11:2 2020

ASTAT

Algebraic Statistics

msp

msp.org/astat

See inside back cover or msp.org/astat for submission instructions.

Algebraic Statistics (ISSN 2693-3004 electronic, 2693-2997 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

AStat peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

Algebraic Statistics
Vol. 11, No. 2, 2020
https://doi.org/10.2140/astat.2020.11.107

MAXIMUM LIKELIHOOD DEGREE OF THE TWO-DIMENSIONAL LINEAR GAUSSIAN COVARIANCE MODEL

JANE IVY COONS, ORLANDO MARIGLIANO AND MICHAEL RUDDY

In algebraic statistics, the maximum likelihood degree of a statistical model is the number of complex critical points of its log-likelihood function. A priori knowledge of this number is useful for applying techniques of numerical algebraic geometry to the maximum likelihood estimation problem. We compute the maximum likelihood degree of a generic two-dimensional subspace of the space of $n \times n$ Gaussian covariance matrices. We use the intersection theory of plane curves to show that this number is $2n - 3$.

1. Introduction

A linear Gaussian covariance model is a collection of multivariate Gaussian probability distributions whose covariance matrices are linear combinations of some fixed symmetric matrices. In this paper, we will focus on the *two-dimensional linear Gaussian covariance model*, in which all of the covariance matrices in the model lie in a two-dimensional linear space. Linear Gaussian covariance models were first studied by 1] in the context of the analysis of time series models. They continue to be studied towards this end, for example, in [23]. These models also have applications in a variety of other contexts.

One of the most common types of linear Gaussian covariance models consists of covariance matrices with some prescribed zeros. Given a Gaussian random vector $(X_1, \ldots, X_n)$ with mean μ and positive-definite covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$, we can discern independence statements from the zeros in Σ. In particular, the disjoint subvectors $(X_{i_1}, \ldots, X_{i_k})$ and $(X_{j_1}, \ldots, X_{j_l})$ are independent if and only if the submatrix of Σ that consists of rows $i_1, \ldots, i_k$ and columns $j_1, \ldots, j_l$ is the zero matrix [22, Proposition 2.4.4].

Maximum likelihood estimation for covariance matrices with a fixed independence structure was studied in [7]. These types of models find applications in the study of gene expression using relevance networks [5]. In these networks, genes are connected with an edge if their expressions are sufficiently correlated. The edges and nonedges in the resulting graph dictate the sparsity structure of the covariance matrix. Problems related to estimation of sparse covariance matrices have been studied in [3; 18].

Linear Gaussian covariance models are also applicable to the field of phylogenetics. In particular, Brownian motion tree models, which model evolution of normally distributed traits along an evolutionary tree, are linear Gaussian covariance models [11]. The covariance matrices of Brownian motion tree models require linear combinations of more than two matrices. However, the authors believe that the

MSC2010: 13P25, 14C17, 14H50, 62H12.
Keywords: algebraic geometry, algebraic statistics, linear Gaussian covariance models, intersection theory, plane curves, maximum likelihood estimation, maximum likelihood degree.

results in this paper will find applications to mixtures of Brownian motion tree models. These apply, for example, to models of trait evolution that consider two genes instead of just one [16].

Algorithms for computing the maximum likelihood estimate for generic linear Gaussian covariance models have been the subject of much study [1; 2; 3; 7]. Zwiernik, Uhler, and Richards have shown that when the number of data points is sufficiently large, maximum likelihood estimation for such models behaves like a convex optimization problem in a large convex region containing the maximum likelihood estimate [24].

In this paper, we are concerned with computing the maximum likelihood degree (ML-degree) of the two-dimensional linear Gaussian covariance model for generic parameters and data. This is the number of complex critical points of the log-likelihood function, and it is considered to be a measurement of the difficulty of computing the maximum likelihood estimate (MLE) [21, Table 3]. Knowledge of the ML-degree of a model is important when applying numerical algebraic geometry methods to solve the MLE problem; in particular, it gives a stopping criterion for monodromy methods [21, §5]. For more background on ML-degrees, we refer the reader to [6; 10, Chapter 2].

2. Preliminaries

Let n be a natural number, and let $\mathrm{PD}_n \subset \mathbb{R}^{\binom{n+1}{2}}$ be the cone of all $n \times n$ symmetric positive-definite matrices. We view PD_n as the space of covariance matrices of all normal distributions $\mathcal{N}(0, \Sigma)$ with mean zero.

In algebraic statistics, a Gaussian statistical model is an algebraic subset of PD_n. In this paper, we consider models of the form

$$\mathcal{M}_{A,B} = \{xA + yB \mid x, y \in \mathbb{R}\} \cap \mathrm{PD}_n$$

for symmetric matrices A and B, whenever the intersection is not empty. That is, $\mathcal{M}_{A,B}$ is the intersection of the positive-definite cone with the linear span of A and B. We call $\mathcal{M}_{A,B}$ the *two-dimensional linear Gaussian covariance model* with respect to A and B.

Given independent, identically distributed (i.i.d.) samples $u_1, \ldots, u_r \in \mathbb{R}^n$ from some normal distribution, the maximum likelihood estimation problem for $\mathcal{M}_{A,B}$ is to find a covariance matrix $\widehat{\Sigma} \in \mathcal{M}_{A,B}$, if one exists, that maximizes the value of the likelihood function

$$L(\Sigma \mid u_1, \ldots, u_r) = \prod_{i=1}^{r} f_\Sigma(u_i),$$

where f_Σ is the density of $\mathcal{N}(0, \Sigma)$. Let S denote the *sample covariance matrix*

$$S = \frac{1}{r} \sum_{i=1}^{r} u_i u_i^T.$$

Since for all Σ the value $L(\Sigma \mid u_1, \ldots, u_r)$ only depends on S, we identify the data given by r i.i.d. samples from a normal distribution with their sample covariance matrix S. The logarithm is a concave function, so the maximizer of the likelihood function is also the maximizer of its natural log, the *log-likelihood*

function. This function can be written in terms of S:

$$\begin{aligned}\ell(\Sigma \mid S) &:= \log L(\Sigma \mid S) \\ &= -\frac{rn}{2}\log(2\pi) - \frac{r}{2}\log\det(\Sigma) - \frac{r}{2}\operatorname{tr}(S\Sigma^{-1}).\end{aligned}$$

Note that the maximizer of this function is equal to the minimizer of

$$\tilde{\ell}(\Sigma \mid S) := \log\det(\Sigma) + \operatorname{tr}(S\Sigma^{-1}).$$

When we restrict to the model $\mathcal{M}_{A,B}$, we require that $\Sigma = xA + yB$ for some $x, y \in \mathbb{R}$ such that $xA + yB$ is positive-definite. So the maximum likelihood estimation problem in this case is equivalent to

$$\operatorname*{argmin}_{x,y} \tilde{\ell}(xA + yB \mid S) \quad \text{subject to} \quad xA + yB \in \mathrm{PD}_n.$$

To find local extrema of the log-likelihood function, we set its gradient equal to 0 and solve for x and y. The two resulting equations are called the score equations.

Definition 2.1. The *score equations* for $\mathcal{M}_{A,B}$ are the partial derivatives of the function $\tilde{\ell}(xA + yB \mid S)$ with respect to x and y. The *maximum likelihood degree* or *ML-degree* of $\mathcal{M}_{A,B}$ is the number of complex solutions to the score equations, counted with multiplicity, for a generic sample covariance matrix S.

Definition 2.1 makes reference to a *generic* sample covariance matrix. We give a detailed explanation of this term from algebraic geometry at the end of this section.

One benefit of working with $\tilde{\ell}$ is that the score equations are rational functions of the data. This allows us to use tools from algebraic geometry to analyze their solutions. Let $\Sigma = xA + yB$. For the sake of brevity, we will denote $P(x, y) = \det\Sigma$ and $T(x, y) = \operatorname{tr}(S \operatorname{adj}\Sigma)$, where $\operatorname{adj}\Sigma$ is the classical adjoint. With this notation, the function $\tilde{\ell}$ takes the form

$$\tilde{\ell}(\Sigma \mid S) = \log P + \frac{T}{P}.$$

Accordingly, the score equations are

$$\begin{aligned}\tilde{\ell}_x(x, y) &= \frac{P_x}{P} + \frac{PT_x - TP_x}{P^2}, \\ \tilde{\ell}_y(x, y) &= \frac{P_y}{P} + \frac{PT_y - TP_y}{P^2}.\end{aligned}$$

Here and throughout, the notation h_x is used for the derivative of a function h with respect to the variable x. We are concerned with values of $(x, y) \in \mathbb{C}^2$ where both of the score equations are zero. We clear denominators by multiplying $\tilde{\ell}_x$ and $\tilde{\ell}_y$ by P^2 to obtain two polynomials,

$$\begin{aligned}f(x, y) &:= PP_x + PT_x - TP_x, \\ g(x, y) &:= PP_y + PT_y - TP_y.\end{aligned} \tag{1}$$

We note the *degrees* of each relevant term for generic A, B, and S. Specifically, their total degrees with respect to their variables x and y are

$$\begin{aligned} \deg P &= n, \\ \deg P_x = \deg P_y = \deg T &= n-1, \\ \deg T_x = \deg T_y &= n-2. \end{aligned}$$

A polynomial h is called a *homogeneous form* if each of its terms has the same degree. The polynomials f and g can be written as a sum of a homogeneous degree $2n-1$ form with a homogeneous degree $2n-2$ form.

The critical points of $\tilde{\ell}$ are in the variety $V(f,g)$. However, this variety also contains points at which $\tilde{\ell}$ and the score equations are not defined since we cleared denominators. The ideal whose variety is exactly the critical points of $\tilde{\ell}$ is the saturation

$$\begin{aligned} J &= \mathcal{J}(f,g) : \langle P \rangle^{\infty} \\ &:= \{h \in \mathbb{C}[x,y] \mid hP^N \in \mathcal{J}(f,g) \text{ for some } N\}. \end{aligned}$$

Saturating with $P = \det \Sigma$ removes all points in $V(f,g)$ where the determinant is zero and $\tilde{\ell}$ is undefined. For more details on the geometric content of saturation, we refer the reader to Chapter 7 of [22]. We will show that $\mathcal{J}(f,g)$ and hence J are zero-dimensional in Lemmas 3.3 and 3.4. The ML-degree of the model is hence the degree of J. This is the number of isolated points in the variety of J counted with multiplicity. For more background on degrees of general varieties, see [15, Lecture 13; 20, Chapter 4, §1.4].

We now state the main result and offer an outline for its proof, which we follow in the remaining sections.

Theorem 2.2. *For generic $n \times n$ symmetric matrices A and B, the maximum likelihood degree of the two-dimensional linear Gaussian covariance model $\mathcal{M}_{A,B}$ is $2n-3$.*

A key tool used in the proof of Theorem 2.2 is Bézout's theorem, a proof of which can be found in Chapter 5.3 of [13].

Theorem 2.3 (Bézout's theorem). *Let H and K be projective plane curves of degrees d_1 and d_2, respectively. Suppose further that H and K share no common component. Then the intersection of H and K is zero-dimensional and the number of intersection points of H and K, counted with multiplicity, is $d_1 d_2$.*

Let $F(x,y,z)$ and $G(x,y,z)$ denote the homogenizations of f and g with respect to z. Then F and G both define projective plane curves of degree $2n-1$. Lemmas 3.3 and 3.4 will show that F and G do not share a common component. So we can apply Bézout's theorem to count their intersection points.

Let $q = [x : y : z]$ be a point in $\mathbb{CP}^2$. Then by Bézout's theorem,

$$(2n-1)^2 = \sum_{q \in V(F,G)} I_q(F,G),$$

where $I_q(F, G)$ denotes the *intersection multiplicity* of F and G at q. The definition and properties of the intersection multiplicity of a pair of algebraic curves at a point can be found in [13, §3, Theorem 3]. For affine points $(x, y) \in V(f, g)$ we sometimes denote the intersection multiplicity as $I_{(x,y)}(f, g) := I_{[x:y:1]}(F, G)$.

We show in Proposition 3.5 that saturating the ideal $\mathcal{J}(f, g)$ with $\det \Sigma$ corresponds to removing only the origin from the affine variety of f and g. This in turn corresponds to removing the point $[0:0:1]$ from the projective variety $V(F, G)$. Since we are only interested in affine intersection points of F and G outside of the origin, we split the sum on the right-hand side of the above equation as follows:

$$(2n-1)^2 = I_{[0:0:1]}(F, G) + \sum_{\substack{q \in V(F,G) \\ q \notin \{[0:0:1]\} \cup V(F,G,z)}} I_q(F, G) + \sum_{q \in V(F,G,z)} I_q(F, G). \tag{2}$$

The middle term of the right-hand side of (2) is exactly the degree of the saturated ideal $J = \mathcal{J}(f, g) : \langle \det \Sigma \rangle^\infty$. Thus one can find the degree of J by computing the intersection multiplicities of F and G at the origin and at their intersection points at infinity. We compute the former in Section 4 and the latter in Section 5 to obtain

$$I_{[0:0:1]}(F, G) = (2n-2)^2 \quad \text{and} \quad \sum_{q \in V(F,G,z)} I_q(F, G) = 2n$$

for generic A, B, and S. Thus, by rearranging (2),

$$\sum_{\substack{q \in V(F,G) \\ q \notin \{[0:0:1]\} \cup V(F,G,z)}} I_q(F, G) = (2n-1)^2 - (2n-2)^2 - 2n = 2n - 3,$$

which implies $\deg J = 2n - 3$.

Example 2.4. Let $n = 3$ and consider the model $\mathcal{M}_{A,B}$ defined by the positive-definite matrices

$$A = \begin{pmatrix} 5 & 1 & 0 \\ 1 & 3 & -2 \\ 0 & -2 & 6 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 1 \end{pmatrix}.$$

Using the *Julia* software package `LinearCovarianceModels.jl` [21] we find that the maximum likelihood degree of $\mathcal{M}_{A,B}$ is indeed $2 \cdot 3 - 3 = 3$, meaning that for a generic sample covariance matrix there will be three solutions over the complex number to the score equations. If we take the sample covariance matrix,

$$S = \begin{pmatrix} 1 & 2 & -2 \\ 2 & 6 & -7 \\ -2 & -7 & 9 \end{pmatrix},$$

then the equations f, g from (1) for $\mathcal{M}_{A,B}$ and S are given explicitly by

$$f = 12288x^5 + 57600x^4y + 74272x^3y^2 + 20172x^2y^3 + 1729xy^4 + 37y^5 \\ - 10496x^4 - 33792x^3y - 45484x^2y^2 - 7232xy^3 - 513y^4$$

and

$$g = 11520x^5 + 37136x^4y + 20172x^3y^2 + 3458x^2y^3 + 185xy^4 + 3y^5 \\ - 12624x^4 - 9448x^3y - 6480x^2y^2 - 528xy^3 - 21y^4.$$

We used the numerical polynomial solver package `HomotopyContinuation.jl` [4] to find the solutions to the system of equations $f = 0$, $g = 0$. The solution set consisted of the origin (with multiplicity 16) and three points corresponding to the critical points of the log-likelihood function,

$$\{(0.6897, 0.1773), (0.2655 + 0.3071i, 0.9865 - 2.4601i), (0.2655 - 0.3071i, 0.9865 + 2.4601i)\}.$$

The number of critical points and the multiplicity at the origin are predicted by Theorem 2.2 and Corollary 4.2, respectively. This fits into (2), which for $n = 3$ becomes $5^2 = 16 + 3 + 6$. Thus the maximum likelihood estimate for $\mathcal{M}_{A,B}$ and S is the real point in the list above, which corresponds to the positive-definite covariance matrix

$$\Sigma = \begin{pmatrix} 3.6257 & 0.5124 & 0 \\ 0.5124 & 3.1329 & -1.7340 \\ 0 & -1.7340 & 4.3154 \end{pmatrix}$$

that maximizes the likelihood function $L(\Sigma \mid S)$.

Properties that hold generically. In Example 2.4, it was important to choose the matrices A, B, and S to be "generic enough." We explain the precise notion of genericity in classical algebraic geometry below.

Let X be an algebraic variety and $\mathcal{P}$ a property of the points of X. One says that $\mathcal{P}(x)$ *holds for generic* $x \in X$, or *holds generically* on X, if there exists a nonempty Zariski open set U of X such that $\mathcal{P}(x)$ holds for all $x \in U$.

Consider the case $X = \mathbb{C}^N$. A Zariski open set in $\mathbb{C}^N$ is the complement of a set $V = V(f_1, \ldots, f_k)$ of common zeros of a collection of polynomials $f_1, \ldots, f_k$ in N variables.

Thus, to verify that some property $\mathcal{P}$ holds generically on $\mathbb{C}^N$, we first have to find such a set V with the property that for all x, if $\mathcal{P}(x)$ does not hold, then $x \in V$. This verifies that $\mathcal{P}(x)$ holds for all $x \in U$, where $U = \mathbb{C}^N \setminus V$. We also have to verify that U is nonempty, which amounts to finding a specific element x_0 such that $x_0 \notin V$.

Note that $\dim V$ is at most $N - 1$, which justifies the term "generic". In particular it is expected that a point $x \in \mathbb{C}^N$ taken at random[1] will lie in U.

Suppose that $\mathcal{Q}$ is another property of the points of X and we want to show that both $\mathcal{P}(x)$ and $\mathcal{Q}(x)$ hold generically on X. Then it is enough to show separately that $\mathcal{P}(x)$ holds for generic x and that $\mathcal{Q}(x)$ holds for generic x. This follows from the fact that the intersection of two nonempty Zariski open sets U_1, U_2 is always a nonempty Zariski open set. In practice, this means that after finding U_1 and U_2 it is enough to find two separate elements $x_1 \in U_1$ and $x_2 \in U_2$, which could be easier than finding an element $x_0 \in U_1 \cap U_2$.

[1] Say, according to the multivariate normal distribution.

In this article, the notion of a property holding generically is important for two reasons. First, it is needed for the definition of the ML-degree. Indeed, the number (with multiplicities) of solutions $(\hat{x}, \hat{y})$ to the score equations $\tilde{\ell}_i(x, y)$ given an empirical covariance matrix S could vary with S. Nevertheless, it is constant for generic S, which justifies the use of a single number. Second, we consider a family of models $\mathcal{M}_{A,B}$ parametrized by pairs of symmetric matrices (A, B) and compute its ML-degree only for generic A, B. To perform the computation, we use several properties that hold for generic A, B, and S. We prove this separately for each one of them and use them together at the same time, as explained above.

3. Geometry of the score equations

In this section, we use Bézout's theorem to derive a formula for computing $\deg J$. Lemma 3.2 will be used throughout the paper for all arguments involving generic A, B, and S. We will use Euler's homogeneous function theorem, which says that if $H(x, y)$ is a homogeneous function of degree m, then $mH = xH_x + yH_y$. We will also use the following fact about binary forms.

Proposition 3.1. *Let $H(x, y) \in \mathbb{C}[x, y]$ be a homogeneous polynomial in two variables. Then $H(x, y)$ factors as a product of linear forms.*

Proof. Suppose that $H(x, y)$ is homogeneous of degree d. Let $h(x) := H(x, 1)$ have degree c. Since $\mathbb{C}$ is algebraically closed, by the fundamental theorem of algebra, $h(x) = a\prod_{i=1}^{c}(x - r_i)$ for some $a, r_1, \ldots, r_c \in \mathbb{C}$. Then $H(x, y) = ay^{d-c}\prod_{i=1}^{c}(x - r_i y)$. □

We further note that a generic $h \in \mathbb{C}[x, y]$ factors as a product of *distinct* linear forms. A binary form has a multiple root if and only if its discriminant vanishes, which is a closed condition on the space coefficients [17, §0.12].

Lemma 3.2. *For generic A, B, and S, the following projective varieties are empty*:

(1) $V(P, P_x), V(P, T), V(P, P_y)$,

(2) $V(P_x, P_y)$,

(3) $V(P_x, T_x), V(P_y, T_y), V(T, T_x), V(T, T_y)$.

Proof. The emptiness of the varieties in the statement is an open condition in the space of parameters (A, B, S). For instance, the subset of the parameter space $\mathbb{A}_{(A,B,S)}$ where $V(P, P_x)$ is nonempty is the image of the variety defined by P and P_x in the space $\mathbb{A}_{(A,B,S)} \times \mathbb{P}^1_{[x:y]}$ under the first projection. This is a Zariski-closed subset of the parameter space by the projective elimination theorem [8, Chapter 8.5].

To show that the projective varieties in the statement are empty, we show that the polynomials defining them have no common factors. This makes use of Proposition 3.1, which states that every homogeneous form in two variables factors as a product of linear forms.

First, consider the case where A is the $n \times n$ identity matrix, B is the diagonal matrix with diagonal entries $1, \ldots, n$, and $S = uu^T$ where u is the vector of all ones. We have

$$P = \prod_{k=1}^{n}(x + ky) \quad \text{and} \quad P_x = \sum_{k=1}^{n}\prod_{j\neq k}(x + jy).$$

From this we deduce that if $p = x + ky$ is a linear form that divides P, then it does not divide P_x. This shows that $V(P, P_x)$ is empty. The variety $V(P, T)$ is empty as well since $P_x = T$ in this case. Similarly, one shows that $V(P, P_y)$ is empty.

Euler's homogeneous function theorem applied to P says that $nP = xP_x + yP_y$. Since $V(P, P_x)$ is generically empty, the same holds for $V(P_x, P_y)$.

To prove the rest of the statements, we switch to an element (A, B, S) that makes the form of T particularly simple. Let A and B be as before and $u = (1, 0, \ldots, 0)$. This is allowed when combining generic properties as we explained at the end of Section 2. In this case we have

$$T = \prod_{k \neq 1} (x + ky) \quad \text{and} \quad P_x = T + (x + y)T_x.$$

Assume p divides P_x and T_x. Then p divides T; hence, we may assume $p = x + ky$ with $k \neq 1$. However, we have $p \nmid P_x$ as before. This contradiction shows that $V(P_x, T_x)$ is empty. Similarly, $V(P_y, T_y)$ is empty. This example also has T with no common roots; hence, $V(T, T_x)$ and $V(T, T_y)$ are generically empty. □

Now we will show that the projective curves defined by F and G satisfy the hypothesis of Bézout's theorem, that is, that they do not share a common component. This justifies our application of Bézout's theorem and allows us to count the points in their variety. To prove this, we show that the polynomials f and g in (1) generically are irreducible and do not share a common factor.

Lemma 3.3. *The polynomials f and g in (1) are irreducible for generic A, B, and S.*

Proof. We prove the statement for f. The proof for g is analogous. Write $f = F_{2n-1} + F_{2n-2}$, where

$$F_{2n-1} = PP_x \quad \text{and} \quad F_{2n-2} = PT_x - TP_x$$

and $\deg F_i = i$. If f decomposes into a product of two polynomials, then at least one of them is homogeneous and we call it h. Indeed, otherwise the degrees of F_{2n-1} and F_{2n-2} would be at least two apart, when in fact they differ by one. Since h is homogeneous and divides a nonzero sum of homogeneous polynomials, h divides each of the summands F_{2n-1} and F_{2n-2}. Using Proposition 3.1, let h_0 be a linear factor of h. Since h_0 divides F_{2n-1} and is irreducible, h_0 divides P or P_x. In the first case, since h_0 divides F_{2n-2}, it would have to divide either T or P_x. This would imply that one of the projective varieties $V(P, T)$ and $V(P, P_x)$ is nonempty. By Lemma 3.2 this does not happen generically. In the second case, it would have to divide either P or T_x, which for the same reason does not happen generically. □

Lemma 3.4. *For generic A, B, and S, the polynomials f and g in (1) are not constant multiples of one another.*

Proof. If f and g are constant multiples of each other, then so are their highest-degree terms PP_x and PP_y. This does not happen generically since by Lemma 3.2 the projective variety $V(P_x, P_y)$ is generically empty. □

Furthermore, we can describe exactly which points are removed from the affine variety $V(f, g)$ after we saturate with the determinant. For generic parameters, the only point that is removed after saturation is the origin.

Proposition 3.5. *For generic A, B, and S, we have*

$$V(f, g) \setminus V(\det \Sigma) = V(f, g) \setminus \{(0, 0)\}.$$

Proof. Let $q \in V(P, f, g)$. Then $f(q) = -T(q)P_x(q)$ and $g(q) = -T(q)P_y(q)$. In order to have $f(q) = g(q) = 0$, we must either have both $P_x(q) = P_y(q) = 0$ or $T(q) = 0$. By Lemma 3.2, for generic A, B, and S, both of these imply $q = (0, 0)$. □

Proposition 3.6. *For generic A and B, the* ML*-degree of the model $\mathcal{M}_{A,B}$ is*

$$(2n-1)^2 - I_{[0:0:1]}(F, G) - \sum_{q \in V(F,G,z)} I_q(F, G).$$

Proof. The ML-degree of $\mathcal{M}_{A,B}$ is defined as the degree of the ideal $J = \langle f, g \rangle : (\det \Sigma)^\infty$. The affine variety $V(J)$ embeds in projective space as

$$V(F, G) \setminus (V(F, G, z) \cup V(\det \Sigma)).$$

By Bézout's theorem (Theorem 2.3), Lemmas 3.3 and 3.4 imply that the variety $V(F, G)$ is zero-dimensional. Using Proposition 3.5 we have

$$\begin{aligned} \deg J &= \sum_{\substack{q \in V(F,G) \\ q \notin \{[0:0:1]\} \cup V(F,G,z)}} I_q(F, G) \\ &= \sum_{q \in V(F,G)} I_q(F, G) - I_{[0:0:1]}(F, G) - \sum_{q \in V(F,G,z)} I_q(F, G). \end{aligned}$$

Both F and G have degree $2n-1$. Applying Theorem 2.3 to F and G gives the desired equality. □

4. Multiplicity at the origin

In this section we compute the intersection multiplicity of the polynomials f, g in (1) at the origin, denoted by $I_{[0:0:1]}(F, G)$ and also $I_{(0,0)}(f, g)$.

For a polynomial in two variables h there is a notion of *multiplicity* of h at the origin, denoted $m_{(0,0)}(h)$. This is the degree of the lowest-degree summand in the decomposition of h as a sum of homogeneous polynomials (for details, see [13, §3.1]). Since the polynomials f, g can be written as the sum of a homogeneous degree $2n-2$ form with a homogeneous degree $2n-1$ form, we have $m_{(0,0)}(f) = m_{(0,0)}(g) = 2n-2$. We have the identity

$$I_{(0,0)}(f, g) = m_{(0,0)}(f) \cdot m_{(0,0)}(g) \tag{3}$$

if the lowest-degree homogeneous forms of f and g share no common factors [13, §3.3]. The degree $2n-2$ parts of f and g are $Q = PT_x - TP_x$ and $R = PT_y - TP_y$, respectively.

Proposition 4.1. *For generic A, B, and S, the polynomials Q and R share no common factor.*

Proof. By the definition of Q and R and two applications of Euler's homogeneous function theorem we have

$$\begin{aligned} xQ + yR &= (xT_x + yT_y)P - (xP_x + yP_y)T \\ &= (2n-2)TP - (2n-1)PT \\ &= -PT. \end{aligned}$$

Assume that Q and R share a common factor p, which we may assume is irreducible. Then p divides PT. So p divides P or p divides T, but not both by Lemma 3.2. If p divides P, then since $Q = PT_x - TP_x$ and p is a factor of Q, p also divides TP_x. Similarly if p divides T, then p also divides PT_x. But then either P and TP_x share a common factor, or T and PT_x do. Each of the resulting four further cases does not occur generically by Lemma 3.2. □

Corollary 4.2. *For generic A, B, and S, the intersection multiplicity of f and g at the origin is $(2n-2)^2$.*

Proof. By Proposition 4.1, this follows from (3). □

5. Multiplicity at infinity

In this section we compute the intersection multiplicity at a point at infinity for the curves $V(f)$ and $V(g)$ defined by the polynomials in (1) for generic A, B, and S. To do this we use the connection between intersection multiplicity of curves and their series expansions about an intersection point.

Consider an irreducible polynomial h in two variables such that $h(0,0) = 0$ and $h_y(0,0) \neq 0$. By [12, §7.11, Corollary 2], there exists an infinite series $\alpha = \sum_{m=1}^{\infty} a_m t^m$ and an open neighborhood $U \subset \mathbb{C}$ containing $t = 0$ such that $h(t, \alpha(t)) = 0$ for all $t \in U$. The series α is called the *series expansion* of h at the origin. The *valuation* of a series is the number M such that $a_M \neq 0$ and $a_m = 0$ for all $m < M$.

Proposition 5.1. *Let h and k be irreducible polynomials in two variables such that h and k vanish at $(0,0)$ and h_y and k_y do not. Let α and β be infinite series expansions of h and k at $(0,0)$, respectively. The intersection multiplicity $I_{(0,0)}(h,k)$ is the valuation of the series $\alpha - \beta$.*

Proof. By [12, §8.7], the intersection multiplicity of h and k at $(0,0)$ is the valuation of the infinite series $h(t, \beta(t))$. We prove that this is the same as the valuation of $\alpha - \beta$. First, let $s(t) = \sum_{m=1}^{\infty} s_m t^m$ be any infinite series and write $h = \sum_{i,j} c_{i,j} x^i y^j$, where the sum ranges over the pairs (i, j) with $0 < i + j \leq \deg h$. We have

$$\begin{aligned} h(t, s(t)) = \sum_{i,j} c_{i,j} t^i \left(\sum_{m=1}^{\infty} s_m t^m \right)^j &= \sum_{i,j} c_{i,j} t^i \left(\sum_{\nu=0}^{\infty} \left(\sum_{|a|=\nu} s_{a_1} \cdots s_{a_j} \right) t^\nu \right) \\ &= \sum_{i,j} \sum_{\nu=0}^{\infty} \sum_{|a|=\nu} c_{i,j} s_{a_1} \cdots s_{a_j} t^{\nu+i}. \end{aligned}$$

The coefficient r_m of t^m in this infinite series is a finite sum of products of the form $c_{i,j} s_{a_1} \cdots s_{a_j}$ with $a_j \leq m$ and $|a| + i = m$. The term s_m only appears in r_m when $j = 1$ and $i = 0$. Hence, we have $r_m = c_{0,1} s_m + p(s_1, \ldots, s_{m-1})$ for some polynomial p, where $c_{0,1} \neq 0$ since $h_y(0,0) \neq 0$. For example,

the coefficient r_0 is zero since h, k vanishing at the origin implies that $c_{0,0}$ and s_0 are zero, and the coefficient of the first nonzero term is given by $r_1 = c_{0,1}s_1 + c_{1,0}$.

Write $\alpha(t) = \sum_{m=1}^{\infty} a_m t^m$ and $\beta(t) = \sum_{m=1}^{\infty} b_m t^m$. Suppose that the valuation of the series $\alpha - \beta$ is M. Then $a_M \neq b_M$ and $a_m = b_m$ for all $m < M$. We show that this is equivalent to $h(t, \beta(t)) = \sum_{m=1}^{\infty} r_m t^m$ having valuation M. Suppose that $M = 1$; then $a_1 \neq b_1$. Since $h(t, \alpha(t))$ is identically zero in a neighborhood of $t = 0$, we have $r_m(a_1, \ldots, a_m) = 0$ for all m. In particular $r_1(a_1) = c_{0,1}a_1 + c_{1,0} = 0$. Since $a_1 \neq b_1$ this implies that $r_1(b_1) = c_{0,1}b_1 + c_{1,0} \neq 0$ and $h(t, \beta(t))$ has valuation one. Similarly if $h(t, \beta(t))$ has valuation one, then $r_1(a_1) \neq r_1(b_1)$ implying $a_1 \neq b_1$. Thus $\alpha - \beta$ has valuation one if and only if $h(t, \beta(t))$ has valuation one.

Now suppose $M > 1$. By the form of r_m it now follows from an inductive argument on m that a_m and b_m agree up to $m = M$ and differ at $m = M + 1$ if and only if $r_m(a_1, \ldots, a_m)$ and $r_m(b_1, \ldots, b_m)$ agree up to $m = M$ and differ at $m = M + 1$. Since $r_m(a_1, \ldots, a_m) = 0$ for all m, the latter is equivalent to $h(t, \beta(t))$ having valuation M. □

Remark 5.2. In the context of Proposition 5.1, consider instead polynomials h and k defining the curves $\mathscr{X}$ and $\mathscr{Y}$, respectively, such that $\mathscr{X}$ and $\mathscr{Y}$ meet at a nonsingular point q. Also, let v be a vector such that the directional derivatives h_v and k_v do not vanish at q. Choose an affine-linear transformation $\varphi : \mathbb{C}^2 \to \mathbb{C}^2$ sending $(0, 0)$ to q and $(0, 1)$ to v. Then $I_q(h, k) = I_{(0,0)}(h \circ \varphi, k \circ \varphi)$ and the polynomials $h \circ \varphi$ and $k \circ \varphi$ satisfy the hypotheses of Proposition 5.1. Thus we can compute the intersection multiplicity at any nonsingular intersection point of h, k using Proposition 5.1.

Remark 5.3. When the series $\alpha - \beta$ has valuation M, one says that h and k have *contact order* or *order of tangency* $M - 1$ at q. Therefore the contact order of two curves at an intersection point is always one less than the intersection multiplicity. For more on contact order of algebraic curves see [19, Chapter 5].

Remark 5.4. The fact that the curves $\mathscr{X}$ and $\mathscr{Y}$ have intersection multiplicity one at q if and only if the gradients of h and k at q are linearly independent (see, e.g., [13, §3.3]) arises as a special case of Proposition 5.1 once one computes the first terms of the series α and β.

Returning to the expressions in (1), recall that F and G denote the homogenizations of f and g with respect to the new variable z. The intersection points of $V(f)$ and $V(g)$ at infinity are exactly the variety $V(F, G, z)$.

Lemma 5.5. *For generic A, B, and S, the projective variety $V(F, G, z)$ consists of n points of the form $[q_1 : q_2 : 0]$ such that $P(q_1, q_2) = 0$.*

Proof. Let $q = [q_1 : q_2 : 0]$ be a projective point of $V(F, G)$. We have

$$F = PP_x + z(PT_x - TP_x),$$
$$G = PP_y + z(PT_y - TP_y),$$

and hence $V(F, G, z)$ consists of points q where $[q_1 : q_2] \in V(PP_x, PP_y)$. Clearly if $P(q_1, q_2) = 0$, then $q \in V(F, G, z)$. These are the only such points since, by Lemma 3.2, for generic A, B, and S the variety $V(P_x, P_y)$ is empty. By Proposition 3.1 $P(x, y)$ factors in n linear forms. These forms are distinct,

since a repeated factor would divide both P and P_x, while $V(P, P_x)$ is empty by Lemma 3.2. Thus there are n distinct points in $V(F, G, z)$. □

Lemma 5.6. *For generic A, B, and S, the projective variety $V(P, P_yT_x - P_xT_y)$ is empty.*

Proof. Let $H = P_yT_x - P_xT_y$. By applying Euler's homogeneous function theorem twice in the following chain of equalities, we have

$$nT_xP - yH = T_x(nP - yP_y) + yP_xT_y = P_x(yT_y + xT_x) = (n-1)P_xT.$$

If P and H have an irreducible common factor p, then $p \mid P_xT$. This does not happen generically by Lemma 3.2. □

Lemma 5.7. *For generic A, B, and S, if $q \in V(F, G, z)$, then $I_q(F, G) = 2$.*

Proof. By Lemma 5.5, such points are of the form $q = [q_1 : q_2 : 0]$ where $P(q_1, q_2) = 0$. Fix such a point q and assume for simplicity that $q_1 \neq 0$. This is not a restriction since the conditions $q_1 = 0$ and $P(q) = 0$ imply $\det B = 0$, which is a closed condition on the parameter space. Thus we can assume q is of the form $[1 : q_2 : 0]$.

Since intersection multiplicity at a point is a local quantity, we may dehomogenize with respect to x and consider the intersection multiplicity of the affine curves $V(F(1, y, z))$ and $V(G(1, y, z))$ at q. We can compute the partial derivatives with respect to y and z:

$$\begin{aligned} F_y(x, y, z) &= P_yP_x + PP_{xy} + z\left(\frac{d}{dy}(PT_x - TP_x)\right), & F_z(x, y, z) &= PT_x - TP_x, \\ G_y(x, y, z) &= P_y^2 + PP_{yy} + z\left(\frac{d}{dy}(PT_y - TP_y)\right), & G_z(x, y, z) &= PT_y - TP_y. \end{aligned} \tag{4}$$

Consider the translated polynomials obtained by translating q to $[1 : 0 : 0]$ given by $\widetilde{F} = F(1, y + q_2, z)$ and $\widetilde{G} = G(1, y + q_2, z)$. Then $\widetilde{F}_z(1 : 0 : 0), \widetilde{G}_z(1 : 0 : 0) \neq 0$ if and only if $F_z(q), G_z(q) \neq 0$, and from (4), we have that

$$F_z(q) = (-TP_x)(1, q_2) \quad \text{and} \quad G_z(q) = (-TP_y)(1, q_2).$$

Since $P(1, q_2) = 0$, Lemma 3.2 implies that $F_z(q), G_z(q) \neq 0$. Thus there exist series expansions $\alpha = \sum_{m=1}^{\infty} a_m t^m$ and $\beta = \sum_{m=1}^{\infty} b_m t^m$ such that, for all t in a neighborhood of $t = 0$,

$$\widetilde{F}\left(1, t, \sum_{m=1}^{\infty} a_m t^m\right) = 0 \quad \text{and} \quad \widetilde{G}\left(1, t, \sum_{m=1}^{\infty} b_m t^m\right) = 0,$$

and hence

$$F\left(1, t + q_2, \sum_{m=1}^{\infty} a_m t^m\right) = 0 \quad \text{and} \quad G\left(1, t + q_2, \sum_{m=1}^{\infty} a_m t^m\right) = 0, \tag{5}$$

in this same neighborhood. Since $I_{[1:0:0]}(\widetilde{F}, \widetilde{G}) = I_q(F, G)$, by Proposition 5.1 we can compute the valuation of the series $\alpha - \beta$ to determine $I_q(F, G)$. Differentiating the expressions in (5) with respect to t, then substituting $t = 0$ yields

$$F_y(q) + F_z(q)a_1 = 0 \quad \text{and} \quad G_y(q) + G_z(q)b_1 = 0.$$

Thus $a_1 = -F_y(q)/F_z(q)$ and $b_1 = -G_y(q)/G_z(q)$, and $(F_yG_z - F_zG_y)(q) = 0$ implies that $a_1 - b_1 = 0$. By differentiating (5) twice with respect to t and substituting these values for a_1 and b_1, one can similarly show that

$$\begin{aligned} a_2 &= \left(\frac{-F_{yy}F_z^2 + 2F_{yz}F_yF_z - F_{zz}F_y^2}{2F_z^3}\right)\Bigg|_q, \\ b_2 &= \left(\frac{-G_{yy}G_z^2 + 2G_{yz}G_yG_z - G_{zz}G_y^2}{2G_z^3}\right)\Bigg|_q. \end{aligned} \tag{6}$$

Since we know that $a_1 - b_1 = 0$, the valuation of $\alpha - \beta = \sum_{m=1}^{\infty}(a_m - b_m)t^m$ is *at least* two; we now show that the valuation is *exactly* two for generic A, B, and S. We verified that $a_2 - b_2 \neq 0$ with the help of the computer algebra system *Maple* by the following steps. First, we computed all second-order derivatives of F and G with respect to y and z in terms of partial derivatives of P and T, by differentiating the expressions in (4). Then, we substituted $P = 0$ and $z = 0$ in these expressions, which corresponds to evaluation at q. Thus from (6) we can obtain expressions for a_2 and b_2 evaluated at q in terms of partial derivatives of P and T. Next we cleared denominators in the resulting expression for $a_2 - b_2$, yielding

$$(a_2 - b_2)(q) = (T^4P_x^2P_y^4(P_yT_x - P_xT_y))(1, q_2).$$

Since $P(1, q_2) = 0$, this expression does not generically evaluate to 0 by Lemmas 3.2 and 5.6. □

Corollary 5.8. *For generic A, B, and S, we have $\sum_{q\in V(F,G,z)} I_q(F, G) = 2n$.*

Proof. This follows from Lemmas 5.5 and 5.7. □

Now we can prove our main result that $\deg J = 2n - 3$:

Proof of Theorem 2.2. Combining Proposition 3.6 with Corollaries 4.2 and 5.8 shows that the ML-degree of $\mathcal{M}_{A,B}$ for generic A and B is

$$(2n - 1)^2 - (2n - 2)^2 - 2n = 2n - 3.$$ □

6. Discussion

Sturmfels, Timme, and Zwiernik [21] use numerical algebraic geometry methods implemented in the Julia package `LinearGaussianCovariance.jl` to compute the ML-degrees of linear Gaussian covariance models for several values of n and m, where n is the size of the covariance matrix and m is the dimension of model. We have proven that for $m = 2$ and arbitrary n, the ML-degree is $2n - 3$, which agrees with the computations in Table 1 of [21].

For higher-dimensional linear spaces, where $m > 2$, the score equations consist of the partial derivatives of $\tilde{\ell}$ with respect to the m parameters of the linear space. Again, in this case, these are rational functions of the data and the parameters. For instance when $m = 3$, we can consider the linear span of three $n \times n$

matrices A, B, and C. Then if $\Sigma = xA + yB + zC$, $P = \det \Sigma$, and $T = \operatorname{tr}(S \operatorname{adj} \Sigma)$, the score equations are

$$\tilde{\ell}_x(x, y, z) = \frac{P_x}{P} + \frac{PT_x - TP_x}{P^2},$$
$$\tilde{\ell}_y(x, y, z) = \frac{P_y}{P} + \frac{PT_y - TP_y}{P^2},$$
$$\tilde{\ell}_z(x, y, z) = \frac{P_z}{P} + \frac{PT_z - TP_z}{P^2},$$

and we can similarly define polynomials

$$f(x, y, z) := PP_x + PT_x - TP_x,$$
$$g(x, y, z) := PP_y + PT_y - TP_y,$$
$$h(x, y, z) := PP_z + PT_z - TP_z,$$

such that the ML-degree of the model is the degree of $J = \mathcal{J}(f, g, h) : \langle \det \Sigma \rangle^{\infty}$. 21] conjecture that the ML-degree in this case is $3n^2 - 9n + 7$. To prove this conjecture as we did for $m = 2$, one might turn to a higher-dimensional generalization of Bézout's theorem, which says that the number of solutions to $V(f, g, h)$ counted with multiplicity is the product $\deg(f) \deg(g) \deg(h)$ *provided that* $V(f, g, h)$ *is zero-dimensional* (see for example [9, §3, Chapter 3] or [20, §2.1, Chapter 3]). This zero-dimensionality restriction is necessary for equality; otherwise the product of the degrees in this case simply gives an upper bound for the number of zero-dimensional solutions counted with multiplicity [14, Theorem 12.3].

Indeed the variety $V(f, g, h)$ contains the one-dimensional affine variety $V(P, T)$ as well as a "curve at infinity" corresponding to the vanishing of P. When $m = 2$, the variety $V(P, T) \subset \mathbb{C}^2$ consisted of only the origin and the elements at infinity were points whose multiplicity we were able to ascertain using properties of curves. This illustrates the added difficulties in counting solutions when moving from planar intersection theory to higher-dimensional intersections.

21] also consider the *generic diagonal model*, in which the linear space that comprises the model consists of diagonal matrices. Their computations show that for $m = 2$, the ML-degree of the generic diagonal model for the first several values of n is also $2n - 3$ [21, Table 2]. It follows from the proof of our result that this ML-degree is indeed $2n - 3$ for all n, as the witnesses for the nonemptiness of the open dense sets that we produced in the proof of Lemma 3.2 were all diagonal matrices. For $m > 2$ and $n > 3$, the ML-degree of the generic diagonal model is conjectured in [21] to be strictly less than the corresponding generic linear Gaussian covariance model. This suggests that the study of linear Gaussian covariance models of arbitrary dimension will require us to look beyond diagonal matrices as witnesses to the nontriviality of some open conditions.

Indeed, many of the projective varieties in Lemma 3.2 are nonempty for diagonal matrices when $m > 2$. For example, when $m \geq 3$, the determinant of P for a diagonal Σ has a nonempty singular locus. Let $m = 3$ and let

$$\Sigma = xA + yB + zC$$

where A, B, and C are the diagonal matrices with diagonal entries $(a_1, \ldots, a_n)$, $(b_1, \ldots, b_n)$, and $(c_1, \ldots, c_n)$, respectively. Then we have

$$P = \prod_{i=1}^{n} (a_i x + b_i y + c_i z).$$

The derivatives of P are of the form

$$P_x = \sum_{i=1}^{n} a_i \prod_{\substack{j=1 \\ j \neq i}}^{n} (a_j x + b_j y + c_j z),$$

and similarly for P_y and P_z. So any projective point in the intersection of linear spaces of the form

$$V(a_i x + b_i y + c_i z) \cap V(a_j x + b_j y + c_j z)$$

for $i \neq j$ is a singular point of P. When $m > 2$, these intersections are nonempty, so such singular points exist.

Thus, when Σ is not defined by diagonal matrices, the problem of finding witnesses to the emptiness of the varieties in Lemma 3.2 for arbitrary n is nontrivial, which adds another layer of difficulty for establishing the ML-degree when $m > 3$. Nevertheless we believe that examining the structure of the score equations for $m = 2$ provides a possible blueprint for approaching the problem for $m > 2$, although it will require different tools from intersection theory.

For the purposes of statistical inference, one is most interested in *real* solutions to the score equations, as these are the ones that may have statistical meaning. Furthermore, it would be nice to understand whether there are truly $2n - 3$ distinct (complex) solutions to the score equations, as opposed to some having higher multiplicity. Based on the examples we have computed, we conjecture an affirmative answer. We thus still have the following open questions regarding the $m = 2$ case.

Problem. How many real solutions can the score equations of a generic two-dimensional linear Gaussian covariance model have?

Conjecture. For generic values of A, B, and S, the score equations of $\mathcal{M}_{A,B}$ with sample covariance matrix S have $2n - 3$ *distinct* solutions.

Acknowledgements

The authors would like to thank Carlos Amendola, Irina Kogan, Emre Sertöz, Bernd Sturmfels, Seth Sullivant, Sascha Timme, Caroline Uhler, Cynthia Vinzant, and Piotr Zwiernik for many helpful conversations. We would also like to thank the anonymous reviewers for their detailed comments on the manuscript. All three authors were supported by the Max-Planck-Institute for Mathematics in the Sciences. Jane Coons was supported by the US National Science Foundation (DGE-1746939).

References

[1] T. W. Anderson, "Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices", pp. 1–24 in *Essays in probability and statistics*, edited by R. C. Bose et al., University of North Carolina, Chapel Hill, NC, 1970.

[2] T. W. Anderson, "Asymptotically efficient estimation of covariance matrices with linear structure", *Ann. Statist.* **1** (1973), 135–141.

[3] J. Bien and R. J. Tibshirani, "Sparse estimation of a covariance matrix", *Biometrika* **98**:4 (2011), 807–820.

[4] P. Breiding and S. Timme, "HomotopyContinuation.jl: a package for homotopy continuation in Julia", pp. 458–465 in *Mathematical software: ICMS 2018*, edited by J. H. Davenport et al., Lecture Notes in Computer Science **10931**, Springer, 2018.

[5] A. J. Butte, P. Tamayo, D. Slonim, T. R. Golub, and I. S. Kohane, "Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks", *Proc. Nat. Acad. Sci.* **97**:22 (2000), 12182–12186.

[6] F. Catanese, S. Hoşten, A. Khetan, and B. Sturmfels, "The maximum likelihood degree", *Amer. J. Math.* **128**:3 (2006), 671–697.

[7] S. Chaudhuri, M. Drton, and T. S. Richardson, "Estimation of a covariance matrix with zeros", *Biometrika* **94**:1 (2007), 199–216.

[8] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*, Springer, 1992.

[9] D. Cox, J. Little, and D. O'Shea, *Using algebraic geometry*, Graduate Texts in Mathematics **185**, Springer, 1998.

[10] M. Drton, B. Sturmfels, and S. Sullivant, *Lectures on algebraic statistics*, Oberwolfach Seminars **39**, Birkhäuser, Basel, 2009.

[11] J. Felsenstein, "Maximum-likelihood estimation of evolutionary trees from continuous characters", *Am. J. Hum. Genet.* **25**:5 (1973), 471–492.

[12] G. Fischer, *Plane algebraic curves*, Student Mathematical Library **15**, American Mathematical Society, Providence, RI, 2001.

[13] W. Fulton, *Algebraic curves: an introduction to algebraic geometry*, Addison-Wesley, Redwood City, CA, 1989.

[14] W. Fulton, *Intersection theory*, 2nd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **2**, Springer, 1998.

[15] J. Harris, *Algebraic geometry: a first course*, Graduate Texts in Mathematics **133**, Springer, 1995.

[16] H. Jiang, *Modeling trait evolutionary processes with more than one gene*, Ph.D. thesis, University of New Mexico, 2017, available at https://digitalrepository.unm.edu/math_etds/106/.

[17] V. L. Popov and E. B. Vinberg, "Invariant theory", pp. 123–278 in *Algebraic geometry*, vol. IV, edited by A. N. Parshin and I. R. Shafarevich, Encyclopaedia of Mathematical Sciences **55**, Springer, 1994.

[18] A. J. Rothman, E. Levina, and J. Zhu, "A new approach to Cholesky-based covariance regularization in high dimensions", *Biometrika* **97**:3 (2010), 539–550.

[19] J. W. Rutter, *Geometry of curves*, Chapman & Hall, Boca Raton, FL, 2000.

[20] I. R. Shafarevich, *Basic algebraic geometry*, vol. 1: Varieties in projective space, 3rd ed., Springer, 2013.

[21] B. Sturmfels, S. Timme, and P. Zwiernik, "Estimating linear covariance models with numerical nonlinear algebra", *Alg. Stat.* **11**:1 (2020), 31–52.

[22] S. Sullivant, *Algebraic statistics*, Graduate Studies in Mathematics **194**, American Mathematical Society, Providence, RI, 2018.

[23] W. B. Wu and H. Xiao, "Covariance matrix estimation in time series", Chapter 8, pp. 187–209 in *Time series analysis: methods and applications*, edited by T. Subba Rao et al., Handbook of Statistics **30**, North-Holland, Amsterdam, 2012.

[24] P. Zwiernik, C. Uhler, and D. Richards, "Maximum likelihood estimation for linear Gaussian covariance models", *J. R. Stat. Soc. B. Stat. Methodol.* **79**:4 (2017), 1269–1292.

Received 2019-09-10. Revised 2020-05-20. Accepted 2020-06-08.

JANE IVY COONS: jicoons@ncsu.edu
Department of Mathematics, North Carolina State University, Raleigh, NC, United States

ORLANDO MARIGLIANO: orlando.marigliano@mis.mpg.de
Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany

MICHAEL RUDDY: michael.ruddy@mis.mpg.de
Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany

Algebraic Statistics
Vol. 11, No. 2, 2020
https://doi.org/10.2140/astat.2020.11.125

HOLONOMIC GRADIENT METHOD FOR TWO-WAY CONTINGENCY TABLES

YOSHIHITO TACHIBANA, YOSHIAKI GOTO, TAMIO KOYAMA AND NOBUKI TAKAYAMA

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 [8] 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., [18], [25]) 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

MSC2010: 33C90, 65Q10, 62B05, 62H17.
Keywords: holonomic gradient method, two-way contingency tables, modular method, conditional maximum likelihood estimation.

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 [20] 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 ($r_1 \times r_2$ (two-way) contingency table). An $r_1 \times r_2$ matrix with nonnegative 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 $\beta^c = \left(\sum_i u_{i1}, \cdots, \sum_i u_{ir_2}\right)^T$. 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 $u = \left(\begin{smallmatrix} 5 & 3 & 6 \\ 7 & 2 & 4 \end{smallmatrix}\right)$ the row sum vector and the column sum vector are

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

The corresponding vector expressions of u^f and β are

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

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

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

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

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

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

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

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

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

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 $(r_1, r_1 + r_2)$ [29] (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 [8].

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

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

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:

$$\frac{1}{a}(a + \theta_x) \bullet f(a) = f(a+1), \tag{4}$$

$$\big(\theta_x(c - 1 + \theta_x) - x(a + \theta_x)(b + \theta_x)\big) \bullet f(a) = 0. \tag{5}$$

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

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

Next, we have (3) as

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

where E is the identity matrix.

A relation like $F(a)=M(a)F(a+1)$ is called a *contiguity relation*. In [8], the vector valued function $F(a)$ is called the *Gauss–Manin vector*.

There are several algorithms to obtain contiguity relations [28], [22], [21], [8]. Among them, we choose to use the method of twisted cohomology groups given in [8], because it is the most efficient method for the case of two-way contingency tables.

We briefly summarize the method given in [8]. Consider the hypergeometric series $f(\alpha;x)$ of type (r_1,r_1+r_2). Here, the parameter $\alpha=(\alpha_1,\dots,\alpha_{r_1+r_2-1})$ 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 $\varphi_1,\dots,\varphi_r$ of H such that the "integral" of $(\varphi_1,\dots,\varphi_r)^T$ gives a constant multiple of the Gauss–Manin vector

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

where $\delta^{(i)}$ is some differential operator with respect to $x=(x_{ij})$. There exist a basis $\varphi'_1,\dots,\varphi'_r$ of H' and a linear map $\mathcal{U}_i:H'\to H$ such that the integral of $\left(\mathcal{U}_i(\varphi'_1),\dots,\mathcal{U}_i(\varphi'_r)\right)^T$ 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 $\{\varphi_j\}$:

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

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 [8]. 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 (see the Appendix) for more details):

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

Apart from the diagonal matrix $\mathrm{diag}(1, 1-x)$, the matrices 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. We 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 $F(-1) = \left(1 - \frac{b}{c}x, -\frac{b}{c}x\right)^T$ 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 [8].

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 $p = \left(\begin{smallmatrix} 1 & 1/2 & 1/3 & 1/5 & 1/7 \\ 1 & 1 & 1 & 1 & 1 \end{smallmatrix}\right)$ 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., [6]). 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 the same as $F(\alpha; x)$ in the above (for the correspondence between $(\beta; p)$ and $(\alpha; x)$, see [8, Proposition 7.1]). In fact, we give an improvement of Step 2–4 of [8, Algorithm 7.8].

Algorithm 1 (A modified version of [8, Step 1–4 of Algorithm 7.8]).

Input: $\beta = (\beta_1^{(1)}, \dots, \beta_{r_1}^{(1)}; \beta_1^{(2)}, \dots, \beta_{r_2}^{(2)})$: 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 $B_0 = (1, \dots, 1, \beta_1^{(1)} + \cdots + \beta_{r_1}^{(1)} - r_1 + 1; \beta_1^{(2)}, \dots, \beta_{r_2}^{(2)})$. 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.)

(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 [8, Step 5–7 of Algorithm 7.8]).

Example 6 (cf. [8, 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 [8], 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

$$C_1(0)F(1, 1, 6; 1, 3, 4; p) = F(2, 1, 5; 1, 3, 4; p),$$
$$C_2(0)F(2, 1, 5; 1, 3, 4; p) = F(2, 2, 4; 1, 3, 4; p), \quad C_2(1)F(2, 2, 4; 1, 3, 4; p) = F(2, 3, 3; 1, 3, 4; p).$$

(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.[1]

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

Remark 1. The algorithm given in [8] requires more matrix multiplications than Algorithm 1. As [8, 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):

$$F(1, 1, 2; 2, 1, 1;\, p) \mapsto F(1, 1, 3; 2, 2, 1;\, p) \mapsto F(1, 1, 4; 2, 3, 1;\, p)$$
$$\mapsto F(1, 1, 5; 2, 3, 2;\, p) \mapsto F(1, 1, 6; 2, 3, 3;\, p) \mapsto F(1, 1, 7; 2, 3, 4;\, p)$$
$$\mapsto F(1, 1, 6; 1, 3, 4;\, p) \mapsto F(2, 1, 5; 1, 3, 4;\, p) \mapsto F(2, 2, 4; 1, 3, 4;\, p) \mapsto F(2, 3, 3; 1, 3, 4;\, p).$$

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 will help the reader follow the argument. By [8, 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 [8], 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

[1] This 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)).
```

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) This 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 [8, 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 [8, 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 [8, 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 [24]. The function names in this section are those in this package.

5.1. ***Our benchmark problems.*** We use four benchmark problems to compare the various methods. The timing data are taken on a machine with

CPU	Intel(R) Xeon(R) CPU E5-4650 2.70 GHz
the number of CPU's	32
the number of cores	8
OS	Debian 9.8
memory	256 GB
software system	Risa/Asir (2018) version 20190328 with GMP [9]

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 given by

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

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 given by

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

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 p given by

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

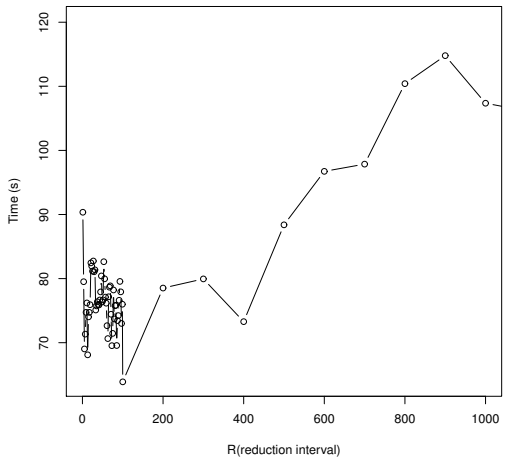
5.2. ***Floating point arithmetic.*** If we can evaluate the composition of linear transformations (7) accurately over floating point numbers, we can utilize GPU's or other hardware for efficient evaluation. Unfortunately, we lose the precision during the iteration of linear transformations in general. 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 $O(n^2)$.[2]

One way to avoid reductions in $\mathbb{Q}$ in our iterations 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

[2]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.

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

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. Figure 1 shows 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). [3]

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".

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

(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.

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

$$\left(\tfrac{2}{3} + o(1)\right)M(d_pN_p)\max\left(\frac{\log N_p}{\log\log(d_pN_p)}, 1 + O(N_p^{-1})\right).$$

(3) The rational reconstruction algorithm `IntegerToRational`, see, e.g., [5], [19], 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 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 why we make

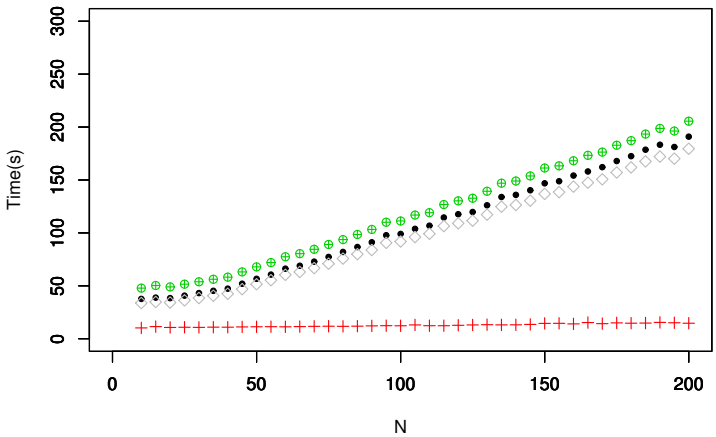


Figure 2. 5×5 contingency table, the Benchmark Problem 3 with 32 processes.

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 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 the 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. Figure 3 shows timing data of our Benchmark Problem 4 of 7×7 contingency tables with the multimodular method by

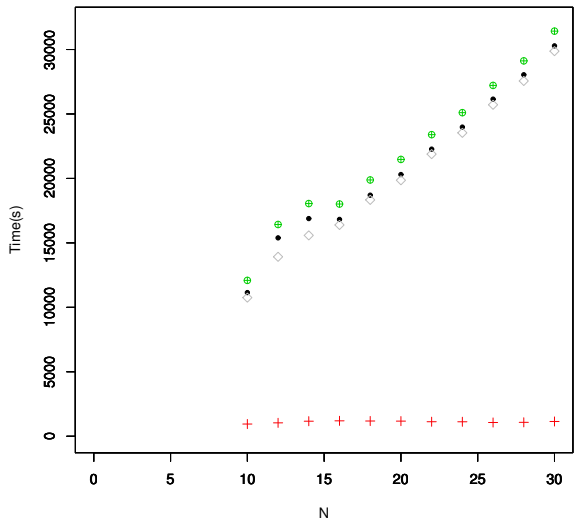


Figure 3. 7×7 contingency table, the Benchmark Problem 4 with 32 processes.

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 when the size of the matrix is not small and the length of multiplication is not very long. However, 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)
```

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

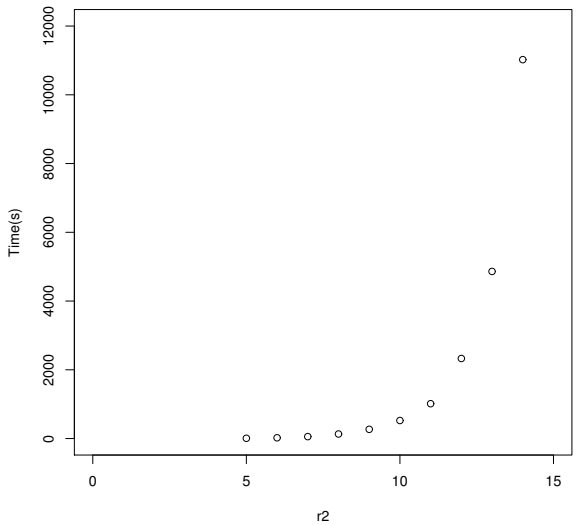


Figure 4. Time to obtain contiguity relations.

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 [8] 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 [27], [23]. Then, we can obtain the exact value by interpolation. □

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

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

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 nonzero 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 zero 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 [7], 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

Often 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 formulation of parameters of interest, nuisance parameters, and sufficient statistics. Theorems 3, 4, and 5 explain what sufficient statistics are for the two-way contingency tables admitting zero cells. In order to prove these theorems, we utilize the notion of sufficient σ-algebra.

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 enables 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. A Bayesian approach to sufficient statistics is presented in, e.g., Chapter 2 of the textbook by M. Schervish [26]. This 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 and, in this section and the next, 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 (1803.04170). 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 salient and of interest versus those which are not, may seem easy. However, 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 [30].

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

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

The measurable space $(\Omega, \mathcal{F})$ is deemed to be the product measurable space of $(\Omega', \mathcal{F}')$ and $(\Theta, \mathcal{B}(\Theta))$ [30, p75]. For $A \in \mathcal{F}'$, let $f_A : \Theta \to \mathbb{R}$ 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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}), N(\mu, \sigma^2))$, where $N(\mu, \sigma^2)$ is the Gaussian distribution on $\mathbb{R}$ with mean μ and variance σ^2. In this case, the parameter space is

$$\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 \mid \sigma^2 > 0\}$$

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 state 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, Y is $\sigma(X)$-measurable if and only if Y equals to $f(X)$ for a Borel measurable function f. See, e.g., [30, p206].

Let X and Y be $\mathbb{R}$-valued random variables and θ be a Θ-valued random variable, which we will call a parameter. We assume that X is integrable. The conditional expectation $\mathbf{E}(X|Y, \theta)$ 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

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

Because the equation $f(y, c_1) = f(y, c_2)$ may hold even if $c_1 \neq c_2$, the conditional expectation $\mathbf{E}(X|Y, \theta)$ is 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 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 $\mathbf{E}(X|\mathcal{G})$ can be measurable for a sub-σ-algebra which is strictly smaller than $\mathcal{G}$.

Definition 2. A 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 $X \in \mathcal{L}^1(\mathcal{D})$, there exists a version of $\mathbf{E}(X|\mathcal{G})$ which 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 2. Note that the condition of Definition 3 is equivalent to stating that the equation

$$\mathbf{E}(X|\sigma(\mathcal{S}, \mathcal{N})) = \mathbf{E}(X|\mathcal{S}) \quad \text{a.s.} \tag{8}$$

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

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

Remark 3. 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 measure theory in Definition 3. In our definition, we use σ-algebra instead of statistics and parameters. Traditional definitions can be reduced to ours by

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

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) $0 \le \theta \le 1$, and

(2) the conditional probability of $X_1, \ldots, X_n$ for given θ is

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

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 $(\mathcal{D}, \mathcal{N})$.

In order to clarify our formulation by the σ-algebra, we will prove that $\mathcal{S}$ is sufficient. For $x = (x_1, \ldots, x_n)^\top \in \mathbf{R}^n$, we denote by $|x|$ the sum of elements of x. Put $X := (X_1, \ldots, X_n)^\top$ and $T := |X| = X_1 + \cdots + X_n$. By [30, p206], for any $Y \in \mathcal{L}^1(\mathcal{D})$, we can take a Borel measurable function $f : \mathbf{R}^d \to \mathbf{R}$ such that $Y = f(X)$. Let $g : \{0, 1, \ldots, n\} \to \mathbf{R}$ be a function defined by

$$g(t) := \binom{n}{t}^{-1} \sum_{x \in \{0,1\}^n} \delta_{t,|x|} f(x).$$

Then, $g(T)$ is $\mathcal{S}$-measurable. For any $B, C \in \mathcal{B}(\mathbf{R})$, we have (with I_B and I_C the indicator functions of B and C)

$$\begin{aligned}
\mathbf{E}(Y; T \in B, \theta \in C) &= \mathbf{E}(Y I_B(T) I_C(\theta)) = \mathbf{E}(f(X) I_B(|X|) I_C(\theta)) \\
&= \int \sum_{x \in \{0,1\}^n} f(x) I_B(|x|) I_C(\theta) \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} p(\theta)\, d\theta \\
&= \int \sum_{x \in \{0,1\}^n} \sum_{t=0}^{n} \delta_{t,|x|} f(x) I_B(|x|) I_C(\theta) \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} p(\theta)\, d\theta \\
&= \int \sum_{x \in \{0,1\}^n} \sum_{t=0}^{n} \delta_{t,|x|} f(x) I_B(t) I_C(\theta) \theta^t (1-\theta)^{n-t} p(\theta)\, d\theta
\end{aligned}$$

$$
\begin{aligned}
&= \int \sum_{t=0}^{n} \binom{n}{t}^{-1} \sum_{x\in\{0,1\}^n} \delta_{t,|x|} f(x) I_B(t) I_C(\theta) \binom{n}{t} \theta^t (1-\theta)^{n-t} p(\theta)\, d\theta \\
&= \mathbf{E}\left(g(T) I_B(t) I_C(\theta)\right) \\
&= \mathbf{E}\left(g(T); T \in B, \theta \in C\right).
\end{aligned}
$$

Since $\sigma(\mathcal{S}, \mathcal{N})$ is generated by $\{T \in B\} \cap \{\theta \in C\}$ $(B, C \in \mathcal{B}(\mathbf{R}))$, by [30, 1.6. Lemma (a)], we have $\mathbf{E}(Y; A) = \mathbf{E}(g(T); A)$ for any $A \in \sigma(\mathcal{S}, \mathcal{N})$. Consequently, $g(T)$ is a version of $\mathbf{E}(Y|\sigma(\mathcal{S}, \mathcal{N}))$ and $\mathcal{S}$-measurable. Hence, $\mathcal{S}$ 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

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

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

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

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

An application of conditional distributions in statistics is the elimination of nuisance parameters. By Definition 3 and Remark 3, 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 the $\mathcal{A}$-distribution moves on the set $\Theta := \mathbb{R}^n_{\geq 0}$. Consider the action of the multiplicative group $G := \mathbb{R}^d_{>0}$ on the space Θ defined as

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

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

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

Theorem 3. *The sub-σ-algebra*

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

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

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

For a vector $v = (v_1, \ldots, v_n)^\top \in \mathbb{R}^n$, 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 $J(v)^c := \{j \in \mathbb{N} \mid j \notin J(v)\}$.

For $\alpha = (\alpha_1, \ldots, \alpha_n)^\top \in \mathbb{R}^n$, let R_α be the function from $\Theta = \mathbb{R}^n_{\geq 0}$ to $\mathbb{R}$ 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 $Z : \Theta \to \mathbb{R}^n$ be the function defined by $Z(c) := (Z_1(c), \ldots, Z_n(c))^\top$ $(c \in \Theta)$ where

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

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

$$\mathcal{O} = \sigma\big(AX,\ R_\alpha(\theta),\ Z(\theta); \alpha \in \ker A\big). \tag{9}$$

This theorem implies that the 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 $\Theta := \mathbb{R}^{r_1 \times r_2}_{\geq 0}$. 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

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

For all $X \in \mathcal{L}^1(\mathcal{D})$, the conditional expectation $\mathbf{E}(X|\mathcal{G})$ is invariant under the action of the multiplicative group $G := \mathbb{R}_{>0}^{r_1+r_2}$ 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 $1 \le i, k \le r_1$ and $1 \le j, \ell \le r_2$, let $R_{ijk\ell} : \Theta \to \mathbb{R}$ 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 $1 \le i \le r_1$ and $1 \le j \le r_2$, we define a function $Z_{ij} : \Theta \to \mathbb{R}$ 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} $(1 \le i \le r_1,\ 1 \le j \le r_2)$ 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.

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

Theorem 5. $\sigma(AX, \theta) = \sigma(A\theta, \mathcal{O})$.

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

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

9. Examples of CMLE problems

In the first part of this paper, we propose some efficient methods to evaluate the normalizing constant of the conditional distribution of fixed row and column sums for solving CMLE problems. In the second part, we clarify a statistical meaning of considering the conditional distribution. When the independence of rows and columns (the null model) is rejected under a test, it will be natural to estimate parameters of interest under the alternative hypothesis based on CMLE we have discussed. More precisely, 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[4] 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

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

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 [29, §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 second column and the third row are 1, we have $g_3 h_1 = g_3 h_2 = g_3 h_3 = 1$ and $g_1 h_2 = g_2 h_2 = g_3 h_2 = 1$. Then, we have $g_i h_j = 1$ 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

$$(p_{ij}) \mapsto \begin{pmatrix} \frac{p_{11}p_{32}}{p_{12}p_{31}} & 1 & \frac{p_{13}p_{32}}{p_{12}p_{33}} \\ \frac{p_{21}p_{32}}{p_{22}p_{31}} & 1 & \frac{p_{23}p_{32}}{p_{22}p_{33}} \\ 1 & 1 & 1 \end{pmatrix} \tag{10}$$

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 (10) 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 [29, 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.[5]
Categorical data:

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

[5]Pharmaceuticals and Medical Devices Agency, Japan, 2000, https://www.pmda.go.jp/files/000148557.pdf

We omit titles and express this table as $\left(\begin{smallmatrix}4 & 7 & 2\\ 32 & 5 & 6\end{smallmatrix}\right)$. By applying our algorithm and the method in [29], 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:[6]

	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:

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

Appendix

We will explain the derivation of the matrix U_2 of Example 5 with twisted cohomology groups by following [8] 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 replace the parameters a, b, c by

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

where $\alpha_0 = -\alpha_1 - \alpha_2 - \alpha_3$ stands for the exponent at infinity. 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

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

[6] http://idsc.nih.go.jp/disease/influenza/iencepha.html

in [8, 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 [8, 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 [8, 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 $\tilde{x} = \left(\begin{smallmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & x & 1 \end{smallmatrix}\right)$, where the numbering starts with 0 (see [8] 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.

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

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 [8, 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

$D_2(x) =$ `repMatrix(2,1,1)`, $Q_2(\alpha)/(2\pi\sqrt{-1}) =$ `intMatrix([0,2],[0,3],1,1)`,
$P_2(\alpha)/(2\pi\sqrt{-1}) =$ `intMatrix([2,0],[0,3],1,1)`, $(2\pi\sqrt{-1})P_2(\alpha)^{-1} =$ `invintMatrix_k([2,0],[0,3],1,1)`,
$C(\alpha)/(2\pi\sqrt{-1}) =$ `intMatrix([0,3],[0,3],1,1)`, $(2\pi\sqrt{-1})C(\alpha)^{-1} =$ `invintMatrix_k([0,3],[0,3],1,1)`.

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, JST CREST Grant Number JP19209317 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 ed., Wiley-Interscience, Hoboken, NJ, 2013.

[2] S. Amari, *Information geometry and its applications*, Applied Mathematical Sciences **194**, Springer, 2016.

[3] P. Billingsley, *Probability and measure*, 3rd ed., Wiley, 1995.

[4] R. P. Brent and P. Zimmermann, *Modern computer arithmetic*, Cambridge Monographs on Applied and Computational Mathematics **18**, Cambridge University Press, 2011.

[5] J. von zur Gathen and J. Gerhard, *Modern computer algebra*, 2nd ed., Cambridge University Press, Cambridge, 2003.

[6] Y. Goto, "Contiguity relations of Lauricella's F_D revisited", *Tohoku Math. J.* (2) **69**:2 (2017), 287–304.

[7] Y. Goto, "Intersection numbers of twisted cycles and cocycles for degenerate arrangements", 2018. arXiv 1805.01714

[8] Y. Goto and K. Matsumoto, "Pfaffian equations and contiguity relations of the hypergeometric function of type $(k+1, k+n+2)$ and their applications", *Funkcial. Ekvac.* **61**:3 (2018), 315–347.

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

[10] T. Hibi et al., *Gröbner Bases: statistics and software systems*, Springer, 2013.

[11] J. van der Hoeven, "Faster chinese remaindering", 2016. hal-01403810.

[12] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, Graduate Texts in Mathematics **113**, Springer, 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", *Ann. Math. Statist.* **43** (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", *Math. Comp.* **77**:261 (2008), 589–607.

[17] H. Nakayama, K. Nishiyama, M. Noro, K. Ohara, T. Sei, N. Takayama, and A. Takemura, "Holonomic gradient descent and its application to the Fisher–Bingham integral", *Adv. in Appl. Math.* **47**:3 (2011), 639–658.

[18] M. Noro and K. Yokoyama, "A modular method to compute the rational univariate representation of zero-dimensional ideals", *J. Symbolic Comput.* **28**:1-2 (1999), 243–263.

[19] M. Noro and K. Yokoyama, *Computation of Gröbner bases: introduction to computational algebra*, University of Tokyo Press, 2003. In Japanese.

[20] M. Ogawa, *Algebraic statistical methods for conditional inference of discrete statistical models*, PhD thesis, University of Tokyo, 2015.

[21] K. Ohara and N. Takayama, "Pfaffian systems of A-hypergeometric systems, II: Holonomic gradient method", 2015. arXiv 1505.02947

[22] T. Oshima, *Fractional calculus of Weyl algebra and Fuchsian differential equations*, MSJ Memoirs **28**, Mathematical Society of Japan, Tokyo, 2012.

[23] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes: the art of scientific computing*, 3rd ed., Cambridge University Press, 2007.

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

[25] T. Sasaki and T. Takeshima, "A modular method for Gröbner-basis construction over **Q** and solving system of algebraic equations", *J. Inform. Process.* **12**:4 (1989), 371–379.

[26] M. J. Schervish, *Theory of statistics*, Springer, 1995.

[27] J. Stoer and R. Bulirsch, *Introduction to numerical analysis*, Springer, 1980.

[28] N. Takayama, "Gröbner basis and the problem of contiguous relations", *Japan J. Appl. Math.* **6**:1 (1989), 147–160.

[29] N. Takayama, S. Kuriki, and A. Takemura, "A-hypergeometric distributions and Newton polytopes", *Adv. in Appl. Math.* **99** (2018), 109–133.

[30] D. Williams, *Probability with martingales*, Cambridge University Press, 1991.

Received 2018-06-14. Revised 2020-01-05. Accepted 2020-03-24.

YOSHIHITO TACHIBANA: tatibana@math.kobe-u.ac.jp
Kobe University, Kobe 657-8501, Japan

YOSHIAKI GOTO: goto@res.otaru-uc.ac.jp
Otaru University of Commerce, Otaru 047-8501, Japan

TAMIO KOYAMA: koyama@wakhok.ac.jp
Wakkanai Hokusei Gakuen University, Wakkanai 097-0013, Japan

NOBUKI TAKAYAMA: takayama@math.kobe-u.ac.jp
Kobe University, Kobe 657-8501, Japan

Algebraic Statistics
Vol. 11, No. 2, 2020
https://doi.org/10.2140/astat.2020.11.155

TROPICAL GAUSSIANS: A BRIEF SURVEY

NGOC MAI TRAN

We review the existing analogues of the Gaussian measure in the tropical semiring and outline various research directions.

1. Introduction

Tropical mathematics has found many applications in both pure and applied areas, as documented by a growing number of monographs on its interactions with various other areas of mathematics: algebraic geometry [Baker and Payne 2016; Gross 2011; Huh 2018; Maclagan and Sturmfels 2015], discrete event systems [Baccelli et al. 1992; Butkovič 2010], large deviations and calculus of variations [Kolokoltsov and Maslov 1997; Puhalskii 2001], and combinatorial optimization [Joswig ≥ 2020]. At the same time, new applications are emerging in phylogenetics [Monod et al. 2018; Yoshida et al. 2019; Page et al. 2020], statistics [Hook 2017], economics [Baldwin and Klemperer 2019; Crowell and Tran 2016; Elsner and van den Driessche 2004; Gursoy et al. 2013; Joswig 2017; Shiozawa 2015; Tran 2013; Tran and Yu 2019], game theory, and complexity theory [Allamigeon et al. 2018; Akian et al. 2012]. There is a growing need for a systematic study of probability distributions in tropical settings. Over the classical algebra, the Gaussian measure is arguably the most important distribution to both theoretical probability and applied statistics. In this work, we review the existing analogues of the Gaussian measure in the tropical semiring. We focus on the three main characterizations of the classical Gaussians central to statistics: invariance under orthonormal transformations, independence and orthogonality, and stability. We show that some notions do not yield satisfactory generalizations, others yield the classical geometric or exponential distributions, while yet others yield completely different distributions. There is no single notion of a 'tropical Gaussian measure' that would satisfy multiple tropical analogues of the different characterizations of the classical Gaussians. This is somewhat expected, for the interaction between geometry and algebra over the tropical semiring is rather different from that over $\mathbb{R}$. Different branches of tropical mathematics lead to different notions of a tropical Gaussian, and it is a worthy goal to fully explore all the options. We conclude with various research directions.

The author would like to thank Bernd Sturmfels for raising the question that inspired this paper. The author would also like to thank Yue Ren and Martin Ulirsch, the organizers of the Tropical Panorama conference at the Max-Planck Institute for Mathematics in the Sciences, for the opportunity to speak about this work while it was still in progress.
Keywords: Gaussian, normal distribution, tropical semiring, p-adic, idempotent probability.

2. Three characterizations of the classical Gaussian

The Gaussian measure $\mathcal{N}(\mu, \Sigma)$, also called the normal distribution with mean $\mu \in \mathbb{R}^n$ and covariance $\Sigma \in \mathbb{R}^{n\times n}$ is the probability distribution with density

$$f_{\Sigma,\mu}(x) \propto \exp(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)), \quad x \in \mathbb{R}^n.$$

Let $\mathbf{I}$ denote the identity matrix, and $\mathbf{0} \in \mathbb{R}^n$ the zero vector. Measures $\mathcal{N}(\mathbf{0}, \Sigma)$ are called centered Gaussians, while $\mathcal{N}(\mathbf{0}, \mathbf{I})$ is the standard Gaussian. Any Gaussian can be standardized by an affine linear transformation.

Lemma 2.1. *Let $\Sigma = U\Lambda U^\top$ be the eigendecomposition of Σ. Then $X \sim \mathcal{N}(\mu, \Sigma)$ if and only if $(U\Lambda^{1/2})^{-1}(X-\mu) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.*

The standard Gaussian has two important properties. First, if X is a standard Gaussian in $\mathbb{R}^n$, then its coordinates $X_1, \ldots, X_n$ are n independent and identically distributed (i.i.d) random variables. Second, for any orthonormal matrix A, $AX \overset{d}{=} X$. These two properties completely characterize the standard Gaussian [Kallenberg 2002, Proposition 11.2]. This result was first formalized in dimension three by Maxwell [1860] when he studied the distribution of gas particles, though the essence of his argument was made by Herschel [1850] ten years earlier, as pointed out in [Bryc 1995, p10].

Theorem 2.2 (Maxwell). *Let $X_1, \ldots, X_n$ be i.i.d univariate random variables, where $n \geq 2$. Then the distribution of $X = (X_1, \ldots, X_n)$ is spherically symmetric iff the X_i's are centered Gaussians on $\mathbb{R}$.*

From a statistical perspective, Lemma 2.1 and Theorem 2.2 reduces working with data from the Gaussian measure to doing linear algebra. In particular, if data points come from a Gaussian measure, then they are the affine linear transformation of data points from a standard Gaussian, whose coordinates are always independent regardless of the orthonormal basis that it is represented in. These properties are fundamental to principal component analysis, an important statistical technique whose tropical analogue is actively being studied [Yoshida et al. 2019].

There are numerous other characterizations of the Gaussian measure whose ingredients are only orthogonality and independence, see [Bogachev 1998, §1.9] and references therein. One famous example is Kac's theorem [1939]. It is a special case of the Darmois–Skitovich theorem [Darmois 1953; Skitovich 1953], which characterizes Gaussians (not necessarily centered) in terms of independence of linear combinations. A multivariate version of this theorem is also known; see [Kagan et al. 1972].

Theorem 2.3 (Darmois–Skitovich). *Let $X_1, \ldots, X_n$ be independent univariate random variables. Then the X_i's are Gaussians if and only if there exist $\alpha, \beta \in \mathbb{R}^n$, $\alpha_i, \beta_i \neq 0$ for all $i = 1, \ldots, n$, such that $\sum_i \alpha_i X_i$ and $\sum_i \beta_i X_i$ are independent.*

Another reason for the wide applicability of Gaussians in statistics is the Central Limit Theorem. An interesting historical account of its development can be found in [Kallenberg 2002, §4]. From the Central Limit Theorem, one can derive yet other characterizations of the Gaussian, such as the distribution which maximizes entropy subject to a fixed variance [Barron 1986]. The appearance of the Gaussian in the Central Limit Theorem is fundamentally linked to its characterization as the unique 2-stable distribution.

This is expressed in the following theorem by Pólya [1923]. There are a number of variants of this theorem; see [Bogachev 1998; Bryc 1995] and discussions therein.

Theorem 2.4 (Pólya). *Suppose $X, Y \in \mathbb{R}^n$ are independent random variables. Then X, Y and $(X+Y)/\sqrt{2}$ have the same distribution iff this distribution is the centered Gaussian.*

3. Tropical analogues of Gaussians

3.1. ***Tropicalizations of p-adic Gaussians.*** Evans [2001] used Kac's Theorem as the definition of Gaussians to extend them to local fields. Local fields are finite algebraic extensions of either the field of p-adic numbers or the field of formal Laurent series with coefficients drawn from the finite field with p elements [Evans 2001]. In particular, local fields come with a tropical valuation val, and thus one can define a tropical Gaussian to be the tropicalization of the Gaussian measure on a local field. A direct translation of [Evans 2001, Theorem 4.2] shows that the tropicalization of the one-dimensional p-adic Gaussian is the classical geometric distribution.

Proposition 3.1 (tropicalization of the p-adic Gaussian). *For a prime $p \in \mathbb{N}$, let X be a $\mathbb{Q}_p$-valued Gaussian with index $k \in \mathbb{Z}$. Then* $\operatorname{val}(X)$ *is a random variable supported on $\{k, k+1, k+2, \dots\}$, and it is distributed as $k +$* geometric$(1-p^{-1})$. *That is,*

$$\mathbb{P}(\operatorname{val}(X) = k+s) = p^{-s}(1-p^{-1}) \text{ for } s = 0, 1, 2, \dots.$$

Proof. Recall that a nonzero rational number $r \in \mathbb{Q}\backslash\{0\}$ can be uniquely written as $r = p^s(a/b)$ where a and b are not divisible by p, in which case the valuation of r is $|r| := p^{-s}$. The completion of $\mathbb{Q}$ under the metric $(x, y) \mapsto |x-y|$ is the field of p-adic numbers, denoted $\mathbb{Q}_p$. The tropical valuation of r is $\operatorname{val}(r) := s$. By [Evans 2001, Theorem 4.2], the family of $\mathbb{Q}_p$-valued Gaussians is indexed by $\mathbb{Z}$. For each $k \in \mathbb{Z}$, there is a unique $\mathbb{Q}_p$-valued Gaussian supported on the ball $p^k\mathbb{Z}_p := \{x \in \mathbb{Q}_p : |x| \le p^{-k}\}$. Furthermore, the Gaussian is the normalized Haar measure on this support. As $p^k\mathbb{Z}_p$ is made up of p translated copies of $p^{k+1}\mathbb{Z}_p$, which in turn is made up of p translated copies of $p^{k+2}\mathbb{Z}_p$, a direct computation yields the density of $\operatorname{val}(X)$. □

There is a large and growing literature surrounding probability on local fields, or more generally, analysis on ultrametric spaces. They have found diverse applications, from spin glasses, protein dynamics, and genetics, to cryptography and geology; see the recent comprehensive review [Dragovich et al. 2017] and references therein. The p-adic Gaussian was originally defined as a step towards building Brownian motions on $\mathbb{Q}_p$ [Evans 2001]. It would be interesting to use tools from tropical algebraic geometry to revisit and expand results involving random p-adic polynomials, such as the expected number of zeroes in a random p-adic polynomial system [Evans 2006], or properties of determinants of matrices with i.i.d p-adic Gaussians [Evans 2002]. Previous work on random p-adic polynomials from a tropical perspective tends to consider systems with uniform valuations [Avendaño and Ibrahim 2011]. Proposition 3.1 hints that to connect the two literatures, the geometric distribution may be more suitable.

3.2. ***Gaussians via tropical linear algebra.*** Consider arithmetic done in the tropical algebra $(\overline{\mathbb{R}}, \oplus, \odot)$, where $\overline{\mathbb{R}}$ is $\mathbb{R}$ together with the additive identity. In the max-plus algebra $(\overline{\mathbb{R}}, \overline{\oplus}, \odot)$ where $a \,\overline{\oplus}\, b = \max(a, b)$, for instance, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$. In the min-plus algebra $(\overline{\mathbb{R}}, \underline{\oplus}, \odot)$ where $a \,\underline{\oplus}\, b = \min(a, b)$, we have $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. To avoid unnecessary technical details, in this section we focus on vectors taking values in $\mathbb{R}$ instead of $\overline{\mathbb{R}}$.

Tropical linear algebra was developed by several communities with different motivations. It evolved as a linearization tool for certain problems in discrete event systems, queueing theory and combinatorial optimization; see the monographs [Baccelli et al. 1992; Butkovič 2010], as well as the recent survey [Komenda et al. 2018] and references therein. A large body of work focuses on using the tropical setting to find analogous versions of classical results in linear algebra and convex geometry. Many fundamental concepts have rich tropical analogues, including the spectral theory of matrices [Akian et al. 2006; Baccelli et al. 1992; Butkovič 2010], linear independence and projectors [Allamigeon et al. 2011; Akian et al. 2011; Butkovič et al. 2007; Sergeev 2009], separation and duality theorems in convex analysis [Briec and Horvath 2008; Cohen et al. 2004; Gaubert and Katz 2011; Nitica and Singer 2007], matrix identities [Gaubert 1996; Hollings and Kambites 2012; Morrison and Tran 2016; Simon 1994], matrix rank [Chan et al. 2011; Develin et al. 2005; Izhakian and Rowen 2009; Shitov 2011], and tensors [Butkovic and Fiedler 2018; Tsukerman 2015]. Another research direction focuses on the combinatorics of objects arising in tropical convex geometry, such as polyhedra and hyperplane arrangements [Akian et al. 2012; Develin and Sturmfels 2004; Joswig and Loho 2016; Joswig et al. 2007; Sturmfels and Tran 2013; Tran 2017]. These works have close connections to matroid theory and are at the interface of tropical linear algebra and tropical algebraic geometry [Ardila and Develin 2009; Fink and Rincón 2015; Giansiracusa and Giansiracusa 2018; Hampe 2015; Loho and Smith 2020].

Despite the rich theory of tropical linear algebra, in this section we shall show that there is currently no satisfactory way to define the tropical Gaussian as a classical probability measure based on the characterizations of Gaussians via orthogonality and independence as in Section 2. This is somewhat surprising, for there are good analogues of norms and orthogonal decomposition in the tropical algebra. In hindsight, the main difficulty stems from the fact that such tropical analogues are compatible with tropical arithmetic, while classical measure theory was developed with the usual algebra. In Section 3.3 we consider the idempotent probability measure theory, where there is a well-defined Gaussian measure complete with a quadratic density function analogous to the classical case.

The natural definition for tropical linear combinations of $v_1, \ldots, v_m \in \mathbb{R}^n$ is the set of vectors of the form

$$[v_1, \ldots, v_m] := \{a_1 \odot v_1 \oplus \cdots \oplus a_m \odot v_m \text{ for } a_1, \ldots, a_m \in \mathbb{R}\}, \tag{1}$$

where scalar-vector multiplication is defined pointwise. That is, for $a \in \mathbb{R}$ and $v \in \mathbb{R}^n$, $a \odot v \in \mathbb{R}^n$ is the vector with entries

$$(a \odot v)_i = a + v_i \text{ for } i = 1, \ldots, n.$$

We shall also write $a + v$ for $a \odot v$, with the convention scalar-vector addition is defined pointwise.

For finite m, $V := [v_1, \ldots, v_m]$ is always a compact set in $\mathbb{TP}^{n-1} := \mathbb{R}^n/\mathbb{R}\mathbf{1}$ [Develin and Sturmfels 2004]. Unfortunately, this means one cannot hope to have finitely many vectors to 'tropically span' $\mathbb{R}^m$. Nonetheless, there is a well-defined analogue orthogonal projection in the tropical algebra. Associated to a tropical polytope $V := [v_1, \ldots, v_m]$ defined by (1) is the canonical projector $P_V : \mathbb{R}^n \to V$ that plays the role of the orthogonal projection onto V [Cohen et al. 2004]. This projection is compatible with the projective Hilbert metric d_H [Cohen et al. 2001; 2004], in the sense that $P_V(x)$ is a best-approximation under the projective Hilbert metric of x by points in V [Cohen et al. 2004; Akian et al. 2011]. When V is a polytrope, that is, a tropical polytope that is also classically convex, then P_V can be written as a tropical matrix-vector multiplication. This is analogous to classical linear algebra, where best-approximations in the Euclidean distance can be written as a matrix-vector multiplication.

In the max-plus algebra, the projective Hilbert metric is defined by

$$d_H(x, y) = \max_{i,j\in[n]} (x_i - y_i + y_j - x_j).$$

It induces the Hilbert projective norm $\|\cdot\|_H : \mathbb{R}^m \to \mathbb{R}$ via $\|x\|_H = d_H(x, 0)$. Since $d_H(x, y) = \max_i(x_i - y_i) - \min_j(x_j - y_j)$, one finds that

$$\|x\|_H = \|x - \min_i x_i\|_\infty.$$

This formulation shows that the projective Hilbert norm plays the role of the ℓ_∞-norm on $\mathbb{TP}^{n-1}$. The appearance of ℓ_∞, instead of ℓ_2, agrees with the conventional 'wisdom' that generally in the tropical algebra, ℓ_2 is replaced by ℓ_∞ [Evans 2001].

To generalize Maxwell's characterization of the classical Gaussians, we need a concept of orthogonality. One could attempt to mimic orthogonality via the orthogonal decomposition theorem, as done in [Evans 2001] for the case of local fields discussed in Section 3.1. Namely, over a normed space $(\mathcal{Y}, \|\cdot\|)$ over some field K, say that $y_1, \ldots, y_m \in \mathcal{Y}$ are orthogonal if and only if for all $\alpha_i \in K$, the norm of the vector $\sum_i \alpha_i y_i$ equals the norm of the vector $(|\alpha_1|\|y_1\|, \ldots, |\alpha_m|\|y_m\|)$, that is,

$$\left\|\sum_i \alpha_i y_i\right\| = \big\|(|\alpha_1|\|y_1\|, \ldots, |\alpha_m|\|y_m\|)\big\|. \tag{2}$$

In the Euclidean case, this is the Pythagorean identity

$$\left\|\sum_i \alpha_i y_i\right\|_2 = \left(\sum_i |\alpha_i|^2\|y_i\|^2\right)^{1/2},$$

for example. The ℓ_∞-norm, unfortunately, does not work well with the usual notion of independence in probability. In the Hilbert projective norm, (2) can be interpreted either as

$$\|\max_i(\alpha_i + y_i)\|_H = \max_i \|\alpha_i + y_i\|_H - \min_i \|\alpha_i + y_i\|_H = \max_i \|y_i\|_H - \min_i \|y_i\|_H \tag{3}$$

or

$$\|\max_i(\alpha_i + y_i)\|_H = \max_i(\alpha_i + \|y_i\|_H) - \min_i(\alpha_i + \|y_i\|_H). \tag{4}$$

Unfortunately, neither formulation give a satisfactory notion of orthogonality. In (3), as the norm is projective, the coefficients α_i have disappeared from the RHS. This does not support the notion that over an orthogonal set of vectors in the classical sense, computing the norm of linear combinations is the same as computing norm of the vector of coefficients. In (4), for sufficiently large α_1, the RHS increases without bound whereas the LHS is bounded, and thus equality cannot hold for all $\alpha_i \in \mathbb{R}$ over *any* generating set of y_i's.

The Darmois–Skitovich characterization for Gaussians also does not generalize well. Note that the additive identity in $(\overline{\mathbb{R}}, \oplus, \odot)$ is either $-\infty$ or $+\infty$, so the condition that $\alpha_i, \beta_i \neq 0$ becomes redundant. The following lemma states that the any compact distribution will satisfy the Darmois–Skitovich condition.

Lemma 3.2. *Let $X_1, \ldots, X_n$ be independent random variables on $\mathbb{R}^n$. Then there exist $\alpha, \beta \in \mathbb{R}^n$ such that $\bigoplus_{i=1}^n \alpha_i \odot X_i$ and $\bigoplus_{i=1}^n \beta_i \odot X_i$ are independent if and only if $X_1, \ldots, X_n$ have compact support.*

Proof. Let us sketch the proof for $n = 2$ under the min-plus algebra. Let $X = (X_1, X_2) \in \mathbb{R}^2$ and $Y = (Y_1, Y_2) \in \mathbb{R}^2$ be two independent variables. Define $\bar{F}_X, \bar{F}_Y : \mathbb{R}^2 \to [0, 1]$ via $\bar{F}_X(t) = \mathbb{P}(X \geq t)$ and $\bar{F}_Y(t) = \mathbb{P}(Y \geq t)$. Fix $\alpha, \beta \in \mathbb{R}^2$. For $t \in \mathbb{R}^2$,

$$\begin{aligned}
\mathbb{P}(\alpha_1 \odot X \oplus \alpha_2 \odot Y \geq t) &= \mathbb{P}(\min(\alpha_1 + X, \alpha_2 + Y) \geq t) && \text{by definition} \\
&= \mathbb{P}(X \geq t - \alpha_1)\mathbb{P}(Y \geq t - \alpha_2) && \text{by independence} \\
&= \bar{F}_X(t - \alpha_1)\bar{F}_Y(t - \alpha_2).
\end{aligned}$$

Meanwhile,

$$\begin{aligned}
&\mathbb{P}(\alpha_1 \odot X \oplus \alpha_2 \odot Y \geq t, \beta_1 \odot X \oplus \beta_2 \odot Y \geq t) \\
&\qquad = \mathbb{P}(\min(\alpha_1 + X, \alpha_2 + Y) \geq t, \min(\beta_1 + X, \beta_2 + Y) \geq t) && \text{by definition} \\
&\qquad = \mathbb{P}(X \geq t - \alpha_1, X \geq t - \beta_1)\mathbb{P}(Y \geq t - \alpha_2, Y \geq t - \beta_2) && \text{by independence} \\
&\qquad = \min(\bar{F}_X(t - \alpha_1), \bar{F}_X(t - \beta_1)) \min(\bar{F}_Y(t - \alpha_2), \bar{F}_Y(t - \beta_2)).
\end{aligned}$$

Therefore, for $\alpha_1 \odot X \oplus \alpha_2 \odot Y$ and $\beta_1 \odot X \oplus \beta_2 \odot Y$ to be independent, for all $t \in \mathbb{R}^2$, we need

$$\begin{aligned}
\bar{F}_X(t-\alpha_1)\bar{F}_X(t-\beta_1)\bar{F}_Y(t-\alpha_2)\bar{F}_Y(t-\beta_2) &= \min(\bar{F}_X(t-\alpha_1), \bar{F}_X(t-\beta_1)) \min(\bar{F}_Y(t-\alpha_2), \bar{F}_Y(t-\beta_2)) \cdot \\
&\quad \max(\bar{F}_X(t-\alpha_1), \bar{F}_X(t-\beta_1)) \max(\bar{F}_Y(t-\alpha_2), \bar{F}_Y(t-\beta_2)) \\
&= \min(\bar{F}_X(t-\alpha_1), \bar{F}_X(t-\beta_1)) \min(\bar{F}_Y(t-\alpha_2), \bar{F}_Y(t-\beta_2)).
\end{aligned}$$

But $\bar{F}_X$ and $\bar{F}_Y$ are nonincreasing functions taking values between 0 and 1. So

$$\bar{F}_X(t-\alpha_1)\bar{F}_X(t-\beta_1)\bar{F}_Y(t-\alpha_2)\bar{F}_Y(t-\beta_2) \leq \min(\bar{F}_X(t-\alpha_1), \bar{F}_X(t-\beta_1)) \min(\bar{F}_Y(t-\alpha_2), \bar{F}_Y(t-\beta_2)),$$

and equality holds if and only if

$$\min(\bar{F}_X(t-\alpha_1), \bar{F}_X(t-\beta_1)) = 0, \text{ or } \min(\bar{F}_Y(t-\alpha_2), \bar{F}_Y(t-\beta_2)) = 0.$$

As either of these scenarios must hold for each $t \in \mathbb{R}^2$, we conclude that X and Y must have compact supports. Conversely, suppose that X and Y have compact supports. Then one can choose $\alpha_1 = \beta_2 = 0$

and $\alpha_2 = \beta_1$ be a sufficiently large number, so that

$$\alpha_1 \odot X \oplus \alpha_2 \odot Y = X, \text{ and } \beta_1 \odot X \oplus \beta_2 \odot Y = Y.$$

In this case, the Darmois–Skitovich condition holds trivially, as desired. □

Now consider Pólya's condition. Here the Gaussian is characterized via stability under addition. When addition is replaced by minimum, it is well-known that this leads to the classical exponential distribution. One such characterization, which generalizes to distributions on arbitrary lattices, is the following [Bryc 1995, Theorem 3.4.1].

Theorem 3.3. *Suppose X, Y are independent and identically distributed nonnegative random variables. Then this distribution is the exponential if and only if for all $a, b > 0$ such that $a + b = 1$, $\min(X/a, Y/b)$ has the same distribution as X.*

By considering $\log(X)$ and $\log(Y)$, one could restate this theorem in terms of the min-plus algebra, though the condition $a + b = 1$ does not have an obvious tropical interpretation. This shows that the tropical analogue of Gaussian is the classical exponential distribution.

3.3. ***Gaussians in idempotent probability.*** Idempotent probability is a branch of idempotent analysis, which is functional analysis over idempotent semirings [Kolokoltsov and Maslov 1997]. Idempotent semirings are characterized by the additive operation being idempotent, that is, $a \oplus a = a$. The tropical semirings used in the previous sections are idempotent, but there are others, such as the Boolean semiring in semigroup theory. Idempotent analysis was developed by Litvinov, Maslov and Shipz [Litvinov et al. 1998] in relation to problems of calculus of variations. Closely related are the work on large deviations [Puhalskii 2001], which has found applications in queueing theory, as well as fuzzy measure theory and logic [Dubois and Prade 2000; Wang and Klir 1992]. The work we discussed in this section is based on that of Akian, Quadrat and Viot and coauthors [Akian et al. 2011; 1994], whose goal was to develop idempotent probability as a theory completely in parallel to classical probability. Following their convention, we work over the min-plus algebra.

All fundamental concepts of probability have an idempotent analogue, see [Akian et al. 1994] and references therein. For a flavor of this theory, we compare the concept of a measure. In classical settings, a probability measure μ is a map from the σ-algebra on a space Ω to $\mathbb{R}_{\geq 0}$ that satisfies three properties: (i) $\mu(\varnothing) = 0$, (ii) $\mu(\Omega) = 1$, and (iii) for a countable sequence (E_i) of pairwise disjoint sets,

$$\mu\Big(\bigcup_{i=1}^{\infty} E_i\Big) = \sum_{i=1}^{\infty} \mu(E_i).$$

The analogous object in the min-plus probability is the cost measure $\mathbb{K}$ defined by three axioms: (i) $\mathbb{K}(\varnothing) = +\infty$, (ii) $\mathbb{K}(\Omega) = 0$, and

$$\mathbb{K}\Big(\bigcup_i E_i\Big) = \bigoplus_i \mathbb{K}(E_i) = \inf_i \mathbb{K}(E_i).$$

Idempotent probability is rich and has interesting connections with dynamic programming and optimization. For instance, tropical matrix-vector multiplication can be interpreted as an update step in a Markov chain,

so the Bellman equation plays the analogue of the Kolmogorov–Chapman equation. Most notably, the classical quadratic form $(x-y)^2/2\sigma^2$ defines a stable distribution [Akian et al. 1994]. Furthermore, it is the unique density that is invariant under the Legendre–Fenchel transform [Akian et al. 1994], which is the tropical analogue of the Fourier transform [Kolokoltsov and Maslov 1997]. This is in parallel to the characterization of the scaled version of the Gaussian density $x \mapsto \exp(-\pi x^2)$ being invariant under the Fourier transform. While $x \mapsto \exp(-\pi x^2)$ is not the unique function to possess this property [Duffin 1948], the fact that the Fourier transform of a $\mathcal{N}(\mu, \sigma^2)$ univariate Gaussian has the form $x \mapsto \exp(i\mu x - \frac{x^2}{2})$ is frequently employed to prove independence of linear combinations of Gaussians. Under this light, one can regard the idempotent measure correspond to the density $(x-y)^2/2\sigma^2$ to be the idempotent analogue of the classical Gaussian.

4. Open directions

4.1. ***Tropical curves, metric graphs and Gaussians via the Laplacian operator.*** From the perspective of stochastic analysis, the Gaussian measure can be characterized as the unique invariant measure for the Ornstein–Uhlenbeck semigroup [Bogachev 1998, §1]. This semigroup is a powerful tool in proving hypercontractivity and log-Sobolev inequalities. In particular, the Gaussian density can be characterized as the function that satisfies such inequalities with the best constants [Bogachev 1998]. One useful characterization of the Ornstein–Uhlenbeck semigroup is by its generator, whose definition has two ingredients: a Laplacian operator and a gradient operator ∇. Let us elaborate. Let γ be a centered Gaussian measure on $\mathbb{R}^n$. The Ornstein–Uhlenbeck semigroup $(T_t, t \geq 0)$ is defined on $L^2(\gamma)$ by the Mehler formula

$$T_t h(x) = \int_{\mathbb{R}^n} h(e^{-t}x + \sqrt{1-e^{-2t}}y)\,\gamma(dy), \quad t > 0$$

and T_0 is the identity operator. It characterizes the Gaussian measure in the following sense [Bogachev 1998, §1].

Lemma 4.1. *γ is the unique invariant probability measure for $(T_t, t \geq 0)$.*

One can arrive at this semigroup without the Mehler's formula as follows. Let $\mathcal{D} = \{h \in L^2(\gamma) : \lim_{t\to 0} \frac{T_t h - h}{t}$ exists in the norm of $L^2(\gamma)\}$. (Recall that $L^2(\gamma)$ is the space of square integrable functions with respect to the measure γ). The linear operator L defined on $\mathcal{D}$ by

$$Lh = \lim_{t\to 0} \frac{T_t h - h}{t}$$

is called the generator of the semigroup $(T_t, t \geq 0)$. The generator of the Ornstein–Uhlenbeck semigroup is given by

$$Lh(x) = \Delta h(x) - \langle x, \nabla h\rangle = \sum_{i=1}^{n} \frac{\partial^2 h}{\partial x_i^2}(x) - \sum_{i=1}^{n} x_i \frac{\partial h}{\partial x_i}(x).$$

This generator uniquely specifies the semigroup. Importantly, the two ingredients needed to define L are the Laplacian operator Δ, and the gradient operator ∇. Thus the semigroup can be defined on Riemannian manifolds, for instance. This opens up ways to define Gaussians on tropical curves.

In tropical algebraic geometry, an abstract tropical curve is a metric graph [Mikhalkin and Zharkov 2008]. There are some minor variants: with vertex weights [Brannetti et al. 2011; Chan 2012], or just the compact part [Baker and Faber 2006]. An embedded tropical curve is a balanced weighted one-dimensional complex in $\mathbb{R}^n$. There are several constructions of tropical curves. In particular, they arise as limits of amoebas through a process called Maslov dequantization in idempotent analysis [Litvinov et al. 1998]. Tropical algebraic geometry took off with the landmark paper of Mikhalkin [2005], who used tropical curves to compute Gromov–Witten invariants of the plane $\mathbb{P}^2$ [Maclagan and Sturmfels 2015]. Since then, tropical curves, and more generally, tropical varieties, have been studied in connection to mirror and symplectic geometry [Gross 2011]. Another heavily explored aspect of tropical curves is their divisors and Riemann–Roch theory [Baker and Norine 2007; Baker and Payne 2016; Gathmann and Kerber 2008; Mikhalkin and Zharkov 2008]. This theory is connected to chip-firing and sandpiles, which were initially conceived as deterministic models of random walks on graphs [Cooper and Spencer 2006].

Metric graphs are Riemannian manifolds with singularities [Baker and Faber 2006]. Brownian motions defined on metric graphs, heat semigroups on graphs, and graph Laplacians are an active research area [Kostrykin et al. 2012; Post 2009]. As of now, however, the author is unaware of an analogue of the Ornstein–Uhlenbeck semigroup and its invariant measure on graphs. It would also be interesting to study what Brownian motion on graphs reveals about tropical curves and their Jacobians.

4.2. ***Further open directions.*** The natural ambient space for doing tropical convex geometry is not $\mathbb{R}^m$, but $\mathbb{TP}^{n-1}$, where a vector $x \in \mathbb{R}^m$ is identified with all of its scalar multiples $a \odot x$. Probability theory on classical projective spaces relies on group representation [Benoist and Quint 2014]. Unfortunately, there is no satisfactory tropical analogue of the general linear group. Every invertible $n \times n$ matrix with entries in $\overline{\mathbb{R}}$ is the composition of a diagonal matrix and a permutation of the standard basis of $\overline{\mathbb{R}}^n$ [Kolokoltsov and Maslov 1997]. We note that several authors have studied tropicalization of special linear group over a field with valuation [Joswig et al. 2007; Werner 2011]. It would be interesting to see whether this can be utilized to define probability measures on $\mathbb{TP}^{n-1}$.

Another approach is to 'fix' the main difficulty with the idempotent algebra, namely, the lack of the additive inverse. Some authors have put back the additive inverse and developed a theory of linear algebra in this new algebra, called the supertropical algebra [Izhakian and Rowen 2010]. It would be interesting to study matrix groups and their actions under this algebra, and in particular, pursue the definition of Gaussians as invariant measures under actions of the orthogonal group.

4.3. ***Beyond Gaussians.*** In a more applied direction, $\mathbb{TP}^{n-1}$ is a natural ambient space to study problems in economics, network flow and phylogenetics. Thus one may want an axiomatic approach to finding distributions on $\mathbb{TP}^{n-1}$ tailored for specific applications. For instance, in shape-constrained density estimation, log-concave multivariate totally positive of ordered two (MTP2) distributions are those whose density $f : \mathbb{R}^d \to \mathbb{R}$ is log-concave and satisfies the inequality

$$f(x)f(y) \leq f(x \vee y)f(x \wedge y) \quad \text{for all } x, y \in \mathbb{R}^d.$$

A variety of distributions belong to this family. Requiring that such inequalities hold for all $x, y \in \mathbb{TP}^{n-1}$ leads to the stronger condition of $L^\natural$-concavity

$$f(x)f(y) \le f((x+\alpha\mathbf{1}) \vee y)f(x \wedge (y-\alpha\mathbf{1})) \quad \text{for all } x, y \in \mathbb{R}^d, \alpha \ge 0.$$

A Gaussian distribution is log-concave MTP2 if and only if the inverse of its covariance matrix is an M-matrix [Lauritzen et al. 2019]. Only diagonally dominant Gaussians are $L^\natural$-concave [Murota 2003, §2]. This subclass of densities has nice properties that make them algorithmically tractable in Gaussian graphical models [Malioutov et al. 2006; Weiss and Freeman 2001]. In particular, density estimation for $L^\natural$-concave distributions is significantly easier than for log-concave MTP2 [Robeva et al. 2018]. It would be interesting to pursue this direction to define distributions on the space of phylogenetic trees.

References

[Akian et al. 1994] M. Akian, J.-P. Quadrat, and M. Viot, "Bellman processes", pp. 302–311 in *11th International Conference on Analysis and Optimization of Systems Discrete Event Systems*, Springer, 1994.

[Akian et al. 2006] M. Akian, R. Bapat, and S. Gaubert, "Max-plus algebra", in *Handbook of linear algebra*, Chapman and Hall, London, 2006.

[Akian et al. 2011] M. Akian, S. Gaubert, V. Niţică, and I. Singer, "Best approximation in max-plus semimodules", *Linear Algebra Appl.* **435**:12 (2011), 3261–3296.

[Akian et al. 2012] M. Akian, S. Gaubert, and A. Guterman, "Tropical polyhedra are equivalent to mean payoff games", *Internat. J. Algebra Comput.* **22**:1 (2012), art. id. 1250001.

[Allamigeon et al. 2011] X. Allamigeon, S. Gaubert, and R. D. Katz, "The number of extreme points of tropical polyhedra", *J. Combin. Theory Ser. A* **118**:1 (2011), 162–189.

[Allamigeon et al. 2018] X. Allamigeon, P. Benchimol, S. Gaubert, and M. Joswig, "Log-barrier interior point methods are not strongly polynomial", *SIAM J. Appl. Algebra Geom.* **2**:1 (2018), 140–178.

[Ardila and Develin 2009] F. Ardila and M. Develin, "Tropical hyperplane arrangements and oriented matroids", *Math. Z.* **262**:4 (2009), 795–816.

[Avendaño and Ibrahim 2011] M. Avendaño and A. Ibrahim, "Multivariate ultrametric root counting", pp. 1–23 in *Randomization, relaxation, and complexity in polynomial equation solving*, Contemp. Math. **556**, Amer. Math. Soc., 2011.

[Baccelli et al. 1992] F. L. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, *Synchronization and linearity: an algebra for discrete event systems*, Wiley, Chichester, 1992.

[Baker and Faber 2006] M. Baker and X. Faber, "Metrized graphs, Laplacian operators, and electrical networks", pp. 15–33 in *Quantum graphs and their applications*, Contemp. Math. **415**, Amer. Math. Soc., 2006.

[Baker and Norine 2007] M. Baker and S. Norine, "Riemann–Roch and Abel–Jacobi theory on a finite graph", *Adv. Math.* **215**:2 (2007), 766–788.

[Baker and Payne 2016] M. Baker and S. Payne (editors), *Nonarchimedean and tropical geometry* (St. John, 2013 and Puerto Rico, 2015), Springer, 2016.

[Baldwin and Klemperer 2019] E. Baldwin and P. Klemperer, "Understanding preferences: "demand types", and the existence of equilibrium with indivisibilities", *Econometrica* **87**:3 (2019), 867–932.

[Barron 1986] A. R. Barron, "Entropy and the central limit theorem", *Ann. Probab.* **14**:1 (1986), 336–342.

[Benoist and Quint 2014] Y. Benoist and J.-F. Quint, "Random walks on projective spaces", *Compos. Math.* **150**:9 (2014), 1579–1606.

[Bogachev 1998] V. I. Bogachev, *Gaussian measures*, Mathematical Surveys and Monographs **62**, American Mathematical Society, 1998.

[Brannetti et al. 2011] S. Brannetti, M. Melo, and F. Viviani, "On the tropical Torelli map", *Adv. Math.* **226**:3 (2011), 2546–2586.

[Briec and Horvath 2008] W. Briec and C. Horvath, "Halfspaces and Hahn–Banach like properties in $\mathbb{B}$-convexity and max-plus convexity", *Pac. J. Optim.* **4**:2 (2008), 293–317.

[Bryc 1995] W. Bryc, *The normal distribution: characterizations with applications*, Lecture Notes in Statistics **100**, Springer, 1995.

[Butkovic and Fiedler 2018] P. Butkovic and M. Fiedler, "Tropical tensor product and beyond", 2018. arXiv 1805.03174

[Butkovič 2010] P. Butkovič, *Max-linear systems: theory and algorithms*, Springer, 2010.

[Butkovič et al. 2007] P. Butkovič, H. Schneider, and S. Sergeev, "Generators, extremals and bases of max cones", *Linear Algebra Appl.* **421**:2-3 (2007), 394–406.

[Chan 2012] M. T. Chan, *Tropical curves and metric graphs*, Ph.D. thesis, University of California, Berkeley, 2012, available at https://www.proquest.com/docview/1081722820.

[Chan et al. 2011] M. Chan, A. Jensen, and E. Rubei, "The 4×4 minors of a $5 \times n$ matrix are a tropical basis", *Linear Algebra Appl.* **435**:7 (2011), 1598–1611.

[Cohen et al. 2001] G. Cohen, S. Gaubert, and J.-P. Quadrat, "Hahn–Banach separation theorem for max-plus semimodules", pp. 325–334 in *Optimal control and partial differential equations*, IOS, Amsterdam, 2001.

[Cohen et al. 2004] G. Cohen, S. Gaubert, and J.-P. Quadrat, "Duality and separation theorems in idempotent semimodules", *Linear Algebra Appl.* **379** (2004), 395–422.

[Cooper and Spencer 2006] J. N. Cooper and J. Spencer, "Simulating a random walk with constant error", *Combin. Probab. Comput.* **15**:6 (2006), 815–822.

[Crowell and Tran 2016] R. A. Crowell and N. M. Tran, "Tropical Geometry and Mechanism Design", 2016. arXiv 1606.04880

[Darmois 1953] G. Darmois, "Analyse générale des liaisons stochastiques: étude particulière de l'analyse factorielle linéaire", *Rev. Inst. Internat. Statist.* **21** (1953), 2–8.

[Develin and Sturmfels 2004] M. Develin and B. Sturmfels, "Tropical convexity", *Doc. Math.* **9** (2004), 1–27.

[Develin et al. 2005] M. Develin, F. Santos, and B. Sturmfels, "On the rank of a tropical matrix", pp. 213–242 in *Combinatorial and computational geometry*, Math. Sci. Res. Inst. Publ. **52**, Cambridge Univ. Press, 2005.

[Dragovich et al. 2017] B. Dragovich, A. Y. Khrennikov, S. V. Kozyrev, I. V. Volovich, and E. I. Zelenov, "p-adic mathematical physics: the first 30 years", *p-Adic Numbers Ultrametric Anal. Appl.* **9**:2 (2017), 87–121.

[Dubois and Prade 2000] D. Dubois and H. Prade (editors), *Fundamentals of fuzzy sets*, The Handbooks of Fuzzy Sets Series **7**, Kluwer, 2000.

[Duffin 1948] R. J. Duffin, "Function classes invariant under the Fourier transform", *Duke Math. J.* **15** (1948), 781–785.

[Elsner and van den Driessche 2004] L. Elsner and P. van den Driessche, "Max-algebra and pairwise comparison matrices", *Linear Algebra Appl.* **385** (2004), 47–62.

[Evans 2001] S. N. Evans, "Local fields, Gaussian measures, and Brownian motions", pp. 11–50 in *Topics in probability and Lie groups: boundary theory*, CRM Proc. Lecture Notes **28**, Amer. Math. Soc., 2001.

[Evans 2002] S. N. Evans, "Elementary divisors and determinants of random matrices over a local field", *Stochastic Process. Appl.* **102**:1 (2002), 89–102.

[Evans 2006] S. N. Evans, "The expected number of zeros of a random system of p-adic polynomials", *Electron. Comm. Probab.* **11** (2006), 278–290.

[Fink and Rincón 2015] A. Fink and F. Rincón, "Stiefel tropical linear spaces", *J. Combin. Theory Ser. A* **135** (2015), 291–331.

[Gathmann and Kerber 2008] A. Gathmann and M. Kerber, "A Riemann–Roch theorem in tropical geometry", *Math. Z.* **259**:1 (2008), 217–230.

[Gaubert 1996] S. Gaubert, "On the Burnside problem for semigroups of matrices in the (max, +) algebra", *Semigroup Forum* **52**:3 (1996), 271–292.

[Gaubert and Katz 2011] S. Gaubert and R. D. Katz, "Minimal half-spaces and external representation of tropical polyhedra", *J. Algebraic Combin.* **33**:3 (2011), 325–348.

[Giansiracusa and Giansiracusa 2018] J. Giansiracusa and N. Giansiracusa, "A Grassmann algebra for matroids", *Manuscripta Math.* **156**:1-2 (2018), 187–213.

[Gross 2011] M. Gross, *Tropical geometry and mirror symmetry*, CBMS Regional Conference Series in Mathematics **114**, American Mathematical Society, 2011.

[Gursoy et al. 2013] B. B. Gursoy, O. Mason, and S. Sergeev, "The analytic hierarchy process, max algebra and multi-objective optimisation", *Linear Algebra Appl.* **438**:7 (2013), 2911–2928.

[Hampe 2015] S. Hampe, "Tropical linear spaces and tropical convexity", *Electron. J. Combin.* **22**:4 (2015), art. id. 4.43.

[Herschel 1850] J. F. W. Herschel, "Quetelet on probabilities", *Edinburgh Rev.* **92**:185 (1850), 1–30.

[Hollings and Kambites 2012] C. Hollings and M. Kambites, "Tropical matrix duality and Green's $\mathscr{D}$ relation", *J. Lond. Math. Soc.* (2) **86**:2 (2012), 520–538.

[Hook 2017] J. Hook, "Max-plus algebraic statistical leverage scores", *SIAM J. Matrix Anal. Appl.* **38**:4 (2017), 1410–1433.

[Huh 2018] J. Huh, "Tropical geometry of matroids", pp. 1–46 in *Current developments in mathematics 2016*, International Press, Somerville, MA, 2018.

[Izhakian and Rowen 2009] Z. Izhakian and L. Rowen, "The tropical rank of a tropical matrix", *Comm. Algebra* **37**:11 (2009), 3912–3927.

[Izhakian and Rowen 2010] Z. Izhakian and L. Rowen, "A guide to supertropical algebra", pp. 283–302 in *Advances in ring theory*, Birkhäuser, 2010.

[Joswig 2017] M. Joswig, "The Cayley trick for tropical hypersurfaces with a view toward Ricardian economics", pp. 107–128 in *Homological and computational methods in commutative algebra*, Springer INdAM Ser. **20**, Springer, 2017.

[Joswig $\geq$ 2020] M. Joswig, *Essentials of tropical combinatorics*, In preparation.

[Joswig and Loho 2016] M. Joswig and G. Loho, "Weighted digraphs and tropical cones", *Linear Algebra Appl.* **501** (2016), 304–343.

[Joswig et al. 2007] M. Joswig, B. Sturmfels, and J. Yu, "Affine buildings and tropical convexity", *Albanian J. Math.* **1**:4 (2007), 187–211.

[Kac 1939] M. Kac, "On a characterization of the normal distribution", *Amer. J. Math.* **61** (1939), 726–728.

[Kagan et al. 1972] A. M. Kagan, Y. V. Linnik, and S. R. Rao, Характеризационные задачи математической статистики, Nauka, 1972. Translated as *Characterization problems in mathematical statistics*, 1973.

[Kallenberg 2002] O. Kallenberg, *Foundations of modern probability*, 2nd ed., Springer, 2002.

[Kolokoltsov and Maslov 1997] V. N. Kolokoltsov and V. P. Maslov, *Idempotent analysis and its applications*, Mathematics and its Applications **401**, Kluwer, 1997.

[Komenda et al. 2018] J. Komenda, S. Lahaye, J.-L. Boimond, and T. van den Boom, "Max-plus algebra in the history of discrete event systems", *Annu. Rev. Control* **45** (2018), 240–249.

[Kostrykin et al. 2012] V. Kostrykin, J. Potthoff, and R. Schrader, "Brownian motions on metric graphs", *J. Math. Phys.* **53**:9 (2012), art. id. 095206.

[Lauritzen et al. 2019] S. Lauritzen, C. Uhler, and P. Zwiernik, "Maximum likelihood estimation in Gaussian models under total positivity", *Ann. Statist.* **47**:4 (2019), 1835–1863.

[Litvinov et al. 1998] G. L. Litvinov, V. P. Maslov, and G. B. Shpiz, "Linear functionals on idempotent spaces: an algebraic approach", *Dokl. Akad. Nauk* **363**:3 (1998), 298–300. In Russian; translated in *Dokl. Math.* **58** (1998), pp. 389–391.

[Loho and Smith 2020] G. Loho and B. Smith, "Matching fields and lattice points of simplices", *Adv. Math.* **370** (2020), art. id. 107232.

[Maclagan and Sturmfels 2015] D. Maclagan and B. Sturmfels, *Introduction to tropical geometry*, Graduate Studies in Mathematics **161**, American Mathematical Society, 2015.

[Malioutov et al. 2006] D. V. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models", *J. Mach. Learn. Res.* **7** (2006), 2031–2064.

[Maxwell 1860] J. C. Maxwell, "V. Illustrations of the dynamical theory of gases, I: On the motions and collisions of perfectly elastic spheres", *Philos. Mag.* **19**:124 (1860), 19–32.

[Mikhalkin 2005] G. Mikhalkin, "Enumerative tropical algebraic geometry in $\mathbb{R}^2$", *J. Amer. Math. Soc.* **18**:2 (2005), 313–377.

[Mikhalkin and Zharkov 2008] G. Mikhalkin and I. Zharkov, "Tropical curves, their Jacobians and theta functions", pp. 203–230 in *Curves and abelian varieties*, Contemp. Math. **465**, Amer. Math. Soc., 2008.

[Monod et al. 2018] A. Monod, B. Lin, R. Yoshida, and Q. Kang, "Tropical geometry of phylogenetic tree space: a statistical perspective", 2018. arXiv 1805.12400

[Morrison and Tran 2016] R. Morrison and N. M. Tran, "The tropical commuting variety", *Linear Algebra Appl.* **507** (2016), 300–321.

[Murota 2003] K. Murota, *Discrete convex analysis*, SIAM, Philadelphia, PA, 2003.

[Nitica and Singer 2007] V. Nitica and I. Singer, "Max-plus convex sets and max-plus semispaces. I", *Optimization* **56**:1-2 (2007), 171–205.

[Page et al. 2020] R. Page, R. Yoshida, and L. Zhang, "Tropical principal component analysis on the space of phylogenetic trees", *Bioinformatics* (2020).

[Pólya 1923] G. Pólya, "Herleitung des Gaußschen Fehlergesetzes aus einer Funktionalgleichung", *Math. Z.* **18**:1 (1923), 96–108.

[Post 2009] O. Post, "Spectral analysis of metric graphs and related spaces", pp. 109–140 in *Limits of graphs in group theory and computer science*, EPFL Press, Lausanne, 2009.

[Puhalskii 2001] A. Puhalskii, *Large deviations and idempotent probability*, Monographs and Surveys in Pure and Applied Mathematics **119**, Chapman and Hall/CRC, 2001.

[Robeva et al. 2018] E. Robeva, B. Sturmfels, N. Tran, and C. Uhler, "Maximum likelihood estimation for totally positive log-concave densities", 2018. To appear in *Scand. J. Stat.* arXiv 1806.10120

[Sergeev 2009] S. Sergeev, "Multiorder, Kleene stars and cyclic projectors in the geometry of max cones", pp. 317–342 in *Tropical and idempotent mathematics*, Contemp. Math. **495**, Amer. Math. Soc., 2009.

[Shiozawa 2015] Y. Shiozawa, "International trade theory and exotic algebras", *Evolut. Institut. Econ. Rev.* **12**:1 (2015), 177–212.

[Shitov 2011] Y. N. Shitov, "An example of a (6×6)-matrix with different tropical and Kapranov ranks", *Vestnik Moskov. Univ. Ser. I Mat. Mekh.* 5 (2011), 58–61. In Russian; translated in ıtMoscow Univ. Math. Bull. **66**:5 (2011), 227–229.

[Simon 1994] I. Simon, "On semigroups of matrices over the tropical semiring", *RAIRO Inform. Théor. Appl.* **28**:3-4 (1994), 277–294.

[Skitovich 1953] V. P. Skitovich, "On a property of the normal distribution", *Doklady Akad. Nauk SSSR* (*N.S.*) **89** (1953), 217–219. In Russian.

[Sturmfels and Tran 2013] B. Sturmfels and N. M. Tran, "Combinatorial types of tropical eigenvectors", *Bull. Lond. Math. Soc.* **45**:1 (2013), 27–36.

[Tran 2013] N. M. Tran, "Pairwise ranking: choice of method can produce arbitrarily different rank order", *Linear Algebra Appl.* **438**:3 (2013), 1012–1024.

[Tran 2017] N. M. Tran, "Enumerating polytropes", *J. Combin. Theory Ser. A* **151** (2017), 1–22.

[Tran and Yu 2019] N. M. Tran and J. Yu, "Product-mix auctions and tropical geometry", *Math. Oper. Res.* **44**:4 (2019), 1396–1411.

[Tsukerman 2015] E. Tsukerman, "Tropical spectral theory of tensors", *Linear Algebra Appl.* **481** (2015), 94–106.

[Wang and Klir 1992] Z. Y. Wang and G. J. Klir, *Fuzzy measure theory*, Plenum Press, New York, 1992.

[Weiss and Freeman 2001] Y. Weiss and W. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology", *Neural Comput.* **13** (2001).

[Werner 2011] A. Werner, "A tropical view on Bruhat–Tits buildings and their compactifications", *Cent. Eur. J. Math.* **9**:2 (2011), 390–402.

[Yoshida et al. 2019] R. Yoshida, L. Zhang, and X. Zhang, "Tropical principal component analysis and its application to phylogenetics", *Bull. Math. Biol.* **81**:2 (2019), 568–597.

Received 2018-09-25. Accepted 2020-07-06.

NGOC MAI TRAN: ntran@math.utexas.edu
University of Texas at Austin, Austin, TX 78712, United States

and

Hausdorff Center for Mathematics, Bonn 53115, Germany

Algebraic Statistics
Vol. 11, No. 2, 2020
https://doi.org/10.2140/astat.2020.11.169

THE NORM OF THE SATURATION OF A BINOMIAL IDEAL, WITH APPLICATIONS TO MARKOV BASES

DAVID HOLMES

Let B be a finite set of pure binomials in the variables x_i, and write I_B for the ideal generated by these binomials. We define a new measure of the complexity of the saturation of the ideal I_B with respect to the product of the x_i, which we call the *norm* of B. We give a bound on the norm in terms of easily computed invariants of B. We discuss statistical applications, both practical and theoretical.

1. Introduction

1.1. ***Background.*** Let A be a $k \times r$ matrix with integer entries, and let $u \in \mathbb{N}^r$ be a vector with nonnegative entries. The *fibre containing* u is defined as

$$\mathcal{F}(u) = \{v \in \mathbb{N}^r : Au = Av\}. \tag{1.1.1}$$

Understanding the structure of this fibre is important in a number of statistical tests. For example, the vectors in $\mathbb{N}^r$ might represent tables of data, and the matrix A might output the row and column sums of these tables, so the fibre consists of all tables with nonnegative entries and with the same row and column sums as the starting table u. See [Diaconis and Sturmfels 1998] for more details and examples. In particular, one often wants to generate samples from some probability distribution (often uniform or hypergeometric) on the fibre. If the fibre is small it is feasible to simply enumerate all the elements of the fibre. However, in practical applications the fibre is often far too large to enumerate, and the standard approach is to perform a *random walk* in the fibre, generating samples via the Metropolis–Hastings Markov chain Monte Carlo algorithm. In order to perform a random walk, we must upgrade the fibre into a graph (whose vertices are the elements of the fibre). The requirements for the Metropolis–Hastings algorithm are rather mild, the key condition is that the graph must be *connected* (since the random walk will always remain within its starting connected component).

MSC2010: 13P25, 14M25.
Keywords: markov basis, saturation, toric ideals.

1.1.1. *A random walk in the fibre.* The most naive way to convert the fibre into a graph is to choose a generating set B for the kernel $K \subseteq \mathbb{Z}^r$ of A as a $\mathbb{Z}$-module, and then form a (simple, undirected) graph by putting an edge between distinct vertices v_1 and v_2 whenever $v_1 - v_2 \in B$ or $v_2 - v_1 \in B$. We say $\mathcal{F}(u)$ is *connected by* B if the resulting graph is connected. In Section 2 we will see several examples of B that *fail* to connect $\mathcal{F}(u)$. The major innovation of Diaconis and Sturmfels [1998] was to give an algorithm to construct a generating set B which connects *every* fibre of a given matrix A.

1.1.2. *Saturated ideals and connected fibres.* To describe their result, we need a little more notation. Given $b \in B$, we write $b = b^+ - b^-$, both summands having nonnegative entries. In the ring $R = \mathbb{Z}[x_1, \ldots, x_r]$ we form the elements

$$x^{b^+} := \prod_{i=1}^{r} x_i^{b_i^+}, \quad x^{b^-} := \prod_{i=1}^{r} x_i^{b_i^-}, \tag{1.1.2}$$

and define an ideal $\mathcal{I}_B = (x^{b^+} - x^{b^-} : b \in B) \subseteq R$. Then the key theorem is the following (where we use [Sturmfels 1996, Lemma 12.2 p. 114] to interpret toric ideals as saturated ideals).

Theorem 1.1 [Diaconis and Sturmfels 1998]. *Fix a $k \times r$ matrix A, and let B be a generating set for the integral kernel of A. Suppose the ideal $\mathcal{I}_B$ is saturated with respect to the element $x_1 \cdots x_r \in R$. Then for every $u \in \mathbb{N}^r$, the fibre $\mathcal{F}(u)$ is connected by B.*

If $\mathcal{I}_B$ is saturated, B is often called a *Markov basis* (though we use the word "basis", this should not be interpreted as implying linear independence of the elements of B). The theorem then tells us that we can generate samples according to our preferred distribution by following the naive random walk algorithm above using the basis B.

On the other hand, suppose that we have a generating set B such that $\mathcal{I}_B$ is not saturated. We can (at least in principal) apply a standard saturation algorithm to $\mathcal{I}_B$ to produce a saturated ideal, and moreover the generating set produced will in fact consist of pure difference binomials (i.e. differences of monomials; see Definition 4.1). Reversing the procedure (1.1.2) we can recover a new generating set B' for the kernel K of A, and following the above theorem of Diaconis–Sturmfels, this generating set will connect all fibres, enabling efficient sampling.

Thus, when it is possible to compute this saturation, the problem is essentially solved. However, the standard algorithm for saturation involves r computations of Gröbner bases (where r is the number of columns as above), and is at present only practical for relatively small examples. General purpose algorithms (not taking advantage of the toric structure) are available in many packages (such as MAGMA [Bosma et al. 1997] and Singular [Decker et al. 2019]), and also specialised implementations for the toric case are available (CoCoa [Bigatti et al. 1999], 4ti2 [Hemmecke et al. 2001–]).

1.1.3. *Connected fibres without saturation.* The difficulty of computing the saturation motivated Aoki, Hara and Takemura [Hara et al. 2012] to suggest an algorithm for generating samples without needing to compute the saturation. They begin in the same way, with a generating set $B = \{b_1, \ldots, b_n\}$ for the integral kernel, but instead of making moves consisting of addition or subtraction of a single element of B, they instead generate n nonnegative integers a_i from a Poisson distribution with some chosen mean λ,

and n elements $\epsilon_i \in \{+1, -1\}$, and their move consists of addition of $\sum_i \epsilon_i a_i b_i$ if the result lies in the fibre, and staying put otherwise. Since the Poisson distribution generates every nonnegative integer with nonzero probability it is immediate that the resulting fibre is connected; in fact, the graph on the fibre is a complete graph, but with highly nonuniform probability of selecting moves from among edges.

They then perform a number of numerical experiments with various values of λ. In cases where it was possible to compute the saturation, they show that for careful choice of λ their algorithm performs comparably to that coming from a Markov basis, and they also illustrate that their algorithm can be applied in cases where the saturation is too hard to compute (though they can of course provide no guarantee that their algorithm is converging in reasonable time; it appears to do so, but this might be deceptive if the fibre has some connected components that are very hard to hit — see Section 1.4).

There is some tension in the use of this algorithm when it comes to choosing the value of λ. If one chooses λ very large then the algorithm takes a long time before it (appears to) converge. On the other hand, a small value of λ will product more rapid *apparent* convergence, but there is a greater risk that one is simply failing to see one or more connected components of the fibre in the time for which the algorithm is run.

1.2. *Results.*

1.2.1. *A bound on the complexity of the saturation.* In the light of the above discussion it is natural to try to bound how large and complex the saturation of the ideal $\mathcal{I}_B$ can get. To make this more precise, we define the *norm* of the generating set B as follows.

Definition 1.2. Let B be set of $n \geq 1$ vectors in $\mathbb{Z}^r$. We write $\mathcal{I}_B$ for the ideal in $R = \mathbb{Z}[x_1, \ldots, x_r]$ as defined in Section 1.1.2. The *norm* of B is the smallest integer $N \geq 1$ such that there exists a finite generating set G for the saturation of $\mathcal{I}_B$ with respect to $x_1 \cdots x_r$, with the properties that

(1) Every element of G is a pure difference binomial;

(2) Every $g \in G$ can be written in the form

$$g = \sum_{i=1}^{N} \epsilon_i m_i (x^{b_i^+} - x^{b_i^-}). \tag{1.2.1}$$

where the $\epsilon_i \in \{-1, 0, 1\}$, the m_i are Laurent monomials, and the b_i are elements of B.

The main result of this paper is the following explicit bound on the norm. In Sections 1.3.1–1.3.2 we will show how this can be applied to give new algorithms for sampling from fibres without needing to compute the saturation.

Theorem 1.3. *Let B be set of $n \geq 1$ vectors in $\mathbb{Z}^r$. Write β for the maximum of the absolute values of the coefficients of elements of B. Then the norm of B is at most*

$$n^n \beta^{n-1}. \tag{1.2.2}$$

Our proof (see Section 4.1) is constructive; it gives an algorithm to determine a generating set G as in the definition of the norm. We do not know whether this algorithm could be practical; it is a-priori less efficient than procedures using Gröbner bases, but is highly parallelisable.

The connection of the norm to fibre connectivity and Markov chains runs via the following result (proven in Section 4.2).

Proposition 1.4. *Let A be a $k \times r$ integer matrix, and $B = \{b_1, \dots, b_n\}$ a basis of the kernel, with B having norm N. Let $u \in \mathbb{N}^r$, and construct a graph with vertex set the fibre $\mathcal{F}(u)$, and where we draw an edge from v_1 to v_2 if and only if $v_1 - v_2$ can be written as an integer linear combination*

$$v_1 - v_2 = \sum_{i=1}^{n} a_i b_i$$

with $\sum_{i=1}^{n} |a_i| \leq N$. Then this graph is connected.

Remark 1.5. Given a $k \times r$ integral matrix A, note that it is easy to compute a basis B of the integral kernel of A from the Smith normal form of A. Indeed, if $SAT = D$ is the Smith normal form (so S and T are invertible, and D diagonal with $D_{i,i} \mid D_{i+1,i+1}$), then let $1 \leq j \leq r$ be maximal such that $D_{j,j} \neq 0$. Then an integral basis of the kernel of A is given by $Te_{j+1}, \dots, Te_r$, where e_i is the i-th standard basis vector in $\mathbb{Z}^r$.

Conversely, while B does not determine A, it does determine the fibres $\mathcal{F}(u)$, so the matrix A is not really essential, but is very relevant to the statistical applications.

1.2.2. *Comparison to other results in the literature.* Needless to say, we are not the first to try to control the complexity of the saturation of an ideal in a polynomial ring. Indeed, the standard method of computing the saturation reduces to a Gröbner basis computation, whose efficient implementation has been the focus of too much research to begin to list here. Specialising to the case of binomial ideals, the literature is still much too large to give more than a quick glimpse of. There are general theoretical results on the structure of fibre graphs; see, e.g., [Gross and Petrović 2013; Hemmecke and Windisch 2015; Windisch 2016; 2019]. There are also many results bounding the degree of the binomials appearing in the saturation [Haws et al. 2014; Koyama et al. 2015; Sturmfels 1996, Chapter 13], and bounding the *Markov complexity*; this is defined in [Santos and Sturmfels 2003], and studied in [Charalambous et al. 2014] and elsewhere.

However, we are not aware on bounds on the norm 1.2 in the literature. Indeed, from an algebraic point of view it appears a rather unnatural invariant. The reason for studying it comes purely from the application (via Proposition 1.4) to fibre connectivity and Markov bases. In the remainder of Section 1 we hope to justify it from this point of view, and perhaps motivate further research in this direction. An unusual feature of our results is that we do not utilise Gröbner bases; this is not from dislike, but simply because we could not see how to bound the norm from that perspective; we hope that others may have more success.

1.3. *Algorithms.*

1.3.1. *Bounded-AHT algorithms.* Aoki, Hara and Takemura connect the fibre by allowing arbitrarily large integer linear combinations of elements of the basis B. This is guaranteed to connect the fibre (since it eventually hits every integer vector), but risks wasting time searching far away from the fibre.

Proposition 1.4 shows that it actually suffices to take combinations with coefficients bounded by the norm N of B; this allows us to improve the efficiency of their algorithm, by truncating the Poisson distribution at N, spending less time exploring far from the fibre. A second algorithm they present (where the coefficients of the b_i are chosen from a multinomial distribution) can be enhanced in a similar way. An even simpler variant is to choose uniformly at random at each step a vector of L^∞-length bounded by the norm.[1]

We will refer to this class of algorithm as *bounded-AHT* algorithms, as they are characterised by the distribution used to select random vectors being of bounded support. We will see in Section 2.1 that, when a good bound on the norm is available, such an algorithm can be substantially faster than the conventional AHT algorithm.

The bound on the norm coming from Theorem 1.3 is in general large, so using it for truncation will not have a large impact on the runtime (though we hope that better bounds on the norm can be found in the future). On the other hand, if a Markov basis can be computed one can obtain a very tight bound on the norm, and our algorithm seems to converge substantially faster than that of Diaconis–Sturmfels, so it is plausible that these bounded-AHT algorithms give the best performance in these cases also.

Another application might be to predicting good values of the constant λ in the AHT algorithm, or giving heuristic bounds on the convergence time for a given value of λ. The norm N can be seen as the maximum distance between connected components of the fibre, thus to have a reasonable chance of hitting all components we should take a number of steps that is very large compared to $1/\mathbb{P}(\text{Poisson}_\lambda \geq N)$.

1.3.2. *The stepping-out algorithm.* In the naive algorithm of Section 1.1.1, one starts at a vector $v \in \mathcal{F}(u)$, and chooses at random an element $b \in \pm B$, and considers the step $v + b$. If $v + b$ is in $\mathcal{F}(u)$ then this is returned as the next element of the Markov chain. If $v + b \notin \mathcal{F}(u)$, then the algorithm simply returns v. However, if we have a bound on the norm then we can modify the algorithm so that the fibre will always be connected; if $v + b \notin \mathcal{F}(u)$ then, rather than returning v, we choose another element b_1 from $\pm B$, and consider the vector $v + b + b_1$. If $v + b + b_1$ lies in $\mathcal{F}(u)$ we return $v + b + b_1$ as the next step in the Markov chain, otherwise we repeat, until we either hit $\mathcal{F}(u)$ again, or we have taken N consecutive steps outside the fibre, in which case we return v again. Alternatively, this can be viewed as a weighted random walk in a certain graph with vertex set $\mathcal{F}(u)$. To define this graph, we first define a graph $\mathcal{F}_{\mathbb{Z}}(u)$ with vertex set $\{v \in \mathbb{Z}^r : Au = Av\}$ and with an edge between v_1 and v_2 whenever $v_1 - v_2 \in \pm B$. Then we define a graph with vertex set $\mathcal{F}(u)$ by putting an edge between two vertices whenever they can be connected by a path in $\mathcal{F}_{\mathbb{Z}}(u)$ of length at most N, and which does not intersect $\mathcal{F}(u)$ except at its endpoints. Again, by Proposition 1.4 this new graph is guaranteed to be connected.

In the examples in Section 2.1, the best performance seems to be obtained by bounded-AHT algorithms. However, we include the stepping-out algorithm because it is an example of a general technique where one can choose any algorithm to efficiently explore the interior of the fibre, and then add some "small" extra steps on the boundary to ensure that the resulting graph is connected.

1.3.3. *Speed comparisons.* In Section 2.1 we describe some numerical experiments to compare the performance of the four algorithms:

[1] That is, with the maximum entry bounded by the norm.

(1) the algorithm of Diaconis–Sturmfels using a Markov basis (DS),

(2) the algorithm of Aoki, Hara and Takemura (AHT);

(3) the bounded-AHT algorithm;

(4) the stepping-out algorithm.

The best results are obtained with the new bounded-AHT algorithm, and the worst with the DS algorithm. In between, the AHT algorithm is faster than the stepping-out algorithm. But one should not extrapolate too much from this small collection of examples.

More generally, with Theorem 1.3 and Proposition 1.4 in hand it is easy to propose new sampling algorithms which guarantee to connect the fibre. The challenge is to design algorithms with reasonable runtime, at least heuristically (rigorous runtime analysis seems hard but very interesting).

If the fibre $\mathcal{F}(u)$ is large with respect to the norm N then designing reasonably efficient algorithms is not hard, since the runtime will be dominated by time spent in the "interior" of the fibre. On the other hand, if the fibre is small compared to N then the runtime will be dominated by time spent around the edge of the fibre looking for new connected components, and will depend sensitively on the norm (or more precisely, on our bound on the norm).

1.4. *Practical consequences.*

(1) The norm bounds coming from Theorem 1.3 are in general rather large, so our new algorithms are unlikely to work very using them. We hope that these bounds can be improved, but in the meantime we note that one way to get a very good bound on the norm is simply to find a Markov basis. When this is possible it is conventional to run the algorithm of Diaconis–Sturmfels, but in Section 2.1 we illustrate that it may in fact be faster to obtain a norm bound form the Markov basis and then apply the bounded-AHT algorithm.

(2) The AHT algorithm of Section 1.1.3 is proven to converge, and in practice the Markov chain is often observed to settle down quite fast. Indeed, in practice it is the latter which will generally be relied upon; people run algorithms until the chain appears to converge. However, there is a critical problem here. Namely, we see in Section 2.2 examples where the chain will *appear* to converge very rapidly, but this "apparent" limit will not be the true limit (the runtime required to achieve true convergence may easily be arranged to exceed the lifespan of the solar system). We hope that this kind of pathological behaviour will be very rare in practice, but at present this seems hard to verify. Our aim in this paper is to get an idea of how long the algorithm should be run in order to be reasonably confident that the "apparent limit" of the chain is in fact the true limit. We are not completely successful in this, partly because our bound on the norm is rather large for practical use (and probably not sharp), and also because passing from the bound in Theorem 1.3 to an estimate on the convergence time needs substantial further work. We think it is interesting and useful to investigate this further. In the meantime, we would encourage people using this type of algorithm to let it run for as long as possible, even after the chain appears to have settled down, to maximise the change of hitting new connected components.

2. Examples

2.1. ***A very simple example.*** Consider the matrix

$$A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{bmatrix}.$$

An integral basis for the kernel of A is then given by $B = \{b, b'\}$ where

$$b = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \quad b' = \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \end{bmatrix}.$$

The fibre containing the vector $[2\ \ 2\ \ 2\ \ 2]^T$ is illustrated in Figure 1, where red arrows (pointing up and to the right) correspond to addition of b, and blue arrows (pointing down and to the right) to addition of b'. Evidently, this fibre is not connected, since the element $[4\ \ 0\ \ 0\ \ 4]^T$ is isolated. Thus if our chain begins anywhere in the large component it will never hit the isolated vertex, and if it begins at the isolated vertex it will remain there. This has practical consequences, since it is common to simply run such a Markov chain until it appears (by eye) to have converged; in this example, convergence will be rapid, but the resulting distribution will not be the expected one (see Section 1.4).

The approach of Diaconis–Sturmfels is to replace the basis B by a larger generating set which makes the fibre connected. The ideal $\mathcal{I}_B$ is generated by $x_1x_3 - x_2^2$ and $x_2x_4 - x_3^2$, and its saturation can be generated by these two polynomials together with the polynomial $x_1x_4 - x_2x_3$, the latter corresponding to the vector $[1\ \ -1\ \ -1\ \ 1]^T$. Clearly one can step from $[3\ \ 1\ \ 1\ \ 3]^T$ to $[4\ \ 0\ \ 0\ \ 4]^T$ by addition of this new vector, so the fibre is indeed connected by this new generating set for the integral kernel of A.

Our approach is to allow the chain to step briefly outside the fibre while it hunts for vectors with nonnegative entries. As long as we allow two negative steps the fibre will become connected, as we can step from $[3\ \ 1\ \ 1\ \ 3]^T$ to $[4\ \ 0\ \ 0\ \ 4]^T$ via $[4\ \ -1\ \ 2\ \ 3]^T$ or $[3\ \ 2\ \ -1\ \ 4]^T$; one sees easily that the norm is 2. Let us compute the bound resulting from Theorem 1.3: we have $\beta = 2$ and $n = 2$, so our bound is 8. Thus if we use the bound from the theorem we should allow 8 negative steps; it is clear that this will be sufficient to connect the fibre, but also that this bound is not optimal.

Remark 2.1. This is an opportune moment to illustrate the necessity of allowing $\epsilon_i = 0$ in Definition 1.2. In the above example the norm is 2. However, there does not exist a generating set G for the saturation of $\mathcal{I}_B$ with respect to $x_1 \cdots x_r$, with the properties that

(1) Every element of G is a pure difference binomial;

(2) Every $g \in G$ can be written in the form

$$g = \sum_{i=1}^{2} \epsilon_i m_i (x^{b_i^+} - x^{b_i^-}). \tag{2.1.1}$$

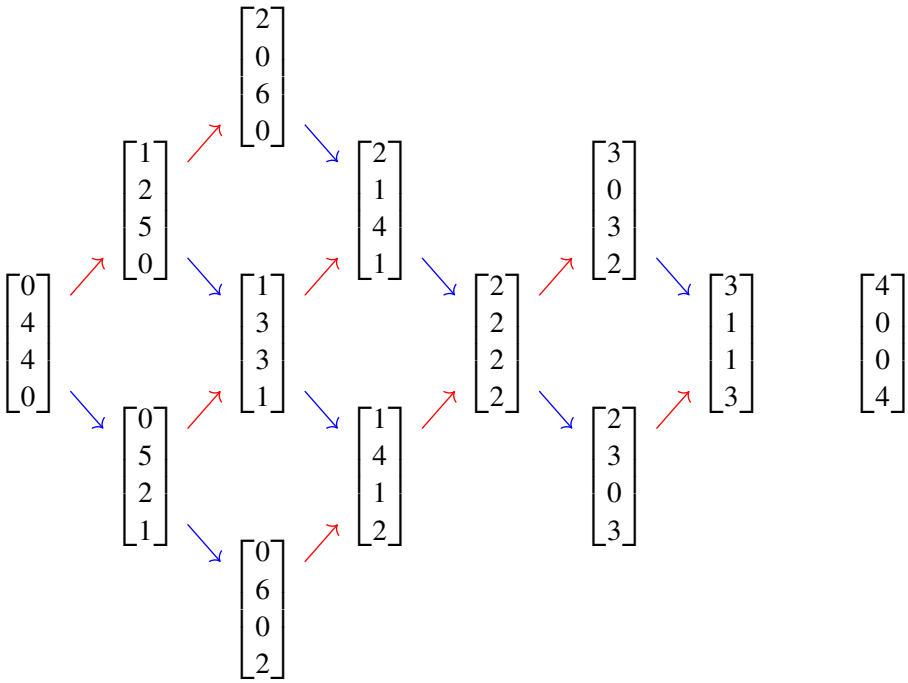


Figure 1. A (nonconnected) fibre.

where the $\epsilon_i \in \{-1, 1\}$, the m_i are Laurent monomials, and the b_i are elements of B (this differs from Definition 1.2 exactly by requiring $\epsilon_i \in \{-1, 1\}$).

To see this, suppose that G is such a generating set. Since $N = 2$ and $\#G = 2$, elementary considerations yield that every element of G is of one of the following forms:

(1) px_b where p is a polynomial consisting of two monomials with coefficients in ± 1;

(2) $px_{b'}$ where p is a polynomial consisting of two monomials with coefficients in ± 1;

(3) $m(x_1x_4 - x_2x_4)$ where m is a monomial.

We know that x_b lies in the ideal generated by G; translating into vectors, this means that b can be written as a linear combination $b = ab + a'b'$ with a, a' integer vectors whose entries sum together to an even number. This is evidently impossible.

2.2. ***Families where the fibres are arbitrarily badly connected.*** Consider the 1×3 matrix $A = [1 \;\; 1 \;\; 1]$, and write e_i for the i-th standard basis vector in $\mathbb{Z}^3$. Let $u = e_2$. Then the fibre $\mathcal{F}(u) = \{e_1, e_2, e_3\}$. For a positive integer n, choose the basis

$$B_n = \left\{ \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ n \\ 1-n \end{bmatrix} \right\}$$

of the kernel of A. Then the fibre consists of two connected components, namely $\{e_2, e_3\}$ and $\{e_1\}$. Moreover, to step between the connected components requires $(n-1)$ consecutive negative steps. Thus for every positive integer M and every real number λ there exists an integer n such that the algorithm of Aoki, Hara and Takemura presented in Section 1.1.3 applied to the above basis B_n will appear to converge immediately, but will take M steps before the probability of hitting the other connected component rises above any given positive threshold. This issue may be well-known, but this particular example appears to be new.

This example is quite artificial, as the fibre is essentially simple, but we have made a poor choice of generating set B_n. We can also construct a slightly less artificial example of the same phenomenon, by generalising the example in Section 2.1. For an integer $n \geq 2$, let

$$A_n = \begin{bmatrix} 1 & 2 & \cdots & n-1 & n \\ n & n-1 & \cdots & 2 & 1 \end{bmatrix},$$

and consider the basis of the integral kernel given by

$$B_n = \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ -2 \\ 1 \end{bmatrix} \right\},$$

where we denote the elements of B_n by $b_2, \ldots, b_{n-1}$ in the given order. Then the fibre of $[2 \ \cdots \ 2]^T$ contains the vector $v = [n \ \ 0 \ \cdots \ \ 0 \ \ n]^T$. This vector v is at least $n-2$ steps distant from any other point in the fibre; more precisely, if $c_1, \ldots, c_r \in \pm B_n$ are such that

$$v + \sum_{i=1}^{r} c_i \in \mathcal{F}(v),$$

then either $r \geq n-2$ or $v + \sum_{i=1}^{r} c_i = v$ (the bound $n-2$ is in fact sharp). We leave the elementary verification to the interested reader. Again we see that, though the algorithm of Section 1.1.3 (and variants) may appear to converge rapidly, there are connected components which take an arbitrarily long time to hit.

2.3. ***The no-three-factor-interaction model.*** This model is described in detail (in particular, its statistical interpretation) in [Aoki et al. 2012]. It depends on a choice of three positive integers I, J and K; we will often take $I = J = K$ for simplicity. The matrix A is then an $(IJ + JK + KI) \times IJK$ matrix, described in a slightly complicated way. Define Id_I to be the $I \times I$ identity matrix, and 1_I to be a row vector of length I with all entries equal to 1. Then

$$A = \begin{bmatrix} Id_I \otimes Id_J \otimes 1_K \\ Id_I \otimes 1_J \otimes Id_K \\ 1_I \otimes Id_J \otimes Id_K \end{bmatrix},$$

where $\otimes$ represents the Kroneker product of matrices.

Hara et al. [2012] numerically tested their algorithm 1.1.3 on the no-three-factor-interaction model in the cases $I = J = K = 3$, 5, and 10. In the case $I = 3$ the saturation can be computed by Gröbner basis techniques, but seems presently out of reach $I = 5$, and worse for $I = 10$. In each case they compute a basis for the integral kernel, then run numerical tests of their algorithm for several values of the Poisson parameter λ, and also occasionally replacing the Poisson with a different distribution (we are not completely clear on how they chose these parameters and distributions). In the case $I = 3$ they compare their results to those obtained using a saturated basis, and observe that the Markov chains coming from their algorithm converge similarly to those coming from a saturated basis (though for $\lambda = 50$ the convergence is rather slow).

For $I = 10$ their algorithm does not converge well, but for $I = 5$ it appears to converge fairly rapidly. As throughout this paper, the question we are interested in is whether this apparent convergence can be trusted, or is it possible that there is some connected component of the fibre which their chain has never hit? Of course, their algorithm will find every component with probability 1 if allowed to run for unlimited time, but there is no a-priori reason to assume that the time required for this will be in any way comparable to the time required for the chain to appear to settle down.

To try to get a handle on this, let us compute our upper bound on the number of negative steps required to walk between components (the "distance between" connected components of the fibre). Using SAGE we compute the smith normal form of the 75×125 matrix A, obtaining an integral basis B with $n = 64$ elements. The largest absolute value of an entry in B is $\beta = 1$. This leads to an upper bound on the norm by

$$N' = n^n \beta^{n-1} = 64^{64} \approx 3.9 \times 10^{115}. \tag{2.3.1}$$

Now, in this example Aoki, Hara and Takemura replace the Poisson distribution with a geometric distribution (for reasons which are unclear to us), and try parameters $p = 0.1$, 0.5. The proportion of steps in their algorithm which will exceed N' in length is then so small that it is likely never to occur before the sun runs cold. This means that *if* the bound N' were to be close to the true norm, then this algorithm will in practice never converge to the correct solution. In practice, our bound on the norm is surely very far from sharp, but we gave this example to illustrate the difficulty in guaranteeing convergence (despite the fact that the algorithm might appear to the human eye to have converged).

3. Computational experiments

3.1. ***The very simple example.*** For the example in Section 2.1 we implemented four algorithms:

(1) The algorithm of [Diaconis and Sturmfels 1998] using the Markov basis described in Section 2.1 (we refer to this algorithm as *DS*);

(2) The algorithm of Aoki, Hara and Takemura described in Section 1.1.3 (referred to as *AHT*);

(3) The bounded-AHT algorithm described in Section 1.3.1, generating vectors uniformly at random of L^∞-length up to some integer at least the norm;

(4) The stepping-out algorithm of Section 1.3.2,

so that we could compare their results. We considered the fibre containing the vector $[10, 10, 10, 10]^T$ which has 211 elements, as this small example allowed us to run many simulations to get reasonably accurate timings of the algorithms.

We use the Kolmogorov Smirnov statistic to decide how well a chain has converged. For a chain of length n in the fibre $\mathcal{F} := \mathcal{F}([10, 10, 10, 10]^T)$, a perfectly uniform distribution would sample each point $n/211$ times. Given a function $d\colon \mathcal{F} \to \mathbb{Z}$ with $\sum_{v\in\mathcal{F}} d(v) = n$ we define

$$\mathrm{KS}(d) = \max_{v\in\mathcal{F}} |d(v) - \frac{n}{211}|, \tag{3.1.1}$$

so a larger value of $\mathrm{KS}(d)$ indicates that d is further from being uniform.

There are two subtleties to comparing the outputs of the algorithms:

(1) The AHT, bounded-AHT and stepping-out algorithms have parameters that can be tuned: the mean of the Poisson distribution for AHT and the bound on the norm used in the latter two (even when the norm is known, as in this example, it is not obvious that using it as the bound will yield the best convergence). To work around this, we will tune the parameters of all three algorithms to try to get the best performance out of each for our example.

(2) When comparing runtimes, counting the number of steps in the chain is not a very good measure. Each step in AHT requires repeated sampling from a Poisson distribution, and steps in the stepping-out algorithm can involve a number of substeps outside the fibre. Because of this, we will also compare the actual runtimes, though this is then sensitive to implementation issues.

We produce a chain of $n = 211,000$ samples, so that each site expects 1000 samples. Table 1 compares the Kolmogorov Smirnov statistic and runtime for the four algorithms, making optimised choices of parameters for the AHT, bounded-AHT and stepping-out algorithms. Table 2, left, shows how the Kolmogorov Smirnov statistic and runtime for the bounded-AHT algorithm vary with the bound used, and Table 2, right, shows the same for the stepping-out algorithm.

algorithm	DS	AHT	bounded-AHT	stepping-out
KS statistic (*lower is better*)	309	210.9	162.8	270.5
runtime (s) (*lower is better*)	174	161.7	98.3	156.1
optimised parameter	-	$\lambda = 2$	$N = 3$	$N = 4$

Table 1. Comparison of algorithms (averaged over 20 runs).

Norm bound	2	3	8
KS statistic	177.3	162.8	201.4
runtime (s)	98.8	98.3	96.0

Norm bound	2	4	8
KS statistic	360.8	270.5	305.0
runtime (s)	147.6	156.1	174.3

Table 2. Comparison of norm bounds for bounded-AHT (left) and stepping-out (right), averaged over 20 runs.

We make a number of comments on these results; all come with the serious caveat that this is only a small, simple example.

(1) The improvement obtained by using the bounded-AHT algorithm in place of the original DS algorithm is quite substantial; both the KS statistic and runtime are close to being halved. This suggests that the bounded-AHT algorithm is worth investigating even in cases where a Markov basis can be computed.

(2) While bounded-AHT performs best with a norm bound $N = 3$, its performance with the bound of $N = 8$ coming from Theorem 1.3 is still better than any of the other algorithms.

(3) The worst performance is achieved by the DS algorithm (using a Markov basis), perhaps somewhat surprisingly. Even though this algorithm should explore the boundary of the fibre in a more efficient way, it probably looses out by exploring the interior less efficiently.

3.2. ***The no-three-factor-interaction model.*** Here we took $I = J = K = 5$, as this beyond the range where the saturation can currently be computed, and hence it is interesting to investigate other approaches to sampling. We implemented the stepping-out algorithm described in Section 1.3.2 for this example. Now, with the given norm bound of order 10^{115} it is clear that this algorithm will not work well. However, we remain optimistic that bounds on the norm can be improved, so it seems interesting to investigate how the runtime of the algorithm depends on the given bound. We do this in a very crude way; we simply measure the proportion of steps in the algorithm which take place within the fibre (as opposed to searching for new components outside the fibre). We interpret this as giving a very rough idea of how much slower the algorithm of Section 1.3.2 will be compared to what could be done if one had a Markov basis. The results were as follows.

(1) For a fixed fibre, when the norm bound is *large* compared to the diameter of the fibre, the runtime seems to be very roughly linear in the given bound on the norm.

(2) For a fixed fibre, when the norm bound is *small* compared to the diameter of the fibre, the runtime seems to be relatively insensitive to the size of the bound.

In other words, this might be interpreted as suggesting that the algorithm of Section 1.3.2 will work reasonably well when the norm bound is not too large compared to the diameter of the fibre.

We did not implement the bounded-AHT algorithm here; for formal reasons it is clear that it must perform slightly better than AHT, but the size of the norm bound also makes it clear that the difference will be entirely imperceptible for any practical computation.

4. Proof of the main results

4.1. ***Proof of Theorem 1.3.*** Let $B = \{b_1, \dots, b_n\}$ be a set of vectors in $\mathbb{Z}^r$. Following the notation of (1.1.2), we write

$$f_i^+ = x^{b_i^+}, \quad f_i^- = x^{b_i^-}, \quad f_i = f_i^+ - f_i^-$$

in the ring $R = \mathbb{Z}[x_1, \dots, x_r]$. Then $\mathcal{I}_B = (f_1, \dots, f_n) \subseteq R$, and our goal is to bound how far the saturation

$$\mathrm{Sat}_{x_1 \cdots x_r} \mathcal{I}_B = \{a \in R : \exists m > 0 : a(x_1 \cdots x_r)^m \in \mathcal{I}_B\} \tag{4.1.1}$$

can be from $\mathcal{I}_B$.

Definition 4.1. A *monomial* in R is an element of the form $\prod_{i=1}^r x_i^{m_i}$ with $m_i \in \mathbb{Z}_{\geq 0}$. A *pure binomial* in R is an element of the form $m_1 - m_2$ where the m_i are monomials. An ideal $I \subseteq R$ is called *pure binomial* if it admits a generating set consisting of pure binomials; evidently, $\mathcal{I}_B$ is a pure binomial ideal.

Lemma 4.2 [Herzog et al. 2018, Proposition 3.18]. *The saturation of $\mathcal{I}_B$ with respect to $x_1 \cdots x_r$ is also a pure binomial ideal.*

Definition 4.3. Given pure binomials $f = f^+ - f^-$ and $g = g^+ - g^-$, we define the *subtraction polynomial* (again a pure binomial)

$$S(f, g) = g^+ f + f^- g = f^+ g^+ - f^- g^-.$$

If $f, g \in \mathcal{I}_B$ then clearly $S(f, g)$ lies in $\mathcal{I}_B$.

We make the unsurprising notational conventions that $--=+$, $+-=-+=-$ and $++=+$; thus we interpret $f^{--} = f^+$, which is less usual, but makes for efficient and hopefully comprehensible notation in what follows.

Definition 4.4. Let $\epsilon\colon \{1, \dots, n\} \to \{+, -\}$, and let $t\colon \{1, \dots, n\} \to \mathbb{N}$. Define

$$S(\epsilon, t) = \prod_{i=1}^n (f_i^{\epsilon(i)})^{t(i)} - \prod_{i=1}^n (f_i^{-\epsilon(i)})^{t(i)} \in \mathcal{I}_B, \tag{4.1.2}$$

(here we use our convention that $--=+$ when we write $f_i^{-\epsilon(i)}$).

Lemma 4.5. *Let P be a pure binomial in $\mathcal{I}_B$. Then there exist ϵ, t, and monomials m and n such that*

$$nP = mS(\epsilon, t).$$

Proof. For the purposes of the proof, we will simplify notation by assuming that for every $b_i \in B$, the element $-b_i$ also lies in B.

Let $P \in \mathcal{I}_B$ be a pure binomial. Write $P = \sum_{j=1}^k m_j f_{i_j}$, where the m_j are monomials. We can and do assume that k is chosen minimal, and we proceed by induction on k. The case $k = 1$ is trivial.

Up to harmless sign changes, there exists a j_0 such that $m_{j_0} f_{i_{j_0}}^+ = P^+$. Reordering, we may assume that $j_0 = 1$, so

$$P - m_1 f_{i_1} = \sum_{j=2}^k m_j f_{i_j}$$

is again a pure difference binomial. By the induction hypothesis there exist monomials m and n and vectors ϵ, t with

$$m \sum_{j=2}^k m_j f_{i_j} = nS(\epsilon, t).$$

Write $S(\epsilon, t) = S^+ - S^-$. Then

$$mP = nS^+ - nS^- + m_1 f_{i_1}^+ - m_1 f_{i_1}^-.$$

Since this is a binomial, up to signs we may assume without loss of generality that $nS^- = m_1 f_{i_1}^+$. We can then write

$$f_{i_1}^+ mP = n\left(f_{i_1}^+ S^+ - f_{i_1}^- S^-\right) = nS'$$

where S' is an iterated subtraction binomial of the f_i. □

Theorem 4.6. *There exist a positive integer M, functions $\epsilon_1, \dots, \epsilon_M$ and $t_1, \dots, t_M$ as in Definition 4.4, and monomials $m_1, \dots, m_M \in R$, such that*

(1) *for all $1 \le j \le M$ we have $m_j \mid S(\epsilon_j, t_j)$;*

(2)

$$\mathrm{Sat}_{x_1 \cdots x_r} \mathcal{I}_B = \left(\frac{S(\epsilon_j, t_j)}{m_j} : 1 \le j \le M \right).$$

Proof. Combine Lemma 4.2 and Lemma 4.5. □

Given $t\colon \{1, \dots, n\} \to \mathbb{N}$ we define the L^1-length of t to be the sum of its values. To prove Theorem 1.3 it suffices to show that we can choose each of the vectors t_j in Theorem 4.6 to have L^1-length bounded by $N = n^n \beta^{n-1}$, where β is the maximum of the absolute values of entries of vectors in B; compare (1.2.2)). Given vectors ϵ of signs and t of natural numbers as in Definition 4.4, observe that the power of x_j dividing $S(t, \epsilon)$ is given by

$$\min\left(\sum_{i=1}^n t(i) \operatorname{ord}_{x_j} f_i^{\epsilon(i)}, \sum_{i=1}^n t(i) \operatorname{ord}_{x_j} f_i^{-\epsilon(i)} \right); \tag{4.1.3}$$

here $\operatorname{ord}_x f$ denotes the largest power of x which divides f. We say the *minimum in (4.1.3) is achieved on the + side* if

$$\sum_{i=1}^n t(i) \operatorname{ord}_{x_j} f_i^{\epsilon(i)} \le \sum_{i=1}^n t(i) \operatorname{ord}_{x_j} f_i^{-\epsilon(i)},$$

and we say the *minimum in (4.1.3) is achieved on the − side* if

$$\sum_{i=1}^n t(i) \operatorname{ord}_{x_j} f_i^{\epsilon(i)} \ge \sum_{i=1}^n t(i) \operatorname{ord}_{x_j} f_i^{-\epsilon(i)}.$$

Definition 4.7. Given $\epsilon\colon \{1, \dots, n\} \to \{+, -\}$ and $\delta\colon \{1, \dots, r\} \to \{+, -\}$, we define

$$T_{\epsilon,\delta} = \{t \in \mathbb{N}^n : \forall 1 \le i \le r, \text{ the minimum in (4.1.3) is achieved on the } \delta(i) \text{ side}\}.$$

This set $T_{\epsilon,\delta}$ is a rational polyhedral cone in $\mathbb{N}^n$, and for fixed ϵ we have

$$\bigcup_{\delta} T_{\epsilon,\delta} = \mathbb{N}^n. \tag{4.1.4}$$

Given $t \in T_{\epsilon,\delta}$, we write

$$\varphi_t = \frac{S(\epsilon, t)}{\prod_{j=1}^{r} x_j^{\sum_{i=1}^{n} t(i) \operatorname{ord}_{x_j} f_i^{\epsilon(i)\delta(i)}}}, \tag{4.1.5}$$

which we write as a difference of monomials $\varphi_t = \varphi_t^+ - \varphi_t^-$ in the usual way. From the definition of $T_{\epsilon,\delta}$ we see that $\varphi_t \in R$, i.e. all exponents of the x_i are nonnegative.

Lemma 4.8. *Fix ϵ and δ as above, and let $t, t_1, \ldots, t_a \in T_{\epsilon,\delta}$ such that $t = t_1 + \cdots + t_a$. Then*

$$\varphi_t \in (\varphi_{t_1}, \ldots, \varphi_{t_a}) \subseteq R.$$

Proof. Elementary manipulations yield

$$\varphi_t = \prod_{\alpha=1}^{a} \varphi_{t_\alpha}^+ - \prod_{\alpha=1}^{a} \varphi_{t_\alpha}^- = S(\cdots S(S(\varphi_{t_1}, \varphi_{t_2})\varphi_{t_3}) \cdots \varphi_{t_a}). \qquad \square$$

Theorem 4.9. *For each ϵ and each δ, choose a generating set $\tau_{\epsilon,\delta}$ for the cone $T_{\epsilon,\delta}$. Then*

$$\bigcup_{\epsilon,\delta} \{\varphi_t : t \in \tau_{\epsilon,\delta}\} \tag{4.1.6}$$

is a generating set for $\operatorname{Sat}_{x_1 \cdots x_r} \mathcal{I}_B$.

Proof. Let $t \in \mathbb{N}^n$, then $S(\epsilon, t) \in \mathcal{I}_B$, and $\varphi_t \in R$, hence by definition of the saturation we see that $\varphi_t \in \operatorname{Sat}_{x_1 \cdots x_r} \mathcal{I}_B$. Conversely, Theorem 4.6 tells us that the φ_t generate $\operatorname{Sat}_{x_1 \cdots x_r} \mathcal{I}_B$ as t ranges over $\mathbb{N}^n$. We must justify why it suffices to consider only t ranging over the set in (4.1.6). Fixing ϵ, we note that every $t \in \mathbb{N}^r$ lies in some $T_{\epsilon,\delta}$ by (4.1.4), and then by Lemma 4.8 it suffices to range over elements of a generating set for $T_{\epsilon,\delta}$. $\square$

Fixing ϵ and δ, it remains to show that $T_{\epsilon,\delta}$ can be generated by vectors of length bounded by $N = n^n \beta^{n-1}$. First, we have an elementary lemma.

Lemma 4.10. *Let $v_1, \ldots, v_a \in \mathbb{N}^n$, and let C be the intersection of $\mathbb{N}^n$ with the* rational *cone spanned by the v_i. Then C is generated by*

$$C \cap \left\{ \sum_{i=1}^{a} \lambda_i v_i : \lambda_i \in [0, 1) \right\} \cup \{v_1, \ldots, v_a\}.$$

Observe that the faces of $T_{\epsilon,\delta}$ are defined by the equations

$$\sum_{i=1}^{n} t(i) \operatorname{ord}_{x_j} f_i^{\epsilon(i)} = \sum_{i=1}^{n} t(i) \operatorname{ord}_{x_j} f_i^{-\epsilon(i)}; \tag{4.1.7}$$

thus the extremal rays of $T_{\epsilon,\delta}$ are obtained by solving $n-1$ equations of the form (4.1.7). Let β be the maximum of the absolute values of the $\operatorname{ord}_{x_j} f_i = b_{i,j}$ as i and j vary. Observing that for any given i and j at least one of $\operatorname{ord}_{x_j} f_i^{\epsilon(i)}$ and $\operatorname{ord}_{x_j} f_i^{-\epsilon(i)}$ is equal to zero, we can rearrange these equations to the form $\sum_i \beta_{i,j,\epsilon} t(i) = 0$ with $\beta_{i,j,\epsilon}$ an integer of absolute value not greater than β. By Siegel's lemma,

the L^1-length of such a (nonzero) solution is then bounded above by $(n\beta)^{n-1}$. From Lemma 4.10, and cutting into simplicial cones, we see that $T_{\epsilon,\delta}$ can be generated by vectors of length at most $N = n^n\beta^{n-1}$, concluding the proof.

4.1.1. *Detailed description of the $T_{\epsilon,\delta}$.* The $T_{\epsilon,\delta}$ for fixed ϵ and varying δ resemble the cones of a complete polyhedral fan in $\mathbb{N}^n$ in the sense of [Fulton 1993]. More precisely, they form a collection of polyhedral cones in $\mathbb{N}^n$ which cover $\mathbb{N}^n$ and such that the intersection of any two cones is a face of both. However, they do not quite form a fan, for two reasons:

(1) it can happen that $T_{\epsilon,\delta} = T_{\epsilon,\delta'}$ for $\delta \neq \delta'$;

(2) the intersection of two $T_{\epsilon,\delta}$ does not necessarily occur among the $T_{\epsilon,\delta}$.

However, by throwing away duplicate cones and appending the intersections of cones, one does obtain a complete fan. The corresponding toric variety is then a toric blowup of affine space $\mathbb{A}^n$.

In the example of Section 2.1 we have $n = 2$ and $r = 4$, and so the fans can readily be drawn for each ϵ. We use this to illustrate the above comments in Table 3.

To explain this in more detail for the case $\epsilon = (+, +)$ (i.e. ϵ taking the constant value $+$), the fan is obtained by subdividing $\mathbb{N}^2$ along the rays through $(1, 2)$ and $(2, 1)$. For each δ we describe in Table 4 the fan $T_{(+,+),\delta}$.

Our main work in this proof is to bound the lengths of generators for these cones. The general bound we obtained is $N = n^n\beta^{n-1}$, which in this case yields $N = 8$. However, just from studying the last row of Table 3 we see that we can take a generating set to be

$$(1, 0),\ (0, 1),\ (1, 2),\ (2, 1),\ (1, 1). \tag{4.1.8}$$

In particular, we obtain a bound on the norm of 3. This is very close to sharp, as we saw in Section 2.1 that the norm is 2. This illustrates that a major source of nonsharpness in our bound is the application of Siegel's lemma below. It seems reasonable to hope that one can find better bounds on the norm by studying Hilbert bases for the cones $T_{\epsilon,\delta}$.

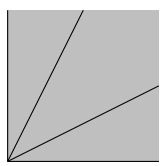
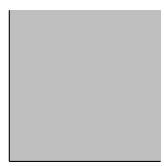
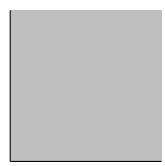
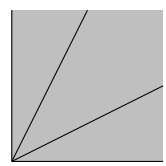
ϵ	$(+, +)$	$(+, -)$	$(-, +)$	$(-, -)$
rays generating fan	(1,2), (2,1)	-	-	(1,2), (2,1)
fan				
generating set for all cones	(0,1), (1,0), (1,2), (2,1), (1,1)	(1,0), (0,1)	(1,0), (0,1)	(0,1), (1,0), (1,2), (2,1), (1,1)

Table 3. The fans generated by the $T_{\epsilon,\delta}$.

$\delta(1)\delta(2)\delta(3)\delta(4)$	cone
$++++$	$\{(0,0)\}$
$+++-$	$\{(0,0)\}$
$++-+$	$\{(0,0)\}$
$++--$	$\{(0,0)\}$
$+-++$	$\{(0,0)\}$
$+-+-$	$\langle(0,1)\rangle$
$+--+$	$\{(0,0)\}$
$+---$	$\{(0,0)\}$
$-+++$	$\{(0,0)\}$
$-++-$	$\langle(1,2),(2,1)\rangle$
$-+-+$	$\langle(1,0)\rangle$
$-+--$	$\langle(0,1),(1,2)\rangle$
$--++$	$\{(0,0)\}$
$--+-$	$\langle(1,0),(2,1)\rangle$
$---+$	$\{(0,0)\}$
$----$	$\{(0,0)\}$

Table 4. The cones $T_{(++),\delta}$.

4.2. ***Proof of Proposition 1.4.*** Let G be a generating set for the saturation as in Definition 1.2. Each $g \in G$ is a pure difference binomial, say $g = x^{c^+} - x^{c^-}$ with $c^+, c^- \in \mathbb{N}^r$, and can be written in the form

$$g = \sum_{i=1}^{N} \epsilon_i m_i f_{j_i},$$

with $\epsilon_i \in \{1, 0, -1\}$, m_i monomials, and f_j as in Section 4.1. Writing $c = c^+ - c^-$, it suffices (by Theorem 1.1) to show that c can be written as $c = \sum_{i=1}^{n} a_i b_i$ with $\sum_{i=1}^{n} |a_i| \le N$.

We wish to prove this by induction on N, but this makes no sense as N is the norm. Instead we rephrase things slightly so that induction makes sense, resulting in the following lemma.

Lemma 4.11. *Let M be a positive integer, and suppose that the expression*

$$\sum_{i=1}^{M} \epsilon_i m_i f_{j_i}, \tag{4.2.1}$$

is a pure binomial $x^{c^+} - x^{c^-}$, where $\epsilon_i \in \{1, -1\}$, and the m_i are monomials. Then there exist integers $a_1, \ldots, a_n$ with $\sum_{i=1}^{n} |a_i| \le M$ and $c^+ - c^- = \sum_{i=1}^{n} a_i b_i$.

It is clear that the lemma (applied with $M = N$) implies Proposition 1.4, so it only remains to verify the lemma.

Proof. For a warmup we treat first the case $M = 1$. Then

$$x^{c^+} - x^{c^-} = \pm m(x^{b^+_{j_1}} - x^{b^-_{j_1}}) = \pm(x^{d+b^+_{j_1}} - x^{d+b^-_{j_1}})$$

where we write $m = x^d$ for some $d \in \mathbb{N}^r$. Hence

$$c^+ - c^- = \pm((d + b_{j_1}^+) - (d + b_{j_1}^-)) = \pm b_{j_1}$$

as required.

We prove the general case by induction on M. First, up to changing some signs, observe that we can reorder the terms in the expression (4.2.1) so that $m_M f_{j_M}^+ = x^{c^+}$, hence we can assume that $\sum_{i=1}^{M-1} \epsilon_i m_i f_{j_i}$ is also a pure binomial, say

$$\sum_{i=1}^{M-1} \epsilon_i m_i f_{j_i} = x^{c'^+} - x^{c'^-}.$$

By our induction hypothesis we can write $c'^+ - c'^- = \sum_{i=1}^n a_i' b_i$ with $\sum_{i=1}^n |a_i'| \leq M - 1$. Then

$$\sum_{i=1}^{M-1} \epsilon_i m_i f_{j_i} = x^{c'^+} - x^{c'^-} = x^{c^+} - x^{c^-} - \epsilon_M m_M (x^{b_{j_M}^+} - x^{b_{j_M}^-}),$$

and we can (again changing some signs, without loss of generality) assume that $\epsilon_M = +1$ and that $x^{c^-} = m_M x^{b_{j_M}^+}$. Writing $m_M = x^d$, we see

- $x^{c^+} = x^{c'^+}$, so $c^+ = c'^+$;
- $x^{c^-} = x^{d + b_{j_M}^+}$, so $c^- = d + b_{j_M}^+$;
- $x^{c'^-} = m_M x^{b_{j_M}^-} = x^{d + b_{j_M}^-}$, so $c'^- = d + b_{j_M}^-$.

Putting these together we see

$$c^+ - c^- = c'^+ - c^- = (c'^+ - c'^-) + (b_{j_M}^+ - b_{j_M}^-) = (c'^+ - c'^-) + b_{j_M},$$

from which the result is immediate. □

Acknowledgements

This work owes its existence to a seminar on algebraic statistics organised in Leiden in the Autumn of 2018 by Garnet Akeyr, Rianne de Heide, and Rosa Winter. I am very grateful to them for organising it, for the many expert speakers who took the time to patiently explain basic ideas of probability and statistics to us, and especially to the determined participants who survived to the end, and offered very useful comments on a presentation of the results contained here.

I am also grateful to the referees for encouraging me to improve the paper, and to the editors for their graceful handling of the apoptosis and apocatastasis of the journal.

References

[Aoki et al. 2012] S. Aoki, H. Hara, and A. Takemura, *Markov bases in algebraic statistics*, Springer, 2012.

[Bigatti et al. 1999] A. M. Bigatti, R. La Scala, and L. Robbiano, "Computing toric ideals", *J. Symbolic Comput.* **27**:4 (1999), 351–365.

[Bosma et al. 1997] W. Bosma, J. Cannon, and C. Playoust, "The Magma algebra system, I: The user language", *J. Symbolic Comput.* **24**:3-4 (1997), 235–265.

[Charalambous et al. 2014] H. Charalambous, A. Thoma, and M. Vladoiu, "Markov complexity of monomial curves", *J. Algebra* **417** (2014), 391–411.

[Decker et al. 2019] W. Decker, G.-M. Greuel, G. Pfister, and H. Schönemann, "Singular 4-1-2 — a computer algebra system for polynomial computations", 2019, available at http://www.singular.uni-kl.de.

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

[Fulton 1993] W. Fulton, *Introduction to toric varieties*, Annals of Mathematics Studies **131**, Princeton University Press, 1993.

[Gross and Petrović 2013] E. Gross and S. Petrović, "Combinatorial degree bound for toric ideals of hypergraphs", *Internat. J. Algebra Comput.* **23**:6 (2013), 1503–1520.

[Hara et al. 2012] H. Hara, S. Aoki, and A. Takemura, "Running Markov chain without Markov basis", pp. 45–62 in *Harmony of Gröbner bases and the modern industrial society*, edited by T. Hibi, World Sci., Hackensack, NJ, 2012.

[Haws et al. 2014] D. Haws, A. Martín del Campo, A. Takemura, and R. Yoshida, "Markov degree of the three-state toric homogeneous Markov chain model", *Beitr. Algebra Geom.* **55**:1 (2014), 161–188.

[Hemmecke and Windisch 2015] R. Hemmecke and T. Windisch, "On the connectivity of fiber graphs", *J. Algebr. Stat.* **6**:1 (2015), 24–45.

[Hemmecke et al. 2001–] R. Hemmecke, R. Hemmecke, M. Köppe, P. Malkin, and M. Walter, "4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces", 2001–, available at https://4ti2.github.io/.

[Herzog et al. 2018] J. Herzog, T. Hibi, and H. Ohsugi, *Binomial ideals*, Graduate Texts in Mathematics **279**, Springer, 2018.

[Koyama et al. 2015] T. Koyama, M. Ogawa, and A. Takemura, "Markov degree of configurations defined by fibers of a configuration", *J. Algebr. Stat.* **6**:2 (2015), 80–107.

[Santos and Sturmfels 2003] F. Santos and B. Sturmfels, "Higher Lawrence configurations", *J. Combin. Theory Ser. A* **103**:1 (2003), 151–164.

[Sturmfels 1996] B. Sturmfels, *Gröbner bases and convex polytopes*, University Lecture Series **8**, American Mathematical Society, Providence, RI, 1996.

[Windisch 2016] T. Windisch, "Rapid mixing and Markov bases", *SIAM J. Discrete Math.* **30**:4 (2016), 2130–2145.

[Windisch 2019] T. Windisch, "The fiber dimension of a graph", *Discrete Math.* **342**:1 (2019), 168–177.

Received 2019-09-17. Revised 2020-05-26. Accepted 2020-07-06.

DAVID HOLMES: holmesdst@math.leidenuniv.nl
, *Mathematisch Instituut Leiden, Niels Bohrweg 1, 2333 CA Leiden, Netherlands*

Algebraic Statistics
Vol. 11, No. 2, 2020
https://doi.org/10.2140/astat.2020.11.189

ALGEBRAIC ANALYSIS OF ROTATION DATA

MICHAEL F. ADAMER, ANDRÁS C. LŐRINCZ,
ANNA-LAURA SATTELBERGER AND BERND STURMFELS

We develop algebraic tools for statistical inference from samples of rotation matrices. This rests on the theory of D-modules in algebraic analysis. Noncommutative Gröbner bases are used to design numerical algorithms for maximum likelihood estimation, building on the holonomic gradient method of Sei, Shibata, Takemura, Ohara, and Takayama. We study the Fisher model for sampling from rotation matrices, and we apply our algorithms to data from the applied sciences. On the theoretical side, we generalize the underlying equivariant D-modules from SO(3) to arbitrary Lie groups. For compact groups, our D-ideals encode the normalizing constant of the Fisher model.

1. Introduction

Many of the multivariate functions that arise in statistical inference are holonomic. Being holonomic roughly means that the function is annihilated by a system of linear partial differential operators with polynomial coefficients whose solution space is finite-dimensional. Such a system of PDEs can be written as a left ideal in the Weyl algebra, or D-ideal, for short. This representation allows for the application of algebraic geometry and algebraic analysis, including the use of computational tools, such as Gröbner bases in the Weyl algebra [28; 30].

This important connection between statistics and algebraic analysis was first observed by a group of scholars in Japan, and it led to their development of the *holonomic gradient method* (HGM) and the *holonomic gradient descent* (HGD). We refer to [10; 16; 31] and to further references given therein. The point of departure for the present article is the work of Sei et al. [29], who developed HGD for data sampled from the rotation group SO(n), and the article of Koyama [16] who undertook a study of the associated equivariant D-module.

The statistical model we examine in this article is the Fisher distribution on the group of rotations, defined in (1) and (2). The aim of maximum likelihood estimation (MLE) is to learn the model parameters Θ that best explain a given data set. In our case, the MLE problem is difficult because there is no simple formula for evaluating the normalizing constant of the distribution. This is where algebraic analysis comes in. The normalizing constant is a holonomic function of the model parameters, and we can use its holonomic D-ideal to derive an efficient numerical scheme for solving the maximum likelihood estimation problem.

ORCID: Adamer 0000-0002-8996-7167 / Sattelberger 0000-0002-2308-4070.
MSC2020: 14F10, 33F10, 62F10, 62H11, 62R01, 90C90.
Keywords: algebraic analysis, Fisher model, maximum likelihood estimation, directional statistics, holonomic gradient method, Fourier transform.

The present article addresses diverse audiences and it offers multiple points of entry. First, we give an introduction to the use of D-module methods in statistics. We focus on data in the group of rotations in 3-space, and we advance both the theory and the practice of this application. Readers with an interest in the applied sciences may start in Section 5, as that displays a panorama of occurrences of rotation data in the real world. Experts in representation theory can jump straight to Section 6, where our approach is developed for arbitrary Lie groups. For such readers, the particular group SO(3) is merely an example.

Our presentation is organized as follows. Section 2 is purely expository. Here, we introduce the Fisher model, and we express its log-likelihood function in terms of the sufficient statistics of the given data. These are obtained from the singular value decomposition of the sample mean. In Section 3, we turn to algebraic analysis. We review the holonomic D-ideal in [29] that annihilates the normalizing constant of the Fisher distribution, and we derive its associated Pfaffian system. Passing to $n \geq 3$, we next study the D-ideals on $\mathrm{SO}(n)$ given in [16]. First new results can be found in Theorem 3.4 and in Propositions 3.5 and 3.6.

Section 4 is concerned with numerical algorithms for maximum likelihood estimation. We develop and compare holonomic gradient ascent (HGA), holonomic BFGS (H-BFGS), and a holonomic Newton method. We implemented these methods in the language R. Section 5 highlights how samples of rotation matrices arise in the sciences and engineering. Topics range from materials science and geology to astronomy and biomechanics. We apply holonomic methods to data from the literature, and we discuss both successes and challenges.

The D-ideal of the normalizing constant is of independent interest from the perspective of representation theory, as it generalizes naturally to other Lie groups. The development of that theory is our main new mathematical contribution. This work is presented in Section 6.

2. The Fisher model for random rotations

In this section, we introduce the Fisher model on the rotation group, building on [29]. The group $\mathrm{SO}(3)$ consists of all real 3×3 matrices Y that satisfy $Y^{\mathrm{t}}Y = \mathrm{Id}_3$ and $\det(Y) = 1$. This is a smooth algebraic variety of dimension 3 in the 9-dimensional space $\mathbb{R}^{3\times 3}$. See [5] for a study of rotation groups from the perspective of combinatorics and algebraic geometry.

The Haar measure on $\mathrm{SO}(3)$ is the unique probability measure μ that is invariant under the group action. The *Fisher model* is a family of probability distributions on $\mathrm{SO}(3)$ that is parametrized by 3×3 matrices Θ. For a fixed Θ, the density of the *Fisher distribution* equals

$$f_\Theta(Y) \;=\; \frac{1}{c(\Theta)} \cdot \exp(\mathrm{tr}(\Theta^{\mathrm{t}} \cdot Y)) \qquad \text{for all} \quad Y \in \mathrm{SO}(3). \tag{1}$$

This is the density with respect to the Haar measure μ. The denominator is the *normalizing constant*. It is chosen such that $\int_{\mathrm{SO}(3)} f_\Theta(Y)\mu(dY) = 1$. This requirement is equivalent to

$$c(\Theta) \;=\; \int_{\mathrm{SO}(3)} \exp(\mathrm{tr}(\Theta^{\mathrm{t}} \cdot Y))\mu(dY). \tag{2}$$

This function is the Fourier–Laplace transform of the Haar measure μ; see Remark 6.6. The Fisher model is an exponential family. It is one of the simplest statistical models on SO(3). The task at hand is the accurate numerical evaluation of the integral (2) for given Θ in $\mathbb{R}^{3\times 3}$. We begin with the observation that, since integration is against the Haar measure, the function (2) is invariant under multiplying Θ on the left or right by a rotation matrix:

$$c(Q \cdot \Theta \cdot R) \;=\; c(\Theta) \qquad \text{for all} \;\; Q, R \in \mathrm{SO}(3).$$

In order to evaluate (2), we can therefore restrict to the case of diagonal matrices. Namely, given any 3×3 matrix Θ, we first compute its *sign-preserving singular value decomposition*

$$\Theta \;=\; Q \cdot \mathrm{diag}(x_1, x_2, x_3) \cdot R.$$

Sign-preserving means that $Q,\ R \in \mathrm{SO}(3)$ and $|x_1| \geq x_2 \geq x_3 \geq 0$. For nonsingular Θ this implies that $x_1 > 0$ whenever $\det(\Theta) > 0$ and $x_1 < 0$ otherwise.

The normalizing constant $c(\Theta)$ is the following function of the three singular values:

$$\tilde{c}(x_1, x_2, x_3) \;:=\; c(\mathrm{diag}(x_1, x_2, x_3)) \;=\; \int_{\mathrm{SO}(3)} \exp(x_1 y_{11} + x_2 y_{22} + x_3 y_{33}) \mu(dY). \tag{3}$$

The statistical problem we address in this paper is parameter estimation for the Fisher model. Suppose we are given a finite sample $\{Y_1, Y_2, \ldots, Y_N\}$ from the rotation group SO(3). We refer to Figure 1 for a concrete example. Our aim is to find the parameter matrix Θ whose Fisher distribution f_Θ best

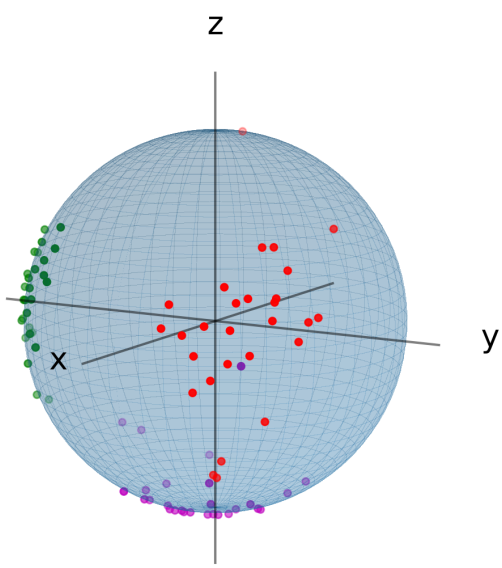


Figure 1. A dataset of 28 rotations from a study in vectorcardiography [7], a method in medical imaging. Each point represents the rotation of the unit standard vector on the x-axis (depicted in red color), the y-axis (green), and the z-axis (purple). This sample from the group SO(3) will be analyzed in Section 5.1.

explains the data. We work in the classical framework of likelihood inference, i.e. we seek to compute the maximum likelihood estimate (MLE) for the given data $\{Y_1, Y_2, \ldots, Y_N\}$. By definition, the MLE is the 3×3 parameter matrix $\hat{\Theta}$ which maximizes the log-likelihood function. Thus, we must solve an optimization problem.

From our data we obtain the *sample mean* $\bar{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k$. Of course, the sample mean $\bar{Y}$ is generally not a rotation matrix anymore. We next compute the sign-preserving singular value decomposition of the sample mean, i.e., we determine $Q, R \in \mathrm{SO}(3)$ such that

$$\bar{Y} \;=\; Q \cdot \mathrm{diag}(g_1, g_2, g_2) \cdot R.$$

The signed singular values g_1, g_2, g_3 together with Q and R are sufficient statistics for the Fisher model. The sample $\{Y_1, \ldots, Y_N\}$ enters the log-likelihood function only via g_1, g_2, g_3.

Lemma 2.1 [29, Lemma 2]. The log-likelihood function for the given sample from $\mathrm{SO}(3)$ is

$$\ell : \mathbb{R}^3 \longrightarrow \mathbb{R}, \quad x \;\mapsto\; x_1 g_1 + x_2 g_2 + x_3 g_3 - \log(\tilde{c}(x_1, x_2, x_3)). \tag{4}$$

If $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ is the maximizer of the function ℓ, then the matrix $\hat{\Theta} = Q\,\mathrm{diag}(\hat{x}_1, \hat{x}_2, \hat{x}_3)R$ is the MLE of the Fisher model (1) of the sample $\{Y_1, \ldots, Y_N\}$ from the rotation group $\mathrm{SO}(3)$.

Lemma 2.1 says that we need to maximize the function (4) in order to compute the MLE in the Fisher model. We note that a local maximum is already a global one since (4) is a strictly concave function. The maximum is attained at a unique point in $\mathbb{R}^3$. We shall compute this point using tools from algebraic analysis that are discussed in the next section.

Remark 2.2. The singular values of the sample mean $\bar{Y}$ are bounded from above and below, namely $1 \geq |g_1| \geq g_2 \geq g_3 \geq 0$. If g_3 is close to 1, i.e., the average of the rotation matrices is almost a rotation matrix, then the data is typically concentrated about a preferred rotation. In this case the normalizing constant becomes very large and MLE on $\mathrm{SO}(3)$ is numerically intractable; see also Remark 4.3. However, due to the small spread of the data around a point in $\mathrm{SO}(3)$, a matrix valued Gaussian model on $\mathbb{R}^3$ is an accurate approximation.

3. Holonomic representation

We shall represent the normalizing constant $\tilde{c}$ by a system of linear differential equations it satisfies. This is known as the holonomic representation of this function. We work in the *Weyl algebra* D and in the *rational Weyl algebra* R with complex coefficients:

$$D \;=\; \mathbb{C}[x_1, x_2, x_3]\langle \partial_1, \partial_2, \partial_3 \rangle \qquad \text{and} \qquad R \;=\; \mathbb{C}(x_1, x_2, x_3)\langle \partial_1, \partial_2, \partial_3 \rangle.$$

We refer to [28; 30] for basics on these two noncommutative algebras of linear partial differential operators with polynomial and rational function coefficients, respectively. In order to stress the number of variables, we sometimes write D_3 instead of D and R_3 instead of R. By a *D-ideal* we mean a left ideal in D, and by an *R-ideal* a left ideal in R. The use of these algebras in statistical inference was pioneered by

Takemura, Takayama, and their collaborators [10; 16; 17; 29; 31]. We begin with an exposition of their results from [29].

The normalizing constant $\tilde{c}$ is closely related to the hypergeometric function ${}_0F_1$ of a matrix argument. In [29], annihilating differential operators of $\tilde{c}$ are derived from

$$H_i \;=\; \partial_i^2 - 1 + \sum_{j\neq i} \frac{1}{x_i^2 - x_j^2}(x_i\partial_i - x_j\partial_j) \qquad \text{for } i = 1,2,3. \tag{5}$$

These in turn can be obtained from Muirhead's differential operators in [21, Theorem 7.5.6] by a change of variables. In the notation of [28], we have $H_i \bullet \tilde{c} = 0$ for $i = 1,2,3$. Written in the more familiar form of linear PDEs, this says

$$\frac{\partial^2\tilde{c}}{\partial x_i^2} + \sum_{j\neq i} \frac{1}{x_i^2 - x_j^2}\left(x_i\frac{\partial\tilde{c}}{\partial x_i} - x_j\frac{\partial\tilde{c}}{\partial x_j}\right) \;=\; \tilde{c} \qquad \text{for } i = 1,2,3.$$

Note that the operators H_i are elements in the rational Weyl algebra R. Clearing the denominators, we obtain elements G_i in the Weyl algebra D that annihilate $\tilde{c}$, namely

$$G_i \;=\; \prod_{j\neq i}(x_i^2 - x_j^2)\cdot H_i. \tag{6}$$

By [29, Theorem 1], the following three additional differential operators in D annihilate $\tilde{c}$:

$$L_{ij} \;:=\; (x_i^2 - x_j^2)\partial_i\partial_j \;-\; (x_j\partial_i - x_i\partial_j) \;-\; (x_i^2 - x_j^2)\partial_k. \tag{7}$$

Here the indices are chosen to satisfy $1 \le i < j \le 3$ and $\{i,j,k\} = \{1,2,3\}$.

Let us consider the D-ideal that is generated by the six operators in (6) and (7):

$$I \;:=\; \langle G_1, G_2, G_3, L_{12}, L_{13}, L_{23}\rangle. \tag{8}$$

In the rational Weyl algebra, we have $RI = \langle H_1, H_2, H_3, L_{12}, L_{13}, L_{23}\rangle$ as R-ideals. We enter the D-ideal I into the computer algebra system Singular:Plural as follows:

```
ring r = 0,(x1,x2,x3,d1,d2,d3),dp;
def D = Weyl(r); setring D;
poly L12 = (x1^2-x2^2)*d1*d2 - (x2*d1-x1*d2)-(x1^2-x2^2)*d3;
poly L13 = (x1^2-x3^2)*d1*d3 - (x3*d1-x1*d3)-(x1^2-x3^2)*d2;
poly L23 = (x2^2-x3^2)*d2*d3 - (x3*d2-x2*d3)-(x2^2-x3^2)*d1;
poly G1 = (x1^2-x2^2)*(x1^2-x3^2)*d1^2 + (x1^2-x3^2)*(x1*d1-x2*d2)
            + (x1^2-x2^2)*(x1*d1-x3*d3) - (x1^2-x2^2)*(x1^2-x3^2);
poly G2 = (x2^2-x1^2)*(x2^2-x3^2)*d2^2 + (x2^2-x3^2)*(x2*d2-x1*d1)
            + (x2^2-x1^2)*(x2*d2-x3*d3) - (x2^2-x1^2)*(x2^2-x3^2);
poly G3 = (x3^2-x1^2)*(x3^2-x2^2)*d3^2 + (x3^2-x2^2)*(x3*d3-x1*d1)
            + (x3^2-x1^2)*(x3*d3-x2*d2) - (x3^2-x1^2)*(x3^2-x2^2);
ideal I = L12,L13,L23,G1,G2,G3;
```

We can now perform various symbolic computations in the Weyl algebra D. We used the libraries `dmodloc` [1] and `dmod` [18], due to Andres, Levandovskyy, and Martín-Morales. In particular, the following two lines confirm that I is holonomic and its holonomic rank is 4:

```
isHolonomic(I);
holonomicRank(I);
```

The rank statement means algebraically that $\dim_{\mathbb{C}(x_1,x_2,x_3)}(R/RI) = 4$. In terms of analysis, it means that the set of holomorphic solutions to I on a small open ball $\mathcal{U} \subset \mathbb{C}^3$ is a 4-dimensional vector space. Here $\mathcal{U}$ is chosen to be disjoint from the singular locus

$$\mathrm{Sing}(I) \;=\; \bigl\{ x \in \mathbb{C}^3 \,:\, (x_1^2 - x_2^2)(x_1^2 - x_3^2)(x_2^2 - x_3^2) \,=\, 0 \bigr\}. \tag{9}$$

We note that the normalizing constant $\tilde{c} = \tilde{c}(x_1, x_2, x_3)$ is a real analytic function on $\mathbb{R}^3 \backslash \mathrm{Sing}(I)$ that extends to a holomorphic function on all of complex affine space $\mathbb{C}^3$.

Using Gröbner bases in the rational Weyl algebra R, we find that the initial ideal of RI for the degree reverse lexicographic order is generated by the symbols of our six operators:

$$\mathrm{in}(RI) \;=\; \langle\, \partial_1\partial_2\,,\; \partial_1\partial_3\,,\; \partial_2\partial_3\,,\; \partial_1^2\,,\; \partial_2^2\,,\; \partial_3^2 \,\rangle.$$

The set of standard monomials equals $S = \{1, \partial_1, \partial_2, \partial_3\}$. This is a $\mathbb{C}(x_1, x_2, x_3)$-basis for the vector space R/RI. In this situation, we can associate a *Pfaffian system* to the D-ideal I. For the general theory, we refer the reader to [30] and specifically to [28, (23)].

The Pfaffian system is a system of first-order linear differential equations associated to the holonomic function $\tilde{c}$. It consists of three 4×4 matrices P_1, P_2, P_3 whose entries are rational functions in x_1, x_2, x_3. We introduce the column vector $C = (\,\tilde{c},\, \partial_1 \bullet \tilde{c},\, \partial_2 \bullet \tilde{c},\, \partial_3 \bullet \tilde{c}\,)^{\mathrm{t}}$.

Theorem 3.1 [29, Theorem 2]. The Pfaffian system associated to the normalizing constant $\tilde{c}$ of the Fisher distribution (1) consists of the following three vector equations:

$$\partial_i \bullet C \;=\; P_i \cdot C \qquad \text{for } i \,=\, 1, 2, 3, \tag{10}$$

where the matrices $P_1, P_2, P_3 \in \mathbb{C}(x_1, x_2, x_3)^{4\times 4}$ are

$$P_1 \;=\; \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & \frac{x_1(-2x_1^2+x_2^2+x_3^2)}{(x_1^2-x_3^2)(x_1^2-x_2^2)} & \frac{x_2}{x_1^2-x_2^2} & \frac{x_3}{x_1^2-x_3^2} \\ 0 & \frac{x_2}{x_1^2-x_2^2} & \frac{-x_1}{x_1^2-x_2^2} & 1 \\ 0 & \frac{x_3}{x_1^2-x_3^2} & 1 & \frac{-x_1}{x_1^2-x_3^2} \end{pmatrix}, \quad P_2 \;=\; \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & \frac{-x_2}{x_2^2-x_1^2} & \frac{x_1}{x_2^2-x_1^2} & 1 \\ 1 & \frac{x_1}{x_2^2-x_1^2} & \frac{x_2(x_1^2-2x_2^2+x_3^2)}{(x_2^2-x_1^2)(x_2^2-x_3^2)} & \frac{x_3}{x_2^2-x_3^2} \\ 0 & 1 & \frac{x_3}{x_2^2-x_3^2} & \frac{-x_2}{x_2^2-x_3^2} \end{pmatrix},$$

$$\text{and} \quad P_3 \;=\; \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & \frac{-x_3}{x_3^2-x_1^2} & 1 & \frac{x_1}{x_3^2-x_1^2} \\ 0 & 1 & \frac{-x_3}{x_3^2-x_2^2} & \frac{x_2}{x_3^2-x_2^2} \\ 1 & \frac{x_1}{x_3^2-x_1^2} & \frac{x_2}{x_3^2-x_2^2} & \frac{x_3(x_1^2+x_2^2-2x_3^2)}{(x_3^2-x_1^2)(x_3^2-x_2^2)} \end{pmatrix}.$$

We reproduced this Pfaffian system from the operators $G_1, G_2, G_3, L_{12}, L_{13}, L_{23}$ with the Mathematica package `HolonomicFunctions` [15]. This was done by running Gröbner basis computations in the rational Weyl algebra R with the degree reverse lexicographic order. See [28, Example 3.4] for an illustration on how this is done.

The Pfaffian system (10) allows us to recover the ith partial derivative of the normalizing constant as the first coordinate of the column vector $P_i \cdot C$. In symbols we have $\partial_i \bullet \tilde{c} = (P_i \cdot C)_1$. We make extensive use of this fact when computing the MLE in Section 4. In the same vein, we can recover the Hessian of $\tilde{c}$ from the Pfaffian system of $\tilde{c}$ as follows:

$$\begin{aligned} \partial_1^2 \bullet \tilde{c} &= (P_1 \cdot C)_2, \quad \partial_1\partial_2 \bullet \tilde{c} = (P_2 \cdot C)_2, \quad \partial_1\partial_3 \bullet \tilde{c} = (P_3 \cdot C)_2, \\ \partial_2^2 \bullet \tilde{c} &= (P_2 \cdot C)_3, \quad \partial_2\partial_3 \bullet \tilde{c} = (P_3 \cdot C)_3, \quad \partial_3^2 \bullet \tilde{c} = (P_3 \cdot C)_4. \end{aligned} \tag{11}$$

This allows for the use of second order optimization algorithms, see Section 4.

An object of interest—from the algebraic analysis perspective—is the Weyl closure of the D-ideal I. By definition, the *Weyl closure* is the following D-ideal which clearly contains I:

$$W(I) := RI \cap D.$$

In general, it is a challenging problem to compute the Weyl closure of a D-ideal. This computation is reminiscent of finding the radical of a polynomial ideal, which, according to Hilbert's Nullstellensatz, consists of all polynomials that vanish on the complex solutions to the given polynomials. The Weyl closure plays a similar role for holonomic functions. It turns out that computing $W(I)$ is fairly benign for the D-ideal I studied in this section.

Lemma 3.2. *Let I be the holonomic D-ideal in (8). Then the Weyl closure $W(I)$ is generated by I and the one additional operator $x_1\partial_1\partial_3 + x_2\partial_2\partial_3 + x_3\partial_3^2 - x_2\partial_1 - x_1\partial_2 - x_3 + 2\partial_3$.*

Proof. We used the Singular library `dmodloc` [1] to compute the Weyl closure of I. We found that I is not Weyl-closed, i.e., $I \subsetneq W(I)$. Moreover, by Gröbner basis reductions in the Weyl algebra, we find that adding the claimed operator results in a Weyl-closed ideal. □

Following [16; 29], we now consider the Fisher distribution on $\mathrm{SO}(n)$. The normalizing constant $c(\Theta)$ is defined as in (2), with the integral taken over $\mathrm{SO}(n)$ with its Haar measure. Let D_{n^2} be the Weyl algebra whose variables are the entries of the $n \times n$ matrix $\Theta = (t_{ij})$. The corresponding $n \times n$ matrix of differential operators in D_{n^2} is denoted by $\partial = (\partial_{ij})$. The following result was established by Koyama [16], based on earlier work of Sei et al. [29]. We shall prove a more general statement for arbitrary compact Lie groups in Section 6.

Theorem 3.3. *The annihilator of $c(\Theta)$ is the D-ideal J generated by the following operators*:

$$\begin{aligned} d &= 1 - \det(\partial), \quad g_{ij} = \delta_{ij} - \textstyle\sum_{k=1}^n \partial_{ik}\partial_{jk} && \text{for } 1 \le i \le j \le n, \\ P_{ij} &= \textstyle\sum_{k=1}^n \big(t_{ik}\partial_{jk} - t_{jk}\partial_{ik}\big) && \text{for } 1 \le i < j \le n. \end{aligned}$$

Above we omitted half of the equations given in [16, (12)], which is justified by the results in [25, Section 8.7.3]. Also, the operators P_{ij} are induced from left matrix multiplication (as in (26)) rather than right multiplication as in [16, (11)].

A problem that was left open in [16; 29], even for $n = 3$, is the determination of the holonomic rank of J. We now address this by introducing dimensionality reduction via invariant theory. The same approach makes sense for the general definition of J seen in (30).

Let J' be the D-ideal generated by the operators P_{ij}, g_{ij}. This is the analogue of J for the orthogonal group $\mathrm{O}(n)$ in its standard representation in $\mathrm{GL}_n(\mathbb{C})$ (see Section 6). Since $\mathrm{O}(n)$ has two connected components, the corresponding module in Theorem 6.2 is a direct sum of two simple holonomic D_{n^2}-modules. By symmetry, we obtain

$$\operatorname{rank}(J') \;=\; 2 \cdot \operatorname{rank}(J). \tag{12}$$

The ring of $\mathrm{O}(n)$-invariant polynomials on $\mathbb{C}^{n\times n}$ is generated by the $\binom{n+1}{2}$ entries $\{y_{kl}\}_{1\le k\le l\le n}$ of the symmetric matrix $Y = \Theta^{\mathrm{t}} \cdot \Theta$ (see [25, Section 11.2.1]). These matrix entries y_{kl} are algebraically independent quadratic forms in the n^2 unknowns t_{ij}.

We now work in the Weyl algebra $D_{\binom{n+1}{2}}$ with the convention $y_{kl} = y_{lk}$ and $\partial_{kl} = \partial_{lk}$. Let K denote the left ideal in that Weyl algebra which is generated by the operators

$$h_{ij} \;=\; 2^{\delta_{ij}}\, n \cdot \partial_{ij} \;-\; \delta_{ij} \;+\; \sum_{k,\,l=1}^{n} 2^{\delta_{ki}+\delta_{lj}}\, y_{kl} \cdot \partial_{ik}\partial_{jl} \qquad \text{for } 1 \le i \le j \le n. \tag{13}$$

Theorem 3.4. *A holomorphic function is a solution to J' if and only if it is of the form $\Theta \mapsto \phi(y_{ij}(\Theta))$, where ϕ is a solution to K. In particular,* $\operatorname{rank}(K) = 2 \cdot \operatorname{rank}(J)$.

Proof. The Lie algebra operators P_{ij} express left invariance under $\mathrm{SO}(n)$. The fact that every solution to J' is expressible in Y follows from Luna's Theorem [19] (see also [11, Section 6.4]). We note that the determinant $\det(\Theta)$ is an $\mathrm{SO}(n)$-invariant that we may omit, due to the relation $\det(\Theta)^2 = \det Y$. The D-ideal K is the invariant version of J'. The operator h_{ij} is derived from g_{ij} by the chain rule. The result therefore follows from (12). □

As an application of Theorem 3.4, we answer a question left open in [29, Proposition 2].

Proposition 3.5. *For $n = 3$, we have* $\operatorname{rank}(J) = 4$.

Proof. We used the computer algebra system Macaulay2 [9]. Unlike for $\operatorname{rank}(J)$, the calculations for $\operatorname{rank}(K)$ finished, and we found $\operatorname{rank}(K) = 8$. We conclude by Theorem 3.4. □

For arbitrary $n \ge 2$, we define I to be the D-ideal generated by the n operators G_i analogous to (6) and the $\binom{n}{2}$ operators L_{ij} analogous to (7). We saw this D-ideal in (8) for $n = 3$. We now explain how the D_{n^2}-ideal J and the D_n-ideal I are connected. We use the construction of the *restriction ideal*. For the general definition see [28, (13)]. In our case, the construction works as follows. We set $x_i = t_{ii}$ for $i = 1, \ldots, n$ and we write D_n for the corresponding Weyl algebra. Then

$$J_{\mathrm{diag}} \;:=\; (J \,+\, \{t_{ij} \,:\, 1 \le i \ne j \le n\} \cdot D_{n^2}) \;\cap\; D_n \tag{14}$$

is the D_n-ideal obtained by restricting the annihilator of $c(\Theta)$ to the diagonal entries of the matrix Θ. Note that the second summand in (14) is a *right* ideal in the Weyl algebra D_{n^2}.

If $f(\Theta)$ is a function in the n^2 variables t_{ij} that is annihilated by J, then the restriction ideal J_{diag} annihilates the function $f(\text{diag}(x_1, \ldots, x_n))$ in n variables. Therefore, J_{diag} annihilates the restricted normalizing constant $\tilde{c}(x_1, \ldots, x_n)$. The result to be presented next guarantees that the Pfaffian system in Theorem 3.1 is indeed of minimal size.

Proposition 3.6. *The following inclusions hold among holonomic D_n-ideals representing $\tilde{c}$:*

$$I \subseteq J_{\text{diag}} \subseteq W(J_{\text{diag}}) \subseteq \text{ann}_{D_n}(\tilde{c}).$$

Equality holds for $n \leq 3$ in the rightmost inclusion.

Sketch of proof. The proof of [29, Theorem 1] shows that I is contained in J_{diag}. Note that for $n = 3$ the middle inclusion is strict by Lemma 3.2. We have $W(J_{\text{diag}}) \subseteq \text{ann}_{D_n}(\tilde{c})$ because the annihilator of a smooth function such as $\tilde{c}$ is Weyl-closed, by an argument spelled out in [8].

The equality on the right for $n = 3$ is shown by proving $W(I) = \text{ann}_{D_3}(\tilde{c})$. We use the following argument and computations. The Fourier transform $W(I)^{\mathcal{F}}$ is the D-ideal obtained by switching ∂_i and x_i (up to sign). We find that its holonomic rank is 1. We next compute the *holonomic dual* of the module $D_3/W(I)^{\mathcal{F}}$. This is another D_3-module, as defined in [12, Section 2.6]. There is a built-in command for the holonomic dual in Macaulay2 [9]. Another computation, using localization techniques, verifies that both $D_3/W(I)^{\mathcal{F}}$ and its holonomic dual are torsion-free as $\mathbb{C}[x_1, x_2, x_3]$-modules. These facts imply that $D_3/W(I)^{\mathcal{F}}$ is a simple D-module, and hence so is $D_3/W(I)$. From this we conclude that $W(I) = \text{ann}_{D_3}(\tilde{c})$. □

We conjecture that the inclusion on the right is an equality for all positive integers n. Using results from Section 6, we can argue that $W(J_{\text{diag}})^{\mathcal{F}}$ is regular holonomic for any n. It appears that its singular locus is a hyperplane arrangement. The special combinatorial structure encountered in this arrangement gives strong evidence for the conjecture above.

4. Maximum likelihood estimation

We now proceed to computing the maximum of the log-likelihood function of Lemma 2.1 for given datasets. Since the objective function (4) is strictly concave, a local maximum is the global maximizer and attained at a unique point $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3) \in \mathbb{R}^3$. In order to compute $\hat{x}$, we run a number of algorithms, each using the holonomic gradient method. This is based on the results presented in the previous section, especially on Theorem 3.1 and equation (11). These are used to compute the function values, gradients, and Hessians in each iteration. The code for the numerical computations of this section can be found at https://github.com/MikeAdamer/hgm_MLE.

A critical step in running any local optimization method is finding a suitable starting point. As mentioned in Section 3, solutions to the D-ideal I are analytic outside the singular locus $\text{Sing}(I)$. Starting points need to be chosen in $\mathbb{R}^3 \backslash \text{Sing}(I)$. For the Fisher model on SO(3), the singular locus $\text{Sing}(I)$ is the arrangement (9) of six planes through the origin in $\mathbb{R}^3$. This partitions $\mathbb{R}^3$ into 24 distinct chambers. For the algorithms described below, we choose starting points in each of the 24 connected components of $\mathbb{R}^3/\text{Sing}(I)$, and we evaluate the vector C at these points. This initialization can be done

either via the series expansion method of [29, Section 3.2] or, equivalently, using the package `hgm` [31] in the statistical software R, which uses a series expansion in conjunction with HGM.

In this section, we present three optimization methods based on algebraic analysis, building on the methods given in [22]. The holonomic part of the algorithms stems from the basic structure of most optimization schemes. In essence, there are always two steps: a gradient evaluation step and a gradient descent step.

In this paper, we show how HGD is used in the first step for exact evaluation of the gradients. For further details on the second step in a number of optimization schemes we refer to [23]. The simplest algorithm is *holonomic gradient ascent* (HGA). This is a straightforward adaptation of the HGD method in [29]. Second, we introduce a holonomic version of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method [23, Chapter 6, §1]. BFGS is a quasi-Newton method that requires the gradient and the function value as inputs. Both can be calculated directly using (10). This turns BFGS into *holonomic BFGS* (H-BFGS). Our third algorithm is a *holonomic Newton method*. This second-order method exploits the fact that the Hessian is easy to calculate from (11) and that the objective function is strictly concave.

To get started, we need an expression for the gradient of the log-likelihood function ℓ and a holonomic gradient method (HGM) for evaluating that expression. By Lemma 2.1,

$$\nabla \ell(x) \;=\; \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} - \frac{1}{\tilde{c}(x)} \cdot \nabla \tilde{c}(x). \tag{15}$$

Note that $C(x) = (\tilde{c}(x), \nabla \tilde{c}(x))^{\mathrm{t}}$. Hence, our task to evaluate $\nabla \ell$ at any point amounts to evaluating the vector-valued function C at any point. This is where the HGM comes in.

In general, we approximate the function C at a point $x^{(n+1)}$ given its value at a previous point $x^{(n)}$. To this end, a path $x^{(n)} \to x^{(n)} + \delta^{(1)} \to x^{(n)} + \delta^{(2)} \to \cdots \to x^{(n)} + \delta^{(K)} \to x^{(n+1)}$ is chosen, where $\delta^{(1)}, \ldots, \delta^{(K)} \in \mathbb{R}^3$ with $\|\delta^{(m+1)} - \delta^{(m)}\|$ sufficiently small. The linear part of the Taylor series expansion of C at $x^{(n)}$ yields the following approximations:

$$\begin{aligned} C(x^{(n)} + \delta^{(m+1)}) \;&\approx\; C(x^{(n)} + \delta^{(m)}) + \sum_{i=1}^{3} (\delta_i^{(m+1)} - \delta_i^{(m)})\, (\partial_i \bullet C)(x^{(n)} + \delta^{(m)}) \\ &=\; C(x^{(n)} + \delta^{(m)}) + \sum_{i=1}^{3} (\delta_i^{(m+1)} - \delta_i^{(m)})\, P_i \cdot C(x^{(n)} + \delta^{(m)}). \end{aligned} \tag{16}$$

We choose a path consisting of points, separated by intervals of size Δt, on the line segment $x(t) = x^{(n)}(1-t) + x^{(n+1)}t$ with $t \in [0, 1]$. With this notation, (16) becomes

$$C(x((m+1)\Delta t)) \;\approx\; C(x(m\Delta t)) + \textstyle\sum_{i=1}^{3} (x_i^{(n+1)} - x_i^{(n)})\, \Delta t \cdot\; P_i \cdot C(x(m\Delta t)). \tag{17}$$

If we take the limit $\Delta t \to 0$, then the equation above becomes the differential equation

$$\frac{dC(t)}{dt} \;=\; \sum_{i=1}^{3} \frac{\partial x_i}{\partial t} \frac{\partial C}{\partial x_i} \;=\; \sum_{i=1}^{3} \big(x_i^{(n+1)} - x_i^{(n)}\big) P_i \cdot C.$$

This ordinary differential equation can be solved using any numerical ODE solver, e.g., an Euler scheme or Runge–Kutta scheme. This leads to the following algorithm.

Algorithm 1: Holonomic gradient method.

Input: $x^{(n)}$, $x^{(n+1)}$, $C(x^{(n)})$, a Pfaffian system P_1, P_2, P_3
Output: $C(x^{(n+1)})$
1 Set $x(t) = x^{(n)}(1-t) + x^{(n+1)}t$.
2 Let $\frac{dC(t)}{dt} = \sum_{i=1}^{3} \frac{\partial x_i}{\partial t}\frac{\partial C}{\partial x_i} = \sum_{i=1}^{3}(x_i^{(n+1)} - x_i^{(n)})P_i \cdot C$.
3 Numerically integrate line 2 from $t = 0$ to $t = 1$.

We employ Algorithm 1 as a subroutine for the holonomic gradient ascent algorithm, which will be described next. HGA is analogous to other gradient ascent/descent methods, however, with the special feature that the gradients are calculated via the HGM algorithm. A description of the algorithm, adapted for data from SO(3), is outlined below.

Algorithm 2: Holonomic gradient ascent.

Input: Matrices Q and R, singular values g_1, g_2, g_3 and a starting point $x^{(0)} \in \mathbb{R}^3$
Result: A maximum likelihood estimate for the data in the Fisher model (1)
1 Choose a learning rate γ_n.
2 Choose a threshold δ.
3 Evaluate C at the starting point $x^{(0)}$.
4 Evaluate $\nabla\ell$ at the starting point $x^{(0)}$.
5 Set $n = 0$.
6 **while** $\max|\nabla\ell(x^{(n)})| < \delta$ **do**
7 $\quad x^{(n+1)} = x^{(n)} + \gamma_n \nabla\ell(x^{(n)})$.
8 $\quad$ Calculate $C(x^{(n+1)})$ via HGM using Algorithm 1.
9 $\quad$ Calculate $\nabla\ell(x^{(n+1)})$ from $C(x^{(n+1)})$.
10 $\quad$ Set $n = n+1$.
11 **end**
12 Output the vector $x^{(n)} \in \mathbb{R}^3$ as our approximation for $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$.
13 Output the rotation matrix $\hat{\Theta} = Q \cdot x^{(n)} \cdot R$ as our approximation for the MLE.

The given data is a list of rotation matrices $Y_1, \ldots, Y_N$ in SO(3). As explained in Section 2, we encode these in the singular values g_1, g_2, g_3 of the sample mean $\bar{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k$. Thus, the input for HGA consists primarily of just three numbers g_1, g_2, g_3. They are used in the evaluation in the first terms of $\nabla\ell$, as seen in (15). The second term is evaluated by matrix multiplication with P_1, P_2, P_3, as seen in (10). Part of the input are also the matrices Q and R that diagonalize the sample mean $\bar{Y}$. They are needed in the last step to recover $\hat{\Theta}$ from $\hat{x}_1, \hat{x}_2, \hat{x}_3$ as in Lemma 2.1. The HGA algorithm has two

parameters, namely the threshold δ which indicates a termination condition, and the learning rate γ_n. While δ can be chosen freely depending on the desired accuracy, choosing the learning rate can have significant effects on the convergence of the algorithm. In our computations we chose $\gamma_n = 10^{-2}$. This can clearly be improved. However, the standard technique of performing line searches to find a good γ_n is not recommended as evaluating C at a new point is costly.

To employ more advanced methods such as BFGS, and to avoid integrating along a path crossing the singular locus, we use [29, Corollary 1]. This states that the value of the column vector C at a point (x_1, x_2, x_3) can be obtained by fixing (x_1, x_2, x_3) and integrating the following ODE from $t = \epsilon \ll 1$ to $t = 1$. Here C is regarded as a function of the parameter t:

$$C'(t) \;=\; \begin{pmatrix} 0 & x_1 & x_2 & x_3 \\ x_1 & -2/t & x_3 & x_2 \\ x_2 & x_3 & -2/t & x_1 \\ x_3 & x_2 & x_1 & -2/t \end{pmatrix} \cdot C(t). \tag{18}$$

Using this approach for calculating C, we can employ BFGS optimization using HGM as a subroutine to calculate the gradients and function values required as inputs. This also prevents the accumulation of numerical errors as the initial conditions of the ODE are exact. The H-BFGS method achieves faster convergence rates than the simple HGA Algorithm 2.

A final very powerful algorithm for concave (or convex) functions is the Newton method which uses the Hessian matrix. Often, finding the Hessian matrix $\mathbf{H}[\ell(x)]$ of a function is a difficult task. However, using holonomic methods the Hessian is obtained for free via

$$\partial_i \partial_j \bullet \ell \;=\; \frac{1}{\tilde{c}^2} \left(\partial_i \bullet \tilde{c}\right) \left(\partial_j \bullet \tilde{c}\right) - \frac{1}{\tilde{c}} \partial_i \partial_j \bullet \tilde{c},$$

and the relations in (10) and (11). We found that the Newton method,

$$x^{(n+1)} \;=\; x^{(n)} - \mathbf{H}[\ell(x)]^{-1} \cdot \nabla \ell(x),$$

gives the fastest convergence. We refer to this approach as the *Holonomic Newton method*.

We implemented the H-BFGS method in a script in the software R. Interested readers may obtain our implementation at https://github.com/MikeAdamer/hgm_MLE. This code is custom-tailored for rotations in 3-space. The function C is evaluated at the starting point $x^{(0)}$ using the series expansion method that is described in [29, Section 3.2]. Here we truncate the series at order 41.

Example 4.1. We created a synthetic dataset consisting of $N = 500$ rotation matrices. These were sampled from the Fisher distribution with parameter matrix

$$\Theta \;=\; \begin{pmatrix} -1.178 & 0.2804 & 1.037 \\ -0.3825 & 0.9181 & 0.6016 \\ -0.0955 & 0.9037 & 1.695 \end{pmatrix}. \tag{19}$$

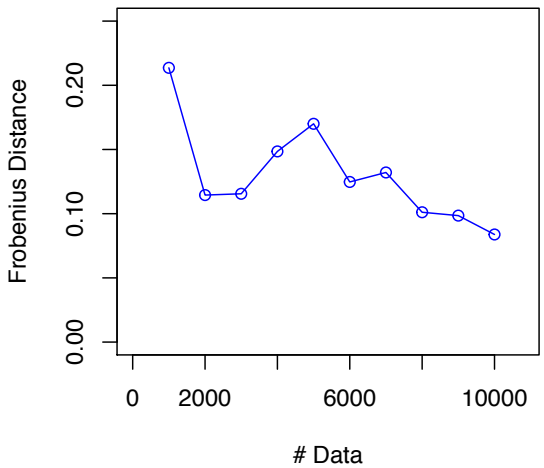


Figure 2. The Frobenius distance of the MLE parameters to the true parameter values. The convergence is relatively slow. This suggests that the MLE problem is not well conditioned.

The sample mean and its sign-preserving singular value decomposition are found to be

$$\bar{Y} = \begin{pmatrix} -0.2262 & 0.1021 & 0.2260 \\ -0.0233 & 0.0611 & 0.2779 \\ -0.0364 & 0.2802 & 0.3529 \end{pmatrix} = Q \cdot \begin{pmatrix} 0.5946 & 0.0000 & 0.0000 \\ 0.0000 & 0.1838 & 0.0000 \\ 0.0000 & 0.0000 & 0.1059 \end{pmatrix} \cdot R,$$

$$\text{with} \quad Q = \begin{pmatrix} -0.4977 & 0.8589 & 0.1211 \\ -0.4518 & -0.1376 & -0.8815 \\ -0.7404 & -0.4934 & 0.4565 \end{pmatrix}, \quad R = \begin{pmatrix} 0.2524 & -0.4808 & -0.8397 \\ -0.9419 & -0.3209 & -0.0993 \\ -0.2217 & 0.8160 & -0.5339 \end{pmatrix}.$$

Running H-BFGS on this input, the MLE is found to be

$$\hat{\Theta} = \begin{pmatrix} -0.8972 & 0.3446 & 0.9682 \\ -0.2392 & 0.7777 & 0.7856 \\ -0.0763 & 0.8664 & 1.616 \end{pmatrix} = Q \cdot \begin{pmatrix} 2.422 & 0.0000 & 0.0000 \\ 0.0000 & 0.7432 & 0.0000 \\ 0.0000 & 0.0000 & -0.3043 \end{pmatrix} \cdot R. \tag{20}$$

While the entries of the MLE $\hat{\Theta}$ have the correct sign and order of magnitude, the actual values are not very close to those in Θ. In order to isolate the effect of the sample size on the MLE, we extended the data to 10000 matrices. In the iterations we recorded the Frobenius distance (FD) from $\hat{\Theta}$ to Θ. Our findings are outlined in Figure 2.

Remark 4.2. The authors in [29] report that the HGD algorithm becomes numerically unstable when it is close to the singular locus of the Pfaffian system. They recommend picking a starting point in the same connected component of $\mathbb{R}^3 \setminus \mathrm{Sing}(I)$ where the MLE is suspected. In contrast, our computations suggest that the output of the HGA does not depend on the connected component which the starting point lies in, when a sufficiently stable numerical integration method (e.g. `lsode` from the R package `deSolve`) is

chosen in Algorithm 1. Therefore, the starting point of the optimization can be chosen arbitrarily, as long as it is close enough to the origin so that the series expansion converges.

Remark 4.3. The sample mean matrix $\bar{Y}$ lies in the convex hull of the rotation group. This convex body, denoted conv(SO(3)), was studied in [27, Section 4.4], and an explicit representation as a spectrahedron was given in [27, Proposition 4.1]. It follows from the theory of orbitopes [27] that the singular values of matrices in conv(SO(3)) are precisely the triples that satisfy $1 \geq |g_1| \geq g_2 \geq g_3 \geq 0$. These inequalities define two polytopes, which are responsible for the facial description of conv(SO(3)) found in [27, Theorem 4.11].

We can think of the MLE as a map from the interior of the orbitope conv(SO(3)) to $\mathbb{R}^3$. Using the singular value decomposition, we restricted this map to the open polytopes given by $1 > |g_1| > g_2 > g_3 > 0$. Note that the coordinates of the vector $\hat{x}$ goes off to infinity as the maximum of $\{g_1, g_2, g_3\}$ approaches 1. This follows from [14, (4.12)], where the analogue for $\mathrm{O}(n)$ was derived. This divergence can cause numerical problems.

In this section, we have turned the earlier results on D-ideals into numerical algorithms. This is just a first step. The success of any local method relies heavily on a clear understanding of the numerical analysis that is relevant for the problem at hand. A future study of condition numbers from the perspective of holonomic representations would be desirable.

5. Rotation data in the sciences

Rotation data arise in any field of science in which the orientation of an object in 3-space is important. Occurrences include a diverse number of research areas such as medical imaging, biomechanics, astronomy, geology, and materials science. In this section, we apply our methods to a prominent dataset of vectorcardiograms and to biomechanical data. We also review previous findings on rotation data in astronomy, geology, and materials science.

5.1. ***Medical imaging.*** One important occurrence of rotational data in the applied sciences stems from medical imaging, and more precisely from vectorcardiography. In that field, the electrical forces generated by the heart are studied and their magnitude and direction are recorded.

The dataset presented in [7] is a famous example of directional data. It contains the orientation of the vectorcardiogram (VC) loop of 98 children aged $2-19$. In particular, the orientation is measured using two different techniques. Both measurements are given in the form of two vectors. The first identifies the VC loop of greatest magnitude and the second is the normal direction to the loop. We add as a third vector the cross product of the magnitude and normal vector to form a right handed set and, therefore, a rotation matrix.

This dataset has been used to exemplify a range of methods in directional statistics, see, e.g., [24]. We applied the optimization methods from Section 4 to the same dataset. In other words, we computed the maximum of the log-likelihood function (4) for the orientations of the VC loop. In order to match our analysis with the results of [24], we only consider the 28 data points of the boys aged $2-10$. A colorful illustration of the action of these 28 rotation matrices on the coordinate axes is shown in Figure 1.

We now proceed to the MLE. The sample mean has the singular value decomposition

$$\bar{Y} = \begin{pmatrix} 0.6868 & 0.5756 & 0.1828 \\ 0.5511 & -0.7372 & -0.0045 \\ 0.1216 & 0.1417 & -0.8630 \end{pmatrix} = Q \cdot \begin{pmatrix} 0.9469 & 0.0000 & 0.0000 \\ 0.0000 & 0.8962 & 0.0000 \\ 0.0000 & 0.0000 & 0.8737 \end{pmatrix} \cdot R, \tag{21}$$

where

$$Q = \begin{pmatrix} 0.6112 & 0.7636 & 0.2079 \\ -0.7498 & 0.4748 & 0.4608 \\ 0.2532 & -0.4376 & 0.8628 \end{pmatrix}, \quad R = \begin{pmatrix} 0.03941 & 0.99324 & -0.1092 \\ 0.81778 & 0.03072 & 0.5747 \\ 0.57418 & -0.11194 & -0.8110 \end{pmatrix}. \tag{22}$$

By forming the matrix product QR we recover the result of [24]. The matrix QR, however, is only one part of the MLE as described in [14]. By using H-BFGS, we can find the full MLE of the Fisher model. We compared H-BFGS to other methods. For that, we estimated x_1, x_2, x_3 with a BFGS optimization of the log-likelihood using the series expansion of the normalizing constant. We then compare the resulting estimate to the output of H-BFGS. The H-BFGS algorithm finds the MLE

$$\begin{aligned} \hat{x}_1 &= 20.072407, \\ \hat{x}_2 &= 12.513841, \\ \hat{x}_3 &= -6.510704, \end{aligned}$$

which corresponds to a log-likelihood of $\hat{\ell} = 3.97299$. The runtime of the algorithm is highly dependent on the number of nonzero terms in the series expansion for $\tilde{c}$. In this calculation, the first 6000 nonzero terms are used and the runtime is about 4 seconds. To improve the runtime one could try to truncate the series at lower order. For further refinement of the MLE one can combine H-BFGS with H-Newton by using the output of H-BFGS as the starting point of H-Newton. The classical BFGS method is not convergent if only the first 6000 nonzero terms are used. Hence, we need to truncate the series expansion at higher order. If we use the first 48000 nonzero terms, then the series expansion BFGS method finds the MLE $\hat{x}_1 = 17.604156$, $\hat{x}_2 = 10.024591$, $\hat{x}_3 = -3.881811$, which gives $\hat{\ell} = 3.96330$. The computation takes about 20 seconds. Hence, the holonomic BFGS outperformed the classical method by finding a better likelihood value in much shorter time.

5.2. ***Biomechanics.*** Rotational data is ubiquitous in the biomedical sciences. A prominent experiment in this area is the human kinematics study of [26]. In this experiment, the rotations of four different upper body parts were tracked while the subject was drilling holes into six different locations of a vertical panel. In [3], this dataset was studied and maximum likelihood and Bayesian point estimates for the orientation of the wrist were obtained and credible regions constructed.

A further experiment concerns the heel orientation of primates. In the experiments, the rotation of the calcaneus bone (the heel) and the cuboid bone, which is horizontally adjacent to the heel and closer to the toes, was measured. A load was applied to three sedentary primates, a human, a chimpanzee, and a baboon and the rotation of their ankle was recorded. While the data is actually a time series, the

simplifying assumption of independent identically distributed data is made in its analysis [4]. We study this dataset which was kindly provided by Melissa Bingham. The sample mean for the human data equals

$$\bar{Y} = \begin{pmatrix} -0.1013 & -0.9127 & -0.3811 \\ 0.3275 & -0.3895 & 0.8535 \\ -0.9335 & -0.0358 & 0.3475 \end{pmatrix} = Q \cdot \begin{pmatrix} 0.9997 & 0.0000 & 0.0000 \\ 0.0000 & 0.9926 & 0.0000 \\ 0.0000 & 0.0000 & 0.9923 \end{pmatrix} \cdot R, \tag{23}$$

with

$$Q = \begin{pmatrix} 0.4771 & 0.8753 & -0.0791 \\ -0.4320 & 0.1552 & -0.8884 \\ -0.7654 & 0.4580 & 0.4521 \end{pmatrix} \text{ and } R = \begin{pmatrix} 0.5248 & -0.2399 & -0.8167 \\ -0.4690 & -0.8822 & -0.0422 \\ -0.7104 & 0.4051 & -0.5754 \end{pmatrix}.$$

We see on the right hand side in (23) that the singular values for this dataset only differ in the third significant figure and the smallest singular value is approximately 1. We found that the normalizing constant gets too large to be computed directly. Indeed, our simulations returned a value error when $\tilde{c} \approx 10^{308}$. This is a serious numerical issue, arising in any MLE algorithm that attempts to directly calculate $\tilde{c}$ when the sample mean is almost a rotation matrix. Singular values close to 1 imply that the samples are concentrated on the unit sphere. One could either use a rotational Maxwell distribution [13] as a local model or the approximation in [4]. The data for the baboon and the chimpanzee show similar traits.

Progress can be made by applying a gauge transform in (18), aimed at scaling the input for H-BFGS. Let λ_0 be the largest eigenvalue of

$$A = \begin{pmatrix} 0 & x_1 & x_2 & x_3 \\ x_1 & 0 & x_3 & x_2 \\ x_2 & x_3 & 0 & x_1 \\ x_3 & x_2 & x_1 & 0 \end{pmatrix}.$$

We can derive an ODE for the function $E = C \cdot \exp(-\lambda_0 t)$ from (18). The function E is guaranteed to have smaller values than C. Furthermore, the ratio $(\partial_i \bullet \tilde{c})/\tilde{c} = C_i/C_0 = D_i/D_0$ is invariant. Despite being able to compute $\log(\tilde{c})$ using the gauge transformation, MLE becomes very unstable due to the numerical accuracy required. Finding the MLE from a random starting point using H-BFGS proved intractable. However, to provide a suitable starting point for H-BFGS, we use the asymptotic formula of [14], which gives $\hat{x}_1 = 5543.102$, $\hat{x}_2 = 3753.025$, $\hat{x}_3 = -3685.298$. We refined this result with H-BFGS and found the MLE $\hat{x}_1 = 5543.106$, $\hat{x}_2 = 3753.078$, $\hat{x}_3 = -3685.242$ corresponding to a log-likelihood of $\hat{\ell} = 10.59342$. Calculating the log-likelihood using HGM and the asymptotic values yields $\hat{\ell} = 10.52366$. Hence, H-BFGS finds a slightly better MLE than the asymptotic formula.

5.3. ***Astronomy, geology, and materials science.*** Astronomical applications of the matrix Fisher model on SO(3) are often concerned with the orbits of near-earth objects [20; 29]. Such objects are comets or asteroids in an elliptic orbit around the sun with the sun in their focus. The data comes as sets of vectors in $\mathbb{R}^3$ taking the sun as the origin. The first vector, X_1, is the perihelion direction, which points to the location on the orbit closest to the sun. The second vector, X_2, is the unit normal to the orbit. Together

with their cross product these vectors form a right handed set. Therefore, they define a rotation matrix. Questions of astronomical interest are whether the perihelion direction is uniformly distributed on the sphere and whether the orbit orientations are uniform on $\mathrm{SO}(3)$. To answer the latter question the Raleigh statistic can be used [20; 29].

Sei et al. [29] studied a dataset of rotations representing 151 comets and 6496 asteroids. They computed maximum likelihood estimates using the holonomic gradient method and also series expansions. The Raleigh statistic for the dataset was calculated and the null hypothesis of a uniform distribution was strongly rejected. Further, the hypothesis of the data originating from a Fisher distribution on a Stiefel manifold was tested against the hypothesis of $\mathrm{SO}(3)$, and the evidence strongly suggested to reject the Stiefel manifold.

Rotations arise in geology and earth sciences in the study of earthquake epicenters [13] and the analysis of plate tectonics [6]. Davis and Titus [6] studied a dataset of the deformation of a shear zone in northern Idaho. However, this was done in the context of invalidating a geology inspired model that had been used previously to explain the shear deformations.

Kagan [13] studied rotational data describing the earthquake focal mechanism orientation. Various models, including the Fisher model, were discussed in this article. However, the Fisher model was dismissed due to the difficulty of normalization for small spread data as discussed in Remark 2.2. The alternative model used in [13] was a rotational Maxwell distribution as a local approximation. Our results offer a chance to revisit the Fisher model.

One important source of rotational data is materials science, where patterns from electron backscatter diffraction (EBSD) are analyzed (see, e.g., [2]). This type of data provides information about the orientation of grains within a material. Crystal orientation has important implications on the properties of polycrystalline materials. One issue with EBSD data is the fact that orientations of the crystals can only be determined within a coset of the crystallographic group the grain belongs to. This is due to the fact that a crystal is a lattice and every lattice comes with certain translational and rotational symmetries. Orientations can only be determined up to the rotational invariance of the lattice. Hence, the data, although giving information about rotations, is strictly speaking not on $\mathrm{SO}(3)$, but on its quotient by a discrete symmetry subgroup. To adapt our analysis, an appropriate parametrization or embedding for such a quotient needs to be found. This, however, is beyond the scope of this paper and is left for future work. Before going to such manifolds, we start with Lie groups.

6. Compact Lie groups

The Fisher model on $\mathrm{SO}(n)$ generalizes naturally to other compact Lie groups. We define the Fisher distribution and the normalizing constant as in (1) and (2), but with integration over the Haar measure on the Lie group. In this section, we introduce these objects and their holonomic representation. In particular, we establish the analogue of Theorem 3.3 for compact Lie groups. This opens up the possibility of applying algebraic analysis to data sampled from manifolds other than $\mathrm{SO}(n)$ provided these have the structure of a group.

Let G be a compact connected Lie group and fix a real representation $\pi : G \to \mathrm{GL}_n(\mathbb{R})$. We can assume that π is injective, i.e., the representation is faithful. We note that any compact Lie group admits a faithful representation [25, Section 8.3.4]. The matrix group $\pi(G) \subset \mathbb{R}^{n\times n}$ is a closed algebraic subvariety (see [25, Section 8.7]). If one starts with a complex representation instead, the situation can be studied in the polynomial ring over $\mathbb{C}$.

For our algebraic approach, the ambient setting is the complex affine space $X := \mathbb{C}^{n\times n}$. The complexification $G_{\mathbb{C}}$ of our group G is a complex connected reductive algebraic group [25, Section 8.7.2]. The extension $\pi : G_{\mathbb{C}} \to X$ is a closed embedding. Its image, the matrix group $\pi(G_{\mathbb{C}})$, is the complex affine variety in X, cut out by the same polynomials as the ones defining $\pi(G)$. We denote by I_G the ideal generated by these polynomials in $\mathbb{C}[X]$. The quotient ring $\mathbb{C}[G] := \mathbb{C}[X]/I_G$ is the ring of polynomial functions on the group $\pi(G_{\mathbb{C}})$.

Let $\mathfrak{g}$ denote the complex Lie algebra of $G_{\mathbb{C}}$. This is the complexification of the real Lie algebra of the given Lie group G. We write $U(\mathfrak{g})$ for the universal enveloping algebra of $\mathfrak{g}$. For any affine variety, one can define the ring of algebraic differential operators on that variety. This is generally a complicated object, but things are quite nice in our case.

Let D_G denote the ring of differential operators on $G_{\mathbb{C}}$. We have natural inclusions

$$\mathfrak{g} \subset U(\mathfrak{g}) \subset D_G \qquad \text{and} \qquad \mathbb{C}[G] \subset D_G.$$

These inclusions exhibit desirable properties. Namely, we have canonical isomorphisms

$$D_G \;\cong\; \mathbb{C}[G] \otimes U(\mathfrak{g}) \;\cong\; U(\mathfrak{g}) \otimes \mathbb{C}[G]. \tag{24}$$

This holds because left (or right) invariant vector fields of $G_{\mathbb{C}}$ trivialize the tangent bundle. Recall that $G_{\mathbb{C}}$ acts on $X = \mathbb{C}^{n\times n}$ by left matrix multiplication via π. Through this action, elements in the Lie algebra $\mathfrak{g}$ induce vector fields on X. This gives an injective map

$$\phi \,:\, U(\mathfrak{g}) \hookrightarrow D_{n^2}. \tag{25}$$

We now proceed to describing the algebra map ϕ explicitly. Fix an arbitrary element $\xi \in \mathfrak{g}$. Let $-M_\xi$ be the $n \times n$ matrix corresponding to ξ via the inclusion $\mathfrak{g} \hookrightarrow \mathfrak{gl}(n)$. The following is the vector field encoding the Lie algebra action of M_ξ on the space $\mathfrak{gl}(n) \simeq \mathbb{C}^{n\times n}$:

$$\phi(\xi) \;=\; \sum_{i,\,j\,=1}^{n} (M_\xi)_{ij} \cdot \sum_{k=1}^{n} t_{jk}\partial_{ik} \quad \in D_{n^2}. \tag{26}$$

Example 6.1. Fix $G = \mathrm{SO}(n)$ and let $\pi : G \to \mathrm{GL}_n(\mathbb{R})$ be the standard representation on $\mathbb{R}^n$. The associated Lie algebra $\mathfrak{g}$ is the space of skew-symmetric $n \times n$ matrices over $\mathbb{C}$. A canonical basis of $\mathfrak{g}$ consists of the rank 2 matrices $e_{ij} - e_{ji}$ for $1 \le i < j \le n$. The operator $P_{ij} \in D_{n^2}$ in Theorem 3.3 is Fourier dual to the vector field (26) if we take $\xi = e_{ji} - e_{ij}$.

As seen in [12, Section 1.3], the morphism of varieties $\pi : G_{\mathbb{C}} \to X$ induces a pushforward functor of D-modules $\pi_+ : \mathrm{Mod}(D_G) \to \mathrm{Mod}(D_{n^2})$ satisfying the following key property.

Theorem 6.2. *If we regard* $\mathbb{C}[G]$ *as a left* D_G*-module, then we have the isomorphism*

$$\pi_+(\mathbb{C}[G]) \;\cong\; D_{n^2}/\langle\, I_G,\, \phi(\mathfrak{g})\,\rangle.$$

In particular, this quotient is a regular holonomic simple D_{n^2}*-module.*

Proof. By (24), we have the following isomorphism of right D_G-modules:

$$\mathbb{C}[G] \;\cong\; \mathbb{C} \otimes_{U(\mathfrak{g})} D_G. \tag{27}$$

On the right, $\mathbb{C}$ denotes the trivial representation of the universal enveloping algebra $U(\mathfrak{g})$.

Let $D_{G\to X} := \mathbb{C}[G] \otimes_{\mathbb{C}[X]} D_X$ denote the transfer bimodule. This is a left D_G-module and a right D_X-module. Since the action of $\mathfrak{g}$ extends to the whole space X, we have $\mathbb{C}[G] \cong \mathbb{C}[X]/I_G$ as $\mathfrak{g}$-modules, and the left $U(\mathfrak{g})$-structure of $D_{G\to X}$ is induced by the Leibniz rule via the map (25) on the second factor. We obtain the isomorphism of bimodules

$$D_{G\to X} \;\cong\; D_X/(I_G \cdot D_X). \tag{28}$$

By (27) and (28), we have the following isomorphisms of right D_X-modules:

$$\begin{aligned}\pi_+(\mathbb{C}[G]) := \mathbb{C}[G] \otimes_{D_G} D_{G\to X} &\;\cong\; (\mathbb{C} \otimes_{U(\mathfrak{g})} D_G) \otimes_{D_G} D_{G\to X} \\ &\;\cong\; \mathbb{C} \otimes_{U(\mathfrak{g})} D_X/(I_G \cdot D_X) \;\cong\; D_X/((I_G + \phi(\mathfrak{g})) \cdot D_X).\end{aligned}$$

The fist claim now follows by switching to left D_X-modules. By Kashiwara's Equivalence Theorem [12, Section 1.6], the module $D_X/\langle I_G, \phi(\mathfrak{g})\rangle$ is regular holonomic and simple. □

Remark 6.3. The assumption that G is compact is not needed in Theorem 6.2. The proof works for any representation $\pi : H \to \mathrm{GL}_n(\mathbb{C})$ of a complex connected algebraic group such that $\pi(H)$ is closed in $\mathbb{C}^{n\times n}$. Such a representation exists for all semisimple groups H. Another natural setting is that of orbits of a compact group G acting linearly on a real vector space, with left-invariant measures used in Corollary 6.5. In our view, the theory of orbitopes [27] should be of interest for statistical inference with data sampled from orbits.

Remark 6.4. Here is a more conceptual argument for Theorem 6.2. The D-module $M = D_{n^2}/\langle I_G,\, \phi(\mathfrak{g})\rangle$ is equivariant and supported on $\pi(G_{\mathbb{C}})$ (see [12, Section 11.5]). By Kashiwara's Equivalence Theorem, it is the pushforward of a coherent equivariant D-module on $G_{\mathbb{C}}$. This is a direct sum of copies of the module $\mathbb{C}[G]$, by the Riemann–Hilbert Correspondence. Hence, M is a direct sum of copies of $\pi_+(\mathbb{C}[G])$. The existence of a unique left-invariant measure on G implies that there is only one such summand in M.

Let μ_π be the distribution on $\mathbb{R}^{n\times n}$ given by integration against the Haar measure on G. The following corollary generalizes [16, Theorem 1] from $\mathrm{SO}(n)$ to other Lie groups G.

Corollary 6.5. *The annihilator in* D_{n^2} *of this distribution equals*

$$\mathrm{ann}_{D_{n^2}}(\mu_\pi) \;=\; \langle\, I_G,\, \phi(\mathfrak{g})\,\rangle.$$

Proof. Since $\mathrm{supp}(\mu_\pi) = \pi(G)$, we have $I_G \subset \mathrm{ann}_{D_{n^2}}(\mu_\pi)$. Since μ_π is a left-invariant distribution, we have also $\phi(\mathfrak{g}) \subset \mathrm{ann}_{D_{n^2}}(\mu_\pi)$. By Theorem 6.2, the D-ideal $\langle I_G, \phi(\mathfrak{g}) \rangle$ is a maximal left ideal in D_{n^2}, since its quotient is simple. It is therefore equal to $\mathrm{ann}_{D_{n^2}}(\mu_\pi)$. □

The following observation establishes the connection to statistics, as in [16, Section 4].

Remark 6.6. The Fourier–Laplace transform of μ_π has a complex analytic continuation to a holomorphic function on $\mathbb{C}^{n\times n}$ by the Paley–Wiener–Schwartz Theorem, namely

$$c(\Theta) \;=\; \int_G \exp(\mathrm{tr}(\Theta^{\mathrm{t}}\pi(Y)))\mu(dY). \tag{29}$$

This is the normalizing constant of the Fisher distribution on the group $\pi(G) \subset \mathrm{GL}_n(\mathbb{R})$. Note that this can be defined for a complex representation $\pi(G) \subset \mathrm{GL}_n(\mathbb{C})$ as well.

The Fourier transform, denoted by $(\bullet)^{\mathcal{F}}$, switches the operators t_{ij} and ∂_{ij} in the Weyl algebra D_{n^2}, with a minus sign involved. We consider the image of the D-ideal in Corollary 6.5 under this automorphism of D_{n^2}. This image is a D-ideal J_π that is defined over $\mathbb{R}$:

$$J_\pi \;=\; \langle I_G, \phi(\mathfrak{g}) \rangle^{\mathcal{F}}. \tag{30}$$

The following result generalizes Theorem 3.3 to compact Lie groups other than $\mathrm{SO}(n)$.

Corollary 6.7. *The D-module D_{n^2}/J_π is simple holonomic and* $\mathrm{ann}_{D_{n^2}}(c(\Theta)) = J_\pi$.

Proof. By Corollary 6.5, Remark 6.6, and the defining property of the Fourier transform, we see that J_π annihilates the integral in (29). The proof concludes by recalling that the Fourier transform induces an auto-equivalence on the category of (holonomic) D_{n^2}-modules. □

We saw in Section 5 that sampling from $\mathrm{SO}(3)$ is ubiquitous in the applied sciences. It would be worthwhile to explore such scenarios also for other matrix groups $\pi(G)$, and to apply holonomic methods to maximum likelihood estimation in their Fisher model. An example of such a model is the complex matrix Fisher distribution for unitary groups [20, Section 13.2.4].

One promising context for data applications is the unitary group in quantum physics. The following example is as an invitation to mathematical physicists to develop this further.

Example 6.8. The compact group $G = \mathrm{SU}(2)$ consists of complex 2×2 matrices of the form

$$\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}, \qquad \text{with } |\alpha|^2 + |\beta|^2 \;=\; 1. \tag{31}$$

Note that G is a double cover of $\mathrm{SO}(3)$. While the odd-dimensional (complex) representations of G descend to real-valued representations of $\mathrm{SO}(3)$, this is not true for the even-dimensional (spin) representations. Consider the standard representation $G \subset \mathbb{C}^{2\times 2}$.

The complexification of the matrix group in (31) is simply the group $\mathrm{SL}_2(\mathbb{C}) \subset \mathbb{C}^{2\times 2}$. The associated

(maximal, holonomic) ideal J_π is generated by four operators:

$$d = \det(\partial) - 1, \qquad h = t_{11}\partial_{11} + t_{12}\partial_{12} - t_{21}\partial_{21} - t_{22}\partial_{22},$$
$$e = t_{21}\partial_{11} + t_{22}\partial_{12}, \qquad f = t_{11}\partial_{21} + t_{12}\partial_{22}.$$

A computation shows that $\operatorname{rank} J_\pi = 2$ and $\operatorname{Sing}(J_\pi) = \{\Theta \in \mathbb{C}^{2\times 2} \mid \det(\Theta) = 0\}$. The Lie algebra operators e, f, h ensure that every holomorphic solution to J_π is SL_2-invariant. By [19], every solution has the form $\Theta \mapsto \phi(\det(\Theta))$, for some analytic function ϕ in a domain of $\mathbb{C}^*$. This is annihilated by d (hence, by J_π) if and only if $\phi(x)$ is annihilated by

$$x\partial^2 + 2\partial - 1 \in D_1.$$

This has only one (up to scaling) entire solution ϕ, with series expansion at $x = 0$ given by

$$\phi(x) = \sum_{n=0}^{\infty} \frac{1}{n! \cdot (n+1)!}\, x^n.$$

By comparing constant terms, we conclude that $c(\Theta) = \phi\big(\det(\Theta)\big)$. It is straightforward to generalize the above considerations to the fundamental representation of the special unitary group $\mathrm{SU}(m)$ for any $m \geq 1$. In that setting, we find that $\operatorname{rank}(J_\pi) = m$.

In conclusion, the D-ideal J_π is an interesting object that deserves further study, not just for the rotation group $\mathrm{SO}(n)$, but for arbitrary Lie groups G. Sections 3 and 6 offer numerous suggestions for future research. For instance, what is the holonomic rank of J_π? Furthermore, it would be desirable to experiment with data sampled from groups G other than $\mathrm{SO}(3)$, so as to broaden the applicability of algebraic analysis in statistical inference.

Acknowledgments

We thank Mathias Drton for helpful discussions on statistics, Ralf Hielscher for discussions on materials science, and Max Pfeffer for improvements in our numerical methods. We are grateful to Nobuki Takayama and his collaborators for many insightful discussions, and to Charles Wang for getting us started on the material for $\mathrm{SO}(n)$.

References

[1] D. Andres, "dmodloc_lib: a Singular:Plural library for localization of algebraic D-modules and applications", available at https://www.singular.uni-kl.de/Manual/4-1-3/sing_661.htm#SEC715.

[2] F. Bachmann, R. Hielscher, P. E. Jupp, W. Pantleon, H. Schaeben, and E. Wegert, "Inferential statistics of electron backscatter diffraction data from within individual crystalline grains", *J. Appl. Crystallogr.* **43** (2010), 1338–1355.

[3] M. A. Bingham, D. J. Nordman, and S. B. Vardeman, "Finite-sample investigation of likelihood and Bayes inference for the symmetric von Mises–Fisher distribution", *Comput. Statist. Data Anal.* **54**:5 (2010), 1317–1327.

[4] M. A. Bingham, D. J. Nordman, and S. B. Vardeman, "Bayes inference for a tractable new class of non-symmetric distributions for 3-dimensional rotations", *J. Agric. Biol. Environ. Stat.* **17**:4 (2012), 527–543.

[5] M. Brandt, J. Bruce, T. Brysiewicz, R. Krone, and E. Robeva, "The degree of SO(n, $\mathbb{C}$)", pp. 229–246 in *Combinatorial algebraic geometry*, edited by G. G. Smith and B. Sturmfels, Fields Inst. Commun. **80**, Fields Inst. Res. Math. Sci., Toronto, ON, 2017.

[6] J. R. Davis and S. J. Titus, "Modern methods of analysis for three-dimensional orientational data", *J. Struct. Geol.* **96** (2017), 65–89.

[7] T. Downs, J. Liebman, and W. Mackay, "Statistical methods for vectorcardiogram orientations", pp. 216–222 in *Proceedings of the XIth International Symposium on Vectorcardiography*, North-Holland, Amsterdam, 1974.

[8] P. Görlach, C. Lehn, and A.-L. Sattelberger, "Algebraic analysis of the hypergeometric function ${}_1F_1$ of a matrix argument", *Beitrğe Algebra Geom.* (published online November 2020).

[9] D. R. Grayson and M. E. Stillman, "Macaulay2, a software system for research in algebraic geometry", available at http://www.math.uiuc.edu/Macaulay2/.

[10] H. Hashiguchi, Y. Numata, N. Takayama, and A. Takemura, "The holonomic gradient method for the distribution function of the largest root of a Wishart matrix", *J. Multivariate Anal.* **117** (2013), 296–312.

[11] P. Heinzner, "Geometric invariant theory on Stein spaces", *Math. Ann.* **289**:4 (1991), 631–662.

[12] R. Hotta, K. Takeuchi, and T. Tanisaki, *D-modules, perverse sheaves, and representation theory*, Progress in Mathematics **236**, Birkhäuser, Boston, 2008.

[13] Y. Y. Kagan, "Double-couple earthquake source: symmetry and rotation", *Geophys. J. Int.* **194**:2 (2013), 1167–1179.

[14] C. G. Khatri and K. V. Mardia, "The von Mises–Fisher matrix distribution in orientation statistics", *J. Roy. Statist. Soc. Ser. B* **39**:1 (1977), 95–106.

[15] C. Koutschan, "HolonomicFunctions: A Mathematica package for dealing with multivariate holonomic functions, including closure properties, summation, and integration", available at https://www3.risc.jku.at/research/combinat/software/ergosum/RISC/HolonomicFunctions.html.

[16] T. Koyama, "The annihilating ideal of the Fisher integral", preprint, 2015. arXiv 1503.05261

[17] T. Koyama, H. Nakayama, K. Nishiyama, and N. Takayama, "The holonomic rank of the Fisher–Bingham system of differential equations", *J. Pure Appl. Algebra* **218**:11 (2014), 2060–2071.

[18] V. Levandovskyy and J. Martin Morales, "dmod_lib: A Singular:Plural library for algorithms for algebraic D-modules", available at https://www.singular.uni-kl.de/Manual/4-1-3/sing_537.htm#SEC591.

[19] D. Luna, "Fonctions différentiables invariantes sous l'opération d'un groupe réductif", *Ann. Inst. Fourier* (*Grenoble*) **26**:1 (1976), 33–49.

[20] K. V. Mardia and P. E. Jupp, *Directional statistics*, Wiley Series in Probability and Statistics **494**, Wiley, Chichester, 2009.

[21] R. J. Muirhead, *Aspects of multivariate statistical theory*, Wiley, New York, 1982.

[22] H. Nakayama, K. Nishiyama, M. Noro, K. Ohara, T. Sei, N. Takayama, and A. Takemura, "Holonomic gradient descent and its application to the Fisher–Bingham integral", *Adv. in Appl. Math.* **47**:3 (2011), 639–658.

[23] J. Nocedal and S. J. Wright, *Numerical optimization*, 2nd ed., Springer, 2006.

[24] M. J. Prentice, "Orientation statistics without parametric assumptions", *J. Roy. Statist. Soc. Ser. B* **48**:2 (1986), 214–222.

[25] C. Procesi, *Lie groups: an approach through invariants and representations*, Springer, 2007.

[26] D. Rancourt, L.-P. Rivest, and J. Asselin, "Using orientation statistics to investigate variations in human kinematics", *J. Roy. Statist. Soc. Ser. C* **49**:1 (2000), 81–94.

[27] R. Sanyal, F. Sottile, and B. Sturmfels, "Orbitopes", *Mathematika* **57**:2 (2011), 275–314.

[28] A.-L. Sattelberger and B. Sturmfels, "D-Modules and Holonomic Functions", preprint, 2019. arXiv 1910.01395

[29] T. Sei, H. Shibata, A. Takemura, K. Ohara, and N. Takayama, "Properties and applications of Fisher distribution on the rotation group", *J. Multivariate Anal.* **116** (2013), 440–455.

[30] N. Takayama, "Gröbner basis for rings of differential operators and applications", pp. 279–344 in *Gröbner bases*, edited by T. Hibi, Springer, 2013.

[31] N. Takayama, T. Koyama, T. Sei, H. Nakayama, and K. Nishiyama, "hgm: an R package for the holonomic gradient method", available at https://cran.r-project.org/web/packages/hgm/hgm.pdf.

Received 2019-12-03. Revised 2020-04-09. Accepted 2020-04-30.

MICHAEL F. ADAMER: michael.adamer@bsse.ethz.ch
D-BSSE, ETH Zürich, 4058 Basel, Switzerland

ANDRÁS C. LŐRINCZ: lorincza@hu-berlin.de
Humboldt-Universität zu Berlin, Institut für Mathematik, 10099 Berlin, Germany

ANNA-LAURA SATTELBERGER: anna-laura.sattelberger@mis.mpg.de
Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany

BERND STURMFELS: bernd@math.berkeley.edu
University of California, Berkeley, CA 94720, United States

and

Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany

Algebraic Statistics
Vol. 11, No. 2, 2020
https://doi.org/10.2140/astat.2020.11.213

COMPATIBILITY OF DISTRIBUTIONS IN PROBABILISTIC MODELS: AN ALGEBRAIC FRAME AND SOME CHARACTERIZATIONS

LUIGI BURIGANA AND MICHELE VICOVARO

A probabilistic model may be formed of distinct distributional assumptions, and these may specify admissible distributions on distinct (not necessarily disjoint) subsets of the whole set of random variables of concern in the model. Such distributions on subsets of variables are said to be mutually compatible if there exists a distribution on the whole set of variables that precisely subsumes all of them. In Section 2 of this paper, an algebraic frame for this compatibility concept is constructed, by first observing that all marginal and/or conditional distributions (also called "probability kernels") that are implicit in a global distribution form a lattice, and then by highlighting the properties of useful operations that are internal to this algebraic structure. In Sections 3, 4, and 5, characterizations of the concept of compatibility are presented; first a characterization that depends only on set-theoretic relations between the variables involved in the distributions under judgment; then characterizations that are applicable only to pairs of candidate distributions; and then a characterization that is applicable to any set of candidate distributions when the variables involved in each of these are exhaustive of the set of variables in the model. Lastly, in Section 6, different categories of models are mentioned (a model of classical statistics, a corresponding hierarchical Bayesian model, Bayesian networks, Markov random fields, and the Gibbs sampler) to illustrate why the compatibility problem may have different levels of saliency and solutions in different kinds of probabilistic models.

1. Introduction

Several of the probabilistic models used in statistics and in other areas of applied probability are presented in modular form. By this, we mean that a model may be defined by a number of assumptions $A_1, \ldots, A_m$ concerning the distributions acting on definite subsets $X_1, \ldots, X_m$ of the total set $T = \{T_1, \ldots, T_n\}$ of the elementary random variables of concern in the model. Some of these subsets may be disjoint, and others may overlap with one another. Any assumption A_i may specify a single distribution $p_i(X_i)$ or, more typically, it may fix only a constraint on an unknown $p_i(X_i)$ such that it in fact specifies a class of admissible distributions for X_i. Furthermore, each distribution $p_i(X_i)$ imposed or allowed for by an assumption A_i may be a marginal distribution; alternatively, it may be a conditional distribution that expresses how some of the variables in the set X_i are expected to be stochastically influenced by some other variables within the same X_i. Assumptions that specify or constrain the local conditional distributions are characteristic of hierarchical Bayesian models, Bayesian networks, Markov random fields, and probabilistic graphical models in general (Lauritzen, 1996; Koller and Friedman, 2009).

Burigana is the corresponding author.
MSC2020: 62H10, 62H22, 62R01, 60E99.
Keywords: probability kernel, conditional distribution, compatibility, lattice, graphical model.

When considering a model that is presented in modular form, the following question naturally arises. Suppose that $p_1(X_1), \ldots, p_m(X_m)$ are local distributions specified (or allowed for) by the assumptions constituting the model. Are we assured that there exists a global distribution $p(T)$ that acts on the total set of variables and faithfully "assembles" these local distributions, in the sense that each $p_i(X_i)$ can be deduced from $p(T)$ through marginalization and/or conditioning? This is known as the *compatibility problem* for distributional assumptions (Berti, Dreassi and Rigo, 2014, p. 191). Compatibility is an essential requirement for the consistency and plausibility of a model as a whole. Indeed, if the local distributions $p_1(X_1), \ldots, p_m(X_m)$ that comply with these assumptions were not mutually compatible, then analyses guided by the model (so far as these concern the whole set T of variables) would be disqualified as efforts towards a non-existing target, as there would be no $p(T)$ consistent with all $p_1(X_1), \ldots, p_m(X_m)$. The compatibility problem thus conceived has been the subject of systematic research over the past three decades (Arnold, Castillo and Sarabia, 1999, 2001). Interest in this problem is particularly related to the study of so-called "conditionally specified statistical models", as the difficulty of testing compatibility greatly increases when the local distributions under judgment are in conditional form and act on sets of variables that overlap with one another.

With this study, we intend to contribute to the discussion of the compatibility requirement by setting this concept within an algebraic frame and presenting characterizations of it, some of which are taken and reformulated from the literature, and others we believe are new. The algebraic frame is defined in Section 2, and relies on the lattice structure possessed by the set of all marginal and conditional distributions that are deducible from a full joint distribution $p(T)$. The characterizations are presented in the next three sections; we examine a characterization that depends only on set-theoretic relations between the variables in the distributions under consideration (Section 3); characterizations limited to pairs of conditional distributions (Section 4); and a characterization that is applicable to any collection of conditional distributions such that the variables involved in each distribution exhaust the total set T (Section 5). Finally, in Section 6, we comment on simple examples to illustrate that in probabilistic models of different kinds, the compatibility problem may attain different saliency and may require different arguments for its solution.

The main reason for characterizing the frame of this study as an "algebraic" one is that our principal analyses will be conducted by working on structures that are lattices, as defined in abstract algebra, and by discussing relations and operations of algebraic character definable within those structures. In particular, this aspect will become apparent in Sections 2 and 3. Research on the compatibility problem, however, has also produced studies that can be categorized as "algebraic" for a complementary reason, that is, the mathematical tools used in them are of a kind familiar to contemporary algebraic statistics, such as analytical and computational tools from the theory of polynomials and algebraic geometry. Selected references to these studies will be presented in Sections 4 and 5 of our paper.

2. An algebraic view of variable pairs and probability kernels

In dealing with any probabilistic model, we assume that the elementary random quantities (individual data, parameters, hyper-parameters, etc.) involved in the model are exhaustively enumerated in a set

$T = \{T_1, \ldots, T_n\}$. The word *variable* is used here for any subset of T, including the whole T (the full variable), the empty set $\varnothing$ (the empty variable), any singleton $\{T_i\}$ (an elementary variable, also denoted by T_i), and arbitrary multiple variables (that is, sets of two or more elements of T). As they are understood as subsets of T, arbitrary variables may be compared or combined in set-theoretical manner, so that if X and Y are variables, then $X \cup Y$, $X \cap Y$, or $X \setminus Y$ are also variables.

For each elementary variable T_i, we assume that a measure space $(T_i^\circ, \mathcal{B}_i, \mu_i)$ is specified, in which T_i° is a standard set that includes all possible values of T_i (e.g., T_i° could be the real axis, or a definite subset of this), $\mathcal{B}_i$ is a sigma-field of subsets of T_i°, and μ_i is a reference measure on this sigma-field. Through multiplication, this construction associated with elementary variables is inherited by multiple variables. Specifically, a definite measure space $(X^\circ, \mathcal{B}_X, \mu_X)$ may be associated with any variable $X = \{T_{i_1}, \ldots, T_{i_k}\} \subseteq T$, in which $X^\circ = T_{i_1}^\circ \times \cdots \times T_{i_k}^\circ$ is the product of the spaces characteristic of the individual components (the space X° includes all possible values of X), $\mathcal{B}_X = \mathcal{B}_{i_1} \times \cdots \times \mathcal{B}_{i_k}$ is the product of the corresponding sigma-fields, and $\mu_X = \mu_{i_1} \times \cdots \times \mu_{i_k}$ is the product of the reference measures defined on these sigma-fields (Billingsley, 1995, § 18).

Discussions of conditional probability distributions imply reference to ordered pairs $(Y|X)$ in which X and Y are disjoint variables. Specifically, the term X (on the right of the bar) has the role of the conditioning variable and may be empty, whereas the term Y (on the left of the bar) has the role of the conditioned variable and is non-empty. We call $(Y|X)$ a *variable pair* and denote by $\mathcal{O}(T)$ the collection of such pairs. Simple combinatorics shows that if n is the cardinality of T, then $3^n - 2^n$ is the cardinality of $\mathcal{O}(T)$. For the purposes of our analysis, we make no substantial difference between any pairs $(\varnothing|X)$ and $(\varnothing|U)$ that have the empty variable on the left: both are symbols of *one* formal entity, called *null variable pair* and generally denoted by $\perp$. The symbol $\widetilde{\mathcal{O}}(T)$ stands for the set $\mathcal{O}(T) \cup \{\perp\}$.

For the utilities that will appear in the next paragraphs, the following criterion for comparing variable pairs is adopted.

Definition 1. A non-null variable pair $(V|U)$ is *dominated* by another non-null variable pair $(Y|X)$ (notation $(V|U) \preceq (Y|X)$) if the inclusions $V \cup U \subseteq Y \cup X$ and $U \supseteq X$ are both true. Furthermore, the null variable pair is dominated by any other variable pair (that is, $\perp \preceq (Y|X)$ for all $(Y|X) \in \mathcal{O}(T)$).

For example, assuming $T = \{T_1, \ldots, T_5\}$, if $(V|U) = (T_1, T_2|T_4, T_5)$, $(Y|X) = (T_1, T_2, T_4|T_5)$, and $(Z|W) = (T_1, T_2|T_5)$, then both $(V|U) \preceq (Y|X)$ and $(Z|W) \preceq (Y|X)$, but neither $(V|U) \preceq (Z|W)$ (condition $V \cup U \subseteq Z \cup W$ is violated) nor $(Z|W) \preceq (V|U)$ (condition $W \supseteq U$ is violated). Figure 1 allows us to present a useful characterization of the dominance defined above. This figure describes the crossing between two generic pairs $(V|U)$ and $(Y|X)$, by labeling the intersections and differences between the variables constituting each pair (for example, A stands for the difference $Y \setminus (V \cup U)$, E for the intersection $Y \cap U$, etc.). In these terms, it is readily seen that

$$(V|U) \preceq (Y|X) \text{ if and only if } B \cup C \cup D \cup G = \varnothing.$$

Figure 1 also plays a crucial role in the proofs of Theorems 1 and 2 stated below.

From the stated definition, any relation $(V|U) \preceq (Y|X)$ is the logical conjunction of the inclusions $V \cup U \subseteq Y \cup X$ and $U \supseteq X$. Due to this and to the fact that inclusion is a partial order (a reflexive,

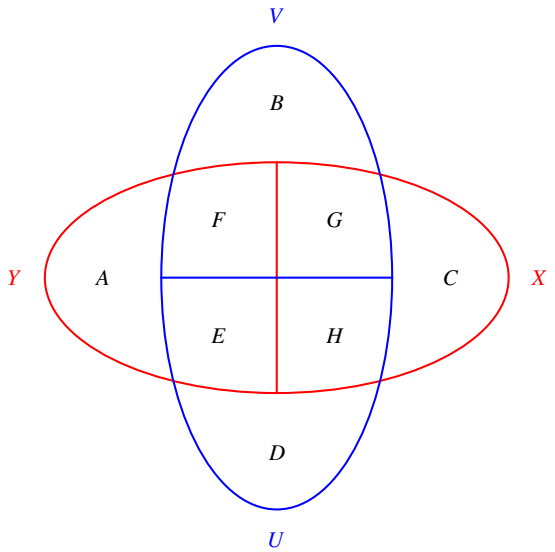


Figure 1. Crossing of two generic variable pairs $(V|U) = (B \cup F \cup G|D \cup E \cup H)$ and $(Y|X) = (A \cup E \cup F|C \cup G \cup H)$. Some of the eight subvariables $A, \ldots, H$ may be empty.

transitive, and antisymmetric relation), we obtain that the relation $\preceq$ is itself a partial order, and that the structure $(\widetilde{\mathcal{O}}(T), \preceq)$ is a partially ordered set. More specifically, the following properties can be proved (Burigana and Vicovaro, 2020, Proposition 1).

Proposition 1. *The structure* $(\widetilde{\mathcal{O}}(T), \preceq)$ *is a lattice in which the pair* $(T|\varnothing)$ *is the supremum, the null pair* $\perp$ *is the infimum, and for all non-null pairs* $(V|U)$ *and* $(Y|X)$ *their join* (*least upper bound*) *and meet* (*greatest lower bound*) *are given by the following equations*:

$$(V|U) \vee (Y|X) = (V \cup Y \cup (U + X)|U \cap X) \tag{1}$$
with $U + X = (U \setminus X) \cup (X \setminus U)$;

$$(V|U) \wedge (Y|X) = (V \cap Y|U \cup X) \text{ or } = \perp \tag{2}$$
depending on whether the conditions $V \cap Y \neq \varnothing,\ U \subseteq Y \cup X,\ X \subseteq V \cup U$ *are or are not jointly true.*

The following additional properties are easily recognized: the atoms in the lattice $\widetilde{\mathcal{O}}(T)$ are the pairs $(Y|X)$ with $|Y| = 1$ (i.e., the conditioned variable is elementary, so that if $n = |T|$, then $n2^{n-1}$ is the number of atoms); the lattice is atomic, in that each member of $\mathcal{O}(T)$ can be expressed as the join of a suitable set of atoms; it is rankable, the rank of any pair $(Y|X)$ being simply the cardinality $|Y|$ of the conditioned variable; and it is locally distributive, by which we mean that the distributive laws hold true on any triple of variable pairs whose pairwise meets are all different from $\perp$ (the null variable pair). Figure 2 shows the Hasse diagrams (in three-dimensional form) of the lattices $\widetilde{\mathcal{O}}(T)$ for $|T| = 2, 3, 4$. In each diagram, a downward line represents a covering $(V|U) \prec\cdot (Y|X)$ in which $|Y \setminus V| = 1$ and

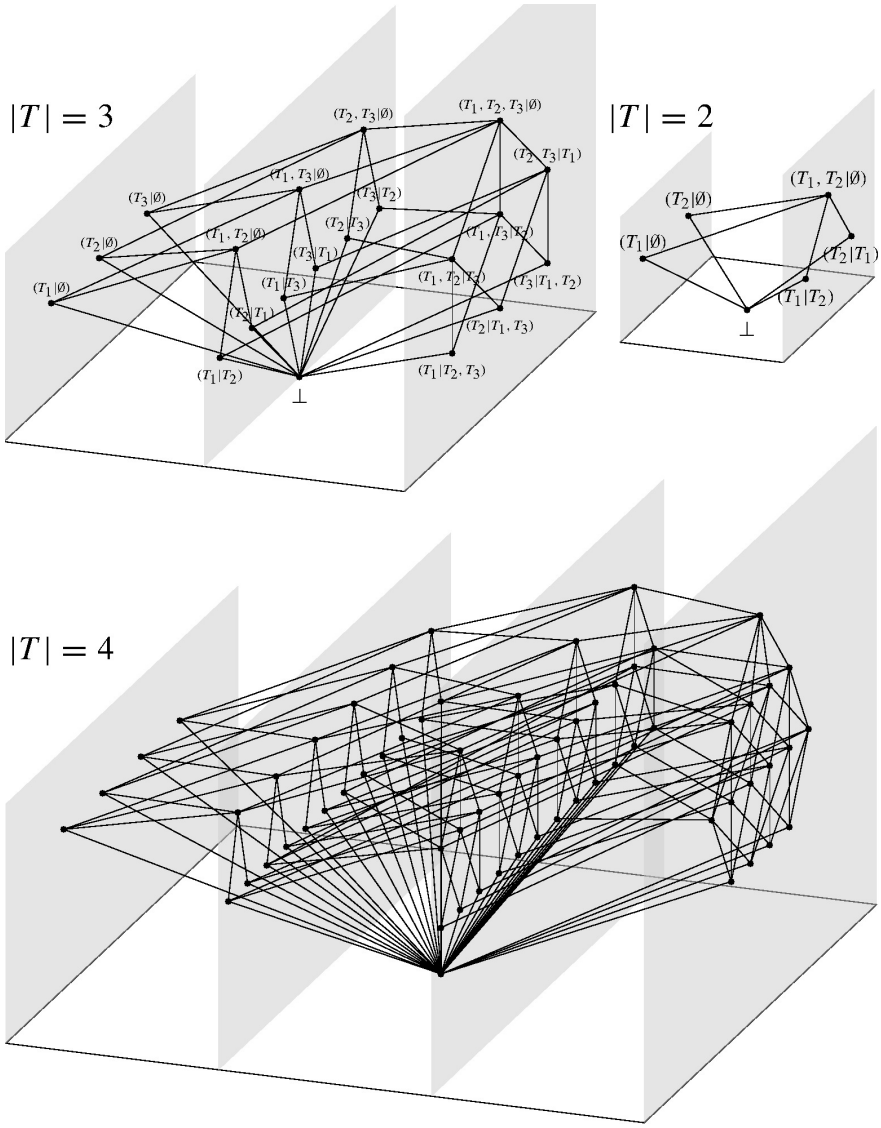


Figure 2. Three-dimensional Hasse diagrams of the lattices $\widetilde{O}(T)$ for $|T| = 2$ (top right), $|T| = 3$ (top left), and (bottom) $|T| = 4$.

$Y \setminus V = U \setminus X$ (that is, $(V|U)$ is derived from $(Y|X)$ by *transferring* one elementary variable from the left to the right component of the pair), whereas a backward line represents a covering $(V|U) \prec\cdot (Y|X)$ in which $|Y \setminus V| = 1$ and $U = X$ (that is, $(V|U)$ is derived from $(Y|X)$ by *cancelling* one elementary variable from the left component of the pair).

The next definition introduces the basic probabilistic objects of our study.

Definition 2. For any variable pair $(Y|X)$, a *probability kernel* associated with it is any family $\{p(Y|x) : x \in X^{\circ}\}$ (indexed by the values of X) in which each member $p(Y|x)$ is a non-negative valued function

defined on the space Y°, measurable relative to the sigma-field $\mathcal{B}_Y$, and such that $\int p(y|x)\mu_Y(dy) = 1$. The family is also denoted by the symbol $p(Y|X)$.

The terms Y°, $\mathcal{B}_Y$, and μ_Y in this definition are the components of the measure space associated with the conditioned variable Y in the kernel. Following common usage, any member $p(Y|x)$ of a given family $p(Y|X)$ is referred to here as a *density*, irrespective of the kind (discrete, continuous, mixed, or other) of the variable Y[1]. The symbol $S(p(Y|x))$ will denote the *support* for any density $p(Y|x)$, that is, the set of points in the space Y° on which the density is positive. Furthermore, simply by considering the set-theoretic characteristics of the variables involved, basic kinds of kernels may be distinguished, which are assigned distinctive names. In particular, a kernel $p(Y|X)$ is *elementary* if Y is an elementary variable (a singleton in T), *saturated* if $X \cup Y = T$ (the full variable), *marginal* if $X = \varnothing$ (a marginal kernel $p(Y|\varnothing)$ is tantamount to one density $p(Y)$ on the space Y°), *full* if $Y = T$ (one density $p(T)$ over the space T°), *null* if $Y = \varnothing$ (symbol $\sharp$ is used here for the null kernel).

Probability kernels are subject to peculiar operations. The following are two basic exemplars.

Definition 3. Let $p(Y \cup Z|X)$ be a kernel in which Y and Z are non-empty disjoint variables.
(i) The result of *projecting* $p(Y \cup Z|X)$ relative to Z, denoted by $J[p(Y \cup Z|X), Z]$, is the kernel that is formed on the variable pair $(Y|X)$ by setting for all $(y, x) \in Y^\circ \times X^\circ$

$$p(y|x) = \int p(y, z|x)\mu_Z(dz).$$

(ii) The result of *conditioning* $p(Y \cup Z|X)$ relative to Z, denoted by $C[p(Y \cup Z|X), Z]$, is the kernel that is formed on the variable pair $(Y|Z \cup X)$ by setting for all $(y, z, x) \in Y^\circ \times Z^\circ \times X^\circ$

$$p(y|z, x) = \begin{cases} \dfrac{p(y, z|x)}{p(z|x)} & \text{if } p(z|x) \neq 0, \\ q(y) & \text{if } p(z|x) = 0, \end{cases}$$

where $p(Z|X) = J[p(Y \cup Z|X), Y]$ and $q(Y)$ is a freely chosen density on the space Y°.

The projection operation (symbol J) is tantamount to marginalization, as applicable to multivariate density functions. The conditioning operation (symbol C) is determined here as a division (the ratio $p(y, z|x)/p(z|x)$ in the formula), thus imitating the concept of conditional probability in its elementary version. For completeness, the definition may include an arbitrary density $q(Y)$, which however does not affect the univocal recovery of $p(Y \cup Z|X)$ from $p(Y|Z \cup X)$ and $p(Z|X)$ through the promotion operation in Definition 5 hereafter[2]. In the stated form, projection and conditioning are only defined for

[1] Based on any kernel $p(Y|X) = \{p(Y|x) : x \in X^\circ\}$, which according to Definition 2 is a family of point functions, a family $P_{Y|X} = \{P_{Y|x} : x \in X^\circ\}$ of set functions on the sigma-field $\mathcal{B}_Y$ can be constructed by setting $P_{Y|x}(B) = \int_B p(y|x)\mu_Y(dy)$ for all $B \in \mathcal{B}_Y$ and $x \in X^\circ$. In fact, in the measure-theoretic approach to probability, it is precisely a family like $P_{Y|X}$ that is called a probability kernel and is taken as a primitive structure, whereas $p(Y|X)$ is deduced as a corresponding family of Radon-Nikodym derivatives (Pollard, 2002, pp. 84, 119). Besides the term "probability kernel", other expressions are also used in different contexts to indicate probabilistic structures of the stated kind, such as "transition probability measure" (Parthasarathy, 2005, p. 174), "probability potential" (Koski and Noble, 2009, p. 58), "characteristic" (Griffeath, 1976, p. 426), "conditional probability distribution" (Koller and Friedman, 2009, p. 47), or simply "conditional" (Gelfand and Smith, 1990, p. 400).

[2] A discussion of the formal difficulties implicit in the intuitive notion of conditional probability and alternative ways of addressing them is given by Chang and Pollard (1997).

any $p(Y \cup Z|X)$ such that Y and Z are non-empty. Both concepts, however, may consistently be extended beyond this boundary by assuming these equations:

$$J[p(Y|X), \varnothing] = p(Y|X) = C[p(Y|X), \varnothing], \tag{3}$$
$$J[p(Z|X), Z] = \sharp = C[p(Z|X), Z].$$

For example, $J[p(Z|X), Z] = J[p(\varnothing \cup Z|X), Z] = p(\varnothing|X)$, which is the null kernel $\sharp$. Lastly, the following equations, in which X, Y, W, and Z are disjoint variables, are easily deduced from Definition 3:

$$J[J[p(Y \cup Z \cup W|X), Z], W] = p(Y|X) = J[J[p(Y \cup Z \cup W|X), W], Z],$$
$$C[C[p(Y \cup Z \cup W|X), Z], W] = p(Y|Z \cup W \cup X) = C[C[p(Y \cup Z \cup W|X), W], Z],$$
$$J[C[p(Y \cup Z \cup W|X), Z], W] = p(Y|Z \cup X) = C[J[p(Y \cup Z \cup W|X), W], Z].$$

They express commutativity properties of projection and conditioning.

These operations determine a key relation among kernels.

Definition 4. A kernel $p(V|U)$ is *dominated* by a kernel $p(Y|X)$ (notation $p(V|U) \preceq p(Y|X)$) if the former can be obtained from the latter by projection, conditioning, or a combination of projection and conditioning.

Note that, if $p(V|U) \preceq p(Y|X)$, meaning that $p(V|U) = J[C[p(Y|X), W], Z]$ for some W and Z disjoint sub-variables of Y, then $V = Y \setminus (W \cup Z)$ and $U = W \cup X$ according to Definition 3, so that $V \cup U \subseteq Y \cup X$ and $U \supseteq X$, and therefore $(V|U) \preceq (Y|X)$ on applying Definition 1. In other words:

$$\text{if } p(V|U) \preceq p(Y|X) \text{ then } (V|U) \preceq (Y|X). \tag{4}$$

Hence, dominance between variable pairs (the symbol $\preceq$ in the consequent of this implication) is a necessary condition for dominance between kernels (the symbol $\preceq$ in the antecedent).

From a general perspective, given a full density $p(T) = p(T|\varnothing)$, we may consider the set of *all* kernels that are *dominated* by $p(T)$, a set denoted here by $\mathcal{P}(T)$. On the one hand, implication (4) shows that there is a natural one-to-one correspondence between this set $\mathcal{P}(T)$ and the set $\mathcal{O}(T)$ of all variable pairs in T, and there is also correspondence between the null kernel $\sharp$ and the null variable pair $\perp$. On the other hand, it can be seen that if comparison is limited to kernels belonging to a definite set $\mathcal{P}(T)$, then the implication (4) is reinforced as a bi-implication, that is:

$$\text{for all } p(V|U) \text{ and } p(Y|X) \text{ in } \mathcal{P}(T)$$
$$p(V|U) \preceq p(Y|X) \text{ if and only if } (V|U) \preceq (Y|X),$$

so that the abovementioned correspondence is in fact an isomorphism between the structure $(\widetilde{\mathcal{P}}(T), \preceq)$ (with $\widetilde{\mathcal{P}}(T) = \mathcal{P}(T) \cup \{\sharp\}$) and the structure $(\widetilde{\mathcal{O}}(T), \preceq)$ (with $\widetilde{\mathcal{O}}(T) = \mathcal{O}(T) \cup \{\perp\}$). Proposition 1 ensures that the latter structure is a lattice. This means that the former is also a lattice, referred to here as the *lattice of kernels* generated by the assumed full density $p(T)$. This full density is the supremum in the lattice $\widetilde{\mathcal{P}}(T)$, while the infimum is the null kernel $\sharp$, and the atoms are the elementary kernels. The join and meet operations are expressed by formulas that are similar to (1) and (2) but involve arbitrary

kernels $p(V|U)$ and $p(Y|X)$, rather than variable pairs. In the Hasse diagram of a lattice $\widetilde{\mathcal{P}}(T)$ (see Figure 2) each backward line represents the move from a $p(Y|X)$ to $p(Y \setminus \{T_i\}|X) = J[p(Y|X), \{T_i\}]$ by projection relative to an elementary variable $T_i \in Y$, whereas each downward line represents the move from a $p(Y|X)$ to $p(Y \setminus \{T_i\}|\{T_i\} \cup X) = C[p(Y|X), \{T_i\}]$ by elementary conditioning.

For later use, we note the following special property of any lattice of kernels.

Lemma 1. *Let $p(T)$ be a full density and suppose that $q(Y|Z \cup X)$ and $q(Y|X)$ are kernels such that $q(Y|z, x) = q(Y|x)$ for all $(z, x) \in Z^\circ \times X^\circ$. In these conditions, if $q(Y|Z \cup X)$ belongs to the lattice $\widetilde{\mathcal{P}}(T)$ generated by $p(T)$, then $q(Y|X)$ also belongs to the same lattice.*

Proof. Consider the kernels $p(Y|Z \cup X) = J[C[p(T), Z \cup X], T \setminus (Y \cup Z \cup X)]$ and $p(Y|X) = J[C[p(T), X], T \setminus (Y \cup X)]$, which surely belong to $\widetilde{\mathcal{P}}(T)$. If $q(Y|Z \cup X) \in \widetilde{\mathcal{P}}(T)$, then $q(Y|Z \cup X) = p(Y|Z \cup X)$, due to the one-to-one correspondence between $\widetilde{\mathcal{P}}(T)$ and $\widetilde{\mathcal{O}}(T)$. The hypothesized relation between $q(Y|Z \cup X)$ and $q(Y|X)$ implies that for each $x \in X^\circ$, the density $p(Y|z, x)$ is invariant relative to z varying in Z°. Thus, under the density $p(T)$ the variables Y and Z are conditionally independent given X, which means $p(Y|z, x) = p(Y|x)$ for all $(z, x) \in Z^\circ \times X^\circ$ (Lauritzen, 1996, p. 29). We then have $q(Y|x) = q(Y|z, x) = p(Y|z, x) = p(Y|x)$ for all $(z, x) \in Z^\circ \times X^\circ$, so that $q(Y|X) = p(Y|X)$, which combined with $p(Y|X) \in \widetilde{\mathcal{P}}(T)$ implies $q(Y|X) \in \widetilde{\mathcal{P}}(T)$. □

In our analyses, in addition to the projection and conditioning operations (which have a kernel and a variable as operands), two further operations are considered that have two kernels as operands. These are constrained operations, since to be applied they require that the operands (and, in particular, the variable pairs in these) comply with definite conditions.

Definition 5. (i) Given two kernels $p(Y|V \cup U)$ and $p(V|U)$, the result of *promoting* the former by the latter, denoted by $M[p(Y|V \cup U), p(V|U)]$, is the kernel that is formed on the variable pair $(Y \cup V|U)$ by setting for all $(y, v, u) \in Y^\circ \times V^\circ \times U^\circ$

$$p(y, v|u) = p(y|v, u) \cdot p(v|u).$$

(ii) Given two kernels $p(Y|V \cup U)$ and $p(V|Y \cup U)$ that are dominated by some full density $p(T)$ under which the equation $S(p(Y \cup V \cup U)) = S(p(Y)) \times S(p(V)) \times S(p(U))$ concerning the supports is satisfied, the result of *lightening* the former kernel by the latter, denoted by $L[p(Y|V \cup U), p(V|Y \cup U)]$, is the kernel that is formed on the variable pair $(Y|U)$ by setting for all $(y, u) \in Y^\circ \times U^\circ$

$$p(y|u) = \frac{1}{\displaystyle\int \frac{p(v|y, u)}{p(y|v, u)} \mu_V(dv)}.$$

With regard to the variable pairs, the applicability conditions of these two operations are implicit in the notation used in their definition. In particular, promotion is applicable only if the variable pair in the second operand $p(V|U)$ (the promoter) is a bipartition of the conditioning variable in the first operand $p(Y|V \cup U)$; in relation to this, the operation has the effect of *transferring* the variable V from the right to the left of the bar, yielding $p(Y \cup V|U)$ as the result. Lightening is applicable only if the variable pairs in the operands $p(Y|V \cup U)$ and $p(V|Y \cup U)$ have the same union and

the conditioned variables are disjoint; the operation has the effect of *cancelling* the variable V from $p(Y|V \cup U)$, thus yielding $p(Y|U)$ as the result. Furthermore, lightening is applicable only if the factorability $S(p(Y \cup V \cup U)) = S(p(Y)) \times S(p(V)) \times S(p(U))$ of the support of $p(Y \cup V \cup U)$ is satisfied, meaning that for each $(y, v, u) \in Y^\circ \times V^\circ \times U^\circ$ the value $p(y, v, u)$ is positive if (and only if) all three values $p(y)$, $p(v)$, and $p(u)$ are positive, where $p(Y \cup V \cup U)$, $p(Y)$, $p(V)$, and $p(U)$ are densities deducible from some full density $p(T)$ through projection[3]. Under this condition, for each $(y, v, u) \in S(p(Y \cup V \cup U))$ the ratio $p(v|y, u)/p(y|v, u)$ does exist as a positive real number and for each $(y, u) \in S(p(Y \cup U))$ the following equations are true:

$$\frac{1}{\int \frac{p(v|y,u)}{p(y|v,u)}\mu_V(dv)} = \frac{1}{\int \frac{p(y,v,u)/p(y,u)}{p(y,v,u)/p(v,u)}\mu_V(dv)} = \frac{p(y,u)}{\int p(v,u)\mu_V(dv)} = \frac{p(y,u)}{p(u)} = p(y|u).$$

This shows that the result $L[p(Y|V \cup U), p(V|Y \cup U)]$ of lightening is precisely the kernel $p(Y|U)$ belonging to the same lattice to which the operands $p(Y|V \cup U)$ and $p(V|Y \cup U)$ belong[4].

With regard to the promotion operation, it is readily seen that for any two kernels $p(Y|V \cup U)$ and $p(V|U)$ its result $p(Y \cup V|U) = M[p(Y|V \cup U), p(V|U)]$ is itself a kernel on the indicated variable pair. In particular, for all $u \in U^\circ$,

$$\begin{aligned}\int p(y, v|u)(\mu_Y \times \mu_V)(d(y, v)) &= \int p(y|v, u) \cdot p(v|u)(\mu_Y \times \mu_V)(d(y, v)) \\ &= \int \Big[\int p(y|v, u)\mu_Y(dy)\Big] \cdot p(v|u)\mu_V(dv) = \int 1 \cdot p(v|u)\mu_V(dv) = 1\end{aligned}$$

where the second step is justified by Fubini's theorem. When reference is made to a definite lattice of kernels, the promotion operation (as well as the lightening operation) may then be viewed as a binary operation internal to the lattice and subject to a specific applicability condition. Indeed, when applicable, promotion produces the same results as the join operation in the lattice, as may be inferred from Equation (1). Furthermore, its definition may be consistently refined by assuming these equations:

$$M[\sharp, p(V|U)] = p(V|U), \qquad M[p(Y|X), \sharp] = p(Y|X). \tag{5}$$

These characterize the null kernel $\sharp$ as the left and right identity term for promotion and can be justified by replacing $\sharp$ by $p(\varnothing|V \cup U)$ in one case and by $p(\varnothing|X)$ in the other. Lastly, the following equation is easily deduced from Definition 5(i):

$$\begin{aligned}&M[M[p(Y|W \cup V \cup U), p(W|V \cup U)], p(V|U)] = p(Y \cup W \cup V|U) = \\ &M[p(Y|W \cup V \cup U), M[p(W|V \cup U), p(V|U)]].\end{aligned} \tag{6}$$

This characterizes promotion as an associative operation.

[3]This factorability requirement is also known as the "positivity condition" regarding multivariate densities (Besag, 1974, p. 195).

[4]The operation we call "lightening" was considered, for example, by Gourieroux and Monfort (1979) and Robert and Casella (2004, p. 344).

On the whole, we have four basic operations on probability kernels, with the symbols J (proJection), C (Conditioning), M (proMotion), and L (Lightening). In the next lemma, several equations are highlighted that arise from the combined use of these operations and will be applied in the following.

Lemma 2. *In each of the following equations, the kernels involved are assumed to belong to one lattice* $\widetilde{\mathcal{P}}(T)$.

(i) $J[M[p(Y|V \cup U),\, p(V|U)],\, V] = p(Y|U)$ (*relative to the first operand, part* V *of the conditioning variable is cancelled*).

(ii) $C[M[p(Y|V \cup U),\, p(V|U)],\, V] = p(Y|V \cup U)$ (*recovery of the first operand of a promotion*).

(iii) $J[M[p(Y|V \cup U),\, p(V|U)],\, Y] = p(V|U)$ (*recovery of the second operand of a promotion*).

(iv) $J[M[p(Y \cup X|V \cup U),\, p(V|U)],\, X] = M[J[p(Y \cup X|V \cup U),\, X],\, p(V|U)]$ (*a kind of commutativity between projection and promotion*).

(v) $M[p(Y|V \cup U),\, L[p(V|Y \cup U),\, p(Y|V \cup U)]] = p(Y \cup V|U)$ (*simulation of the join operation*).

(vi) $M[C[p(Y \cup Z|X),\, Z],\, J[p(Y \cup Z|X),\, Y)]] = p(Y \cup Z|X)$ (*recovery of a kernel through promotion*).

Proof. Each of these equations can be proved by noting that the kernel resulting from the composite operation on the left-hand side acts on the same variable pair as the kernel specified on the right-hand side, and then considering the one-to-one correspondence between variable pairs and kernels in a lattice (as implied by Equation (4)). Consider, for example, statement (v). From Definition 5(ii), the result $L[p(V|Y \cup U),\, p(Y|V \cup U)]$ is a kernel on the variable pair $(V|U)$, so that from Definition 5(i), the result $M[p(Y|V \cup U),\, L[p(V|Y \cup U),\, p(Y|V \cup U)]]$ is a kernel on the variable pair $(Y \cup V|U)$. This is the same variable pair as that for the kernel $p(Y \cup V|U)$, so that from the mentioned one-to-one correspondence, that result must be equal to this kernel. In turn, according to Equation (1), $p(Y \cup V|U)$ is equal to $p(Y|V \cup U) \vee p(V|Y \cup U)$, which is the join of the two input kernels on the left-hand side of the equation. □

As a point that is relevant to the following, let us consider this task: for any full density $p(T)$, find a (preferably small) set of kernels that are dominated by $p(T)$ and that form a *sufficient basis* for the univocal recovery of $p(T)$, using available operations on kernels. As they are all dominated by $p(T)$, the kernels in any such sufficient basis are compatible with one another. We shall see that the stated task is related to the problem of compatibility among kernels, and especially to the possible uniqueness of a consensus density for compatible kernels. The next lemma presents two exemplary answers to the task that describe sufficient bases of different forms, one of which follows a "cumulative scheme" and the other an "alternating scheme" as regards the variable pairs in the kernels.

Lemma 3. *Let* $p(T)$ *be a full density and* $\{Y_1, \ldots, Y_m\}$ *be a partition of the full variable* T.

(i) *Suppose* $X_1 = \varnothing$ *and* $X_i = Y_1 \cup \cdots \cup Y_{i-1}$ *for* $i = 2, \ldots, m$. *Then the set* $\{p(Y_1|X_1), \ldots, p(Y_m|X_m)\}$ *of kernels dominated by* $p(T)$ *is a sufficient basis for the recovery of* $p(T)$.

(ii) *Suppose that* $X_i = T \setminus Y_i = Y_1 \cup \cdots \cup Y_{i-1} \cup Y_{i+1} \cup \cdots \cup Y_m$ *for* $i = 1, \ldots, m$ *and that* $S(p(T)) = S(p(Y_1)) \times \cdots \times S(p(Y_m))$ (*factorability of the support for the density* $p(T)$). *Then the set* $\{p(Y_1|X_1), \ldots, p(Y_m|X_m)\}$ *of kernels dominated by* $p(T)$ *is a sufficient basis for the recovery of* $p(T)$.

Proof. (i) Consider this sequence of densities, all deducible (by projection) from the full density in question:

$$p(Y_1), \ldots, p(Y_1 \cup \cdots \cup Y_{i-1}), p(Y_1 \cup \cdots \cup Y_{i-1} \cup Y_i), \ldots, p(Y_1 \cup \cdots \cup Y_m).$$

The density $p(Y_1)$ is equal to $p(Y_1|\varnothing) = p(Y_1|X_1)$, which is available in the assumed set of kernels. Consider any $1 < i \leq m$ and suppose (as an inductive hypothesis) that the density $p(Y_1 \cup \cdots \cup Y_{i-1})$ is uniquely determined by the available kernels. Then, $p(Y_1 \cup \cdots \cup Y_{i-1} \cup Y_i)$ is also uniquely determined, since

$$p(Y_1 \cup \cdots \cup Y_{i-1} \cup Y_i) = M[p(Y_i|Y_1 \cup \cdots \cup Y_{i-1}), p(Y_1 \cup \cdots \cup Y_{i-1})]$$

by Definition 5(i) and $p(Y_i|Y_1 \cup \cdots \cup Y_{i-1}) = p(Y_i|X_i)$ is one of the available kernels. In particular, we can then conclude for $i = m$ that the full density $p(T) = p(Y_1 \cup \cdots \cup Y_m)$ is univocally recoverable (through iterated promotion) from the available kernels.

(ii) For each $i = 1, \ldots, m$, let us denote by Z_i the variable $Y_1 \cup \cdots \cup Y_i$, and then consider this sequence of saturated kernels, which are all deducible from $p(T)$ by conditioning:

$$p(Z_1|T \setminus Z_1), \ldots, p(Z_{i-1}|T \setminus Z_{i-1}), p(Z_i|T \setminus Z_i), \ldots, p(Z_m|T \setminus Z_m).$$

The first member equals $p(Y_1|T \setminus Y_1)$, which is one of the available kernels. We then consider any $1 < i \leq m$, and assume (as an inductive hypothesis) that the kernel $p(Z_{i-1}|T \setminus Z_{i-1})$ is uniquely determined by the available kernels. According to Lemma 2(v), and since $Z_i = Z_{i-1} \cup Y_i$, the following equation is true:

$$p(Z_i|T \setminus Z_i) = M[p(Z_{i-1}|T \setminus Z_{i-1}), L[p(Y_i|T \setminus Y_i), p(Z_{i-1}|T \setminus Z_{i-1})]].$$

Hence, from the inductive hypothesis concerning $p(Z_{i-1}|T \setminus Z_{i-1})$ and the fact that $p(Y_i|T \setminus Y_i)$ is one of the available kernels, we can deduce that $p(Z_i|T \setminus Z_i)$ is uniquely determined by the available kernels. In particular, we then have for $i = m$ that the full density $p(T) = p(T|\varnothing) = p(Z_m|T \setminus Z_m)$ is univocally recoverable (through an iterated lightening-and-promotion operation) from the available kernels. Note that this property can also be proved using the odds-product method that is discussed in Section 5. □

The partition hypothesized in Lemma 3 could be the subdivision $\{\{T_1\}, \ldots, \{T_n\}\}$ of the full variable $T = \{T_1, \ldots, T_n\}$ into singletons, meaning that all kernels mentioned in the lemma would be elementary kernels, that is, atoms in a lattice $\widetilde{\mathcal{P}}(T)$. Part (ii) of the lemma would then state that a full density $p(T)$ is unambiguously recoverable from the corresponding set $\{p(T_1|T \setminus \{T_1\}), \ldots, p(T_n|T \setminus \{T_n\})\}$ of elementary saturated kernels, which is a well-known result in the theory of conditionally specified probabilistic models (Besag, 1974, pp. 195–196). In addition to corollaries, Lemma 3 also admits significant generalizations. It is proven, for example, that if $\{Y_1, \ldots, Y_m\}$ is a partition of T and X_i includes (but does not necessarily equal) $Y_1 \cup \cdots \cup Y_{i-1}$ for all $i = 1, \ldots, m$, then $\{p(Y_1|X_1), \ldots, p(Y_m|X_m)\}$ is a sufficient basis for the

recovery of $p(T)$ (Gelman and Speed, 1993). Furthermore, if given variables $Y_1, \ldots, Y_m$ cover the whole of T (but may overlap with one another), then the set $\{p(Y_1|T \setminus Y_1), \ldots, p(Y_m|T \setminus Y_m)\}$ of saturated kernels is a sufficient basis for the recovery of $p(T)$ (see the "if" part of Theorem 3 in Section 5). Lastly, the conditions for the recovery of a full density $p(T)$ are easily adaptable to the recovery of any kernel $p(Y|X)$. For example, if $\{Z_1, \ldots, Z_k\}$ is any partition of the variable Y, then $p(Y|X)$ is unambiguously recoverable both from the set $\{p(Z_1|X), p(Z_2|Z_1 \cup X), \ldots, p(Z_k|Z_{k-1} \cup \cdots \cup Z_1 \cup X)\}$ and from the set $\{p(Z_1|(Y \setminus Z_1) \cup X), \ldots, p(Z_k|(Y \setminus Z_k) \cup X)\}$ of kernels dominated by it in a lattice.

3. Compatibility of probability kernels and sure compatibility of variable pairs

The algebraic frame outlined in the preceding section allows us to assign a suitable place to the main concept of our study.

Definition 6. Given probability kernels $p(Y_1|X_1), \ldots, p(Y_m|X_m)$ on variable pairs within a full variable T are *compatible* with one another if there exists a full density $p(T)$ that dominates each of them. Such $p(T)$ is said to be a *consensus density* for the given kernels.

In other words, compatibility within a set of kernels means that there exists a lattice of kernels that includes that set. This concept is a topic in the literature concerning conditionally specified probabilistic models (Arnold, Castillo and Sarabia, 1999, p. 5; Kaiser, 2002, p. 1213). Terms such as "candidate, putative, proposed conditionals" are used in relation to kernels whose mutual compatibility is under judgement (Arnold, Castillo and Sarabia, 2004, pp. 147, 157).

As a first comment on the above definition, we remark that there are alternative ways of expressing the compatibility relation. In particular, we can refer to the join $(Z|W) = (Y_1|X_1) \vee \cdots \vee (Y_m|X_m)$ of the variable pairs in the candidate kernels, and state that these are compatible if there exists a kernel $p(Z|W)$ that dominates each of them. As a second comment, we note that a uniqueness problem and a construction problem are naturally associated with the compatibility problem (concerning existence). That is, if there are reasons for stating that given kernels are compatible, then one may ask whether there is only one consensus density for them, or a (possibly infinite) number of such densities, and look for practical procedures for discovering or constructing such a density. The concept of a "sufficient basis", which was discussed at the end of the preceding section, is related to these problems. As a third comment, we remark that compatibility is not a transitive relation in general. For example, if X and Y are disjoint variables, then any density $p(X)$ is compatible both with any density $p(Y)$ and with any kernel $p(Y|X)$, although these two may fail to be compatible unless $p(Y) = J[M[p(Y|X), p(X)], X]$ (see Lemma 2(i)). This lack of transitivity implies that compatibility within a set of three or more kernels cannot be approved solely on the basis of pairwise compatibility, and is a sign of the difficulty of the problem. Indeed, the compatibility problem may become quite tricky, as revealed by paradoxical situations noted in the literature. It is surprising, for example, that if $p(Y|X)$ and $p(X|Y)$ are deduced (by conditioning) from a certain density $p(X \cup Y)$, a collection $\mathcal{Q}$ of kernels may nevertheless exist such that compatibility is true within $\mathcal{Q} \cup \{p(Y|X), p(X|Y)\}$ but false within $\mathcal{Q} \cup \{p(X \cup Y)\}$, even though $\{p(Y|X), p(X|Y)\}$ and $p(X \cup Y)$ are substantially equivalent, since the latter (under suitable conditions) may be recovered

from the former through lightening-and-promotion, as shown by Lemma 2(v) (Kuo and Wang, 2011, pp. 2460–2461).

The next definition focuses on variable pairs as the set-theoretic carriers of probability kernels.

Definition 7. A set $\{(Y_1|X_1), \ldots, (Y_m|X_m)\}$ of variable pairs has *sure compatibility* if every set

$$\{p(Y_1|X_1), \ldots, p(Y_m|X_m)\}$$

of kernels definable on those pairs satisfies the compatibility condition.

Note that given any set $\{(Y_1|X_1), \ldots, (Y_m|X_m)\}$ of variable pairs in a full variable T, there is certainly *some* set $\{p(Y_1|X_1), \ldots, p(Y_m|X_m)\}$ of kernels on those pairs that are mutually compatible: we can simply consider any density $p(T)$ on the full variable and then derive the kernels from it by suitable projections and/or conditionings. However, Definition 7 demands much more than this: it demands that *every* possible set $\{p(Y_1|X_1), \ldots, p(Y_m|X_m)\}$ of kernels on the given variable pairs satisfies compatibility, meaning that the root of compatibility is to be found not in the numerical characteristics of the particular kernels considered but in the set-theoretic characteristics of the variable pairs themselves. For example, if X and Y are disjoint variables, then the set $\{(X|\varnothing), (Y|\varnothing)\}$ has sure compatibility, since for any densities $p(X) = p(X|\varnothing)$ and $q(Y) = q(Y|\varnothing)$ we can consider (for example) the product density $r(X \cup Y) = p(X) \cdot q(Y)$, which dominates (by projection) both $p(X|\varnothing)$ and $q(Y|\varnothing)$, so that these two are compatible with each other. Conversely, if X and Y are not disjoint, then any given densities $p(X) = p(X|\varnothing)$ and $q(Y) = q(Y|\varnothing)$ are compatible only if $p(X \cap Y) = q(X \cap Y)$ (with $p(X \cap Y) = J[p(X), X \setminus Y]$ and $q(X \cap Y) = J[q(Y), Y \setminus X]$), so that the set $\{(X|\varnothing), (Y|\varnothing)\}$ has not sure compatibility.

The next lemma highlights two further counterexamples to sure compatibility.

Lemma 4. (i) *For all variable pairs* $(V|U)$ *and* $(Y|X)$ *such that* $V \cap Y \neq \varnothing$, *the set* $\{(V|U), (Y|X)\}$ *has not sure compatibility.*

(ii) *Any circular set* $\{(Z_m|Z_{m-1}), (Z_{m-1}|Z_{m-2}), \ldots, (Z_2|Z_1), (Z_1|Z_m)\}$ *of non-null variable pairs has not sure compatibility.*

Proof. (i) Since $V \cap Y$ is a non-empty variable, there are densities $p(V \cap Y)$ and $q(V \cap Y)$ that are different from each other. Choose any densities $p(V)$ and $q(Y)$ such that $p(V \cap Y) \preceq p(V)$ and $q(V \cap Y) \preceq q(Y)$, and then construct the kernels $p(V|U)$ and $q(Y|X)$ by setting $p(V|u) = p(V)$ for all $u \in U^\circ$ and $q(Y|x) = q(Y)$ for all $x \in X^\circ$. We claim that these kernels are not compatible with each other (which then implies that the set $\{(V|U), (Y|X)\}$ has not sure compatibility). Indeed, suppose (ad absurdum) that they are compatible, so that there is a full density $r(T)$ such that $p(V|U) \preceq r(T)$ and $q(Y|X) \preceq r(T)$. Then, considering the way in which $p(V|U)$ and $q(Y|X)$ are constructed, we should also have $p(V) \preceq r(T)$ and $q(Y) \preceq r(T)$ by Lemma 1, and then $p(V \cap Y) \preceq r(T)$ and $q(V \cap Y) \preceq r(T)$ by transitivity. However, this is impossible, because $p(V \cap Y)$ and $q(V \cap Y)$ act on the same variable and are different.

(ii) Given a circular set $\{(Z_m|Z_{m-1}), \ldots, (Z_2|Z_1), (Z_1|Z_m)\}$ of variable pairs, let us construct a corresponding set $\{p_{m-1}(Z_m|Z_{m-1}), \ldots, p_1(Z_2|Z_1), p_m(Z_1|Z_m)\}$ of kernels with these characteristics:

for each $i = 1, \ldots, m-1$, the kernel $p_i(Z_{i+1}|Z_i)$ is of deterministic type,
which means that there is a function f_i from Z_i° to Z_{i+1}° such that
$p_i(v|u) = 1$ or $= 0$, depending on whether v equals or does not equal $f_i(u)$,
for all $u \in Z_i^\circ$ and $v \in Z_{i+1}^\circ$;

the kernel $p_m(Z_1|Z_m)$ is such that
there are $t \in Z_1^\circ$ and $w \neq w' \in Z_m^\circ$ with $p_m(t|w) > 0$ and $p_m(t|w') > 0$.

Our claim is that such kernels are not mutually compatible (which then implies that the given circular set of variable pairs has not sure compatibility). Suppose the contrary, that is, the existence of a consensus density $r(T)$, so that

$$r(Z_{i+1}|Z_i) = p_i(Z_{i+1}|Z_i) \text{ for all } i = 1, \ldots, m-1 \quad \text{and} \quad r(Z_1|Z_m) = p_m(Z_1|Z_m).$$

It can be seen that, due to the deterministic character of the kernels $r(Z_m|Z_{m-1}), \ldots, r(Z_2|Z_1)$, under the density $r(T)$ for each point $u \in Z_1^\circ$ there must be a single point $g(u) \in Z_m^\circ$ with positive conditional probability given the hypothesis $Z_1 = u$. In other words, the kernel $r(Z_m|Z_1)$ that is deducible from $r(T)$ must itself be deterministic. In regard to the above-mentioned points $t \in Z_1^\circ$ and $w \neq w' \in Z_m^\circ$, this implies, in particular, that we cannot have both $r(t, w) > 0$ and $r(t, w') > 0$, whereas from the construction of $p_m(Z_1|Z_m)$ and the equality $r(Z_1|Z_m) = p_m(Z_1|Z_m)$ we should have both $r(t, w) > 0$ and $r(t, w') > 0$, which is a contradiction. Therefore, the constructed kernels cannot have a consensus density; that is, they are not mutually compatible. □

In the next paragraph, we will present a characterization of the concept of "sure compatibility". In expressing the characterization, use will be made of a simple binary relation between variable pairs that is different from the dominance relation specified in Definition 1.

Definition 8. A variable pair $(V|U)$ is *incident* on a variable pair $(Y|X)$ (notation $(V|U) \to (Y|X)$) if $V \cap X \neq \varnothing$.

This relation is areflexive, simply because in any variable pair the two variables are assumed to be disjoint. Besides, it is free as regards other possible formal properties of binary relations, such as symmetry, asymmetry, transitivity, acyclicity, and so on. For example, if X and Y are disjoint non-empty variables, then the set $\{(Y|X), (X|Y)\}$ of two symmetric variable pairs forms a cycle of length two according to incidence. If $T = \{T_1, T_2, T_3, T_4, T_5\}$, then the pairs $\{(T_1, T_3|T_4, T_5), (T_4|T_2), (T_2, T_3|T_1)\}$ form a cycle of length three, whereas within the set $\{(T_5|T_3, T_4), (T_4|T_1, T_2), (T_3|T_1), (T_1|\varnothing)\}$ the incidence relation is acyclic.

The following is a salient result of our discussion of compatibility in this article.

Theorem 1. *A set* $\{(Y_1|X_1), \ldots, (Y_m|X_m)\}$ *of variable pairs in a full variable* T *has sure compatibility of kernels if and only if* (i) *the conditioned variables* $Y_1, \ldots, Y_m$ *in the pairs are disjoint from one another and* (ii) *the incidence* $\rightarrow$ *between the pairs is an acyclic relation.*

Proof. "Only if" part. The necessity of condition (i) follows directly from Lemma 4(i), since if a set of variable pairs has sure compatibility, then each of its subsets must also have this property. To prove the necessity of condition (ii), let us first suppose that the given set of variable pairs forms a $\rightarrow$-cycle, specifically

$$(Y_m|X_m) \rightarrow (Y_1|X_1) \rightarrow (Y_2|X_2) \rightarrow \cdots \rightarrow (Y_{m-2}|X_{m-2}) \rightarrow (Y_{m-1}|X_{m-1}) \rightarrow (Y_m|X_m),$$

which means that the following variables are all non-empty

$$Z_1 = X_1 \cap Y_m,\ Z_2 = X_2 \cap Y_1, \ldots, Z_{m-1} = X_{m-1} \cap Y_{m-2},\ Z_m = X_m \cap Y_{m-1}.$$

Thus, $\{(Z_m|Z_{m-1}), \ldots, (Z_2|Z_1), (Z_1|Z_m)\}$ is a circular set of non-null variable pairs and Lemma 4(ii) ensures the existence of a set of kernels $\{p_{m-1}(Z_m|Z_{m-1}), \ldots, p_1(Z_2|Z_1),\ p_m(Z_1|Z_m)\}$ that are not mutually compatible. For each $i = 1, \ldots, m$ (with $i - 1 = m$ for $i = 1$, and $i + 1 = 1$ for $i = m$), we can expand the kernel

$$p_i(Z_{i+1}|Z_i) = p_i(X_{i+1} \cap Y_i | X_i \cap Y_{i-1})$$

into a kernel

$$p_i(Y_i|X_i) = p_i(Z_{i+1} \cup (Y_i \setminus X_{i+1}) | Z_i \cup (X_i \setminus Y_{i-1}))$$

by first constructing

$$p_i(Z_{i+1} \cup (Y_i \setminus X_{i+1})|Z_i) = M[p_i(Y_i \setminus X_{i+1}|Z_{i+1} \cup Z_i),\ p_i(Z_{i+1}|Z_i)]$$

where $p_i(Y_i \setminus X_{i+1}|Z_{i+1} \cup Z_i)$ is an arbitrarily chosen kernel, and then setting

$$p_i(Z_{i+1} \cup (Y_i \setminus X_{i+1})|Z_i, u) = p_i(Z_{i+1} \cup (Y_i \setminus X_{i+1})|Z_i) \text{ for every } u \in (X_i \setminus Y_{i-1})^\circ.$$

It can be seen that if we have a full density $p(T)$ such that $p_i(Y_i|X_i) \preceq p(T)$ for all $i = 1, \ldots, m$, then, from Lemmas 1 and 2(iii), we also have $p_i(Z_{i+1}|Z_i) \preceq p(T)$ for all $i = 1, \ldots, m$, which contradicts the assumption that the kernels $\{p_i(Z_{i+1}|Z_i)\}$ are not compatible. Hence, the kernels $\{p_i(Y_i|X_i)\}$ constructed in this way are not compatible, which proves that the set of pairs $\{(Y_i|X_i)\}$ has not sure compatibility. Lastly, if the set of pairs $\{(Y_1|X_1), \ldots, (Y_m|X_m)\}$ were not a $\rightarrow$-cycle but included a subset that formed a $\rightarrow$-cycle, then the argument developed above could be applied to that subset, thus showing that not only the subset but also the entire set including it has not sure compatibility.

"If" part. Suppose that $\{(Y_1|X_1), \ldots, (Y_m|X_m)\}$ is a set of variable pairs that complies with conditions (i) and (ii) in the theorem. Property (ii) implies that there is a permutation $((Y_{s(1)}|X_{s(1)}), \ldots, (Y_{s(m)}|X_{s(m)}))$ of the given set such that not$((Y_{s(i)}|X_{s(i)}) \rightarrow (Y_{s(j)}|X_{s(j)}))$ for all $1 \leq j < i \leq m$. Together with property (i), this implies

$$Y_{s(i)} \cap (Y_{s(i-1)} \cup X_{s(i-1)} \cup \cdots \cup Y_{s(1)} \cup X_{s(1)}) = \varnothing \quad \text{for all } i = 2, \ldots, m. \tag{7}$$

The proof of sure compatibility is by induction on the number $m \geq 2$ of variable pairs.

First step: For $m = 2$, let any two variable pairs $(Y_{s(1)}|X_{s(1)}) = (Y|X)$ and $(Y_{s(2)}|X_{s(2)}) = (V|U)$ be given such that $V \cap (Y \cup X) = \varnothing$, that is $FG = \varnothing$ in the terms of Figure 1, so that $(Y|X) \vee (V|U) = (ABCDE|H)$ according to Equation (1) (here and in the rest of this proof the symbol $\cup$ is omitted for simplicity, so that FG and $ABCDE$ stand for $F \cup G$ and $A \cup B \cup C \cup D \cup E$, respectively). Let $p(Y|X) = p(AE|CH)$ and $p(V|U) = p(B|DEH)$ be arbitrary kernels on the variable pairs. First, we extend $p(B|DEH)$ into $p(B|ACDEH)$ by setting

$$p(B|a, c, DEH) = p(B|DEH), \text{ for all } (a, c) \in A^\circ \times C^\circ. \tag{8}$$

Then by multiple promotion we can construct this kernel

$$p(ABCDE|H) = p(B|ACDEH) \; M \; p(D|ACEH) \; M \; p(AE|CH) \; M \; p(C|H)$$

where $p(D|ACEH)$ and $p(C|H)$ are freely chosen kernels (this writing takes account of the associativity of the promotion operation, as noted in Equation (6)). The kernel $p(ABCDE|H)$ thus constructed dominates $p(AE|CH)$ due to parts (ii) and (iii) of Lemma 2, and dominates $p(B|ACDEH)$ due to part (ii) of that lemma. Hence, it also dominates $p(B|DEH)$ on account of Equation (8) and of Lemma 1. The kernels $p(Y|X) = p(AE|CH)$ and $p(V|U) = p(B|DEH)$ (which are arbitrary) are therefore compatible with each other, as there is a kernel $p(ABCDE|H)$ on the join variable pair $(Y|X) \vee (V|U) = (ABCDE|H)$ that dominates both of them. Note that $p(ABCDE|H)$ does not involve any variables besides those involved in $p(AE|CH)$ or in $p(B|DEH)$.

Inductive step: Let us now consider any list

$$(p_{s(1)}(Y_{s(1)}|X_{s(1)}), \ldots, p_{s(m-1)}(Y_{s(m-1)}|X_{s(m-1)}), p_{s(m)}(Y_{s(m)}|X_{s(m)}))$$

of $m > 2$ kernels whose variable pairs comply with condition (7), and suppose (as an inductive hypothesis) that the first $m - 1$ members in the list are compatible with one another, so that there is a kernel $p(Z|W)$ that dominates all of them. Based on the remark that concludes the preceding step in the current proof, we may presume that $Z \cup W \subseteq Y_{s(m-1)} \cup X_{s(m-1)} \cup \cdots \cup Y_{s(1)} \cup X_{s(1)}$, so that $Y_{s(m)} \cap (Z \cup W) = \varnothing$ due to hypothesis (7). Thus, the conditions are satisfied that make it possible to apply the argument in the preceding step to the kernels $p(Z|W)$ and $p_{s(m)}(Y_{s(m)}|X_{s(m)})$. This argument ensures the existence of a kernel that dominates both $p(Z|W)$ (and hence

$$p_{s(1)}(Y_{s(1)}|X_{s(1)}), \ldots, p_{s(m-1)}(Y_{s(m-1)}|X_{s(m-1)})$$

by transitivity) and $p_{s(m)}(Y_{s(m)}|X_{s(m)})$. The m kernels are therefore all compatible with one another. As these are arbitrary, this proves that the given set of variable pairs has sure compatibility. □

The theorem thus proved characterizes the sure compatibility of a set of variable pairs by referring only to the set-theoretic properties of those pairs, rather than to the numerical properties of the probability kernels definable on them. The "if" part of the theorem ensures that if a given set of pairs satisfies the set-theoretic conditions (i) and (ii), then no further test is required to accept the compatibility

hypothesis of any set of kernels acting on those pairs. The "only if" part signifies, complementarily, that if conditions (i) and (ii) are not both satisfied, then the acceptance (or refusal) of the compatibility hypothesis requires further tests of the numerical characteristics of the kernels under judgement. Let us consider, for example, the two schemes presented in Lemma 3. We can see that the cumulative scheme (e.g., $\{(Y_1|\varnothing), (Y_2|Y_1), (Y_3|Y_1 \cup Y_2)\}$ for $m = 3$) complies with conditions (i) and (ii), so that arbitrary kernels defined on the variable pairs in the scheme are mutually compatible. From Lemma 3(i), there is a single consensus density for the given kernels, which can be constructed from these by multiple promotion. On the contrary, the alternating scheme (e.g., $\{(Y_1|Y_2 \cup Y_3), (Y_2|Y_1 \cup Y_3), (Y_3|Y_1 \cup Y_2)\}$ for $m = 3$) violates condition (ii) (indeed, the incidence relation within that scheme forms a complete directed graph, thus containing cycles) so that compatibility is not generally ensured for kernels definable on the variable pairs in the scheme. From Lemma 3(ii), if kernels defined on the variable pairs in an alternating scheme admit a consensus density whose support is factorable, then this density is unique and can be constructed from the given kernels through lightening-and-promotion operations.

4. Compatibility beyond structural assurance: the two kernels case

In light of the preceding discussion, we can expect that the situations explored in research on the compatibility of distributions are those in which the two conditions in Theorem 1 are not both satisfied, so that compatibility is not structurally guaranteed. The simplest of these situations involves a set $\{p(Y|X), q(X|Y)\}$ of two kernels on *symmetric variable pairs*. The pairs $(Y|X)$ and $(X|Y)$ form a cycle (of length two) according to the incidence relation (Definition 8), and thus they falsify condition (ii) of Theorem 1. In the first half of this section, we review some results from the literature that characterize compatibility within a pair of kernels on symmetric variable pairs. We review them from the standpoint defined in the preceding section and, for simplicity, limit ourselves to results applicable to kernels $p(Y|X)$ and $q(X|Y)$ that satisfy the following *positivity condition* (see footnote 3):

$$p(y|x) > 0 \text{ and } q(x|y) > 0 \text{ for all } (x, y) \in X^\circ \times Y^\circ. \tag{9}$$

In the literature, however, there are also generalizations of these results that apply to kernels satisfying the following, less restrictive condition:

$$p(y|x) > 0 \text{ if and only if } q(x|y) > 0 \text{ for all } (x, y) \in X^\circ \times Y^\circ. \tag{10}$$

Note that this condition is necessary for compatibility, because if kernels $p(Y|X)$ and $q(X|Y)$ are both dominated by a density $r(X \cup Y)$ such that $S(r(X)) = X^\circ$ and $S(r(Y)) = Y^\circ$, then, according to Definition 3(ii), $p(y|x) = r(x, y)/r(x)$ and $q(x|y) = r(x, y)/r(y)$ for all $(x, y) \in X^\circ \times Y^\circ$, so that $p(y|x)$ and $q(x|y)$ are positive precisely when $r(x, y)$ is positive. In the second half of the current section, we will present a result of our own analysis, which characterizes compatibility between kernels on arbitrary variable pairs $(Y|X)$ and $(V|U)$, thus going beyond the special case of symmetric variable pairs.

One characterization is expressed by the following statement (Arnold and Press, 1989).

Proposition 2. *Two kernels $p(Y|X)$ and $q(X|Y)$ on symmetric variable pairs are compatible if and only if there are densities $p(X)$ and $q(Y)$ such that $p(y|x) \cdot p(x) = q(x|y) \cdot q(y)$ for all $(x, y) \in X^\circ \times Y^\circ$.*

In the notation used in Definition 5(i), the equation in this proposition can be rewritten as

$$M[p(Y|X),\, p(X)] = M[q(X|Y),\, q(Y)]. \tag{11}$$

The truth of the proposition is clarified by noting that the existence of densities $p(X)$ and $q(Y)$ that satisfy equation (11) is tantamount to the existence of a common upper bound $p(X \cup Y) = q(X \cup Y)$ for $p(Y|X)$ and $q(X|Y)$ within a lattice of kernels, which is precisely the meaning of compatibility between the given kernels. The stated characterization is of *existential* type, as it links the compatibility between $p(Y|X)$ and $q(X|Y)$ to the existence of a solution to Equation (11) in the unknowns $p(X)$ and $q(Y)$.

A second characterization involves two functions that are deducible from the kernels under consideration. Specifically, once a reference point $(x', y') \in X^\circ \times Y^\circ$ has been arbitrarily chosen, two functions $f(X, Y)$ and $g(X, Y)$ can be separately derived from the kernels $p(Y|X)$ and $q(X|Y)$ by setting, for each $(x, y) \in X^\circ \times Y^\circ$,

$$f(x, y) = \frac{p(y|x) \cdot p(y'|x')}{p(y'|x) \cdot p(y|x')}, \quad g(x, y) = \frac{q(x|y) \cdot q(x'|y')}{q(x'|y) \cdot q(x|y')}.$$

In other words, $f(X, Y)$ is obtained as an odd-ratio function based on $p(Y|X)$, and similarly $g(X, Y)$ from $q(X|Y)$ (all ratios exist as real numbers, under the positivity condition (9)). In these terms, the following relationship is true (Arnold and Press, 1989, p. 52; Chen, 2010, p. 672).

Proposition 3. *Kernels $p(Y|X)$ and $q(X|Y)$ are compatible if and only if $f(X,Y)=g(X,Y)$.*

The truth of the "only if" part is easily seen: if $p(Y|X)$ and $q(X|Y)$ are dominated by the same density $r(X \cup Y)$, then $p(y|x) = r(x, y)/r(x)$ and $q(x|y) = r(x, y)/r(y)$, so that $f(x, y) = r(x, y) \cdot r(x', y')/r(x', y) \cdot r(x, y') = g(x, y)$ for all $(x, y) \in X^\circ \times Y^\circ$. The given characterization is of *deductive* type: it expresses compatibility in terms of the equality between two functions $f(X, Y)$ and $g(X, Y)$ that are deducible from $p(Y|X)$ and $q(X|Y)$ in the described way.

A third characterization involves another function that is deducible from the kernels in question. Specifically, based on $p(Y|X)$ and $q(X|Y)$, a ratio function $h(X, Y)$ can be constructed by setting, for all $(x, y) \in X^\circ \times Y^\circ$,

$$h(x, y) = \frac{p(y|x)}{q(x|y)}$$

which again is a legitimate computation under the positivity condition. The following statement holds true (Arnold and Press, 1989, pp. 152–153; Tian, Tan, Ng and Tang, 2009, p. 119):

Proposition 4. *Kernels $p(Y|X)$ and $q(X|Y)$ are compatible if and only if there are functions $a(X)$ and $b(Y)$ such that $h(x, y) = a(x) \cdot b(y)$ for all $(x, y) \in X^\circ \times Y^\circ$.*

As above, the "only if" part is easily proved: if $p(Y|X)$ and $q(X|Y)$ are dominated by the same density $r(X \cup Y)$, then $h(x, y) = p(y|x)/q(x|y) = (r(x, y)/r(x))/(r(x, y)/r(y)) = (1/r(x)) \cdot r(y)$ for all $(x, y) \in X^\circ \times Y^\circ$, so that by setting $a(X) = 1/r(X)$ and $b(Y) = r(Y)$ a factorization of $h(X, Y)$ in the asserted form is obtained.

In particular, if sets X° and Y° have finite cardinality, then the function $h(X, Y)$ can be represented as a matrix of $|X^\circ|$ rows and $|Y^\circ|$ columns. The equation $h(X, Y) = a(X) \cdot b(Y)$ would then mean that

this matrix is expressible as the product of a column vector $a(X)$ by a row vector $b(Y)$, implying that all rows in the matrix are proportional to one another (and similarly for the columns). As a consequence, when referring to variables with finite sets of possible values, the above connection can be reformulated as follows (Arnold, Castillo and Sarabia, 2004, p. 137; Kuo, Song and Jiang, 2017, pp. 117–118).

Proposition 5. *Kernels $p(Y|X)$ and $q(X|Y)$ are compatible if and only if the matrix $h(X, Y)$ has rank* 1.

Note that, although the general characterization in Proposition 4 is of the existential type (it demands the existence of functions $a(X)$ and $b(Y)$ satisfying a definite equation), the specific version in Proposition 5 is of the deductive type, as it concerns a possible property (unit rank) of the matrix $h(X, Y)$ that is deducible from the given kernels. We also remark that, besides this elementary result, there are other ways in which linear algebra and associated geometrical arguments have proved of use in discussing compatibility of distributions. For example, Arnold, Castillo, and Sarabia (2002, pp. 235–239) on considering any pair of kernels $\{p(Y|X), q(X|Y)\}$ that comply with (10) but may violate (9), show how the search for a consensus distribution $r(X \cup Y)$ can be formalized as the task of solving a definite system of linear equations and inequalities. Thus, the set of possible solutions is tantamount to a convex subset of a geometric space and may be explored using methods of linear programming.

The characterizations reviewed so far are limited in scope, as they apply only to any pair of kernels $\{p(Y|X), q(X|Y)\}$ on symmetric variable pairs. With the next theorem, we contribute to the topic of compatibility by presenting a characterization that is applicable to any pair of kernels $\{p(V|U), q(X|Y)\}$ that are free of constraints on the variable pairs. In stating and proving this theorem, use will be made of the set-theoretic labeling represented in Figure 1 and the basic operations on kernels defined in Section 2. As in the proof of Theorem 1, the symbol $\cup$ will be omitted for brevity when specifying composite variables (e.g., BFG stands for $B \cup F \cup G$).

Theorem 2. *Any two kernels $p(V|U) = p(BFG|DEH)$ and $q(Y|X) = q(AEF|CGH)$ on variable pairs in a full variable T are compatible with each other if and only if there are kernels $p(DE|H)$ and $q(CG|H)$ that satisfy the equation*

$$J[M[p(BFG|DEH), p(DE|H)], BD] = J[M[q(AEF|CGH), q(CG|H)], AC], \tag{12}$$

where J and M are the projection and promotion operations.

Proof. "Only if" part. If the given kernels are compatible, then there is a consensus full density $r(T)$ for them, such that $p(BFG|DEH) = r(BFG|DEH)$ and $q(AEF|CGH) = r(AEF|CGH)$. By setting $p(DE|H) = r(DE|H)$ and $q(CG|H) = r(CG|H)$, we find that both sides of Equation (12) specify the kernel $r(EFG|H)$, so that the equation is satisfied.

"If" part. Let $p(V|U) = p(BFG|DEH)$ and $q(Y|X) = q(AEF|CGH)$ be arbitrary kernels on the indicated variable pairs, and suppose that there are kernels $p(DE|H)$ and $q(CG|H)$ that when combined with them satisfy Equation (12). By promotion, we can first construct the kernels

$$p(BDEFG|H) = M[p(BFG|DEH), p(DE|H)], \tag{13}$$

$$q(ACEFG|H) = M[q(AFE|CGH), q(CG|H)], \tag{14}$$

from which we may derive the following further kernels by conditioning

$$p(BD|EFGH) = C[p(BDEFG|H), EFG], \tag{15}$$

$$q(AC|EFGH) = C[q(ACEFG|H), EFG]. \tag{16}$$

Hypothesis (12) means that the projection (relative to BD) of the kernel defined in (13) equals the projection (relative to AC) of the kernel defined in (14), so that the same symbol $r(EFG|H)$ may be used for both projections:

$$r(EFG|H) = J[p(BDEFG|H), BD] = J[q(ACEFG|H), AC]. \tag{17}$$

Furthermore, the kernels defined in (15) and (16) act on variable pairs that have the same conditioning variable $EFGH$ and disjoint conditioned variables BD and AC. Thus, by Theorem 1 the pair of such variable pairs has sure compatibility, implying that there exists some kernel $r(ABCD|EFGH)$ that dominates both kernels. More precisely

$$p(BD|EFGH) = J[r(ABCD|EFGH), AC], \tag{18}$$

$$q(AC|EFGH) = J[r(ABCD|EFGH), BD]. \tag{19}$$

Lastly, the kernels $r(ABCD|EFGH)$ and $r(EFG|H)$ are suitable for promotion, thus producing the result

$$r(ABCDEFG|H) = M[r(ABCD|EFGH), r(EFG|H)]. \tag{20}$$

We now prove that this kernel (which acts on the join variable pair $(ABCDEFG|H) = (V|U) \vee (Y|X)$) dominates both $p(V|U)$ and $q(Y|X)$, so that these are compatible. Indeed:

$C[J[r(ABCDEFG|H), AC], DE] =$ by (20)

$C[J[M[r(ABCD|EFGH), r(EFG|H)], AC], DE] =$ by Lemma 2(iv)

$C[M[J[r(ABCD|EFGH), AC], r(EFG|H)], DE] =$ by (18)

$C[M[p(BD|EFGH), r(EFG|H)], DE] =$ by (15) and (17)

$C[M[C[p(BDEFG|H), EFG], J[p(BDEFG|H), BD]], DE] =$ by Lemma 2(vi)

$C[p(BDEFG|H), DE] =$ by Definition 3(ii)

$p(BFG|DEH) = p(V|U),$

so that $r(ABCDEFG|H)$ dominates $p(V|U)$. The dominance of $r(ABCDEFG|H)$ over $q(Y|X)$ is proved by a similar argument. □

The characterization in Theorem 2 is of the existential type, as it demands the existence (and the discovery in actual applications) of kernels $p(DE|H)$ and $q(CG|H)$ such that when combined with the given kernels $p(V|U)$ and $q(Y|X)$ through promotion-and-projection, Equation (12) is verified. From Theorem 2, several corollaries may be deduced by setting constraints on the kernels $p(V|U)$ and $q(Y|X)$

under consideration, more precisely by assuming that certain parts of the variables they involve are empty. For example, if $ABCDFH = \varnothing$, then Equation (12) becomes

$$J[M[p(G|E), p(E|\varnothing)], \varnothing] = J[M[q(E|G), q(G|\varnothing)], \varnothing],$$

that is, due to (3),

$$M[p(G|E), p(E)] = M[q(E|G), q(G)],$$

which is a rewriting of (11). Hence, the characterization in Proposition 2 amounts to a special case of the characterization in Theorem 2. As another example, if $CDEG = \varnothing$, then Equation (12) becomes

$$J[M[p(BF|H), p(\varnothing|H)], B] = J[M[q(AF|H), q(\varnothing|H)], A],$$

that is, due to (5),

$$J[p(BF|H), B] = J[q(AF|H), A].$$

This formally corroborates the following intuitive principle: any two kernels with the same conditioning variable and partially overlapping conditioned variables are compatible if and only if their projections on the intersection of the conditioned variables are equal.

5. Compatibility beyond structural assurance: the multiple kernels case

A natural generalization of the case discussed in the first half of the preceding section (that is, a pair of kernels $\{p(Y|X), q(X|Y)\}$ on symmetric variable pairs) is given by any set $\{p_1(Y_1|X_1), \ldots, p_m(Y_m|X_m)\}$ of *saturated* kernels whose conditioned variables are *exhaustive* of the full variable T. A notable example of this is the alternating scheme represented in Lemma 3(ii), in which the conditioned variables $Y_1, \ldots, Y_m$ more precisely form a partition of T. In this section, we then refer to any set of kernels $\{p_1(Y_1|X_1), \ldots, p_m(Y_m|X_m)\}$ whose variable pairs comply with these conditions:

$$X_i = T \setminus Y_i \text{ for each } i = 1, \ldots, m \text{ (the kernels are saturated);} \tag{21}$$

$$Y_1 \cup \cdots \cup Y_m = T \text{ (the conditioned variables are exhaustive).} \tag{22}$$

Note that within a set of variable pairs with these properties, the incidence relation $\rightarrow$ in Definition 8 can give rise to cycles (for example, within an alternating scheme, it determines a complete directed graph that obviously has cycles), so that based on Theorem 1, such a set of variable pairs could fail to have sure compatibility. In that situation, a decision concerning the compatibility between given kernels should then be taken by examining the numerical properties of the kernels themselves, as families of density functions. Here, we review an exemplary decision criterion limited to densities on finite domains, a criterion that has been variously studied in the literature.

Let $T = \{T_1, \ldots, T_n\}$ be a full variable whose elements are variables with *finite* sets of possible values. For any point t in the space T° and any sub-variable U of T, let t_U denote the *projection* of the point t

on the space U°. In formal terms:

> for any $t = (t_1, \ldots, t_n) \in T_1^\circ \times \cdots \times T_n^\circ = T^\circ$ and any $U = \{T_{g(1)}, \ldots, T_{g(k)}\} \subseteq T$
> t_U stands for $(t_{g(1)}, \ldots, t_{g(k)})$.

For example, if $T = \{T_1, T_2, T_3, T_4, T_5\}$, $U = \{T_2, T_3, T_5\}$, and $t = (2, 3, 1, 3, 2)$, then $t_U = (3, 1, 2)$. Using this notation and referring to any set $\{p_1(Y_1|X_1), \ldots, p_m(Y_m|X_m)\}$ of saturated kernels, we first remark that if these kernels are compatible, then the following condition must be true:

$$\begin{aligned} &\text{for all } t \in T^\circ \text{ and all } 1 \leq i, j \leq m \\ &p_i(t_{Y_i}|t_{X_i}) > 0 \text{ if and only if } p_j(t_{Y_j}|t_{X_j}) > 0. \end{aligned} \tag{23}$$

This condition generalizes requirement (10), and its necessity for compatibility can be proved by an argument similar to that used for that requirement. Our discussion in this section, however, is focused on sets of kernels that satisfy the *positivity condition*

$$p_i(y_i|x_i) > 0 \quad \text{for all } x_i \in X_i^\circ,\ y_i \in Y_i^\circ, \text{ and } i = 1, \ldots, m, \tag{24}$$

which is stronger than (23) and in turn generalizes (9). After presenting the main result, in the last paragraph we will comment on the complications that may arise when the kernels comply with (23) but not with (24), that is, when there are "structural zeros" in the kernels under consideration.

In the assumed conditions, for each $i = 1, \ldots, m$ an *adjacency* relation E_i within the space T° can be determined by setting

$$E_i = \{(s, i, t) : s, t \in T^\circ, s_{X_i} = t_{X_i}\}. \tag{25}$$

In other words, any two points $s = (s_1, \ldots, s_n)$ and $t = (t_1, \ldots, t_n)$ in the space T° are adjacent according to E_i if they coincide in all the coordinates for the elementary variables in X_i (and thus may only differ in some of the coordinates for the elementary variables in $Y_i = T \setminus X_i$). For example, suppose $T = \{T_1, \ldots, T_5\}$, $T^\circ = \{1, 2, 3\}^5$, and $X_i = \{T_2, T_3\}$, and consider the points $s = (1, 3, 1, 2, 2)$, $t = (2, 3, 1, 3, 2)$, and $u = (1, 2, 1, 2, 2)$. Then $(s, i, t) \in E_i$ because $s_{X_i} = (3, 1) = t_{X_i}$, whereas $(s, i, u) \notin E_i$ because $s_{X_i} = (3, 1) \neq (2, 1) = u_{X_i}$. Overall, we can then consider a relational structure

$$(T^\circ, E) = (T^\circ, E_1 \cup \cdots \cup E_m)$$

which formally amounts to a graph with the space T° as the set of points and the relation $E = E_1 \cup \cdots \cup E_m$ as the set of lines.

We note the following properties of this graphical structure. First, the lines in the graph are *labeled* and *directed*. Indeed, each line (s, i, t) has the label i, which indicates the adjacency E_i to which the line belongs, and it is counted as distinct from the inverse line (t, i, s), which also belongs to E_i. Secondly, each adjacency E_i has the formal properties of an *equivalence* (reflexivity, symmetry, and transitivity). However, the pooled adjacency $E = E_1 \cup \cdots \cup E_m$ may fail to be transitive because for any points s, t, and u, the existence of some E_i and E_j such that $(s, i, t) \in E_i$ and $(t, j, u) \in E_j$ does not ensure the existence of some E_h such that $(s, h, u) \in E_h$. Thirdly, any two points in T° may have *multiple*

adjacency. For example, referring to the case mentioned in the preceding paragraph with $s = (1, 3, 1, 2, 2)$ and $t = (2, 3, 1, 3, 2)$, and assuming $X_i = \{T_2, T_3\}$ and $X_j = \{T_3, T_5\}$, then both $(s, i, t) \in E_i$ (since $s_{X_i} = (3, 1) = t_{X_i}$) and $(s, j, t) \in E_j$ (since $s_{X_j} = (1, 2) = t_{X_j}$). Fourthly, the description of any *walk* within the graph has the following generic form

$$(t^0, i(1), t^1, i(2), t^2, \ldots, t^{k-1}, i(k), t^k)$$

which records not only the points $t^0, t^1, \ldots, t^k$ touched on by the walk, but also the adjacencies $E_{i(1)}, \ldots, E_{i(k)}$ used in passing from point to point. For example, assuming $X_i = \{T_2, T_3\}$, $X_j = \{T_3, T_5\}$, and $X_h = \{T_1, T_4, T_5\}$, the following expressions describe two different walks with the same initial and terminal points:

$$((2, 1, 3, 2, 3), j, (2, 2, 3, 1, 3), h, (2, 3, 1, 1, 3)),$$
$$((2, 1, 3, 2, 3), h, (2, 3, 1, 2, 3), i, (1, 3, 1, 2, 3), j, (2, 3, 1, 1, 3)).$$

Lastly, the assumption (22) ensures that the graph is *connected*. Indeed, the fact that the conditioned variables $Y_1, \ldots, Y_m$ (on which any difference in coordinates is permitted) exhaust the full variable T allows us to transform any point s into any other point t through a sequence of changes each of which preserves adjacency.

The graphical structure described thus far is only determined by the set $\{X_1, \ldots, X_m\}$ of the conditioning variables in the assumed set of kernels $\{p_1(Y_1|X_1), \ldots, p_m(Y_m|X_m)\}$. As families of density functions, these kernels allow us to endow that structure with a *valuation function*. Specifically, let us consider any kernel $p_i(Y_i|X_i)$ in the set, the corresponding adjacency E_i (determined by X_i according to (25)), and any line (s, i, t) belonging to E_i (so that the projections s_{X_i} and t_{X_i} are equal and are a point in X_i°, whereas the projections s_{Y_i} and t_{Y_i} are possibly different points in Y_i°). The kernel $p_i(Y_i|X_i)$ provides definite values $p_i(s_{Y_i}|s_{X_i})$ and $p_i(t_{Y_i}|t_{X_i})$, which under the positivity condition (24) are both positive real numbers. Thus, they may be combined by division to obtain a positive value associated with the line in question:

$$R(s, i, t) = \frac{p_i(t_{Y_i}|t_{X_i})}{p_i(s_{Y_i}|s_{X_i})}. \tag{26}$$

This value may be interpreted as an odds quantity, being the ratio between the probabilities of two events $Y_i = t_{Y_i}$ and $Y_i = s_{Y_i}$ concerning the variable Y_i, both conditional on the event $X_i = t_{X_i} = s_{X_i}$ concerning the variable X_i. By applying this method in relation to each kernel $p_i(Y_i|X_i)$ in the given set and each line (s, i, t) in the corresponding adjacency E_i, a positive-valued function R on the pooled adjacency E is generated that upgrades the graphical structure in the following form:

$$(T^\circ, E, R) = (T^\circ, E_1 \cup \cdots \cup E_m, R).$$

In graph-theoretic terms, this is a line-valued directed multi-graph (Yao, Chen and Wang, 2014, p. 2).

The valuation R, which is first defined on single lines in the graph, can be consistently extended to any

walk by setting

$$R(t^0, i(1), t^1, i(2), t^2, \ldots, t^{k-1}, i(k), t^k) = \tag{27}$$
$$R(t^0, i(1), t^1) \cdot R(t^1, i(2), t^2) \cdots R(t^{k-1}, i(k), t^k).$$

Under condition (24), all factors in this product exist as positive real numbers, meaning that the product itself is a positive real number. The characterization of compatibility expressed in the next theorem specifically refers to the values that may result from this multiplicative formula (a product of odds).

Theorem 3. *Let* $\{p_1(Y_1|X_1), \ldots, p_m(Y_m|X_m)\}$ *be a set of kernels that satisfy the conditions of saturation* (21), *exhaustiveness* (22), *and positivity* (24), *and let* (T°, E, R) *be the line-valued directed graph that can be constructed based on the kernels in the way described above. The kernels are mutually compatible if and only if the valuation* R *assigns the value* 1 *to every closed walk in the graph.*

Proof. "Only if" part. Suppose that the saturated kernels $\{p_1(Y_1|X_1), \ldots, p_m(Y_m|X_m)\}$ are mutually compatible, that is, there exists a full density $r(T)$ such that $p_i(Y_i|X_i) = r(Y_i|X_i)$ for all $i = 1, \ldots, m$, which means

$$p_i(t_{Y_i}|t_{X_i}) = r(t_{Y_i}|t_{X_i}) = \frac{r(t)}{r(t_{X_i})} \text{ for all } t \in T^\circ. \tag{28}$$

If $(t^0, i(1), t^1, i(2), t^2, \ldots, t^{k-1}, i(k), t^0)$ is any closed walk in the graph (T°, E, R) (note the term t^0 in the place of t^k), then

$$\begin{aligned}
R(t^0, i(1), t^1, \ldots, t^{k-1}, i(k), t^0) &= R(t^0, i(1), t^1) \cdots R(t^{k-1}, i(k), t^0) \\
&= \frac{p_{i(1)}(t^1_{Y_{i(1)}}|t^1_{X_{i(1)}})}{p_{i(1)}(t^0_{Y_{i(1)}}|t^0_{X_{i(1)}})} \cdots \frac{p_{i(k)}(t^0_{Y_{i(k)}}|t^0_{X_{i(k)}})}{p_{i(k)}(t^{k-1}_{Y_{i(k)}}|t^{k-1}_{X_{i(k)}})} \\
&= \frac{r(t^1)/r(t^1_{X_{i(1)}})}{r(t^0)/r(t^0_{X_{i(1)}})} \cdots \frac{r(t^0)/r(t^0_{X_{i(k)}})}{r(t^{k-1})/r(t^{k-1}_{X_{i(k)}})} \\
&= \frac{r(t^1)}{r(t^0)} \cdots \frac{r(t^0)}{r(t^{k-1})} = \frac{r(t^1) \cdots r(t^{k-1}) \cdot r(t^0)}{r(t^0) \cdot r(t^1) \cdots r(t^{k-1})} = 1
\end{aligned}$$

where the first three equalities are justified by (27), (26), and (28), respectively, and the fourth is derived from the fact that $(t^h, i(h+1), t^{h+1}) \in E_{i(h+1)}$ for every $h = 0, \ldots, k-1$, meaning that $t^h_{X_{i(h+1)}} = t^{h+1}_{X_{i(h+1)}}$.

"If" part. Let $\{p_1(Y_1|X_1), \ldots, p_m(Y_m|X_m)\}$ be a set of saturated kernels such that in the resulting graph (T°, E, R) the valuation R assigns value 1 to every closed walk. First note that definition (26) implies $R(s, i, t) = 1/R(t, i, s)$ for each pair $\{(s, i, t), (t, i, s)\}$ of symmetric lines. In turn, this implies that our assumption concerning R can be reformulated as follows:

$$\begin{aligned}
&\text{for all walks } (t^0, i(1), t^1, \ldots, t^{k-1}, i(k), t^k) \text{ and } (u^0, j(1), u^1, \ldots, u^{h-1}, j(h), u^h) \\
&\text{if } t^0 = u^0 \text{ and } t^k = u^h \\
&\text{then } R(t^0, i(1), t^1, \ldots, t^{k-1}, i(k), t^k) = R(u^0, j(1), u^1, \ldots, u^{h-1}, j(h), u^h).
\end{aligned} \tag{29}$$

Consider the following constructive procedure: (arbitrarily) choose a reference point o in the space T°; construct a function f on T° by setting

$$f(t) \text{ equal to the value assigned by } R \text{ to any walk from } o \text{ to } t, \text{ for all } t \in T^\circ; \tag{30}$$

and then define

$$r(t) = \frac{f(t)}{\sum_{s \in T^\circ} f(s)} \text{ for all } t \in T^\circ. \tag{31}$$

The connectedness of the graph (ensured by (22)), the positivity assumption (24), and the hypothesis (29) imply that f is a well-defined positive-valued function over the space T° (in particular, hypothesis (29) ensures that for each $t \in T^\circ$, the value $f(t)$ is invariant relative to the available walks from o to t). Hence, the function $r(T)$ specified by (31) is a well-defined full density over the space T°. We will now show that $r(T)$ is indeed a consensus density for the given kernels, that is

$$p_i(t_{Y_i}|t_{X_i}) = r(t_{Y_i}|t_{X_i}) = \frac{r(t)}{r(t_{X_i})} \text{ for all } t \in T^\circ \text{ and all } i = 1, \ldots, m \tag{32}$$

which allows us to conclude that the given kernels are mutually compatible. Consider any $i = 1, \ldots, m$ and any $t \in T^\circ$, and define

$$T^\circ|(i,t) = \{s \in T^\circ : (t,i,s) \in E_i\} = \{s \in T^\circ : t_{X_i} = s_{X_i}\}$$

so that

$$\sum_{s \in T^\circ|(i,t)} p_i(s_{Y_i}|s_{X_i}) = 1 \quad \text{and} \quad \sum_{s \in T^\circ|(i,t)} r(s) = r(t_{X_i}).$$

If $W = (o, i(1), w^1, \ldots, w^{k-1}, i(k), t)$ is any walk from o to t, then $f(t) = R(W)$, and for each $s \in T^\circ|(i,t)$ the list $(o, i(1), w^1, \ldots, w^{k-1}, i(k), t, i, s)$ describes a walk from o to s, so that from (26), (27), and (30)

$$f(s) = R(W) \cdot \frac{p_i(s_{Y_i}|s_{X_i})}{p_i(t_{Y_i}|t_{X_i})}.$$

Therefore

$$\frac{r(t)}{r(t_{X_i})} = \frac{r(t)}{\sum\limits_{s \in T^\circ|(i,t)} r(s)} = \frac{f(t)}{\sum\limits_{s \in T^\circ|(i,t)} f(s)} = \frac{R(W)}{\sum\limits_{s \in T^\circ|(i,t)} R(W) \cdot \dfrac{p_i(s_{Y_i}|s_{X_i})}{p_i(t_{Y_i}|t_{X_i})}} = \frac{p_i(t_{Y_i}|t_{X_i})}{\sum\limits_{s \in T^\circ|(i,t)} p_i(s_{Y_i}|s_{X_i})} = p_i(t_{Y_i}|t_{X_i}),$$

which verifies Equation (32). □

A key idea underlying the proof of Theorem 3 is that the *product* of a suitable *chain of ratios* between *conditional* probabilities (that is, a product of odds) equals the *ratio* between two *joint* probabilities associated with the first and the last links in the chain. The paper generally cited as the source of this idea is Besag (1974), where this odds-product method is applied to problems of spatial statistics. The same idea occurs as a crucial principle in several studies concerning compatibility of distributions, although from one study to another there may be differences in the mathematical context in which it is embedded and the form in which it is expressed (Gurevich, 1992, pp. 373–374; Cressie, 1993, pp. 412–414; Hobert

and Casella, 1998, pp. 48–49; Slavkovic and Sullivant, 2006, pp. 198 and 206; Yao, Chen and Wang, 2014). In particular, Slavkovic and Sullivant (2006) address the compatibility problem using tools from the algebra of polynomials, show that the compatibility between any two or more saturated probability kernels can be characterized in terms of a system of binomial equations, and come to conclude that for any fixed combination $\{(Y_1|X_1), \ldots, (Y_m|X_m)\}$ of saturated variable pairs, the set of all combinations $\{p_1(Y_1|X_1), \ldots, p_m(Y_m|X_m)\}$ of kernels that are mutually compatible (within each combination) is a "unimodular toric variety". The algebraic method applied in the cited study is different from the graph-theoretic view we took in preparing and proving Theorem 3 above. Nevertheless, there is a common root in both approaches, which may be recognized by comparing the "circuits" used in constructing the required binomial equations with the "closed walks" mentioned in Theorem 3, and by considering this simple fact: if $(v_1, v_2, v_3, v_4, \ldots, v_{r-1}, v_r)$ is a list of an even number of non-null values ("indeterminates" in polynomials), then the binomial equation $v_1 v_3 \cdots v_{r-1} - v_2 v_4 \cdots v_r = 0$ is equivalent to the equation $(v_1/v_2)(v_3/v_4) \cdots (v_{r-1}/v_r) = 1$ concerning a product of ratios. Ultimately, that common root is related to the Besag's (1974) idea of a consistent chain of pairs of conditional probabilities mentioned above. The Slavkovic and Sullivant's (2006) method has the additional merit of wider generality: it can also be applied to saturated kernels that (while satisfying the necessary condition (23)) may violate the condition (24) of positivity.

One advantage of the characterization in Theorem 3 (also shared by that in Slavkovic and Sullivant, 2006) is that it is of deductive type: the mutual compatibility of the given kernels $\{p_1(Y_1|X_1), \ldots, p_m(Y_m|X_m)\}$ can be decided through a series of tests on the graphical structure (T°, E, R), which is directly deducible from the kernels in question. A limitation is that it is applicable only to saturated kernels, or more generally to kernels in which the united variable $Y_i \cup X_i$ is the same for all $i = 1, \ldots, m$. Another limitation is its expensiveness, as the acceptance of the compatibility hypothesis requires a test (with positive result) on each cycle within the graph (T°, E, R). Algorithms may be devised to simplify this testing process by exploiting redundancies implicit in the graph (Wang and Kuo, 2010; Kuo and Wang, 2011; Yao, Chen and Wang, 2014). Lastly we remark that, of the cases covered by Theorem 3, special notice should be given to the case in which the conditioned variables $Y_1, \ldots, Y_m$ in the kernels, in addition to being exhaustive, are also mutually disjoint, so that they form a partition of the full variable T – this is the "alternating scheme" focused on by Lemma 3(ii). In this case, a formal simplification is available, since if $Y_i \cap Y_j = \varnothing$, $X_i = T \setminus Y_i$, and $X_j = T \setminus Y_j$, then for no points $s \neq t$ in T° can we have both $s_{X_i} = t_{X_i}$ and $s_{X_j} = t_{X_j}$; hence, the directed graph (T°, E) is simple, that is, a graph with at most one line directed from a point s to another point t.

The proof of the "if" part of Theorem 3 implies that, under the stated conditions, the given kernels admit a single consensus density. Indeed, under those conditions, the graph (T°, E, R) is connected, so that the rule in (30) determines a single valuation $f(T)$ over T°, which in turn by (31) determines a single full density $r(T)$ dominating over the given kernels (this argument, if applied to an alternating scheme of kernels, provides a supplementary proof of Lemma 3(ii)). The connectedness of the graph and the positive valuation of all lines in it are ensured by assumptions (22) and (24). Now suppose that the kernels $\{p_1(Y_1|X_1), \ldots, p_m(Y_m|X_m)\}$ satisfy (22) (as well as (23)) but violate (24), so that there are points u in

T° such that $p_i(u_{Y_i}|u_{X_i}) = 0$ for some $i = 1, \ldots, m$. In this looser situation, the pooled adjacency E will contain lines (s, i, t) such that the ratio $R(s, i, t)$ as defined by (26) is either null or nonexistent as a real number; hence, these lines are of no use for the odds-product method and should be cancelled from E, which then degrades into a poorer adjacency G. Unlike E, this G could fail to be connected, so that the space T° would be divided into two or more connected components $(T^\circ)_1, \ldots, (T^\circ)_k$ according to G. If the graph satisfies the condition stated in Theorem 3 (concerning closed walks), then by the odds-product method we would still be able to uniquely construct densities $r_1(T), \ldots, r_k(T)$ separately defined on $(T^\circ)_1, \ldots, (T^\circ)_k$. These could be assembled into a full density $r(T)$ by choosing a set $(c_1, \ldots, c_k)$ of positive numbers with unit sum and then setting

$$r(t) = c_1 r'_1(t) + \cdots + c_k r'_k(t) \quad \text{for all } t \in T^\circ, \tag{33}$$

where (for $h = 1, \ldots, k$) we set $r'_h(t) = r_h(t)$ if $t \in (T^\circ)_h$ and $r'_h(t) = 0$ otherwise. Following the reasoning developed for Theorem 3 and considering that any two $(T^\circ)_h \neq (T^\circ)_{h'}$ have disjoint projections on each space X_i° (for $i = 1, \ldots, m$), it can be seen that a density $r(T)$ thus constructed dominates each of the given kernels, so that these are mutually compatible. The main point of this discussion is that if the diminished graph (T°, G) fails to be connected, then there are several (infinitely many) consensus densities that are constructible according to (33), simply because the set $(c_1, \ldots, c_k)$ may be chosen as any k-tuple of positive numbers with unit sum. Thus, the possible violation of the positivity condition (24) is detrimental not to the existence of a consensus density (which is still guaranteed by the condition on closed walks stated in Theorem 3), but to the uniqueness of such a density.

6. Concluding remarks: the varying saliency of the compatibility problem

In principle, as noted in the Introduction, compatibility of distributions is a basic requirement of any probabilistic model. Indeed, if the distributional assumptions in a model were not fully compatible with one another, then the data analyses guided by the model could in fact be directed towards a nonexistent ideal target, as there could be no global distribution that consistently encompasses all the local distributions postulated by the assumptions. In practice, however, the compatibility problem does not appear to have equal saliency for different kinds of probabilistic models. There are even models for which that problem may seem an idle question, since compatibility appears implicit in the basic structure of such models, regardless of the mathematical form of the distributions involved. In this concluding section, we will use some of the results of our study (in particular, those in Section 3) to illustrate the reasons for the different saliency of compatibility relative to different elementary kinds of probabilistic models.

First, we consider a simple model of classical statistics: the model for comparing the means of two normal populations under the assumption of equal variance. The full variable in the model is the set $T = \{T_{1,1}, \ldots, T_{1,n_1}, T_{2,1}, \ldots, T_{2,n_2}\}$, formed of a sample taken from one population and a sample taken from the other. In addition to the assumptions of stochastic independence within and between both samples, the model includes distributional assumptions, which are expressed by these assignments

$$T_{1,i} \sim \text{Normal}(\mu_1, \sigma^2) \text{ and } T_{2,j} \sim \text{Normal}(\mu_2, \sigma^2) \quad \text{for } i = 1, \ldots, n_1 \text{ and } j = 1, \ldots, n_2, \tag{34}$$

where μ_1, μ_2, and σ^2 are quantities that are unknown but (in the classical view) not treated as random variables. In the terms used in this study, the distributional assumptions are about the following marginal densities

$$p(T_{1,1}), \ldots, p(T_{1,n_1}), p(T_{2,1}), \ldots, p(T_{2,n_2})$$

which may be viewed as probability kernels on the variable pairs

$$(T_{1,1}|\varnothing), \ldots, (T_{1,n_1}|\varnothing), (T_{2,1}|\varnothing), \ldots, (T_{2,n_2}|\varnothing)$$

which in turn are atoms in the lattice $\widetilde{\mathcal{O}}(T)$. The (elementary) conditioned variables in the pairs are disjoint from one another, and the incidence relation $\rightarrow$ when referred to these pairs is acyclic (indeed, it is empty), so that from Theorem 1, the $n_1 + n_2$ kernels (or marginal densities) are compatible with one another for any values of μ_1, μ_2, and σ^2. Thus, distributional compatibility is assured here by the basic structure of the model (regardless of the form of the distributions) and this may explain why the compatibility question is not generally raised when presenting this and other similar models of classical statistics. Of course, there is another, more familiar way of reaching the same conclusion, that is, to observe that the product function $p(T) = p(T_{1,1}) \cdots p(T_{1,n_1}) \cdot p(T_{2,1}) \cdots p(T_{2,n_2})$ is certainly a consensus density for the $n_1 + n_2$ marginal densities, as each of these can be deduced from $p(T)$ by projection (operation J in Definition 3). Note that if the assumptions of stochastic independence are left out of the model, then the product function is not the only consensus density for the given marginal densities.

Our second example is a Bayesian expansion of the preceding classical model. Specifically, let us suppose that the parameters μ_1, μ_2, and σ^2 are themselves conceived of as random variables and that the system (34) becomes enriched by the following distributional assumptions:

$$\mu_1 \sim \text{Normal}(\nu, \tau^2), \quad \mu_2 \sim \text{Normal}(\nu, \tau^2), \quad \sigma^2 \sim \text{InverseGamma}(\alpha, \beta),$$
$$\nu \sim \text{Uniform}(0, 50), \quad \tau^2 \sim \text{Uniform}(0, 10), \quad \alpha \sim \text{Uniform}(0, 1), \quad \beta \sim \text{Uniform}(0, 1).$$

The full set of elementary random variables in the model then becomes

$$T = \{T_{1,1}, \ldots, T_{1,n_1}, T_{2,1}, \ldots, T_{2,n_2}, \mu_1, \mu_2, \sigma^2, \nu, \tau^2, \alpha, \beta\}$$

in which the first $n_1 + n_2$ elements are the data, the next three are first-level parameters, and the last four are second-level parameters (or hyper-parameters). On the whole, the distributional assumptions in the model are constraints on these probability kernels

$$p(T_{1,1}|\mu_1, \sigma^2), \ldots, p(T_{1,n_1}|\mu_1, \sigma^2), \; p(T_{2,1}|\mu_2, \sigma^2), \ldots, p(T_{2,n_2}|\mu_2, \sigma^2),$$
$$p(\mu_1|\nu, \tau^2), \; p(\mu_2|\nu, \tau^2), \; p(\sigma^2|\alpha, \beta), \; p(\nu|\varnothing), \; p(\tau^2|\varnothing), \; p(\alpha|\varnothing), \; p(\beta|\varnothing).$$

Note that the assumptions precisely specify the densities of the four hyper-parameters, and thus the conditioning variable in the last four kernels is the empty variable. Figure 3 is the directed graph generated by the incidence relation $\rightarrow$ when this is applied to the set of variable pairs in the kernels in question. It is seen that the graph has no cycle. This property and the fact that the conditioned variables in the pairs are mutually disjoint (they are distinct elementary variables) guarantee (again by Theorem 1) that this set

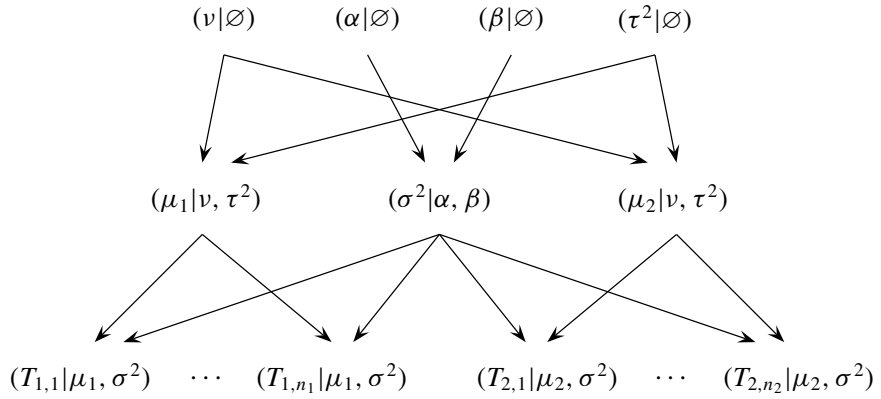


Figure 3. Incidence relation within a set of $n_1 + n_2 + 7$ variable pairs.

of variable pairs has sure compatibility, so that the distributional assumptions in the model are compatible. This appears to be a general property of hierarchical Bayesian models (Gelman et al., 2014, chapter 5; Lunn et al., 2013, chapter 10): the hierarchical form of a model implies absence of cycles among variable pairs in the primitive kernels, and then mutual compatibility of the kernels themselves.

Similar arguments may be used for the probabilistic models known as "Bayesian networks" (Pearl, 1988). Any Bayesian network rests on a graphical structure called a directed acyclic graph (DAG). A DAG differs from the kind of structures illustrated in Figure 3, since the nodes in it are individual elementary variables, rather than pairs of variables. However, a DAG can be faithfully translated into a graph of variable pairs, simply by replacing each elementary variable T_i by the pair $(T_i|X_i)$, where X_i is the set of "parents" of T_i (i.e., variables sending an arrow towards T_i in the DAG). For example, using this criterion, the graph of individual variables in the left-hand part of Figure 4 becomes translated into the graph of variable pairs in the right-hand part of the same figure. A DAG is acyclic, and this implies that the corresponding graph of variable pairs (related by the incidence relation $\rightarrow$) is also acyclic. Hence, by virtue of Theorem 1, we have that any assignment of kernels $\{p_1(T_1|X_1), \ldots, p_n(T_n|X_n)\}$ (which specify how each variable T_i is expected to *depend* on the set X_i of its parents in the network) will be mathematically consistent, that is, there is a consensus density $p(T)$ for the kernels. Within this structural assurance of consistency presumably lies a reason for the importance of the acyclicity requirement for Bayesian networks. Moreover, the DAG of a Bayesian network is also intended to represent a set of conditional stochastic *independencies* between the variables in the network. Specifically, each T_i is assumed to be independent of $T \setminus (T_i \cup X_i \cup Z_i)$ conditional on X_i, where X_i is the set of parents and Z_i is the set of descendants of T_i in the DAG. For example, the DAG in Figure 4 is intended to represent the following conditional independencies:

$$I(T_1, T_2|\varnothing),\ I(T_2, T_1T_3|\varnothing),\ I(T_3, T_2T_4|T_1),\ I(T_4, T_3|T_1T_2),\ I(T_5, T_1T_2|T_3T_4).$$

Based on these, any set of hypothesized kernels

$$p(T_1|\varnothing),\ p(T_2|\varnothing),\ p(T_3|T_1),\ p(T_4|T_1T_2),\ p(T_5|T_3T_4)$$

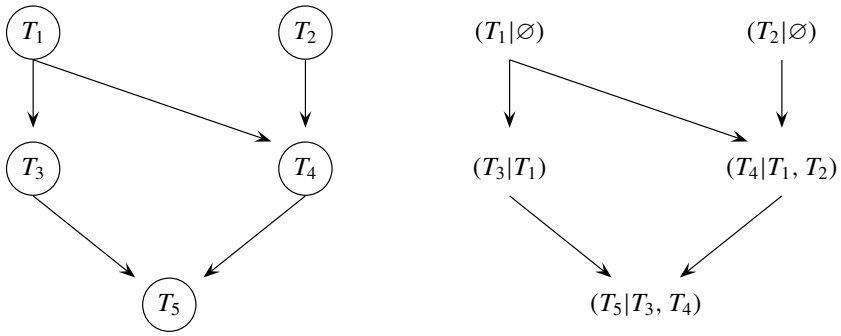


Figure 4. The DAG of a Bayesian network for five variables (left) and its rewriting in terms of incidence between variable pairs (right).

may equivalently be presented as

$$p(T_1|\varnothing),\ p(T_2|T_1),\ p(T_3|T_1T_2),\ p(T_4|T_1T_2T_3),\ p(T_5|T_1T_2T_3T_4).$$

In this form, the kernels constitute a cumulative scheme, so that by Lemma 3(i) there is a single full density $p(T)$ that is admissible for the network, which is obtainable from the kernels through multiple promotion (or the "chain rule" for Bayesian networks: Pearl, 1988, pp. 119–120; Kjærulff and Madsen, 2008, pp. 58–60).

As for Bayesian networks, the definition of a Markov random field is a mix of stochastic independence assumptions (each elementary variable T_i is assumed to be stochastically independent of $T \setminus (T_i \cup X_i)$ conditional on X_i, this being the set of elementary variables adjacent to T_i in the field) and stochastic dependence assumptions (a kernel $p(T_i|X_i)$ or a class of such kernels is associated with each elementary variable T_i and expresses how T_i is expected to be affected by its neighborhood X_i) (Kindermann and Snell, 1980; Koller and Friedman, 2009, chapter 4). Unlike the DAG of a Bayesian network, however, the graphical structure (an undirected graph) implicit in a Markov field does not generally possess the acyclicity character required for directly ensuring (by Theorem 1) compatibility between the kernels. Let us consider, for example, the graph in Figure 5. To specify a Markov field on this graph is tantamount to specifying nine local kernels

$$\begin{aligned}
&p(T_1|T_2T_4),\ p(T_2|T_1T_3T_5),\ p(T_3|T_2T_6),\\
&p(T_4|T_1T_5T_7),\ p(T_5|T_2T_4T_6T_8),\ p(T_6|T_3T_5T_9),\\
&p(T_7|T_4T_8),\ p(T_8|T_5T_7T_9),\ p(T_9|T_6T_8).
\end{aligned}$$

Each of these expresses how one of the nine elementary variables is assumed to be affected by its neighborhood in the graph. Evidently, there are $\rightarrow$-cycles within the set of variable pairs – for example, $(T_1|T_2T_4) \rightarrow (T_2|T_1T_3T_5) \rightarrow (T_1|T_2T_4)$ is a cycle – so that Theorem 1 cannot be invoked to directly conclude in favor of compatibility between the kernels. To answer the compatibility question in this situation, we need to consider the numerical properties of the kernels themselves, as families of density functions. The theory of Markov random fields especially explores the stochastic independence properties

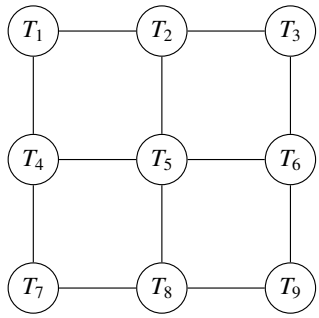


Figure 5. The neighborhood system of a possible Markov random field on nine variables.

implicit in such fields. For the reasons now suggested, however, advancements regarding compatibility, beyond structural assurance and in the directions illustrated in Sections 4 and 5, may also be relevant in dealing with such probabilistic models (Kaiser and Cressie, 2000; Kaiser, 2002).

Our last comment is on the relevance of the compatibility requirement for the so-called "Gibbs sampler" in probability simulations (Geman and Geman, 1984, pp. 730–732; Robert and Casella, 2004, chapter 10). A typical context for a Gibbs sampling is formed of a set $T = \{T_1, \ldots, T_n\}$ of elementary variables and a corresponding complete set $\{p(T_1|T \setminus \{T_1\}), \ldots, p(T_n|T \setminus \{T_n\})\}$ of elementary saturated kernels (so-called "full conditionals"). The method produces simulations of the full density $p(T)$ implied by the n input kernels, as well as simulations of other densities or probability kernels dominated by $p(T)$. In most applications, the input kernels are specified in a deductive manner. This means that a researcher first specifies the analytical expression for a full density $p(T)$, and then deduces (by conditioning) the analytical expression for each input kernel $p(T_i|T \setminus \{T_i\})$, which should be "available to sampling", that is, for each $x_i \in (T \setminus \{T_i\})^\circ$, it is practically possible to simulate samples from the distribution $p(T_i|x_i)$. In this approach, the mutual compatibility of the input kernels is true by construction, simply because their analytical expressions are deduced from the formula of $p(T)$, so that the input kernels are jointly dominated by $p(T)$. Furthermore, the kernels $\{p(T_1|T \setminus \{T_1\}), \ldots, p(T_n|T \setminus \{T_n\})\}$ thus determined form an alternating scheme, so that under the conditions stated in Lemma 3(ii) they constitute a sufficient basis for the univocal recovery of $p(T)$. A different route, however, could be taken. A researcher could directly specify (rather than deduce from a full density $p(T)$) a set $\{p(T_1|T \setminus \{T_1\}), \ldots, p(T_n|T \setminus \{T_n\})\}$ of elementary saturated kernels, and then apply to these the Gibbs sampler for simulation purposes. It is from this perspective that the critical role of the compatibility requirement appears more clearly. The mutual compatibility of the input kernels thus proposed has neither deductive assurance (they are not deduced from a common parent density $p(T)$) nor structural assurance (the incidence $\rightarrow$ among their variable pairs forms a complete directed graph, certainly containing cycles, so that the "if" part of Theorem 1 cannot be invoked). Thus, the proposed input kernels may fail to be compatible, in which case applying the Gibbs sampler would be tantamount to trying to simulate a nonexistent full density, resulting in erratic non-converging output sequences (Heckerman et al., 2000, pp. 56–57; Robert and Casella, 2011, p. 108). All of this demonstrates the saliency of the compatibility problem for the Gibbs sampler in current use for probability simulations.

References

Arnold, B. C., Castillo, E. and Sarabia, J. M. (1999), *Conditional specification of statistical models*. New York: Springer.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2001), "Conditionally specified distributions: An introduction (with comments and a rejoinder by the authors)", *Stat. Science* **16**, 249–274. DOI 10.1214/ss/1009213728

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2002), "Exact and near compatibility of discrete conditional distributions", *Comput. Stat. Data Anal.* **40**, 231–252. DOI 10.1016/S0167-9473(01)00111-6

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2004), "Compatibility of partial or complete conditional probability specifications", *J. Stat. Planning and Inference* **123**, 133–159. DOI 10.1016/S0378-3758(03)00137-X

Arnold, B. C. and Press, S. J. (1989), "Compatible conditional distributions", *J. Amer. Stat. Assoc.* **84**, 152–156. DOI 10.2307/2289858

Berti, P., Dreassi, E. and Rigo, P. (2014), "Compatibility results for conditional distributions", *J. Multivariate Anal.* **125**, 190–203. DOI 10.1016/j.jmva.2013.12.009

Besag, J. E. (1974), "Spatial interaction and the statistical analysis of lattice systems (with discussion)", *J. Royal Stat. Society, Series B* **36**, 192–236. DOI 10.1111/j.2517-6161.1974.tb00999.x

Billingsley, P. (1995), *Probability and measure*. New York: Wiley.

Burigana, L. and Vicovaro, M. (2020), "Inferring properties of probability kernels from the pairs of variables they involve", *Algebraic Stat.* **11**, 79–97. DOI 10.2140/astat.2020.11.79

Chang, J. T. and Pollard, D. (1997), "Conditioning as disintegration", *Stat. Neerlandica* **51**, 287–317. DOI 10.1111/1467-9574.00056

Chen, H. Y. (2010), "Compatibility of conditionally specified models", *Stat. Prob. Let.* **80**, 670–677. DOI 10.1016/j.spl.2009.12.025

Cressie, N. (1993), *Statistics for spatial data*. New York: Wiley.

Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-based approaches to calculating marginal densities", *J. Amer. Stat. Assoc.* **85**, 398–409. DOI 10.2307/2289776

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2014), *Bayesian data analysis*. Boca Raton, FL: CRC Press.

Gelman, A. and Speed, T. P. (1993), "Characterizing a joint probability distribution by conditionals", *J. Royal Stat. Society, Series B* **55**, 185–188. DOI 10.1111/j.2517-6161.1993.tb01477.x

Geman, S. and Geman, D. (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6*, 721–741. DOI 10.1109/TPAMI.1984.4767596

Gourieroux, C. and Monfort, A. (1979), "On the characterization of a joint probability distribution by conditional distributions", *J. Econometrics* **10**, 115–118. DOI 10.1016/0304-4076(79)90070-8

Griffeath, D. (1976), "Introduction to random fields", pp. 425–458 in *Denumerable Markov chains*, edited by J. G. Kemeny, J. L. Snell and A. W. Knapp. Berlin: Springer.

Gurevich, B. M. (1992), "On the joint distribution of random variables with given cross conditional distributions: discrete case", *Theory of Probability and its Applications* **36**, 371–375. DOI 10.1137/1136041

Heckerman, D., Chickering, D. M., Meek, C., Rounthwaite, R. and Kadie, C. (2000), "Dependency networks for inference, collaborative filtering, and data visualization", *J. Machine Learning Research* **1**, 49–75.

Hobert, J. P. and Casella, G. (1998), "Functional compatibility, Markov chains, and Gibbs sampling with improper posteriors", *J. Comput. Graphical Stat.* **7**, 42–60. DOI 10.1080/10618600.1998.10474760

Kaiser, M. S. (2002), "Markov random field models", pp. 1213–1224 in *Encyclopedia of Environmetrics*, vol. 3, edited by A. H. El-Shaarawi and W. W. Piegorsch, New York: Wiley. 10.1002/9781118445112.stat07479

Kaiser, M. S. and Cressie, N. (2000), "The construction of multivariate distributions from Markov random fields", *J. Multivariate Anal.* **73**, 199–220. DOI 10.1006/jmva.1999.1878

Kindermann, R. and Snell, J. L. (1980), *Markov random fields and their applications*. Providence, RI: American Mathematical Society.

Kjærulff, U. B. and Madsen, A. L. (2008), *Bayesian networks and influence diagrams: A guide to construction and analysis*. New York: Springer.

Koller, D. and Friedman, N. (2009), *Probabilistic graphical models: Principles and techniques*. Cambridge, MA: MIT Press.

Koski, T. and Noble, J. M. (2009), *Bayesian networks: An introduction*. Chichester, UK: Wiley.

Kuo, K. L, Song, C. C. and Jiang, T. J. (2017), "Exactly and almost compatible joint distributions for high-dimensional discrete conditional distributions", *J. Multivariate Anal.* **157**, 115–123. DOI 10.1016/j.jmva.2017.03.005

Kuo, K. L. and Wang, Y. J. (2011), "A simple algorithm for checking compatibility among discrete conditional distributions", *Comput. Stat. Data Anal.* **55**, 2457–2462. DOI 10.1016/j.csda.2011.02.017

Lauritzen, S. L. (1996), *Graphical models*. Oxford, UK: Oxford University Press.

Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2013), *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press.

Parthasarathy, K. R. (2005), *Introduction to probability and measure*. New Delhi: Hindustan Book Agency.

Pearl, J. (1988), *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.

Pollard, D. (2002), *A user's guide to measure theoretic probability*. Cambridge, UK: Cambridge University Press.

Robert, C. P. and Casella, G. (2004), *Monte Carlo statistical methods*. New York: Springer.

Robert, C. P. and Casella, G. (2011), "A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data", *Stat. Science* **26**, 102–115. DOI 10.1214/10-STS351

Slavkovic, A. B. and Sullivant, S. (2006), "The space of compatible full conditionals is a unimodular toric variety", *J. Symbolic Computation* **41**, 196–209. DOI 10.1016/j.jsc.2005.04.006

Tian, G. L., Tan, M., Ng, K. W. and Tang, M. L. (2009), "A unified method for checking compatibility and uniqueness for finite discrete conditional distributions", *Commun. Stat. Theory Methods* **28**, 115–129. DOI 10.1080/03610920802169586

Wang, Y. J. and Kuo, K. L. (2010), "Compatibility of discrete conditional distributions with structural zeros", *J. Multivariate Anal.* **101**, 191–199. DOI 10.1016/j.jmva.2009.07.007

Yao, Y. C., Chen, S. C. and Wang, S. H. (2014), "On compatibility of discrete full conditional distributions: A graphical representation approach", *J. Multivariate Analysis* **124**, 1–9. DOI 10.1016/j.jmva.2013.10.007

Received 2020-01-25. Revised 2020-06-11. Accepted 2020-07-06.

LUIGI BURIGANA: luigi.burigana@unipd.it
Department of General Psychology, University of Padua, I-35131 Padova, Italy

MICHELE VICOVARO: michele.vicovaro@unipd.it
Department of General Psychology, University of Padua, I-35131 Padova, Italy

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the submission page.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles are usually in English or French, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not refer to bibliography keys. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and a Mathematics Subject Classification for the article, and, for each author, affiliation (if appropriate) and email address.

Format. Authors are encouraged to use LaTeX and the standard amsart class, but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should normally be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages — Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc. — allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with as many details as you can about how your graphics were generated.

Bundle your figure files into a single archive (using zip, tar, rar or other format of your choice) and upload on the link you been provided at acceptance time. Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebraic Statistics

2020 11 : 2