Journal of the Korean Statistical Society 43 (2014) 31–45

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Exact inference in contingency tables via stochastic approximation Monte Carlo

Byoung Cheol Jung [b], Sunha So [c], Sooyoung Cheon [a,*]

[a] Department of Informational Statistics, Korea University, 2511 Sejong-ro, Sejong-city, 339-700, South Korea
[b] Department of Statistics, University of Seoul, Seoul 130-743, South Korea
[c] Risk Model Validation Team, Woori Bank, Seoul 100-792, South Korea

ARTICLE INFO

Article history:
Received 7 January 2013
Accepted 22 June 2013
Available online 12 July 2013

AMS 2000 subject classifications:
62-04
90-08

Keywords:
Complete or incomplete contingency table
Exact inference
Structural zero cells
Importance sampling
Markov chain Monte Carlo
Stochastic approximation Monte Carlo

ABSTRACT

Monte Carlo methods for the exact inference have received much attention recently in complete or incomplete contingency table analysis. However, conventional Markov chain Monte Carlo, such as the Metropolis–Hastings algorithm, and importance sampling methods sometimes generate the poor performance by failing to produce valid tables. In this paper, we apply an adaptive Monte Carlo algorithm, the stochastic approximation Monte Carlo algorithm (SAMC; Liang, Liu, & Carroll, 2007), to the exact test of the goodness-of-fit of the model in complete or incomplete contingency tables containing some structural zero cells. The numerical results are in favor of our method in terms of quality of estimates.

1. Introduction

In the contingency table analysis, the log-linear model has usually been used to test the goodness-of-fit of the model using, in general, the Pearson chi-squared statistic (χ^2) or the Wilks likelihood-ratio statistic (G^2). However, when the expected cell frequencies are small, it is preferred to use exact tests, such as Fisher's exact test, which requires enumerating the set of all tables with the observed margins. Following the literature, the set of all tables with the observed margins is called a reference set, and each table in the reference set is called a valid table. Complete enumeration of the reference set is often infeasible for large tables. McCullagh (1986) and Paul and Deng (2000) made efforts to get more accurate normal approximations with the conditional moments of the likelihood ratio statistic. Their methods produce sometimes accurate p-values, but they are still not satisfactory. Haberman (1988) and Kreiner (1987) argued against the use of large-sample χ^2-approximations for goodness-of-fit tests when tables are large sparse ones, or expected cell counts are small and large.

Monte Carlo methods for the exact inference have received much attention recently in the complete contingency table analysis. Since it is generally difficult for Monte Carlo methods to draw valid samples from the large reference set, Booth and Butler (1999) applied importance sampling to exact tests for general multi-way contingency tables using a rounded normal trial distribution with its moments matching with the target conditional distribution. But their method does not often generate valid tables, especially when the degrees of freedom of the test are high. Chen, Dinwoodie, and Sullivant (2006)

* Corresponding author. Tel.: +82 448601552.
E-mail addresses: bcjung@uos.ac.kr (B.C. Jung), scheon@korea.ac.kr (S. Cheon).

1226-3192/$ – see front matter © 2013 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jkss.2013.06.002

and Dinwoodie and Chen (2011) proposed to use sequential importance sampling with linear programming to generate valid tables. However, solving linear programming could be time-consuming sometimes. Diaconis and Sturmfels (1998) employed the Metropolis–Hastings algorithm (MH; Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) to draw valid tables under the Markov basis from the reference set directly. The Markov basis refers to a finite set of data swaps which allow any two tables in the reference set to be connected and thus guarantee the irreducibility of the Markov chain. However, the computation of Markov basis can be difficult, in even three-way tables (Deloera & Onn, 2006). Caffo and Booth (2001) proposed a Markov chain Monte Carlo (MCMC) method not relying on the Markov basis with a rounded normal candidate to update some randomly selected cells. However, this use of the rounded normal candidate can make certain tables practically inaccessible. Del Moral, Doucet, and Jasra (2006) developed the sequential Monte Carlo (SMC) algorithm based on the multilevel approach of the cross-entropy method (Rubinstein & Kroese, 2007). It is not easy to use a relevant SMC sampler because of many degrees of freedom such as number of steps, type of SMC sequence and type of MCMC moves. Due to the inefficiency of the MH sampler in sample space exploration, the SMC sampler seems only able to draw samples from a small proportion of the sample space, and the resulting p-value estimate is often biased. The SMC sampler may be improved by employing an optimized resampling scheme, an appropriately tempered target distribution, and an optimized number of MCMC steps of each level.

The analysis of incomplete contingency tables with structural zeros has received less attention in the literature of contingency table analysis. Structural zeros arise in situations where it is theoretically impossible for some cells to contain observations. Such cells can occur due to the sampling variability and the relatively very small probabilities of the cell. Such cells can be distinguished from sampling zeros. It is noted that if the probabilities of some cells are regarded as nuisance parameters, these cells can also be treated as structural zero cells. Chen, Diaconis, Holmes, and Liu (2005) studied the problem of sampling zero–one two-way tables with fixed one-way marginals, and Chen, Dinwoodie, and Yoshida (2010) presented algebraic methods for studying connectivity of Markov moves with margin positivity to develop Markov sampling methods when a Markov basis is hard to compute. Sampling two-way tables with structural zeros can be performed using the importance sampling algorithm of Chen (2007) or using the Markov bases developed by Aoki and Takemura (2005). Rapallo (2006) studied Markov bases for incomplete multi-way tables from a theoretical point of view, and Rapallo and Yoshida (2010) examined the necessary and sufficient condition on the set of structural zeros so that the set of basic moves of all 2×2 minors connects all incomplete tables with given positive margins. However, the applicability of all existing methods to higher-dimensional incomplete tables was not discussed. Dobra (2009) proposed a Metropolis–Hastings sampler called the bounds sampling algorithm (BSA) based on computation of lower and upper bounds. BSA requires to compute the bounds, which can be determined by linear programming of the sequential importance sampling method of Chen et al. (2005). However, solving linear programming could be time-consuming sometimes.

In this paper, we consider the analysis of complete or incomplete contingency tables containing structural zero cells. Recently, Cheon, Liang, Chen, and Yu (in press) proposed a method utilizing stochastic approximation Monte Carlo (SAMC; Liang et al., 2007) based importance sampling for approximating exact conditional probabilities in contingency tables. Their method performs well, however, it can apply to only complete tables. This paper proposes a general method using SAMC to draw samples from an enlarged reference set with a known Markov basis for the goodness-of-fit test in complete or incomplete contingency tables with structural zero cells, in particular for the test of no three-way interaction. The contingency tables, which belong to the enlarged reference set but not the original reference set, are called auxiliary tables in this paper. The auxiliary tables provide connections to different parts of the reference set and thus can improve the mixing of the SAMC chain. SAMC makes certain of irreducibility on the enlarged reference set by employing a Markov basis as the proposal. Our method also avoids the requirement for the Markov basis of the original reference set by working on an enlarged reference set. The performance of SAMC has been investigated on real datasets, comparing with existing Monte Carlo methods. The numerical results are in favor of our method in terms of quality of estimates.

The remainder of this paper is organized as follows. Section 2 provides a description of exact tests of contingency tables, reviews briefly the SAMC algorithm, and describes how to apply SAMC to contingency table analysis. In Sections 3 and 4, we apply the proposed method to several examples with complete or incomplete tables. In Section 5, we conclude the paper with a brief discussion.

2. Exact inference in contingency tables via the SAMC algorithm

2.1. Exact conditional inference

We consider a general Poisson log-linear model for a given multi-way $(I_1 \times I_2 \times \cdots \times I_N)$ contingency table $\boldsymbol{X} = \{X_{i_1\cdots i_N} : (i_1,\ldots,i_N) \in (1,\ldots,I_1)\times\cdots\times(1,\ldots,I_N)\}$. The cell counts of $\boldsymbol{X}$ are represented as $\{X_{i_1\cdots i_N} = x_{i_1\cdots i_N}\}$. Let $P(\boldsymbol{x}|\boldsymbol{\mu})$ denote the joint Poisson probability mass function of the data, given by

$$P(\boldsymbol{x}|\boldsymbol{\mu}) = \prod_{i_1=1}^{I_1}\cdots\prod_{i_N=1}^{I_N}\frac{\mu_{i_1\cdots i_N}^{x_{i_1\cdots i_N}}\exp\left(-\mu_{i_1\cdots i_N}\right)}{x_{i_1\cdots i_N}!}, \tag{1}$$

where $\mu_{i_1\cdots i_N}$ denotes the expected frequency of the cell $(i_1, \ldots, i_N)$, and $\boldsymbol{\mu} = (\mu_{i_1\cdots i_N})$. The saturated model of expected cell frequencies in log-linear model is usually specified as

$$\log(\mu_{i_1\ldots i_N}) = \mu + \sum_{l_1=1}^{N} \lambda_{i_{l_1}} + \sum_{l_1\neq l_2} \lambda_{i_{l_1} i_{l_2}} + \sum_{l_1\neq l_2\neq l_3} \lambda_{i_{l_1} i_{l_2} i_{l_3}} + \cdots + \sum_{l_1\neq\cdots\neq l_{N-1}} \lambda_{i_{l_1}\cdots i_{l_{N-1}}} + \lambda_{i_1\cdots i_N}, \tag{2}$$

where $\lambda_{i_{l_1}}$, $\lambda_{i_{l_1} i_{l_2}}$, $\lambda_{i_{l_1} i_{l_2} i_{l_3}}$, $\lambda_{i_{l_1}\cdots i_{l_{N-1}}}$ and $\lambda_{i_1\ldots i_N}$ denote respective one-way, two-way, three-way, $N-1$ and N-way interaction effects, respectively.

For example, when $N = 3$, a multi-way contingency table becomes a three-way $(I \times J \times K)$ contingency table $\boldsymbol{X} = \{X_{ijk} : (i, j, k) \in (1, \ldots, I) \times (1, \ldots, J) \times (1, \ldots, K)\}$. The cell counts of $\boldsymbol{X}$ are represented as $\{X_{ijk} = x_{ijk}\}$. Let $P(\boldsymbol{x}|\boldsymbol{\mu})$ denote the joint Poisson probability mass function of the data, given by

$$P(\boldsymbol{x}|\boldsymbol{\mu}) = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K} \frac{\mu_{ijk}^{x_{ijk}} \exp\left(-\mu_{ijk}\right)}{x_{ijk}!}, \tag{3}$$

where μ_{ijk} denotes the expected frequency of the cell (i, j, k), and $\boldsymbol{\mu} = (\mu_{ijk})$. The saturated model of expected cell frequencies in log-linear model is usually specified as

$$\log(\mu_{ijk}) = \mu + \lambda_i + \lambda_j + \lambda_k + \lambda_{ij} + \lambda_{ik} + \lambda_{jk} + \lambda_{ijk}, \tag{4}$$

where λ_i, λ_j and λ_k denote a row, column and layer effect, respectively; λ_{ij}, λ_{jk} and λ_{ik} denote respective two-way interaction effects; and λ_{ijk} denotes the three-way interaction effect.

From (3) and (4), sufficient statistics are the coefficients of the parameters because the Poisson distribution belongs to the exponential family. The conditional distribution conditioning on the sufficient set S of the model such as the margins is given by

$$P(\boldsymbol{x}|S) \propto \frac{1}{\prod_{i,j,k} x_{ijk}!}, \tag{5}$$

where $\boldsymbol{x}$ denotes a three-way contingency table. An exact test can be constructed using this conditional distribution. The exact tests for contingency tables are discussed in detail in Agresti (1992).

In complete contingency tables, since Cheon et al. (in press) discussed the tests of the mutual or conditional independence, or no-three-way interaction models broadly, we focus on the test of no three-way interaction models, which are also called the uniform association models (Faraway, 2006). Testing the no three-way interaction model versus the saturated model (4) corresponds to testing the null hypothesis $H_0 : \lambda_{ijk} = 0$ for all i, j, k. The sufficient statistics of λ_{ij}, λ_{ik} and λ_{jk} are the two-way interaction margins, x_{ij+}, x_{i+k} and x_{+jk}, respectively.

We also consider the test of a goodness-of-fit in incomplete log-linear models containing structural zero cells, e.g., a quasi-independence model. In a complete three-way table $\boldsymbol{X}$, let a proper subset $\boldsymbol{X}^* = \{X_{ijk} : (i, j, k) : 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq j \leq K\}$ be the set of cells that are not structural zeros, and $R = \{(i, j, k) : X_{ijk} \in \boldsymbol{X}^*\}$. For example, for the quasi-independence model, we define the model for the sub-table $\boldsymbol{X}^*$ by setting $H_0 : \lambda_{ij} = \lambda_{ik} = \lambda_{jk} = \lambda_{ijk} = 0$ for all $(i, j, k) \in R$. Under the quasi-independence model, the incomplete table $\boldsymbol{X}^*$ is restricted to $\{\boldsymbol{X}^* | \sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk} = x_{i++}, \sum_{i=1}^{I}\sum_{k=1}^{K} x_{ijk} = x_{+j+}, \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ijk} = x_{++k},\ (i, j, k) \in R$, and $x_{ijk} = 0$ for $(i, j, k) \notin R\}$. An interpretation of this model and restrictions on the parameters are discussed in detail in Bishop, Fienberg, and Holland (1975).

The exact p-value can be calculated based on the complete enumeration of all the elements in $\boldsymbol{X}$ or $\boldsymbol{X}^*$. However, it is usually computationally infeasible for large sparse tables. In this paper, the exact p-value is computed by approximating the p-value using the general method with stochastic approximation Monte Carlo described in Section 2.2. The conditional p-value is then defined as

$$p_h = P(h \geq h_{\text{obs}}|S) = \frac{\sum_{i,j,k} I(h(x_{ijk}) \geq h_{\text{obs}}) \left(\prod_{i,j,k} x_{ijk}!\right)^{-1}}{\sum_{i,j,k} \left(\prod_{i,j,k} x_{ijk}!\right)^{-1}}, \tag{6}$$

where h is any test statistic for which larger values of h support the alternative hypothesis and $I(\cdot)$ is an indicator function. This provides a Monte Carlo estimate of the exact conditional p-value.

As common candidates of h, there are the Wilks likelihood-ratio statistic (G^2) which is called the deviance of the Poisson log-linear model, $h(\boldsymbol{x}, \hat{\boldsymbol{\mu}}) = -2\sum_{ijk}(x_{ijk}\log(x_{ijk}/\hat{\mu}_{ijk}) - (x_{ijk} - \hat{\mu}_{ijk}))$, the Pearson χ^2-statistic, $h(\boldsymbol{x}, \hat{\boldsymbol{\mu}}) = \sum_{ijk}(x_{ijk} - \hat{\mu}_{ijk})^2/\hat{\mu}_{ijk}$, and the negative log-likelihood statistic, $h(\boldsymbol{x}, \hat{\boldsymbol{\mu}}) = \sum_{ijk}\log(x_{ijk}!)$, where $\hat{\boldsymbol{\mu}}$ denotes the maximum likelihood estimate (MLE) of $\boldsymbol{\mu}$ under the null model. In this paper, for comparison, the negative log-likelihood statistic is used for the test of no three-way interaction effects of complete tables and an example 4.4 of incomplete tables, while for the test of a goodness-of-fit of other incomplete tables, the Wilks likelihood-ratio statistic is used.

Unlike other models in complete tables, there is no explicit form of MLE $\hat{\boldsymbol{\mu}}$ for the no three-way interaction effect models. In these models, $\hat{\boldsymbol{\mu}}$ can be estimated by using iterative algorithms such as Newton–Raphson or iterative proportional fitting methods (Agresti, 2002).

For goodness-of-fit models in incomplete tables, there is also no explicit form of $\hat{\boldsymbol{\mu}}$, which can be estimated by iterative methods with estimating the multiplicative parameters as described by Bishop et al. (1975). For example, for the quasi-independence model in two-way contingency tables, the quasi-independence implies that $m_{ij} = \delta_{ij} a_i b_j,\ i = 1, \ldots, I;\ j = 1, \ldots, J$, where $\delta_{ij} = 1$ for cells $(i, j) \in R$, otherwise 0, where a_i and b_j are positive constants for rows and columns, respectively. Then, MLE $\hat{\boldsymbol{\mu}}$ is estimated by the iteration $(n \geq 1)$ of $a_i^{(n)} = \frac{x_{i+}}{\sum_j \delta_{ij} b_j^{(n-1)}}$ for $i = 1, \ldots, I$ and $b_j^{(n)} = \frac{x_{+j}}{\sum_i \delta_{ij} a_i^{(n)}}$ for $j = 1, \ldots, J$. After the nth iteration, the estimates of m_{ij} are $\hat{m}_{ij}^{(2n)} = \delta_{ij} a_i^{(n)} b_j^{(n)}$. The iteration is continued until obtaining the desired accuracy. The resulting $\hat{m}_{ij}$'s are maximum likelihood estimates $\hat{\boldsymbol{\mu}}$. This method can be extensively applied to the multi-way tables. MLE for the goodness-of-fit models in incomplete tables can be also estimated by using the modified Deming–Stephen iterative proportional fitting procedure (Bishop et al., 1975). In this paper, we used this modified Deming–Stephen method to find the MLE $\hat{\boldsymbol{\mu}}$ in testing of no three-way interaction effects in incomplete tables, while an iterative method with estimating the multiplicative parameters is used in testing of other models.

2.2. Use of stochastic approximation Monte Carlo in exact inference

In this section, we describe how to apply the SAMC algorithm (Liang et al., 2007) to exact tests for complete or incomplete contingency tables. We first give a brief description of SAMC.

Suppose that we are working with the following target distribution,

$$f(\boldsymbol{x}) = \frac{1}{Z}\psi(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{X}, \tag{7}$$

where Z is the normalizing constant, $\mathcal{X}$ is the sample space of $\boldsymbol{x}$, and $\psi(\boldsymbol{x})$ is a non-negative function. For contingency tables, $\mathcal{X}$ may correspond to an enlarged reference set, and $\psi(\boldsymbol{x})$ can be defined as in (15). Furthermore, we suppose that the sample space has been partitioned according to $U(\boldsymbol{x})$ into $m+1$ disjoint subregions: $E_0 = \{\boldsymbol{x} : U(\boldsymbol{x}) \leq u_0\}$, $E_1 = \{\boldsymbol{x} : u_0 < U(\boldsymbol{x}) \leq u_1\}, \ldots, E_{m-1} = \{\boldsymbol{x} : u_{m-2} < U(\boldsymbol{x}) \leq u_{m-1}\}$, and $E_m = \{\boldsymbol{x} : U(\boldsymbol{x}) > u_{m-1}\}$. Let $\theta_i = \log\left(\int_{E_i} \psi(\boldsymbol{x}) d\boldsymbol{x}/\pi_i\right)$ for $i = 0, 1, \ldots, m$. In the contingency table simulations, we set $\psi(\boldsymbol{x}) = \exp\{-U(\boldsymbol{x})\}$.

SAMC seeks to draw samples from each of the subregions with a pre-specified frequency. Let $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_m)$, and let Θ denote the space of $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}^{(t)} = (\theta_0^{(t)}, \ldots, \theta_m^{(t)})$ denote the working estimate of $\boldsymbol{\theta}$ obtained at iteration t. Let $\boldsymbol{x}^{(t+1)}$ denote a sample drawn from a MH kernel $K_{\boldsymbol{\theta}^{(t)}}(\boldsymbol{x}^{(t)}, \cdot)$ with the proposal distribution $q(\boldsymbol{x}^{(t)}, \cdot)$ and the stationary distribution

$$f_{\boldsymbol{\theta}^{(t)}}(\boldsymbol{x}) \propto \sum_{i=0}^{m} \frac{\psi(\boldsymbol{x})}{\exp\left(\theta_i^{(t)}\right)} I(\boldsymbol{x} \in E_i). \tag{8}$$

Let $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_m)$ be a $(m+1)$-vector with $0 < \pi_i < 1$ and $\sum_{i=0}^{m} \pi_i = 1$, which defines desired sampling frequencies for the subregions. Henceforth, $\boldsymbol{\pi}$ is called the desired sampling distribution. Define $H(\boldsymbol{\theta}^{(t)}, \boldsymbol{x}^{(t+1)}) = \boldsymbol{e}^{(t+1)} - \boldsymbol{\pi}$, where $\boldsymbol{e}^{(t+1)} = (e_0^{(t+1)}, \ldots, e_m^{(t+1)})$ and $e_i^{(t+1)} = 1$ if $\boldsymbol{x}^{(t+1)} \in E_i$ and 0 otherwise. Note that the dependence of $H(\cdot, \cdot)$ on $\boldsymbol{\theta}^{(t)}$ is implicit through the sample $\boldsymbol{x}^{(t+1)}$. To have the algorithm complied with the notation of stochastic approximation, $\boldsymbol{\theta}^{(t)}$ is still included in the function $H(\cdot, \cdot)$. In this paper, we assume that Θ is compact for the sequence $\{\boldsymbol{\theta}^{(t)}\}$ to keep in a compact set. Extension of this algorithm to the case that $\Theta = \mathbb{R}^{m+1}$ is trivial with the technique of varying truncations studied in Andrieu, Moulines, and Priouret (2005) and Chen (2002), which ensures, almost surely, that the sequence $\{\boldsymbol{\theta}^{(t)}\}$ can be included in a compact set. In simulations, we can set Θ to a huge set, e.g., $\Theta = \left[-10^{100}, 10^{100}\right]^{m+1}$, which, as a practical matter, is equivalent to setting $\Theta = \mathbb{R}^{m+1}$.

Let $\{\gamma_t\}$ be a positive, non-decreasing sequence satisfying the conditions,

$$\text{(i)} \sum_{t=0}^{\infty} \gamma_t = \infty, \qquad \text{(ii)} \sum_{t=0}^{\infty} \gamma_t^{\delta} < \infty, \tag{9}$$

for some $\delta \in (1, 2)$. In the context of stochastic approximation (Robbins & Monro, 1951), $\{\gamma_t\}_{t \geq 0}$ is called the gain factor sequence.

Let $J(\boldsymbol{x})$ denote the index of the subregion that the sample $\boldsymbol{x}$ belongs to, which takes values in $\{0, 1, \ldots, m\}$. With the above notations, one iteration of SAMC can be described as follows.

The SAMC algorithm:

(a) (Sampling) Simulate a sample $\boldsymbol{x}^{(t+1)}$ by a single MH update with the target distribution as defined in (8).
  (a.1) Generate $\boldsymbol{y}$ according to a proposal distribution $q(\boldsymbol{x}^{(t)}, \boldsymbol{y})$.

(a.2) Calculate the ratio

$$r = \exp\left(\theta^{(t)}_{J(\boldsymbol{x}^{(t)})} - \theta^{(t)}_{J(\boldsymbol{y})}\right) \frac{\psi(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x}^{(t)})}{\psi(\boldsymbol{x}^{(t)})q(\boldsymbol{x}^{(t)}, \boldsymbol{y})}. \tag{10}$$

(a.3) Accept the proposal with probability $\min(1, r)$. If it is accepted, set $\boldsymbol{x}^{(t+1)} = \boldsymbol{y}$; otherwise, set $\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)}$.

(b) (θ-updating) Set

$$\boldsymbol{\theta}^{(t+\frac{1}{2})} = \boldsymbol{\theta}^{(t)} + \gamma_{t+1}H(\boldsymbol{\theta}^{(t)}, \boldsymbol{x}^{(t+1)}). \tag{11}$$

If $\boldsymbol{\theta}^{(t+\frac{1}{2})} \in \Theta$, then set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t+\frac{1}{2})}$; otherwise, find a value of c such that $\boldsymbol{\theta}^{(t+\frac{1}{2})} + c\mathbf{1}_{m+1} \in \Theta$ and set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t+\frac{1}{2})} + c\mathbf{1}_{m+1}$, where $\mathbf{1}_{m+1}$ denotes a constant $(m+1)$-vector of ones.

The self-adjusting mechanism of the SAMC algorithm is obvious: If a proposal is rejected, the weight of the subregion that the current sample belongs to will be adjusted to a larger value, and thus the proposal of jumping out from the current subregion will less likely be rejected in the next iteration. This mechanism warrants that the algorithm will not be trapped by local energy minima. The SAMC algorithm represents a significant advance in simulations of complex systems for which the energy landscape is rugged.

The proposal distribution $q(\boldsymbol{x}, \boldsymbol{y})$ used in the MH updates is required to satisfy the following condition: For every $\boldsymbol{x} \in \mathcal{X}$, there exist $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that

$$|\boldsymbol{x} - \boldsymbol{y}| \leq \epsilon_1 \Longrightarrow q(\boldsymbol{x}, \boldsymbol{y}) \geq \epsilon_2, \tag{12}$$

where $|\boldsymbol{x} - \boldsymbol{y}|$ denotes a certain distance measure between $\boldsymbol{x}$ and $\boldsymbol{y}$. This is a natural condition in study of MCMC theory (Roberts & Tweedie, 1996). In practice, this kind of proposals can be easily designed for both discrete and continuum systems as discussed in Liang et al. (2007).

SAMC falls into the category of varying truncation stochastic approximation algorithms (Andrieu et al., 2005; Chen, 2002). Following Liang et al. (2007), we have the following convergence result: Under the conditions (9) and (12), for all non-empty subregions,

$$\theta^{(t)}_i \to C + \log\left(\int_{E_i} \psi(\boldsymbol{x})d\boldsymbol{x}\right) - \log(\pi_i + \upsilon_0), \tag{13}$$

as $t \to \infty$, where $\upsilon_0 = \sum_{j\in\{i:E_i=\emptyset\}} \pi_j/(m + 1 - m_0)$, $m_0 = \#\{i : E_i = \emptyset\}$ is the number of empty subregions, and $C = -\log\left(\int_{E_m} \psi(\boldsymbol{x})d\boldsymbol{x}\right) + \log(\pi_m + \upsilon_0)$.

Let $\widehat{\pi}^{(t)}_i = P(\boldsymbol{x}^{(t)} \in E_i)$ be the probability of sampling from the subregion E_i at iteration t. Eq. (13) implies that as $t \to \infty$, $\widehat{\pi}^{(t)}_i$ will converge to $\pi_i + \upsilon_0$ if $E_i \neq \emptyset$ and 0 otherwise. With an appropriate specification of $\boldsymbol{\pi}$, sampling can be biased to some subregions that are of interest to the user.

Let $(\boldsymbol{x}^{(1)}, \boldsymbol{\theta}^{(1)}), \ldots, (\boldsymbol{x}^{(N)}, \boldsymbol{\theta}^{(N)})$ denote the samples drawn by SAMC in a run. Liang (2009) showed that SAMC is actually a dynamic importance sampler and for any integrable function $\rho(\boldsymbol{x})$,

$$\frac{\sum_{t=1}^{N} \exp\left(\theta^{(t)}_{J(\boldsymbol{x}_t)}\right) \rho(\boldsymbol{x}_t)}{\sum_{t=1}^{N} \exp\left(\theta^{(t)}_{J(\boldsymbol{x}_t)}\right)} \longrightarrow E_f\rho(\boldsymbol{x}), \quad \text{a.s.}, \tag{14}$$

as $N \to \infty$, where $E_f\rho(\boldsymbol{x})$ denotes the expectation of $\rho(\boldsymbol{x})$ with respect to the target distribution $f(\boldsymbol{x})$ given in (7).

Now, we describe how to apply SAMC to exact tests for complete or incomplete contingency tables, and then discuss some practical issues on SAMC implementation.

Let S be the space of constrained tables of sufficient statistics, and T be the space of structural zero cells, which is $T(y) = 1$ if $y \neq 0$ and $y \in \boldsymbol{X} \setminus \boldsymbol{X}^*$, otherwise 0. Instead of sampling directly from (5), we propose to sample from a target distribution $g(\boldsymbol{x}|S)$, which is defined by

$$g(\boldsymbol{x}|S) \propto \frac{1}{\prod_{i,j,k}(x_{ijk} \vee 0)!} \exp\left\{-\sum_{i,j,k}(x_{ijk} \wedge 0)^2 - [S(\boldsymbol{x}) - S(\boldsymbol{x}^0)]^t[S(\boldsymbol{x}) - S(\boldsymbol{x}^0)] - \sum_{i,j,k} T(x_{ijk})\right\}, \tag{15}$$

where $x_{ijk} \vee 0 = \max(x_{ijk}, 0)$, $x_{ijk} \wedge 0 = \min(x_{ijk}, 0)$ and $\boldsymbol{x}^0$ is an observed dataset. The function partitioning the sample space of $\boldsymbol{x}$ is defined by

$$\begin{aligned} U(\boldsymbol{x}) &= \sum_{i,j,k}(x_{ijk} \wedge 0)^2 + [S(\boldsymbol{x}) - S(\boldsymbol{x}^0)]^t[S(\boldsymbol{x}) - S(\boldsymbol{x}^0)] + \sum_{i,j,k} T(x_{ijk}) \\ &\triangleq U_1(\boldsymbol{x}) + U_2(\boldsymbol{x}) + U_3(\boldsymbol{x}), \end{aligned} \tag{16}$$

where $U_1(\boldsymbol{x})$, $U_2(\boldsymbol{x})$ and $U_3(\boldsymbol{x})$ penalize the tables with negative entries, not satisfying the fixed margin for sufficient statistics, and violating structural zero cells, respectively. This paper considers $\Omega = \{\boldsymbol{x} : U(\boldsymbol{x}) = 0\}$ as the valid table. The convergence theory of SAMC implies that the proportion of valid tables will converge to π_0 as the number of iterations goes to infinity. The sample space of (15) is restricted to the following compact set $\mathcal{X}$ with the minimal sufficient statistics,

$$\mathcal{X} = \{\boldsymbol{x} : S(\boldsymbol{x}) = S(\boldsymbol{x}^0), U_1(\boldsymbol{x}) \le 10^{10}, U_2(\boldsymbol{x}) \le 10^{10}, U_3(\boldsymbol{x}) \le 10^{10},\ x_{ijk} \in \mathbb{Z}\} \tag{17}$$

where $\mathbb{Z}$ denotes the set of integers. For example, for the test of no three-way interaction effects, $\mathcal{X} = \{\boldsymbol{x} : x_{ij+} = x^0_{ij+}, x_{i+k} = x^0_{i+k}, x_{+jk} = x^0_{+jk}, U_1(\boldsymbol{x}) \le 10^{10}, U_2(\boldsymbol{x}) \le 10^{10}, U_3(\boldsymbol{x}) \le 10^{10}, x_{ijk} \in \mathbb{Z}\}$. If a Markov basis is available, $U_2(\boldsymbol{x})$ can be set to a small maximum number such as 2. In the case that a Markov basis is unavailable, e.g., a no three-way interaction test, the maximum of $U_2(\boldsymbol{x})$ should be large such that the resulting Markov chain is irreducible on Ω (Cheon et al., in press). The conditional inference of SAMC given these marginals can be made based on the samples drawn from $\mathcal{X}$. When $U_1(\boldsymbol{x}) = U_2(\boldsymbol{x}) = U_3(\boldsymbol{x}) = 0$, SAMC will be reduced to MH. When $U_3(\boldsymbol{x}) = 0$, the sample space of (15) will be reduced to that of complete tables.

For the SAMC sampler to collect valid tables, the sample space is partitioned according to (16) as follows: $E_0 = \{\boldsymbol{x} : U(\boldsymbol{x}) = 0, \boldsymbol{x} \in \mathcal{X}\}$, $E_1 = \{\boldsymbol{x} : U(\boldsymbol{x}) = 1, \boldsymbol{x} \in \mathcal{X}\}$, $E_2 = \{\boldsymbol{x} : U(\boldsymbol{x}) = 2, \boldsymbol{x} \in \mathcal{X}\}, \ldots, E_{m-1} = \{\boldsymbol{x} : U(\boldsymbol{x}) = m-1, \boldsymbol{x} \in \mathcal{X}\}$, $E_m = \{\boldsymbol{x} : U(\boldsymbol{x}) \ge m, \boldsymbol{x} \in \mathcal{X}\}$, where m is a user-specified number. It is easy to see that the samples in E_0 are those we want. The reference set of the valid tables can be represented as $\Omega = \{\boldsymbol{x} : U(\boldsymbol{x}) = 0\} \subset \mathcal{X}$, and thus $\mathcal{X}$ is called an enlarged set. The tables included in an enlarged set $\mathcal{X}$ except Ω provide connections to different parts of Ω in simulation and thus improve the mixing rate of the Markov chain. The importance sampling approach can then be used for making the conditional inference based on the samples in $\mathcal{X}$.

Consider $I \times J \times K$ tables for the data swap proposal. For testing no three-way interaction, we propose the following data swap for a SAMC run to draw samples from (15):

1. Randomly draw i_1 and i_2 from the set $\{1, 2, \ldots, I\}$ without replacement, draw j_1 and j_2 from the set $\{1, 2, \ldots, J\}$ without replacement, and draw k_1 and k_2 from the set $\{1, 2, \ldots, K\}$ without replacement.
2. Set $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ denotes an $I \times J \times K$ table, which has all its elements $\delta_{ijk} = 0$, apart from $\delta_{i_1j_1k_1} = \delta_{i_1j_2k_2} = \delta_{i_2j_1k_2} = \delta_{i_2j_2k_1} = +1$ and $\delta_{i_1j_1k_2} = \delta_{i_1j_2k_1} = \delta_{i_2j_1k_1} = \delta_{i_2j_2k_2} = -1$ with probability 0.5, and $\delta_{i_1j_1k_1} = \delta_{i_1j_2k_2} = \delta_{i_2j_1k_2} = \delta_{i_2j_2k_1} = -1$ and $\delta_{i_1j_1k_2} = \delta_{i_1j_2k_1} = \delta_{i_2j_1k_1} = \delta_{i_2j_2k_2} = +1$ with probability 0.5.

For two-way contingency tables, the above proposal can be simply modified by ignoring the coordinate of layer. This proposal can be easily extended to four or higher-way contingency tables, such as randomly choosing two coordinates in each dimension to vary at each iteration while leaving all others to be fixed.

For testing all effects such as mutual independence and conditional independence, except for no three-way interaction, we use the following data swap proposed by Cheon et al. (in press):

1. Randomly draw one element from the set $\{i, j, k\}$ to determine which is to be fixed, where i, j, and k represent row, column and layer, respectively. Say, i is drawn.
2. Randomly draw a row i_1 from the set $\{1, 2, \ldots, I\}$, draw two columns j_1 and $j_2 (j_1 < j_2)$ from the set $\{1, 2, \ldots, J\}$ without replacement, and draw two layers k_1 and $k_2 (k_1 < k_2)$ from the set $\{1, 2, \ldots, K\}$ without replacement.
3. Set $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ denotes an $I \times J \times K$ table, which has all its elements $\delta_{ijk} = 0$, apart from $\delta_{i_1j_1k_1} = \delta_{i_1j_2k_2} = +1$ and $\delta_{i_1j_1k_2} = \delta_{i_1j_2k_1} = -1$ with probability 0.5, and $\delta_{i_1j_1k_1} = \delta_{i_1j_2k_2} = -1$ and $\delta_{i_1j_1k_2} = \delta_{i_1j_2k_1} = +1$ with probability 0.5.

The above two data swaps are the only moves that need to be included in the Markov basis and the Markov chain induced by these data swaps can be irreducible on $\mathcal{X}$ (Bunea & Besag, 2000; Dobra, 2003). We can use the above both data swaps to draw samples for testing. However, as described in Bunea and Besag (2000), our proposed data swap is more relevant to test the no three-way interaction effects.

The above settings ensure that a pre-specified proportion of valid tables in SAMC can be achieved in simulations. Controlling the sampling proportion of valid tables to a specified value has been beyond the capability of the existing importance sampling and MCMC methods. Given an appropriate choice of the gain factor sequence $\{\gamma_t\}$, the convergence (13) holds for the SAMC simulations.

Let $(\boldsymbol{y}^{(1)}, \boldsymbol{v}^{(1)}), \ldots, (\boldsymbol{y}^{(n)}, \boldsymbol{v}^{(n)})$ denote the samples drawn from $E_0 = \Omega$, which form a subset of $(\boldsymbol{x}^{(1)}, \boldsymbol{\theta}^{(1)}), \ldots, (\boldsymbol{x}^{(N)}, \boldsymbol{\theta}^{(N)})$. Then the conditional p-value in (6) can be estimated by

$$\hat{p_h} = \frac{\sum_{t=1}^{n} I(h(\boldsymbol{y}^{(t)}) \ge h_{\text{obs}}) \exp\left(\nu^{(t)}_{J(\boldsymbol{y}^{(t)})}\right)}{\sum_{t=1}^{n} \exp\left(\nu^{(t)}_{J(\boldsymbol{y}^{(t)})}\right)}, \tag{18}$$

where $\boldsymbol{y}^{(t)} = \left(y^{(t)}_{ijk}\right)$ in terms of entries and $I(\cdot)$ is an indicator function. It follows from (14), the standard theory of importance sampling, that $\hat{p_h} \to p_h$ almost surely as $n \to \infty$.

In the contingency table analysis, we consider several issues for an effective implementation of SAMC.

- *Choice of the desired sampling distribution* $\boldsymbol{\pi}$: According to our aim to draw samples from $E_0 = \Omega$, we choose $\boldsymbol{\pi}$ to bias sampling to the reference set. In this paper, we use $\pi_i \propto \frac{1}{i+1}$, $i = 0, \ldots, m$. In all simulations, we set $m = 20$ to have a large sample space; i.e., $(\pi_0, \pi_1, \ldots, \pi_{20}) = (0.274, 0.137, 0.091, 0.069, 0.055, 0.046, 0.039, 0.034, 0.031, 0.027, 0.025, 0.023, 0.021, 0.020, 0.018, 0.017, 0.016, 0.015, 0.014, 0.014, 0.013)$.
- *Choice of the gain factor sequence and the total number of iterations*: To meet condition (9), we suggest

$$\gamma_t = \left(\frac{T_0}{\max(T_0, t)}\right)^{\eta}, \quad t = 0, 1, 2, \ldots, \tag{19}$$

for pre-specified values of $T_0 > 1$ and $\eta \in (0.5, 1]$. A large value of T_0 will allow the sampler to reach all subregions rapidly, even for a large system. The appropriateness of the choice of T_0 and a number of iterations can be diagnosed by checking the convergence of multiple runs (starting with different points) via an examination for the variation of $\widehat{\boldsymbol{\theta}}$ or $\widehat{\boldsymbol{\pi}}$, where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\pi}}$ denote the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$, respectively, obtained at the end of a run. If the simulation is diagnosed as unconverged, SAMC should be re-run with a larger value of T_0, a larger number of iterations, or both. In this paper, we set $\eta = 1.0$ in all simulations.

3. Examples for testing of no three-way interaction effects in complete contingency tables

In this section, we illustrate the performance of SAMC on three examples for the exact test of no three-way interaction effects in complete tables. For the SAMC sampler, we set $U_3(\boldsymbol{x}) = 0$. We compared SAMC with DaS (MH method by Diaconis & Sturmfels, 1998), BaB (importance sampling method by Booth & Butler, 1999), CaB (MCMC method by Caffo & Booth, 2001), and SMC (sequential Monte Carlo sampling method by Del Moral et al., 2006). The BaB and CaB methods have been implemented in the R package *exactLoglinTest*, where, instead of a rounded normal, a rounded student t-variate is used to generate candidates to update randomly selected cells. As stated in Caffo and Booth (2001), the use of a rounded student t-variate generally improves the performance of the two methods. First, BaB and CaB were applied to these examples a couple of times, but failed to produce any valid tables in all runs. Each run consisted of 5.0×10^6 iterations, but cost much longer CPU time than that of SAMC. Since no valid tables were generated, they reported 0 as the proportion of valid tables and thus failed to get p-values. Hence, SAMC was compared with DaS and SMC.

The SMC method is also applied to these examples. Since SMC has many degrees of freedom such as the type of SMC sequence and type of MCMC moves, it is not easy to find a good SMC sampler, and thus SMC has a difficulty in computation. Although there are much better resampling methods, we used a multinomial resampling method with same target distribution $g(\boldsymbol{x}|S)$ in (15). SMC was run based on the partition of the sample space we made for SAMC. Let $\mathcal{X}_k = \bigcup_{i=0}^{k} E_i$, $k = 0, 1, \ldots, m$, denote a sequence of nested sample spaces, and let $g_k(\boldsymbol{x}|S) = g(\boldsymbol{x}|S)I(\boldsymbol{x} \in \mathcal{X}_k)$ denote the trial distribution defined on $\mathcal{X}_k$, where $E_1, \ldots, E_m$ are as specified in SAMC, and $I(\cdot)$ is an indicator function. Therefore, $\mathcal{X}_k = \mathcal{X}$ and $\mathcal{X}_0 = E_0$ correspond to the reference set of the contingency table. The sampling procedure of SMC is discussed in detail in Del Moral et al. (2006). The SMC sampler is essentially a multi-start MCMC algorithm, which its sample space gradually shrunk as the level moves on. Since the number of levels cannot be very large and the chains performed at each level cannot be very long, the success of SMC depends largely on the performance of the construction of population whether the Markov chain has converged on $\mathcal{X}$. For the SMC sampler, we set $m = 20$ and a population size by $1.0 \times 10^5 (= N_p)$. For short MH chains in each sample, we also set three iterations for cross-cultural study, and six iterations for livestock breeds study. Under the above settings, for cross-cultural study, SMC had a total of 6.5×10^6 MH moves in each run where the first 5×10^5 iterations were discarded for the burn-in process, about 20% more than that done by SAMC. For livestock breeds study, SMC had a total of 1.25×10^7 MH moves in each run, about 20% more than that done by SAMC. Obviously, these settings improve the performance of SMC. SMC was run for 10 times for each example. We tried other settings for SMC, however, we found slightly worse results.

For our comparison, the DaS method was also run for 10 times independently, and each run consisted of same iterations with that done by SAMC.

3.1. Murdock–White's cross-cultural study

This example concerns the test of no three-way interaction for a $2 \times 2 \times 2$ contingency table. The data are shown in Table 9 of Appendix, which was obtained from Murdock and White's (1969) cross-cultural sample. White, Pesner, and Reitz (1983) suggested a method of modification of Fisher's exact test for $2 \times 2 \times 2$ contingency table for a test of the null hypothesis of no three-way statistical interaction among variables, controlling for the two-way or first-order correlations. They found $p_h = 0.0003$ as the exact p-value; i.e., when the worldwide association between patrilineality and bridewealth is broken down by region, there are significant differences between societies in the insular Pacific region and those outside of this region. Within this region, in fact, the direction of the relationship is reversed. This test uses a truncated hypergeometric distribution, limited by the bivariate marginal totals of the variables. The degree of freedom for this problem is small; i.e., it is only 1. The negative log-likelihood ratio statistics for this test is 496.1429.

Table 1
Comparison of the desired sampling frequencies ($\pi_i,\ i = 0, 1, 5, 13, 20$) with the realized sampling frequencies for the cross-cultural study.

Subregion (i)	π_i	Relative frequency (%)
0	0.3764	0.3744
1	0.1477	0.1482
5	0.1151	0.1155
13	0.2392	0.2398
20	0.1216	0.1221

Table 2
Comparison of SAMC with the existing DaS and SMC methods and result of White et al. (1983) for the cross-cultural study. $\hat{p}_h$: estimate of p_h; RMSE: root of mean squared errors of $\hat{p}_h$, which was calculated based on 10 independent runs; p-value*: p-value of t-test for the hypotheses H_0: $p_h = 0.0003$ versus H_1: $p_h \neq 0.0003$; Proportion: proportion of valid tables and its standard deviation (given in parentheses), which were calculated based on 10 independent runs; CPU (m): CPU time (in minutes) cost by a single run on a 1.70 GHz personal computer.

Method	$\hat{p}_h$	RMSE	p-value*	Proportion	CPU (m)
White et al.	3.00×10^{-4}				
DaS	3.20×10^{-4}	6.85×10^{-10}	5.89×10^{-3}	1.00 (0)	1.25
SMC	2.42×10^{-3}	5.07×10^{-6}	8.88×10^{-6}	–	1.51
SAMC	3.08×10^{-4}	2.98×10^{-10}	1.61×10^{-1}	0.374 (4.77×10^{-4})	1.24

SAMC was first applied to this example. For the SAMC sampler, we set $m = 20$ and $\mathcal{X} = \bigcup_{i=0}^{20} E_i$, where $E_0 = \{x : U(x) = 0\}, E_1 = \{x : U(x) = 1\}, \ldots, E_{19} = \{x : U(x) = 19\}$ and $E_{20} = \{x : U(x) \geq 20\}$. SAMC was run for 10 times. Each run consisted of 5.5×10^6 iterations with $T_0 = 2000$, where the first 5×10^5 iterations were discarded for the burn-in process and the remaining iterations were used for inference. Since the empty subregions exist in this example, its desired sampling distribution is different from the other examples; i.e., desired sampling distribution is $(\pi_0, \pi_1, \pi_5, \pi_{13}, \pi_{20}) = (0.376, 0.239, 0.148, 0.122, 0.115)$. Thus the proportion of valid tables is about 37.6%. Table 1 shows the relative sampling frequency of each of the subregions realized in a run, and indicates that each subregion has been bias sampled to low energy regions approximately for each run. This result shows that the SAMC simulation has converged, and our choices of T_0 and the total number of iterations are appropriate for this example.

The numerical results are summarized in Table 2. For comparison, DaS and SMC were also applied to this example, and each was run for ten times. Under above settings, SMC took a little more CPU cost than SAMC and DaS in each run. The comparison indicates that for this example, SAMC works even slightly better than the DaS method, while they both outperform SMC. Given its importance sampling nature, SAMC should be less efficient than DaS at each iteration, but it still outperforms DaS. This implies that sampling from an enlarged set improves the mixing rate of the Markov chain. For this example, SAMC also outperforms SMC. The roots of mean squared errors of p-value estimates indicate that for this example, SAMC can be more efficient than DaS and SMC, respectively. We note that for this example, the p-value estimate of SMC is biased upward and has a high variability compared to the estimates of SAMC and DaS. The t-test indicates that at a significance level of 0.01 the SAMC estimate is significantly close to the exact value, 0.0003, found by White et al. (1983) while the SMC estimate is much larger than 0.0003. Since the MH sampler used in SMC generates the inefficiency in sample space exploration, the SMC sampler has a possibility to draw samples from a small proportion of the sample space, and the resulting estimate is often biased.

3.2. Livestock breeds study

In this section, we consider data on frequency of livestock breeds studied in Dinwoodie and Chen (2011) and Hall and Ruane (1993). The data shown in Table 10 of Appendix are classified in three ways by region, presence (rare, extinct) and type of animal which make up a $7 \times 2 \times 7$ table. Hall and Ruane (1993) classified in three ways by type of animal, presence (common, rare, extinct) and region. Dinwoodie and Chen (2011) removed the common level to focus on rare and extinct combinations to test how well the model of no 3-way interaction, sometimes called all 2-way interactions, fits the data. For this example, Dinwoodie and Chen (2011) employed a sequential importance sampling (SIS) method with linear programming to generate valid tables. They found that the Markov basis for MCMC was not found after 24 h of running time, and estimated a p-value of 0.012 with standard error 0.005, using SIS with 100% valid tables and normal proposal based on the negative log-likelihood ratio statistic. Our interest is also to test of no three-way interaction among three factors. The degrees of freedom of this model is 36. The negative log-likelihood ratio statistics for this test is 3151.5457.

For the SAMC sampler, we set $m = 20$ and $T_0 = 10\,000$. The desired sampling distribution is $(\pi_0, \pi_1, \ldots, \pi_{20}) = (0.274, 0.137, \ldots, 0.013)$; that is, the proportion of valid tables is about 27.4%. SAMC was run for 10 times with 1.05×10^7 iterations, where 5×10^5 iterations were discarded for the burn-in process. After running SAMC for ten times, we found that all realized sampling frequencies are very similar to the desired sampling distribution; i.e., it shows the convergence of the SAMC simulation.

Table 3
Comparison of SAMC with SIS, DaS and SMC methods for the livestock breeds study. $\hat{p}_h$: estimate of p_h by averaging over 10 runs; SD: standard deviation of $\hat{p}_h$; Proportion: proportion of valid tables and its standard deviation (given in parentheses), which were calculated based on 10 independent runs; CPU (m): CPU time (in minutes) cost by a single run on a 1.70 GHz personal computer.

Method	$\hat{p}_h$	SD	Proportion	CPU (m)
SIS	0.0120	5.00×10^{-3}		
DaS	0.0089	4.71×10^{-4}	1.00 (0)	15.02
SMC	0.0011	7.19×10^{-4}	–	16.29
SAMC	0.0102	3.55×10^{-4}	0.274 (9.60×10^{-5})	13.18

Table 4
Comparison of SAMC with SIS and Cheon et al. (in press) methods for the Whittaker survey study. $\hat{p}_h$: estimate of p_h by averaging over 8 runs; SD: standard deviation of $\hat{p}_h$; CPU (m): CPU time (in minutes) cost by a single run on a 2.90 GHz personal computer.

Method	$\hat{p}_h$	SD	CPU (m)
SIS	0.186	0.0410	52.50[a]
Cheon et al.	0.208	0.0056	4.80[b]
SAMC	0.203	0.0025	37.52

[a] The CPU time reported in Dinwoodie and Chen (2011).
[b] The CPU time reported in Cheon et al. (in press).

Table 3 summarizes the numerical results of SAMC, DaS, SMC and SIS for this example. Since the p-value is highly significant, there is a significant association among region, presence and type of animal. The comparison indicates that SAMC has the best performance among all methods in terms of the accuracy of p-value.

As aforementioned, SMC is essentially a multi-start MCMC algorithm and its success depends largely on the ability of the MH sampler to draw samples from different parts of the sample space $\mathcal{X}_m$. Thus it is not easy to find a good SMC sampler. In this example, $\mathcal{X}_m$ is huge and thus it seems that the MH sampler cannot converge within the number of iterations in each step. Therefore, due to this inefficiency of the MH sampler, the p-value estimate of SMC is biased downward compared to the estimates of SAMC, DaS and SIS. When the sample space $\mathcal{X}_m$ is much larger, the performance of SMC becomes worse. However, our method can work well for this example. SAMC is a single chain method which can traverse over the whole sample space very quickly and thus generate enough valid tables in each run because of its self-adjusting mechanism.

Regarding the CPU time, SMC can cost more CPU time than SAMC and DaS in each run. When a population size, N_p, increases, SMC takes much more CPU cost. In the test of no three-way interaction, there is no explicit form of $\hat{\boldsymbol{\mu}}$. This can be estimated by iterative methods and thus calculation of the likelihood-ratio statistic is time consuming. Since SMC is a multiple-chain method, it needs to calculate the test statistic for each of the N_p samples obtained at the level $\mathcal{X}_0$. However, SAMC is a single chain method that the test statistic is calculated only when a new sample is accepted in E_0. Hence, SMC can cost more CPU time than SAMC in each run.

3.3. Whittaker's survey study

Table 11 shows the counts of 8-way binary data relating women's economic activity and husband's unemployment from a survey of households in Rochdale reported in Whittaker (1990) and analyzed by Cheon et al. (in press) and Dinwoodie and Chen (2011). Whittaker (1990) concerns a survey of 665 households with each response classified in one of 2^8 ways based on 8 "no or yes" responses to economic and employment questions. The data are given in lexicographic order and each cell is coded by a 8-digit binary number such as 00000000, 00000001, 00000010, etc. where 0 codes for "no" on a particular question. Cheon et al. (in press) and Dinwoodie and Chen (2011) tested how well the model of no 3-way interaction fits the data. Since the data contain a large number of zeros, the χ^2-asymptotic methods work badly; i.e., the p-value is greater than 0.9999. Dinwoodie and Chen (2011) noticed that the Markov basis is quite too difficult to compute it in 4*ti*2 (4ti2 team, 2006) for the exact analysis. Since the degree of freedom, 219, of this test is very large, it is a greater challenge to the existing Monte Carlo methods. Like in livestock breeds study, Dinwoodie and Chen (2011) employed a SIS method to test the no 3-way interaction effect and estimated a p-value of 0.186 with standard error 0.041, based on the negative log-likelihood ratio statistic. We are also interested in testing of all two-way interactions among eight factors. The negative log-likelihood ratio statistics for this test is 1278.9328.

Cheon et al. (in press) noticed that allowing negative entries for this example will increase the size of $\mathcal{X}$ drastically on the sample space of the SAMC sampler and thus setting $U_1(\boldsymbol{x}) > 0$ hinders the convergence of the Markov chain. We set $U_1(\boldsymbol{x}) = U_2(\boldsymbol{x}) = 0$ which is similar to the DaS method for this test. SAMC was run for 8 times, like in Cheon et al. (in press) and Dinwoodie and Chen (2011). Each run consisted of 5.05×10^7 iterations with $T_0 = 50$ where the first 5×10^5 iterations were discarded for the burn-in process. The numerical results were summarized in Table 4.

Since BaB, CaB and SMC failed for this example (Cheon et al., in press), we compared our method with only SIS and the method of Cheon et al. (in press). Although χ^2-asymptotic p-value is greater than 0.9999, the estimated exact p-value was much smaller, 0.203. The comparison indicates that the proposed proposal is more efficient than that used by Cheon et al.

Table 5
Comparison of SAMC with the existing DaS and SMC methods and result of Aoki and Takemura (2005) for the jury study. $\hat{p}_h$: estimate of p_h by averaging over 10 runs; SD: standard deviation of $\hat{p}_h$; Proportion: proportion of valid tables and its standard deviation (given in parentheses), which were calculated based on 10 independent runs; CPU (m): CPU time (in minutes) cost by a single run on a 1.70 GHz personal computer.

Method	$\hat{p}_h$	SD	Proportion	CPU (m)
χ^2-asymptotic	0.0268			
Aoki and Takemura	0.0444	5.20×10^{-4}		
DaS	0.0452	3.03×10^{-4}	1.00 (0)	4.48
SMC	0.0132	5.95×10^{-3}	–	4.23
SAMC	0.0445	2.79×10^{-4}	0.274 (2.26×10^{-5})	3.22

(in press). Note that the method of Cheon et al. (in press) is efficient for the independent test. In summary, SAMC outperforms the existing methods for this example with much more accurate estimates.

4. Examples for the goodness-of-fit test of log-linear models in incomplete contingency tables

This section shows the performance of SAMC on four examples for the goodness-of-fit test of log-linear models in incomplete tables with structural zero cells. For the goodness-of-fit models in incomplete tables, there is no explicit form of MLE $\hat{\boldsymbol{\mu}}$. This can be estimated by an iterative method of estimating the multiplicative parameters as described by Bishop et al. (1975). Thus the sampler in incomplete tables costs longer CPU time to get estimates than it does in complete tables for the test of mutual independence in particular. To ensure the irreducibility of the resulting Markov chain on the reference set, we set $\mathcal{X}$ the compact set as described in Section 2. Thus, setting the maximum of each $U_1(\boldsymbol{x})$, $U_2(\boldsymbol{x})$ or $U_3(\boldsymbol{x})$ as 10^{10} can lead to a much larger sample space of $\mathcal{X}$ and affect the efficiency of SAMC. For jury data, we set $U_3(\boldsymbol{x}) = 0$ for the SAMC sampler because SAMC fails to produce any valid tables. This is possibly caused by many structural zeros. For Whittaker's survey data, we set $U_1(\boldsymbol{x}) = 0$ for the SAMC sampler because allowing negative entries will increase the size of $\mathcal{X}$ drastically.

Since the BaB and CaB methods cannot be applied to incomplete tables containing structural zero cells, SAMC was compared with DaS and SMC. SAMC was run for 10 times. In the first three examples, each run consisted of 1.05×10^7 iterations, where the first 5×10^5 iterations were discarded for the burn-in process. For the exact inference, we used the remaining iterations 1.0×10^7. In Whittaker's survey study, each run consisted of 5.05×10^7 iterations with the first 5×10^5 iterations as a burn-in process. For comparison, DaS and SMC were applied to this example. The DaS and SMC methods were also run for 10 times independently, and each run consisted of same iterations with that done by SAMC. SAMC took less CPU time than SMC and DaS in each run for all examples.

4.1. Jury study

In this section, we show an example of testing the hypothesis of quasi-independence for a given dataset. Table 12 of Appendix shows a data collected by Vidmar (1972) for discovering the possible effects on decision making of limiting the number of alternatives available to the members of a jury panel. This is a 4×7 contingency table which has 9 structural zero cells. The degrees of freedom for testing quasi-independence is 9. The maximum likelihood estimates μ_{ij}'s under the hypothesis of quasi-independence are calculated by an iterative method as described in Section 2. The likelihood ratio statistic of this model is 18.8155 and the corresponding asymptotic p-value is 0.0268 from the asymptotic distribution $\chi^2(9)$. Aoki and Takemura (2005) performed the Markov chain Monte Carlo method with the minimal Markov basis, and found the estimated exact p-value, 0.0444 with estimated standard deviation 0.00052.

SAMC was first applied to this example. SAMC was run for 10 times with $T_0 = 10\,000$ and setting $m = 20$, which implies that the proportion of valid tables is about 27.4%. For the SMC sampler, we run for 10 times with setting $m = 20$ for comparison with SAMC, $N_p = 1.0 \times 10^5$ and six short MH chains for each sample. The initial population was generated by thinning a MH run of $6N_p$ iterations by a factor of 6. The thinning step makes the samples in the initial population less dependent on each other. Under these setting, SMC had about 20% more than MH moves done by SAMC.

Table 5 summarizes the numeric results. The significant results imply that two factors are dependent each other. SAMC outperforms DaS and SMC in terms of the accuracy of p-values although SAMC works as an importance sampling algorithm and the proportion of valid tables is only set to 0.274. This implies again that sampling from an enlarged set of valid tables improves the mixing of simulations. The p-value estimate of SMC is biased downward compared to the estimates of SAMC and DaS. This is due to the inefficiency of the MH sampler in the sample space exploration. It seems difficult for SMC to draw samples from the high probability region in this example although it can improve by using the better resampling schemes.

4.2. Health concerns study

The data in Table 13 of Appendix are a $4 \times 2 \times 2$ cross-classification of health concerns (H), their sex (S) and their age group (A) in teenagers, which were originally surveyed by Brunswick (1971) and analyzed by Grizzle and Williams (1972), followed by Fienberg (2007). Since males do not menstruate, there are two structural zeros in this table and one structural

Table 6
Comparison of SAMC with DaS, SMC and BSA methods for the health concerns study. DF: degrees of freedom for each model; χ^2-asymptotic: χ^2-asymptotic p-value and likelihood ratio statistic in the parentheses. Each of the other entries of the table is the p-value calculated by averaging over ten independent runs, and the number in the parentheses represents the standard deviation of the corresponding average, and CPU time (in minutes) cost by a single run on a 1.70 GHz personal computer.

Model	DF	χ^2-asymptotic	Method			
			SAMC	DaS	SMC	BSA
(HS, HA, SA)	2	0.3630 (2.0265)	0.3565 (0.00071, 4.45)	0.3564 (0.00079, 6.10)	0.3939 (0.00343, 5.12)	0.357 (0.0010)
(HS, HA)	3	0.1825 (4.8580)	0.1946 (0.00041, 4.48)	0.1968 (0.00127, 6.09)	0.0704 (0.00173, 5.05)	0.175 (0.0005)
(HS, SA)	5	0.0932 (9.4260)	0.1103 (0.00019, 4.32)	0.1099 (0.00147, 5.56)	0.0528 (0.00193, 4.43)	0.085 (0.0004)
(HA, SA)	4	0.0093 (13.4473)	0.0121 (0.00012, 4.27)	0.0126 (0.00048, 5.14)	0.0006 (0.00015, 4.59)	0.010 (0.0002)
(HS, A)	6	0.0158 (15.6441)	0.0213 (0.00039, 5.28)	0.0215 (0.00051, 8.15)	0.0516 (0.00230, 5.18)	
(HA, S)	5	0.0037 (17.4567)	0.0051 (0.00011, 5.13)	0.0052 (0.00016, 7.43)	0.0150 (0.00216, 5.17)	
(SA, H)	7	0.0025 (22.0247)	0.0034 (0.00008, 5.07)	0.0035 (0.00011, 7.43)	0.0093 (0.00258, 5.14)	
(H, S, A)	8	0.0004 (28.2428)	0.0006 (0.00004, 4.42)	0.0006 (0.00005, 6.36)	0.0217 (0.00410, 5.24)	

zero in the marginal of health concerns and sex. Fienberg (2007) argued that if the marginal of health concerns and sex is not fitted in a log-linear model, the number of degrees of freedom is reduced by two, and if their marginal is fitted, the number of degrees of freedom is reduced by one. In this example, we are interested in fitting various log-linear models to these data, as discussed in Dobra (2009) and Fienberg (2007). The likelihood ratio statistics and χ^2-asymptotic p-value of various log-linear models are given in Table 6.

Dobra (2009) proposed the bounds sampling algorithm (BSA) based on computation of lower and upper bounds determined by linear programming. He found the exact p-value estimates by BSA in some models, given in Table 6.

SAMC was first applied to this dataset. For the SAMC sampler, we set $T_0 = 1000$, $m = 10$ and $\mathcal{X} = \bigcup_{i=0}^{10} E_i$, where $E_0 = \{x : U(x) = 0\}$, $E_1 = \{x : U(x) = 1\}, \ldots, E_9 = \{x : U(x) = 9\}$ and $E_{10} = \{x : U(x) \geq 10\}$. The proportion of valid tables in SAMC is about 33.1%. For the SMC sampler, we run for 10 times with setting $m = 10$ for comparison with SAMC, $N_p = 1.5 \times 10^5$ and eight short MH chains for each sample. Under these settings, SMC had about 20% more than MH moves done by SAMC.

As discussed in Fienberg (2007) for fitting various log-linear models, the results of SAMC at the significance level 0.05 also suggest four acceptable models which contain some or all two-way interaction effects; i.e., $\lambda_{HS} + \lambda_{HA}$, $\lambda_{HS} + \lambda_{SA}$, $\lambda_A + \lambda_{HS}$, and $\lambda_{HS} + \lambda_{HA} + \lambda_{SA}$ (Table 6). Fienberg (2007) suggested that the model with λ_{HS} and λ_{HA} is an appropriate one for the data; i.e., given a particular health concern, there is no relationship between the age and sex of individuals with that concern.

For comparison, DaS and SMC were also applied to this example. Table 6 shows that all methods suggest the similar results in all models except SMC. However, SAMC can cost less CPU time than SMC and DaS in each run, producing much more accurate estimates. The resulting almost all estimates of SMC are biased compared to the estimates of SAMC and DaS.

4.3. The study of national bureau of economic research

Table 14 is a $4 \times 5 \times 4$ cross-classification of 4345 individuals by occupational groups, aptitude levels (A) and educational levels (E), which was collected in a 1969 survey of the National Bureau of Economic Research (Dobra, 2009; Fienberg, 2007). The data contain the 12 structural zeros, which are associated with professional self-employees and teachers being required to have higher educational levels. We are interested in assessing the fit of all two-way interactions log-linear model; i.e., no three-way interaction effect model. The number of degrees of freedom for this model is 26. Maximum likelihood estimates can be calculated by the Deming–Stephen iterative proportional fitting procedure (Bishop et al., 1975). The observed value of the likelihood-ratio test statistic is 15.906 which leads to an asymptotic p-value for the all two-way interactions model of 0.938. Dobra (2009) found that BSA converges to the p-value estimate, 0.968, of the exact p-value for the likelihood-ratio test.

For comparison, SAMC, DaS and SMC were applied to this example. SAMC was run for 10 times with $T_0 = 10\,000$ and $m = 20$. For the SMC sampler, we also run for 10 times with setting $m = 20$, $N_p = 1.0 \times 10^5$ and six short MH chains for each sample. Under these settings, SMC had about 20% more than MH moves done by SAMC. Table 7 shows that all methods also suggest the insignificant results for testing all two-way interactions log-linear model; i.e., there is no three-way interaction effect in this model. Among all methods, SAMC outperforms based on the accuracy of p-value. The estimate of SMC is biased downward compared to the estimates of other methods. Although the BSA method by proposed Dobra (2009) is efficient in incomplete tables, its variability is higher than our method. In this example, one run of SAMC costs about 12.22 min CPU time, which is much shorter than those done by DaS and SMC.

4.4. A revisit of Whittaker's survey study

We consider the contingency table studied in Section 3.3 again. Whittaker (1990) argued that the Rochdale data are sparse and thus any asymptotic results relating to the limiting distributions of goodness-of-fit statistics for loglinear models becomes questionable due to the existence of many zeros. Whittaker (1990) also noted that log-linear models with lower

Table 7
Comparison of SAMC with BSA, DaS and SMC methods for the study of national bureau of economic research. $\hat{p}_h$: estimate of p_h by averaging over 10 runs; SD: standard deviation of $\hat{p}_h$; Proportion: proportion of valid tables and its standard deviation (given in parentheses), which were calculated based on 10 independent runs; CPU (m): CPU time (in minutes) cost by a single run on a 1.70 GHz personal computer.

Method	$\hat{p}_h$	SD	Proportion	CPU (m)
χ^2-asymptotic	0.938			
BSA	0.968	2.00×10^{-3}		24.35
DaS	0.965	7.57×10^{-4}	1.00 (0)	18.45
SMC	0.511	1.80×10^{-2}	–	14.24
SAMC	0.965	3.95×10^{-4}	0.274 (3.15×10^{-5})	12.22

Table 8
Comparison of SAMC with BSA, DaS and SMC methods for the incomplete data of the Whittaker survey study. $\hat{p}_h$: estimate of p_h by averaging over 10 runs; SD: standard deviation of $\hat{p}_h$; Proportion: proportion of valid tables and its standard deviation (given in parentheses), which were calculated based on 10 independent runs; CPU (m): CPU time (in minutes) cost by a single run on a 2.90 GHz personal computer.

Method	$\hat{p}_h$	SD	Proportion	CPU (m)
χ^2-asymptotic	0.200			
BSA	0.197			
DaS	0.956	3.17×10^{-3}	1.00 (0)–	33.14
SMC	0	0.0	–	7.03
SAMC	0.239	2.97×10^{-3}	0.331 (1.41×10^{-2})	24.35

order interactions fit this dataset well. Dobra (2009) insisted that determining whether a contingency table is sparse is crucial for a statistician who needs to choose whether to consider only log-linear models with lower-order interactions or to assess goodness-of-fit based exclusively on exact tests. Dobra (2009) proposed an alternate way of analyzing this sparse tables, assuming that only the cells with positive observed counts exist and the others are missing. Treating counts of zero as structural zeros makes Whittaker's survey data an incomplete table with the 165 cells of zero. He considered the test of the all two-way interactions model using the BSA method. The observed value of the χ^2-test statistic is 62.50 which leads to an asymptotic p-value for the all two-way interactions model of 0.2. After running five instances of BSA for 100 000 iterations with a burn-in of 1000 iterations, he found the estimated exact p-value is 0.197. We consider the all two-way interactions model, like in Dobra (2009). The negative log-likelihood ratio statistics for this test is 1278.9328.

SAMC was first applied to this example. For the SAMC sampler, we set $U_1(\boldsymbol{x}) = 0$, like in example 3.3. SAMC was run for 10 times with $T_0 = 50$ and $m = 10$. The proportion of valid tables in SAMC is about 33.1%. DaS was also applied to this example with running for 10 times. For the SMC sampler, we also run for 10 times with setting $m = 10$, $N_p = 1.0 \times 10^5$ and 50 MH chains for each sample. Table 8 shows the results of SAMC, DaS and SMC for this example. The comparison indicates that for this example, SAMC outperforms DaS and SMC with more accurate estimates. SMC failed for this example due to its inability in generating valid tables. Although DaS can generate valid tables, DaS also failed due to the inefficiency of the MH sampler in sample space exploration. The resulting estimate of p-value by DaS is up-biased. All methods suggest there is no three-way interaction effect in this model.

5. Conclusion

In this paper, we have proposed to use a general method via a stochastic approximation Monte Carlo algorithm for the exact inference in complete or incomplete contingency table analysis. Our method includes the SAMC sampler, a known Markov basis, and an enlarged reference set containing the tables with negative entries, not satisfying the fixed margin for sufficient statistics, and violating structural zero cells. Our method ensures irreducibility by working on an enlarged reference set and employing the known Markov basis as the proposal. The self-adjust mechanism of SAMC can control the proportion of valid tables, which has been beyond the ability of other importance sampling and MCMC methods.

Our method was compared with DaS, SIS, SMC and BSA algorithms on six real datasets. The numerical results indicate that our method using the SAMC algorithm takes much less CPU time in obtaining estimates which are much more accurate than other methods, such as the MH algorithm by Diaconis and Sturmfels (1998), the sequential importance sampling method by Dinwoodie and Chen (2011), the sequential Monte Carlo method by Del Moral et al. (2006), and the bounds sampling algorithm by Dobra (2009). No matter whether the Markov basis is available or unavailable, SAMC can be more efficient than all other methods. SAMC less likely gets trapped by local minima, and thus has more chance of inferring correctly from the desired distribution for the contingency tables.

Acknowledgments

Cheon's research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0015000). Jung's research was supported by the research fund of the University of Seoul in 2011.

Table 9
Data for cross-cultural study example.
Source: Murdock and White (1969).

Society	Bridewealth groups	Patrilineal groups	
		Present	Absent
Insular pacific	Present	5	8
	Absent	9	9
Outside of insular pacific	Present	46	11
	Absent	22	76

Table 10
Data for livestock breeds study example.
Source: Hall and Ruane (1993).

Region	Presence	Animal						
		Ass	Water Buffalo	Cattle	Goat	Horse	Pig	Sheep
Africa	Rare	0	0	10	0	2	0	4
	Extinct	0	0	22	0	2	0	1
Asia	Rare	0	2	8	4	14	2	1
	Extinct	0	0	5	1	3	8	2
Europe	Rare	10	0	101	29	49	37	109
	Extinct	5	0	154	19	58	79	98
North and Central America	Rare	0	0	8	4	9	5	7
	Extinct	0	0	1	1	4	17	10
South America	Rare	1	0	4	0	0	0	1
	Extinct	0	0	19	0	0	0	0
Oceania	Rare	0	0	1	0	1	1	2
	Extinct	0	0	2	0	1	1	5
Ex-USSR	Rare	0	0	9	4	23	2	11
	Extinct	0	0	21	6	20	21	32

Table 11
Data for Whittaker's survey study example.
Source: Whittaker (1990).

5	0	2	1	5	1	0	0	4	1	0	0	6	0	2	0
8	0	11	0	13	0	1	0	3	0	1	0	26	0	1	0
5	0	2	0	0	0	0	0	0	0	0	0	0	0	1	0
4	0	8	2	6	0	1	0	1	0	1	0	0	0	1	0
17	10	1	1	16	7	0	0	0	2	0	0	10	6	0	0
1	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0
4	7	3	1	1	1	2	0	1	0	0	0	1	0	0	0
0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
18	3	2	0	23	4	0	0	22	2	0	0	57	3	0	0
5	1	0	0	11	0	1	0	11	0	0	0	29	2	1	1
3	0	0	0	4	0	0	0	1	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	25	0	1	37	26	0	0	15	10	0	0	43	22	0	0
0	0	0	0	2	0	0	0	0	0	0	0	3	0	0	0
2	4	0	0	2	1	0	0	0	1	0	0	2	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 12
Data for the jury study example.
Source: Vidmar (1972).

Alternative	Condition						
	1	2	3	4	5	6	7
First-degree	11	[0]	[0]	2	7	[0]	2
Second-degree	[0]	20	[0]	22	[0]	11	15
Manslaughter	[0]	[0]	22	[0]	16	13	5
Not guilty	13	4	2	0	1	0	2

Appendix

See Tables 9–14.

Table 13
Data for health concerns study example.
Source: Fienberg (2007, p. 148).

Health concerns	Sex	Age	
		12–15	16–17
Sex, reproduction	Male	4	2
	Female	9	7
Menstrual problems	Male	[0]	[0]
	Female	4	8
How healthy I am	Male	42	7
	Female	19	10
Nothing	Male	57	20
	Female	71	31

Table 14
Data for the study of national bureau of economic research.
Source: Fienberg (2007).

Occupation	Aptitude	Education			
		E1	E2	E3	E4
Self-employed, business	A1	42	55	22	3
	A2	72	82	60	12
	A3	90	106	85	25
	A4	27	48	47	8
	A5	8	18	19	5
Self-employed, professional	A1	1	2	8	19
	A2	1	2	15	33
	A3	2	5	25	83
	A4	2	2	10	45
	A5	[0]	[0]	12	19
Teacher	A1	[0]	[0]	1	19
	A2	[0]	[0]	3	60
	A3	[0]	[0]	5	86
	A4	[0]	[0]	2	36
	A5	[0]	[0]	1	14
Salary-employed	A1	172	151	107	42
	A2	208	198	206	92
	A3	279	271	331	191
	A4	99	126	179	97
	A5	36	35	99	79

References

4ti2 team, (2006). 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at: www.4ti2.de.
Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*, *7*, 131–153.
Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Wiley.
Andrieu, C., Moulines, É, & Priouret, P. (2005). Stability of stochastic approximation under verifiable conditions. *SIAM Journal on Control and Optimization*, *44*, 283–312.
Aoki, S., & Takemura, A. (2005). Markov chain Monte Carlo exact tests for incomplete two-way contingency tables. *Journal of Statistical Computation and Simulation*, *75*(10), 787–812.
Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice*. Cambridge: The MIT Press.
Booth, J. G., & Butler, R. W. (1999). An importance sampling algorithm for exact conditional test in log-linear models. *Biometrika*, *86*(2), 321–332.
Brunswick, A. F. (1971). Adolescent health, sex, and fertility. *American Journal of Public Health*, *61*, 711–720.
Bunea, F., & Besag, J. (2000). MCMC in $I \times J \times K$ contingency tables. In *Fields institute communications* (pp. 25–36). American Mathematical Society.
Caffo, B. S., & Booth, J. G. (2001). A Markov chain Monte Carlo algorithm for approximating exact conditional probabilities. *Journal of Computational and Graphical Statistics*, *10*(4), 730–745.
Chen, H. F. (2002). *Stochastic approximation and its applications*. Dordrecht: Kluwer Academic Publishers.
Chen, Y. (2007). Conditional inference on tables with structural zeros. *Journal of Computational and Graphical Statistics*, *16*, 445–467.
Chen, Y., Diaconis, P., Holmes, S. P., & Liu, J. S. (2005). Sequential Monte Carlo methods for statistical analysis of tables. *Journal of the American Statistical Association*, *100*, 109–120.
Chen, Y., Dinwoodie, I. H., & Sullivant, S. (2006). Sequential importance sampling for multiway tables. *The Annals of Statistics*, *34*, 523–545.
Chen, Y., Dinwoodie, I. H., & Yoshida, R. (2010). Markov chains, quotient ideals and connectivity with positive margins. In P. Gibilisco, E. Riccomagno, M. P. Rogantin, & H. P. Wynn (Eds.), *Algebraic and geometric methods in statistics* (pp. 99–110). Cambridge: The MIT Press.
Cheon, S., Liang, F., Chen, Y., & Yu, K. (2013). Stochastic approximation Monte Carlo importance sampling for approximating exact conditional probabilities. *Statistics and Computing* (in press).
Del Moral, P., Doucet, A., & Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(3), 411–436.
Deloera, J. A., & Onn, S. (2006). Markov basis of three-way tables are arbitrarily complicated. *Journal of Symbolic Computation*, *41*, 173–181.
Diaconis, P., & Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, *26*, 363–397.
Dinwoodie, I. H., & Chen, Y. (2011). Sampling large tables with constraints. *Statistica Sinica*, *21*, 1591–1609.

Dobra, A. (2003). Markov bases for decomposable graphical models. *Bernoulli*, *9*, 1093–1108.
Dobra, A. (2009). *Computing exact p-values in incomplete multi-way tables. Technical report no. 548*. Department of Statistics, University of Washington.
Faraway, J. J. (2006). *Extending the linear model with R*. Chapman & Hall/CRC.
Fienberg, S. E. (2007). *The analysis of cross-classified categorical data* (2nd ed.). New York: Springer Science.
Grizzle, J. E., & Williams, O. D. (1972). Loglinear models and tests of independence for contingency tables. *Biometrics*, *28*, 137–156.
Haberman, S. J. (1988). A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. *Journal of the American Statistical Association*, *83*, 555–560.
Hall, S. J. G., & Ruane, J. (1993). Livestock breeds and their conservation: a global overview. *Conservation Biology*, *7*, 815–825.
Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*, 97–109.
Kreiner, S. (1987). Analysis of multidimensional contingency tables by exact conditional tests: techniques and strategies. *Scandinavian Journal of Statistics*, *14*, 97–112.
Liang, F. (2009). On the use of stochastic approximation Monte Carlo for Monte Carlo integration. *Statistics & Probability Letters*, *79*, 581–587.
Liang, F., Liu, C., & Carroll, R. (2007). Stochastic approximation in Monte Carlo computation. *Journal of American Statistical Association*, *102*(477), 305–320.
McCullagh, P. (1986). The conditional distribution of goodness-of-fit statistics for discrete data. *Journal of the American Statistical Association*, *81*, 104–107.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*, 1087–1091.
Murdock, G. P., & White, D. R. (1969). Standard cross-cultural sample. *Ethnology*, *8*, 329–369.
Paul, S., & Deng, D. (2000). Goodness of fit of generalized linear models to sparse data. *Journal of the Royal Statistical Society: Series B*, *62*, 323–333.
Rapallo, F. (2006). Markov bases and structural zeros. *Journal of Symbolic Computation*, *41*, 164–172.
Rapallo, F., & Yoshida, R. (2010). Markov bases and subbases for bounded contingency tables. *Annals of the Institute of Statistical Mathematics*, *62*(4), 785–805.
Robbins, H., & Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*, *22*, 400–407.
Roberts, G. O., & Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, *83*, 95–110.
Rubinstein, R. Y., & Kroese, D. P. (2007). *Simulation and the Monte Carlo method* (2nd ed.). New York: Wiley.
Vidmar, N. (1972). Effects of decision alternatives on the verdicts and social perceptions of simulated jurors. *Journal of Personality and Social Psychology*, *22*, 211–218.
White, D. R., Pesner, R., & Reitz, K. P. (1983). An exact significance test for three-way interaction effects. *Behavior Science Research*, *18*(2), 103–122.
Whittaker, J. (1990). *Graphical models in applied mathematical statistics*. New York: Wiley.