_{\mathbf{k}_1}^2+\sigma_{\mathbf{k}_2}^2)\right)(\exp(\sigma_{\mathbf{k}_1\mathbf{k}_2})-1), \tag{15}$$

where $\mu_{\mathbf{k}_1}$ *and* $\sigma_{\mathbf{k}_1}^2$ *are the mean and variance of* $\ln n_{\mathbf{k}_1}$*, respectively, and* $\sigma_{\mathbf{k}_1\mathbf{k}_2}$ *is the covariance between* $\ln n_{\mathbf{k}_1}$ *and* $\ln n_{\mathbf{k}_2}$.

PROOF. The expectation in (13) can be derived as a simple extension of the properties of the univariate log-normal distribution. See Aitchison and Brown (1957), for example. Clearly the variance in (14) will be obtained as a simple corollary of the covariance result of (15). To prove the final result, suppose that we let the $|K|$-vector $\ln\mathbf{Z}$ have a multivariate normal distribution with

mean $\boldsymbol{\mu}$, and covariance matrix Σ. Then, for any $|K|$-vector $\mathbf{a}$, say, we have the standard property

$$\mathbf{a}^T \ln \mathbf{Z} \sim N(\mathbf{a}^T \boldsymbol{\mu}, \mathbf{a}^T \Sigma \mathbf{a}).$$

But, by definition and using the general properties of logarithms,

$$\mathbf{a}^T \ln \mathbf{Z} = \sum_{i=1}^{|K|} a_i \ln Z_i = \ln\left(\prod_{i=1}^{|K|} Z_i^{a_i} \right),$$

so that

$$\ln\left(\prod_{i=1}^{|K|} Z_i^{a_i} \right) \sim N(\mathbf{a}^T \boldsymbol{\mu}, \mathbf{a}^T \Sigma \mathbf{a}). \tag{16}$$

[cf. Aitchison and Brown (1957), Theorem 2.4]

We can now easily calculate the covariances of the expected cell counts using (16). Suppose that we want to calculate the covariance between the expected cell counts $n_{\mathbf{k}_1}$ and $n_{\mathbf{k}_2}$. Then, in (16), we consider using the vector $\mathbf{a}$, such that the elements corresponding to cells $\mathbf{k}_1$ and $\mathbf{k}_2$ are set to unity, with all other entries equal zero. We then obtain

$$\ln(n_{\mathbf{k}_1} n_{\mathbf{k}_2}) \sim N(\mu_{\mathbf{k}_1} + \mu_{\mathbf{k}_2}, \sigma^2_{\mathbf{k}_1} + \sigma^2_{\mathbf{k}_2} + 2\sigma_{\mathbf{k}_1 \mathbf{k}_2}).$$

So that the product $n_{\mathbf{k}_1} n_{\mathbf{k}_2}$ has a log-normal distribution, and from (13) we have

$$\mathbb{E}(n_{\mathbf{k}_1} n_{\mathbf{k}_2}) = \exp\left((\mu_{\mathbf{k}_1} + \mu_{\mathbf{k}_2}) + \frac{(\sigma^2_{\mathbf{k}_1} + \sigma^2_{\mathbf{k}_2} + 2\sigma_{\mathbf{k}_1 \mathbf{k}_2})}{2} \right). \tag{17}$$

Then, we can calculate the covariance of $n_{\mathbf{k}_1}$ and $n_{\mathbf{k}_2}$ to be

$$\begin{aligned}
\mathrm{Cov}(n_{\mathbf{k}_1}, n_{\mathbf{k}_2}) &= \mathbb{E}(n_{\mathbf{k}_1} n_{\mathbf{k}_2}) - \mathbb{E}(n_{\mathbf{k}_1}) \mathbb{E}(n_{\mathbf{k}_2}) \\
&= \exp\left((\mu_{\mathbf{k}_1} + \mu_{\mathbf{k}_2}) + \frac{(\sigma^2_{\mathbf{k}_1} + \sigma^2_{\mathbf{k}_2} + 2\sigma_{\mathbf{k}_1 \mathbf{k}_2})}{2} \right) \\
&\quad - \exp\left(\mu_{\mathbf{k}_1} + \frac{\sigma^2_{\mathbf{k}_1}}{2} \right) \exp\left(\mu_{\mathbf{k}_2} + \frac{\sigma^2_{\mathbf{k}_2}}{2} \right) \\
&= \exp\left(\mu_{\mathbf{k}_1} + \mu_{\mathbf{k}_2} + \frac{1}{2}(\sigma^2_{\mathbf{k}_1} + \sigma^2_{\mathbf{k}_2}) \right) (\exp(\sigma_{\mathbf{k}_1 \mathbf{k}_2}) - 1). \qquad \square
\end{aligned}$$

A corollary to Lemma 4.1 provides the prior mean and variance of the total cell count induced by the multivariate normal prior placed on the log-linear parameters.

COROLLARY 4.2. *With a multivariate normal prior placed on the log-linear parameters, the corresponding mean and variance of the total cell count, denoted by N, are given by*

$$\mathbb{E}(N) = \sum_{\mathbf{k}} \mathbb{E}(n_{\mathbf{k}}), \quad \text{and} \quad \text{Var}(N) = \sum_{\mathbf{k}_1, \mathbf{k}_2} \text{Cov}\,(n_{\mathbf{k}_1}, n_{\mathbf{k}_2}).$$

The proof follows immediately from standard formulae for the mean and variance for the sum of dependent random variables.

These results (in particular, Corollary 4.2 and Lemma 4.1) provide us with a usable form for the prior induced on the cell counts by the corresponding multivariate normal prior on the log-linear parameters. These can be used for the purposes of prior elicitation as discussed in the next section.

5. Prior specification—examples. In practice, there are two distinct forms of prior that one might want to use: an informative prior based upon expert opinion, or a vague prior which reflects broad a priori uncertainty with regard to the model parameters. We shall illustrate how our results can be of use in either context, beginning with the use of vague priors placed on the log-linear parameters.

5.1. *Placing vague priors on the log-linear parameters.* It is common practice to want to be able to place so-called vague or diffuse priors on model parameters. However, it is important to check that the prior is vague in terms of all important parameterizations, that is, both the log-linear parameters and the corresponding cell counts. This consistency property is not apparent in all modelling structures and so it is important to check that a vague prior placed on the log-linear parameters induces a similarly vague prior on all interpretable parameters. As an illustration of when this is not the case, suppose that we had a probability p about which we wished to be vague a priori, we might consider placing a standard uniform prior on p. However, if we were to set logit $p = \mu$, say, then in order to be vague, we might place a normal prior on μ with zero mean and a large variance. However, it is well known that this induces a prior on p which, asymptotically as the variance for μ increases, has point masses at the values of zero and one. The following lemma provides reassurance that the same problem does not arise in the context of our modelling of a general k-way contingency table by illustrating that by placing a vague prior on the log-linear parameters we induce a similarly vague prior on the cell counts.

LEMMA 5.1. *Suppose that for model $m \subseteq \mathscr{P}(S)$, $(m \neq \varnothing)$, we place a multivariate normal prior on the log-linear parameters, with finite mean and covariances. Then, as the prior variances on the log-linear parameters tend to infinity, the corresponding prior for the cell counts becomes flat over the positive real line and has correlation matrix asymptotically equal to the identity;*

that is,

$$\operatorname{Corr}(n_{\mathbf{k}_1}, n_{\mathbf{k}_2}) = \begin{cases} 1, & \mathbf{k}_1 = \mathbf{k}_2, \\ 0, & \mathbf{k}_1 \neq \mathbf{k}_2. \end{cases}$$

PROOF. From (11) in the proof of Theorem 4.1 we have that

$$\Sigma = \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \sum_{d \in m} \sum_{\mathbf{j} \in \mathbf{M}^d} \sigma_{\mathbf{ij}}^{cd} \mathbf{x}_{\mathbf{j}}^d (\mathbf{x}_{\mathbf{i}}^c)^T.$$

So that the covariance between $n_{\mathbf{k}_1}$ and $n_{\mathbf{k}_2}$ can be expressed as

$$\begin{aligned} \sigma_{\mathbf{k}_1\mathbf{k}_2} &= \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \sum_{d \in m} \sum_{\mathbf{j} \in \mathbf{M}^d} \sigma_{\mathbf{ij}}^{cd} x_{\mathbf{i}}^c(\mathbf{k}_2) x_{\mathbf{j}}^d(\mathbf{k}_1) \\ &= \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \sum_{d \in m} \sum_{\mathbf{j} \in \mathbf{M}^d} \sigma_{\mathbf{ij}}^{cd} I_{\mathbf{i}}^c(\mathbf{k}_2) I_{\mathbf{j}}^d(\mathbf{k}_1) \quad \text{by equation (10)} \\ &= \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \sigma_{\mathbf{ii}}^{cc} I_{\mathbf{i}}^c(\mathbf{k}_2) I_{\mathbf{i}}^c(\mathbf{k}_1) + \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \sum_{\mathbf{j} \in \mathbf{M}^c \setminus \{\mathbf{i}\}} \sigma_{\mathbf{ij}}^{cc} I_{\mathbf{i}}^c(\mathbf{k}_2) I_{\mathbf{j}}^c(\mathbf{k}_1) \\ &\quad + \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \sum_{d \in m \setminus c} \sum_{\mathbf{j} \in \mathbf{M}^d} \sigma_{\mathbf{ij}}^{cd} I_{\mathbf{i}}^c(\mathbf{k}_2) I_{\mathbf{j}}^d(\mathbf{k}_1). \end{aligned} \tag{18}$$

However, since we have a finite covariance between each of the log-linear parameters, the last two terms in (18) are equal to a finite constant, which we denote by $D(\mathbf{k}_1, \mathbf{k}_2)$. So that for cell $\mathbf{k} \in K$, we can write the variance of the expected cell count as,

$$\sigma_{\mathbf{k}}^2 = \sum_{c \in m} \sum_{\mathbf{i} \in \mathbf{M}^c} \sigma_{\mathbf{ii}}^{cc} (I_{\mathbf{i}}^c(\mathbf{k}))^2 + D(\mathbf{k}, \mathbf{k}).$$

Clearly, as $\sigma_{\mathbf{ii}}^{cc}$ tends to infinity for each log-linear parameter within the model, then $\sigma_{\mathbf{k}}$ also tends to infinity [since $I_{\mathbf{i}}^c(\mathbf{k})$ takes at least one nonzero value for $c \in m$ and $\mathbf{i} \in \mathbf{M}^c$].

Then, from (13) and (14) of Lemma 4.1, it is clear that as the prior variance on the log-linear parameter tends to infinity, then both the mean and variance also tend towards infinity. However, by considering the ratio of the mean and standard deviation for each possible cell $\mathbf{k} \in K$, it is clear that the mean tends to infinity more slowly, since,

$$\frac{\mathbb{E}(n_{\mathbf{k}})}{SD(n_{\mathbf{k}})} = \frac{\exp\left(\mu_{\mathbf{k}} + \frac{\sigma_{\mathbf{k}}^2}{2}\right)}{\sqrt{\exp(2\mu_{\mathbf{k}} + \sigma_{\mathbf{k}}^2)(\exp(\sigma_{\mathbf{k}}^2) - 1)}} = \frac{1}{\sqrt{\exp(\sigma_{\mathbf{k}}^2) - 1}} \to 0,$$

as $\sigma_{\mathbf{ii}}^{cc} \to \infty$ for all $c \in m$ and $\mathbf{i} \in \mathbf{M}^c$, and this implies that $\sigma_{\mathbf{k}}^2 \to \infty$, by the above argument. This means that regardless of the model imposed, the corresponding prior on the expected cell counts is flat over the positive real line.

We now consider the correlation structure of the prior for the expected cell counts. We have that for the general cells $\mathbf{k}_1, \mathbf{k}_2 \in K$,

$$\begin{aligned}
\operatorname{Corr}(n_{\mathbf{k}_1}, n_{\mathbf{k}_2}) &= \frac{\operatorname{Cov}(n_{\mathbf{k}_1}, n_{\mathbf{k}_2})}{\sqrt{\operatorname{Var}(n_{\mathbf{k}_1})\operatorname{Var}(n_{\mathbf{k}_2})}} \\
&= \frac{\exp(\mu_{\mathbf{k}_1}+\mu_{\mathbf{k}_2}+\frac{1}{2}(\sigma^2_{\mathbf{k}_1}+\sigma^2_{\mathbf{k}_2}))(\exp(\sigma_{\mathbf{k}_1\mathbf{k}_2})-1)}{\sqrt{\exp(2\mu_{\mathbf{k}_1}+\sigma^2_{\mathbf{k}_1})(\exp(\sigma^2_{\mathbf{k}_1}-1))\exp(2\mu_{\mathbf{k}_2}+\sigma^2_{\mathbf{k}_2})(\exp(\sigma^2_{\mathbf{k}_2}-1))}} \\
&\qquad \text{from (15)} \\
&= \frac{\exp(\sigma_{\mathbf{k}_1\mathbf{k}_2})-1}{\sqrt{(\exp(\sigma^2_{\mathbf{k}_1})-1)(\exp(\sigma^2_{\mathbf{k}_2})-1)}} \\
&\to \frac{\exp(\sigma_{\mathbf{k}_1\mathbf{k}_2})}{\exp(\frac{1}{2}[\sigma^2_{\mathbf{k}_1}+\sigma^2_{\mathbf{k}_2}])} \qquad (19) \\
&= \exp\left(\sigma_{\mathbf{k}_1\mathbf{k}_2}-\frac{1}{2}\left[\sigma^2_{\mathbf{k}_1}+\sigma^2_{\mathbf{k}_2}\right]\right). \qquad (20)
\end{aligned}$$

The limit here is taken as $\sigma^{cc}_{\mathbf{ii}} \to \infty$ for all $c \in m$ and $\mathbf{i} \in \mathbf{M}^c$, which implies that $\sigma^2_{\mathbf{k}} \to \infty$. We can use the expression for the elements of Σ, given in (18), to compute this asymptotic correlation.

The numerator in (19) can be expressed as

$$\exp(\sigma_{\mathbf{k}_1\mathbf{k}_2}) = \exp\left(\sum_{c\in m}\sum_{\mathbf{i}\in\mathbf{M}_c}\sigma^{cc}_{\mathbf{ii}} I^c_{\mathbf{i}}(\mathbf{k}_1) I^c_{\mathbf{i}}(\mathbf{k}_2) + D(\mathbf{k}_1, \mathbf{k}_2)\right)$$

and the denominator as

$$\begin{aligned}
\exp\left(\tfrac{1}{2}[\sigma^2_{\mathbf{k}_1}+\sigma^2_{\mathbf{k}_2}]\right) = \exp\Bigg(&\tfrac{1}{2}\sum_{c\in m}\sum_{\mathbf{i}\in\mathbf{M}^c}\sigma^{cc}_{\mathbf{ii}}\left[(I^c_{\mathbf{i}}(\mathbf{k}_1))^2+(I^c_{\mathbf{i}}(\mathbf{k}_2))^2\right] \\
&+\tfrac{1}{2}\left[D(\mathbf{k}_1,\mathbf{k}_1)+D(\mathbf{k}_2,\mathbf{k}_2)\right]\Bigg).
\end{aligned}$$

Substituting these expressions into (20), and noting that as the variances of the log-linear parameters tend to infinity, the contribution of the constant finite terms $D(\mathbf{k}, \mathbf{k})$ become negligible for all $\mathbf{k} = \mathbf{k}_1, \mathbf{k}_2 \in K$, and we obtain

$$\begin{aligned}
&\operatorname{Corr}(n_{\mathbf{k}_1}, n_{\mathbf{k}_2}) \\
&\quad \to \exp\left(\sum_{c\in m}\sum_{\mathbf{i}\in\mathbf{M}_c}\sigma^{cc}_{\mathbf{ii}}\left\{I^c_{\mathbf{i}}(\mathbf{k}_1)I^c_{\mathbf{i}}(\mathbf{k}_2)-\tfrac{1}{2}\left[(I^c_{\mathbf{i}}(\mathbf{k}_1))^2+(I^c_{\mathbf{i}}(\mathbf{k}_2))^2\right]\right\}\right). \qquad (21)
\end{aligned}$$

Using trivial algebra, expanding the equation $(a+b)^2 \geq 0$ and setting $a = I^c_{\mathbf{i}}(\mathbf{k}_1)$ and $b = I^c_{\mathbf{i}}(\mathbf{k}_2)$, we obtain

$$I^c_{\mathbf{i}}(\mathbf{k}_1)I^c_{\mathbf{i}}(\mathbf{k}_2) \leq \tfrac{1}{2}\Big[(I^c_{\mathbf{i}}(\mathbf{k}_1))^2 + (I^c_{\mathbf{i}}(\mathbf{k}_2))^2\Big]. \tag{22}$$

Clearly, the equality is true if and only if $I^c_{\mathbf{i}}(\mathbf{k}_1) = I^c_{\mathbf{i}}(\mathbf{k}_2)$ $(= \pm 1, 0)$. So that, for $\mathbf{k}_1 \neq \mathbf{k}_2$ and $m \neq \varnothing$,

$$\operatorname{Corr}(n_{\mathbf{k}_1}, n_{\mathbf{k}_2}) \to \exp\left(\sum_{c\in m}\sum_{\substack{\mathbf{i}\in\mathbf{M}^c:\\ I^c_{\mathbf{i}}(\mathbf{k}_1)\neq I^c_{\mathbf{i}}(\mathbf{k}_2)}} \sigma^{cc}_{\mathbf{ii}}\left\{I^c_{\mathbf{i}}(\mathbf{k}_1)I^c_{\mathbf{i}}(\mathbf{k}_2) - \tfrac{1}{2}\Big[(I^c_{\mathbf{i}}(\mathbf{k}_1))^2 + (I^c_{\mathbf{i}}(\mathbf{k}_2))^2\Big]\right\}\right).$$

However, by (22),

$$I^c_{\mathbf{i}}(\mathbf{k}_1)I^c_{\mathbf{i}}(\mathbf{k}_2) - \tfrac{1}{2}\Big[\big(I^c_{\mathbf{i}}(\mathbf{k}_1)\big)^2 + \big(I^c_{\mathbf{i}}(\mathbf{k}_2)\big)^2\Big] < 0,$$

for all $c \in m$ and $\mathbf{i} \in \mathbf{M}^c$ such that $I^c_{\mathbf{i}}(\mathbf{k}_1) \neq I^c_{\mathbf{i}}(\mathbf{k}_2)$. So that as $\sigma^{cc}_{\mathbf{ii}} \to \infty$, if $\mathbf{k}_1 \neq \mathbf{k}_2$, then

$$\operatorname{Corr}(n_{\mathbf{k}_1}, n_{\mathbf{k}_2}) \to \exp(-\infty) = 0.$$

Similarly, when $\mathbf{k}_1 = \mathbf{k}_2$, we get equality in (22) and $\operatorname{Corr}(n_{\mathbf{k}_1}, n_{\mathbf{k}_2}) \to \exp(0) = 1$.

Hence, the correlation matrix tends to the identity as the variances on the log-linear parameters tend to infinity. □

A simple corollary to this result, based also upon Corollary 4.2, is that the distribution induced on the total cell count is also diffuse so that asymptotically, we obtain similarly vague priors on all interpretable parameters. This is an extremely useful result that ensures consistency across parameterizations in the absence of strong prior information.

Of course, we may also place more informative priors on the log-linear parameters and we discuss the application of our results in the context of a particular example.

5.2. *Placing informative priors on the log-linear parameters.* In practice, informative prior information often consists of knowledge about the magnitude or at least the direction of different effects within the model, together with some idea of the total cell count. Since priors may only be placed on either the log-linear parameters (expressing knowledge about interactions and effects) or the cell counts (reflecting knowledge about the total cell count) it may be difficult to find a prior which properly reflects all of the prior information available. However, using the results from Section 4 we can place a sensible prior on the log-linear parameters and check that this is consistent with the prior information available on the cell counts. In order to illustrate how the results from the previous section may be usefully applied in this context, we will consider a practical example.

We consider the data presented by Hook, Albright and Cross (1980) [see also Hook and Regal (1995)] concerning the number of black babies born with spina bifida in upstate New York between 1969 and 1974. There are three distinct sources of data, namely birth certificates (BC), death certificates (DC) and medical records (MR), respectively. The data are summarized by an incomplete 2^3 contingency table which records the number of individuals identified by any combination of sources. The total population size is unknown and so there is no data on the cell corresponding to the number of individuals not observed by any source. This missing cell, and thus the total population size, is of primary interest in this analysis.

A previous study of the white population in New York during the same years provides us with a good deal of prior information, as the relationship between the two populations is well understood. Consultation with experts in the area suggests that a priori we might expect a negative interaction between the death certificates and medical records. However, a positive interaction (though unlikely) could not be discounted. In particular, it was believed that with probability 0.9, the effect of this interaction would decrease the corresponding cell counts by a proportion in the interval $[0.1, 0.9]$. There was little information on the magnitude and sign of the remaining main effects and interactions. In addition to the information on the log-linear parameters, it was expected that the number of black babies born with spina bifida during the study period would lie somewhere in the interval $[9, 56]$ with probability 0.95 and a prior with mean somewhere in the interval $[29, 35]$ seemed sensible.

Since we have prior knowledge on only a single interaction term, we consider the corresponding prior on the cell counts, given the model that contains the constant term, the main effects terms and the corresponding interaction term, that is, model $(\varnothing, BC, DC, MR, DC–MR)$. For illustrative purposes, we shall restrict our attention to this single model. We consider independent normal priors for the log-linear parameters, as these are consistent with the information available a priori and we have no prior information to the contrary. We construct a prior for the interaction term from the information above and consider a normal prior with mean -1.2 and variance $\frac{4}{9}$, which provides a 90% credible interval for the proportional effect on the expected cell counts of $[0.10, 0.91]$. We have no other prior information concerning the remaining log-linear parameters. Thus, we place independent and identical normal priors on these parameters with mean zero and a variance of 0.5. This corresponds to the assumption that with probability 0.95, each of these effects will increase or decrease the corresponding cell counts by a factor of no more than four. These priors appeared to be a fair summary of the information available on the log-linear parameters.

In order to check that these priors are also consistent with the information available on the cell counts, we can use the results of Section 4 to summarize the priors induced on these parameters. However, from Corollary 4.2 alone we can see that the prior specified above produces an expected total population of 59, with a variance of 7183 (or standard deviation of 85). This is well outside the 95% interval given for the population size above, with approximately twice

the expected number in the population and too large a standard deviation. However, the general correlation structure of the cell counts did not seem unreasonable under this prior:

$$\mathrm{Corr}(\mathbf{n}) = \begin{pmatrix} 1.000 & 0.308 & 0.071 & -0.034 & 0.071 & -0.034 & 0.053 & -0.041 \\ & 1.000 & -0.034 & 0.071 & -0.034 & 0.071 & -0.041 & 0.053 \\ & & 1.000 & 0.308 & 0.053 & -0.041 & 0.071 & -0.034 \\ & & & 1.000 & -0.041 & -0.053 & -0.034 & 0.071 \\ & & & & 1.000 & 0.308 & 0.071 & -0.034 \\ & & & & & 1.000 & -0.034 & 0.071 \\ & & & & & & 1.000 & 0.308 \\ & & & & & & & 1.000 \end{pmatrix}.$$

In this case, what appear to be reasonable priors on the log-linear parameters produces a corresponding prior on the cell counts which is at odds with prior knowledge. Taking this into consideration, we can alter our priors, in order to obtain a distribution that better reflects our beliefs. We can only reasonably consider altering the priors on the parameters for which we do not have strong prior information, that is, the constant and main-effect terms. We have no preference as to the negative or positive effect that these terms have, so we need to maintain a prior mean of zero. However, we may reconsider the variance associated with each of these terms, in order to obtain a compromise prior consistent with the information available both upon the log-linear parameters and the total population size. By placing a prior variance of 0.25 on each of these terms, we assume that with probability 0.95 each of these effects will increase or decrease the corresponding cell counts by a factor of no more than two, which still appears consistent with prior beliefs about these parameters, though slightly less vague than the previous prior. Further consultation with experts confirmed this to be reasonable.

With this new prior, we use Theorem 4.1 and the associated results from Section 4 to calculate the mean and correlation of the expected cell counts. We obtain

$$\mathbb{E}(\mathbf{n}) = \begin{pmatrix} 0.62 \\ 0.62 \\ 6.84 \\ 6.84 \\ 6.84 \\ 6.84 \\ 0.62 \\ 0.62 \end{pmatrix}$$

and

$$\mathrm{Corr}(\mathbf{n}) = \begin{pmatrix} 1.000 & 0.485 & 0.018 & -0.111 & 0.118 & -0.111 & 0.173 & -0.017 \\ & 1.000 & -0.111 & 0.018 & -0.111 & 0.188 & -0.017 & 0.173 \\ & & 1.000 & 0.485 & 0.173 & -0.017 & 0.188 & -0.111 \\ & & & 1.000 & -0.017 & 0.173 & -0.111 & 0.188 \\ & & & & 1.000 & 0.485 & 0.018 & -0.111 \\ & & & & & 1.000 & -0.111 & 0.018 \\ & & & & & & 1.000 & 0.485 \\ & & & & & & & 1.000 \end{pmatrix}.$$

The prior mean structure for the cell counts here is unsurprising since this simply reflects the influence of the negative interaction between sources DC and MR, so that more individuals are expected to be observed on combinations of sources containing only one of sources DC and MR. The symmetry between the different combinations of sources results from the same priors being placed on all the main-effect terms in the model.

The patterns within the correlation matrix are similarly understandable. Here, the interaction term is once again partitioning the cells into groups according to either the level of BC source or the level of the DC and MR sources combined. The fact that each value appears several times is a result of the fact that we took identical priors for all of the log-linear parameters except the interaction term. We also note that the correlations between cells for the second prior are larger than those for the first prior considered. This is an illustration of the effect of decreasing the prior variance and is consistent with Lemma 5.1 which describes the effect of taking increasingly diffuse priors on parameters about which we have little knowledge a priori.

From these priors for the expected cell counts, we can derive the mean and variance for the total population and we find that $\mathbb{E}(N) = 29.8$ and $SD(N) = 31.31$. This prior is now consistent with the prior information available on this parameter with a mean well within the range specified but with a slightly larger standard deviation. If we are willing to further compromise on the variances associated with some of the log-linear parameters we may iterate this procedure once again. However, in this example this latest prior is considered adequate to describe our expert opinion and the analysis of the data can proceed with these values.

6. Discussion. In this paper we have focused upon the estimation of cell counts rather than cell probabilities. If the cell probabilities are themselves of interest, it is possible to derive the form of the induced prior directly from Theorem 4.1 using a change of variables argument.

If we let

$$p_i = \frac{n_i}{\sum_{i=1}^{|K|} n_i} \quad \text{for } i = 1, \ldots, |K| - 1,$$

$$N = \sum_{i=1}^{|K|} n_i$$

and, for notational convenience, set $p_{|K|} = 1 - \sum_{i=1}^{|K|-1} p_i$ then the determinant of the Jacobian for this transformation can be shown to be $N^{|K|-1}$. The corresponding joint density induced on $(p_1, \ldots, p_{|K|-1}, N)$ is given by

$$f_{\mathbf{p}}(p_1, \ldots, p_{|K|-1}, N) = \frac{1}{|\Sigma|^{1/2}(2\pi)^{|K|/2}} \prod_{i=1}^{|K|} \frac{1}{p_i} N^{-(h(N)/2+1)} \times \exp\left(-\frac{1}{2}(\ln \mathbf{p} - \boldsymbol{\mu})^T \Sigma^{-1} (\ln \mathbf{p} - \boldsymbol{\mu})\right),$$

where

$$h(N) = \ln N \mathbf{1}^T \Sigma^{-1} \mathbf{1} + 2(\ln \mathbf{p} - \boldsymbol{\mu})^T \Sigma^{-1} \mathbf{1},$$

and $\mathbf{p}$ denotes the complete vector of cell probabilities, $(p_1, \ldots, p_{|K|})$. The marginal distribution of the cell probabilities can then be easily shown to be

$$f_{\mathbf{p}}(p_1, \ldots, p_{|K|-1}) = \frac{\sqrt{2\pi}}{\sqrt{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}} \exp\left(\frac{((\ln \mathbf{p} - \boldsymbol{\mu})^T \Sigma^{-1} \mathbf{1})^2}{2\mathbf{1}^T \Sigma^{-1} \mathbf{1}}\right) \times \frac{1}{|\Sigma|^{1/2}(2\pi)^{|K|/2}} \prod_{i=1}^{|K|} \frac{1}{p_i} \times \exp\left(-\frac{1}{2}(\ln \mathbf{p} - \boldsymbol{\mu})^T \Sigma^{-1} (\ln \mathbf{p} - \boldsymbol{\mu})\right). \tag{23}$$

Thus it is possible, using the results in this paper, to obtain the form of the joint distribution of the cell probabilities induced by a multivariate normal prior placed on the log-linear parameters. Unfortunately, as with previous approaches [Knuiman and Speed (1988), Dellaportas and Forster (1999)], the moments for this distribution are not analytically tractable. However, since we have already derived an explicit expression for both the mean and covariance matrix of the expected cell counts these will generally be sufficient for the purpose of prior elicitation. Thus, this problem is largely overcome through the results presented in this paper.

Another advantage of our approach is that the density given in (23) corresponds to a *general* multivariate normal prior placed on the log-linear parameters. Previous approaches have been limited to special forms of multivariate normality in order to impose the necessary "sum to one" constraints on the cell probabilities. Thus, our approach permits a more general family of prior distributions to be used.

We would argue that knowledge of the distribution of the cell counts is sufficient for the purposes of prior elicitation and this is certainly borne out by our experience in this area. This paper provides the results required to actually undertake that process. We derive the form of the prior distribution induced on the cell counts by placing a general multivariate normal distribution on the log-linear parameters. We further derive the means and covariances of that distribution, together with the mean and variance of the prior induced on the total cell count. In the case of some degree of prior knowledge this latter result

will often be extremely valuable, as demonstrated in Section 5. We have also shown that diffuse multivariate normal priors placed on the log-linear parameters are consistently diffuse across parameterizations in that they are similarly vague in terms of the induced distributions on the individual cell counts and on the total cell count. We have discussed how this property is vital in any realistic application but cannot automatically be assumed. Finally, we provide an illustration of how our results may be used to aid the process of prior specification. It is clear that naive placement of priors on the log-linear parameters can induce implausible priors on the cell counts. In practice this may have a substantial effect on the results of the analysis. Of course such an influence should be detected in a routine sensitivity analysis, but without these results it is extremely difficult to find a prior which truly reflects our prior beliefs across both parameterizations.

APPENDIX

PROOF OF LEMMA 2.1. From (3) it is clear that $I^c_{\mathbf{1}}(\mathbf{1}) = 1$ since in this case, $k(\gamma) = j(\gamma) = 1$ for all γ, and $|K_\gamma| > 1$. Thus, the identifiability constraint is satisfied by the definition of $I^c_{\mathbf{j}}(\mathbf{k})$ in (3). Now all that remains is the proof that the definition of $I^c_{\mathbf{j}}(\mathbf{k})$ in (3) satisfies the sum to zero condition in (2).

For $c \in \mathscr{P}(S)\backslash\varnothing$ and $l \in \{1, \ldots, |S|\}$ we have, for some $\mathbf{e} \in K$,

$$\sum_{\mathbf{k}\in L^{\mathbf{e}}_l} I^c_{\mathbf{j}}(\mathbf{k}) = \sum_{\mathbf{k}\in L^{\mathbf{e}}_l} \prod_{\gamma:\, S_\gamma \in c} [\mathscr{I}(k(\gamma) = j(\gamma)) - \mathscr{I}(k(\gamma) = |K_\gamma|)]$$

from the definition in (3).

$$= \sum_{\mathbf{k}\in L^{\mathbf{e}}_l} \left\{ \prod_{\substack{\gamma:\, S_\gamma \in c \\ \gamma \neq l}} [\mathscr{I}(e(\gamma) = j(\gamma)) - \mathscr{I}(e(\gamma) = |K_\gamma|)] \times [\mathscr{I}(k(l) = j(l)) - \mathscr{I}(k(l) = |K_l|)] \right\}$$

since $k(\gamma) = e(\gamma)\ \forall\ \gamma \neq l$

$$= \prod_{\substack{\gamma:\, S_\gamma \in c \\ \gamma \neq l}} [\mathscr{I}(e(\gamma) = j(\gamma)) - \mathscr{I}(e(\gamma) = |K_\gamma|)] \times \sum_{\mathbf{k}\in L^{\mathbf{e}}_l} [\mathscr{I}(k(l) = j(l)) - \mathscr{I}(k(l) = |K_l|)] \quad \text{by definition of } L^{\mathbf{e}}_l$$

$$= \prod_{\substack{\gamma:\, S_\gamma \in c \\ \gamma \neq l}} [\mathscr{I}(e(\gamma) = j(\gamma)) - \mathscr{I}(e(\gamma) = |K_\gamma|)] \times \left(\sum_{\mathbf{k}\in L^{\mathbf{e}}_l} \mathscr{I}(k(l) = j(l)) - \sum_{\mathbf{k}\in L^{\mathbf{e}}_l} \mathscr{I}(k(l) = |K_l|) \right)$$

$$= \prod_{\substack{\gamma: S_\gamma \in c \\ \gamma \neq l}} [\mathscr{I}(e(\gamma) = j(\gamma)) - \mathscr{I}(e(\gamma) = |K_\gamma|)]$$

$$\times \left(\sum_{k(l)=1}^{|K_l|} \mathscr{I}(k(l) = j(l)) - \sum_{k(l)=1}^{|K_l|} \mathscr{I}(k(l) = |K_l|) \right)$$

by definition of $L_l^{\mathbf{e}}$

$$= \prod_{\substack{\gamma: S_\gamma \in c \\ \gamma \neq l}} [\mathscr{I}(e(\gamma) = j(\gamma)) - \mathscr{I}(e(\gamma) = |K_\gamma|)] \times (1 - 1)$$

since $j(l) \in \{1, \ldots, |K_l| - 1\}$

$$= 0.$$

Hence both the sum to zero constraint and the identifiability constraint are satisfied by $I_{\mathbf{j}}^c(\mathbf{k})$ as defined in (3). □

PROOF OF LEMMA 3.1. Using the method of induction, we begin with the case when $S = \{S_1\}$ that is, there is only one source with levels $K_1 = \{1, \ldots, |K_1|\}$. Then $\mathscr{P}(S) = \{\varnothing, \{S_1\}\}$, and we have only the matrices $X^{\varnothing}$ and X^{S_1} to consider. First, we consider the design matrix corresponding to the constant term $X^{\varnothing}$ which is of dimension $(|K_1| \times 1)$. We let $x^{\varnothing}(k)$ be the element of $X^{\varnothing}$ corresponding to cell k, for $k \in K_1$, dropping the subscript $\mathbf{j}$ since there are no levels for this term. Then, comparing equations (1) and (4),

$$x^{\varnothing}(k) = I^{\varnothing}(k) \qquad \text{by (10)}$$
$$= 1 \quad \text{by definition of } I^{\varnothing}(k).$$

We therefore have that

$$X^{\varnothing} = \mathbf{1}_{|K_1|} = \bigotimes_{\gamma=1}^{|S|} \mathscr{I}(S_\gamma \notin c)\mathbf{1}_{|K_\gamma|}.$$

Hence, this matrix has the desired form.

Similarly, X^{S_1} is of dimension $(|K_1| \times |K_1| - 1)$, and we denote by $x_j^{S_1}(k)$ the element of the matrix corresponding to cell $k \in K_1$ and source S_1 at level $j \in M^{S_1}$. Then,

$$x_j^{S_1}(k) = I_j^{S_1}(k) \quad \text{by (10)}$$
$$= \mathscr{I}(k = j) - \mathscr{I}(k = |K_1|) \quad \text{by (3)}.$$

We consider three different cases (since $j \neq |K_1|$ by definition of X^{S_1}). For $k = j$,

$$x_j^{S_1}(k) = 1.$$

Similarly, for $k \neq j$ and $k \neq |K_1|$,

$$x_j^{S_1}(k) = 0.$$

Finally, when $k = |K_1|$,

$$x_j^{S_1}(k) = -1 \qquad \forall\ j \in M^{S_1}.$$

These first two conditions describe the first $|K_1| - 1$ rows of X^{S_1} with the only nonzero entries lying along the diagonal. The final condition describes the final row which consists only of the element -1 across all columns (j-values). Thus,

$$X^{S_1} = \begin{pmatrix} I_{|K_1|-1} \\ -\mathbf{1}^T_{|K_1|-1} \end{pmatrix} = L_{|K_1|}.$$

This matrix is clearly of the desired form and so the lemma holds for the case where we have just a single source. To continue our inductive proof, we now consider the case where we have a general number of sources and examine what happens when we add an additional source.

Suppose that the lemma holds when there are $|S|$ sources and consider adding a new source, $\{S_\eta\}$ with $|K_\eta|$ levels to obtain a new set of sources $S' = S \cup \{S_\eta\}$. Note that we shall also express S_η as $S_{|S|+1}$, with $|K_{|S|+1}|$ levels, for notational convenience, at times. The set of cells is then denoted by $K' = K \times \{k(\eta)\colon \eta = 1, \ldots, |K_\eta|\}$ and we define $\mathbf{k}' = \{\mathbf{k}, k'(\eta)\}$. Similarly, for $\mathbf{j}' \in \mathbf{M}_c$, we define $\mathbf{j}' = \{\mathbf{j}, j'(\eta)\}$. We let X'^c and X^c be the design matrices of parameter c for the set of sources S' and S, respectively. The design matrix X'^c is then of dimension

$$\left(\prod_{\gamma=1}^{|S|+1} |K_\gamma| \times \prod_{\gamma\colon S_\gamma \in c} (|K_\gamma| - 1) \right).$$

Note that the second dimension of X'^c is $|K_\eta| - 1$ times larger than that for X^c if $S_\eta \in c$ and the same size if $S_\eta \notin c$. Similarly the first dimension of X'^c is always $|K_\eta|$ times larger than that of X^c.

We denote the element of X'^c corresponding to cell $\mathbf{k}' \in K'$ and at source levels $\mathbf{j}' \in \mathbf{M}^c$ by $x'^c_{\mathbf{j}'}(\mathbf{k}')$. Then we have that

$$\begin{aligned} x'^c_{\mathbf{j}'}(\mathbf{k}') &= I^c_{\mathbf{j}'}(\mathbf{k}') \quad \text{by (10) which holds for any set of sources} \\ &= \prod_{\gamma\colon S_\gamma \in c} [\mathcal{I}(k'(\gamma) = j'(\gamma)) - \mathcal{I}(k'(\gamma) = |K_\gamma|)] \quad \text{by Lemma 2.1,} \end{aligned}$$

where $k'(\gamma)$ and $j'(\gamma)$ are the elements of $\mathbf{k}'$ and $\mathbf{j}'$, respectively, corresponding to source S_γ.

We now consider the two separate cases, depending upon whether or not $S_\eta \in c$.

Case I. $S_\eta \notin c$. If $\mathbf{j}' \in \mathbf{M}^c$ and $k'(\eta) \in K_\eta$, then

$$\begin{aligned} x'^c_{\mathbf{j}'}(\mathbf{k}') &= x'^c_{\mathbf{j}'}([\mathbf{k}, k'(\eta)]) = I^c_{\mathbf{j}'}([\mathbf{k}, k'(\eta)]) \quad \text{by (10)} \\ &= \prod_{\gamma:\, S_\gamma \in c} [\mathscr{I}(k'(\gamma) = j'(\gamma)) - \mathscr{I}(k'(\gamma) = |K_\gamma|)] \\ &= \prod_{\gamma:\, S_\gamma \in c} [\mathscr{I}(k(\gamma) = j(\gamma)) - \mathscr{I}(k(\gamma) = |K_\gamma|)] \\ &\qquad \text{since } S_\eta \notin c \text{ and } k'(\gamma) = k(\gamma) \text{ and } j'(\gamma) = j(\gamma), \quad \forall\, \gamma \neq \eta \\ &= I^c_{\mathbf{j}}(\mathbf{k}) = x^c_{\mathbf{j}}(\mathbf{k}), \quad \text{by (10).} \end{aligned}$$

Thus,

$$x'^c_{\mathbf{j}}([\mathbf{k}, k'(\eta)]) = x^c_{\mathbf{j}}(\mathbf{k}) \quad \forall\, \mathbf{j} \in \mathbf{M}^c \text{ and } k'(\eta) = 1, \ldots, |K_\eta|,$$

which implies that

$$\begin{aligned} X'^c = \begin{pmatrix} X^c \\ \vdots \\ X^c \end{pmatrix} &= \mathbf{1}_{|K_\eta|} \otimes X^c \\ &= \mathbf{1}_{|K_\eta|} \otimes \bigotimes_{\gamma=1}^{|S|} \left[\mathscr{I}(S_\gamma \in c) L_{|K_\gamma|} + \mathscr{I}(S_\gamma \notin c) \mathbf{1}_{|K_\gamma|}\right] \\ &\qquad \text{since the result holds for } |S| \text{ sources} \\ &= \left(\mathscr{I}(S_\eta \in c) L_{|K_\eta|} + \mathscr{I}(S_\eta \notin c) \mathbf{1}_{|K_\eta|}\right) \\ &\qquad \otimes \bigotimes_{\gamma=1}^{|S|} \left[\mathscr{I}(S_\gamma \in c) L_{|K_\gamma|} + \mathscr{I}(S_\gamma \notin c) \mathbf{1}_{|K_\gamma|}\right] \\ &= \bigotimes_{\gamma=1}^{|S|+1} \left[\mathscr{I}(S_\gamma \in c) L_{|K_\gamma|} + \mathscr{I}(S_\gamma \notin c) \mathbf{1}_{|K_\gamma|}\right], \end{aligned}$$

and therefore X'^c is also of the correct form if $S_\eta \notin c$.

Case II. $S_\eta \in c$. For $\mathbf{j}' = (\mathbf{j}, j'(\eta))$, then $k'(\eta) = 1, \ldots, |K_\eta| - 1$,

$$\begin{aligned} x'^c_{\mathbf{j}'}(\mathbf{k}') &= I^c_{\mathbf{j}'}(\mathbf{k}') \\ &= \prod_{\gamma:\, S_\gamma \in c} [\mathscr{I}(k'(\gamma) = j'(\gamma)) - \mathscr{I}(k'(\gamma) = |K_\gamma|)] \quad \text{by Lemma 2.1} \\ &= \mathscr{I}(k'(\eta) = j'(\eta)) \prod_{\gamma:\, S_\gamma \in c \setminus S_\eta} [\mathscr{I}(k'(\gamma) = j'(\gamma)) - \mathscr{I}(k(\gamma) = |K_\gamma|)] \\ &= \mathscr{I}(k'(\eta) = j'(\eta)) I^{c \setminus S_\eta}_{\mathbf{j}}(\mathbf{k}) \\ &= \mathscr{I}(k'(\eta) = j'(\eta)) x^{c \setminus S_\eta}_{\mathbf{j}}(\mathbf{k}). \end{aligned} \tag{24}$$

Also, for $k'(\eta) = |K_\eta|$

$$\begin{aligned} x'^c_{\mathbf{j}'}(\mathbf{k}') &= I^c_{\mathbf{j}'}(\mathbf{k}') \\ &= -\prod_{\gamma:\, S_\gamma \in c\setminus S_\eta} [\mathscr{I}(k(\gamma) = j(\gamma)) - \mathscr{I}(k(\gamma) = |K_\gamma|)] \\ &\quad \text{since } \mathscr{I}(k(\eta) = |K_\eta|) = 1 \text{ and } j(\eta) \neq |K_\eta| \text{ by definition} \\ &= -I^{c\setminus S_\eta}_{\mathbf{j}}(\mathbf{k}) \\ &= -x^{c\setminus S_\eta}_{\mathbf{j}}(\mathbf{k}). \qquad (25) \end{aligned}$$

Combining the results for these two cases, we obtain:

$$x'^c_{\mathbf{j}'}(\mathbf{k}') = \begin{cases} x^{c\setminus S_\eta}_{\mathbf{j}}(\mathbf{k}), & \text{if } k'(\eta) = j'(\eta) \text{ by (24)}, \\ -x^{c\setminus S_\eta}_{\mathbf{j}}(\mathbf{k}), & \text{if } k'(\eta) = |K_\eta| \text{ by (25)}, \\ 0, & \text{if } k'(\eta) \neq |K_\eta| \text{ and } k'(\eta) \neq j'(\eta) \text{ by (24)}. \end{cases} \qquad (26)$$

The design matrix corresponding to source c is therefore given by

$$\begin{aligned} X'^c &= \begin{pmatrix} x'^c_{(\mathbf{j},1)}([\mathbf{k},1]) & \cdots & x'^c_{(\mathbf{j},|K_\eta|)}([\mathbf{k},1]) \\ x'^c_{(\mathbf{j},1)}([\mathbf{k},2]) & \cdots & x'^c_{(\mathbf{j},|K_\eta|)}([\mathbf{k},2]) \\ \vdots & & \vdots \\ x'^c_{(\mathbf{j},1)}([\mathbf{k},|K_\eta|]) & \cdots & x'^c_{(\mathbf{j},|K_\eta|)}([\mathbf{k},|K_\eta|]) \end{pmatrix} \\ &= \begin{pmatrix} X^{c\setminus S_\eta} & 0 & \cdots & 0 \\ 0 & X^{c\setminus S_\eta} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & X^{c\setminus S_\eta} \\ -X^{c\setminus S_\eta} & -X^{c\setminus S_\eta} & \cdots & -X^{c\setminus S_\eta} \end{pmatrix} \quad \text{by (26)} \\ &= \begin{pmatrix} \operatorname{diag}(X^{c\setminus S_\eta} \cdots X^{c\setminus S_\eta}) \\ -X^{c\setminus S_\eta} \cdots - X^{c\setminus S_\eta} \end{pmatrix} = \begin{pmatrix} I_{|K_\eta|-1} \\ -\mathbf{1}^T_{|K_\eta|-1} \end{pmatrix} \otimes X^{c\setminus S_\eta} \\ &= \begin{pmatrix} I_{|K_\eta|-1} \\ -\mathbf{1}^T_{|K_\eta|-1} \end{pmatrix} \otimes \bigotimes_{\gamma=1}^{|S|} \big[\mathscr{I}(S_\gamma \in c) L_{|K_\gamma|} + \mathscr{I}(S_\gamma \notin c)\mathbf{1}_{|K_\gamma|}\big] \\ &\quad \text{since the result holds for } |S| \text{ sources} \\ &= \big(\mathscr{I}(S_\eta \in c) L_{|K_\eta|} + \mathscr{I}(S_\eta \notin c)\mathbf{1}_{|K_\eta|}\big) \otimes \bigotimes_{\gamma=1}^{|S|} \big[\mathscr{I}(S_\gamma \in c) L_{|K_\gamma|} + \mathscr{I}(S_\gamma \notin c)\mathbf{1}_{|K_\gamma|}\big] \\ &\quad \text{by definition of } L_{|K_\eta|} \\ &= \bigotimes_{\gamma=1}^{|S|+1} \big[\mathscr{I}(S_\gamma \in c) L_{|K_\gamma|} + \mathscr{I}(S_\gamma \notin c)\mathbf{1}_{|K_\gamma|}\big]. \end{aligned}$$

Thus we have shown that if the lemma holds for a set of sources S, then it also holds if we add an extra source, S_η. Since we have already demonstrated that the result holds when we have just a single source it therefore holds for any number of sources greater than one, by induction. □

In order to prove Lemma 3.2 we first introduce some notation and state and prove two further lemmas. The first lemma establishes a result which provides a usable form for the elements of the matrix formed as the product of $W_{\mathbf{ij}}^{cd}$ with $V_{\mathbf{i}}^{c}$. We then derive the form for the elements of the matrix formed as the product of $\mathbf{x}_{\mathbf{j}}^{d}$ with $(\mathbf{x}_{\mathbf{i}}^{c})^T$. Finally, we use these results to show that these two product matrices are in fact equal, which is the result given in Lemma 3.2. We begin by establishing a notation for the elements of the product matrices.

For $\mathbf{k}_1, \mathbf{k}_2 \in K$, we denote the $(\mathbf{k}_1, \mathbf{k}_2)$th element of $W_{\mathbf{ij}}^{cd}$ by $w_{\mathbf{ij}}^{cd}(\mathbf{k}_1, \mathbf{k}_2)$, for $c, d \in \mathscr{P}(S)$ and $\mathbf{i} \in \mathbf{M}^c$, $\mathbf{j} \in \mathbf{M}^d$. Similarly, we denote by $v_{\mathbf{ij}}^{cd}(\mathbf{k}_1, \mathbf{k}_2)$ the $(\mathbf{k}_1, \mathbf{k}_2)$ element of the matrix $V_{\mathbf{i}}^{c}$. Finally, we let $y_{\mathbf{ij}}^{cd}(\mathbf{k}_1, \mathbf{k}_2)$ be the $(\mathbf{k}_1, \mathbf{k}_2)$ element of the product $W_{\mathbf{ij}}^{cd} V_{\mathbf{i}}^{c}$. For notational convenience, when $c = \varnothing$, or $d = \varnothing$, we will remove the subscript referring to the respective levels of that parameter. We will also use the convention that if $S_\eta \notin c$, then $c \backslash S_\eta = c$, and similarly for $S_\eta \notin d$.

The first of our two new results establishes the form of the element $y_{\mathbf{ij}}^{cd}(\mathbf{k}_1, \mathbf{k}_2)$. We let S be the set of sources $\{S_1 \ldots, S_{|S|}\}$ and let $S' = S \cup \{S_\eta\}$. We also define $S_\eta = S_{|S|+1}$ for notational convenience at times. For $c \in \mathscr{P}(S')$ and $d \in \mathscr{P}(S')$ we let $\mathbf{i}' \in \mathbf{M}^c$ and $\mathbf{j}' \in \mathbf{M}^d$. We use the convention that all matrices and vectors corresponding to the set of sources S' are indexed by a "prime" that is, $'$, while all other terms correspond to S.

LEMMA A.1. *For $\mathbf{k}'_1, \mathbf{k}'_2 \in K'$, so that $\mathbf{k}'_1 = (\mathbf{k}_1, k'_1(\eta))$ and $\mathbf{k}'_2 = (\mathbf{k}_2, k'_2(\eta))$,*

$$y_{\mathbf{i'j'}}^{'cd}(\mathbf{k}'_1, \mathbf{k}'_2) = \begin{cases} y_{\mathbf{ij}}^{(c\backslash S_\eta)(d\backslash S_\eta)}(\mathbf{k}_1, \mathbf{k}_2) \\ \quad \times \sum_{k'(\eta)\in K_\eta} \left[a_{\mathbf{i'j'}}^{cd}(k'_1(\eta), k'(\eta)) g_{\mathbf{i}'}^{\eta}(k'(\eta), k'_2(\eta))\right], & \text{if } S_\eta \in c, \\ y_{\mathbf{ij}}^{c(d\backslash S_\eta)}(\mathbf{k}_1, \mathbf{k}_2) \\ \quad \times \sum_{k'(\eta)\in K_\eta} a_{\mathbf{i'j'}}^{cd}(k'_1(\eta), k'(\eta)), & \text{if } S_\eta \notin c, S_\eta \in d, \\ y_{\mathbf{ij}}^{cd}(\mathbf{k}_1, \mathbf{k}_2), & \text{if } S_\eta \notin c, S_\eta \notin d, \end{cases}$$

where $\mathbf{i}' = \mathbf{i}$ if $S_\eta \notin c$ and $\mathbf{i}'$ is of the form $(\mathbf{i}, i'(\eta))$ when $S_\eta \in c$, with the analogous definition for $\mathbf{j}'$. The terms $a_{\mathbf{i'j'}}^{cd}(k'_1(\eta), k'(\eta))$ refer to the $(k'_1(\eta), k'(\eta))$ element of the matrix $A_{\mathbf{i'j'}}^{cd}(\eta)$ and similarly for g.

PROOF. We initially consider the matrices $W_{\mathbf{i'j'}}^{'cd}$ and $V_{\mathbf{i}'}^{'c}$, and their corresponding elements. We then combine the results to calculate the individual

elements of the product of the matrices. Let us first consider the matrix $W'^{cd}_{\mathbf{i}'\mathbf{j}'}$. From (7), we have that,

$$\begin{aligned} W'^{cd}_{\mathbf{i}'\mathbf{j}'} &= \bigotimes_{\gamma=1}^{|S'|}\left[\mathscr{I}(S_\gamma \in c\cup d)A^{cd}_{\mathbf{i}'\mathbf{j}'}(\gamma) + (1-\mathscr{I}(S_\gamma \in c\cup d))I_{|K_\gamma|}\right] \\ &= \left[\mathscr{I}(S_\eta \in c\cup d)A^{cd}_{\mathbf{i}'\mathbf{j}'}(\eta) + (1-\mathscr{I}(S_\eta \in c\cup d))I_{|K_\eta|}\right] \\ &\quad \otimes \bigotimes_{\gamma=1}^{|S|}\left[\mathscr{I}(S_\gamma \in c\cup d)A^{cd}_{\mathbf{i}'\mathbf{j}'}(\gamma) + (1-I(S_\gamma \in c\cup d))I_{|K_\gamma|}\right] \\ &= \left[\mathscr{I}(S_\eta \in c\cup d)A^{cd}_{\mathbf{i}'\mathbf{j}'}(\eta) + (1-\mathscr{I}(S_\eta \in c\cup d))I_{|K_\eta|}\right] \otimes W^{(c\backslash S_\eta)(d\backslash S_\eta)}_{\mathbf{ij}}, \end{aligned} \tag{27}$$

since $\mathbf{i}' = (\mathbf{i}, i'(\eta))$ and $\mathbf{j}' = (\mathbf{j}, j'(\eta))$ by definition.

We now consider the individual elements of $W'^{cd}_{\mathbf{i}'\mathbf{j}'}$ for the two different cases, depending upon whether $S_\eta \in c\cup d$.

Case I. $S_\eta \in c\cup d$. From (27) we have that,

$$\begin{aligned} W'^{cd}_{\mathbf{i}'\mathbf{j}'} &= \left(\left[\mathscr{I}(S_\eta \in c\cup d)A^{cd}_{\mathbf{i}'\mathbf{j}'}(\eta) + (1-\mathscr{I}(S_\eta \in c\cup d))I_{|K_\eta|}\right] \otimes W^{(c\backslash S_\eta)(d\backslash S_\eta)}_{\mathbf{ij}}\right) \\ &= A^{cd}_{\mathbf{i}'\mathbf{j}'}(\eta) \otimes W^{(c\backslash S_\eta)(d\backslash S_\eta)}_{\mathbf{ij}} \quad \text{since } S_\eta \in c\cup d \\ &= \begin{pmatrix} a^{cd}_{\mathbf{i}'\mathbf{j}'}(1,1)W^{(c\backslash S_\eta)(d\backslash S_\eta)}_{\mathbf{ij}} & \cdots & a^{cd}_{\mathbf{i}'\mathbf{j}'}(1,|K_\eta|)W^{(c\backslash S_\eta)(d\backslash S_\eta)}_{\mathbf{ij}} \\ \vdots & & \vdots \\ a^{cd}_{\mathbf{i}'\mathbf{j}'}(|K_\eta|,1)W^{(c\backslash S_\eta)(d\backslash S_\eta)}_{\mathbf{ij}} & \cdots & a^{cd}_{\mathbf{i}'\mathbf{j}'}(|K_\eta|,|K_\eta|)W^{(c\backslash S_\eta)(d\backslash S_\eta)}_{\mathbf{ij}} \end{pmatrix}. \end{aligned}$$

So that the $W'^{cd}_{\mathbf{i}'\mathbf{j}'}$ can be expressed as a matrix of submatrices $a^{cd}_{\mathbf{i}'\mathbf{j}'}(m,n) \times W^{(c\backslash S_\eta)(d\backslash S_\eta)}_{\mathbf{ij}}$ for $m = 1,\ldots,|K_\eta|$ and $n = 1,\ldots,|K_\eta|$. For submatrix (m,n), the $(\mathbf{k}_1,\mathbf{k}_2)$ element, for $\mathbf{k}_1,\mathbf{k}_2 \in K$, is just the corresponding element in $W^{(c\backslash S_\eta)(d\backslash S_\eta)}_{\mathbf{ij}}$ multiplied by the given scalar, $a^{cd}_{\mathbf{i}'\mathbf{j}'}(m,n)$. So that we have

$$\begin{aligned} w'^{cd}_{\mathbf{i}'\mathbf{j}'}(\mathbf{k}'_1,\mathbf{k}'_2) &= w'^{cd}_{\mathbf{i}'\mathbf{j}'}([\mathbf{k}_1, k'_1(\eta)],[\mathbf{k}_2, k'_2(\eta)]) \\ &= a^{cd}_{\mathbf{i}'\mathbf{j}'}(k'_1(\eta), k'_2(\eta))w^{(c\backslash S_\eta)(d\backslash S_\eta)}_{\mathbf{ij}}(\mathbf{k}_1,\mathbf{k}_2), \end{aligned} \tag{28}$$

with obvious notational changes if $S_\eta \notin c$ so that $c\backslash S_\eta = c$, or if $S_\eta \notin d$, so that $d\backslash S_\eta = d$.

Case II. $S_\eta \notin c\cup d$. From (7), we have that

$$W'^{cd}_{\mathbf{i}'\mathbf{j}'} = I_{|K_\eta|} \otimes W^{cd}_{\mathbf{ij}},$$

and therefore,

$$w'^{cd}_{\mathbf{i}'\mathbf{j}'}(\mathbf{k}'_1,\mathbf{k}'_2) = w^{cd}_{\mathbf{ij}}(\mathbf{k}_1,\mathbf{k}_2)\mathscr{I}(k'_1(\eta) = k'_2(\eta)). \tag{29}$$

We next consider the matrix $V'^{c}_{\mathbf{i}'}$ and from (8), we have

$$
\begin{aligned}
V'^{c}_{\mathbf{i}'} &= \bigotimes_{\gamma=1}^{|S'|}\Big[\mathscr{I}(S_\gamma \in c)G^{\gamma}_{\mathbf{i}'} + (1-\mathscr{I}(S_\gamma \in c))J_{|K_\gamma|}\Big] \quad \text{by definition} \\
&= \Big[\mathscr{I}(S_\eta \in c)G^{\eta}_{\mathbf{i}'} + (1-\mathscr{I}(S_\eta \in c))J_{|K_\eta|}\Big] \\
&\quad \otimes \bigotimes_{\gamma=1}^{|S|}\Big[\mathscr{I}(S_\gamma \in c)G^{\gamma}_{\mathbf{i}'} + (1-\mathscr{I}(S_\gamma \in c))J_{|K_\gamma|}\Big] \\
&= \big[\mathscr{I}(S_\eta \in c)G^{\eta}_{\mathbf{i}'} + (1-\mathscr{I}(S_\eta \in c))J_{|K_\eta|}\big] \otimes V_{\mathbf{i}}^{c\setminus S_\eta}. \qquad (30)
\end{aligned}
$$

As before, there are two separate cases, this time depending upon whether or not $S_\eta \in c$.

Case I. $S_\eta \in c$. From (30) we have that,

$$
\begin{aligned}
V'^{c}_{\mathbf{i}'} &= G^{\eta}_{\mathbf{i}'} \otimes V_{\mathbf{i}}^{c\setminus S_\eta} \quad \text{since } S_\eta \in c \\
&= \begin{pmatrix} g^{\eta}_{\mathbf{i}'}(1,1)V_{\mathbf{i}}^{c\setminus S_\eta} & \cdots & g^{\eta}_{\mathbf{i}'}(1,|K_\eta|)V_{\mathbf{i}}^{c\setminus S_\eta} \\ \vdots & & \vdots \\ g^{\eta}_{\mathbf{i}'}(|K_\eta|,1)V_{\mathbf{i}}^{c\setminus S_\eta} & \cdots & g^{\eta}_{\mathbf{i}'}(|K_\eta|,|K_\eta|)V_{\mathbf{i}}^{c\setminus S_\eta} \end{pmatrix}.
\end{aligned}
$$

Thus, $V'^{c}_{\mathbf{i}'}$ is a matrix of submatrices $g^{\eta}_{\mathbf{i}'}(m,n)V_{\mathbf{i}}^{c\setminus S_\eta}$ for $m = 1,\ldots,|K_\eta|$ and $n = 1,\ldots,|K_\eta|$. For $\mathbf{k}_1, \mathbf{k}_2 \in K$, the element $(\mathbf{k}_1, \mathbf{k}_2)$ of the submatrix indexed by (m,n), is the $(\mathbf{k}_1, \mathbf{k}_2)$ element of the matrix $V_{\mathbf{i}}^{c\setminus S_\eta}$ multiplied by the corresponding scalar $g^{\eta}_{\mathbf{i}'}(m,n)$. Hence, using $\mathbf{k}'_1 = (\mathbf{k}_1, k'_1(\eta))$ and $\mathbf{k}'_2 = (\mathbf{k}_2, k'_2(\eta))$, we can write,

$$
v'^{cd}_{\mathbf{i}'}(\mathbf{k}'_1, \mathbf{k}'_2) = g^{\eta}_{\mathbf{i}'}\big(k'_1(\eta), k'_2(\eta)\big)v_{\mathbf{i}}^{c\setminus S_\eta}\big(\mathbf{k}_1, \mathbf{k}_2\big). \qquad (31)
$$

Case II: $S_\eta \notin c$. Again, using (30),

$$
V'^{c}_{\mathbf{i}'} = J_{|K_\eta|} \otimes V^{c}_{\mathbf{i}},
$$

where $J_{|K_\eta|}$ is the matrix of dimension $(|K_\eta| \times |K_\eta|)$ with each element equal to unity. Trivially, the $(\mathbf{k}'_1, \mathbf{k}'_2)$ element of $V'^{c}_{\mathbf{i}'}$ is therefore given by,

$$
v'^{c}_{\mathbf{i}'}(\mathbf{k}'_1, \mathbf{k}'_2) = v^{c}_{\mathbf{i}}\big(\mathbf{k}_1, \mathbf{k}_2\big). \qquad (32)
$$

These two separate results providing the elements of the two matrices $W'^{cd}_{\mathbf{i}'\mathbf{j}'}$ and $V'^{c}_{\mathbf{i}'}$ can then be combined to find the elements of their product as follows.

Clearly, the $(\mathbf{k}'_1, \mathbf{k}'_2)$ element of the product is given by

$$
\begin{aligned}
y'^{cd}_{\mathbf{i}'\mathbf{j}'}(\mathbf{k}'_1, \mathbf{k}'_2) &= \sum_{\mathbf{k}' \in K'} w'^{cd}_{\mathbf{i}'\mathbf{j}'}(\mathbf{k}'_1, \mathbf{k}') v'^{c}_{\mathbf{i}'}(\mathbf{k}', \mathbf{k}'_2) \\
&= \sum_{\mathbf{k} \in K} \sum_{k'(\eta) \in K_\eta} w'^{cd}_{\mathbf{i}'\mathbf{j}'}(\mathbf{k}'_1, \mathbf{k}') v'^{c}_{\mathbf{i}}(\mathbf{k}', \mathbf{k}'_2),
\end{aligned} \tag{33}
$$

with obvious notation for $\mathbf{k}' = \big(\mathbf{k}, k'(\eta)\big)$.

There are clearly three separate cases, depending upon the sets c and d and whether or not they contain S_η. Considering first the case where $S_\eta \in c$, then substituting equations (28) and (31) into equation (33) gives,

$$
\begin{aligned}
y'^{cd}_{\mathbf{i}'\mathbf{j}'}(\mathbf{k}'_1, \mathbf{k}'_2) &= \sum_{\mathbf{k} \in K} \sum_{k'(\eta) \in K_\eta} a^{cd}_{\mathbf{i}'\mathbf{j}'}(k'_1(\eta), k'(\eta)) w^{(c\setminus S_\eta)(d\setminus S_\eta)}_{\mathbf{ij}}(\mathbf{k}_1, \mathbf{k}) \\
&\qquad \times g^{\eta}_{\mathbf{i}'}(k'(\eta), k'_2(\eta)) v^{c\setminus S_\eta}_{\mathbf{i}}(\mathbf{k}, \mathbf{k}_2) \\
&= \sum_{\mathbf{k} \in K} w^{(c\setminus S_\eta)(d\setminus S_\eta)}_{\mathbf{ij}}(\mathbf{k}_1, \mathbf{k}) v^{c\setminus S_\eta}_{\mathbf{i}}(\mathbf{k}, \mathbf{k}_2) \\
&\qquad \times \sum_{k'(\eta) \in K_\eta} a^{cd}_{\mathbf{i}'\mathbf{j}'}(k'_1(\eta), k'(\eta)) g^{\eta}_{\mathbf{i}'}(k'(\eta), k'_2(\eta)) \\
&= y^{(c\setminus S_\eta)(d\setminus S_\eta)}_{\mathbf{ij}}(\mathbf{k}_1, \mathbf{k}_2) \sum_{k'(\eta) \in K_\eta} a^{cd}_{\mathbf{i}'\mathbf{j}'}(k'_1(\eta), k'(\eta)) g^{\eta}_{\mathbf{i}'}(k'(\eta), k'_2(\eta)),
\end{aligned} \tag{34}
$$

by definition of $y^{(c\setminus S_\eta)(d\setminus S_\eta)}_{\mathbf{ij}}(\mathbf{k}_1, \mathbf{k}_2)$.

The remaining two cases (for $S_\eta \notin c, S_\eta \in d$ and $S_\eta \notin c, S_\eta \notin d$) can be shown similarly, using equations (28), (29) and (32). □

The next lemma provides a similar result, identifying the elements of the product matrix of $\mathbf{x}^d_{\mathbf{j}}$ with $(\mathbf{x}^c_{\mathbf{i}})^T$.

LEMMA A.2. *We can express the product of element* $\mathbf{k}'_1$ *of* $\mathbf{x}'^{d}_{\mathbf{j}'}$, *with element* $\mathbf{k}'_2$ *of* $\mathbf{x}'^{c}_{\mathbf{i}'}$ *as follows*:

$$
x'^{d}_{\mathbf{j}'}(\mathbf{k}'_1) x'^{c}_{\mathbf{i}'}(\mathbf{k}'_2)
$$

$$
= \begin{cases}
x^{(d\setminus S_\eta)}_{\mathbf{j}}(\mathbf{k}_1) x^{(c\setminus S_\eta)}_{\mathbf{i}}(\mathbf{k}_2) & \\
\quad \times [\mathscr{I}(k'_1(\eta) = j'(\eta)) - \mathscr{I}(k'_1(\eta) = |K_\eta|)] & \\
\quad \times [\mathscr{I}(k'_2(\eta) = j'(\eta)) - \mathscr{I}(k'_2(\eta) = |K_\eta|)], & \text{if } S_\eta \in c,\ S_\eta \in d, \\
x^{d}_{\mathbf{j}}(\mathbf{k}_1) x^{(c\setminus S_\eta)}_{\mathbf{i}}(\mathbf{k}_2) & \\
\quad \times [\mathscr{I}(k'_2(\eta) = j'(\eta)) - \mathscr{I}(k'_2(\eta) = |K_\eta|)], & \text{if } S_\eta \in c,\ S_\eta \notin d, \\
x^{(d\setminus S_\eta)}_{\mathbf{j}}(\mathbf{k}_1) x^{c}_{\mathbf{i}}(\mathbf{k}_2) & \\
\quad \times [\mathscr{I}(k'_1(\eta) = j'(\eta)) - \mathscr{I}(k'_1(\eta) = |K_\eta|)], & \text{if } S_\eta \notin c,\ S_\eta \in d, \\
x^{d}_{\mathbf{j}}(\mathbf{k}_1) x^{c}_{\mathbf{i}}(\mathbf{k}_2), & \text{if } S_\eta \notin c,\ S_\eta \notin d.
\end{cases}
$$

PROOF. Let us examine the case where $S_\eta \in c$, $S_\eta \in d$. From (10), we have that

$$
\begin{aligned}
(35)\quad x_{\mathbf{j}'}^{'d}(\mathbf{k}_1')x_{\mathbf{i}'}^{'c}(\mathbf{k}_2') &= I_{\mathbf{j}'}^{'d}(\mathbf{k}_1')I_{\mathbf{i}'}^{'c}(\mathbf{k}_2') \\
&= \prod_{\gamma: S_\gamma \in d} [\mathscr{I}(k_1'(\gamma) = j'(\gamma)) - \mathscr{I}(k_1'(\gamma) = |K_\gamma|)] \\
&\quad \times \prod_{\gamma: S_\gamma \in c} [\mathscr{I}(k_2'(\gamma) = i'(\gamma)) - \mathscr{I}(k_2'(\gamma) = |K_\gamma|)] \quad \text{by (3)} \\
&= \prod_{\gamma: S_\gamma \in d\setminus S_\eta} [\mathscr{I}(k_1'(\gamma) = j'(\gamma)) - \mathscr{I}(k_1'(\gamma) = |K_\gamma|)] \\
&\quad \times \prod_{\gamma: S_\gamma \in c\setminus S_\eta} [\mathscr{I}(k_2'(\gamma) = i'(\gamma)) - \mathscr{I}(k_2'(\gamma) = |K_\gamma|)] \\
&\quad \times [\mathscr{I}(k_1'(\eta) = j'(\eta)) - \mathscr{I}(k_1'(\eta) = |K_\eta|)] \\
&\quad \times [\mathscr{I}(k_2'(\eta) = i'(\eta)) - \mathscr{I}(k_2'(\eta) = |K_\eta|)] \\
&= x_{\mathbf{j}}^{d\setminus S_\eta}(\mathbf{k}_1)x_{\mathbf{i}}^{c\setminus S_\eta}(\mathbf{k}_2) \\
(36)\quad &\quad \times [\mathscr{I}(k_1'(\eta) = j'(\eta)) - \mathscr{I}(k_1'(\eta) = |K_\eta|)] \\
&\quad \times [\mathscr{I}(k_2'(\eta) = i'(\eta)) - \mathscr{I}(k_2'(\eta) = |K_\eta|)].
\end{aligned}
$$

Thus, the lemma holds for the case $S_\eta \in c$, $S_\eta \in d$ and the remaining cases follow similarly. □

PROOF OF LEMMA 3.2. We use the method of induction. We assume that the lemma is true for the set of sources S, and consider the set of sources $S' = S \cup \{S_\eta\}$. As usual, we use the notation that all the matrices and vectors corresponding to the set of sources S' be indexed by a "prime," that is, $'$, while all other terms correspond to the set of sources S. We establish the result by considering the individual elements of both product matrices in (9) and show them to be the same for any combination of cases for S_η being in only c or d, neither or both. If each of the elements are the same, then clearly so are the corresponding matrices that they comprise.

We begin with the case where $S_\eta \in c$ and $S_\eta \in d$. From Lemma A.1, we have that for $\mathbf{k}_1' = (\mathbf{k}_1, k_1'(\eta))$ and $\mathbf{k}_2' = (\mathbf{k}_2, k_2'(\eta))$, the $(\mathbf{k}_1', \mathbf{k}_2')$ element of the product $W_{\mathbf{i}'\mathbf{j}'}^{'cd} V_{\mathbf{i}'}^{'c}$, for sources S' is given by

$$
\begin{aligned}
y_{\mathbf{i}'\mathbf{j}'}^{'cd}(\mathbf{k}_1', \mathbf{k}_2') &= y_{\mathbf{ij}}^{(c\setminus S_\eta)(d\setminus S_\eta)}(\mathbf{k}_1, \mathbf{k}_2) \sum_{k'(\eta)\in K_\eta} a_{\mathbf{i}'\mathbf{j}'}^{cd}(k_1'(\eta), k'(\eta)) g_{\mathbf{i}'}^{\eta}(k'(\eta), k_2'(\eta)) \\
&= x_{\mathbf{j}}^{d\setminus S_\eta}(\mathbf{k}_1)x_{\mathbf{i}}^{c\setminus S_\eta}(\mathbf{k}_2) \sum_{k'(\eta)\in K_\eta} a_{\mathbf{i}'\mathbf{j}'}^{cd}(k_1'(\eta), k'(\eta)) g_{\mathbf{i}'}^{\eta}(k'(\eta), k_2'(\eta)),
\end{aligned}
$$

since the result is assumed to hold for the set of sources S. Then, from Lemma A.2, it is clear that $x_{\mathbf{j}'}^{'d}(\mathbf{k}_1')x_{\mathbf{i}'}^{'c}(\mathbf{k}_2') = y_{\mathbf{i}'\mathbf{j}'}^{'cd}(\mathbf{k}_1', \mathbf{k}_2')$ if

$$\begin{aligned}\sum_{k'(\eta)\in K_\eta} & a_{\mathbf{i}'\mathbf{j}'}^{cd}(k_1'(\eta), k'(\eta))g_{\mathbf{i}'}^{\eta}(k'(\eta), k_2'(\eta)) \\ &= [\mathscr{I}(k_1'(\eta) = j'(\eta)) - \mathscr{I}(k_1'(\eta) = |K_\eta|)] \\ &\quad \times [\mathscr{I}(k_2'(\eta) = i'(\eta)) - \mathscr{I}(k_2'(\eta) = |K_\eta|)].\end{aligned} \tag{37}$$

Since $S_\eta \in c$ and $S_\eta \in d$, then

$$a_{\mathbf{i}'\mathbf{j}'}^{cd}(m, n) = \begin{cases} 1, & \text{if } (m, n) = (i'(\eta), j'(\eta)); (j'(\eta), i'(\eta)); \text{ or } (|K_\eta|, |K_\eta|), \\ 0, & \text{otherwise,} \end{cases}$$

and

$$g_{\mathbf{i}'}^{\eta}(m, n) = \begin{cases} 1, & \text{if } (m, n) = (i'(\eta), i'(\eta)); \text{ or } (|K_\eta|, |K_\eta|), \\ -1, & \text{if } (m, n) = (i'(\eta), |K_\eta|); \text{ or } (|K_\eta|, i'(\eta)), \\ 0, & \text{otherwise,} \end{cases}$$

from their definitions given at the beginning of this section.

To show that (37) is true, we must consider all of the different possible combinations of the indicator functions on the right-hand side of the equation. For convenience we set

$$I_1(j'(\eta)) = I(k_1'(\eta) = j'(\eta)); \qquad I_1(|K_\eta|) = \mathscr{I}(k_1'(\eta) = |K_\eta|);$$

and

$$I_2(j'(\eta)) = \mathscr{I}(k_2'(\eta) = j'(\eta)); \qquad I_2(|K_\eta|) = \mathscr{I}(k_2'(\eta) = |K_\eta|).$$

The full proof considers all possible combinations of the values of these indicator functions and the corresponding value of the product on the right-hand side of (37). For clarity, we list all possible values in Table 1 and for brevity we shall consider only the first case, where $k_1'(\eta) = j'(\eta)$ and $k_2'(\eta) = i'(\eta)$.

TABLE 1

All possible combinations of the values of the indicator functions of interest and the corresponding product of $[I_1(j'(\eta)) - I_1(|K_\eta|)] \times [I_2(j'(\eta)) - I_2(|K_\eta|)]$

Case	$I_1(j(\eta))$	$I_1(\|K_\eta\|)$	$I_2(j(\eta))$	$I_2(\|K_\eta\|)$	**Product**
1	1	0	1	0	1
2	1	0	0	1	−1
3	0	1	1	0	−1
4	0	1	0	1	1
5	0	0	0 or 1	0 or 1	0
6	0 or 1	0 or 1	0	0	0

The terms "0" and "1" in the table indicate the value taken by the corresponding indicator function.

If $k'_1(\eta) = j'(\eta)$ and $k'_2(\eta) = i'(\eta)$, then

$$a^{cd}_{\mathbf{i}'\mathbf{j}'}(k'_1(\eta), k'(\eta)) = \begin{cases} 1, & \text{if } k'(\eta) = i'(\eta), \\ 0, & \text{otherwise,} \end{cases}$$

and

$$g^{\eta}_{\mathbf{i}'}(k'(\eta), k'_2(\eta)) = \begin{cases} 1, & \text{if } k'(\eta) = i'(\eta), \\ -1, & \text{if } k'(\eta) = |K_\eta|, \\ 0, & \text{otherwise.} \end{cases}$$

This implies that

$$a^{cd}_{\mathbf{i}'\mathbf{j}'}(k'_1(\eta), k'(\eta)) g^{\eta}_{\mathbf{i}'}(k'(\eta), k'_2(\eta)) = \begin{cases} 1, & \text{if } k'(\eta) = i'(\eta), \\ 0, & \text{if } k'(\eta) \neq i'(\eta). \end{cases}$$

So that

$$\sum_{k'(\eta)\in K_\eta} a^{cd}_{\mathbf{i}'\mathbf{j}'}(k'_1(\eta), k'(\eta)) g^{\eta}_{\mathbf{i}'}(k'(\eta), k'_2(\eta)) = 1,$$

since $k'(\eta) = i'(\eta)$ exactly once in the sum, and which corresponds to the value of the indicator function product given in Table 1. Therefore, the result holds in this case. The remaining five cases in Table 1 follow similarly and hence for the case $S_\eta \in c$ and $S_\eta \notin d$, (37) is satisfied.

There remain three other cases which work somewhat similarly. In the case that $S_\eta \in c$; $S_\eta \notin d$, we need only show that

$$\begin{aligned} &\sum_{k'(\eta)\in K_\eta} a^{cd}_{\mathbf{i}'\mathbf{j}'}(k'_1(\eta), k'(\eta)) g^{\eta}_{\mathbf{i}'}(k'(\eta), k'_2(\eta)) \\ &\qquad = [\mathscr{I}(k'_2(\eta) = i'(\eta)) - \mathscr{I}(k'_2(\eta) = |K_\eta|)], \end{aligned} \tag{38}$$

which can be proved using similar arguments to the previous case. For the case where $S_\eta \notin c$; $S_\eta \in d$, we need only show that

$$\sum_{k'(\eta)\in K_\eta} a^{cd}_{\mathbf{i}'\mathbf{j}'}(k'_1(\eta), k'(\eta)) = [\mathscr{I}(k'_1(\eta) = j'(\eta)) - \mathscr{I}(k'_1(\eta) = |K_\eta|)]. \tag{39}$$

Both these proofs follow similar lines to the first case described above and the details are omitted for brevity.

The final case in which $S_\eta \notin c, d$ follows directly from the final case of Lemma A.1, so that

$$\begin{aligned} y'^{cd}_{\mathbf{i}'\mathbf{j}'}(\mathbf{k}'_1, \mathbf{k}'_2) &= y^{cd}_{\mathbf{ij}}(\mathbf{k}_1, \mathbf{k}_2) \\ &= x^{d}_{\mathbf{j}}(\mathbf{k}_1) x^{c}_{\mathbf{i}}(\mathbf{k}_2) \quad \text{since the lemma is true for } S \\ &= x'^{d}_{\mathbf{j}'}(\mathbf{k}_1) x'^{c}_{\mathbf{i}'}(\mathbf{k}_2) \quad \text{since } \mathbf{i} = \mathbf{i}' \text{ and } \mathbf{j} = \mathbf{j}' \text{ with } S_\eta \notin c \text{ and } S_\eta \notin d. \end{aligned}$$

Hence, we have shown that if the result holds for a set of sources S, then it will also hold if we add a new source S_η, that is,

$$W'^{cd}_{\mathbf{ij}} V'^{c}_{\mathbf{i}} = \mathbf{x}'^{d}_{\mathbf{j}} (\mathbf{x}'^{c}_{\mathbf{i}})^T.$$

All that remains of our inductive proof is to show that the result holds when we have just one source and the general result will follow.

Consider the simple case where $S = \{S_\eta\}$. For $k_1, k_2 \in K_\eta$, we have $y^{cd}_{ij}(k_1, k_2)$ is the (k_1, k_2)th element of $W^{cd}_{ij} V^{c}_{i}$, where $c, d \in \mathscr{P}(S) = \{\varnothing, S_\eta\}$, and i and j represent the level of sources c and d, respectively.

In the case where $c = d = S_\eta$, we have,

$$\begin{aligned}
y^{cd}_{ij}(k_1, k_2) &= \sum_{k \in K_\eta} w^{cd}_{ij}(k_1, k) v^{c}_{i}(k, k_2) \\
&= \sum_{k \in K_\eta} a^{cd}_{ij}(k_1, k) g^{\eta}_{i}(k, k_2) \quad \text{since } |S| = 1 \text{ and using equation (34)} \\
&= \begin{cases} 1, & \text{if } (k_1, k_2) = (i, j) \text{ or } (|K_\eta|, |K_\eta|), \\ -1, & \text{if } (k_1, k_2) = (|K_\eta|, j); \text{ or } (i, |K_\eta|), \\ 0, & \text{otherwise} \end{cases} \\
&\quad \text{using the definitions of } a^{cd}_{ij}(k_1, k) \text{ and } g^{\eta}_{i}(k, k_2) \\
&= (\mathscr{I}(k_1 = i) - \mathscr{I}(k_1 = |K_\eta|)) \times (\mathscr{I}(k_2 = j) - \mathscr{I}(k_2 = |K_\eta|)) \\
&= I^{c}_{i}(k_1) I^{d}_{j}(k_2) \quad \text{by (3)} \\
&= x^{c}_{i}(k_1) x^{d}_{j}(k_2) \quad \text{by (10)}.
\end{aligned}$$

The remaining cases (where $c = S_\eta, d = \varnothing$; $c = \varnothing, d = S_\eta$; $c = \varnothing, d = \varnothing$) follow similarly. Hence,

$$\mathbf{x}^{d}_{j} (\mathbf{x}^{c}_{i})^T = W^{cd}_{ij} V^{c}_{i},$$

and the lemma holds for the case in which we have just a single source and, by induction, for any collection of sources. □

Acknowledgments. The authors gratefully acknowledge the contribution of two anonymous referees, an Associate Editor and Jim Berger as Editor for their constructive criticism of an earlier version of this paper. In addition we would like to thank Jon Forster for his extremely detailed and helpful feedback on earlier drafts and also Ernie Hook for providing us with expert priors for the analysis of the spina bifida data discussed in this paper. Finally, we would like to acknowledge the support of the U.K. Engineering and Physical Sciences Research Council in funding the research of both authors.

REFERENCES

AITCHISON, J. and BROWN, J. A. C. (1957). *The Lognormal Distribution: With Special Reference to Its Uses in Economics*. Cambridge Univ. Press.

DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317.

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

EVANS, M., GILULA, Z. and GUTTMAN, I. (1993). Computational issues in the Bayesian analysis of categorical data: log-linear and Goodman's RC model. *Statist. Sinica* **3** 391–406.

FORSTER, J. J. (1992). Models and marginal densities for multiway contingency tables. Ph.D. thesis, Univ. Nottingham.

GIUDICI, P., GREEN, P. J. and TARANTOLA, C. (1999). Efficient model determination for discrete graphical models. Technical report, Univ. Pavia, Italy.

HOOK, E. B., ALBRIGHT, S. G. and CROSS, P. K. (1980). Use of binomial census and log-linear methods for estimating the prevalence of spina bifida in livebirths and the completeness of vital record reports in New York State. *J. Amer. Epidemiology* **112** 750–758.

HOOK, E. B. and REGAL, R. R. (1995). Capture–recapture methods in epidemiology: methods and limitations. *Epidemiologic Rev.* **17** 243–264.

KING, R. and BROOKS, S. P. (2001). On the Bayesian analysis of population size. *Biometrika* **88** 317–336.

KNUIMAN, M. W. and SPEED, T. P. (1988). Incorporating prior information into the analysis of contingency tables. *Biometrics* **44** 1061–1071.

MADIGAN, D. and YORK, J. C. (1997). Bayesian methods for estimation of the size of a closed population. *Biometrika* **84** 19–31.

STATISTICAL LABORATORY
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD
CAMBRIDGE, CB3 0WB
UNITED KINGDOM
E-MAIL: S.P.Brooks@statslab.cam.ac.uk