Holonomic Gradient Method for Two Way Contingency Tables

Yoshihito Tachibana, Yoshiaki Goto, Tamio Koyama, Nobuki Takayama

June 19, 2019

Abstract

The holonomic gradient method gives an algorithm to efficiently and accurately evaluate normalizing constants and their derivatives. We apply the holonomic gradient method in the case of the conditional Poisson or multinomial distribution on two way contingency tables. We utilize the modular method in computer algebra or some other tricks for an efficient and exact evaluation, and we compare them and discuss on their implementation. We also discuss on a theoretical aspect of the distribution from the viewpoint of the conditional maximum likelihood estimation. We decompose parameters of interest and nuisance parameters in terms of sigma algebras for general two way contingency tables with arbitrary zero cell patterns.

1 Introduction

The holonomic gradient method (HGM) proposed in [17] provides an algorithm to efficiently and accurately evaluate normalizing constants and their derivatives. This algorithm utilizes holonomic differential equations or holonomic difference equations. Y. Goto and K. Matsumoto [7] determined a system of difference equations for the hypergeometric system of type (k, n). The normalizing constant of the conditional Poisson or multinomial distribution on two-way contingency tables is a polynomial solution of this hypergeometric system. Thus, we can apply these difference equations to exactly evaluate the normalizing constant and its derivatives by HGM. However, there is a difficulty: numerical evaluation errors, incurred by repeatedly applying these difference equations or recurrence relations, increase rapidly if we use floating point number arithmetic. Accordingly, we evaluate the normalizing constant by exact rational arithmetic. However, in general, exact evaluation is slow. The modular method in computer algebra (see, e.g., [19], [27]) has been used for efficient and exact evaluation over the field of rational numbers. We apply the modular method or some other tricks to our evaluation procedure. We compare these methods and explore implementation of these algorithms in Sections 4 and 5.

We then turn from computation to a theoretical question before presenting statistical applications. An interesting application of the evaluation of the normalizing constant is the conditional maximum likelihood estimation (CMLE) of parameters of interest with fixed marginal sums. Broadly speaking, the parameters of interest in this case are (generalized) odds ratios. However, we could not identify a rigorous formulation on parameters of interest for contingency tables with zero cells in the literature. In Sections 7 and 8, we introduce $\mathcal{A}$-distributions as a conditional distribution. The conditional Poisson or multinomial distribution on contingency tables with fixed marginal sums is a special and important case of $\mathcal{A}$-distributions. We will decompose parameters of interest and nuisance parameters in terms of σ-algebras. We note that the conditional distribution of a statistic given the occurrence of a sufficient statistic of a nuisance parameter does not depend on the value of the nuisance parameter. Hence, by the conditional distribution, we can estimate the parameter of interest without being affected by the nuisance parameter.

Finally, we apply our method to a CMLE problem for contingency tables. This problem is discussed in [21] for the case of $2 \times n$ contingency tables and the work presented here generalizes this to two-way contingency tables of any size and with any pattern of zero cells.

2 Two Way Contingency Tables

We introduce our notation for contingency tables and review how the normalizing constant for a conditional distribution is expressed by a hypergeometric polynomial of type (k, n). There are several salient references on contingency tables. Among them, we will refer to [1] and [10, Chap 4] herein.

2.1 $r_1 \times r_2$ Contingency Table

Definition 1 **($r_1 \times r_2$ (2 way) contingency table)** *An $r_1 \times r_2$ matrix with non-negative integer entries is called an $r_1 \times r_2$ contingency table. For a contingency table $u = (u_{ij})$, we define the row sum vector by $\beta^r = \left(\sum_j u_{1j}, \cdots, \sum_j u_{r_1 j}\right)^T$, and the column sum vector by $\beta^c = (\sum_i u_{i1}, \cdots, \sum_i u_{ir_2})^T$. A contingency table u is also written as a column vector of length $r_1 \times r_2$, denoted by u^f. The column vector obtained by joining β^r and β^c is denoted by β, which is called the row column sum vector or the marginal sum vector.*

Example 1 **(2×3 contingency table and the row sum and the column sum)** For the 2×3 contingency table $u = \begin{pmatrix} 5 & 3 & 6 \\ 7 & 2 & 4 \end{pmatrix}$ the row sum vector and the column sum vector are

$$\beta^r = \begin{pmatrix} 5+3+6=14 \\ 7+2+4=13 \end{pmatrix}, \ \beta^c = \begin{pmatrix} 5+7=12 \\ 3+2=5 \\ 6+4=10 \end{pmatrix}.$$

The corresponding vector expressions of u^f and β are

$$u^f = \begin{pmatrix} 5 & 3 & 6 & 7 & 2 & 4 \end{pmatrix}^T, \ \beta = \begin{pmatrix} 14 & 13 & 12 & 5 & 10 \end{pmatrix}^T.$$

We fix $p = (p_{ij}) \in \mathbb{R}_{>0}^{r_1 \times r_2}$, $N \in \mathbb{N}_0$ and consider the multinomial distribution

$$\frac{N! p^u}{u! |p|^N}, \ p^u = \prod_{i,j} p_{ij}^{u_{ij}}, \ u! = \prod_{i,j} u_{ij}!$$

on contingency tables satisfying $|u| = \sum_{i,j} u_{ij} = N$. The conditional distribution obtained by fixing the row sum vector β^r and the column sum vector β^c is

$$\frac{p^u}{u! Z(\beta; p)}, \qquad Z(\beta; p) = \sum_{Au^f = \beta, \ u \in \mathbb{N}_0^{r_1 \times r_2}} \frac{p^u}{u!}. \tag{1}$$

Here, the polynomial $Z(\beta; p)$ is the normalizing constant of this conditional distribution. The matrix A satisfies the following conditions: (1) entries are 0 or 1; (2) Au^f is the marginal sum vector (see Example 2). The expectation of the u-value at (i, j) of this conditional distribution is equal to

$$E[U_{ij}] = p_{ij} \frac{\partial \log Z}{\partial p_{ij}}. \tag{2}$$

Exact evaluation of the conditional probability of getting a contingency table u and evaluation of the expectation is reduced to the evaluation of the normalizing constant Z and its derivatives. For given rational numbers p_{ij}, we provide an efficient and exact method to evaluate Z and its derivatives.

Example 2 **(example of A)** When $u^f = \begin{pmatrix} 5 & 3 & 6 & 7 & 2 & 4 \end{pmatrix}^T$, the matrix A is

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

and we have

$$Au^f = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 \\ 3 \\ 6 \\ 7 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 14 \\ 13 \\ 12 \\ 5 \\ 10 \end{pmatrix} = \beta.$$

Example 3 We consider 2×2 contingency tables with the marginal sum vector $\beta = \begin{pmatrix} 5 & 7 & 8 & 4 \end{pmatrix}^T$. All contingency tables u satisfying $Au^f = \beta$ are

$$\begin{pmatrix} 5 & 0 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 4 & 3 \end{pmatrix}, \begin{pmatrix} 3 & 2 \\ 5 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 3 \\ 6 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 4 \\ 7 & 0 \end{pmatrix}.$$

These u are written as

$$\begin{pmatrix} 5 & 0 \\ 3 & 4 \end{pmatrix} + i \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \ (i = 0, 1, 2, 3, 4).$$

3 The Normalizing Constant of 2×2 Tables

It is known that the normalizing constant for the conditional distribution for $r_1 \times r_2$ tables is A-hypergeometric polynomial (see, e.g., [10, Section 6.13]). We will illustrate this correspondence for 2×2 contingency tables.

Consider the marginal sum vector $\beta = (u_{11}, u_{21} + u_{22}, u_{11} + u_{21}, u_{22})$ with $u_{ij} \geq 0$. The 2×2 contingency tables with the marginal sum vector β are

$$u = \begin{pmatrix} u_{11} & 0 \\ u_{21} & u_{22} \end{pmatrix} + i \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \ (i = 0, 1, 2, \cdots, n).$$

Here, we have $n = \min\{u_{11}, u_{22}\}$. The normalizing constant is

$$\begin{aligned} Z(\beta; p) &= \sum_{i=0}^{n} \frac{p_{11}^{u_{11}-i} p_{12}^{i} p_{21}^{u_{21}+i} p_{22}^{u_{22}-i}}{(u_{11}-i)!(i)!(u_{21}+i)!(u_{22}-i)!} \\ &= \frac{p_{11}^{u_{11}} p_{21}^{u_{21}} p_{22}^{u_{22}}}{u_{11}! u_{21}! u_{22}!} \sum_{i=0}^{n} \frac{(-u_{11})_i (-u_{22})_i}{(u_{21}+1)_i (1)_i} \left(\frac{p_{12} p_{21}}{p_{11} p_{22}} \right)^i, \end{aligned}$$

where $(a)_i = a(a+1) \cdots (a+i-1)$. Then, it can be expressed in terms of the Gauss hypergeometric function

$${}_2F_1(a, b, c; x) = \sum_{i=0}^{\infty} \frac{(a)_i (b)_i}{(c)_i (1)_i} x^i.$$

Note that when $a, b \in \mathbb{Z}_{\leq 0}$, it is a polynomial. The normalizing constant can also be expressed in terms of ${}_2F_1$ for other types of marginal sum vectors. A consequence of this observation is that we can utilize several formulae of the hypergeometric function to evaluate the normalizing constant.

4 Contiguity relation

In the previous section, we expressed the normalizing constant for 2×2 contingency tables with a fixed marginal sum vector in terms of the Gauss hypergeometric function. For $r_1 \times r_2$ contingency tables, the normalizing constant with a fixed marginal sum vector can be expressed in terms of the Aomoto-Gel'fand hypergeometric function of type $(r_1, r_1 + r_2)$ [32] (the function ${}_2F_1$ is of type $(2, 4)$). This hypergeometric function is also called the A-hypergeometric function for the product of the $(r_1 - 1)$-simplex and $(r_2 - 1)$-simplex. The difference holonomic gradient method for these hypergeometric functions utilizes contiguity relations. We illustrate this for the case of the Gauss hypergeometric function; for the general case, see [7].

Example 4 (the case of ${}_2F_1$) Put $f(a) = {}_2F_1(a, b, c; x)$ and

$$F(a) = \begin{pmatrix} f(a) \\ \theta_x f(a) \end{pmatrix}, \ M(a) = \frac{1}{a - c + 1} \begin{pmatrix} bx + a - c + 1 & x - 1 \\ -abx & a(1 - x) \end{pmatrix},$$

where θ_x is the Euler operator $x\partial_x$. Then, we have

$$F(a) = M(a) F(a + 1). \tag{3}$$

Now, note the following relations:

$$\begin{aligned} \frac{1}{a}(a+\theta_x)\bullet f(a) &= f(a+1), &(4)\\ \big(\theta_x(c-1+\theta_x)-x(a+\theta_x)(b+\theta_x)\big)\bullet f(a) &= 0. &(5)\end{aligned}$$

The first relation can be shown from the series expansion and the second relation is the Gauss hypergeometric differential equation. It follows from (4), (5) that we have

$$\begin{aligned} \frac{1}{a}(a+\theta_x)\bullet F(a) &= F(a+1),\\ \theta_x F(a) &= \begin{pmatrix} 0 & 1\\ \frac{abx}{1-x} & \frac{ax+bx-c+1}{1-x}\end{pmatrix} F(a)\\ &= A(a)F(a).\end{aligned}$$

Next, we have (3) as

$$\begin{aligned} \frac{1}{a}(a+\theta_x)\bullet F(a) &= \frac{1}{a}(aE+A(a))F(a),\\ F(a) &= \left(\frac{1}{a}(aE+A(a))\right)^{-1}F(a+1)\\ &= M(a)F(a+1),\end{aligned}$$

where E is the identity matrix.

A relation like $F(a)=M(a)F(a+1)$ is called a *contiguity relation*. In [7], the vector valued function $F(a)$ is called the *Gauss-Manin vector*.

There are several algorithms to obtain contiguity relations [31], [23], [22], [7]. Among them, we choose to use the method of twisted cohomology groups given in [7], because it is the most efficient method for the case of two-way contingency tables.

We briefly summarize the method given in [7]. Consider the hypergeometric series $f(\alpha;x)$ of type (r_1, r_1+r_2). Here, the parameter $\alpha=(\alpha_1,\ldots,\alpha_{r_1+r_2-1})$ stands for the marginal sum vector β and the variable $x=(x_{ij})_{1\le i\le r_1-1, 1\le j\le r_2-1}$ stands for p. It follows from the twisted cohomology group (a vector space spanned by equivalence classes of differential forms) associated to the integral representation of f that the contiguity relation for $\alpha_i\to\alpha_i+1$ can be obtained as follows.

We consider the twisted cohomology group H (resp. H') standing for the function $f(\alpha;x)$ (resp. $f(\alpha;x)|_{\alpha_i\to\alpha_i+1}$). Both twisted cohomology groups are of dimension $r=\binom{r_1+r_2-2}{r_1-1}$. We take a basis $\varphi_1,\ldots,\varphi_r$ of H such that the "integral" of $(\varphi_1,\ldots,\varphi_r)^T$ gives a constant multiple of the Gauss-Manin vector

$$F(\alpha;x)=(f(\alpha;x),\delta^{(2)}\bullet f(\alpha;x),\ldots,\delta^{(r)}\bullet f(\alpha;x))^T,$$

where $\delta^{(i)}$ is some differential operator with respect to $x=(x_{ij})$. There exist a basis $\varphi'_1,\ldots,\varphi'_r$ of H' and a linear map $\mathcal{U}_i: H'\to H$ such that the integral of $\big(\mathcal{U}_i(\varphi'_1),\ldots,\mathcal{U}_i(\varphi'_r)\big)^T$ gives a constant multiple of the shifted Gauss-Manin vector $F(\alpha;x)|_{\alpha_i\to\alpha_i+1}$. Let $U_i(\alpha;x)$ be a representation matrix of $\mathcal{U}_i$ with respect to the bases $\{\varphi'_i\}$ and $\{\varphi_j\}$:

$$\big(\mathcal{U}_i(\varphi'_1),\ldots,\mathcal{U}_i(\varphi'_r)\big)^T=U_i(\alpha;x)\cdot(\varphi_1,\ldots,\varphi_r)^T.$$

Integrating both sides, we thus obtain the contiguity relation

$$F(\alpha;x)|_{\alpha_i\to\alpha_i+1}=\tilde{U}_i(\alpha;x)F(\alpha;x),$$

where $\tilde{U}_i$ is a constant multiple of U_i. It turns out that the representation matrix U_i can be expressed in terms of a simple diagonal matrix and base transformation matrices which can be obtained by evaluating intersection numbers among differential forms. The contiguity relation for $\alpha_i\to\alpha_i-1$ can be derived analogously. For more details, see [7]. Here, we illustrate this method in the case of ${}_2F_1$.

Example 5 (the case of $_2F_1$ ($r_1 = r_2 = 2$, $r = 2$)) For the parameter (a, b, c) of $_2F_1$, we put

$$(\alpha_1, \alpha_2, \alpha_3) = (b, -a, c - b - 1).$$

Here, we set $\alpha_0 = -\alpha_1 - \alpha_2 - \alpha_3 = a - c + 1$ for convenience. Since the move $a + 1 \to a$ corresponds to $\alpha_2 - 1 \to \alpha_2$ (and $\alpha_0 + 1 \to \alpha_0$) in the new parametrization, the matrix $M(a)$ in Example 4 stands for $U_2(\alpha; x)$. The representation matrix U_2 has the following decomposition.[1]

$$\begin{aligned} U_2 = & \frac{\alpha_1(\alpha_2 - 1)}{\alpha_3} \begin{pmatrix} \frac{1}{\alpha_0} + \frac{1}{\alpha_1} & \frac{1}{\alpha_0} \\ \frac{1}{\alpha_0} & \frac{1}{\alpha_0} + \frac{1}{\alpha_2} \end{pmatrix} \begin{pmatrix} \alpha_1 & -\alpha_1 \\ 0 & -\alpha_2 \end{pmatrix} \\ & \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 - x \end{pmatrix} \begin{pmatrix} \frac{1}{\alpha_0+1} + \frac{1}{\alpha_1} & \frac{1}{\alpha_0+1} \\ \frac{1}{\alpha_0+1} & \frac{1}{\alpha_0+1} \end{pmatrix} \begin{pmatrix} \frac{\alpha_1+\alpha_3}{\alpha_2-1} & 1 \\ 1 & \frac{\alpha_2-1+\alpha_3}{\alpha_1} \end{pmatrix}. \end{aligned}$$

The matrices except the diagonal matrix $\mathrm{diag}(1, 1 - x)$ are expressed by intersection numbers. Since we have $\delta^{(2)} = \frac{1}{\alpha_2}\theta_x$, the matrix U_2 has a small difference with $M(a)$ in Example 4 and we obtain $M(a)$ by adjusting the scale factor $1/\alpha_2$ of θ_x.

By the contiguity relation, we can evaluate the normalizing constant Z and its derivatives. Let us explain the procedure for the case of $_2F_1$.

Suppose $a \in \mathbb{Z}_{<-1}$. By the contiguity relation (3), we have

$$\begin{aligned} F(a) &= M(a)F(a + 1) \\ &= M(a)M(a + 1)F(a + 2) \\ &\vdots \\ &= M(a)M(a + 1) \cdots M(-2)F(-1). \end{aligned} \tag{6}$$

Then, we can obtain the value of $F(a)$ from the initial value $F(-1) = (1 - \frac{b}{c}x, -\frac{b}{c}x)^T$ by applying linear transformations. Values of the normalizing constant and its derivatives can be obtained from $F(a)$ with the differential equation for the Gauss hypergeometric function. This method is called the *difference holonomic gradient method* (difference HGM) and can be generalized to the case of $r_1 \times r_2$ contingency tables with the Gauss-Manin vector and contiguity relations given in [7].

We note that a naive evaluation of the polynomial Z is very slow. For example, the polynomial Z of the 2×5 contingency table with the row sum $(4n, 5n)$, the column sum $(5n, n, n, n, n)$ and $p = \begin{pmatrix} 1 & 1/2 & 1/3 & 1/5 & 1/7 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$ can be expressed in terms of the Lauricella function $F_D(-4n; -n, -n, -n, -n; n+1; 1/2, 1/3, 1/5, 1/7)$ of 4 variables (see, e.g., [8]). The number of terms is $O(n^4)$. Here is a comparison of the naive summation of F_D and our HGM implementation discussed in the next section.

n	20	30	40
Naive summation (in seconds)	16.0	111.7	456.6
HGM (in seconds)	0.28	0.276	0.284

Thus, the HGM is worth researching.

We briefly introduce an algorithm of difference HGM for $r_1 \times r_2$ contingency tables. The following algorithm computes the Gauss-Manin vector $F(\beta; p)$ which is essentially same as $F(\alpha; x)$ in the above (for the correspondence between $(\beta; p)$ and $(\alpha; x)$, see [7, Proposition 7.1]). In fact, we give improvement of Step 2–4 of [7, Algorithm 7.8].

Algorithm 1 (A modified version of [7, Step 1–4 of Algorithm 7.8])

Input: $\beta = (\beta_1^{(1)}, \ldots, \beta_{r_1}^{(1)}; \beta_1^{(2)}, \ldots, \beta_{r_2}^{(2)})$*: a marginal sum vector,* $p = (p_{ij}) \in \mathbb{Q}_{>0}^{r_1 \times r_2}$*: probabilities of the cells.*

Output: the Gauss-Manin vector $F(\beta; p)$ *(which is a vector of size* $r = \binom{r_1+r_2-2}{r_1-1}$*).*

1. *Set* $B_0 = (1, \ldots, 1, \beta_1^{(1)} + \cdots + \beta_{r_1}^{(1)} - r_1 + 1; \beta_1^{(2)}, \ldots, \beta_{r_2}^{(2)})$*. Compute* $F(B_0; p)$ *by the definition. (In this case, the normalizing constant* $Z(B_0; p)$ *is a polynomial of small degree, and hence the Gauss-Manin vector* $F(B_0; p)$ *is easily computed.)*

[1]see the appendix (Section 10) for more details.

2. *For $k = 1, \dots, r_1 - 1$, define B_k inductively as $B_k = B_{k-1} + (\beta_k^{(1)} - 1) \cdot \delta_k$, where*

$$\delta_k = (0, \dots, 0, \underset{k\text{-th}}{1}, 0, \dots, 0, -1; 0, \dots, 0)$$

(note that B_{r_1-1} is β). Evaluate the contiguity matrices $C_k(t)$ that satisfy

$$F(B_{k-1} + (T+1)\delta_k; p) = C_k(T) \cdot F(B_{k-1} + T\delta_k; p), \quad T = 0, 1, \dots, \beta_k^{(1)} - 2.$$

Here, t is an indeterminate and each entry of $C_k(t)$ is an element of $\mathbb{Q}(t)$.

3. *For $k = 1, \dots, r_1 - 1$, compute $F(B_k; p)$ inductively as*

$$F(B_k; p) = C_k(\beta_k^{(1)} - 2) \cdots C_k(1) C_k(0) F(B_{k-1}; p). \tag{7}$$

4. *Return $F(B_{r_1-1}; p)$.*

By using $F(\beta; p)$, we can compute the normalizing constant $Z(\beta; p)$ and the expectations $E[U_{ij}]$ (see [7, Step 5–7 of Algorithm 7.8]).

Example 6 (cf. [7, Example 7.10]) *We consider 3×3 contingency tables whose marginal sum vector is $\beta = (2, 3, 3; 1, 3, 4)$. In this case, the Gauss-Manin vector is of size $\binom{3+3-2}{3-1} = 6$.*

1. *We set $B_0 = (1, 1, 6; 1, 3, 4)$, and compute $F(B_0; p)$ by the definition. In this case, the normalizing constant $Z(B_0; p)$ has only eight terms.*

2. *We set $B_1 = (2, 1, 5; 1, 3, 4)$, $B_2 = (2, 3, 3; 1, 3, 4)(= \beta)$. By using notations in [7], we put*

$$C_1(t) = U_1^{-1}(-5+t, -2-t, -1, 3, 4, 1; x), \quad C_2(t) = U_2^{-1}(-4+t, -2, -2-t, 3, 4, 1; x).$$

Here, $x \in \mathbb{Q}^{(r_1-1)\times(r_2-1)}$ is defined from p. We have

$$\begin{aligned} &C_1(0)F(1,1,6;1,3,4;p) = F(2,1,5;1,3,4;p), \\ &C_2(0)F(2,1,5;1,3,4;p) = F(2,2,4;1,3,4;p), \quad C_2(1)F(2,2,4;1,3,4;p) = F(2,3,3;1,3,4;p). \end{aligned}$$

3. *We compute the product*

$$\begin{aligned} C_2(1)C_2(0)C_1(0)F(B_0;p) &= C_2(1)C_2(0)C_1(0)F(1,1,6;1,3,4;p) \\ &= C_2(1)C_2(0)F(2,1,5;1,3,4;p) \ (= C_2(1)C_2(0)F(B_1;p)) \\ &= C_2(1)F(2,2,4;1,3,4;p) \\ &= C_2(1)F(2,3,3;1,3,4;p) \ (= F(B_2;p)). \end{aligned}$$

4. *We obtain the Gauss-Manin vector $F(B_2; p) = F(\beta; p)$.*

For example, when $p = \begin{pmatrix} 1 & 1/2 & 1/3 \\ 1 & 1/5 & 1/7 \\ 1 & 1 & 1 \end{pmatrix}$, the 6×6 matrix $C_2(t)$ is given as follows[2].

$$C_2(t) = \begin{pmatrix} \frac{-(35t+29)}{35(t+2)} & \frac{12}{5(t+2)} & \frac{24}{7(t+2)} & \frac{-12}{5(t+2)} & \frac{-24}{7(t+2)} & 0 \\ \frac{1}{5} & \frac{-1}{5} & 0 & \frac{1}{5} & 0 & 0 \\ \frac{1}{7} & 0 & \frac{-1}{7} & 0 & \frac{1}{7} & 0 \\ \frac{-8}{5(t+2)} & \frac{8}{5(t+2)} & 0 & \frac{21t-73}{35(t+2)} & \frac{-88}{35(t+2)} & \frac{88}{35(t+2)} \\ \frac{-6}{7(t+2)} & 0 & \frac{6}{7(t+2)} & \frac{-33}{35(t+2)} & \frac{10t-47}{35(t+2)} & \frac{-33}{35(t+2)} \\ 0 & 0 & 0 & \frac{-1}{35} & \frac{1}{35} & \frac{-1}{35} \end{pmatrix}.$$

[2] It is obtained by our program `gtt_ekn3` as

```
gtt_ekn3.downAlpha3(2,2,2 | arule=gtt_ekn3.alphaRule_num([-5+t,-2,-1-t,3,4,1],2,2),
                           xrule=gtt_ekn3.xRule_num([[1,1/2,1/3],[1,1/5,1/7],[1,1,1]],2,2)).
```

Remark 1 *The algorithm given in [7] requires more matrix multiplications than Algorithm 1. As [7, Example 7.10], the former algorithm computes the above $F(2,3,3;1,3,4;p)$ by nine matrix multiplications (each "$\mapsto$" means one multiplication):*

$$
\begin{aligned}
&F(1,1,2;2,1,1;p) \mapsto F(1,1,3;2,2,1;p) \mapsto F(1,1,4;2,3,1;p) \\
&\mapsto F(1,1,5;2,3,2;p) \mapsto F(1,1,6;2,3,3;p) \mapsto F(1,1,7;2,3,4;p) \\
&\mapsto F(1,1,6;1,3,4;p) \mapsto F(2,1,5;1,3,4;p) \mapsto F(2,2,4;1,3,4;p) \mapsto F(2,3,3;1,3,4;p).
\end{aligned}
$$

On the other hand, Algorithm 1 needs only the last three steps.

We give the complexity to construct the matrix $C_k(t)$. The appendix (Section 10) will help to follow the following argument. By [7, Theorem 5.3], the matrix $U_k^{\pm 1}$ for the contiguity relation is the product of five matrices of size $r = \binom{r_1+r_2-2}{r_1-1} = \frac{(r_1+r_2-2)!}{(r_1-1)!(r_2-1)!}$:

(a) one diagonal matrix whose entries are rational functions in p,

(b) two intersection matrices whose entries are rational functions in β,

(c) two inverse matrices of intersection matrices

(cf. Example 5). For U_k^{-1}, by substituting

- $\beta_k^{(1)}$ and $\beta_{r_1}^{(1)}$ with certain polynomials in t of degree 1,
- the other $\beta_j^{(i)}$'s and p with certain rational numbers,

we obtain the matrix $C_k(t)$. By this construction and the formula for (a), (b), (c) in [7], it turns out that when we construct $C_k(t)$, we treat rational functions in t whose denominator and numerator are of degree at most 12. As long as we have tried on a computer for cases $5 \times r_i$, $r_i \leq 12$, the degrees of numerators and denominators are much smaller than 12 and no big number (large number so that FFT multiplication algorithms are used) appears in the matrix $C_k(t)$; when we use the modular method, all numbers in the matrix are elements in a finite field. Thus, we assume in the following theorem that the complexity of arithmetics of polynomials in one variable is $O(1)$.

Theorem 1 *Let $r_1, r_2 \geq 2$. Assume that the complexity of arithmetics is $O(1)$, the complexities of multiplying two $n \times n$ matrices and evaluating the determinant of an $n \times n$ matrix are $O(n^\omega)$ for some $2 \leq \omega < 3$. The complexity of obtaining the matrix $C_k(t)$ in Algorithm 1 for $r_1 \times r_2$ contingency tables is $O(r^\omega)$, where $r = \binom{r_1+r_2-2}{r_1-1}$. Especially, it is*

1. *$O(r_2^{\omega r_1})$ when r_1 is fixed,*
2. *$O(r_1^{\omega r_2})$ when r_2 is fixed,*
3. *$O(2^{2\omega r_1})$ when $r_1 = r_2$.*

Proof As explained later, the complexity to construct the above matrices (a), (b) and (c) are $O(r_1^\omega r)$, $O(r_1^2 r^2)$ and $O(r_1^2 r^2)$, respectively. Since the size of each matrix is r, the complexity of multiplication is $O(r^\omega)$. Thus, the complexity to obtain a contiguity relation is $O(r^\omega) + O(r_1^\omega r) + O(r_1^2 r^2)$. Since r is larger than r_1^2 in general, the complexity is equal to $O(r^\omega)$.

1. We fix r_1 and assume $r_2 \gg r_1$. By the Stirling formula $\log n! \sim n \log n - n$, we have

$$
\begin{aligned}
\log r &\sim (r_1 + r_2)\log(r_1 + r_2) - r_2 \log r_2 \\
&= r_1 \log r_2 + r_1 \log\left(1 + \frac{r_1}{r_2}\right) + r_2 \log\left(1 + \frac{r_1}{r_2}\right) \sim r_1 \log r_2.
\end{aligned}
$$

Then we obtain $r \sim r_2^{r_1}$ and the complexity is $O(r_2^{\omega r_1})$.

2. Claim 2 can be obtained by a similar argument to Claim 1.

3. If $r_1 = r_2$, then by the Stirling formula, we have

$$\log r \sim 2r_1 \log 2r_1 - 2r_1 \log r_1 = 2r_1 \log 2,$$

which implies $r \sim 2^{2r_1}$. Thus, the complexity is $O(2^{2\omega r_1})$.

Now, we explain the complexity of obtaining the matrices (a), (b), (c).

(a) As [7, Theorem 5.3], each nonzero entry of the diagonal matrix is the ratio of determinants of two $r_1 \times r_1$ matrices. Thus the complexity of evaluation is $O(r_1^{\omega} r)$.

(b) The entries of intersection matrices are intersection numbers of $(r_1 - 1)$-th twisted cohomology groups, which can be evaluated by the formula in [7, Fact 3.2]. The complexity of evaluating an intersection number by this formula is $O(r_1^2)$, and hence the complexity of obtaining the intersection matrix is $O(r_1^2 r^2)$.

(c) By the proof of [7, Proposition A.1], the inverse matrix of an intersection matrix is expressed as a product of two diagonal matrices and one intersection matrix. The complexity of obtaining the diagonal matrices is $O(r_1 r)$, since that of their nonzero entry is $O(r_1)$. Therefore, the complexity of obtaining the inverse matrix of the intersection matrix is dominated by the complexity $O(r_1^2 r^2)$ of obtaining the intersection matrix.

□

In this section we conducted a complexity analysis of the method for obtaining the contiguity relation. The theoretical complexity is of a polynomial order when r_i is fixed and our implementation shows that this step is efficient for small sized contingency tables. However, a naive evaluation of the composition of linear transformations (6) is slow, even for small contingency tables, because of large numbers when $|a|$ is large.

5 Efficient Evaluation of a Composition of Linear Transformations

To perform exact and efficient evaluations by the difference HGM, we need a fast and exact evaluation of a composition of linear transformations for vectors with rational number entries. This problem has hitherto been explored and there are several implementations, e.g., LINBOX [15]. For the purposes of empirical application, we study several methods to evaluate the composition of linear transformations such as (6) or (7). Our implementation is published as the package `gtt_ekn3` for Risa/Asir [26]. The function names in this section are those in this package.

5.1 Our Benchmark Problems

In order to compare several methods, we will use the following 4 benchmark problems. The timing data are taken on a machine with

CPU	Intel(R) Xeon(R) CPU E5-4650 2.70GHz
the number of CPU's	32
the number of cores	8
OS	Debian 9.8
memory	256GB
software system	Risa/Asir (2018) version 20190328 with GMP [33]

Benchmark Problem 1 Evaluate

$$f = {}_2F_1\left(-36N, -11N, 2N; \frac{1-\frac{1}{N}}{56}\right), \ N \in \mathbb{N}.$$

It stands for the 2×2 contingency tables with the row sums $(36N, 13N - 1)$ and the column sums $(38N - 1, 11N)$. The parameter (p_{ij}) is set to $\begin{pmatrix} 1 & \frac{1-1/N}{56} \\ 1 & 1 \end{pmatrix}$.

Benchmark Problem 2 Evaluate the expectation for the 3×5 contingency tables with the row sums $(N, 2N, 12N)$, the column sums $(N, 2N, 3N, 4N, 5N)$, and the parameter p

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{5} & \frac{1}{7} \\ 1 & \frac{1}{11} & \frac{1}{13} & \frac{1}{17} & \frac{1}{19} \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Benchmark Problem 3 Evaluate the expectation for the 5×5 contingency tables with the row sums $(4N, 4N, 4N, 4N, 4N)$, the column sums $(2N, 3N, 5N, 5N, 5N)$, and the parameter p

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{5} & \frac{1}{7} \\ 1 & \frac{1}{11} & \frac{1}{13} & \frac{1}{17} & \frac{1}{19} \\ 1 & \frac{1}{23} & \frac{1}{29} & \frac{1}{31} & \frac{1}{37} \\ 1 & \frac{1}{37} & \frac{1}{41} & \frac{1}{43} & \frac{1}{47} \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Benchmark Problem 4 Evaluate the expectation for the 7×7 contingency tables with the row sums $(N, 2N, 3N, 4N, 5N, 6N, 7N)$, the column sums $(N, 2N, 3N, 4N, 5N, 6N, 7N)$, and the parameter

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{5} & \frac{1}{7} & \frac{1}{11} & \frac{1}{13} \\ 1 & \frac{1}{17} & \frac{1}{19} & \frac{1}{23} & \frac{1}{29} & \frac{1}{31} & \frac{1}{37} \\ 1 & \frac{1}{41} & \frac{1}{43} & \frac{1}{47} & \frac{1}{53} & \frac{1}{59} & \frac{1}{61} \\ 1 & \frac{1}{67} & \frac{1}{71} & \frac{1}{73} & \frac{1}{79} & \frac{1}{83} & \frac{1}{89} \\ 1 & \frac{1}{97} & \frac{1}{101} & \frac{1}{103} & \frac{1}{107} & \frac{1}{109} & \frac{1}{113} \\ 1 & \frac{1}{127} & \frac{1}{131} & \frac{1}{137} & \frac{1}{139} & \frac{1}{149} & \frac{1}{151} \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

5.2 Floating Point Arithmetic

If we can evaluate the composition of linear transformations (7) accurately over floating point numbers, we can utilize GPU's or other hardware for efficient evaluation. Unfortunately, we lose the precision during the iteration of linear transformations in general. For example, let us evaluate the case of $N = 100$ for our 2×2 benchmark problem 1 with double arithmetic. The output by the double precision floating point arithmetic is `4.08315e+94`, but the answer is `4.48194745579962e+94` where we use the double value expression in the standard form, e.g., `4.08e+94` means 4.08×10^{94}. The output by double has only one digit of accuracy.

5.3 Intermediate Swell of Integers

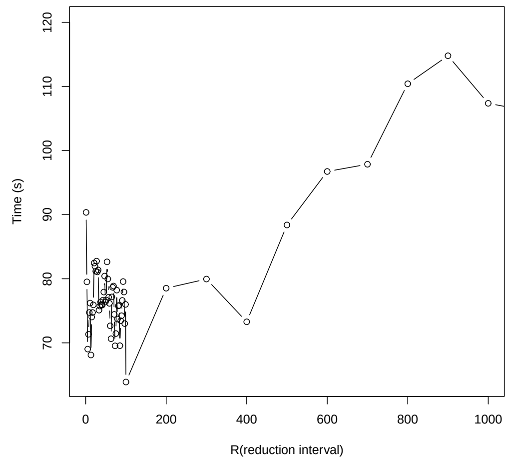
We denote by $M(n)$ the complexity of the multiplication of two n-digits integers. The book [4] is a survey on algorithms and complexities on integer arithmetic.

Arithmetic over $\mathbb{Q}$ is more expensive than arithmetic over $\mathbb{Z}$, because the reduction of a rational number needs the computation of GCD of the numerator and the denominator. The best known complexity of the operation of GCD is $O(M(n) \log n)$ for two n-digits numbers (see, e.g., [16], [4]). The complexity of the Euclidean algorithm for GCD is $O(n^2)$ [3].

One way to avoid reductions in $\mathbb{Q}$ in our interations of linear transformations (7) is to evaluate numerators and denominators separately and compute the GCD of the numerator and the denominator every R step of the linear transformations. We will call this sequential method `g_mat_fac_int` (generalized matrix factorial over integers). A reduction performing in every R step is necessary. In fact, our evaluation problems make intermediate swell of integers by the method `g_mat_fac_int`. For example, the table below shows sizes of the numerators and the denominators by the separate evaluation without the intermediate reduction in our benchmark problem 1;

N	digits of num./den.	digits of num./den. after reduction	time
300	$1.97 \times 10^5/1.96 \times 10^5$	$3.35 \times 10^4/3.28 \times 10^4$	0.92s
500	$3.47 \times 10^5/3.47 \times 10^5$	$5.87 \times 10^4/5.76 \times 10^4$	1.56s

[3] Timing data over $\mathbb{Q}$ in the version 1 of this paper at arxiv is very slow, because asir 2000 uses the Euclidean algorithm for the reductions in $\mathbb{Q}$ as default. The system asir 2018 based on GMP uses faster GCD algorithms as default.



R (reduction interval)	1	3	5	7	9	11	13	15
Time(s)	90.352	79.5147	69.024	71.335	74.7312	76.2025	68.1058	74.0283

Figure 1: Intermediate reduction

After the reduction, the numerators and the denominators become smaller as shown in the second column of the table.

We have no theoretical estimate for the best choice of R for intermediate reductions. The Figure 1 is timing data of our benchmark problem 2 with $N = 100$. The horizontal axis is the interval R of the intermediate reduction and the vertical axis is the timing. The graph indicates that we should choose R such that $5 \leq R \leq 100$.

5.4 Multimodular Method

It may be standard to use the modular method when we have an intermediate swell of integers. We refer to, e.g., [11] and its references for the complexity analysis on modular methods.

Algorithm 2 (`g_mat_fac_itor` (generalized matrix factorial by itor), modular method) [4]

Input: $M(k)$ (matrix), F (vector), $S < E$ (indices), P_{list} (a list of prime numbers), C_{list} (a list of processes for a distributed computation).

Output: A candidate value of $M(E)\cdots M(S+2)M(S+1)M(S)F$ or "failure".

1. *Let F_n, F_d (scalar), M_n, M_d (scalar) be numerators and denominators of F and M respectively.*
2. *For each prime number P_i in P_{list}, perform the linear transformations $\prod_{i=0}^{E-S}(M_n(S+i)M_d(S+i)^{-1})F_nF_d^{-1}$ of F over $\mathbb{F}_{P_i}$. If the integer F_d or M_d is not invertible modulo P_i (unlucky case), then skip this prime number P_i and set P_{list} to $P_{\text{list}} \setminus \{P_i\}$. Let the output be G_i. This step may be distributed to processes in the C_{list}.*
3. *Apply the Chinese remainder theorem to construct a vector G over $\mathbb{Z}/P\mathbb{Z}$ satisfying $G \equiv G_i \mod P_i$ where $P = \prod_{P_i \in P_{\text{list}}} P_i$.*
4. *Return a candidate value by the procedure `IntegerToRational`(G, P) (rational reconstruction).*

The complexity of the modular method `g_mat_fac_itor` is estimated as follows.

[4] We use "itor" as an abbreviation of the procedure `IntegerToRational`.

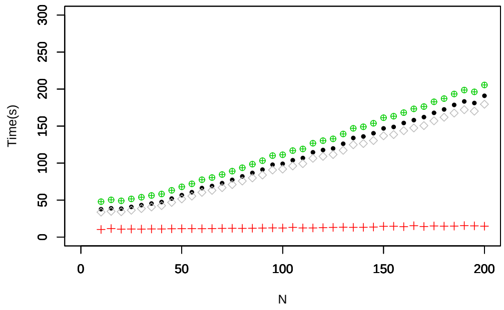


Figure 2: 5×5 contingency table, the benchmark problem 3 with 32 processes

Theorem 2 *Let n be the number of the linear transformations and the size of the square matrix $r = \binom{r_1+r_2-2}{r_1-1}$. Suppose that each prime number P_i is d_p digits number and we use N_p prime numbers. C is the number of processes. The complexity of* `g_mat_fac_itor` *is approximated as*

$$\max\left\{O\left(\frac{nr^2N_pM(d_p)}{C}\right), O\left(r(d_pN_p)^2\right)\right\}$$

when n is in a bounded region where the rational reconstruction succeeds and the asymptotic complexity of the Chinese remainder theorem approximates well the corresponding exact complexity in the region.

Proof We estimate the complexity of each step of `g_mat_fac_itor`.

1. The complexity of one linear transformation is $O(r^2M(d_p))$. The linear transformation is performed n times for N_p prime numbers. Then the complexity is $O(nr^2N_pM(d_p))$ on a single process. This step can be distributed into C processes, then the complexity is $O(\frac{nr^2N_pM(d_p)}{C})$.

2. The complexity to find an integer x such that $x \equiv x_i \mod p_i$ $(i = 1, \ldots, N_p)$ is discussed in [11, Theorem 6] under the assumption that an inborn FFT scheme is used. It follows from the estimate that the reconstruction complexity $C_n(N_p)$ of N_p primes of d_p digits is bounded by $(2/3 + o(1))M(d_pN_p)\max\left(\frac{\log N_p}{\log\log(d_pN_p)}, 1 + O(N_p^{-1})\right)$

3. The rational reconstruction algorithm `IntegerToRational`, see, e.g., [6], [20], is a variation of the Euclidean algorithm and its complexity is bounded by $O((N_pd_p)^2)$. We have r numbers to reconstruct.

Since the complexity of the step 2 is smaller than other parts, we obtain the conclusion.

The complexity is linear with respect to n (which is proportional to the size of the marginal sum vector in our benchmark problems) when the first argument of the "max" in the theorem is dominant. However, when n becomes larger, the rational reconstruction fails or gives a wrong answer. This is the reason why we give the assumption that n is in a bounded region. Note that the complexity estimate in the theorem is not an asymptotic complexity and is an approximate evaluation of it.

Let us present an example that this approximate evaluation works. Figure 2 is a graph of the timing data for the benchmark problem 3 with $N_p = 400$ and $d_p = 100$ by the decimal digits. The top point graph is the total time, the second top point graph is the time of the generalized matrix factorial (the execution time of Algorithm 2), the third point graph is the time of the distributed generalized matrix factorial by modulo P_i's (the step 2 of Algorithm 2). The last point graph is the time to obtain contiguity relations. Contiguity relations for several directions are obtained by distributing the procedures into 32 processes. Note that the point graph is linear with respect to N, which is proportional to the number of the linear transformations n. The timing data imply that the first argument of "max" of Theorem 2 is

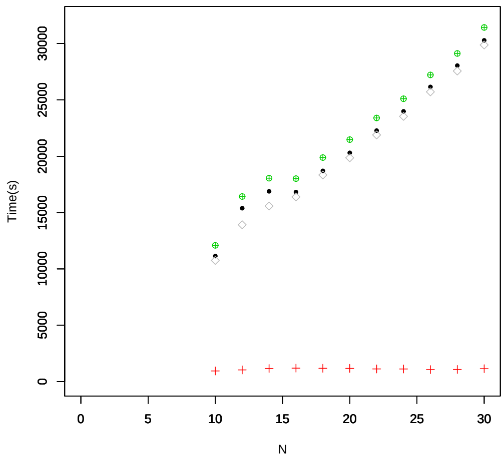


Figure 3: 7×7 contingency table, the benchmark problem 4 with 32 processes

dominant in this case. In fact, when $N = 200$, the step for reconstructing rational numbers only takes about 8 seconds and linear transformations over finite fields take from 35 seconds to 52 seconds.

We should ask if our multimodular method is efficient on real computer environments. The following table is a comparison of timing data of the sequential method `g_mat_fac_int` (with a distributed computation of contiguity relations by 32 processors) and the multimodular method `g_mat_fac_itor` by 32 processors for the benchmark problem 3.

N	90	200
`g_mat_fac_int` with the reduction interval $R = 100$	21.57	45.40
`g_mat_fac_int` without the intermediate reduction	68.17	227.23
`g_mat_fac_itor` by 32 processors	103.23	205.57

Unfortunately, the multimodular method is slower than the sequential method `g_mat_fac_int` with a relevant choice of R on our best computer, however it is faster than the case of a bad choice of $R = \infty$.

When the size of contingency table becomes larger, the rank r becomes larger rapidly. For example, $r = 20$ for the 5×5 contingency tables and $r = 924$ for the 7×7 contingency tables. The Figure 3 shows timing data of our benchmark problem 4 of 7×7 contingency tables with the multimodular method by 32 processors. We can also see linear timing with respect to N, but the slope is much larger than the 5×5 case as shown in our complexity analysis.

5.5 Binary Splitting Method

It is well-known that the binary splitting method for the evaluation of the factorial $m!$ of a natural number m is faster method than a naive evaluation of the factorial by $m! = m \times (m-1)!$. The binary splitting method evaluates $m(m-1)\cdots(\lfloor m/2 \rfloor + 1)$ and $\lfloor m/2 \rfloor(\lfloor m/2 \rfloor - 1)\cdots 1$ and obtains $m!$. This procedure can be recursively executed. This binary splitting can be easily generalized to our generalized matrix factorial; we may evaluate, for example, $M(a)M(a+1)\cdots M(\lfloor a/2 \rfloor - 1)$ and $M(\lfloor a/2 \rfloor)\cdots M(-2)$ to obtain $M(a)M(a+1)\cdots M(-2)$, $a < -2$ in (6). This procedure can be recursively applied. However, what we want to evaluate is the application of the matrix to the vector $F(-1)$. The matrix multiplication is slower than the linear transformation. Then, we cannot expect that this method is efficient for our problem. However, when the size of the matrix is relatively small, there are cases that the binary splitting method is faster. Here is an output by our package `gtt_ekn3.rr`.

```
[1828] import("gtt_ekn3.rr")$
[4014] cputime(1)$
0sec(1.001e-05sec)
[4015] gtt_ekn3.expectation(Marginal=[[1950,2550,5295],[1350,1785,6660]],
                   P=[[17/100,1,10],[7/50,1,33/10],[1,1,1]]|bs=1)$ //binary splitting
3.192sec(3.19sec)
[4016] gtt_ekn3.expectation(Marginal,P)$
4.156sec(4.157sec)
```

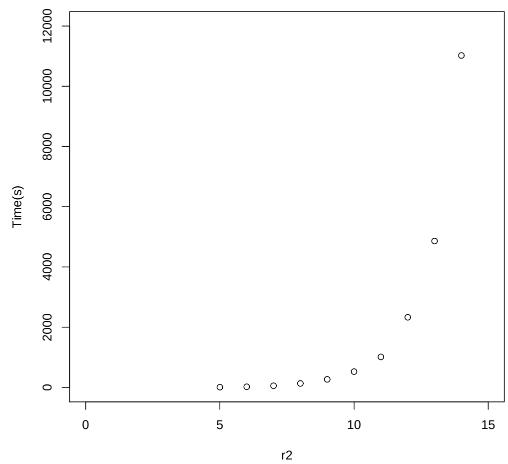


Figure 4: Time to obtain contiguity relations

5.6 Benchmark of Constructing Contiguity Relations

We gave a complexity analysis of finding contiguity relations. When r_1 is fixed, it is $O(r_2^{3r_1})$. The Figure 4 shows timing data to obtain contiguity relations for $5 \times r_2$ contingency tables where the parameter p is $\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1/p_1 & 1/p_2 & \cdots & 1/p_{r_2-1} \\ 1 & 1/p_{r_2} & 1/p_{r_2+1} & \cdots & 1/p_{2(r_2-1)} \\ 1 & \cdots & & & \\ 1 & 1/p_{(r_1-1)(r_2-1)+1} & \cdots & & \end{pmatrix}$ (p_i is the i-th prime number), the row sum vector is $(a_1, 400, 400, 400, 400)$, and the column sum vector is $(200, 300, 500, 500, \ldots, 500)$. As is shown by our complexity analysis, when r_2 becomes larger, it rapidly becomes harder to obtain contiguity relations.

6 Zero Cells

The contiguity relations derived by [7] are valid only when there are no zero cells in the contingency table. If there is a zero ($p_{ij} = 0$ and $u_{ij} = 0$) in the contingency table, a denominator of the contiguity relation is zero in general and therefore we cannot use their identity. One method to avoid this difficulty is interpolation. Note that the normalizing constant Z is a rational function in p_{ij} and the expectation $E[U_{ij}] = p_{ij} \frac{\partial \log Z}{\partial p_{ij}}$ is also a rational function. Because it is a rational function, we can obtain the exact value by evaluating it on a sufficient number of rational p_{ij}'s.

Proposition 1 *Let β be the marginal sum vector and L a generic line in p-space. If we evaluate $E[U_{ij}]$ at $2\beta_1$ points $p \in \mathbb{R}_{>0}^{r_1 \times r_2}$ on a line L, then the exact value of $E[U_{ij}]$ can be obtained at any point on L.*

Proof When we restrict $E[U_{ij}]$ to the line L, it is a rational function in one variable. The degree of the denominator and the numerator is β_1 at most. Apply an interpolation algorithm by rational function, e.g., Stoer-Bulirsch algorithm [29], [24]. Then, we can obtain the exact value by interpolation. $\square$

Example 7 Let the marginal sums and the parameter p (cell probability) be

$$\begin{array}{ccc|c} * & * & * & 3 \\ * & * & * & 4 \\ * & * & * & 3 \\ \hline 3 & 4 & 3 & \end{array}, p = \begin{pmatrix} 1 & 1/2 & 0 \\ 1 & 1/3 & 1/4 \\ 1 & 1 & 1 \end{pmatrix}$$

Then, we can evaluate the expectation matrix $(E[U_{ij}])$ by the difference HGM and interpolation. Below is an output of our package `gtt_ekn3`. Here the `randinit` parameter specifies an interval of random non-zero p_{ij}'s where (i, j)'s are positions of zero cells.

```
[5150] import("gtt_ekn3.rr");
0
[5151] E=gtt_ekn3.cBasistoE_0(0,[[3,4,3],[3,4,3]],[[1,1/2,0],[1,1/3,1/4],[1,1,1]] | randinit=20);
[ 71076/56575 98649/56575 0 ]
[ 157581/113150 28069/22630 77337/56575 ]
[ 39717/113150 114957/113150 92388/56575 ]
// Expectation (exact value)
[5153] number_eval(E);  // Expectation (approximate value)
[ 1.25631462660186 1.74368537339814 0 ]
[ 1.39267344233319 1.2403446752099 1.36698188245692 ]
[ 0.351011931064958 1.01596995139196 1.63301811754308 ]
```

Although the interpolation method is applicable to any pattern of 0-cells, a more efficient method involves utilizing hypergeometric functions restricted on some $p_{ij} = 0$'s. In general, contiguity relations and Pfaffian systems for such hypergeometric functions become complicated. In [9], a method is put forward to evaluate intersection numbers and contiguity relations when only one p_{ij} is zero.

7 Sufficient Statistics as σ-algebra

It is often that we decompose parameters for contingency tables into row and column probabilities and odds ratios. When only odds ratios are the parameters of interest, CMLE is an appropriate method to estimate those odds ratios. However, this decomposition is no longer elementary when contingency tables contain zero cells. To facilitate a mathematically clear discussion of CMLE in the next section, we offer a new formulation of parameters of interest, nuisance parameters, and sufficient statistics.

Classical formulations of sufficient statistics as σ-algebras appear in, e.g., [3], [14]. Our formulation is different because we treat parameters as random variables instead of considering a family of probability measures. This Bayesian statistical approach enable us to consider σ-algebras on parameter spaces. We express nuisance parameters and parameters of interest as sub σ-algebras of the σ-algebra generated by all parameters and data. A Bayesian approach to sufficient statistics is presented in, e.g., Chapter 2 of the text book by M.Schervish [28]. This text book studies sufficient statistics by conditional probabilities given parameter valued random variables. We study them by a more general approach of conditional expectations given σ-algebras. The technical details are lengthy to be precise and, in this section and the next section, we state only fundamental notions and theorems which we need to study two way contingency tables. Proofs for them are given in the preprint of this paper at arxiv [5]. A general framework of the theory will be given in [13].

The treatment of nuisance parameters and parameters of interest is an important issue in statistics. The distinction between those parameters which are of interest versus those which are nuisance, may seem easy. In fact, it seems to be only a matter of declaring that μ is a parameter of interest or ν is a nuisance parameter. As we will see in the next section, when a group acts on parameter spaces and the group is regarded as the space of nuisance parameters, the distinction between them is not trivial. From a geometric perspective, the cause of this difficulty is that determining whether a parameter is "of interest" or a "nuisance" depends on a coordinate system. To formulate the "of interest" notion independently of a specific coordinate system, we will consider σ-algebras on parameter spaces. In probability theory and stochastic processes, σ-algebra is important as a natural way to express information (see, e.g. [12]). Discussions in this section are based on conditional expectations with respect to σ-algebra. For basic properties of conditional expectation, see [34].

Let Θ be a set. The set Θ stands for the parameter spaces. Let $\mathcal{B}(\Theta)$ be a σ-algebra on Θ, then $(\Theta, \mathcal{B}(\Theta))$ is a measure space. In the case where Θ is a topological space, we assume that $\mathcal{B}(\Theta)$ is the Borel algebra on Θ.

In standard parameter estimation, we assume a probability space $(\Omega', \mathcal{F}', \mathbf{P}'_c)$ with a parameter $c \in \Theta$. Let us define our probability space from the standard setting. Suppose $(\Theta, \mathcal{B}(\Theta), \mu)$ is a probability space. Put $\Omega := \Omega' \times \Theta$. Let $\mathcal{F}$ be the σ-algebra on Ω generated by

$$A \times B := \{(\omega, c) \in \Omega | \omega \in A,\, c \in B\} \quad (A \in \mathcal{F}',\, B \in \mathcal{B}(\Theta)).$$

[5] `https://arxiv.org/abs/1803.04170`

The measurable space $(\Omega, \mathcal{F})$ is deemed to be the product measurable space of $(\Omega', \mathcal{F}')$ and $(\Theta, \mathcal{B}(\Theta))$ [34, p75]. For $A \in \mathcal{F}'$, let $f_A : \Theta \to \mathbf{R}$ be the function defined by $f_A(c) := \int_A \mathbf{P}'_c(d\omega)$ $(c \in \Theta)$. If f_A is $\mathcal{B}(\Theta)$-measurable for any $A \in \mathcal{F}'$, we can define a measure $\mathbf{P}$ on $\mathcal{F}$ by $\mathbf{P}(A \times B) := \int_B f_A(c)\mu(dc)$ $(A \in \mathcal{F}',\ B \in \mathcal{B}(\Theta))$. Thus, our probability space is defined as the product space under the measurable condition of f_A.

Let θ be a measurable map from Ω to Θ defined by

$$\theta : \Omega \ni (\omega', c) \mapsto c \in \Theta.$$

This implies that parameters can be regarded as a Θ-valued random variable. Although random variables are usually denoted by capital letters, we use lower case letters to denote random variables that are regarded as parameters.

Example 8 Let $(\Omega', \mathcal{F}', \mathbf{P}'_c)$ be the probability space $(\mathbf{R}, \mathcal{B}(\mathbf{R}), N(\mu, \sigma^2))$, where $N(\mu, \sigma^2)$ is the Gaussian distribution on $\mathbf{R}$ with mean μ and variance σ^2. In this case, the parameter space is $\Theta = \{(\mu, \sigma^2) \in \mathbf{R}^2 | \sigma^2 > 0\}$ and the parameter θ as a measurable map is defined by

$$\theta : \Omega \ni (x, (\mu, \sigma^2)) \mapsto (\mu, \sigma^2) \in \Theta.$$

We restart from a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, which is not necessarily a product space. For a sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$, we use $\mathcal{L}^1(\mathcal{G})$ to denote the linear space of random variables which are integrable and $\mathcal{G}$-measurable. When two elements X and Y of $\mathcal{L}^1(\mathcal{G})$ satisfy $X(\omega) = Y(\omega)$ for all $\omega \in \Omega$, we say that X and Y are equal and denote $X = Y$. Note that $X = Y$ almost surely does not imply that $X = Y$. Let ϑ be the sub σ-algebra of $\mathcal{F}$ generated by a random variable θ. It represents the information of θ. We formulate notions of nuisance parameters, sufficient parameters, and parameters of interest as sub σ-algebras of ϑ.

For a pair of random variables X and Y, it is equivalent that Y is $\sigma(X)$-measurable, to the condition that Y equals to $f(X)$ for a Borel measurable function f. In orther words, we have the following propostion:

Lemma 1 *For a random variable $X \in \mathcal{L}^1$, we have*

$$\mathcal{L}^1(\sigma(X)) = \{f(X) \mid f \in \mathrm{m}\mathcal{B}(\mathbf{R}),\ \mathbf{E}\left(()\, f(X)\right) < \infty\}. \tag{8}$$

Here, $\mathrm{m}\mathcal{B}(\mathbf{R})$ *is the set of all of Borel measurable functions on* $\mathbf{R}$*.*

Proof Since the conposition of measurable functions is also measurable, the left-hand side of Equation (8) includes the hand side.

Suppose $Y \in \mathcal{L}^1(\sigma(X))$.

In the case where Y is an indicator function, $A := \{Y = 1\} \in \sigma(X)$ implies that there exists $B \in \mathcal{B}(\mathbf{R})$ such that $A = \{X \in B\}$. Put $f : \mathbf{R} \to \mathbf{R}$ as

$$f(x) := \begin{cases} 1 & (x \in B) \\ 0 & (x \in B^c) \end{cases} \quad (x \in \mathbf{R}),$$

then f is Borel measurable, and $Y = f(X)$ and $\mathbf{E}\left(()\, f(X)\right) = \P(A) < \infty$ hold. Hence Y is an element of the right-hand side of Equation (8).

In the case where Y is a linear combination of indicator functions $Y_1, \ldots, Y_n$, i.e., $Y = c_1Y_1 + \cdots + c_nY_n$ holds for some $c_i \in \mathbf{R}$, there exists $f_i \in \mathbf{m}\mathcal{B}(\mathbf{R})$ such that $Y_i = f_i(X)$. Then, $f := c_1f_1 + \cdots + c_nf_n$ is an element of the right-hand side of Equation (8).

In the case where $Y \geq 0$, take a sequence of random variables $\{Y_n\}$ as follows:

- each Y_n is a linear combination of indicator functions.
- $\{Y_n\}$ is monotonically increasing.
- $Y_n \to Y$ $(n \to \infty)$.

For each n, take $f_n \in \mathbf{m}\mathcal{B}(\mathbf{R})$ such that $Y_n = f_n(X)$, and put $f := \sup f_n$. Then, we have $Y = f(X)$ and $f \in \mathrm{m}\mathbf{B}(\mathbf{R})$.

For general Y, decompose as $Y = Y_+ - Y_-$ ($Y_+ \geq 0, Y_- \geq 0$). Then there exist f_+ and f_- such that $Y_+ = f_+(X)$ and $Y_- = f_-(X)$. Then, $f := f_+ - f_-$ holds $Y = f(X)$ and $f \in \mathrm{m}\mathbf{B}(\mathbf{R})$.

Remark 2 Theorem A. 41 in [28] may seem like to a generalization of this lemma. However, for $Y \in \mathcal{L}^1(\sigma(X))$, this theorem implies the existence of a fucntion f *only on the image of* X such that $Y = f(X)$, while Lemma 1 gives the existence of a function f on $\mathbf{R}$.

Let X and Y be $\mathbf{R}$-valued random variables and θ be a Θ-valued random variable, which we will call a parameter. We assume that X is integrable. By Lemma 1, the conditional expectation $\mathbf{E}\left(X|Y,\theta\right)$ can be regarded as a function of (Y, θ), i.e., we can take a Borel measurable function f from $\mathbf{R} \times \Theta$ to $\mathbf{R}$ such that

$$f(Y, \theta) = \mathbf{E}\left(X|Y,\theta\right) \quad a.s.$$

Because the equation $f(y, c_1) = f(y, c_2)$ may hold even if $c_1 \neq c_2$, the conditional expectation $\mathbf{E}\left(X|Y,\theta\right)$ can be measurable with respect to a sub σ-algebra strictly smaller than $\sigma(Y, \theta)$. This suggests that taking conditional expectation can reduce the information of θ.

Let us express this loss of information of θ in terms of σ-algebra. Let $\mathcal{D}$ and $\mathcal{G}$ be a sub σ-algebras of $\mathcal{F}$. In some applications, such as Theorem 3 discussed later, it is assumed that $\mathcal{D}$ is the sub σ-algebra generated by all observable statistics and $\mathcal{G}$ is a sub σ-algebra generated by a fraction of the observable statistics and a fraction of the parameters. Note that $\mathcal{G}$ may include some information of parameters. For $X \in \mathcal{L}^1(\mathcal{D})$, the conditional expectation $\mathbf{E}\left(X|\mathcal{G}\right)$ can be measurable for a sub σ-algebra which is strictly smaller than $\mathcal{G}$.

Definition 2 Sub σ-algebra $\mathcal{I}$ is said to be *of interest* with respect to a pair of sub σ-algebras $(\mathcal{D}, \mathcal{G})$ if, for all $X \in \mathcal{L}^1(\mathcal{D})$, $\mathbf{E}\left(X|\mathcal{G}\right)$ is $\mathcal{I}$-measurable.

Notions of nuisance and sufficiency describe a special case of such information loss.

Definition 3 Let $\mathcal{D}$, $\mathcal{S}$ and $\mathcal{N}$ be sub σ-algebras of $\mathcal{F}$. When $\mathcal{S}$ is of interest with respect to $(\mathcal{D}, \sigma(\mathcal{S}, \mathcal{N}))$, we deem that $\mathcal{S}$ is sufficient for $(\mathcal{D}, \mathcal{N})$ or that $\mathcal{N}$ is nuisance for $(\mathcal{D}, \mathcal{S})$.

Remark 3 Note that the condition of Definition 3 is equivalent to stating that the equation

$$\mathbf{E}\left(X|\sigma(\mathcal{S}, \mathcal{N})\right) = \mathbf{E}\left(X|\mathcal{S}\right) \quad \text{a.s.} \tag{9}$$

holds for any $X \in \mathcal{L}^1(\mathcal{D})$. In fact, we have

$$\begin{aligned}
\mathbf{E}\left(X|\sigma(\mathcal{S}, \mathcal{N})\right) &= \mathbf{E}\left(\mathbf{E}\left(X|\sigma(\mathcal{S}, \mathcal{N})\right)|\mathcal{S}\right) && \left(\mathbf{E}\left(X|\sigma(\mathcal{S}, \mathcal{N})\right) \in \mathcal{L}^1(\mathcal{S})\right) \\
&= \mathbf{E}\left(X|\mathcal{S}\right) && (\text{tower property})\,.
\end{aligned}$$

Remark 4 In statistics, a statistic T is sufficient with respect to a parameter θ if the conditional distribution of observed data X given the statistic $T = t$ does not depend on the parameter θ. This condition is formally expressed as

$$p(x|t, \theta) = p(x|t).$$

In similar tests and the Neyman–Scott Problem, θ is denoted as a nuisance parameter or an uninteresting parameter [2]. We express this condition in terms of the measure theory in Definition 3. In our definition, we use σ-fields instead of statistics and parameters. Traditional definitions can be reduced to our definition by

$$\mathcal{D} = \sigma\left(X\right), \qquad \mathcal{S} = \sigma\left(T\right), \qquad \mathcal{N} = \sigma\left(\theta\right).$$

Intuitively, $\mathcal{D}$, $\mathcal{S}$, and $\mathcal{N}$ denote the information of the observed data, the sufficient statistics, and the nuisance parameters, respectively.

In addition, we utilize conditional expectations instead of conditional probabilities because the latter can only be defined for a limited class of probability space and conditions.

Fundamental theorems on sufficient statistics can be generalized in our formulation on the sufficient sigma field [13].

Example 9 *For random variables $X_1, \ldots, X_n, \theta$, suppose that*

1. $0 \le \theta \le 1$

2. *The conditional probability of $X_1, \ldots, X_n$ for given θ is*

$$\mathbf{P}\left(X_1 = x_1, \ldots, X_n = x_n | \theta\right) = \prod_{i=1}^{n} \theta^{x_i} \left(1-\theta\right)^{1-x_i} \quad (x_i \in \{0,1\})$$

Then, putting $\mathcal{D} := \sigma(X_1, \ldots, X_n)$, $\mathcal{N} := \sigma(\theta)$, $\mathcal{S} := \sigma(X_1 + \cdots + X_n)$, $\mathcal{S}$ is sufficient for $(\mathcal{D}, \mathcal{N})$.

To describe a sub σ-algebra of interest in our application to the $\mathcal{A}$-distribution, we consider orbits of some group action. Suppose that a group G acts on a measurable space (S, Σ). For $B \subset S$ and $g \in G$, we put

$$g \cdot B := \{g \cdot b \,|\, b \in B\}, \qquad G \cdot B := \{g \cdot b \,|\, g \in G,\, b \in B\}.$$

Note that $G \cdot B = B$ holds if and only if $g \cdot B = B$ for any $g \in G$.

Let $\mathcal{O}^*$ be the family of the element in Σ invariant under the action of G, i.e., we put

$$\mathcal{O}^* := \{B \in \Sigma : G \cdot B = B\}.$$

Lemma 2 *$\mathcal{O}^*$ is a sub σ-algebra of Σ.*

Proof Obviously, $\mathcal{O}^*$ includes Σ. Let B be an element in $\mathcal{O}^*$. Take any $g \in G$ and $b' \in B^c$. Suppose that $g \cdot b' \in B$. Then, we have $b' = g^{-1} g b' \in G \cdot B = B$. This is a contradition. Hene, $g \cdot b'$ is an element of B^c, and we have $G \cdot B^c \subset B^c$. Since $G \cdot B^c$ includes B^c obviously, we have $G \cdot B^c = B^c$. Consequently, $\mathcal{O}^*$ includes B^c.

Suppose that B_n $(n \in \mathbf{N})$ is an element of $\mathcal{O}^*$. Since we have $G \cdot \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} G \cdot B_n = \bigcup_{n=1}^{\infty} B_n$, $\bigcup_{n=1}^{\infty} B_n$ is an element of $\mathcal{O}^*$. $\square$

A measurable map $X : (\Omega, \mathcal{F}) \to (S, \Sigma)$ induces a sub σ-algebra of $\mathcal{F}$ by

$$\mathcal{O} := \{\{X \in B\} | B \in \mathcal{O}^*\}.$$

Note that $\{X \in B\}$ is the inverse image $X^{-1}(B) = \{\omega \in \Omega | X(\omega) \in B\}$. This notation is often used in the probability theory and we use it in the sequel. We call $\mathcal{O}$ as the σ-algebra generated by the orbits of group G.

Lemma 3 *Let $f : S \to \mathbf{R}$ be a function. Suppose that a measurable map $X : (\Omega, \mathcal{F}) \to (S, \Sigma)$ is surjective. Then, all of the following four conditions are equivalent:*

(a). *f is $\mathcal{O}^*$-measurable.*

(b). *$f(g \cdot x) = f(x)$ holds for any $g \in G$ and any $x \in S$.*

(c). *$f(X)$ is $\mathcal{O}$-measurable.*

(d). *$f(g \cdot X) = f(X)$ holds for any $g \in G$.*

Proof [(a)⇒(c)]. Suppose that f is $\mathcal{O}^*$-measurable. For any $B \in \mathcal{B}(\mathbf{R})$, we have $f^{-1}(B) \in \mathcal{O}^*$. By the definition of $\mathcal{O}$, $X^{-1}(f^{-1}(B)) = \{f(X) \in B\} \in \mathcal{O}$ holds. Hence, $f(X)$ is $\mathcal{O}$-measurable.

[(c)⇒(d)]. Suppose that $f(X)$ is $\mathcal{O}$-measurable. Take an arbitrary $a \in \mathbf{R}$. Then, $\{f(X) = a\} \in \mathcal{O}$ implies that there exists $B \in \mathcal{O}^*$ such that $\{X \in B\} = \{f(X) = a\} = \{X \in f^{-1}(a)\}$. Since $X : \Omega \to S$ is surjective, we have $B = f^{-1}(a)$. Thus, $f^{-1}(a)$ is an element of $\mathcal{O}^*$, and we have $G \cdot f^{-1}(a) = f^{-1}(a)$. This implies $g \cdot f^{-1}(a) = f^{-1}(a)$ holds for any $g \in G$, and we have

$$\begin{aligned}\{f(g \cdot X) = a\} &= \{g \cdot X \in f^{-1}(a)\} = \{X \in g^{-1} \cdot f^{-1}(a)\} \\ &= \{X \in f^{-1}(a)\} = \{f(X) = a\}.\end{aligned}$$

Hence, $f(g \cdot X) = f(X)$ holds for all $g \in G$.

[(d)⇔(b)]. Since $X : \Omega \to S$ is surjective, $f(g \cdot x) = f(x)$ holds for any $x \in S$ and $g \in G$ if and only if $f(g \cdot X) = f(X)$ holds for any $g \in G$.

[(b)⇒(a)]. Suppose the function f is invariant under the action of G. Take an arbitrary $B \in \mathcal{B}(\mathbf{R})$. By the invariance of the function f, we have

$$G \cdot f^{-1}(B) = \{g \cdot x | g \in G,\, x \in S,\, f(x) \in B\} = \{x \in S | f(x) \in B\} = f^{-1}(B).$$

Hence, $f^{-1}(B)$ is included in $\mathcal{O}^*$. This implies that f is $\mathcal{O}^*$-measurable. □

Since $\mathcal{O}$ is a sub σ-algebra of $\sigma(X)$, $Y \in \mathcal{L}^1(\mathcal{O})$ can be regarded as a function of X. By Lemma 3, we say that a random variable Y is invariant under the action of group G if Y is $\mathcal{O}$-measurable.

We apply the above discussion on group actions to sufficient σ-algebras. Let $\mathcal{D}$ be a sub σ-algebra of $\mathcal{F}$. Let Θ be a topological space, and $\theta : (\Omega, \mathcal{F}) \to (\Theta, \mathcal{B}(\Theta))$ be a measurable map. We regard θ and Θ as the parameter and the space of parameters respectively. Let S be a measurable space, and $T : \Omega \to S$ be an $\mathcal{D}$-measurable map. For $X \in \mathcal{L}^1(\mathcal{D})$, $\mathbf{E}(X|T, \theta)$ can be regarded as a function on $S \times \Theta$. In other words, there exists a function $f_X : S \times \Theta \to \mathbf{R}$ such that $f_X(T(\omega), \theta(\omega)) = \mathbf{E}(X|T, \theta)(\omega)$ for all $\omega \in \Omega$.

Lemma 4 *We assume the same notation as above. Suppose that an action of group G on Θ satisfies*

$$f_X(t, g \cdot c) = f_X(t, c)$$

for all $t \in S$, $c \in \Theta$, $g \in G$, and $X \in \mathcal{L}^1(\mathcal{D})$, and put

$$\mathcal{O} := \{\{(T, \theta) \in B\} \,|\, B \in \Sigma \times \mathcal{B}(\Theta),\, G \cdot B = B\}.$$

Then, $\mathcal{O}$ is of interest with respect to $(\mathcal{D}, \sigma(T, \theta))$.

Proof The group action on Θ induces an group action on the Cartesian product $S \times \Theta$ by

$$g \cdot (t, c) = (t, g \cdot c) \quad (g \in G,\, (t, c) \in S \times \Theta).$$

Applying Lemma 3 in the case of the group action on $S \times \Theta$, $f_X(T, \theta)$ is $\mathcal{O}$-measurable for any $X \in \mathcal{L}^1(\mathcal{D})$. Hence, $\mathcal{O}$ is of interest with respect to $(\mathcal{D}, \sigma(T, \theta))$. □

Although the following lemmas may be well known, we could not find a proof in the literature. Therefore, we present a proof here. We will use these lemmas in the next section.

Lemma 5 *Let a measurable function $\theta : \Omega \to \Theta$ be surjective and $\mathcal{G}$ be a sub σ-algebra of $\sigma(\theta) := \{\theta^{-1}B | B \in \mathcal{B}(\Theta)\}$. Then, $\theta\mathcal{G} := \{\theta(B) | B \in \mathcal{G}\}$ is a sub σ-algebra of $\mathcal{B}(\Theta)$.*

Proof Since θ is surjective, $\Theta = \theta(\Omega)$ is an element of $\theta\mathcal{G}$.

Let $A \in \theta\mathcal{G}$. There exists $B \in \mathcal{G}$ such that $A = \theta(B)$. By $\mathcal{G} \subset \sigma(\theta)$, there exists $C \in \mathcal{B}(\Theta)$ such that $B = \theta^{-1}C$. Since surjectivity of θ implies that $\theta(\theta^{-1}S) = S$ holds for any $S \subset \Theta$, we have $A = \theta(B) = \theta(\theta^{-1}C) = C$. By $\theta^{-1}A = \theta^{-1}C = B \in \mathcal{G}$, we have $\theta^{-1}A^c = (\theta^{-1}A)^c \in \mathcal{G}$. By surjectivity of θ, $A^c = \theta(\theta^{-1}A^c)$ is an element of $\theta\mathcal{G}$.

Suppose $A_n \in \theta\mathcal{G}$ for $n \in \mathbf{N}$. Analogously, we have $\theta^{-1}A_n \in \mathcal{G}$. Consequently, $\theta^{-1}\bigcup_{n\in\mathbf{N}} A_n = \bigcup_{n\in\mathbf{N}} \theta^{-1}A_n \in \mathcal{G}$ implies $\bigcup_{n\in\mathbf{N}} A_n = \theta\left(\theta^{-1}\bigcup_{n\in\mathbf{N}} A_n\right) \in \theta\mathcal{G}$. □

Lemma 6 *Suppose that a measurable function $\theta : \Omega \to \Theta$ is surjective. Let $f_\lambda : \Theta \to \mathbf{R}\,(\lambda \in \Lambda)$ be measurable functions. Then, we have*

$$\sigma\left(f_\lambda \circ \theta : \lambda \in \Lambda\right) = \theta^{-1}\sigma\left(f_\lambda : \lambda \in \Lambda\right), \tag{10}$$

where $\sigma\left(f_\lambda : \lambda \in \Lambda\right)$ is the σ-algebra generated by $\{f_\lambda^{-1}B | \lambda \in \Lambda,\, B \in \mathcal{B}(\mathbf{R})\}$.

Proof Obviously, the right hand side of (10) includes the left hand side. We show the opposite inclusion. By the surjectivity of θ, we have $f_\lambda^{-1}B = \theta\theta^{-1}f_\lambda^{-1}B = \theta(f_\lambda \circ \theta)^{-1}B \in \theta\sigma(f_\lambda \circ \theta : \lambda \in \Lambda)$ for any $B \in \mathcal{B}(\mathbf{R})$. By Lemma 5, $\theta\sigma(f_\lambda \circ \theta : \lambda \in \Lambda)$ is a sub σ-algebra of $\mathcal{B}(\Theta)$. Hence, we have

$$\sigma(f_\lambda : \lambda \in \Lambda) \subset \theta\sigma(f_\lambda \circ \theta : \lambda \in \Lambda). \tag{11}$$

Note that we have

$$C = \theta^{-1}\theta C \quad (C \in \sigma(\theta)). \tag{12}$$

In fact, since there exists $C' \in \mathcal{B}(\Theta)$ such that $C = \theta^{-1}C'$, we have $\theta^{-1}\theta C = \theta^{-1}\theta\theta^{-1}C' = \theta^{-1}C' = C$.

Let $A \in \theta^{-1}\sigma(f_\lambda : \lambda \in \Lambda)$. There exists $B \in \sigma(f_\lambda : \lambda \in \Lambda)$ such that $A = \theta^{-1}B$. By (11), there exists $C \in \sigma(f_\lambda \circ \theta : \lambda \in \Lambda)$ such that $B = \theta^{-1}C$. Equation (12) and $\sigma(f_\lambda \circ \theta : \lambda \in \Lambda) \subset \sigma(\theta)$ implies $A = \theta^{-1}\theta C = C \in \sigma(f_\lambda \circ \theta : \lambda \in \Lambda)$. Consequently, the opposite inclusion holds. □

Lemma 7 *Let V and W be finite-dimensional vector spaces over $\mathbf{R}$, $V \oplus W$ be the direct sum of V and W, $\pi : V \oplus W \to V$ be the projection, and V^* be the dual space of V. Then, we have*

$$\{B \in \mathcal{B}(V \oplus W) | B + W = B\} = \sigma(f \circ \pi : f \in V^*). \tag{13}$$

Here, we put $B + W := \{v + w | v \in B,\ w \in W\}$.

Proof Since $\pi^{-1}B + W = \pi^{-1}B$ holds for any $B \in \mathcal{B}(V)$, $\{B \in \mathcal{B}(V \oplus W) | B + W = B\}$ includes $\pi^{-1}\mathcal{B}(V)$. Let $\iota : V \to V \oplus W$ be the canonical injection. Suppose that $B \in \mathcal{B}(V \oplus W)$ satisfies $B + W = B$. Since we have

$$\begin{aligned} x \in \pi^{-1}\iota^{-1}B &\Leftrightarrow \iota\pi(x) \in B \\ &\Leftrightarrow \pi(x) \in B && (\iota(y) = y \text{ holds for } y \in V) \\ &\Rightarrow \pi(x) + (x - \pi(x)) \in B + W && (x - \pi(x) \in W) \\ &\Rightarrow x \in B + W \\ &\Leftrightarrow x \in B && (B + W = B), \end{aligned}$$

$\pi^{-1}\iota^{-1}B \subset B$ holds. Since we can show the opposite inclusion analogously, we have $\pi^{-1}\iota^{-1}B = B$. By $\iota^{-1}B \in \mathcal{B}(V)$, B is an element of $\pi^{-1}\mathcal{B}(V)$. Then, we have

$$\{B \in \mathcal{B}(V \oplus W) | B + W = B\} = \pi^{-1}\mathcal{B}(V).$$

Since $f \in V^*$ is a continuous map from V to $\mathbf{R}$, $\mathcal{B}(V)$ includes $\sigma(f : f \in V^*)$. Let $\{f_1, \dots, f_n\}$ be a basis of V^*. Since any open subsets of $V \cong \mathbf{R}^n$ is a countable union of open sets of the form

$$\bigcap_{i=1}^{n} f_i^{-1}\left(\{x \in \mathbf{R} | a_i < x < b_i\}\right) \quad (a_i, b_i \in \mathbf{Q}),$$

we have $\mathcal{B}(V) \subset \sigma(f_1, \dots, f_n) \subset \sigma(f : f \in V^*)$. Consequently, $\mathcal{B}(V) = \sigma(f : f \in V^*)$ holds and we have

$$\{B \in \mathcal{B}(V \oplus W) | B + W = B\} = \pi^{-1}\sigma(f : f \in V^*).$$

By Lemma 6, the right hand side of the above equation equals to $\sigma(f \circ \pi : f \in V^*)$. □

8 Application to the Conditional MLE problem

In this section, we discuss a conditional MLE problem for $\mathcal{A}$-distributions.

Let A be an integer matrix of size $d \times n$, and b be an integer vector of length n. Suppose that Poisson random variables $X_k \sim \mathrm{Pois}(c_k)$, $(k = 1, \dots, n)$ are mutually independent. We denote the conditional distribution of the random vector $X := (X_1, \dots, X_n)^\top$ given $AX = b$ as an $\mathcal{A}$-distribution. The parameters of $\mathcal{A}$-distribution are $c = (c_1, \dots, c_n)^\top$ and $b = (b_1, \dots, b_n)^\top$. The probability mass function of the $\mathcal{A}$-distribution is given as

$$\mathbf{P}\left(X = x | AX = b, \theta = c\right) = \frac{\prod_{j=1}^{n} \frac{c_j^{x_j}}{x_j!} \exp\left(-c_j\right)}{\sum_{Ay=b} \prod_{j=1}^{n} \frac{c_j^{y_j}}{y_j!} \exp\left(-c_j\right)} = \frac{\prod_{j=1}^{n} \frac{c_j^{x_j}}{x_j!}}{\sum_{Ay=b} \prod_{j=1}^{n} \frac{c_j^{y_j}}{y_j!}}.$$

An application of conditional distributions in statistics is the elimination of nuisance parameters. By Definition 3 and Remark 4, the conditional distribution of a statistic given the occurrence of a sufficient statistic of a nuisance parameter does not depend on the value of the nuisance parameter. This is an important property in similar tests and the Neyman–Scott problems (see, e.g., [2] and [10]). Hence, by the conditional distribution, we can estimate the parameter of interest without being affected by the nuisance parameter. From this perspective, we can regard the $\mathcal{A}$-distribution as the conditional distribution given the sufficient statistic AX, and the nuisance parameter corresponding to AX is $A\theta$. The traditional definition does not offer a mathematically clear description of the parameter of interest for this case. This is the motivation for the discussions in the previous section. The space of parameters of interest is naturally described as a sub σ-algebra under less restrictive conditions on θ and c.

The parameter c of $\mathcal{A}$-distribution moves on the set $\Theta := \mathbf{R}^n_{\geq 0}$. Consider the action of the multiplicative group $G := \mathbf{R}^d_{>0}$ on the space Θ defined as

$$g \cdot c = \left(c_j \prod_{i=1}^{d} g_i^{a_{ij}} \right)_{j=1,\ldots,n} \qquad (g \in G,\, c \in \Theta).$$

This group action on Θ induces group action on $\mathbf{Z}^d_{\geq 0} \times \Theta$ by

$$g \cdot (b, c) = (b, g \cdot c) \quad (g \in G,\, (b, c) \in \mathbf{Z}^d_{\geq 0} \times \Theta).$$

Applying Lemma 4 in the case where $\mathcal{D} = \sigma(X)$, $S = \mathbf{Z}^d_{\geq 0}$, and $T = AX$, we have the following theorem:

Theorem 3 *The sub σ-algebra*

$$\mathcal{O} := \{\{(AX, \theta) \in B\} \mid B \in \mathcal{B}(\mathbf{Z}^d_{\geq 0}) \times \mathcal{B}(\Theta),\, G \cdot B = B\}$$

is of interest with respect to $(\sigma(X), \sigma(AX, \theta))$.

Proof For any $g \in G$, we have

$$\begin{aligned} g \cdot \frac{\prod_{j=1}^n \theta_j^{x_j}/x_j!}{\sum_{Ay=b} \prod_{j=1}^n \theta_j^{y_j}/y_j!} &= \frac{\prod_{j=1}^n \left(\theta_j^{x_j} \prod_{i=1}^d g_i^{a_{ij}x_j}\right)/x_j!}{\sum_{Ay=b} \prod_{j=1}^n \left(\theta_j^{y_j} \prod_{i=1}^d g_i^{a_{ij}y_j}\right)/y_j!} \\ &= \frac{\prod_{i=1}^d g_i^{b_i} \prod_{j=1}^n \theta_j^{x_j}/x_j!}{\sum_{Ay=b} \prod_{i=1}^d g_i^{b_i} \prod_{j=1}^n \theta_j^{y_j}/y_j!} \\ &= \frac{\prod_{j=1}^n \theta_j^{x_j}/x_j!}{\sum_{Ay=b} \prod_{j=1}^n \theta_j^{y_j}/y_j!}. \end{aligned}$$

Since the conditional distribution of X with respect to (AX, θ) is invariant under the action of G on $\mathbf{Z}^d_{\geq 0} \times \Theta$, for any $Y \in \mathcal{L}^1(\sigma(X))$, the conditional expectation $\mathbf{E}(Y|AX, \theta)$ is also invariant under the action. By Lemma 4, $\mathcal{O}$ is of interest with respect to $(\sigma(X), \sigma(AX, \theta))$. $\square$

Note that the quotient space Θ/G by the group action G is not a manifold. Therein lies the difficulty with describing the space of parameters of interest and hence why we utilized the notion of σ-algebra of interest.

For a vector $v = (v_1, \ldots, v_n)^\top \in \mathbf{R}^n$, we use $J(v)$ to denote the set of subscript j that satisfies $v_j \neq 0$. We also use $|J(v)|$ to denote the number of elements in $J(v)$, and we put $J(v)^c := \{j \in \mathbf{N} | j \notin J(v)\}$.

For $\alpha = (\alpha_1, \ldots, \alpha_n)^\top \in \mathbf{R}^n$, let R_α be the function from $\Theta = \mathbf{R}^n_{\geq 0}$ to $\mathbf{R}$ defined by

$$R_\alpha(c) := \begin{cases} \prod_{j \in J(\alpha)} c_j^{\alpha_j} & (c_j \neq 0 \text{ for all } j \in J(\alpha)) \\ 0 & (c_j = 0 \text{ for some } j \in J(\alpha)) \end{cases} \qquad (c = (c_1, \ldots, c_n)^\top \in \Theta).$$

Let $Z : \Theta \to \mathbf{R}^n$ be the function defined by $Z(c) := (Z_1(c), \ldots, Z_n(c))^\top$ $(c \in \Theta)$ where

$$Z_j(c) := \begin{cases} 1 & (c_j > 0) \\ 0 & (c_j = 0). \end{cases}$$

Lemma 8 *The random variables AX, $R_\alpha(\theta)\,(\alpha \in \ker A)$, and $Z(\theta)$ are $\mathcal{O}$-measurable.*

Proof Obviously, AX is $\mathcal{O}$-measurable. Let $\pi : \mathbf{Z}_{\geq 0}^d \times \Theta \to \Theta$ be the projection. By some calculations, we have

$$R_\alpha \circ \pi(g \cdot (t,c)) = R_\alpha \circ \pi((t,c)), \qquad\qquad Z \circ \pi(g \cdot (t,c)) = Z \circ \pi((t,c))$$

for any $\alpha \in \ker A$, $g \in G$, and $(t,c) \in \mathbf{Z}_{\geq 0}^d \times \Theta$. Consequently, the functions $R_\alpha \circ \pi$ and $Z \circ \pi$ are invariant under G. Applying Lemma 3 in the case where $X = (AX, \theta)$, $R_\alpha \circ \pi(X) = R_\alpha(\theta)$ and $Z \circ \pi(X) = Z(\theta)$ are $\mathcal{O}$-measurable. $\square$

Let $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbf{R}^n$, i.e., the i-th component of e_i is 1 and the other components are 0. For the $d \times n$ matrix A, $\ker A$ and $\mathrm{Im}A^\top$ can be written as

$$\ker A = \left\{ \sum_{j=1}^n x_j e_j | \sum_{j=1}^n a_{ij} x_j = 0 \right\}, \qquad\qquad \mathrm{Im}A^\top = \sum_{i=1}^d \mathbf{R} \sum_{j=1}^n a_{ij} e_j,$$

where a_{ij} is the (i,j)-component of A. For $z \in \{0,1\}^n$, let $\mathbf{R}^{J(z)} := \sum_{j \in J(z)} \mathbf{R} e_j$ be the sub vector space of $\mathbf{R}^n$ spanned by e_j $(j \in J(z))$, $p_z : \mathbf{R}^n \to \mathbf{R}^{J(z)}$ $\left(\sum_{j=1}^n x_j e_j \mapsto \sum_{j \in J(z)} x_j e_j\right)$ be the projection, and $\hat{\iota}_z : \mathbf{R}^{J(z)} \to \mathbf{R}^n$ be the canonical injection. For $\alpha \in \mathbf{R}^n$, we denote by L_α the linear map from $\mathbf{R}^n$ to $\mathbf{R}$ defined by

$$L_\alpha(x) = \sum_{j=1}^n \alpha_j x_j \quad \left(x = (x_1, \ldots, x_n)^\top \in \mathbf{R}^n\right).$$

The compositions of L_α and $\hat{\iota}_z$ generate a sub σ-algebra of $\mathcal{B}(\mathbf{R}^{J(z)})$, and we put it as $\sigma(L_\alpha \hat{\iota}_z : \alpha \in \mathbf{R}^{J(z)} \cap \ker A) = \{(L_\alpha \hat{\iota}_z)^{-1}(B) | B \in \mathcal{B}(\mathbf{R})\}$.

Lemma 9 *Under the same notation as above, the following equation holds for any $z \in \{0,1\}^n$:*

$$\left\{ B \in \mathcal{B}(\mathbf{R}^{J(z)}) | B + p_z \mathrm{Im}A^\top = B \right\} = \sigma(L_\alpha \hat{\iota}_z : \alpha \in \mathbf{R}^{J(z)} \cap \ker A). \tag{14}$$

Proof With a map

$$\langle \cdot, \cdot \rangle : \mathbf{R}^{J(z)} \times \mathbf{R}^{J(z)} \to \mathbf{R} \quad \left(\left\langle \sum_{j \in J(z)} x_j e_j, \sum_{j \in J(z)} y_j e_j, \right\rangle = \sum_{j \in J(z)} x_j y_j \right),$$

$\mathbf{R}^{J(z)}$ is an inner product space. By the equation

$$\mathbf{R}^{J(z)} \cap \ker A = \{ v \in \mathbf{R}^{J(z)} | \langle v, w \rangle = 0 \text{ for all } w \in p_z \mathrm{Im}A^\top \},$$

we have

$$\mathbf{R}^{J(z)} = \left(\mathbf{R}^{J(z)} \cap \ker A \right) \oplus p_z \mathrm{Im}A^\top.$$

By Lemma 7, we have

$$\left\{ B \in \mathcal{B}(\mathbf{R}^{J(z)}) | B + p_z \mathrm{Im}A^\top = B \right\} = \sigma\left(f \circ \pi : f \in \left(\mathbf{R}^{J(z)} \cap \ker A \right)^* \right),$$

where $\pi : \mathbf{R}^{J(z)} \to \mathbf{R}^{J(z)} \cap \ker A$ is the projection and $\left(\mathbf{R}^{J(z)} \cap \ker A\right)^*$ denotes the dual space of $\mathbf{R}^{J(z)} \cap \ker A$. Since we have

$$\{ f \circ \pi : f \in \left(\mathbf{R}^{J(z)} \cap \ker A \right)^* \} = \{ L_\alpha \hat{\iota}_z : \alpha \in \mathbf{R}^{J(z)} \cap \ker A) \},$$

Equation (14) holds. $\square$

Lemma 10 *For $z \in \{0,1\}^n$, Let ι_z be the canonical injection from $\Theta_z := \{c \in \Theta | Z(c) = z\}$ to Θ. Then, the inclusion*

$$\iota_z^{-1} \mathcal{O}^{**} \subset \iota_z^{-1} \sigma(R_\alpha : \alpha \in \ker A) \tag{15}$$

holds. Here, we put $\mathcal{O}^{**} := \{B \in \mathcal{B}(\Theta) | G \cdot B = B\}$.

Proof Fix $z = (z_1, \ldots, z_n)^\top \in \{0,1\}^n$.

Let $B \in \mathcal{O}^{**}$. Since ι_z is a continuous map, $\iota_z^{-1}B$ is a Borel set of Θ_z. For any $c \in \iota_z^{-1}B$ and $g \in G$, $\iota_z^{-1}B \subset \Theta_z$ and $G \cdot \Theta_z = \Theta_z$ implies $g \cdot c \in \Theta_z$. Since the equation $\iota_z(g \cdot c) = g \cdot c \in G \cdot B = B$ implies $g \cdot c \in \iota_z^{-1}B$, we have $G \cdot \iota_z^{-1}B = \iota_z^{-1}B$. Hence, we have the following inclusion relation:

$$\iota_z^{-1}\mathcal{O}^{**} \subset \{B \in \mathcal{B}(\Theta_z) | G \cdot B = B\}. \tag{16}$$

Suppose that $B \in \mathcal{B}(\Theta_z)$ satisfies $G \cdot B = B$. Let $\psi : \Theta_z \to \mathbf{R}^{J(z)}$ be a diffeomorphism defined by $\psi(c) := \sum_{j \in J(z)} \log(c_j) e_j$ ($c = (c_1, \ldots, c_n)^\top \in \Theta_z$). Note that c_j is strictly positive for $j \in J(z)$. Since ψ is a diffeomorphism, $\psi(B)$ is a Borel set of $\mathbf{R}^{J(z)}$. Put $v_i := \sum_{j \in J(z)} a_{ij} e_j \in \mathbf{R}^{J(z)}$ $(i = 1, \ldots, d)$. For any $c \in B$ and any $g_i \in \mathbf{R}$ $(i = 1, \ldots, d)$, $\psi(c) + \sum_{i=1}^d g_i v_i$ is an element of $\psi(B)$. In fact, the inverse

$$\begin{aligned}\psi^{-1}\left(\psi(c) + \sum_{i=1}^d g_i v_i\right) &= \psi^{-1}\left(\sum_{j \in J(z)} \left(\log(c_j) + \sum_{i=1}^d a_{ij} g_i\right) e_j\right) = \sum_{j \in J(z)} \exp\left(\log(c_j) + \sum_{i=1}^d a_{ij} g_i\right) e_j \\ &= \sum_{j \in J(z)} \left(c_j \prod_{i=1}^d \exp(g_i a_{ij})\right) e_j = g' \cdot c\end{aligned}$$

is an element of $G \cdot B = B$. Here, we put $g' := (\exp(g_i))_{i=1,\ldots,d} \in G$. This implies $\psi(B) + p_z \mathrm{Im} A^\top = \psi(B)$. Since $\psi^{-1}\psi(B) = B$ holds, we have

$$\{B \in \mathcal{B}(\Theta_z) | G \cdot B = B\} \subset \psi^{-1}\left\{B \in \mathcal{B}(\mathbf{R}^{J(z)}) | B + p_z \mathrm{Im} A^\top = B\right\}. \tag{17}$$

By Lemma 9, we have

$$\psi^{-1}\left\{B \in \mathcal{B}(\mathbf{R}^{J(z)}) | B + p_z \mathrm{Im} A^\top = B\right\} = \psi^{-1}\sigma(L_\alpha \hat{\iota}_z : \alpha \in \mathbf{R}^{J(z)} \cap \ker A) \tag{18}$$

Since the standard exponential mapping $\exp : \mathbf{R} \to \mathbf{R}_{>0}$ $(x \mapsto \exp(x))$ is a diffeomorphism, we have

$$\sigma(L_\alpha \hat{\iota}_z : \alpha \in \mathbf{R}^{J(z)} \cap \ker A) = \sigma(\exp L_\alpha \hat{\iota}_z : \alpha \in \mathbf{R}^{J(z)} \cap \ker A).$$

The mappings $R_\alpha \iota_z$ $(\alpha \in \mathbf{R}^{J(z)} \cap \ker A)$ induce a σ-algebra on Θ_z as

$$\sigma(R_\alpha \iota_z : \alpha \in \mathbf{R}^{J(z)} \cap \ker A) := \sigma\left((R_\alpha \iota_z)^{-1}(B) : \alpha \in \ker A \cap \mathbf{R}^{J(z)},\, B \in \mathcal{B}(\mathbf{R})\right).$$

By Lemma 6 and the equation $\exp L_\alpha \hat{\iota}_z \psi = R_\alpha \iota_z$, we have

$$\psi^{-1}\sigma(\exp L_\alpha \hat{\iota}_z : \alpha \in \mathbf{R}^{J(z)} \cap \ker A) = \sigma(R_\alpha \iota_z : \alpha \in \mathbf{R}^{J(z)} \cap \ker A) \tag{19}$$

Obviously, we have

$$\sigma(R_\alpha \iota_z : \alpha \in \mathbf{R}^{J(z)} \cap \ker A) \subset \sigma(R_\alpha \iota_z : \alpha \in \ker A). \tag{20}$$

For any $B \in \mathcal{B}(\mathbf{R})$ and any $\alpha \in \ker A$, $(R_\alpha \iota_z)^{-1}(B) = \iota_z^{-1} R_\alpha^{-1}(B)$ is an element of $\iota_z^{-1}\sigma(R_\alpha : \alpha \in \ker A)$. Since $\iota_z^{-1}\sigma(R_\alpha : \alpha \in \ker A)$ is a σ-algebra on Θ_z, we have

$$\sigma(R_\alpha \iota_z : \alpha \in \ker A) \subset \iota_z^{-1}\sigma(R_\alpha : \alpha \in \ker A) \tag{21}$$

By (16), (17), (18), (19), (20) and (21), we have (15). □

Lemma 11 *The following equation holds:*

$$\mathcal{O}^{**} = \sigma(R_\alpha,\, Z; \alpha \in \ker A). \tag{22}$$

Proof Since R_α and Z are invariant under the action G, Lemma 3 implies $\mathcal{O}^{**} \supset \sigma(R_\alpha,\, Z; \alpha \in \ker A)$. Let $B \in \mathcal{O}^{**}$. For $z \in \{0,1\}^n$, put $B_z := B \cap \Theta_z \in \mathcal{O}^{**}$. Then, we have $B = \bigcup_{z \in \{0,1\}^n} B_z$. Since Lemma 10 implies $B_z = \iota_z^{-1}B \in \iota_z^{-1}\sigma(R_\alpha,\, Z; \alpha \in \ker A)$, there exists $\hat{B}_z \in \sigma(R_\alpha,\, Z; \alpha \in \ker A)$ such that $B_z = \iota_z^{-1}\hat{B}_z$. By $B_z = \iota_z^{-1}\hat{B}_z = \hat{B}_z \cap \Theta_z$, we have $B_z \in \sigma(R_\alpha,\, Z; \alpha \in \ker A)$. Consequently, $\sigma(R_\alpha,\, Z; \alpha \in \ker A)$ includes B. Hence, we have (22). □

Lemma 12 *Let* $\pi : \mathbf{Z}^d_{\geq 0} \times \Theta \to \Theta$ *and* $\pi' : \mathbf{Z}^d_{\geq 0} \times \Theta \to \mathbf{Z}^d_{\geq 0}$ *be the projections. Put* $\mathcal{O}^* := \{B \in \mathcal{B}(\mathbf{Z}^d_{\geq 0} \times \Theta) | G \cdot B = B\}$*. Then, the following equation holds:*

$$\mathcal{O}^* = \sigma\left(\pi',\, R_\alpha \circ \pi,\, Z \circ \pi; \alpha \in \ker A\right). \tag{23}$$

Proof Since π', $R_\alpha \circ \pi$, and $Z \circ \pi$ are invariant under the action G, Lemma 3 implies

$$\mathcal{O}^* \supset \sigma\left(\pi',\, R_\alpha \circ \pi,\, Z \circ \pi; \alpha \in \ker A\right).$$

Let $B \in \mathcal{O}^*$. For $t \in \mathbf{Z}^d_{\geq 0}$, let $\iota_t : \Theta \to \mathbf{Z}^d_{\geq 0} \times \Theta$ is an inclusion map defined by $\iota_t(c) = (t, c)$, and put $B_t := B \cap \iota_t(\Theta)$. Then, $\iota_t^{-1} B_t$ is a Borel set of Θ. Since the equation $G \cdot B = B$ implies $G \iota_t^{-1} B = \iota^{-1} B$, $\iota_t^{-1} B$ is an element of $\mathcal{O}^{**}$. By Lemma 11 and Lemma 6, we have $\pi^{-1}\iota_t^{-1} B_t \in \pi^{-1}\sigma\left(R_\alpha,\, Z; \alpha \in \ker A\right) = \sigma\left(R_\alpha \circ \pi,\, Z \circ \pi; \alpha \in \ker A\right)$. Hence, $B_t = \pi^{-1}\iota_t^{-1} B_t \cap \iota_t(\Theta) \in \sigma\left(\pi',\, R_\alpha \circ \pi,\, Z \circ \pi; \alpha \in \ker A\right)$ implies $B = \bigcup_{t \in \mathbf{Z}^d_{\geq 0}} B_t \in \sigma\left(\pi',\, R_\alpha \circ \pi,\, Z \circ \pi; \alpha \in \ker A\right)$. We have (23). □

Theorem 4 *Let* $\hat{\theta} : \Omega \to \mathbf{Z}^d_{\geq 0} \times \Theta$ *be the measurable function defined by* $\hat{\theta}(\omega) = (AX(\omega), \theta(\omega))$*. If* $\hat{\theta}$ *is surjective, then the equation*

$$\mathcal{O} = \sigma\left(AX, R_\alpha(\theta),\, Z(\theta); \alpha \in \ker A\right) \tag{24}$$

holds.

Proof By Lemma 6 and Lemma 12, we have

$$\begin{aligned} \mathcal{O} &= \hat{\theta}^{-1}\mathcal{O}^* = \hat{\theta}^{-1}\sigma\left(\pi', R_\alpha \circ \pi,\, Z \circ \pi; \alpha \in \ker A\right) \\ &= \sigma\left(\pi'(\hat{\theta}), R_\alpha \circ \pi(\hat{\theta}),\, Z \circ \pi(\hat{\theta}); \alpha \in \ker A\right) = \sigma\left(AX,\, R_\alpha(\theta),\, Z(\theta); \alpha \in \ker A\right). \end{aligned}$$

□

This theorem implies that sub σ-algebra of interest $\mathcal{O}$ stands for generalized odds ratios, which are, intuitively, parameters of interest. Note that the parameter may lie on the border θ_i.

As an interesting and important case of $\mathcal{A}$-distributions, we consider the $r_1 \times r_2$ contingency table. Let u_{ij} be independent Poisson random variables with parameter $\theta_{ij} \geq 0\, (1 \leq i \leq r_1,\, 1 \leq j \leq r_2)$. The parameter $\theta := (\theta_{ij})$ lies on the set $\Theta := \mathbf{R}^{r_1 \times r_2}_{\geq 0}$. As in the previous section, we regard θ as a measurable function from $(\Omega, \mathcal{F})$ to $(\Theta, \mathcal{B}(\Theta))$. Note that we can assume that θ is surjective without loss of generality. Let $\mathcal{D}$ be the sub σ-algebra generated by all u_{ij}, and $\mathcal{G}$ be the sub σ-algebra generated by

$$\theta_{ij}\,(1 \leq i \leq r_1,\, 1 \leq j \leq r_2), \qquad \sum_{i=1}^{r_1} u_{ij}\,(1 \leq j \leq r_2), \qquad \sum_{j=1}^{r_2} u_{ij}\,(1 \leq i \leq r_1).$$

For all $X \in \mathcal{L}^1(\mathcal{D})$, the conditional expectation $\mathbf{E}\left(X | \mathcal{G}\right)$ is invariant under the action of the multiplicative group $G := \mathbf{R}^{r_1 + r_2}_{>0}$ on Θ defined by

$$g \cdot c := (g_i g_{r_1 + j} c_{ij}) \quad (g = (g_i) \in G,\, c = (c_{ij}) \in \Theta).$$

For $1 \leq i, k \leq r_1$ and $1 \leq j, \ell \leq r_2$, let $R_{ijk\ell} : \Theta \to \mathbf{R}$ be a function defined by

$$R_{ijk\ell}(c) := \begin{cases} \frac{c_{ij} c_{k\ell}}{c_{i\ell} c_{kj}} & (c_{ij} c_{k\ell} c_{i\ell} c_{kj} \neq 0) \\ 0 & (c_{ij} c_{k\ell} c_{i\ell} c_{kj} = 0) \end{cases} \quad (c = (c_{ij}) \in \Theta).$$

Note that $R_{ijk\ell}$ is a function obtained from the odds ratio. For $1 \leq i \leq r_1$ and $1 \leq j \leq r_2$, we define a function $Z_{ij} : \Theta \to \mathbf{R}$ by

$$Z_{ij}(c) := \begin{cases} 1 & (c_{ij} > 0) \\ 0 & (c_{ij} = 0) \end{cases} \quad (c = (c_{ij}) \in \Theta).$$

The functions $Z_{ij}\,(1 \leq i \leq r_1,\, 1 \leq j \leq r_2)$ hold information on the position of zero cells. The functions $R_{ijk\ell}$ and Z_{ij} are invariant with respect to the action of group G. By Lemma 3, random variables $R_{ijk\ell}(\theta)$ and $Z_{ij}(\theta)$ are $\mathcal{O}$-measurable.

The following theorem states that $A\theta$ is a nuisance parameter.

Theorem 5 *The following equation holds:*

$$\sigma(AX,\theta) = \sigma(A\theta,\mathcal{O}).$$

Corollary 1 *$\sigma(A\theta)$ is nuisance for $(\sigma(X),\mathcal{O})$.*

Proof By Theorem 3, for any $Y \in \mathcal{L}^1(\sigma(X))$, $\mathbf{E}\left(Y|\sigma\left(AX,\theta\right)\right)$ is $\mathcal{O}$-measurable. The equation in Therorem 5 implies $\mathbf{E}\left(Y|\sigma\left(AX,\theta\right)\right) = \mathbf{E}\left(Y|\sigma\left(A\theta,\mathcal{O}\right)\right)$. Hence, $\mathcal{O}$ is of interest with respect to $(\sigma(X),\sigma\left(A\theta,\mathcal{O}\right))$. Therefore $\sigma(A\theta)$ is nuisance for $(\sigma(X),\mathcal{O})$. $\square$

To show Theorem 5, we prepare the following lemma:

Lemma 13 *Let $F : \mathbf{R}^n \to \mathbf{R}^d$ be a linear map and $\iota : \mathbf{R}^n_{>0} \to \mathbf{R}^n$ be the inclusion. For $\alpha \in \ker F$, let $R_\alpha : \mathbf{R}^n_{>0} \to \mathbf{R}$ be a function defined by $R_\alpha(x) := \prod_{i=1}^n x_i^{\alpha_i}$. Then, we have*

$$\mathcal{B}(\mathbf{R}^n_{>0}) = \sigma(F\iota, R_\alpha; \alpha \in \ker F).$$

Proof It is ovbious that the left-hand side includes the right-hand side. We show the opposite inclusion. Let $\{\alpha_1,\dots,\alpha_m\}$ be a basis of $\ker F$. Then the differential map

$$\varphi : \mathbf{R}^n_{>0} \to (\mathrm{Im}F\iota) \times \mathbf{R}^m_{>0} \quad (x \mapsto (F\iota(x),\, R_{\alpha_1}(x),\dots,R_{\alpha_m}(x)))$$

is surjective. By the general theory of the exponential family ([32, p. 125], [5, Theorem 3. 6]), φ is also injective. Hence, φ is a diffeomorphism between $\mathbf{R}^n_{>0}$ and $(\mathrm{Im}F\iota) \times \mathbf{R}^m_{>0}$, and we have

$$\begin{aligned}\mathcal{B}(\mathbf{R}^n_{>0}) &= \varphi^{-1}\mathcal{B}\left((\mathrm{Im}F\iota)\times\mathbf{R}^m_{>0}\right) = \varphi^{-1}\sigma\left(p,q_1,\dots,q_m\right) = \sigma\left(p\varphi,q_1\varphi,\dots,q_m\varphi\right)\\ &= \sigma\left(F\iota,\, R_{\alpha_1}(x),\dots,R_{\alpha_m}\right) \subset \sigma(F\iota,R_\alpha;\alpha\in\ker F)\end{aligned}$$

Here, $p : (\mathrm{Im}F\iota)\times\mathbf{R}^m_{>0} \to (\mathrm{Im}F\iota)$ and $q_i : (\mathrm{Im}F\iota)\times\mathbf{R}^m_{>0} \to \mathbf{R}_{>0}$ are the projections. $\square$

Proof of Theorem 5 Recall that, for $z \in \{0,1\}^n$, we put $\Theta_z = \{c \in \Theta \,|\, Z(c) = z\}$ and that $\iota_z : \Theta_z \to \Theta$ is the inclusion. Applying Lemma 13 in the case where $F = A\iota_z$, we have

$$\mathcal{B}(\Theta_z) = \sigma(A\iota_z, R_\alpha\iota_z; \alpha \in \ker A\iota_z). \tag{25}$$

The equation $\ker A\iota_z = \mathbf{R}^{J(z)} \cap \ker A$ implies

$$\sigma\left(A\iota_z,\, R_\alpha\iota_z : \alpha\in\ker A\iota_z\right) = \sigma\left(A\iota_z,\, R_\alpha\iota_z : \alpha \in \mathbf{R}^{J(z)}\cap\ker A\right) \tag{26}$$

By Equations (25) and (26), we have

$$\begin{aligned}\mathcal{B}(\Theta_z) &= \sigma\left(A\iota_z,\, R_\alpha\iota_z : \alpha\in\mathbf{R}^{J(z)}\cap\ker A\right) \subset \iota_z^{-1}\sigma\left(A,\, R_\alpha : \alpha\in\mathbf{R}^{J(z)}\cap\ker A\right)\\ &\subset \iota_z^{-1}\sigma\left(A,\, R_\alpha,\, Z : \alpha\in\ker A\right) \subset \sigma\left(A,\, R_\alpha,\, Z : \alpha\in\ker A\right).\end{aligned} \tag{27}$$

Any $B \in \mathcal{B}(\Theta)$ can be decomposed as $B = \bigcup_{z\in\{0,1\}^n}\left(B\cap\Theta_z\right) = \bigcup_{z\in\{0,1\}^n}\iota_z^{-1}B$. By (27), B is an element of $\sigma\left(A,\, R_\alpha,\, Z : \alpha\in\ker A\right)$. Hence, we have

$$\mathcal{B}(\Theta) = \sigma\left(A,\, R_\alpha,\, Z : \alpha\in\ker A\right) \tag{28}$$

The σ-algebra generated by θ is the pull-back of the left-hand side of (28) with respect to θ. By Lemma 6, the pull-back of the right-hand side of (28) equals to $\sigma\left(A\theta,\, R_\alpha(\theta),\, Z(\theta) : \alpha\in\ker A\right)$. Hence, we have

$$\sigma\left(\theta\right) = \sigma\left(A\theta,\, R_\alpha(\theta),\, Z(\theta) : \alpha\in\ker A\right).$$

This equation implies

$$\sigma\left(AX,\,\theta\right) = \sigma\left(AX,\, A\theta,\, R_\alpha(\theta),\, Z(\theta) : \alpha\in\ker A\right)$$

By Theorem 4, we have

$$\begin{aligned}\sigma\left(A\theta,\mathcal{O}\right) &= \sigma\left(A\theta,\, \sigma\left(AX,\, R_\alpha(\theta),\, Z(\theta) : \alpha\in\ker A\right)\right)\\ &= \sigma\left(A\theta,\, AX,\, R_\alpha(\theta),\, Z(\theta) : \alpha\in\ker A\right) = \sigma\left(AX,\,\theta\right)\end{aligned}$$

$\square$

9 Examples of CMLE problems

Theorem 4 and 5 claim that when AX is given, $\sigma(R_\alpha(\theta), Z(\theta))$ are of interest and $\sigma(A\theta)$ is a nuisance. In the case of contingency tables, generalized odds ratios $R_\alpha(p)$ and positions of zero cells $Z(p)$ are of interest and row and column probabilities Ap are a nuisance when the marginal sums of the table are given. We present examples of estimating generalized odds ratios by CMLE.

Example 10 We generate categorical data concerning the number of hours slept and time of going to bed from a student sample in the LearnBayes package [6] of the system R for statistical computing.

Rows are categorized by time spent sleeping. The categories are sleeping less than 6 hours, 6–7 hours, and more than 7 hours. Columns are categorized by the time of going to bed. The categories are going to bed before midnight, between midnight and 1am, and after 1am. We wish to analyze these categorical data by the Poisson random model $U_{ij} \sim \mathrm{Pois}(p_{ij})$. The independence of rows and columns is rejected by the χ^2 test with the threshold p-value 0.05. Then, we regard the column sum $\sum_i p_{ij}$ and the row sum $\sum_j p_{ij}$ as nuisance parameters. These represent probabilities of the event standing for j-th row and one standing for i-th column when the rows and the columns are independent. We perform CMLE under the condition that column sums $\sum_i u_{ij}$ and row sums $\sum_j u_{ij}$ are given.

Categorical data for all:

Bed time \ Hours slept	less than 6 hour	6–7	more than 7 hours
Before 24	1	6	123
24–25	3	22	145
After 25	86	91	176

We omit titles and express this table as $\begin{pmatrix} 1 & 6 & 123 \\ 3 & 22 & 145 \\ 86 & 91 & 176 \end{pmatrix}$. Categorical data for males:

$$\begin{pmatrix} 1 & 2 & 28 \\ 0 & 4 & 47 \\ 35 & 32 & 71 \end{pmatrix}$$

Categorical data for females:

$$\begin{pmatrix} 0 & 4 & 95 \\ 3 & 18 & 98 \\ 51 & 59 & 105 \end{pmatrix}$$

Because this CMLE can be solved by the $\mathcal{A}$-distribution discussed previously, we apply our algorithm for evaluating normalizing constants and their derivatives to the method for estimating the conditional maximum likelihood in [32, §4]. We obtain the following estimates. CMLE (p_{ij}) for all:

$$\begin{pmatrix} 0.176556059977815 & 1 & 10.5634953362788 \\ 0.144532927997885 & 1 & 3.39969669537228 \\ 1 & 1 & 1 \end{pmatrix}$$

CMLE for males:

$$\begin{pmatrix} 0.458167657900967 & 1 & \underline{6.25676090279981} \\ 0 & 1 & \underline{5.25200491199345} \\ 1 & 1 & 1 \end{pmatrix}$$

CMLE for females:

$$\begin{pmatrix} 0 & 1 & \underline{13.2714773737657} \\ 0.193351042187373 & 1 & \underline{3.04872586155291} \\ 1 & 1 & 1 \end{pmatrix}$$

As explained in the previous section, the space of parameters of interest should be regarded as the collection of different orbits by the torus action. When the parameter value obtained via CMLE is (p_{ij}), values on the orbit $(g_i h_j p_{ij})$, $g_i, h_j \in \mathbb{R}_{>0}$ are equivalent parameters. Since the normalized elements of the

[6] `https://cran.r-project.org/web/packages/LearnBayes/index.html`

second column and the third row are 1, we have $g_3h_1 = g_3h_2 = g_3h_3 = 1$ and $g_1h_2 = g_2h_2 = g_3h_2 = 1$. Then, we have $g_ih_j = 1$ for all i, j. The condition whereby this normalization is possible ($p_{i2} \neq 0$, $p_{3j} \neq 0$) defines a subspace of the parameters of interest. The subspace is isomorphic to $\mathbb{R}^4_{\geq 0}$ by the quotient topology. The correspondence is given by

$$(p_{ij}) \mapsto \begin{pmatrix} \frac{\underline{p_{11}p_{32}}}{p_{12}p_{31}} & 1 & \frac{\underline{p_{13}p_{32}}}{p_{12}p_{33}} \\ \frac{\underline{p_{21}p_{32}}}{p_{22}p_{31}} & 1 & \frac{\underline{p_{23}p_{32}}}{p_{22}p_{33}} \\ 1 & 1 & 1 \end{pmatrix} \tag{29}$$

In this chart, males and females exhibit different tendencies. For example, the underlined values at $(1,3)$ and $(2,3)$ positions are close in the case of males but not for females.

The number obtained by replacing p_{ij} by the frequency u_{ij} in (29) is called a generalized odds ratio. Generalized odds ratios for our data are as follows. Odds ratios for all:

$$\begin{pmatrix} 0.176356589147287 & 1 & 10.5994318181818 \\ 0.144291754756871 & 1 & 3.40779958677686 \\ 1 & 1 & 1 \end{pmatrix}$$

Odds ratios for males:

$$\begin{pmatrix} 0.457142857142857 & 1 & 6.30985915492958 \\ 0 & 1 & 5.29577464788732 \\ 1 & 1 & 1 \end{pmatrix}$$

Odds ratios for females:

$$\begin{pmatrix} 0 & 1 & 13.3452380952381 \\ 0.19281045751634 & 1 & 3.05925925925926 \\ 1 & 1 & 1 \end{pmatrix}$$

Note that, as proved in [32, Theorem 5], these generalized odds ratios approximate CMLE because we have a sufficient sample size.

When the sample size is relatively small, a generalized odds ratio may not approximate the corresponding CMLE well. We present one example.

Example 11 The categorical data below are taken from emergency safety information on diclofenac sodium for influenza encephalitis and encephalopathy[7].
Categorical data:

	acetaminophen	diclofenac sodium	mefenamic acid
death	4	7	2
survival	32	5	6

We omit titles and express this table as $\begin{pmatrix} 4 & 7 & 2 \\ 32 & 5 & 6 \end{pmatrix}$. By applying our algorithm and the method in [32], we obtain the following CMLE.

$$\begin{pmatrix} 1 & \underline{10.5557279737263} & 2.62096714359908 \\ 1 & 1 & 1 \end{pmatrix}$$

Generalized odds ratios are

$$\begin{pmatrix} 1 & \underline{11.2} & 2.66666666666667 \\ 1 & 1 & 1 \end{pmatrix}$$

See the numbers underlined above. We observe that the odds ratio is larger than the CMLE. In other words, the effect of nuisance parameters increases the risk in this case. Finally, we briefly note how subsequent data released from the same institute in 2001 appeared to show that diclofenac sodium was in fact more associated with survival, rather than death. This reminds us of some of the difficulties inherent in statistical analyses. Here are those new data: [8].

[7] Pharmaceuticals and Medical Devices Agency, Japan, 2000, `https://www.pmda.go.jp/files/000148557.pdf`
[8] `http://idsc.nih.go.jp/disease/influenza/iencepha.html`

	acetaminophen	diclofenac sodium	mefenamic acid
death	23	13	6
survival	78	25	9

Our algorithm outputs CMLE

$$\begin{pmatrix} 1 & 1.7567483756645 & 2.24788463785377 \\ 1 & 1 & 1 \end{pmatrix}$$

and odds ratios:

$$\begin{pmatrix} 1 & 1.76347826086957 & 2.26086956521739 \\ 1 & 1 & 1 \end{pmatrix}.$$

10 Appendix

We will explain the derivation of the matrix U_2 of Example 5 with twisted cohomology groups by following [7] and the program `gtt_ekn3/ekn_pfaffian_8.rr` of the package `gtt_ekn3`.

We start with the integral representation of ${}_2F_1$:

$$\frac{\Gamma(b)\Gamma(c-b)}{\Gamma(c)} \cdot {}_2F_1(a,b,c;x) = \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-xt)^{-a}dt = (-1)^b \int_0^{-1} t^b(1+xt)^{-a}(1+t)^{c-b-1}\frac{dt}{t}.$$

We rename the parameters a, b, c by

$$(\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (a-c+1, b, -a, c-b-1).^9$$

The decrement of a stands for an increment of α_2 (and decrement of α_0). The identity we want to derive is $F(a) = M(a)F(a+1)$, which is a special case of

$$\mathbf{S}(\alpha;x) = \frac{1}{\alpha_2}U_2(\alpha_{(2)};x)\mathbf{S}(\alpha_{(2)};x), \quad \alpha_{(2)} := (\alpha_0+1, \alpha_1, \alpha_2-1, \alpha_3)$$

in [7, Corollary 6.3] ($\alpha_{(2)}$ stands for $a+1$). The function `upAlpha(2,1,1)` in the program derives $\frac{1}{\alpha_2}U_2$. $\mathbf{S}(\alpha;x)$ is the vector consisting of the hypergeometric series $S(\alpha;x)$ defined in [7, Section 6] and its derivatives (Gauss-Manin vector). When $c \in \mathbb{N}_0$, it can be expressed in terms of ${}_2F_1$ as

$$\mathbf{S}(\alpha;x) = \begin{pmatrix} S \\ \frac{1}{\alpha_2}\theta_x S \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha_2 \end{pmatrix} \begin{pmatrix} S \\ \theta_x S \end{pmatrix} = \frac{1}{(-a)!(-b)!(c-1)!} \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha_2 \end{pmatrix} \begin{pmatrix} {}_2F_1 \\ \theta_x {}_2F_1 \end{pmatrix}.$$

Hence, the matrix $M(a)$ can be expressed as

$$M(a) = -a \begin{pmatrix} 1 & 0 \\ 0 & \alpha_2 \end{pmatrix} \left(\frac{1}{\alpha_2}U_2(\alpha_{(2)})\right) \begin{pmatrix} 1 & 0 \\ 0 & 1/(\alpha_2-1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \alpha_2 \end{pmatrix} U_2(\alpha_{(2)}) \begin{pmatrix} 1 & 0 \\ 0 & 1/(\alpha_2-1) \end{pmatrix}.$$

It follows from [7, Theorem 5.3] that the representation matrix U_2 can be expressed as

$$U_2(\alpha_{(2)};x) = C(\alpha)P_2(\alpha)^{-1}D_2(x)Q_2(\alpha_{(2)})C(\alpha_{(2)})^{-1}.$$

We use the notation $|\tilde{x}\langle ij\rangle|$, which is the determinant of the minor matrix consisting of the i-th column and the j-th column of the matrix $\tilde{x} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & x & 1 \end{pmatrix}$, where the numbering starts with 0 (see [7] as to details). We put $\varphi\langle ij\rangle = \frac{|\tilde{x}\langle ij\rangle|dt}{L_i L_j}$, where $L_0 = 1$, $L_1 = t$, $L_2 = 1+xt$, and $L_3 = 1+t$. We have the following expressions with these notations.

$$\begin{aligned}
D_2(x) &= \mathrm{diag}\left(\frac{|\tilde{x}\langle 21\rangle|}{|\tilde{x}\langle 01\rangle|}, \frac{|\tilde{x}\langle 23\rangle|}{|\tilde{x}\langle 03\rangle|}\right) = \mathrm{diag}(1, 1-x) = \begin{pmatrix} 1 & 0 \\ 0 & 1-x \end{pmatrix}, \\
C(\alpha) &= \begin{pmatrix} \mathcal{I}(\varphi\langle 01\rangle, \varphi\langle 01\rangle) & \mathcal{I}(\varphi\langle 01\rangle, \varphi\langle 02\rangle) \\ \mathcal{I}(\varphi\langle 02\rangle, \varphi\langle 01\rangle) & \mathcal{I}(\varphi\langle 02\rangle, \varphi\langle 02\rangle) \end{pmatrix} = 2\pi\sqrt{-1}\begin{pmatrix} \frac{1}{\alpha_0}+\frac{1}{\alpha_1} & \frac{1}{\alpha_0} \\ \frac{1}{\alpha_0} & \frac{1}{\alpha_0}+\frac{1}{\alpha_2} \end{pmatrix}, \\
Q_2(\alpha) &= \begin{pmatrix} \mathcal{I}(\varphi\langle 01\rangle, \varphi\langle 01\rangle) & \mathcal{I}(\varphi\langle 01\rangle, \varphi\langle 02\rangle) \\ \mathcal{I}(\varphi\langle 03\rangle, \varphi\langle 01\rangle) & \mathcal{I}(\varphi\langle 03\rangle, \varphi\langle 02\rangle) \end{pmatrix} = 2\pi\sqrt{-1}\begin{pmatrix} \frac{1}{\alpha_0}+\frac{1}{\alpha_1} & \frac{1}{\alpha_0} \\ \frac{1}{\alpha_0} & \frac{1}{\alpha_0} \end{pmatrix}, \\
P_2(\alpha) &= \begin{pmatrix} \mathcal{I}(\varphi\langle 21\rangle, \varphi\langle 01\rangle) & \mathcal{I}(\varphi\langle 21\rangle, \varphi\langle 02\rangle) \\ \mathcal{I}(\varphi\langle 23\rangle, \varphi\langle 01\rangle) & \mathcal{I}(\varphi\langle 23\rangle, \varphi\langle 02\rangle) \end{pmatrix} = 2\pi\sqrt{-1}\begin{pmatrix} \frac{1}{\alpha_1} & -\frac{1}{\alpha_2} \\ 0 & -\frac{1}{\alpha_2} \end{pmatrix},
\end{aligned}$$

[9] $\alpha_0 = -\alpha_1 - \alpha_2 - \alpha_3$ stands for the exponent at infinity.

where $\mathcal{I}$ is the intersection form on the twisted cohomology group. The inverse matrices of them can also be expressed in terms of intersection numbers as in [7, Appendix]. This method is implemented as the function `invintMatrix_k` in our package and it outputs

$$
\begin{aligned}
P_2(\alpha)^{-1} &= \frac{1}{(2\pi\sqrt{-1})^2}\begin{pmatrix}\alpha_1 & 0\\ 0 & \alpha_2\end{pmatrix}\begin{pmatrix}\mathcal{I}(\varphi\langle 31\rangle,\varphi\langle 01\rangle) & \mathcal{I}(\varphi\langle 31\rangle,\varphi\langle 03\rangle)\\ \mathcal{I}(\varphi\langle 32\rangle,\varphi\langle 01\rangle) & \mathcal{I}(\varphi\langle 32\rangle,\varphi\langle 03\rangle)\end{pmatrix}\begin{pmatrix}\alpha_1 & 0\\ 0 & \alpha_3\end{pmatrix}\\
&= \frac{1}{2\pi\sqrt{-1}}\begin{pmatrix}\alpha_1 & 0\\ 0 & \alpha_2\end{pmatrix}\begin{pmatrix}\frac{1}{\alpha_1} & -\frac{1}{\alpha_3}\\ 0 & -\frac{1}{\alpha_3}\end{pmatrix}\begin{pmatrix}\alpha_1 & 0\\ 0 & \alpha_3\end{pmatrix} = \frac{1}{2\pi\sqrt{-1}}\begin{pmatrix}\alpha_1 & -\alpha_1\\ 0 & -\alpha_2\end{pmatrix},\\
C(\alpha)^{-1} &= \frac{1}{(2\pi\sqrt{-1})^2}\begin{pmatrix}\alpha_1 & 0\\ 0 & \alpha_2\end{pmatrix}\begin{pmatrix}\mathcal{I}(\varphi\langle 31\rangle,\varphi\langle 31\rangle) & \mathcal{I}(\varphi\langle 31\rangle,\varphi\langle 32\rangle)\\ \mathcal{I}(\varphi\langle 32\rangle,\varphi\langle 31\rangle) & \mathcal{I}(\varphi\langle 32\rangle,\varphi\langle 32\rangle)\end{pmatrix}\begin{pmatrix}\alpha_1 & 0\\ 0 & \alpha_2\end{pmatrix}\\
&= \frac{1}{2\pi\sqrt{-1}}\begin{pmatrix}\alpha_1 & 0\\ 0 & \alpha_2\end{pmatrix}\begin{pmatrix}\frac{1}{\alpha_3}+\frac{1}{\alpha_1} & \frac{1}{\alpha_3}\\ \frac{1}{\alpha_3} & \frac{1}{\alpha_3}+\frac{1}{\alpha_2}\end{pmatrix}\begin{pmatrix}\alpha_1 & 0\\ 0 & \alpha_2\end{pmatrix} = \frac{\alpha_1\alpha_2}{2\pi\sqrt{-1}\cdot\alpha_3}\begin{pmatrix}\frac{\alpha_1+\alpha_3}{\alpha_2} & 1\\ 1 & \frac{\alpha_2+\alpha_3}{\alpha_1}\end{pmatrix}.
\end{aligned}
$$

These matrices can be obtained in our program as

$$
\begin{aligned}
&D_2(x) = \mathtt{repMatrix}(2,1,1), && C(\alpha)/(2\pi\sqrt{-1}) = \mathtt{intMatrix}([0,3],[0,3],1,1),\\
&P_2(\alpha)/(2\pi\sqrt{-1}) = \mathtt{intMatrix}([2,0],[0,3],1,1), && Q_2(\alpha)/(2\pi\sqrt{-1}) = \mathtt{intMatrix}([0,2],[0,3],1,1),\\
&(2\pi\sqrt{-1})P_2(\alpha)^{-1} = \mathtt{invintMatrix\_k}([2,0],[0,3],1,1), && (2\pi\sqrt{-1})C(\alpha)^{-1} = \mathtt{invintMatrix\_k}([0,3],[0,3],1,1)
\end{aligned}
$$

(the argument $(1,1)$ stands for (r_1-1, r_2-1)).

Acknowledgement: This work was supported by MEXT/JSPS KAKENHI Grant Numbers JP 25220001, 17K05279, 18J01507 and by Research Institute for Mathematical Sciences, a Joint Usage/Research Center located in Kyoto University. We deeply appreciate several constructive criticisms by the reviewers, which made big improvements of our algorithms and implementation.

References

[1] A. Agresti, Categorical Data Analysis, 3rd edition, Wiley Series in Probability and Statistics, Wiley, 2013.

[2] S. Amari, Information Geometry and Its Application, Applied Mathematical Sciences, Springer Japan, 2016.

[3] P. Billingsley, Probability and Measure, Wiley Series in Probability and Statistics, Wiley, 1995.

[4] R. P. Brent, P. Zimmermann, Modern Computer Arithmetic, Cambridge University Press, 2012.

[5] L. D. Brown, Fundamentals of statistical exponential families with applications in statistical decision theory. *Lecture Notes-Monograph Series*, 9:i–279, 1986.

[6] J. von zur Gathen, J. Gerhard, Modern Computer Algebra, Cambridge University Press, 2003.

[7] Y. Goto, K. Matsumoto, Pfaffian equations and contiguity relations of the hypergeometric function of type$(k+1, k+n+2)$ and their applications, Funkcialaj Ekvacioj 61 (2018), 315–347.

[8] Y. Goto, Contiguity relations of Lauricella's F_D revisited, Tohoku Mathematical Journal 69 (2017), 287–304.

[9] Y. Goto, Intersection numbers of twisted cycles and cocycles for degenerate arrangements, arXiv:1805.01714.

[10] T. Hibi et al, Gröbner Bases : Statistics and Software Systems, Springer, 2013.

[11] J. van der Hoeven, Faster chinese remaindering, 2016, hal-01403801

[12] I. Karatzas and S. Shreve, Brownian Motion and Stochastic Calculus, Graduate texts in mathematics, Springer-Verlag, 1988.

[13] T. Koyama, A new formulation of sufficient σ-algebra, in preparation.

[14] D. Landers and L. Rogge, Minimal sufficient σ-fields and minimal sufficient statistics two counterexamples, The Annals of Mathematical Statistics 43, 6 (1972), 2045–2049.

[15] LINBOX: exact computational linear algebra. `http://www.linalg.org`

[16] N. Möller, On Schönhage's algorithm and subquadratic integer gcd computation, Mathematics of Computation 77 (2008), 589–607.

[17] H. Nakayama, K. Nishiyama, M. Noro, K. Ohara, T. Sei, N. Takayama, A. Takemura, Holonomic gradient descent and its application to Fisher-Bingham integral, Advances in Applied Mathematics 47 (2011), 639-658.

[18] M. Noro, N. Takayama, Risa/Asir Drill 2012, `http://www.math.kobe-u.ac.jp/Asir`, 2012 (in Japanese).

[19] M. Noro, K. Yokoyama, A modular method to compute the rational univariate representation of zero-dimensional ideals. Journal of Symbolic Computation 28/1(1999), 243-263.

[20] M. Noro, K. Yokoyama, Computation of Gröbner Basis — introduction to computational algebra, University of Tokyo press, 2003 (in Japanese).

[21] M. Ogawa, Algebraic Statistical Methods for Conditional Inference of Discrete Statistical Models, PhD thesis, University of Tokyo, 2015.

[22] K. Ohara, N. Takayama, Pfaffian systems of A-hypergeometric systems II — holonomic gradient method, arXiv:1505.02947.

[23] T. Oshima, Fractional Calculus of Weyl Algebra and Fuchsian Differential Equations, MSJ Memoirs 28, 2013.

[24] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical Recipes 3rd Edition: The Art of Scientific Computing, Cambridge University Press, 2007.

[25] References for HGM
`http://www.math.kobe-u.ac.jp/OpenXM/Math/hgm/ref-hgm.html`

[26] Risa/Asir, a computer algebra system. `http://www.math.kobe-u.ac.jp/Asir`

[27] T. Sasaki, T. Takeshima, A modular method for Gröbner-basis construction over $\mathbb{Q}$ and solving system of algebraic equations. Journal of Information Processing, 12(1989), no.4, 371-379.

[28] M. Schervish, Theory of Statistics, Springer, 1995.

[29] J. Stoer, R. Bulirsch, Introduction to Numerical Analysis, Springer-Verlag, New York-Heidelberg, 1980.

[30] Y. Tachibana, Y. Goto, N. Takayama, An implementation of the difference holonomic gradient method to two way contingency tables, RIMS Kokyuroku, 2054(2017), 105–117 (in Japanese).

[31] N. Takayama, Gröbner basis and the problem of contiguous relations, Japan Journal of Applied Mathematics, 6 (1989), 147-160.

[32] N. Takayama, S. Kuriki, A. Takemura, A-hypergeometric distributions and Newton polytopes, Advances in Applied Mathematics 99 (2018), 109 – 133.

[33] T. Granlund, and the GMP development team, GNU Multiple Precision Arithmetic Library, 1991–2019, `http://gmplib.org`

[34] D. Williams, Probability with Martingales, Cambridge mathematical textbooks, Cambridge University Press, 1991.