MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Markov chain Monte Carlo tests for designed experiments

Satoshi AOKI and Akimichi TAKEMURA

METR 2006–56   November 2006

Markov chain Monte Carlo tests for designed experiments

Satoshi Aoki

Department of Mathematics and Computer Science

Kagoshima University

and

Akimichi Takemura

Graduate School of Information Science and Technology

University of Tokyo

November, 2006

Abstract

We consider conditional exact tests of factor effects in designed experiments for discrete response variables. Similarly to the analysis of contingency tables, a Markov chain Monte Carlo method can be used for performing exact tests, when large-sample approximations are poor and the enumeration of the conditional sample space is infeasible. For designed experiments with a single observation for each run, we formulate log-linear or logistic models and consider a connected Markov chain over an appropriate sample space. In particular, we investigate fractional factorial designs with 2^{p-q} runs, noting correspondences to the models for 2^{p-q} contingency tables.

1   Introduction

Exact calculations of p values for statistical conditional tests arise mainly in the context of analyzing contingency tables. For example, Fisher's exact test is frequently used for evaluating the hypothesis that the row effect and column effect are independent in the 2×2 contingency tables. Fisher's exact test is generalized to $I \times J$ contingency tables in [11]. Traditionally, statistical tests for contingency tables have relied heavily on large-sample approximations for sampling distribution of the test statistics. However, many works have shown that large-sample approximations can be very poor when the contingency table contains both small and large expected frequencies even when the sample size is large. See [12], for example. Moreover, coupled with rapid development both in computer power and in techniques of algorithms, exact calculations of p values become feasible in various

1

settings for practical use. Consequently, for many types of problems where some ingenious calculation schemes are invented, it is unnecessary to use large-sample approximations for sampling distributions nowadays when their adequacy is in doubt. A typical example is the network algorithm by [18] for calculating exact p values of Freeman-Halton tests in two-way contingency tables. See the survey paper by [2].

At the same time simulation techniques for estimating p values by Monte Carlo procedures have also developed. In particular, for the problems where a closed form expression of the sampling distribution can not be obtained, Monte Carlo method provide powerful tools. Note that, in contrast to the large-sample approximations, we can estimate p values in arbitrary accuracy, theoretically, by increasing simulation sizes. However, for many models, such as general hierarchical log-linear models in multi-way contingency tables, direct generation of random sample is not straightforward. In this case, *Markov chain Monte Carlo* techniques can be used.

For performing Markov chain Monte Carlo methods for sampling from discrete sample space, an important problem is how to construct a connected Markov chain on the given sample space. Note that, if an arbitrary connected Markov chain is constructed, the chain can be modified to give a connected and aperiodic Markov chain with stationary distribution being the desired null distribution by the usual Metropolis procedure ([14], for example). As for this point, the first breakthrough work is given by [8]. The key notion of [8] is *a Markov basis*, which enables to construct a connected Markov chain for arbitrary observed data set. [8] presented a general algorithm for computing a Markov basis in the settings of a general discrete exponential family of distribution. Their approach relies on the existence of a Gröbner basis of a well specified polynomial ideal. After [8], the techniques of the Markov chain Monte Carlo method for sampling from discrete conditional distributions are rapidly developed in the past decade. See for example [9], [10] and the works by Aoki and Takemura ([4], [3], [5], [6], [23], [24]).

In this paper, we consider conditional exact tests in the context of *designed experiments* with counts (or ratios of counts) observations. In most of the classical literatures on designed experiments, the responses are assumed to be normally distributed. However, in many practical situations, the experimental data are not normally distributed. For such non-normal data, the generalized linear models are frequently used. See [13] or Chapter 13 of [25], for example. In these literatures, however, exact testing procedures for non-normal data are not considered. Since the experimental design is used when the cost of obtaining the data is relatively high, it is very important to develop techniques of exact procedures for the case of non-normal responses. Therefore in this manuscript, we consider the exact testing procedures for non-normal responses, based on the theory of the generalized linear models. For discrete responses, the above background and strategies also apply, i.e., to calculate p values for conditional tests,

1. traditionally, large-sample approximations such as the normal distribution or chi-square distribution are used,

2. if the observed data set contains both small and large expected values, the adequacy of the approximation becomes poor,

3. if the sample space is of moderate size, or some ingenious algorithms can be used, an exact calculation of p values is possible,

4. if an exact calculation is not feasible, we can rely on Monte Carlo procedure,

5. if a closed form expression of the null distribution is not given, Markov chain Monte Carlo procedure can be employed, if a Markov basis is available.

The topic we consider in this paper is No. 5 of the above list. In Section 2, we formulate conditional exact tests of factor effects for fractional factorial designs. For designed experiments with a single observation for each run, we formulate log-linear or logistic models and consider how to construct a null model to be tested using the theory of generalized linear models. In Section 3, we consider Markov chain Monte Carlo tests for designed experiments. First we give a definition of Markov bases and a simple algorithm for evaluating p values by the Markov chain Monte Carlo tests in Section 3.1. In Section 3.2, we consider correspondences between the fractional factorial designs *with 2^{p-q} runs* and models for 2^{p-q} contingency tables. We end the paper with some discussions in Section 4.

2 Conditional tests for fractional factorial designs

In this section, we consider the exact conditional tests for the discrete observations derived from fractional factorial designs. We investigate the designs with a single observation for each run, which is either a count or a ratio of counts. For the former case we consider the log-linear models, and for the latter case we consider the logistic models. Since our arguments for the two cases are almost the same, we first explain in detail the log-linear case in Section 2.1, and then give only a short description and a remark for the logistic case in Section 2.2.

2.1 Exact conditional tests for log-linear models of Poisson observations

First we investigate the case that the observations are counts of some events. In this case, it is natural to consider Poisson models. To clarify the procedures of exact tests, we take a close look at an example of fractional factorial design with counts observations. Table 1 is a 1/8 fraction of a full factorial design (i.e., a 2^{7-3} fractional factorial design) defined from the *aliasing relation*

$$ABDE = ACDF = BCDG = I, \tag{1}$$

and response data analyzed in [7] and reanalyzed in [13]. In Table 1, the observation y is the number of defects arising in a wave-soldering process in attaching components to an electronic circuit card. In Chapter 7 of [7], he considered seven factors of a wave-soldering process: (A) prebake condition, (B) flux density, (C) conveyer speed, (D) preheat condition, (E) cooling time, (F) ultrasonic solder agitator and (G) solder temperature,

Table 1: Design and number of defects y for the wave-solder experiment

	Factor								y	
Run	A	B	C	D	E	F	G	1	2	3
1	1	1	1	1	1	1	1	13	30	26
2	1	1	1	2	2	2	2	4	16	11
3	1	1	2	1	1	2	2	20	15	20
4	1	1	2	2	2	1	1	42	43	64
5	1	2	1	1	2	1	2	14	15	17
6	1	2	1	2	1	2	1	10	17	16
7	1	2	2	1	2	2	1	36	29	53
8	1	2	2	2	1	1	2	5	9	16
9	2	1	1	1	2	2	1	29	0	14
10	2	1	1	2	1	1	2	10	26	9
11	2	1	2	1	2	1	2	28	173	19
12	2	1	2	2	1	2	1	100	129	151
13	2	2	1	1	1	2	2	11	15	11
14	2	2	1	2	2	1	1	17	2	17
15	2	2	2	1	1	1	1	53	70	89
16	2	2	2	2	2	2	2	23	22	7

each at two levels with three boards from each run being assessed for defects. The aim of this experiment is to decide which levels for each factors are desirable to reduce solder defects.

In this paper, we only consider designs with a single observation for each run. Therefore, in this example, we focus on the totals for each run in Table 1. This is natural for the settings of Poisson models, since the set of the totals for each run is the sufficient statistics for the parameters. We also ignore the second observation in run 11, which is an obvious outlier as pointed out in [13]. Therefore the weighted total of run 11 is $(28 + 19) \times 3/2 = 70.5 \simeq 71$. By replacing 2 by -1 in Table 1, we rewrite $k \times p$ design

matrix as D, where each element is $+1$ or -1. Consequently, we have

$$
D = \begin{pmatrix}
+1 & +1 & +1 & +1 & +1 & +1 & +1 \\
+1 & +1 & +1 & -1 & -1 & -1 & -1 \\
+1 & +1 & -1 & +1 & +1 & -1 & -1 \\
+1 & +1 & -1 & -1 & -1 & +1 & +1 \\
+1 & -1 & +1 & +1 & -1 & +1 & -1 \\
+1 & -1 & +1 & -1 & +1 & -1 & +1 \\
+1 & -1 & -1 & +1 & -1 & -1 & +1 \\
+1 & -1 & -1 & -1 & +1 & +1 & -1 \\
-1 & +1 & +1 & +1 & -1 & -1 & +1 \\
-1 & +1 & +1 & -1 & +1 & +1 & -1 \\
-1 & +1 & -1 & +1 & -1 & +1 & -1 \\
-1 & +1 & -1 & -1 & +1 & -1 & +1 \\
-1 & -1 & +1 & +1 & +1 & -1 & -1 \\
-1 & -1 & +1 & -1 & -1 & +1 & +1 \\
-1 & -1 & -1 & +1 & +1 & +1 & +1 \\
-1 & -1 & -1 & -1 & -1 & -1 & -1
\end{pmatrix}, \quad
\mathbf{y} = \begin{pmatrix}
69 \\ 31 \\ 55 \\ 149 \\ 46 \\ 43 \\ 118 \\ 30 \\ 43 \\ 45 \\ 71 \\ 380 \\ 37 \\ 36 \\ 212 \\ 52
\end{pmatrix}.
$$

Write $\mathbf{y} = (y_1, \ldots, y_k)'$ and $D = (d_{ij}) = (\mathbf{d}_1, \ldots, \mathbf{d}_p)$ where $\mathbf{d}_j = (d_{1j}, \ldots, d_{kj})' \in \{-1, +1\}^k$ is the j-th column vector of D. k is the number of runs. If there are q aliasing relations defining this design, $k = 2^{p-q}$ holds ($p = 7, q = 3$ for this example). We define $\mathbf{d}_{st}$ and $\mathbf{d}_{stu}$, $1 \leq s < t < u \leq p$, as

$$
\mathbf{d}_{st} = (d_{1s}d_{1t}, \ldots, d_{ks}d_{kt})'
$$

and

$$
\mathbf{d}_{stu} = (d_{1s}d_{1t}d_{1u}, \ldots, d_{ks}d_{kt}d_{ku})'
$$

for later use.

The statistical model for this type of data is constructed from the theory of *generalized linear models* ([17]). Assume that the observations y_i are mutually independently distributed with $\mu_i = E(y_i)$, $i = 1, \ldots, k$. The mean parameter μ_i is expressed as

$$
g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_\nu x_{i\nu},
$$

where $g(\cdot)$ is the link function and $x_{i1}, \ldots, x_{i\nu}$ are the ν covariates defined below. The sufficient statistic is written as $\sum_{i=1}^{k} x_{ij} y_i$. The canonical link for the Poisson distribution is $g(\mu_i) = \log \mu_i$.

Now we define covariates. We write the ν-dimensional parameter β and the covariate matrix X as

$$
\beta = (\beta_0, \beta_1, \ldots, \beta_{\nu-1})'
$$

and

$$
X = \begin{pmatrix}
1 & x_{11} & \cdots & x_{1\nu-1} \\
\vdots & \vdots & \ldots & \vdots \\
1 & x_{k1} & \cdots & x_{k\nu-1}
\end{pmatrix} = \begin{pmatrix} \mathbf{1}_k & \mathbf{x}_1 & \cdots & \mathbf{x}_{\nu-1} \end{pmatrix},
$$

where $\mathbf{1}_k = (1, \ldots, 1)'$ is the k-dimensional column vector. Since the likelihood function is written as

$$
\prod_{i=1}^{k} \frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i} = \left(\prod_{i=1}^{k} \frac{e^{-\mu_i}}{y_i!} \right) \exp\left(\sum_{i=1}^{k} y_i \log \mu_i \right)
$$
$$
= \left(\prod_{i=1}^{k} \frac{e^{-\mu_i}}{y_i!} \right) \exp\left(\beta_0 \mathbf{1}_k' \mathbf{y} + \sum_{j=1}^{\nu-1} \beta_j \mathbf{x}_j' \mathbf{y} \right)
$$
$$
= \left(\prod_{i=1}^{k} \frac{e^{-\mu_i}}{y_i!} \right) \exp\left(\beta' X' \mathbf{y} \right),
$$

the sufficient statistic for β is $X'\mathbf{y} = (\mathbf{1}_k'\mathbf{y}, \mathbf{x}_1'\mathbf{y}, \ldots, \mathbf{x}_{\nu-1}'\mathbf{y})$.

The matrix X is constructed from the design matrix D to reflect the effects of the factors and their interactions which we intend to measure. For example, a simple model which only includes the main effects for each factor is given as $X = (\mathbf{1}_k \ D)$, i.e., $\mathbf{x}_j = \mathbf{d}_j$ for $j = 1, \ldots, \nu - 1 = p$. On the other hand, we can consider a more complicated model containing various interaction effects, under the condition that it is consistent with the aliasing relations. In this example, the aliasing relation up to four-factor interactions is derived from (1) as follows.

$$
\begin{aligned}
&\text{I} = \text{ABDE} = \text{ABFG} = \text{ACDF} = \text{ACEG} = \text{BCDG} = \text{BCEF} = \text{DEFG} \\
&\text{A} = \text{BDE} = \text{BFG} = \text{CDF} = \text{CEG}, \qquad \text{B} = \text{ADE} = \text{AFG} = \text{CDG} = \text{CEF} \\
&\text{C} = \text{ADF} = \text{AEG} = \text{BDG} = \text{BEF}, \qquad \text{D} = \text{ABE} = \text{ACF} = \text{BCG} = \text{EFG} \\
&\text{E} = \text{ABD} = \text{ACG} = \text{DFG}, \qquad \text{F} = \text{ABG} = \text{ACD} = \text{BCF} = \text{DEG} \\
&\text{G} = \text{ABF} = \text{ACE} = \text{BCD} = \text{DEF} \\
&\text{AB} = \text{DE} = \text{FG} = \text{ABDG} = \text{ACDG} = \text{ACEF} = \text{BCDF} = \text{BCEG} \\
&\text{AC} = \text{DF} = \text{EG} = \text{ABEF} = \text{BCDE} = \text{BCFG} \\
&\text{AD} = \text{BE} = \text{CF} = \text{ABCG} = \text{AEFG} = \text{BDFG} = \text{CDEG} \\
&\text{BC} = \text{DG} = \text{EF} = \text{ABDF} = \text{ABEG} = \text{ACDE} = \text{ACFG} \\
&\text{BD} = \text{AE} = \text{CG} = \text{ABCF} = \text{ADFG} = \text{BEFG} = \text{CDEF} \\
&\text{CD} = \text{AF} = \text{BG} = \text{ABCE} = \text{ADEG} = \text{BDEF} = \text{CEFG} \\
&\text{AG} = \text{BF} = \text{CE} = \text{ABCD} = \text{ADEF} = \text{BDEG} = \text{CDFG} \\
&\text{ABC} = \text{ADG} = \text{AEF} = \text{BDF} = \text{BEG} = \text{CDE} = \text{CFG}
\end{aligned}
$$

Subject to the above aliasing relations, we may consider appropriate models where all the parameters are estimable. For example, the saturated model for this example includes $16(= k)$ parameters, when X is the Hadamard matrix of the order 16. One of the interpretations of the saturated model includes seven main effects, seven two-factor interaction effects, $\text{AB}, \text{AC}, \text{AD}, \text{AG}, \text{BC}, \text{BD}, \text{CD}$, and one three-factor interaction effect, ABC. We write this model as

$$
\text{ABC/AD/BD/CD/AG/E/F}
$$

by the manner of the hierarchical models. In this case, as in Table 2, the columns of X can be indexed as

$$
X = (\mathbf{1}_k, \mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_{12}, \mathbf{d}_3, \mathbf{d}_{13}, \mathbf{d}_{23}, \mathbf{d}_{123}, \mathbf{d}_4, \mathbf{d}_{14}, \mathbf{d}_{24}, \mathbf{d}_5, \mathbf{d}_{34}, \mathbf{d}_6, \mathbf{d}_7, \mathbf{d}_{17}).
$$

Table 2: Hadamard matrix of the order 16 and two interpretations.

+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
+1	+1	+1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	−1	−1	−1
+1	+1	+1	+1	−1	−1	−1	−1	+1	+1	+1	+1	−1	−1	−1	−1
+1	+1	+1	+1	−1	−1	−1	−1	−1	−1	−1	−1	+1	+1	+1	+1
+1	+1	−1	−1	+1	+1	−1	−1	+1	+1	−1	−1	+1	+1	−1	−1
+1	+1	−1	−1	+1	+1	−1	−1	−1	−1	+1	+1	−1	−1	+1	+1
+1	+1	−1	−1	−1	−1	+1	+1	+1	+1	−1	−1	−1	−1	+1	+1
+1	+1	−1	−1	−1	−1	+1	+1	−1	−1	+1	+1	+1	+1	−1	−1
+1	−1	+1	−1	+1	−1	+1	−1	+1	−1	+1	−1	+1	−1	+1	−1
+1	−1	+1	−1	+1	−1	+1	−1	−1	+1	−1	+1	−1	+1	−1	+1
+1	−1	+1	−1	−1	+1	−1	+1	+1	−1	+1	−1	−1	+1	−1	+1
+1	−1	+1	−1	−1	+1	−1	+1	−1	+1	−1	+1	+1	−1	+1	−1
+1	−1	−1	+1	+1	−1	−1	+1	+1	−1	−1	+1	+1	−1	−1	+1
+1	−1	−1	+1	+1	−1	−1	+1	−1	+1	+1	−1	−1	+1	+1	−1
+1	−1	−1	+1	−1	+1	+1	−1	+1	−1	−1	+1	−1	+1	+1	−1
+1	−1	−1	+1	−1	+1	+1	−1	−1	+1	+1	−1	+1	−1	−1	+1
I	A	B	AB	C	AC	BC	ABC	D	AD	BD	E	CD	F	G	AG
I	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD

Note that the interpretation of the models is not unique. An extreme interpretation of the saturated model is given by ignoring three main effects, E, F, G, and considering all the higher interaction effects among A, B, C, D, which is written as

$$ABCD.$$

This interpretation is not realistic in the context of designed experiments, but it is useful for understanding Markov bases in terms of contingency tables. These two interpretations of the saturated model are shown in Table 2. Note that the concept of identifiable model in the saturated model is also given in [19]. Using the theory of Gröbner basis, [19] gives a method to obtain one of the maximal identifiable models for a given design. See also [22] and [20].

Since the saturated model cannot be tested, we consider an appropriate submodel of the saturated model. For the purpose of illustration we focus on the model considered in [13], which is given as

$$AC/BD/E/F/G, \tag{2}$$

i.e., the model of seven main effects and two two-factor interactions. We treat this model as the *null model* and consider significance tests. In this case, the null hypothesis can be described as follows. Permuting the columns of Table 2, we partition the covariate matrix X of the saturated model as

$$X = (X_0 \quad X_1),$$
$$X_0 = (\mathbf{1}_k, \mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4, \mathbf{d}_5, \mathbf{d}_6, \mathbf{d}_7, \mathbf{d}_{13}, \mathbf{d}_{24}) = (\mathbf{1}_k, \mathbf{x}_1, \ldots, \mathbf{x}_{\nu-1}),$$
$$X_1 = (\mathbf{d}_{12}, \mathbf{d}_{23}, \mathbf{d}_{123}, \mathbf{d}_{14}, \mathbf{d}_{34}, \mathbf{d}_{17}) = (\mathbf{x}_\nu, \ldots, \mathbf{x}_{k-1}),$$

and consider the corresponding parameter $\beta = (\beta_0, \beta_1, \ldots, \beta_{k-1})$. Then the submodel is specified in the form of a null hypothesis

$$H_0: \quad \beta_\nu = \cdots = \beta_{k-1} = 0.$$

7

Under H_0, the nuisance parameters are $\beta_0, \ldots, \beta_{\nu-1}$ and the sufficient statistic for the nuisance parameters is $X_0'\mathbf{y}$. Then the conditional distribution of $\mathbf{y}$ given the sufficient statistics is written as

$$f(\mathbf{y} \mid X_0'\mathbf{y} = X_0'\mathbf{y}^o) = C(X_0'\mathbf{y}^o) \prod_{i=1}^{k} \frac{1}{y_i!}, \tag{3}$$

where $C(X_0'\mathbf{y}^o)$ is the normalizing constant determined from $X_0'\mathbf{y}^o$ and written as

$$C(X_0'\mathbf{y}^o)^{-1} = \sum_{\mathbf{y} \in \mathcal{F}(X_0'\mathbf{y}^o)} \left(\prod_{i=1}^{k} \frac{1}{y_i!} \right), \tag{4}$$

and

$$\mathcal{F}(X_0'\mathbf{y}^o) = \{\mathbf{y} \mid X_0'\mathbf{y} = X_0'\mathbf{y}^o, \ y_i \text{ is a nonnegative integer for } i = 1, \ldots, k\}. \tag{5}$$

Note that by sufficiency the conditional distribution does not depend on the values of the nuisance parameters.

We can now consider significance tests against various alternatives to H_0. An important alternative is to test the effect of a single additional effect. For example in the above example we can test presence of AB-interaction effect by considering the alternative hypothesis $H_1 : \beta_\nu \neq 0$. Or if we are interested in the goodness-of-fit test, then the alternative hypothesis is $H_1 : (\beta_\nu, \ldots, \beta_{k-1}) \neq (0, \ldots, 0)$. Depending on the alternative hypothesis, we can use appropriate test statistic $T(\mathbf{y})$, such as the likelihood ratio statistic for testing H_0 against H_1. In Section 3 we give a procedure to sample from the conditional distribution (3). Therefore we can assess the conditional distribution of any test statistic $T(\mathbf{y})$ under H_0.

For the purpose of illustration we now consider goodness-of-fit procedures. Traditional χ^2 tests evaluate the upper probability for some discrepancy measures such as the deviance, the likelihood ratio or Pearson goodness-of-fit, based on the asymptotic distribution, $\chi^2_{k-\nu}$. For example, the likelihood ratio statistic

$$T(\mathbf{y}) = G^2(\mathbf{y}) = 2 \sum_{i=1}^{k} y_i \log \frac{y_i}{\hat{\mu}_i} \tag{6}$$

is frequently used, where $\hat{\mu}_i$ is the maximum likelihood estimate for μ_i under the null model (i.e., fitted value), given by

$$\hat{\mu} = (64.53, 47.25, \ldots, 51.42)'$$

for our example. Then for the observed data $\mathbf{y}^o = (y_1^o, \ldots, y_k^o)'$, $T(\mathbf{y}^o) = G^2(\mathbf{y}^o)$ is calculated as $T(\mathbf{y}^o) = 19.096$ and the corresponding asymptotic p value is 0.0040 from the asymptotic distribution χ^2_6. This result tells us that the null hypothesis is highly

Table 3: Design and number of good parts y out of 1000 for the windshield molding slugging experiment

	Factor				
Run	A	B	C	D	y
1	1	1	1	1	338
2	1	1	2	2	826
3	1	2	1	1	350
4	1	2	2	2	647
5	2	1	1	2	917
6	2	1	2	1	977
7	2	2	1	2	953
8	2	2	2	1	972

significant and is rejected. Using the conditional distribution (3), the exact p value is written as

$$p = \sum_{\mathbf{y} \in \mathcal{F}(X_0' \mathbf{y}^o)} f(\mathbf{y} \mid X_0' \mathbf{y} = X_0' \mathbf{y}^o) \mathbf{1}(T(\mathbf{y}) \geq T(\mathbf{y}^o)),$$

where

$$\mathbf{1}(T(\mathbf{y}) \geq T(\mathbf{y}^o)) = \begin{cases} 1, & \text{if } T(\mathbf{y}) \geq T(\mathbf{y}^o), \\ 0, & \text{otherwise.} \end{cases}$$

Unfortunately, however, an enumeration of all the elements in $\mathcal{F}(X_0' \mathbf{y}^o)$ and hence the calculation of the normalizing constant $C(X_0' \mathbf{y}^o)$ is usually computationally infeasible for large sample space. Instead, we consider a Markov chain Monte Carlo method described in Section 3.

2.2 Exact conditional tests for logistic models of binomial observations.

Next we consider the case that the observation for each run is a ratio of counts. Table 3 is a 1/2 fraction of a full factorial design (i.e., a 2^{4-1} fractional factorial design) defined from the relation

$$\text{ACD} = \text{I} \tag{7}$$

and response data given by [16] and reanalyzed in [13]. In Table 3, the observation y is the number of good parts out of 1000 during the stamping process in manufacturing windshield modeling. The purpose of [16] is to decide the levels for four factors, (A) poly-film thickness, (B) oil mixture, (C) gloves and (D) metal blanks, which most improve the slugging condition. Similarly to Section 2.1, we rewrite this data as

$$D = \begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & +1 \\ +1 & -1 & -1 & -1 \\ -1 & +1 & +1 & -1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & +1 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 338 \\ 826 \\ 350 \\ 647 \\ 917 \\ 977 \\ 953 \\ 972 \end{pmatrix}.$$

As for a statistical model for this type of data, it is natural to suppose that the distribution of the observation y_i is the mutually independent binomial distribution $\mathrm{Bin}(\mu_i, n_i)$, $i = 1, \ldots, k$, where $n_i = 1000$, $i = 1, \ldots, k = 8$ for this example. Following the theory of generalized linear models, we also consider the logit link, which is the canonical link for the binomial distribution. It expresses the relation between the mean parameter μ_i and the systematic part as

$$g(\mu_i) = \log \frac{\mu_i}{1 - \mu_i} = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_\nu x_{i\nu}.$$

The covariate matrix and the corresponding parameters are defined similarly as Section 2.1, i.e., permuting the columns of the Hadamard matrix of the order 8, we write $X = (X_0 \; X_1)$ in such a way that X_0 is the covariate matrix for the appropriate null model. In this example, we consider the simple main-effects model, A/B/C/D, which is considered in [13]. For this model, the covariate matrix is written as $X_0 = (\mathbf{1}_k \; D)$. Similarly to Section 2.1, we consider the conditional tests for various alternatives to this model. In this case, since the likelihood function is written as

$$\prod_{i=1}^k \binom{n_i}{y_i} \mu_i^{y_i} (1 - \mu_i)^{n_i - y_i} = \prod_{i=1}^k \binom{n_i}{y_i} (1 - \mu_i)^{n_i} \left(\frac{\mu_i}{1 - \mu_i} \right)^{y_i}$$
$$= \prod_{i=1}^k \binom{n_i}{y_i} (1 - \mu_i)^{n_i} \exp(\beta' X_0' \mathbf{y}),$$

the nuisance parameters under the null hypothesis are $X_0'\mathbf{y}, n_1, \ldots, n_k$. Consequently, the exact conditional tests are based on the conditional distribution,

$$f(\mathbf{y} \mid X_0'\mathbf{y} = X_0'\mathbf{y}^o, n_1, \ldots, n_k) = C(X_0'\mathbf{y}^o, n_1, \ldots, n_k) \prod_{i=1}^k \frac{1}{y_i!(n_i - y_i)!}, \tag{8}$$

where $C(X_0'\mathbf{y}^o, n_1, \ldots, n_k)$ is the normalizing constant determined from $X_0'\mathbf{y}^o, n_1, \ldots, n_k$ and written as

$$C(X_0'\mathbf{y}^o, n_1, \ldots, n_k)^{-1} = \sum_{\mathbf{y} \in \mathcal{F}(X_0'\mathbf{y}^o, n_1, \ldots, n_k)} \left(\prod_{i=1}^k \frac{1}{y_i!(n_i - y_i)!} \right), \tag{9}$$

and

$$\mathcal{F}(X_0' \mathbf{y}^o, n_1, \ldots, n_k) = \{\mathbf{y} \mid X_0' \mathbf{y} = X_0' \mathbf{y}^o, \ y_i \in \{0, 1, \ldots, n_i\}, \ i = 1, \ldots, k\}. \qquad (10)$$

For notational convenience, we extend $\mathbf{y}$ to

$$\tilde{\mathbf{y}} = (y_1, \ldots, y_k, n_1 - y_1, \ldots, n_k - y_k)'$$

for the binomial model. Corresponding to this $\tilde{\mathbf{y}}$, we also extend $\nu \times k$ matrix X_0' to

$$\widetilde{X_0}' = \begin{pmatrix} X_0' & O_{\nu,k} \\ I_k & I_k \end{pmatrix}, \qquad (11)$$

where $O_{\nu,k}$ is the $\nu \times k$ zero matrix and I_k is the identity matrix of the order k. In the theory of the toric ideals, $\widetilde{X_0}'$ is called the *Lawrence lifting* of the configuration X_0'. See [15] for details. Using $\tilde{\mathbf{y}}$ and $\widetilde{X_0}'$, the condition that $X_0' \mathbf{y}$ and $n_1, \ldots, n_k$ are fixed is simply written that $\widetilde{X_0}' \tilde{\mathbf{y}}$ is fixed. Hereafter for notational simplicity, we write $\mathbf{y}$ and X_0' instead of $\tilde{\mathbf{y}}$ and $\widetilde{X_0}'$. Namely, to express the conditional distribution and its support for the binomial model, we use the expression (3)(4)(5), those for the Poisson model, instead of (8)(9)(10), respectively,

3   Markov chain Monte Carlo tests for the designed experiments

In this section, we consider the Markov chain Monte Carlo methods for calculating p values defined in Section 2. In Section 3.1, we give an explanation of Markov chain Monte Carlo methods, along with the definition of Markov bases. We also describe some algorithms and softwares, which are useful in applications. In Section 3.2, we investigate the relation between the fractional factorial designs and contingency tables.

3.1   Markov chain Monte Carlo methods for evaluating p values

To perform the exact tests defined in Section 2, a useful approach is to generate samples from the conditional distribution $f(\mathbf{y} \mid X_0' \mathbf{y} = X_0' \mathbf{y}^o)$ and calculate p values for any test statistic. In particular, when the closed form expression of the null distribution can not be obtained, a Markov chain Monte Carlo approach is a valuable tool. If a connected Markov chain over $\mathcal{F}(X_0' \mathbf{y}^o)$ is constructed, the chain can be modified to give a connected and aperiodic Markov chain with stationary distribution $f(\mathbf{y} \mid X_0' \mathbf{y} = X_0' \mathbf{y}^o)$ by the usual Metropolis procedure.

To construct a connected chain, a frequently used approach is to prepare a *Markov basis* defined in [8]. Let $\mathcal{M}(X_0')$ be the set of integer vectors which are in the kernel of X_0', i.e.,

$$\mathcal{M}(X_0') = \{\mathbf{z} = (z_1, \ldots, z_k)' \mid X_0' \mathbf{z} = \mathbf{0}, z_i \text{ is an integer for } i = 1, \ldots, k\},$$

11

where $\mathbf{0}$ is the zero vector. We call the element in $\mathcal{M}(X_0')$ a *move for X_0'*, in the sense that adding $\mathbf{z} \in \mathcal{M}(X_0')$ to any $\mathbf{y}$ does not change the sufficient statistics, i.e.,

$$X_0'(\mathbf{y} + \mathbf{z}) = X_0'\mathbf{y}.$$

An important point is that $\mathbf{y} + \mathbf{z}$ can include a negative element. On the other hand, if $\mathbf{y} + \mathbf{z} \in \mathcal{F}(X_0'\mathbf{y})$, i.e., $\mathbf{y} + \mathbf{z}$ is still a non-negative vector, we see that $\mathbf{y}$ is *moved* to $\mathbf{y} + \mathbf{z} \in \mathcal{F}(X_0'\mathbf{y})$ by $\mathbf{z}$. Now we give the definition of a Markov basis.

Definition 3.1. *A Markov basis for X_0' is a finite set of moves $\mathcal{B} = \{\mathbf{z}_1, \ldots, \mathbf{z}_L\}, \mathbf{z}_j \in \mathcal{M}(X_0'), j = 1, \ldots, L$, such that, for any $\mathbf{y}, \mathbf{y}^* \in \mathcal{F}(X_0'\mathbf{y}^o)$, there exists $A > 0$, $(\varepsilon_1, \mathbf{z}_{j_1}), \ldots, (\varepsilon_A, \mathbf{z}_{j_A})$ with $\varepsilon_s \in \{-1, +1\}$, $\mathbf{z}_{j_s} \in \mathcal{B}$, $s = 1, \ldots, A$, satisfying*

$$\mathbf{y} = \mathbf{y}^* + \sum_{s=1}^{A} \varepsilon_s \mathbf{z}_{j_s} \text{ and } \mathbf{y}^* + \sum_{s=1}^{a} \varepsilon_s \mathbf{z}_{j_s} \in \mathcal{F}(X_0'\mathbf{y}^o) \text{ for } a = 1, \ldots, A.$$

By definition, a Markov basis enables to construct a connected chain over $\mathcal{F}(X_0'\mathbf{y}^o)$, which is modified so as to have the null distribution $f(\mathbf{y} \mid X_0'\mathbf{y} = X_0'\mathbf{y}^o)$ as the stationary distribution by the Metropolis-Hastings procedure. Therefore we can perform various conditional tests by the Monte Carlo sampling. We give a simple algorithm to calculate p values for some test statistic $T(\cdot)$ based on the Markov chain Monte Carlo sampling.

```
Input: Markov basis B, observed data yᵒ, covariate matrix X₀',
       size of sample N, null distribution f(·), test statistic T(·)
Output: p value
Variables: obs, count, sig, y, y_next
```

Step 1: $\mathrm{obs} = T(\mathbf{y}^o)$. $\mathbf{y} = \mathbf{y}^o$. $\mathrm{count} = 0$. $\mathrm{sig} = 0$.
Step 2: Choose $\mathbf{z}$ from $\mathcal{B}$ randomly. $I = \{n \mid \mathbf{y} + n\mathbf{z} \in \mathcal{F}(X_0'\mathbf{y}^o)\}$.
Step 3: Select $\mathbf{y}_{next}$ from $\{\mathbf{y} + n\mathbf{z} \mid n \in I\}$ with probability

$$p_n = \frac{f(\mathbf{y} + n\mathbf{z})}{\displaystyle\sum_{n \in I} f(\mathbf{y} + n\mathbf{z})}.$$

Step 4: If $T(\mathbf{y}_{next}) \geq \mathrm{obs}$ then $\mathrm{sig} = \mathrm{sig} + 1$.
Step 5: $\mathbf{y} = \mathbf{y}_{next}$. $\mathrm{count} = \mathrm{count} + 1$.
Step 6: If $\mathrm{count} < N$ then Go to Step 2.
Step 7: Estimated p value is $\dfrac{\mathrm{sig}}{N}$

Note that we need not calculate the normalizing constant, $C(X_0'\mathbf{y}^o)$ in (4) or $C(X_0'\mathbf{y}^o, n_1, \ldots, n_k)$ in (9), of the null distribution $f(\cdot)$, since it is canceled in the numerator and denominator in Step 3.

Derivation of Markov bases is itself a very interesting problem. Markov bases can be very complicated and hard to compute for large models. Many works, including the

original work by [8], have relied on the theory of computational algebra and Gröbner bases. See [8], [9], [10]. On the other hand, a series of works by Aoki and Takemura investigates the structure of minimal Markov bases and gives some characterizations. In particular, Aoki and Takemura([4], [3], [5]) give the expression of the minimal Markov bases directly (i.e., not by using algebraic algorithm) for some problems of contingency tables.

In applications, it is most convenient for readers to rely on algebraic computational packages such as 4ti2 ([1]). For example, consider the problem we have seen in Section 2.1. For the model (2), the covariate matrix X_0' is a 10×16 matrix. To calculate the Markov basis for this X_0' by 4ti2, we only have to prepare a datafile

```
10 16
 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
 1  1  1  1  1  1  1  1 -1 -1 -1 -1 -1 -1 -1 -1
 1  1  1  1 -1 -1 -1 -1  1  1  1  1 -1 -1 -1 -1
 1  1 -1 -1  1  1 -1 -1  1  1 -1 -1  1  1 -1 -1
 1 -1  1 -1  1 -1  1 -1  1 -1  1 -1  1 -1  1 -1
 1 -1  1 -1 -1  1 -1  1 -1  1 -1  1  1 -1  1 -1
 1 -1 -1  1  1 -1 -1  1  1 -1 -1  1  1 -1 -1  1
 1 -1 -1  1 -1  1  1 -1  1 -1 -1  1 -1  1  1 -1
 1  1 -1 -1  1  1 -1 -1 -1 -1  1  1 -1 -1  1  1
 1 -1  1 -1 -1  1 -1  1  1 -1  1 -1 -1  1 -1  1
```

and run the command `markov`. Then the list of a minimal Markov basis is instantly given as

```
35 16
-1 -1  0  0  1  1  0  0  0  0  1  1  0  0 -1 -1
-1 -1  0  1  1  1 -1  0  0  0  1  0  0  0  0 -1
-1 -1  1  0  1  1  0 -1  0  0  0  1  0  0 -1  0
....
```

The above output shows that a minimal Markov bases for this X_0' consists of 35 moves, which corresponds to each row.

Using a minimal Markov basis, we perform the likelihood ratio test based on (6) in Section 2.1. After $100,000$ burn-in steps, we construct $1,000,000$ Monte Carlo samples. In contrast to the asymptotic p value 0.0040, the estimated p value is 0.032, with estimated standard deviation 0.0045, where we use a batching method to obtain an estimate of variance, see [14] and [21]. Figure 1 shows a histogram of the Monte Carlo sampling generated from the exact conditional distribution of the likelihood ratio statistic under the null hypothesis, along with the corresponding asymptotic distribution χ_6^2. Figure 1 shows that the asymptotic distribution understates the probability that the values of the test statistic for the samples is not less than the observed value, and overemphasizes the significance.
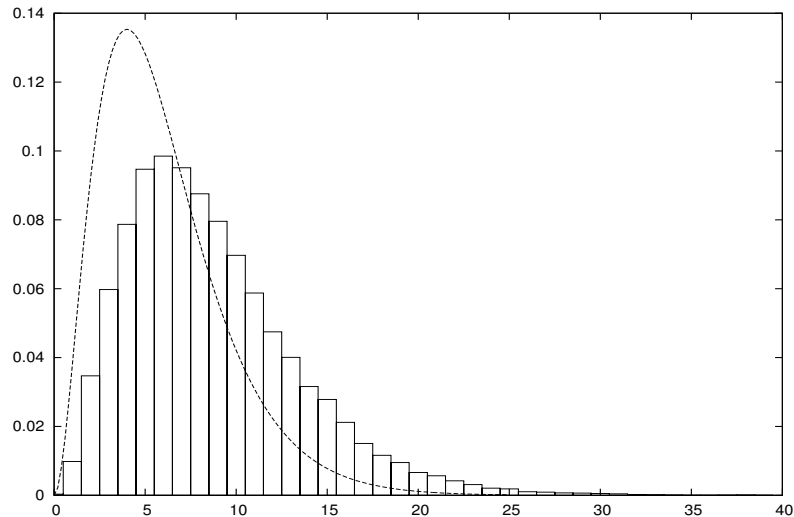
13

Figure 1: Asymptotic and Monte Carlo estimated exact distribution

Table 4: Eight-run 2^{p-q} fractional factorial designs $(p - q = 3)$

Number of factors p	Resolution	Design Generators
4	IV	D = ABC
5	III	D = AB, E = AC
6	III	D = AB, E = AC, F = BC
7	III	D = AB, E = AC, F = BC, G = ABC

3.2 Markov bases and corresponding models for 2^{p-q} contingency tables

In this section, we investigate relationships between contingency tables and fractional factorial designs with 2^{p-q} runs. As noted in Section 1, Markov bases have been mainly investigated in the context of contingency tables. For example, [6] gives an expression of minimal Markov bases of all the hierarchical models for 2^4 contingency tables. This list may be sufficient in application for the analysis of 2^4 contingency tables, since the hierarchical model is the most natural class of models to be considered. We show in this section that, considering the fractional factorial designs, we encounter some new models and Markov bases, which do not correspond to hierarchical models of contingency tables.

Fractional factorial designs with 8 runs. First, we consider fractional factorial designs with 8 runs, i.e., the case of $p - q = 3$. The most frequently used designs are listed in Table 4. We clarify the relationships between these designs and the models of 2^3 contingency tables $\mathbf{y} = (y_{ijk})$, $1 \leq i, j, k \leq 2$, for the Poisson observations, and the models of 2^4 contingency tables $\mathbf{y} = (y_{ijk\ell})$, $1 \leq i, j, k, \ell \leq 2$, for the binomial observations.

14

In the case of Poisson observations, we write eight observations as if they are the frequencies of 2^3 contingency table, i.e.,

$$\mathbf{y} = (y_{111}, y_{112}, y_{121}, y_{122}, y_{211}, y_{212}, y_{221}, y_{222})'.$$

In the case of $p = 5$, for example, the design and the observations are given as follows.

Run	A	B	C	D	E	y
1	1	1	1	1	1	y_{111}
2	1	1	2	1	2	y_{112}
3	1	2	1	2	2	y_{121}
4	1	2	2	2	1	y_{122}
5	2	1	1	2	1	y_{211}
6	2	1	2	2	2	y_{212}
7	2	2	1	1	2	y_{221}
8	2	2	2	1	1	y_{222}

(Factor header spans A B C D E)

For this type of data, we define ν-dimensional parameter β and the covariate matrix X according to an appropriate model we consider, as explained in Section 2. First consider the simple main effect model A/B/C/D/E ($\nu = 5$). To test this model against various alternatives, Markov chain Monte Carlo testing procedure needs a Markov basis for the covariate matrix

$$X_0' = \begin{pmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \end{pmatrix}.$$

Note that each row of the $X_0'\mathbf{y}$ corresponds to the sufficient statistic under the null model A/B/C/D/E. In this case, the sufficient statistic is given as

$$\begin{aligned} & y_{\cdots}, \ y_{1\cdot\cdot}, \ y_{2\cdot\cdot}, \ y_{\cdot1\cdot}, \ y_{\cdot2\cdot}, \ y_{\cdot\cdot1}, \ y_{\cdot\cdot2}, \\ & y_{11\cdot} + y_{22\cdot}, \ y_{12\cdot} + y_{21\cdot}, \ y_{1\cdot1} + y_{2\cdot2}, \ y_{1\cdot2} + y_{2\cdot1}, \end{aligned} \tag{12}$$

where we use the notations such as

$$y_{\cdots} = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} y_{ijk}, \ \ y_{i\cdot\cdot} = \sum_{j=1}^{2}\sum_{k=1}^{2} y_{ijk}, \ \ y_{ij\cdot} = \sum_{k=1}^{2} y_{ijk}.$$

Here we see that the sufficient statistic (12) is nothing but a well-known sufficient statistic for the *conditional independence model* AB/AC, given as

$$\{y_{ij\cdot}\}, \ \{y_{i\cdot k}\}, \ i, j, k = 1, 2. \tag{13}$$

The one-to-one relation between (12) and (13) is easily shown as

$$y_{ij\cdot} = \frac{y_{i\cdot\cdot} + y_{\cdot j\cdot} - (y_{ij^*\cdot} + y_{i^*j\cdot})}{2}, \ \ y_{i\cdot k} = \frac{y_{i\cdot\cdot} + y_{\cdot\cdot k} - (y_{i\cdot k^*} + y_{i^*\cdot k})}{2}, \tag{14}$$

where $\{i, i^*\}, \{j, j^*\}$ and $\{k, k^*\}$ are distinct indices, respectively. This correspondence is, of course, due to the aliasing relation D = AB, E = AC.

Next consider another model. Since there are eight observations, we can estimate eight parameters at most (in the saturated model). Since the saturated model cannot be tested, let us consider the models of seven parameters, i.e., case of $\nu = 6$. If we restrict our attention to the hierarchical models, five main effects and one of the two-factor interaction effects, BC, BE, CD, DE, can be included in the models, since the aliasing relation is given as

$$A = BD = CE, \ B = AD, \ C = AE, \ D = AB, \ E = AC,$$
$$BC = DE, \ BE = CD = ABC.$$

If our null model includes BC or DE, i.e., if our null model is written as A/BC/D/E or A/B/C/DE, we add the column

$$(+1 \ -1 \ -1 \ +1 \ +1 \ -1 \ -1 \ +1)'$$

to the covariate matrix X_0. In this case, the sufficient statistic under the null model includes $y_{\cdot 11} + y_{\cdot 22}$ and $y_{\cdot 12} + y_{\cdot 21}$ in addition to (12), which is nothing but a well-known sufficient statistic for the *no three-factor interaction model*, AB/AC/BC,

$$\{y_{ij\cdot}\}, \ \{y_{i\cdot k}\}, \ \{y_{\cdot jk}\}, \ i, j, k = 1, 2,$$

by the similar relations to (14).

On the other hand, if our null model includes BE or CD, i.e., if our null model is written as A/BE/C/D or A/B/CD/E, we have to add the column

$$(+1 \ -1 \ -1 \ +1 \ -1 \ +1 \ +1 \ -1)'$$

to the covariate matrix X_0. In this case, the sufficient statistic under the null model includes $y_{111} + y_{122} + y_{212} + y_{221}$ and $y_{112} + y_{121} + y_{211} + y_{222}$ in addition to (12). This is one of the models which do not have corresponding models in the hierarchical models of three-way contingency tables. We write this new model as AB/AC + (ABC). The sufficient statistic for this model is

$$\{y_{ij\cdot}\}, \ \{y_{i\cdot k}\}, \ y_{111} + y_{122} + y_{212} + y_{221}, \ y_{112} + y_{121} + y_{211} + y_{222}, \ i, j, k = 1, 2.$$

Similarly, we can specify the corresponding models of three-way contingency tables (for the factors A, B, C) to all the possible models for the designs of Table 4, as if the observations are the frequencies of 2^3 contingency tables. The result is summarized in Table 5.

Table 5: Eight-run 2^{p-q} fractional factorial designs and the corresponding models of three-way contingency tables $(p - q = 3)$

	Design : $p = 4$, D = ABC	
ν	Null model	Corresponding model of 2^3 table
4	A/B/C/D	A/B/C + (ABC)[a]
5	AB/C/D	AB/C + (ABC)[a]
6	AB/AC/D	AB/AC + (ABC)[a]
	Design : $p = 5$, D = AB, E = AC	
ν	Null model	Corresponding model of 2^3 table
5	A/B/C/D/E	AB/AC
6	A/BC/D/E	AB/AC/BC
	A/BE/C/D	AB/AC + (ABC)[a]
	Design : $p = 6$, D = AB, E = AC, F = BC	
ν	Null model	Corresponding model of 2^3 table
6	A/B/C/D/E/F	AB/AC/BC

[a]The sufficient statistic for (ABC) is $y_{111} + y_{122} + y_{212} + y_{221}$, $y_{112} + y_{121} + y_{211} + y_{222}$.

In the case of Binomial observations, there are 16 observations. Similarly to the Poisson case, we treat the observations as if they are the frequencies of 2^4 contingency table. In the case of $p = 5$, for example, the design and the observations are given as follows.

	Factor						
Run	A	B	C	D	E		y
1	1	1	1	1	1	y_{1111}	y_{1112}
2	1	1	2	1	2	y_{1121}	y_{1122}
3	1	2	1	2	2	y_{1211}	y_{1212}
4	1	2	2	2	1	y_{1221}	y_{1222}
5	2	1	1	2	1	y_{2111}	y_{2112}
6	2	1	2	2	2	y_{2121}	y_{2122}
7	2	2	1	1	2	y_{2211}	y_{2212}
8	2	2	2	1	1	y_{2221}	y_{2222}

For this type of data, we also specify parameter β and the covariate matrix according to the appropriate models, by replacing X by $\widetilde{X}$ of (11). Note that the elements of $\mathbf{y}$ is ordered as

$$\mathbf{y} = (y_{1111}, y_{1121}, \ldots, y_{2211}, y_{2221}, y_{1112}, y_{1122}, \ldots, y_{2212}, y_{2222})'.$$

Accordingly, correspondences to the models of 2^4 contingency tables are easily obtained and the result is given in Table 6.

Table 6: Eight-run 2^{p-q} fractional factorial designs and the corresponding models of three-way contingency tables ($p - q = 3$)

Design : $p = 4$, D = ABC		
ν	Null model	Corresponding model of 2^4 table
4	A/B/C/D	AD/BD/CD + (ABC)^a + (ABCD)^b
5	AB/C/D	ABD/CD + (ABC)^a + (ABCD)^b
6	AB/AC/D	ABD/ACD + (ABC)^a + (ABCD)^b
Design : $p = 5$, D = AB, E = AC		
ν	Null model	Corresponding model of 2^4 table
5	A/B/C/D/E	ABD/ACD + (ABC)^a
6	A/BC/D/E	ABD/ACD/BCD/ABC
	A/BE/C/D	ABD/ACD + (ABC)^a + (ABCD)^b
Design : $p = 6$, D = AB, E = AC, F = BC		
ν	Null model	Corresponding model of 2^4 table
6	A/B/C/D/E/F	ABD/ACD/BCD/ABC

^{a}The sufficient statistic for (ABC) is $\{y_{ijk\cdot}\}$, $i, j, k = 1, 2$.
^{b}The sufficient statistic for (ABCD) is $y_{111\ell} + y_{122\ell} + y_{212\ell} + y_{221\ell}$,
$y_{112\ell} + y_{121\ell} + y_{211\ell} + y_{222\ell}$, $\ell = 1, 2$.

Table 6 is automatically converted from Table 5 as follows. By definition, D is added to all the generating sets. Note also that the sufficient statistic for each model includes $\{y_{ijk\cdot}\}, 1 \leq i, j, k \leq 2$, by definition, which yields Table 6. Therefore the models which do not include all of AB, AC and BC do not correspond to hierarchical models.

Fractional factorial designs with 16 runs. Next we consider fractional factorial designs with 16 runs, i.e., the case of $p - q = 4$. Table 7 is a list of sixteen-run 2^{p-q} fractional factorial designs ($p - q = 4, p \leq 10$) from Section 4 of [25]. By the similar considerations to the 8 run cases, we can seek the corresponding models of 2^4 contingency tables for the Poisson observations, and models of 2^5 contingency tables for the Binomial observations. Since the modeling for Binomial observations can be easily obtained from the Poisson case as we have seen, we only consider the Poisson case here.

Since at most sixteen parameters are estimable for the sixteen-run designs, we can consider various models of main effects and interaction effects. For example, the saturated model of the $p = 5$ design, E = ABCD, can include all the main and two-factor interactions,

$$AB/AC/AD/AE/BC/BD/BE/CD/CE/DE.$$

Note that for the models of $p = 5, 6, 7, 8$ in Table 7, each main effect and two-factor interaction is estimable. (On the other hand, for the resolution III models of $p = 9, 10$, some of the two-factor interactions are not estimable.) Among the models which include all the main effects and some of the two-factor interaction effects, some models have

Table 7: Sixteen-run 2^{p-q} fractional factorial designs ($p - q = 4$)

Number of factors p	Resolution	Design Generators
5	V	E = ABCD
6	IV	E = ABC, F = ABD
7	IV	E = ABC, F = ABD, G = ACD
8	IV	E = ABC, F = ABD, G = ACD
		H = BCD
9	III	E = ABC, F = ABD, G = ACD
		H = BCD, J = ABCD
10	III	E = ABC, F = ABD, G = ACD
		H = BCD, J = ABCD, K = CD

the corresponding hierarchical model in the 2^4 contingency tables if we write the sixteen observations as $\mathbf{y} = \{y_{ijk\ell}\}, i, j, k, \ell = 1, 2$. For example, for the $p = 6$ design of E = ABC, F = ABD, the model of 6 main effects and 5 two-factor interaction effects,

$$AB/AC/AD/BC/BD/E/F,$$

has a corresponding model of ABC/ABD in the 2^4 contingency tables. By the aliasing relation

$$AB = CE = DF, \ AC = BE, \ AD = BF, \ AE = BC,$$
$$AF = BD, \ CD = EF, \ CF = DE = ABCD,$$

it is seen that there are $3 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 48$ distinct models such as

$$AB/AC/AD/AE/AF/E/F,$$
$$AB/AC/AD/AE/BD/E/F,$$
$$AB/AC/AD/BC/AF/E/F,$$
$$AB/AC/AD/BC/BD/E/F,$$
$$\vdots$$
$$DF/BE/BF/BC/AF/E/F,$$
$$DF/BE/BF/BC/BD/E/F,$$

which correspond to the model of ABC/ABD in the 2^4 contingency tables. By the similar considerations, we can specify all the models for the designs of Table 7 which correspond some hierarchical models in the 2^4 contingency tables. The result is shown in Table 8.

One of the merits to specify corresponding hierarchical models of contingency tables is a possibility to make use of general already known results on Markov bases of contingency tables. For example, [10] shows that a Markov basis can be constructed by the primitive moves, i.e., degree 2 moves, for the decomposable graphical models in the contingency tables. In our designed experiments, therefore, Markov basis for the models which correspond to decomposable graphical models of contingency tables can be constructed only

Table 8: Sixteen-run 2^{p-q} fractional factorial designs and the corresponding hierarchical models of 2^4 contingency tables ($p - q = 4$)

ν	Representative null model	Num. of the null models	Corresponding hierarchical model of 2^4 table
Design : $p = 6$, E = ABC, F = ABD			
11	AB/AC/AD/BC/BD/E/F	48	ABC/ABD
12	AB/AC/AD/BC/BD/CD/E/F	96	ABC/ABD/CD
Design : $p = 7$, E = ABC, F = ABD, G = ACD			
11	AB/AC/AD/BC/BD/CD/E/F/G	$3^6 = 729$	ABC/ABD/ACD
Design : $p = 8$, E = ABC, F = ABD, G = ACD, H = BCD			
11	AB/AC/AD/BC/BD/CD/E/F/G	$4^6 = 4096$	ABC/ABD/ACD/BCD

by the primitive moves. Among the results of Table 5,6 and 8, there are two models which corresponds to decomposable graphical models in the contingency tables. We can confirm that minimal Markov bases for these models consist of primitive moves as follows. We use the following notational convention for a move $\mathbf{z}$. Consider 2^3 case $\mathbf{z} = (z_{ijk})$. If $z_{i_1 j_1 k_1} = z_{i_2 j_2 k_2} = +1$, $z_{i_3 j_3 k_3} = z_{i_4 j_4 k_4} = +1$, and other elements are zeros, then we denote $\mathbf{z}$ as

$$(i_1 j_1 k_1)(i_2 j_2 k_2) - (i_3 j_3 k_3)(i_4 j_4 k_4).$$

Similar notation is used for 2^4 case.

- 2^{5-2} fractional factorial design of D = AB, E = AC:
  The main effects model A/B/C/D/E corresponds to the decomposable graphical model AB/AC of the 2^3 contingency tables. This is a conditional independence model between B and C given A and a minimal Markov basis is constructed by primitive moves as

  $$(111)(122) - (112)(121), \ (211)(222) - (212)(221).$$

- 2^{6-2} fractional factorial design of E = ABC, F = ABD:
  The model AB/AC/AD/BC/BD/E/F corresponds to the decomposable graphical model ABC/ABD of the 2^4 contingency tables. This is a conditional independence model between C and D given $\{A, B\}$ and a minimal Markov basis is again constructed by primitive moves as

  $$(1111)(1122) - (1112)(1121), \ (1211)(1222) - (1212)(1221),$$
  $$(2111)(2122) - (2112)(2121), \ (2211)(2222) - (2212)(2221).$$

For the other designs of Table 7 ($p = 5, 9, 10$), all the models include the sufficient statistic

$$y_{1111} + y_{1122} + y_{1212} + y_{1221} + y_{2112} + y_{2121} + y_{2211} + y_{2222},$$
$$y_{1112} + y_{1121} + y_{1211} + y_{1222} + y_{2111} + y_{2122} + y_{2212} + y_{2221},$$

and therefore have no corresponding hierarchical models in the 2^4 contingency tables. For example, the sufficient statistic of the the main effect models for 2^{5-1} design of E = ABCD is

$$\{y_{i\cdots}\}, \ \{y_{\cdot j\cdot\cdot}\}, \ \{y_{\cdot\cdot k\cdot}\}, \ \{y_{\cdots\ell}\}, \ i, j, k, \ell = 1, 2,$$
$$y_{1111} + y_{1122} + y_{1212} + y_{1221} + y_{2112} + y_{2121} + y_{2211} + y_{2222},$$
$$y_{1112} + y_{1121} + y_{1211} + y_{1222} + y_{2111} + y_{2122} + y_{2212} + y_{2221},$$

and the sufficient statistic of the main effect models for 2^{10-1} design of

$$\text{E = ABC, \ F = ABD, \ G = ACD, \ H = BCD, \ J = ABCD, \ K = CD}$$

is

$$\{y_{ijk\cdot}\}, \ \{y_{ij\cdot\ell}\}, \ \{y_{i\cdot k\ell}\}, \ \{y_{\cdot jk\ell}\}, \ i, j, k, \ell = 1, 2,$$
$$y_{1111} + y_{1122} + y_{1212} + y_{1221} + y_{2112} + y_{2121} + y_{2211} + y_{2222},$$
$$y_{1112} + y_{1121} + y_{1211} + y_{1222} + y_{2111} + y_{2122} + y_{2212} + y_{2221}.$$

4 Discussion

In this paper, we consider Markov chain Monte Carlo tests for the factor effects in the designed experiments. As is noted in Section 1, Markov chain Monte Carlo procedure is a valuable tool when the adequacy of traditional large-sample tests is doubtful and the enumeration of the conditional sample space is infeasible. Since a closed form expression of the null distribution for the conditional tests considered in Section 2 is not available in general, Markov chain Monte Carlo procedure is valuable in the settings of this paper. Computational experience given in Section 3.1 shows efficacy of our method.

To perform Markov chain Monte Carlo tests, it is often problematic to calculate a Markov basis. Current algorithms may take a very long time to compute Markov basis for 64 run or larger experiments. For the designs of 16 or 32 runs (for the Poisson models), a software such as 4ti2 works quite well and very practical. Nevertheless, the arguments and theoretical considerations in Section 3.2 seem important. One of the merits to specify the corresponding models of 2^p contingency tables is a possibility to make use of general results for the Markov bases of contingency tables as shown in Section 3.2.

It is also important to consider more complicated designs and give appropriate Markov bases for them, such as designs with three levels or balanced incomplete block designs.

The designed experiment is one of the areas in statistics where the applications of the theory of Gröbner basis are first considered. See the works [19], [22] and [20]. In these works, the design is represented as the variety for the set of polynomial equations. On the other hand, this manuscript gives another application of Gröbner basis theory to the designed experiments by considering Markov chain Monte Carlo tests for a discrete response variable.

References

[1] 4ti2 team. 4ti2 – a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at www.4ti2.de.

[2] Alan Agresti. A survey of exact inference for contingency tables. *Statist. Sci.*, 7(1):131–177, 1992. With comments and a rejoinder by the author.

[3] Satoshi Aoki and Akimichi Takemura. The list of indispensable moves of the unique minimal markov basis for $3 \times 4 \times k$ and $4 \times 4 \times 4$ contingency tables with fixed two-dimensional marginals. METR 2003-38, 2003. Submitted for publication.

[4] Satoshi Aoki and Akimichi Takemura. Minimal basis for a connected Markov chain over $3 \times 3 \times K$ contingency tables with fixed two-dimensional marginals. *Aust. N. Z. J. Stat.*, 45(2):229–249, 2003.

[5] Satoshi Aoki and Akimichi Takemura. Markov chain Monte Carlo exact tests for incomplete two-way contingency tables. *J. Stat. Comput. Simul.*, 75(10):787–812, 2005.

[6] Satoshi Aoki and Akimichi Takemura. Minimal invariant markov basis for sampling contingency tables with fixed marginals. *Annals of the Institute of Statistical Mathematics*, 2006. To appear.

[7] L. W. Condra. *Reliability Improvement with Design of Experiments*. Marcel Dekker, New York, NY., 1993.

[8] Persi Diaconis and Bernd Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.*, 26(1):363–397, 1998.

[9] Ian H. Dinwoodie. The Diaconis-Sturmfels algorithm and rules of succession. *Bernoulli*, 4(3):401–410, 1998.

[10] Adrian Dobra. Markov bases for decomposable graphical models. *Bernoulli*, 9(6):1093–1108, 2003.

[11] G. H. Freeman and J. H. Halton. Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika*, 38:141–149, 1951.

[12] Shelby J. Haberman. A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. *J. Amer. Statist. Assoc.*, 83(402):555–560, 1988.

[13] M. Hamada and J. A. Nelder. Generalized linear models for quality-improvement experiments. *Journal of Quality Technology*, 29:292–304, 1997.

[14] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[15] Takayuki Hibi. *Gröbner Bases*. Asakura Shoten, Tokyo, 2003. (In Japanese).

[16] B. Martin, D. Parker, and L. Zenick. Minimize slugging by optimizing controllable factors on topaz windshield molding. In *Fifth Symposium on Taguchi Methods*, pages 519–526. American Supplier Institute, Inc., Dearborn, MI, 1987.

[17] P. McCullagh and J. A. Nelder. *Generalized Linear Models, 2nd ed.* Chapman & Hall, London, 1989.

[18] Cyrus R. Mehta and Nitin R. Patel. A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *J. Amer. Statist. Assoc.*, 78(382):427–434, 1983.

[19] Giovanni Pistone and Henry P. Wynn. Generalised confounding with Gröbner bases. *Biometrika*, 83(3):653–666, 1996.

[20] Giovanni Pistonea, Eva Riccomagno, and Henry P. Wynn. *Algebraic Statistics, Computational Commutative Algebra in Statistics*. Chapman & Hall, 2000.

[21] Brian D. Ripley. *Stochastic Simulation*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1987.

[22] Lorenzo Robbiano and Maria Piera Rogantin. Full factorial designs and distracted fractions. In *Gröbner bases and applications (Linz, 1998)*, volume 251 of *London Math. Soc. Lecture Note Ser.*, pages 473–482. Cambridge Univ. Press, Cambridge, 1998.

[23] Akimichi Takemura and Satoshi Aoki. Some characterizations of minimal Markov basis for sampling from discrete conditional distributions. *Ann. Inst. Statist. Math.*, 56(1):1–17, 2004.

[24] Akimichi Takemura and Satoshi Aoki. Distance reducing markov bases for sampling from a discrete sample space. *Bernoulli*, 11(5):793–813, 2005.

[25] C. F. Jeff Wu and Michael Hamada. *Experiments: Planning, analysis, and parameter design optimization*. Wiley Series in Probability and Statistics: Texts and References Section. John Wiley & Sons Inc., New York, 2000. A Wiley-Interscience Publication.