Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 136 (2006) 355–372
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Data augmentation in multi-way contingency tables with fixed marginal totals

Adrian Dobra[a,*], Claudia Tebaldi[b], Mike West[a]

[a] *Institute of Statistics and Decision Sciences, Duke University, 211c Old Chem, Box 90251, Durham, NC 27708-0251, USA*
[b] *National Center for Atmospheric Research, Boulder, CO 80305, USA*

Received 11 August 2003; accepted 11 July 2004
Available online 11 September 2004

Abstract

We describe and illustrate approaches to data augmentation in multi-way contingency tables for which partial information, in the form of subsets of marginal totals, is available. In such problems, interest lies in questions of inference about the parameters of models underlying the table together with imputation for the individual cell entries. We discuss questions of structure related to the implications for inference on cell counts arising from assumptions about log-linear model forms, and a class of simple and useful prior distributions on the parameters of log-linear models. We then discuss "local move" and "global move" Metropolis–Hastings simulation methods for exploring the posterior distributions for parameters and cell counts, focusing particularly on higher-dimensional problems. As a by-product, we note potential uses of the "global move" approach for inference about numbers of tables consistent with a prescribed subset of marginal counts. Illustration and comparison of MCMC approaches is given, and we conclude with discussion of areas for further developments and current open issues.

MSC: 62F15; 62H17

Keywords: Bayesian inference; Disclosure limitation; Fixed margins problem; Imputation; Log-linear models; Markov basis; Markov chain Monte Carlo; Missing data

* Corresponding author.
E-mail addresses: adobra@stat.duke.edu (Adrian Dobra), tebaldi@ucar.edu (Claudia Tebaldi), mw@stat.duke.edu (Mike West).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.07.002

1. Introduction

The general problem of inference in contingency tables based on partial information in terms of observed counts on a set of margins has become of increasing interest in recent years. We address this problem here in a framework involving inference on parameters of statistical models underlying multi-way tables together with inference about missing cell entries.

Some of our initial motivating interest in this area came from socio-economic and demographic studies that involve and rely on survey and census data representing differing levels of aggregation (hence marginalisation) of population characteristics. Much of the activity in these latter areas has been referred to as micro-simulation, and the development of micro-simulation methods in areas such as transportation policy and planning rely heavily on an ability to impute individual household level counts, for example, from more highly aggregated data from local or population census data.

More recently, the last several years have seen a very significant upsurge of interest in development of methods to aid in the creation, dissemination and use of public data sets from governmental sources, related to serious societal and legal concerns about data confidentiality and security. A range of issues then arise about the potential to infer individual level data from sets of interlinked aggregate-level data, and this can be focused on the problem of inferring cell counts in multi-way tables based on observation of some sets of marginal totals. Here there are questions of the extent to which sets of observed margins can inform on cell counts under varying assumptions about the structure of candidate statistical models (such as log-linear models), with related questions about the role and impact of specific prior distributions on parameters of such models (Knuiman and Speed, 1988; Gelman et al., 2003; Dobra et al., 2003a).

Historically, approaches to fitting models to incomplete contingency tables are discussed in classical texts such as Deming and Stephan (1940) or Bishop et al. (1977). Strong interest in the general problem area has focused on problems of counting tables with prescribed marginal totals, a goal of interest in both data confidentiality studies and in the traditional context of estimating significance levels in testing approaches (Agresti, 1992; Diaconis and Efron, 1985; Smith et al., 1996).

Bayesian inference in this context is in principle straightforward: we aim to compute posterior distributions for the unobserved cell counts and parameters underlying models for cell probabilities, jointly. In practice this may be addressed using Markov chain Monte Carlo (MCMC) simulations; this requires creativity in dealing, in particular, with simulations of the missing cell counts from appropriate conditional posterior distributions. The contributions of this article address a number of questions and needs in support of the practical development of such methods. First, we describe the general problem and context, and develop some theoretical insights into the nature and role of assumptions about model structure in its relation to the problem of inference on individual cell counts based on observation of sets of marginal totals. Then, we discuss MCMC approaches to joint analysis of parameters and missing cell counts. This uses a simple but flexible class of prior distributions on parameters of log-linear models at the parametric level. In imputing cell counts, we discuss "local move" algorithms that rely on Markov basis construction, together with a new class of "global move" approaches that have some relative attractions, especially as problems

increase in dimension and complexity. A detailed discussion of an example demonstrates both the implementation and the efficacy of the "global move" algorithm, and we conclude with some discussion of current open issues and challenges. In addition to innovations in modelling and computation, the work represents a selective overview of some key recent and current issues in this field.

2. Analysis framework and goals

2.1. Definitions and notation

Consider a k-way contingency table of counts over a k-vector of discrete random variables $X=\{X_1, X_2, \ldots, X_k\}$. Let $K=\{1, 2, \ldots, k\}$, and for each $j \in K$, suppose X_j takes values in $\mathscr{I}_j=\{1, \ldots, I_j\}$. Write $\mathscr{I}=\mathscr{I}_1 \times \cdots \times \mathscr{I}_k$ and denote an element of $\mathscr{I}$ by $i=(i_1, \ldots, i_k)$.

The contingency table $n = \{n(i)\}_{i\in\mathscr{I}}$ is a k-dimensional array of non-negative integer numbers, with cell entries $n(i)=\#\{X=i\}$, $(i \in \mathscr{I})$, and a total of $m=I_1 \cdot \ldots \cdot I_k$ cells. Any set of marginal counts is obtained by summation over one or more of the X variables. For any target subset of variables $D \subset K$, the D-marginal n_D of n has cells $i_D \in \mathscr{I}_D = \times_{j\in D} \mathscr{I}_j$, with cell entries

$$n_D(i_D) = \sum_{j\in\mathscr{I}_{K\setminus D}} n(i_D, j).$$

If $D=\emptyset$ then $n_\emptyset$ is the grand total over n. Cells are ordered lexiographically with the index of the kth variable varying fastest, so that $\mathscr{I} = \{i^1, \ldots, i^m\}$, where $i^1 = (1, \ldots, 1)$ is the first cell and $i^m = (I_1, \ldots, I_k)$ is the last cell.

2.2. Tables with sets of fixed margins

Our interest lies in problems in which we observe only subsets of margins from the full table. Suppose the l margins $\mathscr{D}=\{n^1, \ldots, n^l\}$ are recorded. Additional information may be available, such as upper and lower bounds for some of (or all) the cells in the full table, or cases of structural zeroes (that can also be represented by fixed upper and lower bounds, in this case each at zero). In such cases, the constraints can be added to $\mathscr{D}$ without changing the development below. Observing the margins $\mathscr{D}$ induces constraints that imply upper and lower bounds $U(i)=\max\{n(i) : n \in \mathscr{T}\}$ and $L(i)=\min\{n(i) : n \in \mathscr{T}\}$ on each cell entry $n(i)$, $i \in \mathscr{I}$. This limits attention to tables satisfying these constraints. Denote the set of such tables by $\mathscr{T}$–thus $\mathscr{T}$ is the set of k-way tables $n = \{n(i)\}_{i\in\mathscr{I}}$ strictly compatible with $\mathscr{D}$. Write $\mathscr{M}(\mathscr{T})$ for the number of such tables.

2.3. Statistical models over tables and observed margins

Focus on the independent Poisson sampling model that underlies the canonical multinomial distribution for n. That is, cell counts are conditionally independent Poisson, $n(i) \sim$ Poisson$(\lambda(i))$, for each $i \in \mathscr{I}$, with positive means $\lambda = \{\lambda(i)\}$.

The implied probability of the observed set of margins $\mathscr{D}$ is then, theoretically, simply

$$p(\mathscr{D}|\lambda) = \sum_{n' \in \mathscr{T}} p(n'|\lambda). \tag{1}$$

In problems of inference on λ, this defines the likelihood function, but, evidently, in other than trivial, low-dimensional problems, direct evaluation is impossible. The essential role of n as (in part) missing or latent data underlies simulation-based methods using MCMC approaches that, following Tanner and Wong (1987), iteratively resimulate the "missing" components of n and the parameters λ from relevant conditional distributions. For the former, note that

$$p(n, \mathscr{D}|\lambda) = \begin{cases} p(n|\lambda) & \text{if } n \in \mathscr{T}, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Then, conditional on the observed margins and model parameters, inference on the missing components of the table are derived directly from the implied conditional posterior

$$p(n|\mathscr{D}, \lambda) = p(n|\lambda) \Big/ \sum_{n' \in \mathscr{T}} p(n'|\lambda) \tag{3}$$

if $n \in \mathscr{T}$, being zero otherwise. When embedded in a simulation-based analysis, a key technical issue is that of developing methods to simulate this distribution—proportional to the product of Poisson components conditioned by the complicated set of constraints $n(i) \in [L(i), L(i)+1, \ldots, U(i)-1, U(i)]$ over cells $i \in \mathscr{I}$ defined by the observed margins $\mathscr{D}$.

2.4. Population models and parameters

Inference about the underlying population structure is based on choices of model, such as traditional log-linear models, for λ. Any model $\mathscr{A}$ has parameters θ that define $\lambda = \lambda(\theta)$. Then priors are specified on θ, so on λ indirectly and the latter will incorporate deterministic constraints imposed by the model. Thus we work interchangeably between λ and θ in notation. Sampling distributions are conditional on λ, thus implicitly θ. For the given model $\mathscr{A}$, denote a prior density for the implied model parameters θ by $p(\theta|\mathscr{A})$. Posterior inference is theoretically defined through the intractable likelihood function (1). When elaborated to include the missing data to fill out the full table, the more tractable conditional posterior is simply the prior modified by the product of Poisson likelihood components $p(n|\lambda)$.

2.5. MCMC framework

From the above components of the posterior distribution, we may implement MCMC methods to generate, ultimately, samples from $p(n, \theta|\mathscr{D}, \mathscr{A})$ by iteratively re-simulating values of θ (and hence, by direct evaluation, λ), and the missing cell counts in n. Start with

$n^{(0)} \in \mathscr{T}$. At the sth step of the algorithm, do (Tanner and Wong, 1987):

- Simulate $\theta^{(s+1)}$ from $p(\theta|\mathscr{A}, n^{(s)}) \propto p(\theta|\mathscr{A})p(n^{(s)}|\lambda(\theta))$, and compute the implied new value of $\lambda^{(s+1)} = \lambda(\theta^{(s+1)})$.
- Simulate $n^{(s+1)}$ from $p(n|\mathscr{D}, \lambda^{(s+1)})$.

The above data augmentation algorithm represents the basis of Bayesian approaches for analyzing contingency tables with missing data—see, for example, Gelman et al. (2003) or Schafer (1997). This also connects to the well-known EM algorithm of Dempster et al. (1977) and to the classical book about missing data by Little and Rubin (2002).

Section 3 discusses aspects of prior specification and model structure in their impact on $p(n|\mathscr{D}, \lambda)$. Section 4 introduces the Czech autoworkers data used to illustrate the methodology. Section 5 considers specific priors and the resulting analysis in log-linear models. Section 6 discusses algorithms to sampling the critical conditional posterior for the missing elements of n under the relevant distribution (3). Section 7 finalizes the discussion of the Czech autoworkers data example.

3. Log-linear models and structural information

Some initial theoretical results describe structural aspects of the conditional posterior $p(n|\mathscr{D}, \lambda)$ relevant in consideration of classes of log-linear models. We first note that the conditional distribution (3) of course has precisely the same form if we assume multinomial sampling for n.

Suppose $\mathscr{A}$ is a specified log-linear model defining λ in terms of underlying log-linear parameters θ. This may be defined in terms of a specification of the minimal sufficient statistics of $\mathscr{A}$; suppose these to be defined by the index sets $\{C_1, C_2, \ldots, C_q\}$, where $C_j \subset K$. Writing $\mathscr{C} = \{C : \emptyset \neq C \subset C_j \text{ for some } j \in \{1, 2, \ldots, q\}\}$, the log-linear model has the form

$$\lambda(i) = \mu \prod_{C \in \mathscr{C}} \psi_C(i_C), \tag{4}$$

where ψ_C depends on $i \in \mathscr{I}$ only through the indices in C. To make the parameters identifiable, the aliasing constraints are to set each $\psi_C(i_C) = 1$ whenever there exists $p \in C$ with $i_p = 1$—see, for example, Whittaker (1990). One immediate consequence of these aliasing constraints is that $\lambda(i^1) = \mu$. The parameter set is then

$$\theta = \{\mu\} \cup \{\psi_C(i_C) : C \in \mathscr{C},\ i_C \in \mathscr{I}_C \text{ with } i_p \neq 1 \text{ for } p \in C\}. \tag{5}$$

The following theorem now shows that the posterior $p(n|\mathscr{D}, \lambda)$ from (3) simplifies and does not depend on all the parameters θ from (5). This represents an extension of an earlier result by Haberman (1974).

Theorem 1. *If the marginal n_C of the full table is determined from $\mathscr{D}$, then, under model $\mathscr{A}$, the posterior distribution $p(n|\mathscr{D}, \lambda)$ does not depend on the parameters $\psi_C(i_C)$ for all $i_C \in \mathscr{I}_C$.*

Proof. It is not hard to see that, for any table n', we have

$$\prod_{i\in\mathcal{I}} \lambda(i)^{n'(i)} = \mu^{n'_\emptyset} \prod_{C\in\mathcal{C}} \prod_{i_C\in\mathcal{I}_C} \psi_C(i_C)^{n'_C(i_C)}.$$

If $n' \in \mathcal{T}$, it follows that the grand totals of n' and n are equal, i.e., $n'_\emptyset = n_\emptyset$. Moreover, if the marginal n_C is known from $\mathcal{D}$, then the marginals n'_C and n_C coincide. Thus

$$\prod_{i_C\in\mathcal{I}_C} \psi_C(i_C)^{n'_C(i_C)} = \prod_{i_C\in\mathcal{I}_C} \psi_C(i_C)^{n_C(i_C)}$$

for every table $n' \in \mathcal{T}$. It follows that the terms involving $\psi_C(i_C)$, $i_C \in \mathcal{I}_C$, cancel in the denominator and the numerator of (3) and hence $p(n|\mathcal{D}, \lambda)$ does not depend on these parameters. Note that this posterior also does not depend on the grand mean parameter μ. □

A direct consequence of Theorem 1 is that the hypergeometric distribution is a special case of the posterior in (3) obtained by conditioning on a log-linear model whose minimal sufficient statistics are fully determined by the available data $\mathcal{D}$:

$$p(n|\mathcal{D}, \lambda) \equiv p(n|\mathcal{D}) = \left[\prod_{i\in\mathcal{I}} n(i)!\right]^{-1} \Bigg/ \sum_{n'\in\mathcal{T}} \left[\prod_{i\in\mathcal{I}} n'(i)!\right]^{-1}. \tag{6}$$

As an aside, we note that Sundberg (1975) shows that the normalizing constant in (6) can be directly evaluated if the log-linear model $\mathcal{A}$ is decomposable (Whittaker, 1990; Lauritzen, 1996); otherwise, this normalizing constant can be computed only if the set of tables $\mathcal{T}$ can be exhaustively enumerated.

Theorem 1 shows that the terms ψ_C associated with a fixed marginal n_C are redundant in the model since $p(n|\mathcal{D}, \lambda)$ does not depend on them. Only higher-order terms corresponding with marginals that are not known effectively influence the simulation of a new complete table. This means that the log-linear models $\mathcal{A}$ that have to be considered in order to obtain different posterior distributions $p(n|\mathcal{D}, \mathcal{A}, \cdot)$ are those log-linear models whose minimal sufficient statistics embed the set of fixed marginals which constitute part or all the available information $\mathcal{D}$.

4. Example—Czech autoworkers data

The data in Table 1 come from a prospective epidemiological study of 1841 workers in a Czechoslovakian car factory, as part of an investigation of potential risk factors for coronary thrombosis (see, Edwards and Havranek, 1985). In the left-hand panel of Table 1, A indicates whether the worker smokes or not, B corresponds to "strenuous mental work", C corresponds to "strenuous physical work", D corresponds to "systolic blood pressure", E corresponds to "ratio of β and α lipoproteins" and F represents "family anamnesis of coronary heart disease". We focus only on the cell (1, 2, 2, 1, 1, 2) that contains the unique count of 1.

Table 1
Czech autoworkers data from Edwards and Havranek (1985)

				B	No		Yes		B	No		Yes	
F	E	D	C	A	No	Yes	No	Yes	A	No	Yes	No	Yes
Neg	<3	<140	No		44	40	112	67		[35, 45]	[35, 44]	[111, 121]	[63, 72]
			Yes		129	145	12	23		[128, 138]	[141, 150]	[3, 13]	[18, 27]
		$\geqslant 140$	No		35	12	80	33		[29, 39]	[5, 14]	[76, 86]	[31, 40]
			Yes		109	67	7	9		[105, 115]	[65, 74]	[1, 11]	[2, 11]
	$\geqslant 3$	<140	No		23	32	70	66		[16, 25]	[26, 35]	[68, 77]	[63, 72]
			Yes		50	80	7	13		[48, 57]	[77, 86]	[0, 9]	[7, 16]
		$\geqslant 140$	No		24	25	73	57		[19, 28]	[16, 25]	[69, 78]	[57, 66]
			Yes		51	63	7	16		[47, 56]	[63, 72]	[2, 11]	[7, 16]
Pos	<3	<140	No		5	7	21	9		[4, 14]	[3, 12]	[12, 22]	[4, 13]
			Yes		9	17	1	4		[0, 10]	[12, 21]	[0, 10]	[0, 9]
		$\geqslant 140$	No		4	3	11	8		[0, 10]	[1, 10]	[5, 15]	[1, 10]
			Yes		14	17	5	2		[8, 18]	[10, 19]	[1, 11]	[0, 9]
	$\geqslant 3$	<140	No		7	3	14	14		[5, 14]	[0, 9]	[7, 16]	[8, 17]
			Yes		9	16	2	3		[2, 11]	[10, 19]	[0, 9]	[0, 9]
		$\geqslant 140$	No		4	0	13	11		[0, 9]	[0, 9]	[8, 17]	[2, 11]
			Yes		5	14	4	4		[0, 9]	[5, 14]	[0, 9]	[4, 13]

The left-hand panel contains the cell counts and the right-hand panel contains the bounds given the margins $\mathscr{R}_1$. The cell containing the unique count of 1 and its corresponding bounds are marked with a box.

Table 2
Relevant log-linear models for $\mathscr{R}_1$

Log-linear model	Minimal sufficient statistics
$\mathscr{A}_1$	$\mathscr{R}_1 \cup \{n_{\{B,C,D,E,F\}}\}$
$\mathscr{A}_2$	$\mathscr{R}_1 \cup \{n_{\{A,B,C,E,F\}}\}$
$\mathscr{A}_3$	$\mathscr{R}_1 \cup \{n_{\{A,B,C,D,F\}}\}$
$\mathscr{A}_4$	$\mathscr{R}_1 \cup \{n_{\{B,C,D,E,F\}}, n_{\{A,B,C,E,F\}}\}$
$\mathscr{A}_5$	$\mathscr{R}_1 \cup \{n_{\{B,C,D,E,F\}}, n_{\{A,B,C,D,F\}}\}$
$\mathscr{A}_6$	$\mathscr{R}_1 \cup \{n_{\{A,B,C,E,F\}}, n_{\{A,B,C,D,F\}}\}$
$\mathscr{A}_7$	$\mathscr{R}_1 \cup \{n_{\{B,C,D,E,F\}}, n_{\{A,B,C,E,F\}}, n_{\{A,B,C,D,F\}}\}$
$\mathscr{A}_8$	Saturated

We assume that the information we have about this six-way table n consists of the set of marginals

$$\mathscr{R}_1=\{n_{\{A,C,D,E,F\}}, n_{\{A,B,D,E,F\}}, n_{\{A,B,C,D,E\}}, n_{\{B,C,D,F\}}, n_{\{A,B,C,F\}}, n_{\{B,C,E,F\}}\}.$$

Note that $\mathscr{R}_1$ contains most of the marginals of the Czech autoworkers data—the omissions are three five-way marginals. Let $\mathscr{T}_1$ be the set of dichotomous six-way tables consistent with $\mathscr{R}_1$. Using the generalized shuttle algorithm (Dobra et al., 2003b; Dobra, 2002) we find that $\mathscr{T}_1$ contains 810 tables. The upper and lower bounds on cell entries induced by $\mathscr{R}_1$ are given in the right-hand panel of Table 1. Given $\mathscr{R}_1$, every cell in this table can take 10 or 11 possible values.

If the marginals $\mathscr{R}_1$ are fixed, the corresponding set of relevant log-linear models is given in Table 2. These are the models that contain at least one interaction term that is not associated with known marginals.

5. Prior specification and posterior sampling of model parameters

Consider a general log-linear model $\mathscr{A}$ with minimal sufficient statistics specified by the index sets $\{C_1, C_2, \ldots, C_q\}$. Thus, if $\lambda = \{\lambda(i)\}_{i\in\mathscr{I}}$ is consistent with $\mathscr{A}$, then $\lambda(i)$ is represented as in (4). As developed in West (1997), and then extended in Tebaldi and West (1998a) and Tebaldi and West (1998b), independent gamma priors on the multiplicative parameters in the log-linear model representation imply that all complete, univariate conditional posteriors are also of gamma form. Thus drawing new values for the model parameters, and hence the values of the $\lambda(i)$, is immediately accessible using Gibbs sampling. Specifically, we note that:

- $p(\mu|\mathscr{A}, n, \theta\backslash\mu) \propto p(\mu)\mu^{n_\emptyset} \exp\{-\mu \sum_{i\in\mathscr{I}} \prod_{C\in\mathscr{C}} \psi_C(i_C)\}$.
- For each $\alpha := \psi_{C_0}(i^0_{C_0}) \in \Theta\backslash\mu$, the complete conditional posterior for α is proportional to

$$p(\alpha)\alpha^{n_{C_0}(i^0_{C_0})} \exp\left\{-\alpha\mu \sum_{\{i\in\mathscr{I}: i_{C_0}=i^0_{C_0}\}} \prod_{C\in\mathscr{C}} \psi_C(i_C)\right\}. \tag{7}$$

Now, when $\mathscr{A}$ is the saturated log-linear model, $\mathscr{A}$ is specified by one minimal sufficient statistic given by the complete index set K. In this case $\mathscr{C}$ is comprises all the non-empty subsets of K. For $i \in \mathscr{I}$ the conditional posterior for $\beta := \psi_K(i)$ is then proportional to

$$p(\beta)\beta^{n(i)} \exp\left\{-\beta\mu \prod_{\{C:C\subset K,\, C\neq K\}} \psi_C(i_C)\right\}.$$

Thus gamma priors are conditionally conjugate. Alternatively, finite uniform priors might serve as at least initial objective priors. Thus, conditional on a complete table n, we can simulate new parameter values for Poisson rates $\lambda(i)$, $i \in \mathscr{I}$, via sets of independent draws from gamma distributions, or truncated gamma distributions.

6. Imputing cell counts

The major component of MCMC analyses relates to the sampling algorithms to generate complete tables of counts n that are consistent with the constraints defined by the marginal count information $\mathscr{D}$. Direct sampling is computationally infeasible because the normalizing constant in (3) would have to be evaluated at each iteration, and this evaluation cannot be done quickly, if at all, due to the existence of a huge number of tables consistent with $\mathscr{D}$. Hence, some form of embedded Metropolis–Hastings method is required within the overall MCMC that also samples λ.

Given a current state $(n^{(s)}, \lambda^{(s+1)})$, a candidate table n^* is generated from a specified proposal distribution $q(n^{(s)}, n^*)$, and accepted with probability

$$\min\left[1, \frac{p(n^*|\lambda^{(s+1)})q(n^*, n^{(s)})}{p(n^{(s)}|\lambda^{(s+1)})q(n^{(s)}, n^*)}\right]. \tag{8}$$

The only requirement the proposal distribution has to satisfy is that $q(n^{(s)}, n^*) > 0$ if and only if $q(n^*, n^{(s)}) > 0$. In contrast with the direct sampling approaches, it is neither necessary to identify the support $\mathscr{T}$ of $p(n|\mathscr{D}, \lambda)$, nor to evaluate $p(n|\lambda)$ completely across the support. If a proposal distribution generates candidate tables outside $\mathscr{T}$, they will be rejected as they lead to zero acceptance probabilities. Two approaches are considered: "local" and "global" moves methods.

6.1. Local moves

Diaconis and Sturmfels (1998) proposed generating a candidate table $n^* \in \mathscr{T}$ using Markov bases of "local moves". A local move $g = \{g(i)\}_{i\in\mathscr{I}}$ is a multi-way array containing integer entries $g(i) \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. A Markov basis MB($\mathscr{T}$) associated with $\mathscr{T}$ allows any two tables n_1, n_2 in $\mathscr{T}$ to be connected by a series of local moves $g^1, g^2, \ldots, g^r$, i.e.,

$$n_1 - n_2 = \sum_{j=1}^{r} g^j.$$

If the chain is currently at $n^{(s)} \in \mathscr{T}$, a new candidate n^* is generated by uniformly choosing a move $g \in \mathrm{MB}(\mathscr{T})$. The candidate $n^* = n^{(s)} + g$ belongs to $\mathscr{T}$ if and only if $n^*(i) \geqslant 0$ for all $i \in \mathscr{I}$. Such a move g is said to be permissible for the current table $n^{(s)}$. If the selected move is not permissible, the chain stays at $n^{(s)}$. Otherwise, the chain moves to $n^{(s+1)} = n^*$ with probability $\min\{1, \rho\}$, where

$$\rho := \frac{p(n^*|\lambda^{(s+1)})}{p(n^{(s)}|\lambda^{(s+1)})} = \prod_{\{i \in \mathscr{I}: n^*(i) \neq n^{(s)}(i)\}} \frac{n^{(s)}(i)!}{n^*(i)!} \exp\{g(i) \log \lambda^{(s+1)}(i)\}.$$

Note that the proposal distribution $q(\cdot, \cdot)$ induced by a Markov basis is symmetric, i.e., $q(n^*, n^{(s)}) = q(n^{(s)}, n^*)$.

The Markov basis required by the "local move" algorithm constitutes both the strength and the weakness of this sampling procedure. The basis has to be generated before the actual simulation begins. This extra step is likely to involve long and tedious computations in algebra systems (such as Macaulay (Bayer and Stillman, 2002) or Cocoa (CoCoATeam, 2004)) following the approach for computing a Markov basis suggested by Diaconis and Sturmfels (1998). They showed that a Markov basis for a set of tables $\mathscr{T}$ can be determined from a Gröbner basis of a well-specified polynomial ideal.

An alternative to this algebraic approach was proposed by Dobra (2003), who gave direct formulæ for dynamically generating a Markov basis in the special case when the information $\mathscr{D}$ available about the original table n consists of a set of marginals that define a decomposable log-linear model. Although the use of Dobra's formulæ require minimal computational effort, they cannot be extended to more general cases (non-decomposable models). Nevertheless, once a Markov basis is computed, the "local move" method can be very fast since generating a new candidate table is done only by additions and subtractions.

Another possible disadvantage of the "local move" method is that the current table and the candidate table can be very similar, and this is critical if $\mathscr{T}$ is large. The Markov bases associated with such large spaces of tables contain moves that change only few table entries and the change can be as small as ± 1. For example, in the case of two-way tables with fixed row and column totals, the moves have counts of zero everywhere except four cells that contain two counts of 1 and two counts of -1. This type of moves are called primitive. Actually, Dobra (2003) proved that primitive moves are the only moves needed to connect tables in the decomposable case mentioned above. Changing only four cells in a high-dimensional contingency table that might contain millions of cells will undoubtedly lead to high dependencies between consecutive sample tables and the corresponding parameter values.

Therefore, with the "local move" method, one cannot control how far the chain "jumps" in the space of feasible tables because the jumps are pre-specified by the Markov basis employed. We present an alternative to the "local move" method, which we call the "global move" method, that allows one to control and adjust the distance between the current and candidate tables.

6.2. Global moves

The idea behind the "global move" method is straightforward, and utilizes compositional sampling: a new table in $\mathscr{T}$ can be generated by sequentially drawing a value for each cell

in the table from the set of possible values for that cell while updating the corresponding upper and lower bounds for the rest of the cells. This strategy is similar to the approach and algorithm of Tebaldi and West (1998a,b) that was modified to a sequential form as utilized by Liu (2001). Moreover, the same idea constitutes the core of the generalized shuttle algorithm (Dobra et al., 2003b; Dobra, 2002) that calculates sharp upper and lower bounds for cells in a contingency tables only by efficiently exploiting the unique structure of the categorical data.

Using the chain rule, we re-write the target distribution $p(n|\mathscr{D}, \lambda) \propto p(n|\lambda)$ from (3) as

$$\begin{aligned} p(n|\lambda) &= p(n(i^1), \ldots, n(i^m)|\lambda) \\ &= p(n(i^1)|\lambda) \prod_{a=2}^{m} p(n(i^a)|n(i^1), \ldots, n(i^{a-1}), \lambda). \end{aligned} \tag{9}$$

The support of $p(n(i^1)|\lambda)$ is a subset of the set of integers $\mathscr{H}_1$ defined by the upper and lower bounds induced by $\mathscr{D}$ on the cell i^1, i.e., $\mathscr{H}_1 := \{L(i^1), L(i^1)+1, \ldots, U(i^1)-1, U(i^1)\}$. Similarly, the support of $p(n(i^a)|n(i^1), \ldots, n(i^{a-1}), \lambda)$ is a subset of the set of integers $\mathscr{H}_a$ defined by the upper and lower bounds induced on the cell i^a by $\mathscr{D}$ and by the additional constraints resulted from fixing the counts in the cells $i^1, \ldots, i^{a-1}$, i.e., $\mathscr{H}_a := \{L(n; i^a), L(n; i^a)+1, \ldots, U(n; i^a)-1, U(n; i^a)\}$, where $L(n; i^a) = \min\{n'(i^a) : n' \in \mathscr{T}, n'(i^1) = n(i^1), \ldots, n'(i^{a-1}) = n(i^{a-1}))\}$ and $U(n; i^a) = \max\{n'(i^a) : n' \in \mathscr{T}, n'(i^1) = n(i^1), \ldots, n'(i^{a-1}) = n(i^{a-1})\}$.

If a Markov basis associated with the set of feasible tables $\mathscr{T}$ can be constructed such that this basis contains only primitive moves, then the supports of $p(n(i^1)|\lambda)$ and $p(n(i^a)|n(i^1), \ldots, n(i^{a-1}), \lambda)$ will be exactly $\mathscr{H}_1$ and $\mathscr{H}_a$, respectively. Otherwise, to the best of our knowledge, there is no theoretical result which shows that the supports of these two distributions should coincide with $\mathscr{H}_1$ and $\mathscr{H}_a$, although examples when this property does not hold have recently been discovered (Sullivant, 2004).

A candidate table $n^* \in \mathscr{T}$ is generated with the "global move" method as follows:

- Draw a value $n^*(i^1)$ from $\mathscr{H}_1$.
- *for* $a = 2, \ldots, m$ *do*
  (1) Calculate $L(n^*; i^a)$ and $U(n^*; i^a)$.
  (2) Draw a value $n^*(i^a)$ from $\mathscr{H}_a$.
  end for

We still need a way to draw cell values from $\mathscr{H}_1, \mathscr{H}_2, \ldots, \mathscr{H}_m$ such that the resulting candidate tables n^* will be neither "too different" nor "too similar" to the current state $n^{(s)} \in \mathscr{T}$. Candidate tables that are "too different" are very likely to be rejected by the Metropolis–Hastings step, while candidates that are "too similar" will be likely to be accepted, but the chain will not advance fast enough in the target space, so inducing very correlated sample path. One approach to balancing these issues uses an annealing idea.

Consider the scaling factors $v_1, v_2, \ldots, v_m$ with $v_a \in (0, 1)$. For each $a \in \{1, 2, \ldots, m\}$, draw a value $n^*(i^a)$ from the proposal distribution with probabilities

$$q_a(n^{(s)}(i^a), n^*(i^a)) \propto v_a^{|n^*(i^a) - n^{(s)}(i^a)|}. \tag{10}$$

Note that the current value $n^{(s)}(i^a)$ of cell i^a does not have to belong to the support $\mathscr{H}_a$ of $q_a(n^{(s)}(i^a), \cdot)$. However, candidate cell values in $\mathscr{H}_a$ that are closer to $n^{(s)}(i^a)$ receive a higher probability to be selected. This probability increases as the scaling factor v_a decreases towards 0. The full unnormalized proposal distribution can then be written as:

$$q(n^{(s)}, n^*) = \prod_{a=1}^{m} q_a(n^{(s)}(i^a), n^*(i^a)), \tag{11}$$

over contingency tables $n^* \in \mathscr{T}$. Any feasible table $n^* \in \mathscr{T}$ has a strictly positive probability of being sampled given any current state $n^{(s)}$, hence the Markov chain obtained by employing the proposal distribution $q(\cdot, \cdot)$ from (11) will be irreducible.

7. Czech autoworkers example

A numerical example focuses on inference about the rate λ_0 associated with the cell $(1, 2, 2, 1, 1, 2)$ of the Czech autoworkers data given the observed marginals $\mathscr{R}_1$ and using the corresponding set of relevant log-linear models in Table 2. Using the "global move" method, we simulated five samples of size 20,000 from the joint distribution $p(n, \lambda|\mathscr{R}_1, \mathscr{A}_j)$ for each $j = 1, 2, \ldots, 8$. The starting points were tables selected at random from $\mathscr{T}_1$. To reduce the correlation between two consecutive draws, we discarded 25 pairs (n, λ) before selecting another pair in the final sample. The burn-in time was 5000 which should be appropriate given the small number of tables in $\mathscr{T}_1$. The scaling factors v_a were taken to be equal to 0.5 for cell $(1, 1, 1, 1, 1, 1)$ and 0.05 for the rest of the cells. Priors on log-linear model parameters are all taken as uniform on a fixed, large range.

Fig. 1 shows the sample mean of the posterior draws for λ_0 calculated across iterations from each of the five starting points under model $\mathscr{A}_8$. The fact that a relatively large number of iterations are needed until convergence is not surprising since $p(n, \lambda|\mathscr{R}_1, \mathscr{A}_8)$ effectively depends only on the sets of interaction terms $\psi_{\{B,C,D,E,F\}}$, $\psi_{\{A,B,C,E,F\}}$, $\psi_{\{A,B,C,D,F\}}$ and $\psi_{\{A,B,C,D,E,F\}}$. The other ψ-terms cancel in the conditional $p(n|\mathscr{D}, \lambda)$, but are still needed when simulating $\lambda(i)$, $i \in \mathscr{I}$. Fig. 1 provides an excellent proof of the mixing properties of our data augmentation procedure.

Fig. 2 gives the marginal posterior distributions for λ_0 as estimated from the resulting samples of size 100,000 under each log-linear model $\mathscr{A}_j$. The posterior distributions under models $\mathscr{A}_1, \ldots, \mathscr{A}_5$ seem to be unimodal. The other three models contain the combination of four-way interactions $\psi_{\{A,B,C,E,F\}}$ and $\psi_{\{A,B,C,D,F\}}$ that seem to induce a second mode and longer right tails in the posterior of λ_0. The estimated value in cell $(1, 2, 2, 1, 1, 2)$ as given by the posterior modes are 2, 1.8, 1.5, 2.3, 1.6, 2, 2.3 and 1.3, respectively. These estimates are consistent with the true value of 1 in this cell. Remember that the possible values for the $(1, 2, 2, 1, 1, 2)$ count given the marginals $\mathscr{R}_1$ are $\{0, 1, \ldots, 10\}$, thus conditioning on the log-linear models $\mathscr{A}_j$ effectively shrinks the estimates to the actual cell count.

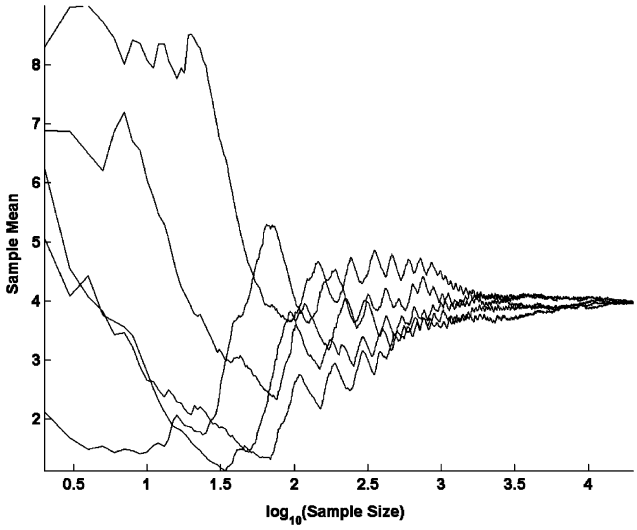

Fig. 1. Convergence of the data augmentation method for the Czech autoworkers data. The *x*-axis represents the iteration number on a $\log_{10}$ scale, while the *y*-axis gives the sample mean of λ_0 from five starting points under model $\mathscr{A}_8$.

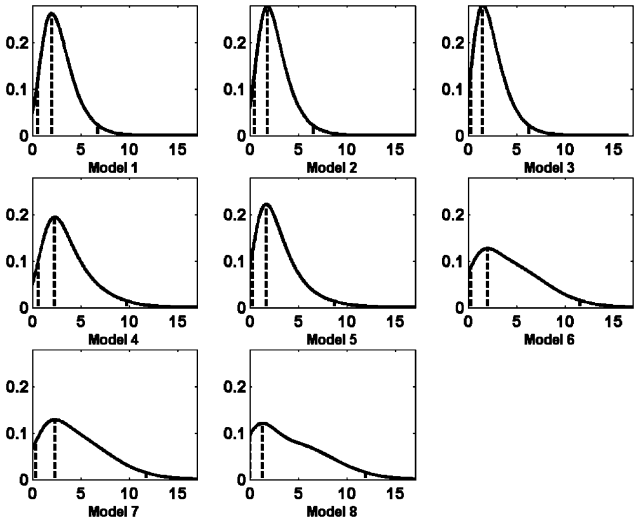

Fig. 2. Approximate posterior distributions for λ_0 under the log-linear models $\mathscr{A}_1, \ldots, \mathscr{A}_8$. The dotted lines represent estimates of the posterior mode and the corresponding 95% confidence intervals.

8. Counting tables

The "global move" method for sampling tables consistent with a set of constraints can be employed to estimate the total number of tables consistent with these constraints. Write $\mathscr{M}(\mathscr{T})$ for the number of tables in the constrained set $\mathscr{T}$. Estimating $\mathscr{M}(\mathscr{T})$ by sampling from the uniform distribution on $\mathscr{T}$, i.e., $p(n|\mathscr{D}) = 1/\mathscr{M}(\mathscr{T})$, is infeasible in any but trivial cases (e.g., two-way tables with fixed one-way marginals). Instead, we simulate tables from the proposal distribution defined in (11) with scaling factors v_a equal (at least eventually) to 1 for all the cells in the table. This particular choice of the scaling factor makes this proposal distribution independent of the table previously sampled, i.e.,

$$q(n) \propto \prod_{a=1}^{m} \frac{1}{U(n; i^a) - L(n; i^a) + 1}.$$

Following Chen et al. (2003) we write:

$$1 = \sum_{n \in \mathscr{T}} \frac{p(n|\mathscr{D})}{q(n)} q(n) = \frac{1}{\mathscr{M}(\mathscr{T})} \sum_{n \in \mathscr{T}} \frac{1}{q(n)} q(n).$$

This suggests an estimate for $\mathscr{M}(\mathscr{T})$ given by

$$\frac{1}{S} \sum_{s=1}^{S} \frac{1}{q(n^{(s)})} = S^{-1} \sum_{s=1}^{S} \prod_{a=1}^{m} [U(n^{(s)}; i^a) - L(n^{(s)}; i^a) + 1], \tag{12}$$

where $n^{(1)}, n^{(2)}, \ldots, n^{(S)}$ are sampled independently from $q(\cdot)$.

As an example, consider two sets of marginals of the Czech autoworkers data. Let $\mathscr{R}_2$ be the 15 four-way marginals of the Czech autoworkers data. The upper and lower bounds induced by $\mathscr{R}_2$ are given in the left-hand panel of Table 3. We generated 100 samples of 5000 tables each from $\mathscr{T}_2$, the set of tables consistent with $\mathscr{R}_2$. Using a modified version of the generalized shuttle algorithm (Dobra et al., 2003b; Dobra, 2002), we determined that the true number of tables in $\mathscr{T}_2$ is 705,884. The mean of our estimates of $\mathscr{M}(\mathscr{T}_2)$ is 703,126, while a 95% confidence interval for $\mathscr{M}(\mathscr{T}_2)$ is 650,000–750,000.

The second example assesses the number of dichotomous six-way tables that have a count of 1 in cell (1, 2, 2, 1, 1, 2) (which identifies the population unique in the Czech autoworkers data n) and that are consistent with the marginals $\mathscr{R}_3 := \{n_{\{B,F\}}, n_{\{A,B,C,E\}}, n_{\{A,D,E\}}\}$. Let $\mathscr{T}_3$ denote this set of tables. The upper and lower bounds associated with $\mathscr{T}_3$ are given in the right-hand panel of Table 3. We generated 1000 samples of 35,000 tables each from $\mathscr{T}_3$. We work on the $\log_{10}$ scale as $\mathscr{M}(\mathscr{T}_3)$ is very large. The mean of $\log_{10}\{\mathscr{M}(\mathscr{T}_3)\}$ is 58% and a 95% confidence interval is 57–59. The true number of tables $\mathscr{M}(\mathscr{T}_3)$ is unknown to us.

9. Concluding comments

Combined parameter inference and missing-data imputation in contingency tables is a very broad-reaching problem. The specific context of inference based on data in terms

Table 3
Bounds for the Czech autoworkers data given the marginals $\mathscr{R}_2$ (left-hand panel) and given the marginals $\mathscr{R}_3$ (right-hand panel)

				B	No		Yes		B	No		Yes	
F	E	D	C	A	No	Yes	No	Yes	A	No	Yes	No	Yes
Neg	<3	<140	No		[27, 58]	[25, 56]	[96, 134]	[44, 82]		[0, 88]	[0, 62]	[0, 224]	[0, 117]
			Yes		[108, 149]	[123, 168]	[0, 22]	[9, 37]		[0, 261]	[0, 246]	[0, 24]	[0, 38]
		$\geqslant 140$	No		[22, 49]	[0, 24]	[60, 96]	[16, 52]		[0, 88]	[0, 62]	[0, 224]	[0, 117]
			Yes		[91, 127]	[45, 85]	[0, 18]	[0, 20]		[0, 261]	[0, 151]	[0, 24]	[0, 38]
	$\geqslant 3$	<140	No		[10, 37]	[17, 44]	[48, 86]	[49, 89]		[0, 58]	[0, 60]	[0, 170]	[0, 148]
			Yes		[30, 68]	[58, 102]	[0, 19]	[0, 25]		[0, 115]	[0, 173]	[0, 20]	[0, 36]
		$\geqslant 140$	No		[13, 37]	[8, 36]	[55, 90]	[38, 76]		[0, 58]	[0, 60]	[0, 170]	[0, 148]
			Yes		[30, 67]	[45, 86]	[0, 19]	[0, 27]		[0, 115]	[0, 173]	[0, 20]	[0, 36]
Pos	<3	<140	No		[0, 15]	[0, 13]	[4, 31]	[0, 23]		[0, 88]	[0, 62]	[0, 125]	[0, 117]
			Yes		[0, 21]	[3, 30]	[0, 10]	[0, 9]		[0, 134]	[0, 134]	[1, 1]	[0, 38]
		$\geqslant 140$	No		[0, 11]	[0, 10]	[0, 24]	[0, 18]		[0, 88]	[0, 62]	[0, 125]	[0, 117]
			Yes		[0, 26]	[2, 30]	[0, 11]	[0, 9]		[0, 134]	[0, 134]	[0, 24]	[0, 38]
	$\geqslant 3$	<140	No		[1, 14]	[0, 9]	[0, 26]	[0, 26]		[0, 58]	[0, 60]	[0, 125]	[0, 125]
			Yes		[0, 19]	[4, 29]	[0, 9]	[0, 9]		[0, 115]	[0, 134]	[0, 20]	[0, 36]
		$\geqslant 140$	No		[0, 9]	[0, 9]	[0, 26]	[0, 22]		[0, 58]	[0, 60]	[0, 125]	[0, 125]
			Yes		[0, 19]	[0, 23]	[0, 9]	[0, 13]		[0, 115]	[0, 134]	[0, 20]	[0, 36]

The bounds for the cell (1, 2, 2, 1, 1, 2) are marked with a box. These bounds were calculated using the generalized shuttle algorithm.

of subsets of marginal counts is pervasive, and is central in problems of data disclosure, dissemination and confidentiality. Issues of prior specification are relevant both for direct analysis, and in connection with questions about potential uses of data released to multiple consumers who will each bring their own priors, or classes of priors, to bear on interpreting the margins. Our work describes some of the basic theoretical and structural issues in the context of log-linear models, and presents a detailed development of MCMC approaches, with examples.

Central to the work reported is the development and implementation of data augmentation under a specific class of structured priors for log-linear model parameters, and the introduction and development of a "global move" simulation approach for imputing missing elements of contingency tables subject to observed margins. Both "local" and "global" move algorithms for sampling tables have their advantages and disadvantages, though we generally prefer the "global move" approach as it is relatively easily set up and implemented. The speed of the above procedure is directly influenced by the number of cells that need to be fixed at a certain value before a full table consistent with the data $\mathscr{D}$ is determined. Every time we assign a value to a cell, we need to update the upper and lower bounds for the rest of the cells in the table. Consequently, the smaller the number of cells we need to fix, the faster the algorithm is. This number is a function of the pattern of constraints induced by the full information $\mathscr{D}$.

One of the requirements of the "global move" approach sampling algorithm is that the bounds defining the values a cell count can take given data $\mathscr{D}$ and given that some other cells have been fixed at a certain value have to be sharp. Gross bounds approximating the corresponding sharp bounds will frequently lead to combinations of cell values that do not correspond to tables in $\mathscr{T}$, and hence the use of gross bounds will significantly decrease the efficiency of this sampling procedure. Unfortunately, computing sharp bounds to determine the admissible values of every cell in the target table could become a serious computational burden in the case of large high-dimensional tables having thousands or possibly millions of cells. Therefore, one very difficult computational problem (the generation of a Markov basis) from the "local move" algorithm is replaced, in the "global move" algorithm, with another very challenging problem—the calculation of sharp integer bounds. Even if sharp bounds are calculated and used at each step, infeasible combinations of cell values might still be generated. Note that the "local move" method also generates candidate tables that are outside $\mathscr{T}$ if a move that is not permissible for the current state of the chain is selected.

A simulation algorithm employing the "global move" method can be started right away without any possibly tedious preliminary computations required in most cases by the "local move" method. There are situations when the generation of a Markov basis could take too long to complete (hence no samples can be drawn with the "local move" method), while the "global move" method might still be applied and samples will still be generated even if these samples might be expensive in term of computing time.

As opposed to the "local move" method that modifies only some small subset of cells affected by the chosen local move, the "global move" method potentially changes the entire table at each step, which leads to sequences of tables with very low correlation from one step to the next, and clearly facilitates an effective exploration of the multidimensional support of the random variables in exam.

Throughout the paper we employed the generalized shuttle algorithm (Dobra et al., 2003b; Dobra, 2002) to compute sharp upper and lower bounds. This is a very flexible algorithm that allows one to specify different types of constraints on the cell entries of a table. These constraints can be, but are not limited to, fixed marginals or cells fixed at a certain value. Moreover, the generalized shuttle algorithm can be modified to exhaustively enumerate the tables consistent with a set of constraints.

We have illustrated how posterior inferences on cell counts can vary based on assumed forms of log-linear models. It may in future be of interest to consider extensions of this work to build analyses across models, utilising model-mixing and model-averaging ideas (Kass and Raftery, 1995; Madigan and York, 1995), and some further developments of the proposed MCMC methods here that extend to reversible-jump methods will be relevant there.

Finally, software for performing the simulations as illustrated is freely available at `http://www.stat.duke.edu/~adobra/sampletable.htm`.

Acknowledgements

We thank Ian Dinwoodie and Stephen Fienberg for comments and useful discussions, and the associate editor and referees for constructive comments on the original version of the paper. Elements of this work formed part of the Stochastic Computation program (2002-03) at the Statistical and Applied Mathematical Sciences Institute, RTP, USA. Partial support was also provided under NSF grant DMS-0102227.

References

Agresti, A., 1992. A survey of exact inference for contingency tables. Statist. Sci. 7, 131–177.

Bayer, D., Stillman, M., 2002. Macaulay: a system for computation in algebraic geometry and commutative algebra. Available at `http://www.math.columbia.edu/~bayer/Macaulay`.

Bishop, Y.M.M., Fienberg, S.E., Holland, P.W., 1977. Discrete Multivariate Analysis: Theory and Practice, MIT Press, Cambridge MA.

Chen, Y., Diaconis, P., Holmes, S.P., Liu, J.S., 2003. Sequential Monte Carlo methods for statistical analysis of tables, iSDS Discussion Paper #03-22, Duke University.

CoCoATeam, 2004. CoCoa: a system for doing Computations in Commutative Algebra. Available at `http://cocoa.dima.unige.it`.

Deming, W.E., Stephan, F.F., 1940. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann. Math. Stat. 11, 427–444.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. J. Roy. Statist. Soc. B 39, 1–38.

Diaconis, P., Efron, B., 1985. Testing for independence in a two-way table: new interpretations of the chi-square statistic. Ann. Statist. 13, 845–874.

Diaconis, P., Sturmfels, B., 1998. Algebraic algorithms for sampling from conditional distributions. Ann. Statist. 26, 363–397.

Dobra, A., 2002. Statistical tools for disclosure limitation in multi-way contingency tables. Ph.D. Thesis, Department of Statistics, Carnegie Mellon University.

Dobra, A., 2003. Markov bases for decomposable graphical models. Bernoulli 9 (6), 1–16.

Dobra, A., Fienberg, S.E., Trottini, M., 2003a. Assessing the risk of disclosure of confidential categorical data (with discussion). In: Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M., West, M. (Eds.), Bayesian Statistics 7. Oxford University Press, Oxford, pp. 125–144.

Dobra, A., Karr, A., Sanil, A., 2003b. Preserving confidentiality of high-dimensional tabulated data: statistical and computational issues. Statist. Comput. 13, 363–370.
Edwards, D.E., Havranek, T., 1985. A fast procedure for model search in multidimensional contingency tables. Biometrika 72, 339–351.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis, 2nd Edition. CRC Press, Boca Raton.
Haberman, S.J., 1974. The Analysis of Frequency Data, University of Chicago Press, Chicago.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Amer. Statist. Assoc. 90, 773–795.
Knuiman, M.W., Speed, T.P., 1988. Incorporating prior information into the analysis of contingency tables. Biometrics 44, 1061–1071.
Lauritzen, S.L., 1996. Graphical Models, Clarendon Press, Oxford.
Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with Missing Data, 2nd Edition. Wiley-Interscience, New York.
Liu, J.S., 2001. Monte Carlo Strategies in Scientific Computing, Springer, New York.
Madigan, D., York, J., 1995. Bayesian graphical models for discrete data. Internat. Statist. Rev. 63, 215–232.
Schafer, J.L., 1997. Analysis of Incomplete Multivariate Data, Chapman & Hall, London.
Smith, P., Forster, J., McDonald, J., 1996. Monte Carlo exact tests for square contingency tables. J. Amer. Statist. Assoc. 2, 309–321.
Sullivant, S., 2004. Small contingency tables with large gaps. Eprint available at http://xxx.lanl.gov/abs/math.OC/0405038.
Sundberg, R., 1975. Some results about decomposable (or markov-type) models for multidimensional contingency tables: distribution of marginals and partitioning of tests. Scand. J. Statist. 2, 71–79.
Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. J. Amer. Statist. Assoc. 82, 528–540.
Tebaldi, C., West, M., 1998a. Bayesian inference on network traffic using link count data (with discussion). J. Amer. Statist. Assoc. 93, 557–576.
Tebaldi, C., West, M., 1998b. Reconstruction of contingency with missing data. ISDS Discussion Paper #98-01, Duke University.
West, M., 1997. Statistical inference for gravity models in transportation flow forecasting. Technical Report #60, National Institute of Statistical Sciences.
Whittaker, J., 1990. Graphical models in applied multivariate statistics, Wiley, New York.