15:2 2024

Algebraic Statistics

NEW DIRECTIONS IN ALGEBRAIC STATISTICS: THREE CHALLENGES FROM 2023

Yulia Alexandr, Miles Bakenhus, Maize Curiel, Sameer K. Deshpande, Elizabeth Gross, Yuqi Gu, Max Hill, Joseph Johnson, Bryson Kagy, Vishesh Karwa, Jiayi Li, Hanbaek Lyu, Sonja Petrović and Jose Israel Rodriguez

msp

Algebraic Statistics
Vol. 15, No. 2, 2024
https://doi.org/10.2140/astat.2024.15.357

NEW DIRECTIONS IN ALGEBRAIC STATISTICS: THREE CHALLENGES FROM 2023

YULIA ALEXANDR, MILES BAKENHUS, MAIZE CURIEL, SAMEER K. DESHPANDE, ELIZABETH GROSS, YUQI GU, MAX HILL, JOSEPH JOHNSON, BRYSON KAGY, VISHESH KARWA, JIAYI LI, HANBAEK LYU, SONJA PETROVIĆ AND JOSE ISRAEL RODRIGUEZ

In the last quarter of a century, algebraic statistics has established itself as an expanding field which uses multilinear algebra, commutative algebra, computational algebra, geometry, and combinatorics to tackle problems in mathematical and computational statistics. These developments have found applications in a growing number of areas, including biology, neuroscience, economics, and social sciences.

Naturally, new connections continue to be made with other areas of mathematics and statistics. We outline three such connections: to statistical models used in educational testing, to a classification problem for a family of nonparametric regression models, and to phase transition phenomena under uniform sampling of contingency tables. We illustrate the motivating problems, each of which is for algebraic statistics a new direction, and demonstrate an enhancement of related methodologies.

1. Introduction

We illustrate three new research directions in algebraic statistics which share the following common philosophy: they connect algebraic statistics to applied problems from another research area, reinterpret that problem, and illustrate that this new connection is effective toward solving a family of challenges. The choice of the three sets of problems is not made by ranking, but rather by opportunity: namely, in fall of 2023, the Institute for Mathematics and Statistics Innovation hosted a long program *Algebraic statistics and our changing world*. The program included two-day working group sessions motivated by a problem presented in the "Questions and Consulting seminar". This paper illustrates three of the new research directions resulting from these interactions. In the spirit of the Oberwolfach *Lectures on algebraic statistics* [20, Chapters 6–7], each section of this survey is self contained.

In Section 2, Yulia Alexandr, Yuqi Gu, Jiayi Li, and Jose Israel Rodriguez study the likelihood geometry of a statistical model motivated by cognitive diagnosis of latent skills in educational and psychological measurement. A Bless model is a discrete statistical model with latent variables. Bless is an acronym for a "binary latent clique star forest." Identifiability of these models have been previously established [26; 27] and reparametrizations have leveraged tools from algebraic statistics. The new direction here is to study the likelihood geometry of the models for the statistical inference using maximum likelihood

MSC2020: 62R01.
Keywords: algebraic statistics, likelihood geometry, educational measurement, identifiability, regression trees, contingency tables, phase transition.

estimation. Since the Bless model can serve as a building block for identifiable deep generative models with multiple latent layers [28], studying the likelihood geometry of the Bless model can pave the way for a deeper understanding of these modern powerful generative models.

In Section 3, Maize Curiel, Sameer Despande, Joe Johnson and Bryson Kagy make progress toward identifying (nearly) equivalent regression trees, which represent piecewise constant step functions (see Figure 1). The new direction aims to leverage ideas from algebraic statistics and combinatorics to improve the Bayesian additive regression trees (BART) [13] model for nonparametric regression.

In Section 4, Miles Bakenhus, Elizabeth Gross, Max Hill, Vishesh Karwa, Hanbaek Lyu, and Sonja Petrović study phase transitions problems on contingency tables through the lens of algebraic statistics. The motivating problem is the appearance of a sharp phase transition in the estimability of the uniform distribution on the space of tables by the hypergeometric distribution. The threshold for this phase transition is expressed as a condition on the margins of the table, and has been solved for the two-dimensional case. Interpreting the problem from the point of view of algebraic statistics, we propose the generalization of this phenomenon to multiway tables and partially solve the problem for the three-dimensional case. The two distributions are not only combinatorially interesting, but have statistical relevance: sampling from the hypergeometric distribution is used for exact conditional tests of model fit, while sampling from the uniform is needed for performing conditional volume tests under the multinomial sampling scheme.

2. Likelihood geometry in a star-forest model with dependent binary latent variables

This section arose from the Questions and Consulting seminar by Yuqi Gu and subsequent discussions amongst Yulia Alexandr, Jiayi Li, and Jose Israel Rodriguez.

2.1. ***Blessed models by parametrization.*** Consider the *Binary Latent cliquE Star foreSt* (Bless) model [27] with the following parametrization. Let $\boldsymbol{A} \in \{0,1\}^K$ denote the latent random vector, and $\boldsymbol{Y} \in \{0,1\}^p$ denote the observed random vector. We next describe the distribution of $\boldsymbol{A}$ and $\boldsymbol{Y} \mid \boldsymbol{A}$, respectively, to complete the model specification. Assume the binary latent variables can be arbitrarily dependent on each other with the following saturated parametrization:

$$\mathbb{P}(\boldsymbol{A} = \boldsymbol{\alpha}) = \nu_{\boldsymbol{\alpha}} \quad \text{for all } \boldsymbol{\alpha} \in \{0,1\}^K,$$

where $\sum_{\boldsymbol{\alpha}\in\{0,1\}^K} \nu_{\boldsymbol{\alpha}} = 1$. Assume the observed $Y_1, \ldots, Y_p$ are conditionally independent given the latent $\boldsymbol{A}$, where each Y_j has exactly one latent parent denoted by $A_{\mathrm{pa}(j)}$ (here $\mathrm{pa}(j) \in \{1, \ldots, K\}$). In other words, the bipartite graph from the latent $\boldsymbol{A}$ to the observed $\boldsymbol{Y}$ is a star-forest graph, and it follows from the conditional independence property of the graphical model that

$$\mathbb{P}(Y_j \mid \boldsymbol{A}) = \mathbb{P}(Y_j \mid A_{\mathrm{pa}(j)}).$$

Parametrize the conditional distribution of $Y_j \mid A_{\mathrm{pa}(j)}$ as follows:

$$\mathbb{P}(Y_j = 1 \mid A_{\mathrm{pa}(j)} = 1) = \theta_{j,+},$$
$$\mathbb{P}(Y_j = 1 \mid A_{\mathrm{pa}(j)} = 0) = \theta_{j,-}.$$

Based on the above assumptions, the marginal distribution of the observed random vector $\boldsymbol{Y}$ can be written as: for all $\boldsymbol{y} \in \{0,1\}^p$, it holds that

$$\begin{aligned}\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}) &= \sum_{\boldsymbol{\alpha}\in\{0,1\}^K} \mathbb{P}(\boldsymbol{A}=\boldsymbol{\alpha}) \prod_{j=1}^{p} \mathbb{P}(Y_j \mid \boldsymbol{A}=\boldsymbol{\alpha}) \\ &= \sum_{\boldsymbol{\alpha}\in\{0,1\}^K} \nu_{\boldsymbol{\alpha}} \prod_{j=1}^{p} \mathbb{P}(Y_j \mid A_{\mathrm{pa}(j)}=\alpha_{\mathrm{pa}(j)}) \\ &= \sum_{\boldsymbol{\alpha}\in\{0,1\}^K} \nu_{\boldsymbol{\alpha}} \prod_{j=1}^{p} \left[\theta_{j,+}^{\alpha_{\mathrm{pa}(j)}} \theta_{j,-}^{1-\alpha_{\mathrm{pa}(j)}}\right]^{y_j} \left[(1-\theta_{j,+})^{\alpha_{\mathrm{pa}(j)}} (1-\theta_{j,-})^{1-\alpha_{\mathrm{pa}(j)}}\right]^{1-y_j}.\end{aligned}$$

We consider the following inequality constraints on the parameters $\boldsymbol{\Theta} = \{\theta_{j,+}, \theta_{j,-} : j \in [p]\}$ and $\boldsymbol{\nu} = (\nu_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \{0,1\}^K)$:

$$\begin{aligned}\theta_{j,+} > \theta_{j,-} \quad &\text{for all } j \in [p], \\ \nu_{\boldsymbol{\alpha}} > 0 \quad &\text{for all } \boldsymbol{\alpha} \in \{0,1\}^K.\end{aligned}$$

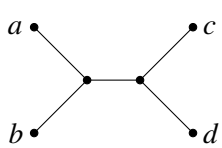
It is known that when each latent variable A_k has exactly two observed variables as children, the model parameters $\boldsymbol{\Theta}$ and $\boldsymbol{\nu}$ are generically identifiable [27]. More specifically, in this case, $\{\theta_{j,+}, \theta_{j,-}\}$ are identifiable if and only if A_k is not independent from $(A_1, \ldots, A_{k-1}, A_{k+1}, \ldots, A_K)$ (i.e., parameters $\boldsymbol{\nu}_{\boldsymbol{\alpha}}$ satisfy certain binomial inequalities). The arbitrary dependence allowed among the K latent variables makes the Bless model an expressive modeling tool, and also makes it possible to extend it to identifiable deep generative models with multiple latent layers.

Given N i.i.d. observed vectors $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$ in a sample, the MLE $(\widehat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\nu}})$ is defined as the maximizer of the likelihood function $\prod_{i=1}^{N} \mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}_i \mid \boldsymbol{\Theta}, \boldsymbol{\nu})$. The identifiability result stated above has a nice consequence that the maximum likelihood estimator (MLE) of parameters $(\boldsymbol{\Theta}, \boldsymbol{\nu})$ is statistically consistent as $N \to \infty$; see, e.g., [29, Proposition 3.4]. However, given a finite sample of size N, the properties of the MLE are not well understood. Moreover, the computation of MLE is often through the iterative EM algorithm, which is sensitive to parameter initialization. Relevant references in algebraic statistics include [2; 3; 22; 40].

2.2. ***Likelihood geometry of Blessed models.*** Let n be the dimension of the ambient space of the model $\mathcal{M}$, i.e., $\overline{\mathcal{M}} \subseteq \mathbb{C}^n$. The *maximum likelihood* (*ML*) *degree* [42, Chapter 7] of an algebraic statistical model $\mathcal{M}$ is the number of complex critical points of ℓ_u on the Zariski closure of $\mathcal{M}$ for generic data $u \in \Delta_{n-1}$. The ML degree measures the algebraic complexity of maximum likelihood estimation.

2.2.1. *A first example and many states.*

Example 1. Consider the Blessed model when $K = 2$, given by the graph



where all random variables are binary. The model has dimension 11 inside Δ_{15}. Its parametrization is given by

$$p_{ij,kl} = v_{00}a_{i0}b_{j0}c_{k0}d_{l0} + v_{10}a_{i1}b_{j1}c_{k0}d_{l0} + v_{01}a_{i0}b_{j0}c_{k1}d_{l1} + v_{11}a_{i1}b_{j1}c_{k1}d_{l1},$$

where

$$a_{ij} = \mathbb{P}(Y_0 = i|\alpha_0 = j), \quad b_{ij} = \mathbb{P}(Y_1 = i|\alpha_0 = j), \quad c_{ij} = \mathbb{P}(Y_0 = i|\alpha_1 = j), \quad d_{ij} = \mathbb{P}(Y_1 = i|\alpha_1 = j),$$

and $p_{ij,kl}$ are the coordinates of the image space.

The below code computes the implicit description of the model, utilizing bounded-degree Gröbner basis computations in Macaulay2:

```
R=QQ[v_(0,0)..v_(1,1),a_(0,0)..a_(1,1),b_(0,0)..b_(1,1),c_(0,0)..c_(1,1),
d_(0,0)..d_(1,1), p_(0,0,0,0)..p_(1,1,1,1), MonomialOrder => Eliminate 20]
probabilities=toList(p_(0,0,0,0)..p_(1,1,1,1))

par = (i,j,k,l)->(
    summ=0;
    for m from 0 to 1 do (
for n from 0 to 1 do (
    s=v_(m,n)*a_(i,m)*b_(j,m)*c_(k,n)*d_(l,n);
    summ=summ+s;
    ););
    return summ;)

gs={}; for p in probabilities do gs=append(gs, p-par((baseName p)#1));
G=ideal(gs); time I=ideal(selectInSubring(1,gens gb(G,DegreeLimit=>15)))
codim I
```

Checking the codimension of the resulting ideal, we find that the model is described by sixteen cubics. Moreover, this model is precisely the 2-mixture of the 4×4 independence model. This can be seen by realizing the parametrization as a product of a 4×2 matrix, 2×2 matrix, and 2×4 matrix, as follows:

$$\begin{bmatrix} p_{00,00} & p_{00,01} & p_{00,10} & p_{00,11} \\ p_{01,00} & p_{01,01} & p_{01,10} & p_{01,11} \\ p_{10,00} & p_{10,01} & p_{10,10} & p_{10,11} \\ p_{11,00} & p_{11,01} & p_{11,10} & p_{11,11} \end{bmatrix} = \begin{bmatrix} a_{00}b_{00} & a_{01}b_{01} \\ a_{00}b_{10} & a_{01}b_{11} \\ a_{10}b_{00} & a_{11}b_{01} \\ a_{10}b_{10} & a_{11}b_{11} \end{bmatrix} \begin{bmatrix} v_{00} & v_{01} \\ v_{10} & v_{11} \end{bmatrix} \begin{bmatrix} c_{00}d_{00} & c_{01}d_{01} \\ c_{00}d_{10} & c_{01}d_{11} \\ c_{10}d_{00} & c_{11}d_{01} \\ c_{10}d_{10} & c_{11}d_{11} \end{bmatrix}^T$$

Implicitly, the model is known to be described by all sixteen 3×3 minors of the matrix of joint probabilities $(p_{ij,kl})$, which confirms our earlier computations. Its ML degree is 191; it was computed in [30].

2.2.2. *Implicit equations for general quartet trees and Blessed cherry orchard models.* We are interested in the generalization of Example 1, where K binary latent variables form a complete graph, and each latent variable is adjacent to two n-category observed variables. Let Y_{i1} and Y_{i2} be the observed variables adjacent to the latent variable A_i. We will refer to these models as *Blessed cherry orchard models*, denoted

by $\mathcal{M}_{K,n}$. Let $T = (v_{i_1 \dots i_K})$ denote the K-way $2 \times \cdots \times 2$ tensor. For each $i \in [K]$, we will let M_i denote the $n^2 \times 2$ matrix with the two columns

$$M_i^{(j)} = \left[\mathbb{P}(Y_{i1} = \ell_1 | \alpha_i = j) \cdot \mathbb{P}(Y_{i2} = \ell_2 | \alpha_i = j)\right]_{\ell \in [n]^2}, \quad \text{one for each value of } j = 0, 1.$$

Note that the parametrization of the model $\mathcal{M}_{K,n}$ can be realized as

$$p_{i_1 \dots i_K} = v_{i_1 \dots i_K} \cdot M_1^{(i_1)} \otimes M_2^{(i_2)} \otimes \cdots \otimes M_K^{(i_K)}.$$

Therefore, $\mathcal{M}_{K,n}$ is the model of K-way $n^2 \times n^2 \times \cdots \times n^2$ tensors with multilinear rank at most $(2, 2, \dots, 2)$. Its equations are just minors of flattenings, similar to Example 1. On the other hand, mixtures of two independence models correspond to border rank at most two matrices, and their equations are much more subtle. Invariants for these cases can be found in [34; 37].

Similarly, when the observed variables have a different number of states, the resulting models are multilinear rank $(2, 2, \dots, 2)$ tensors of different size. A recursive formula can be derived for these models by the results in [38, Section 4].

2.2.3. *EM algorithm for Blessed cherry orchard models.* The EM algorithm is the standard method for maximizing the likelihood function on Blessed models. Fixed points of the EM algorithm (Algorithm 1; also see Algorithm 1 in [27]) on the Blessed cherry orchard model refer to the set of all points

Algorithm 1. Expectation-maximization algorithm.

Data: Observation data $\boldsymbol{Y}$, initial parameters $\boldsymbol{\Theta}^{(0)}$, $\boldsymbol{v}^{(0)}$.
Result: Estimated parameters $\widehat{\boldsymbol{\Theta}}$, $\hat{\boldsymbol{v}}$.
Input: Convergence threshold ϵ.
$\boldsymbol{\Theta} \leftarrow \boldsymbol{\Theta}^{(0)}$;
$\boldsymbol{v} \leftarrow \boldsymbol{v}^{(0)}$;
$t \leftarrow 0$;
repeat
  $t \leftarrow t + 1$;
  E-step: Calculate the expected value of the latent variables;
  $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t-1)}) \leftarrow \mathbb{E}_{\boldsymbol{A}|\boldsymbol{Y},\boldsymbol{\Theta}^{(t-1)},\boldsymbol{v}^{(t-1)}}\left[\log \prod_{i=1}^{N} \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}_i \mid \boldsymbol{\Theta}, \boldsymbol{v})\right]$;
  $Q(\boldsymbol{v}|\boldsymbol{v}^{(t-1)}) \leftarrow \mathbb{E}_{\boldsymbol{A}|\boldsymbol{Y},\boldsymbol{\Theta}^{(t)},\boldsymbol{v}^{(t-1)}}\left[\log \prod_{i=1}^{N} \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}_i \mid \boldsymbol{\Theta}, \boldsymbol{v})\right]$;
  M-step: Find the parameters that maximize this quantity;
  $\boldsymbol{\Theta}^{(t)} \leftarrow \arg\max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t-1)})$;
  $\boldsymbol{v}^{(t)} \leftarrow \arg\max_{\boldsymbol{v}} Q(\boldsymbol{v}|\boldsymbol{v}^{(t-1)})$;
  Check for convergence;
  if $\|\boldsymbol{\Theta}^{(t)} - \boldsymbol{\Theta}^{(t-1)}\| < \epsilon$ and $\|\boldsymbol{v}^{(t)} - \boldsymbol{v}^{(t-1)}\| < \epsilon$ **then**
    break;
until convergence

$(\boldsymbol{\Theta}^*, \boldsymbol{\nu}^*)$ where

$$\boldsymbol{\Theta}^* = \arg\max_{\boldsymbol{\Theta}} \mathbb{E}_{\boldsymbol{A}|\boldsymbol{Y},\boldsymbol{\Theta}^*,\boldsymbol{\nu}^*}\left[\log \prod_{i=1}^{N} \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}_i \mid \boldsymbol{\Theta}^*, \boldsymbol{\nu}^*)\right],$$

$$\boldsymbol{\nu}^* = \arg\max_{\boldsymbol{\nu}} \mathbb{E}_{\boldsymbol{A}|\boldsymbol{Y},\boldsymbol{\Theta}^*,\boldsymbol{\nu}^*}\left[\log \prod_{i=1}^{N} \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}_i \mid \boldsymbol{\Theta}^*, \boldsymbol{\nu}^*)\right].$$

Maximizing the log-likelihood on $\mathcal{M}_{K,n}$ is a nonconvex optimization problem. The output of the EM algorithm $(\boldsymbol{\Theta}^*, \boldsymbol{\nu}^*)$, with respect to any initialization either lies in the relative interior or on the model's boundary. If $(\boldsymbol{\Theta}^*, \boldsymbol{\nu}^*)$ is in the relative interior then $(\boldsymbol{\Theta}^*, \boldsymbol{\nu}^*)$ is a critical point of the log-likelihood function and the number of such points is counted by the ML degree. If $(\boldsymbol{\Theta}^*, \boldsymbol{\nu}^*)$ is on the boundary then $(\boldsymbol{\Theta}^*, \boldsymbol{\nu}^*)$ is generally not a critical point of the log-likelihood requires studying the ML degree of boundary components like in [33].

2.3. ***Remaining open questions.*** The following open questions about the model and the likelihood geometry are of interest. First, note that the Blessed models we have considered so far have high ML degrees, so there is no closed-form MLE. However, if one imposes symmetries like in [30, Section 3] then closed formulas may be derived.

Question 2. Are there any other statistically meaningful restrictions that could guarantee closed-form MLE?

For instance, when entries of the data are zero, this causes the 191 critical points to be partitioned into sets according to their supports. This was previously explored in [25, Section 4.3] under the guise of *ML tables*. A specific challenge problem is to characterize the ML table for the Blessed model in Example 1.

Second, what properties does the EM algorithm have for the Blessed cherry orchard models? An algebraic approach to this would be to characterize the EM fixed points of these models like in [33], however we expect such a characterization to be very challenging.

Third, can we perform formal statistical hypothesis tests of the goodness-of-fit of the Bless model based on the algebraic characterizations? One direction is to adapt the idea of the Monte Carlo goodness-of-fit test developed for the stochastic block model (SBM) in [31] to the Bless model. We conjecture that such an extension is promising because both SBM and Bless share the common features of

(a) having discrete observed and latent variables and

(b) characterizing the joint distribution of the observed and latent variables via an exponential family distribution.

More specifically, one could calculate the χ^2 distance between the marginal distribution of the observed variables of the contingency tables in the fiber and the empirical marginal distribution based on data, and use this distance as a test statistic. Another approach could be to properly sample from logarithmic Voronoi cells [1] (as well as describe their boundary), as the notion of sufficient statistic is not defined for models with latent variables.

Fourth, what is the relationship of Blessed models with mixtures of independence models? Finally, can we determine the boundaries of the image of these models? Are there any nontrivial inequalities?

There are results [40] in algebraic statistics where the two apparently different parametric models have the same image up to Zariski closure. We see another example of this phenomenon in Section 2.2.2 between Blessed cherry orchard models and of multilinear rank $2 \times 2 \times \cdots \times 2$ tensors.

Question 3. What can be said about the inequalities defining the Blessed cherry orchard model, and are they different from those defining the model of multilinear rank $2 \times 2 \times \cdots \times 2$ tensors?

3. Identifying (nearly) equivalent regression trees

This section grew from the Questions and Consulting seminar by Sameer K. Deshpande and follow up discussions with Maize Curiel, Joseph Johnson, and Bryson Kagy.

3.1. ***Introduction.*** **Motivation**. Consider the nonparametric regression problem: given n observations of covariates $\boldsymbol{x} \in \mathbb{R}^p$ and outcomes $y \in \mathbb{R}$ from the model $y \sim \mathcal{N}(f(\boldsymbol{x}), \sigma^2)$, we would like to estimate the function $f : \mathbb{R}^p \to \mathbb{R}$. Bayesian additive regression trees [13] is a Bayesian sum-of-trees model that, at a high-level, approximates f with a large ensemble of regression trees (i.e., piecewise constant step functions). Usually, f is nonlinear and involves complicated high-order interactions. Such nonlinearities and interactions are typically impossible to specify correctly a priori using a parametric model. Using BART, however, users often obtain extremely accurate predictions of function evaluations along with reasonably well-calibrated uncertainty intervals *without prespecifying the functional form of f or tuning several hyperpriors*. The ease-of-use and generally excellent, tuning-free performance have made BART a popular "off-the-shelf" tool to be used within larger modeling workflows.

Formally, BART works by simulating draws from a posterior distribution over tree ensembles using Markov chain Monte Carlo. In each iteration of the sampler, individual trees are grown (by splitting an existing leaf node into two new child nodes) or pruned (by collapsing two leaf nodes to their common parent). It has been observed empirically — and recently demonstrated theoretically [32; 39] — that such local moves result in extremely slow mixing. Intuitively, we might expect to achieve faster mixing by making more radical changes to the tree structure. And this is indeed the case, at least empirically: [36] was able to explore tree space more efficiently using a proposal mechanism that radically changed the overall structure of the tree. At a high-level that proposal simultaneously permuted the order of decision rules within a tree and added or removed new subtrees. Motivated by those results, we conjecture that one can obtain faster mixing by directly transitioning between trees that provide (nearly) identical fits to the data.

3.1.1. *Setting & notation.* To motivate this idea, consider the slightly simpler setting in which we approximate the function $f(\boldsymbol{x})$ with a single binary regression tree and know the residual variance σ^2. Formally, a *regression tree* is a pair $(T, \boldsymbol{\mu})$ consisting of

(i) a finite, rooted binary decision tree T containing several terminal or *leaf* nodes and several nonterminal or *decision* nodes and

(ii) a collection $\boldsymbol{\mu}$ of scalars, one for each leaf node in T.

Every nonterminal node in T is connected to two children nodes, a left child and a right child. Further, associated to every nonterminal node is a decision rule of the form $X_j < c$, where $(X_1, \ldots, X_p)$ is a random vector taking on states in $[0, 1]^p$ and $c \in (0, 1)$.

Given a decision tree T and any point $\boldsymbol{x} = (x_1, \ldots, x_p) \in [0, 1]^p$, we can trace a path from the root to a leaf by following the decision rules. Specifically, starting from the root, whenever the path reaches a decision rule $X_j < c$, it proceeds to the left if $x_j < c$ and to the right otherwise. We will restrict attention only to those decision trees that partition $[0, 1]^p$ in the sense that

(i) every leaf node is reached by the decision-following path of at least one $\boldsymbol{x} \in \mathcal{X}$ and

(ii) the decision-following path of every $\boldsymbol{x} \in [0, 1]^p$ reaches a single, unique leaf.

Given $(T, \boldsymbol{\mu})$ and a point $\boldsymbol{x} \in [0, 1]^p$, let $\ell(\boldsymbol{x}; T)$ denote the leaf reached by $\boldsymbol{x}$'s decision-following path. By associating each leaf of T with its own scalar, the regression tree $(T, \boldsymbol{\mu})$ represents a piecewise constant function of $[0, 1]^p$. Formally, we introduce the evaluation function $g(\boldsymbol{x}; T, \boldsymbol{\mu}) = \mu_{\ell(\boldsymbol{x})}$, which returns the element of $\boldsymbol{\mu}$ associated with the leaf reached by $\boldsymbol{x}$'s decision-following path. Additionally, given N points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in [0, 1]^p$ let $I_\ell(T) = \{i : \ell(\boldsymbol{x}_i; T) = \ell\}$ contain the set of indices of the points that reach leaf ℓ.

With this notation in hand, we can define the single-tree Bayesian model with known residual variance σ^2

$$\begin{aligned} y_i | T, \boldsymbol{\mu}, &\sim \mathcal{N}(g(\boldsymbol{x}_i; T, \boldsymbol{\mu}), \sigma^2) && \text{for } i = 1, \ldots, N, \\ \mu_\ell | T &\sim \mathcal{N}(0, \tau^2) && \text{for } \mu_\ell \in \boldsymbol{\mu}, \\ T &\sim \Pi(T), \end{aligned}$$

where $\sigma, \tau, \nu, \lambda > 0$ are fixed positive constants and $\Pi(T)$ is the decision tree prior used in [13].

Under this model, we can compute the marginal likelihood of the decision tree T in closed-form:

$$p(\boldsymbol{y}|T) \propto \prod_{\ell} \exp\{\text{stuff depending only on } I_\ell(T)\text{'s \& } y\text{'s}\}$$

Importantly, the marginal likelihood of T depends on the decision tree only through the partition of the points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ it induces. We say that two decision trees T and T' are *equivalent* if they induce the same partition of $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$.

We study the following questions: given a single decision tree T and a collection of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in [0, 1]^p$, can we

(1) enumerate or characterize the equivalence class of trees that induce the exact same partition of the points?

(2) sample uniformly from the set of trees inducing the same partition?

(3) enumerate or characterize the set of trees that induce partitions that are close (in some sense) to the one induced by the tree?

(4) sample uniformly from the set of trees inducing nearly the same partition?

Resolving these questions will enable construction of more efficient MCMC sampling techniques for fitting BART. Beyond the motivating Bayesian context, however, answering these questions will more generally facilitate uncertainty quantification about tree models. As one example, suppose we fit a complicated machine learning model (e.g., a deep neural network) to $(\boldsymbol{x}, y)$ data to obtain predicted outcomes $\hat{y}$.

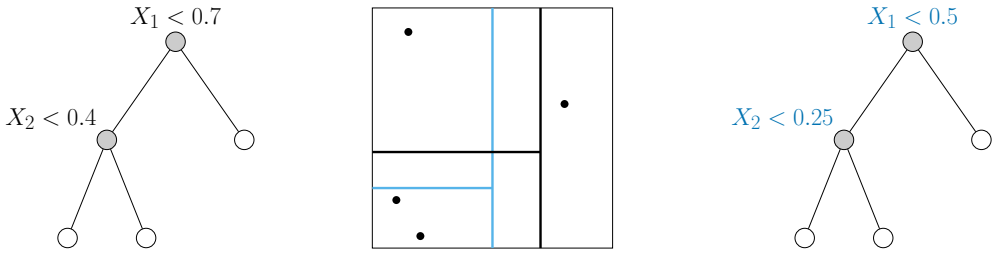


Figure 1. We can form trivially equivalent trees by perturbing decision boundaries. Here the tree topology for both trees is the same, but the decision rules for the tree on the left and right are different but they are equivalent because they differentiate the data points in the same way.

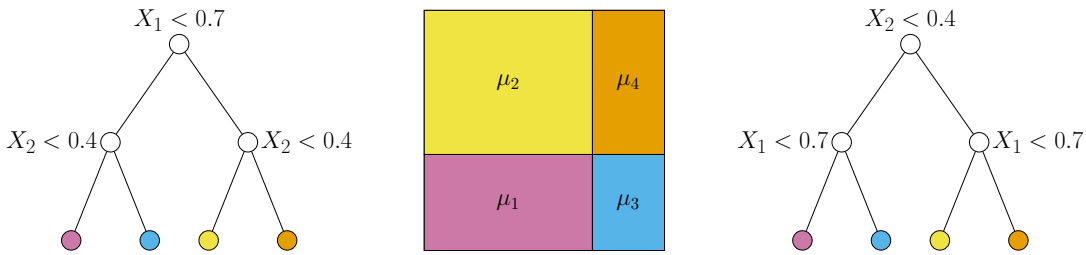


Figure 2. For full decision trees with the same rule at each level, we can form equivalent trees by permuting the decision rules across levels.

Even if the fitted model is difficult to interpret, we can nevertheless obtain a much more interpretable *approximation* by training a regression tree model to the pairs $(\boldsymbol{x}, \hat{y})$. The set of (nearly) equivalent trees provides one avenue to quantify uncertainty about the interpretation of the original fitted model.

3.2. ***Preliminaries & special cases.*** Before proceeding, we verify that equivalent trees do exist generally: given a tree T and set of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, we can trivially obtain equivalent trees by moving the decision boundaries between the data points; see Figure 1 for an example with $p = 2$.

Henceforth, we will focus instead on identifying equivalent trees that do not change the decision boundaries. To this end, consider first the case where T is

(i) a full binary tree of depth D containing 2^D leaf nodes and

(ii) the same decision rule is used at every decision node at depth d.

Given such a tree, we can form an equivalent tree by permuting the decision rules across the levels. Figure 2 shows an example of equivalent trees of depth $D = 2$ with $p = 2$ and the associated partition of $[0, 1]^2$.

3.2.1. *The $p = 1$ setting.* Now suppose that $p = 1$ and that T contains $L \geq 3$ leaf nodes and $L - 1$ decision nodes. In this case, finding equivalent trees is not as simple as permuting decision rules across the levels of T. Notice, however, that we can form an equivalent tree by replacing subtrees of T with equivalent subtrees (see Figure 3 for an example). To see that the two trees are equivalent, notice that the original subtree partitions the interval $[0, 0.5) = [0, 0.25) \cup [0.25, 0.5)$ and then partitions

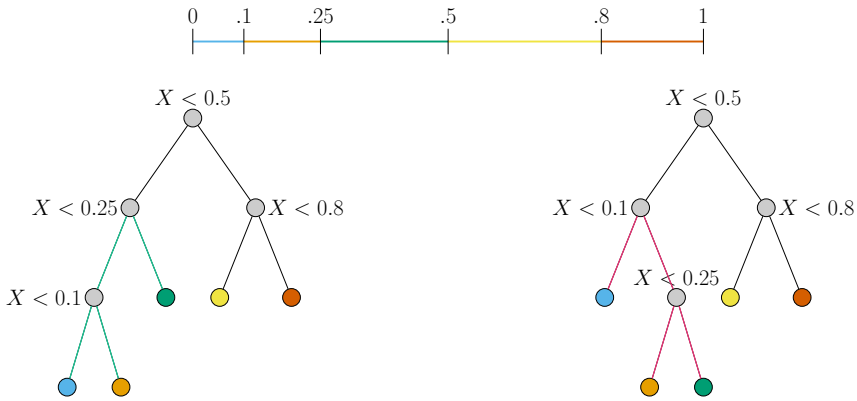


Figure 3. We can form equivalent trees by replacing the green subtree on the left with the equivalent red subtree on the right. Doing so amounts to partitioning $[0, 0.5)$ into $[0, 0.1)$, $[0.1, 0.25)$, and $[0.25, 0.5)$ in different orders.

$[0, 0.25) = [0, 0.1) \cup [0.1, 0.25)$. The equivalent subtree first partitions $[0, 0.5) = [0, 0.1) \cup [0.1, 0.5)$ and then partitions $[0.1, 0.5) = [0.1, 0.25) \cup [0.25, 0.5)$. Essentially, replacing a subtree with an equivalent tree amounts to arranging the decision boundaries in that subtree in a different order.

The example of Figure 3 gives us a strategy for counting the number of equivalent trees. To this end, suppose that T contains $L-1$ internal, decision nodes and let $0 < c_1 < \cdots < c_{L-1} < 1$ be the ordered decision boundaries. For any collection of $\ell - 1$ consecutive decision boundaries let $C_{\ell-1}$ count the number of equivalent trees that contain $\ell - 1$ decision nodes associated with those decision boundaries. Immediately, we know that $\widetilde{C}_{\ell-1} < C_{\ell-1}$, the $(\ell-1)$-st Catalan number, which counts the total number of binary trees with $\ell - 1$ internal nodes.

Further suppose that we have enumerated all $\widetilde{C}_{\ell-1}$ such trees for every collection of $\ell - 1$ consecutive boundaries with $\ell < L$. We can form a new tree $T^\star$ as follows:

(1) Initialize $T^\star$ to be just the root node.

(2) Pick one decision boundary c_k to associate to $T^\star$'s root.

(3) Draw one of the $\widetilde{C}_{k-1}$ trees with $k-1$ decision boundaries $\{c_1, \ldots, c_{k-1}\}$. Call it $T_L^\star$.

(4) Draw one of the $\widetilde{C}_{L-k-1}$ trees with $L-k-1$ decision boundaries $\{c_{k+1}, \ldots, c_{L-1}\}$. Call it $T_R^\star$.

(5) Connect the roots of $T^\star$, $T_L^\star$ and $T_R^\star$ so that $T_L^\star$'s root is the left child of $T^\star$'s root and $T_R^\star$'s root is the right child of $T^\star$'s root.

Using this process, we have

$$\widetilde{C}_{L-1} \geq \sum_{k=1}^{L-1} \widetilde{C}_{k-1} \widetilde{C}_{L-k-1}.$$

A strong induction argument show that, in fact, $\widetilde{C}_{L-1} = C_{L-1}$. So when $p = 1$, there are a Catalan number of equivalent trees with the same number of leaf nodes. Further, given any tree with L leaf nodes, we can

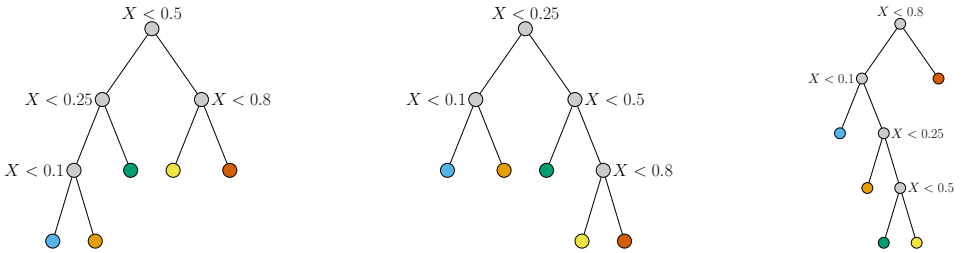


Figure 4. Three trees equivalent to the ones in Figure 3.

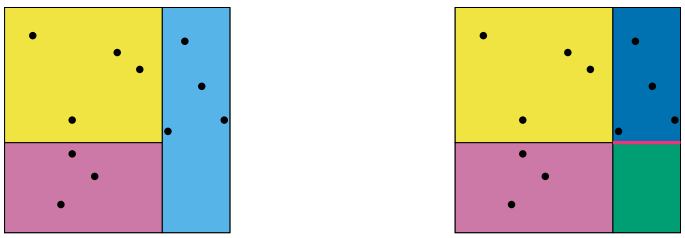


Figure 5. Extending the cut along $x_2 = 0.4$ does not change the underlying partition of the data.

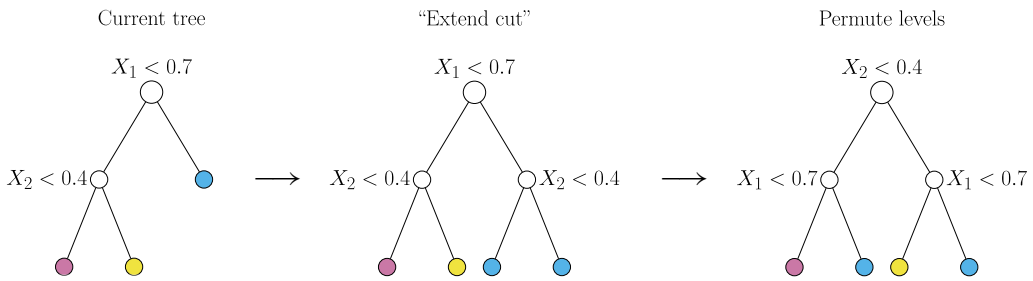


Figure 6. The associated trees that correspond to the extension in Figure 5. The rightmost tree corresponds to permuting the root node with all nodes in the first level.

form every other equivalent tree with the same number of leaf nodes and decision boundaries. Figure 4 shows some examples with $L = 5$.

3.3. ***The $p = 2$ setting.*** Consider the partitions of $[0, 1]^2$ and 10 data points on the left of Figure 5. Notice that we can further partition the blue rectangle $[0.7, 1] \times [0, 1]$ into $[0.7, 1] \times [0, 0.4) \cup [0.7, 1] \times [0.4, 1]$ *without changing the partition of the data points*. That is, we can obtain the partition on the right by "extending" the cut separating pink and yellow rectangles across through the blue rectangle; this extended cut is highlighted in the partition on the right of Figure 5.

The left and right partitions shown in Figure 5 correspond, respectively, to the left and middle trees of Figure 6. Notice further that the middle tree in Figure 6 has the same special form as discussed above: it is a full binary tree with the same decision rule at all nodes at the same level. We can therefore form an equivalent tree by permuting the decision rules across levels, yielding the tree on the right of Figure 6. Unlike in the earlier examples, the left and right trees in Figure 6 represent different partitions of $[0, 1]^2$: the left

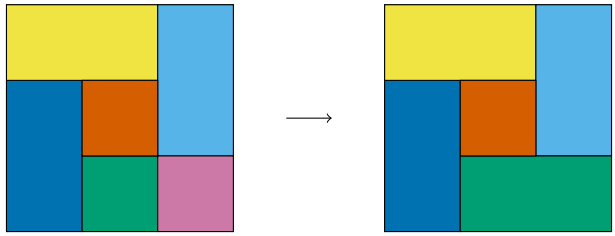


Figure 7. An example of a line which cannot be contracted. The right image cannot be constructed via trees since the decision at the root must create a line spanning opposite sides, but no such spanning line exists.

tree partitions the space into three rectangles while the right partitions it into four. However, because they induce the same partition of the points — and hence, same likelihood values — we view them as equivalent.

To implement an MCMC algorithm, moves must be reversible. Hence it is worth noting that not only can one extend lines in the box diagram, but also one can delete them. It is important when deleting a line, that the resulting diagram still corresponds to a tree. In Figure 7, the right diagram does not correspond to a tree. We conjecture that the condition for a diagram to correspond to a tree is that the diagram must have at least one vertical or horizontal line that goes across the whole diagram, and this same condition must hold recursively in each of the halves that this spanning line creates.

3.4. ***Next steps.*** In higher dimensions (i.e., $p > 2$), constructing equivalent trees is somewhat harder. However, if the tree is a subtree of a full binary tree that uses the same decision at each node at a given level, then we can still permute the decision rules across levels as in Figure 2. Similarly, if we can safely extend or contract cuts as in Figure 5, then we can form equivalent trees in higher dimensions.

4. Phase transition in 3-way contingency tables

This section arose from the Questions and Consulting seminar by Hanbaek Lyu and subsequent discussions amongst Miles Bakenhus, Elizabeth Gross, Max Hill, Vishesh Karwa, and Sonja Petrović. The sharp phase transition phenomenon relates to two distributions on the space of tables with fixed marginal totals, also known as the fiber of a log-linear model with those marginals as sufficient statistics. In this problem, the question is how well the hypergeometric distribution can approximate the uniform distribution on the fiber. In the case of two-dimensional tables, the answer toggles between "really well" and "rather poorly" at a phase transition point that can be explained by a condition on the margins of the table. This project translates the setup for the problem into the contingency table language used in algebraic statistics, allowing us to extend the basic notions needed to generalize the results to multiway tables.

4.1. ***Introduction.*** Log-linear models for cross-classified categorical data — contingency tables — have a long history in statistics [9; 21] and appear in a broad variety of applications including ecology, biology, educational testing, and network science. These models were also some of the first studied in the modern algebraic statistics literature that took off in the 1990s, leading to a vast literature on sampling using

algebraic and hybrid techniques, testing model fit using exact conditional tests, likelihood geometry of log-linear models, the existence of maximum likelihood estimators (MLEs), and connections to classical graphical models and, more recently, colored graphical models. While there are many flavors and variations of how contingency tables appear in the algebraic statistics literature, one typically studies models on tables under the multinomial, Poisson, or product-multinomial sampling scheme. It is well known that the MLEs are the same under all three sampling schemes; see, for example, [21, Chapter 3].

In this work we consider another sampling scheme, namely, geometric. This distribution makes a significant appearance in another vast collection of literature on contingency tables, here referred to as "the phase-transition literature", which is detailed in the next section. One common thread that appears in both the categorical data analysis and phase transition literature is the use of zero-margin tables, commonly referred to as *moves*, to sample the space of tables given fixed marginals. The algebraic statistics literature derives collections of such moves called *Markov bases* using techniques in computational algebraic geometry and combinatorial commutative algebra. Markov bases contain moves guaranteed to connect all sets of tables for a given choice of margins. Markov bases are theoretically defined and studied for many models and types of table margins (see [4] for a recent overview and [19] whose introductory section reviews the early theoretical considerations and statistical applications).

The table margins are sufficient statistics when considering discrete exponential family models. This space of tables with fixed values of sufficient statistics is called *the fiber of the log-linear model*. Sampling fibers provides a bona fide algorithm for testing model goodness of fit of every such exponential family. Fibers are reference sets for the sampling, while the desired distribution for the exact conditional test is the conditional distribution of tables given the marginals. For the Poisson and multinomial sampling schemes, this distribution is hypergeometric. On the other hand, the uniform distribution on the contingency tables also plays a central role in testing for the presence of interactions in cooccurrence tables, particularly cooccurrence (between species i and habitat j) tables arising in ecology [14; 41; 24; 12]. In these contexts, the uniform distribution subject to the margin is taken as a null hypothesis of no interaction among species without any particular modeling assumption. Diaconis [16] uses the uniform distribution on the fiber for *conditional volume tests* under the multinomial sampling scheme. Further, we will see in Section 4.3 that if the cells follow a geometric distribution, then the log-linear model's conditional distribution given the set of margins is, in fact, uniform (see Section 4.3.2).

In the next section we focus on two-dimensional tables and discuss the previous work on phase transitions related to uniformly sampling the space of tables given marginal totals. In Section 4.3 we compute a starting set of examples for three-dimensional tables, and close with a discussion in Section 4.3.6 on connections to algebraic statistics.

4.2. *2-way contingency tables.*

4.2.1. *Phase transition in random two-way tables with given margins.* Two-way contingency tables (CTs) are $m \times n$ matrices of nonnegative integer entries with prescribed row sums $\boldsymbol{r} = (a_1, \ldots, a_m)$ and columns sums $\boldsymbol{c} = (b_1, \ldots, b_n)$ called *margins*, where by $\mathcal{M}(\boldsymbol{r}, \boldsymbol{c})$ we denote the set of all such tables. They are fundamental objects in statistics for studying dependence structure between two or more variables and also correspond to bipartite multigraphs with given degrees and play an important role

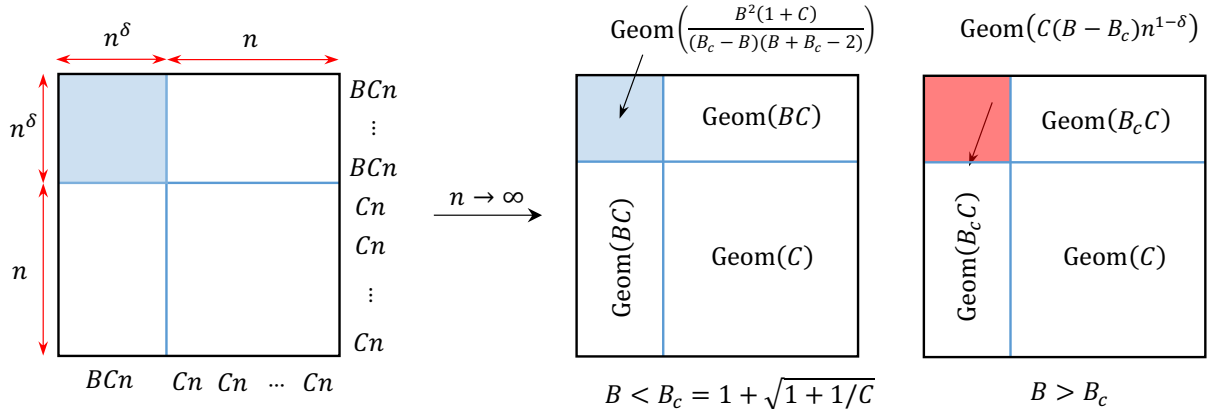


Figure 8. Left: contingency table with parameters n, δ, B and C. First $\lfloor n^\delta \rfloor$ rows and columns have margins $\lfloor BCn \rfloor$, the last n rows and columns have margins $\lfloor Cn \rfloor$. Right: Limiting distributions of the entries in the uniform contingency table X in the subcritical $B < B_c = 1+\sqrt{1+1/C}$, left, and supercritical $B > B_c$, right, regimes for thick bezels $\frac{1}{2} < \delta < 1$. Geom(λ) denotes geometric distribution with mean λ.

in combinatorics and graph theory; see, e.g., [6]. *Counting* their number $|\mathcal{M}(\boldsymbol{r}, \boldsymbol{c})|$ and *sampling* an element from $\mathcal{M}(\boldsymbol{r}, \boldsymbol{c})$ uniformly at random are two fundamental problems concerning CTs with many connections and applications to other fields [10] (e.g., testing hypothesis on cooccurrence of species in ecology [14]). A historic guiding principle to these problems is the *independent heuristic*, which was introduced by I. J. Good as far back as 1950 [23]. The heuristic states that the constraints for the rows and columns of the table are asymptotically independent as the size of the table grows to infinity. This yields a simple yet surprisingly accurate formula that approximates the count $|\mathcal{M}(\boldsymbol{r}, \boldsymbol{c})|$. The independence heuristic also implies the hypergeometric (or Fisher–Yates) distribution should approximate the uniform distribution on $\mathcal{M}(\boldsymbol{r}, \boldsymbol{c})$.

Both of these implications of the independent heuristic have been verified when the margins are constant or have a bounded ratio close to one [11]. However, when the margins are far from being constant, Barvinok [7] conjectured that there is a drastic difference between the uniform and hypergeometric distribution on $\mathcal{M}(\boldsymbol{r}, \boldsymbol{c})$. This was based on investigating CTs with what is known as Barvinok margin, a symmetric linear margin that has two values wherein a vanishing fraction has the larger value. That is, consider an $n \times n$ margin $(\boldsymbol{r}, \boldsymbol{c})$ where $\boldsymbol{r} = \boldsymbol{c} \in \mathbb{N}^n$ and the first $\lfloor n^\delta \rfloor$ coordinates have value $\lfloor BCn \rfloor$ and the rest $\lfloor Cn \rfloor$, where $B \geq 1$, $C > 0$ and $\delta \in [0, 1)$ are parameters. Barvinok and Hartigan [8] showed that the independent heuristic gives a large undercounting of CTs. The work of Dittmer, Lyu, Pak [17] and Lyu and Pak [35] provided the first complete answer to this puzzle; *CTs exhibit a sharp phase transition when the heterogeneity of margins exceeds a certain critical threshold.* For instance, the hypergeometric distribution correctly approximates the uniform distribution for $B < B_c = 1+\sqrt{1+1/C}$, but does so drastically differently for $B > B_c$ (see Figure 8). Such a sharp phase transition gives a probabilistic answer to the statistical question of why sampling a uniformly distributed CT is hard. This result settles

Barvinok's conjecture [7] (except for the special case of $\delta \in [0, \frac{1}{2}]$). In [35], Lyu and Pak obtained a similar phase transition result for CTs, this time from the perspective of the counting problem. Roughly speaking, the rows and columns of CTs are asymptotically independent (and hence the independence heuristic is correct) when the ratio B between the two margins is strictly less than the critical threshold B_c, but suddenly they become positively correlated as soon as B exceeds B_c and the independence heuristic gives exponential undercounting.

4.2.2. *Barvinok's typical table and the mechanism of phase transition.* The key insight in [17; 35] is that the uniformly random CT, say X, with given margins concentrates around a deterministic table called the "typical table", a notion first introduced by Barvinok [7]. Roughly speaking, this is the $m \times n$ *real-valued* table with margin $(\boldsymbol{r}, \boldsymbol{c})$ that maximizes a "geometric entropy function". More precisely, let $\mathcal{P}(\boldsymbol{r}, \boldsymbol{c}) \subseteq \mathbb{R}^{mn}_{\geq 0}$ denote the transportation polytope for margin $(\boldsymbol{r}, \boldsymbol{c})$. For each $X = (X_{ij}) \in \mathcal{P}(\boldsymbol{r}, \boldsymbol{c})$, define a strictly concave function

$$g(X) = \sum_{1 \leq i \leq n,\ 1 \leq j \leq m} f(X_{ij}), \quad \text{where } f(x) = (x+1)\log(x+1) - x \log x. \tag{1}$$

The *typical table* $Z \in \mathcal{P}(\boldsymbol{r}, \boldsymbol{c})$ for $\mathcal{M}(\boldsymbol{r}, \boldsymbol{c})$ is defined to be the unique maximizer of g among all real-valued tables with margin $(\boldsymbol{r}, \boldsymbol{c})$:

$$Z = \operatorname*{arg\,max}_{X \in \mathcal{P}(\boldsymbol{r}, \boldsymbol{c})} g(X). \tag{2}$$

The underlying mechanism of the sharp phase transition of uniformly random CTs with Barvinok margin (in Figure 8) established in [17; 35] is the sharp phase transition of the typical table Z, which can be shown by analyzing how the solution of the strictly concave optimization problem (2) that defines the typical table changes as one varies the margin.

Open problems. Here we state three open problems related to the phase transition problem.

Problem 4. The phase transition in uniformly random CTs with Barvinok margin (Figure 8) is established for the "thick bezel" case $\delta > \frac{1}{2}$. Can one establish a similar phase transition for the "thin bezel" case $0 \leq \delta \leq \frac{1}{2}$? (Phase transition in the typical table is established for all $\delta \in [0, 1]$; cf. [17, Lemma 5.1].)

Problem 5. Can we show the phase transition of the typical table when the margins assume three or more distinct values? For instance, three values An, Bn, Cn for margins for 3×3 block CT of size $n \times n$, phase transition in functions of B/A and C/A (see [17, Lemma 5.1]). In general, can one characterize all phase transitions in typical tables with respect to margin?

Problem 6. Can one show a similar sharp phase transition behavior in [17] for multiway CTs? For example, consider $n \times n \times n$ contingency tensors with margins assuming two values BCn and Cn. *Possible approach*: Develop a parallel "typical tensor" theory for uniformly random contingency tensors, and show phase transition in typical tensors as one varies the margin. Use "transference principle" to derive the behavior of uniformly random CTs from the underlying statistical model with independent entries.

4.3. ***3-way tables with plane-sum margins.*** In this section, we provide a first study for ongoing work on sharp phase transition in 3-way contingency tables and Markov bases [5]. This includes the statement of a preliminary result in sharp phase transition on 3-way plane-sum contingency tables (Theorem 8) and a sketch of proof. For more details we refer the interested readers for the upcoming full paper [5].

For 3-way tables, one can define more than one log-linear model. Here we consider the *model of independence*, which in log-linear model notation from [21] is given by the margins [1][2][3]. We will refer to these margins as 1-margins, or plane-sums of the table. This model is decomposable and as such is known to have a quadratic Markov basis [18]. In other words, the set of contingency tables $Y = (Y_{ijk})$ with fixed 1-margins Y_{i++}, Y_{+j+}, and Y_{++k} is *connected* by moves containing exactly two $+1$'s and two -1's, arranged in pairs of levels of the table so that the 1-margins are zero. It is worth noting that the model of independence stands in stark contrast with the no-three-factor interaction model [21], in which one uses 2-margins — line sums, rather than plane sums — as sufficient statistics. In that model, Markov bases can be arbitrarily complicated [15].

4.3.1. *The model of independence.* Fix a base measure μ on $\mathbb{Z}_{\geq 0}$. For each exponential tilting parameter $\theta \in \mathbb{R}$, define the exponentially tilted measure μ_θ by

$$\frac{d\mu_\theta}{d\mu}(x) = e^{-\theta x - \psi(\theta)}, \quad \psi(\theta) := \log\left(\sum_{k=0}^{\infty} e^{-k\theta}\mu(k)\right), \tag{3}$$

where ψ above is the log partition function. Let $\Theta := \{\theta : \psi(\theta) < \infty\}$, which is the set of exponential tilting parameters that makes the tilted measure μ_θ a probability measure. Note that if $\theta \in \Theta$ and $X \sim \mu_\theta$,

$$\psi'(\theta) = \mathbb{E}_{\mu_\theta}[X] \quad \text{and} \quad \psi''(\theta) = \operatorname{Var}_{\mu_\theta}(X) > 0. \tag{4}$$

The primary example is when the base measure μ is the counting measure, in which case $\Theta = (0, \infty)$ and μ_θ becomes the geometric distribution on nonnegative integers with success probability $1 - e^{-\theta}$, which we denote $\text{Geom}(1 - e^{-\theta})$:

$$\mathbb{P}\big(\text{Geom}(1 - e^{-\theta}) = k\big) = e^{k\theta}(1 - e^{-\theta}) \quad \text{for } k = 0, 1, \ldots \tag{5}$$

$$= e^{k\theta - \psi(\theta)}. \tag{6}$$

Note that in this case,

$$\mathbb{E}\big[\text{Geom}(1 - e^{-\theta})\big] = \psi'(\theta) = \frac{e^{-\theta}}{1 - e^{-\theta}} = \frac{1}{e^{\theta} - 1}. \tag{7}$$

Now define a model of $n_1 \times n_2 \times n_3$ table $Y = (Y_{ijk})$ with independent entries with marginal distribution

$$Y_{ijk} \sim \mu_{\theta_{ijk}}, \tag{8}$$

where θ_{ijk} is the exponential tilting parameter for the (i, j, k) entry. We further make a "rank-1" assumption to the model (8). That is, fix vector parameters $\alpha \in \mathbb{R}^{n_1}_{\geq 0}$, $\beta \in \mathbb{R}^{n_1}_{\geq 0}$, and $\gamma \in \mathbb{R}^{n_1}_{\geq 0}$. Then we assume that each exponential tilting parameter θ_{ijk} is given as the sum

$$\theta_{ijk} = \alpha_i + \beta_j + \gamma_k. \tag{9}$$

In this case, we denote

$$Y = (Y_{ijk}) \sim \mu_{\alpha,\beta,\gamma}. \tag{10}$$

This model is the hierarchical log-linear specification of the model of independence [9], whose sufficient statistics are one-dimensional table marginals. To see this, we note that the model can be written in the exponential family form as follows:

$$\mathbb{P}(\{y_{ijk}\}) = \prod_{ijk} e^{y_{ijk}\theta_{ijk}}(1 - e^{\theta_{ijk}}) \tag{11}$$

$$= \exp\left(\sum_{ijk} y_{ijk}\theta_{ijk} - \sum_{ijk} \psi(\theta_{ijk})\right) \tag{12}$$

$$= \exp\left(\sum_{i} \alpha_i \left(\sum_{jk} y_{ijk}\right) + \sum_{j} \beta_j \left(\sum_{ik} y_{ijk}\right) + \sum_{k} \gamma_k \left(\sum_{ij} y_{ijk}\right) - \sum_{ijk} \psi(\theta_{ijk})\right) \tag{13}$$

$$= \exp\left(\sum_{i} \alpha_i y_{i++} + \sum_{j} \beta_j y_{+j+} + \sum_{k} \gamma_k y_{++k} - \sum_{ijk} \psi(\theta_{ijk})\right). \tag{14}$$

The final equation is in the exponential family form $h(y)\exp\big(t(y)^T\eta(\theta) - \psi(\theta)\big)$, where $h(y) = 1$ is the base measure, $t(y) = \big(\{y_{i++}\}_{i=1}^{n_1}, \{y_{+j+}\}_{j=1}^{n_2}, \{y_{++k}\}_{k=1}^{n_3}\big)$ is the vector of sufficient statistics, and $\eta(\theta) = \big(\{\alpha_i\}_{i=1}^{n_1}, \{\beta_j\}_{j=1}^{n_2}, \{\gamma_k\}_{k=1}^{n_3}\big)$ is the vector of natural parameters.

4.3.2. *Uniform conditional distribution on the space of tables.* Fix vectors $a \in \mathbb{Z}_{\geq 0}^{n_1}$, $b \in \mathbb{Z}_{\geq 0}^{n_2}$, and $c \in \mathbb{Z}_{\geq 0}^{n_3}$ such that $\|a\|_1 = \|b\|_1 = \|c\|_1 =: N$, where N denotes the total sum. Define

$$\mathcal{T}(a, b, c) = \left\{ X \in \mathbb{Z}_{\geq 0}^{n_1 \times n_2 \times n_3} \;\middle|\; \begin{array}{l} X_{i++} = a_i,\ X_{+j+} = b_i,\ X_{++k} = c_i \\ \text{for all } (i, j, k) \in [n_1] \times [n_2] \times [n_3] \end{array} \right\}, \tag{15}$$

which is the set of all 3-way contingency tables with plane-sum margin (a, b, c). This set is called the *fiber* of the given marginal counts under the model of independence.

The probability of Y conditional on satisfying the margin (a, b, c) is uniform over $\mathcal{T}(a, b, c)$. To see this, note that for each $X = (x_{ijk}) \in \mathcal{T}(a, b, c)$, the log likelihood of observing X under the Y-model is

$$\log \mathbb{P}(Y = X) - \sum_{i,j,k} \log \mu(x_{ijk}) \tag{16}$$

$$= (\alpha_i + \beta_j + \gamma_k) - \psi(\alpha_i + \beta_j + \gamma_k) \tag{17}$$

$$= \sum_{i} x_{i++}\alpha_i + \sum_{j} x_{+j+}\beta_j + \sum_{k} x_{++k}\gamma_k - \sum_{i,j,k} \psi(\alpha_i + \beta_j + \gamma_k) \tag{18}$$

$$= \sum_{i} a_i\alpha_i + \sum_{j} b_i\beta_j + \sum_{k} c_k\gamma_k - \sum_{i,j,k} \psi(\alpha_i + \beta_j + \gamma_k) \tag{19}$$

$$=: \ell^{(a,b,c)}(\alpha, \beta, \gamma). \tag{20}$$

Notice that the conditional log-likelihood $\ell^{(a,b,c)}(\alpha, \beta, \gamma)$ defined above does not depend on the particular choice of $X \in \mathcal{T}(a, b, c)$, but only to the margin (a, b, c) and the exponential tilting parameter (α, β, γ). Therefore, if the base measure μ is uniform (i.e., the counting measure), the log-likelihood is uniform over the fiber $\mathcal{T}(\boldsymbol{r}, \boldsymbol{c})$. Thus, in this case, the law of Y conditional on being in $\mathcal{T}(a, b, c)$ is uniform.

In fact, a similar statement is true for general k-way contingency table models under the geometric sampling scheme. For completeness, we provide a short proof of this fact:

Lemma 7. *Let $\boldsymbol{y}$ denote a k-way contingency table in its vectorized form, and let*

$$\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}; \theta) = h(\boldsymbol{y}) \exp\big(t(\boldsymbol{y})^T \eta(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta})\big)$$

define a exponential family model on $\boldsymbol{y}$ where $h(\boldsymbol{y})$ is the base measure, θ is a vector of parameters, $t(\boldsymbol{y})$ is the vector of sufficient statistics. Then

$$\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y} | t(\boldsymbol{y}) = \boldsymbol{b}) = \frac{h(\boldsymbol{y})}{\sum_{\boldsymbol{y}' \in \mathcal{T}(\boldsymbol{b})} h(\boldsymbol{y}')}, \tag{21}$$

where $\mathcal{T}(\boldsymbol{b}) = \{\boldsymbol{y} : t(\boldsymbol{y}) = \boldsymbol{b}\}$ is the set of all tables whose sufficient statistics are equal to $\boldsymbol{b}$.

Under the Poisson and multinomial sampling schemes on the cells, this conditional distribution on the fiber is hypergeometric. Under geometric, the conditional distribution on the fiber is the uniform distribution.

Proof. Note that the distribution of the sufficient statistics has the following form:

$$\mathbb{P}(t(\boldsymbol{y}) = \boldsymbol{b}) = \sum_{\boldsymbol{y}' \in \mathcal{T}(\boldsymbol{b})} \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}') \tag{22}$$

$$= \sum_{\boldsymbol{y}' \in \mathcal{T}(\boldsymbol{b})} h(\boldsymbol{y}') \exp\big(t(\boldsymbol{y}')^T \eta(\theta) - \psi(\theta)\big) \tag{23}$$

$$= \exp\big(\boldsymbol{b}^T \eta(\theta) - \psi(\theta)\big) \Bigg(\sum_{\boldsymbol{y}' \in \mathcal{T}(\boldsymbol{b})} h(\boldsymbol{y}') \Bigg), \tag{24}$$

Hence, we have

$$\mathbb{P}(\boldsymbol{y} | t(\boldsymbol{y}) = \boldsymbol{b}) = \begin{cases} \dfrac{\mathbb{P}(\boldsymbol{y})}{\mathbb{P}(t(\boldsymbol{y}) = \boldsymbol{b})} & \text{if } \boldsymbol{y} \in \mathcal{T}(\boldsymbol{b}), \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

$$= \frac{h(\boldsymbol{y}) \exp\big(\boldsymbol{b}^T \eta(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta})\big)}{\exp\big(\boldsymbol{b}^T \eta(\theta) - \psi(\theta)\big) \big(\sum_{\boldsymbol{y}' \in \mathcal{T}(\boldsymbol{b})} h(\boldsymbol{y}')\big)} = \frac{h(\boldsymbol{y})}{\sum_{\boldsymbol{y}' \in \mathcal{T}(\boldsymbol{b})} h(\boldsymbol{y}')}. \tag{26}$$

It is now a direct consequence of this fact that when the underlying sampling scheme is a Poisson, or multinomial distribution, the base measure

$$h(\boldsymbol{y}) = \frac{N!}{\prod_i y_i!},$$

where $N = \sum_i y_i$ and i loops over the multi-index $\mathcal{I}$ of the k-way contingency table. The corresponding distribution on the fiber is hypergeometric:

$$\frac{\big(\prod_i y_i!\big)^{-1}}{\sum_{\boldsymbol{y}' \in \mathcal{T}(\boldsymbol{b})} \big(\prod_i y_i'\big)^{-1}}. \tag{27}$$

When the sampling scheme is geometric, as in our case, then the base measure is $h(\boldsymbol{y}) = 1$, then the corresponding conditional distribution is uniform:

$$\frac{1}{\sum_{\boldsymbol{y}' \in \mathcal{T}(\boldsymbol{b})} 1} = \frac{1}{|\mathcal{T}(\boldsymbol{b})|}. \tag{28}$$

This completes the proof. □

4.3.3. *The maximum likelihood problem.* The MLE for the exponential tilting parameter (α, β, γ) given margin (a, b, c) is given by

$$\max_{\alpha>0,\beta>0,\gamma>0} \left[\ell^{(a,b,c)}(\alpha, \beta, \gamma) = \sum_i a_i \alpha_i + \sum_j b_i \beta_j + \sum_k c_k \gamma_k - \sum_{i,j,k} \psi(\alpha_i + \beta_j + \gamma_k) \right]. \tag{29}$$

Note that the log-likelihood function $\ell^{(a,b,c)}$ is strictly concave since

$$\psi''(\theta) = \mathrm{Var}(\mathrm{Geom}(1 - e^{\theta})) > 0. \tag{30}$$

Thus if a solution for (29) exists, then it is unique and is a critical point. Thus, solving the optimization problem (29) is equivalent to solving the following MLE equations:

$$\begin{cases} \sum_{j,k} \psi'(\alpha_i + \beta_j + \gamma_k) = a_i, & i = 1, \ldots, n_1, \\ \sum_{k,i} \psi'(\alpha_i + \beta_j + \gamma_k) = b_j, & j = 1, \ldots, n_2, \\ \sum_{i,j} \psi'(\alpha_i + \beta_j + \gamma_k) = c_k, & k = 1, \ldots, n_3. \end{cases} \tag{31}$$

Using (7), the above is equivalent to

$$\begin{cases} \sum_{j,k} \frac{1}{\exp(\alpha_i + \beta_j + \gamma_k) - 1} = a_i, & i = 1, \ldots, n_1, \\ \sum_{k,i} \frac{1}{\exp(\alpha_i + \beta_j + \gamma_k) - 1} = b_j, & j = 1, \ldots, n_2, \\ \sum_{i,j} \frac{1}{\exp(\alpha_i + \beta_j + \gamma_k) - 1} = c_k, & k = 1, \ldots, n_3. \end{cases} \tag{32}$$

More concisely, this is simply requiring that the expected table $\mathbb{E}[Y]$ satisfies the margin (a, b, c):

$$Z := \mathbb{E}[Y] \in \mathcal{T}(a, b, c). \tag{33}$$

This corresponds to the standard critical equation setup in log-linear models, and here the sufficient statistics are the table plane sums.

4.3.4. *The computational problem: plane sums of 3-way tables for 3-way Barvinok margin.* Now we consider the three-way plane-sum Barvinok margin for $(n \times n \times n)$ three-way contingency tables:

$$a = b = c = (Bn^2, n^2, \ldots, n^2) \in \mathbb{R}^n. \tag{34}$$

The resulting MLE problem is to find the parameters (α, β, γ) such that the plane sums of the expected table Z as in (33) match the values Bn^2 or n^2. This amounts to satisfying the MLE equations in (32) for the three-way Barvinok margin (34). By symmetry, without loss of generality we can assume

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_2) = \beta = \gamma. \tag{35}$$

Slices can be cut from Z in each of three directions, however due to (35), Z is symmetric and hence it is sufficient to choose just one direction, as the plane sums of slices cut in the other two directions will be the same. Therefore it will suffice to compute the plane sums Z_{++k}, for $k \in [n]$. Observe that for each k, the slice of Z corresponding to the sum Z_{++k} can be decomposed into four blocks of entries: a 1×1 block, an $(n-1) \times (n-1)$ block, and two blocks with dimensions $1 \times (n-1)$ and $(n-1) \times 1$. Due to (35), all entries in a block have the same value. Using this observation, it follows that the MLE equation (32) in this case reduces to 0

$$Z_{++k} = Z_{11k} + 2(n-1)Z_{12k} + (n-1)^2 Z_{22k} = \begin{cases} Bn^2 & \text{if } k = 1, \\ n^2 & \text{if } k \geq 2. \end{cases} \tag{36}$$

Letting

$$P := e^{\alpha_1} \quad \text{and} \quad Q := e^{\alpha_2},$$

we have

$$Z_{111} = \frac{1}{P^3 - 1}, \quad Z_{121} = Z_{112} = \frac{1}{P^2 Q - 1}, \quad Z_{221} = Z_{122} = \frac{1}{Q^2 P - 1}, \quad Z_{222} = \frac{1}{Q^3 - 1}. \tag{37}$$

Since $Z_{ijk} \geq 0$, we must have $P, Q \geq 1$. Also, this change of variables rewrites (36) as

$$\begin{cases} \frac{1}{P^3-1} + 2(n-1)\frac{1}{P^2Q-1} + \frac{(n-1)^2}{Q^2P-1} = Bn^2, \\ \frac{1}{P^2Q-1} + 2(n-1)\frac{1}{PQ^2-1} + \frac{(n-1)^2}{Q^3-1} = n^2. \end{cases} \tag{38}$$

Since all terms involved are nonnegative, dropping the first two terms in the second equation in (38), it follows that $(n-1)^3/(Q^3-1) \leq n^3$, so

$$Q \geq \left(\left(\frac{n}{n-1}\right)^3 + 1\right)^{1/3} = 2 + O(n^{-1}). \tag{39}$$

Thus $Q^3 > \frac{3}{2}$ for all sufficiently large n. It follows that $Z_{ijk} = O(1)$ whenever $(i, j, k) \neq (1, 1, 1)$. Reusing the second equation in (38) with this fact, we have

$$\frac{(n-1)^2}{Q^3 - 1} = n^2 - O(n), \tag{40}$$

so we deduce

$$Q^3 = 2 + O(n^{-1}). \tag{41}$$

Thus, for all values of the ratio parameter B, $Q \sim 2^{1/3}$. On the contrary, as we will see shortly, the asymptotic of P will depend on B critically.

4.3.5. *Sharp phase transition in the maximum-likelihood expected table.* Below is the result on sharp phase transition in the expected table under the MLE model with three-dimensional Barvinok margin $(Bn^2, n^2, \dots, n^2)$. This result is a three-way extension of [17, Lemma 5.1], which partially addresses Problem 5.

Theorem 8. *Let $Z = Z^{r,c} = \mathbb{E}_{\alpha,\beta,\gamma}[Y]$, where $Y \sim \mu_{\alpha,\beta,\gamma}$ and (α, β, γ) is an MLE for the rank-1 model for the plane-sum margin (a, b, c) with $a = b = c = (Bn^2, n^2, \dots, n^2) \in \mathbb{R}^n$. For $B_c := 1/(2^{2/3} - 1)$, the following hold*:

(i) (*subcritical regime*) *Suppose $B < B_c$. Then*

$$Z_{111} = \frac{1}{\left(\frac{B^{-1}+1}{B_c^{-1}+1}\right)^3 - 1} + O(n^{-1}), \qquad Z_{121} = Z_{112} = \frac{1}{2^{1/3}\left(\frac{B^{-1}+1}{B_c^{-1}+1}\right)^2 - 1} + O(n^{-1}), \tag{42}$$

$$Z_{221} = Z_{122} = \frac{1}{2^{2/3}\left(\frac{B^{-1}+1}{B_c^{-1}+1}\right) - 1} + O(n^{-1}), \qquad Z_{222} = 1 + O(n^{-1}). \tag{43}$$

(ii) (*supercritical regime*) *Suppose $B > B_c$. Then*

$$Z_{111} = (B - B_c)n^2 - O(n), \quad Z_{121} = Z_{112} = \frac{1}{2^{1/3} - 1} + O(n^{-1}), \tag{44}$$

$$Z_{221} = Z_{122} = \frac{1}{2^{2/3} - 1} + O(n^{-1}), \qquad Z_{222} = 1 + O(n^{-1}). \tag{45}$$

Sketch of proof. First assume $B < B_c$. Dropping the first two terms in the first equation in (38) and using the asymptotic of Q in (40),

$$\frac{(n-1)^2}{Q^2 P - 1} \le Bn^2. \tag{46}$$

Solving for P and using $B < B_c$, we get

$$P \ge Q^{-2}\left(B^{-1}\left(\frac{n-1}{n}\right)^2 + 1\right) \sim 2^{-2/3}(B^{-1} + 1) = \frac{B^{-1}+1}{B_c^{-1}+1} > 1. \tag{47}$$

It follows that

$$Z_{111} = \frac{1}{P^3 - 1} = O(1).$$

Hence the first two terms in the first equation in (38) are of order $O(n)$, so we get

$$\frac{(n-1)^2}{Q^2 P - 1} = Bn^2 - O(n). \tag{48}$$

Consequently, we deduce the following asymptotic of P in the subcritical regime:

$$P = \frac{B^{-1}+1}{B_c^{-1}+1} + O(n^{-1}). \tag{49}$$

Now using (37) and the asymptotic of Q in (40), we conclude the asymptotics of the entries of Z stated in (i).

Next, assume $B > B_c$. Recall that taking $P \searrow 1$, the largest n^2 asymptotic that the third term in the first equation in (38) is at most $B_c n^2$. Also recall that the second term in the first equation in (38) is of order $O(n)$. Therefore, since $B > B_c$, it follows that the first term $1/(P^3 - 1)$ must have contributions in the $(B - B_c)n^2$. We can obtain more precise asymptotics as follows. Using the first equation in (38) and that $Z_{121} = O(1)$,

$$\frac{1}{P^3 - 1} + \frac{(n-1)^2}{Q^2 P - 1} = Bn^2 - O(n). \tag{50}$$

Since $Q^3 = 2 + O(n^{-1})$, $P^3 = 1 + O(n^{-2})$, so

$$Z_{111} = \frac{1}{P^3 - 1} = (B - B_c)n^2 - O(n). \tag{51}$$

From this, we get

$$P^3 = 1 + \frac{n^{-2}}{B - B_c} + O(n^{-3}). \tag{52}$$

Then using (37) and the asymptotic of Q in (40), we can also conclude the asymptotics of the entries of Z stated in (ii). □

We add a remark on the practical implications of the phase transition behavior stated in Theorem 8. The phase transition in Theorem 8 tells us that the shape of the set of all contingency tables with a given margin is very "elongated" in the supercritical phase. Thus, sampling uniformly from this set and counting all elements in this set will likely be quite difficult.

In the two-way case, the fact that the counting problem is, in some sense, "easy" in the subcritical phase is established by Barvinok and Hartigan for the class of "δ-tame" margins [8]. These are the class of margins for which the corresponding typical table has entries uniformly bounded between δ and δ^{-1}. Specializing to the Barvinok margin, such a "δ-tame phase" resides in the subcritical phase. A more precise result is known for this case [35]: The independence heuristic [23] provides a good approximation of the number of contingency tables in the subcritical phase, but it results in a significant undercounting in the supercritical phase.

4.3.6. *Algebraic statistics and phase transitions.* We propose to connect the phase transition phenomena to Markov bases for sampling fibers and also to the geometry of marginal polytopes.

First, phase transitions in two-way tables happen for so-called tame fibers, that is, two $m \times n$ margins that are δ-tame, a concept introduced at the end of the previous section. From the point of view of discrete exponential families, this means that the MLE is away from the boundary of the marginal polytope, which is defined as the convex hull of all possible sufficient statistics for the exponential family. This implies that the phase transition can be also interpreted on an infinite sequence of polytopes: for small table sizes, the MLE is well in the relative interior of the polytope, but as the size grows, there is a threshold when the MLE comes too close (within δ) to the face of the polytope. It would be of great interest to perform some initial simulations in this direction.

Second, the moves used to connect the fiber in the 2-way table case provide a basis for a modified Markov chain to sample from the fiber uniformly, until the phase transition occurs with the mass blowup

on the corner (1, 1) cell of the cube. Naturally, we expect that this phenomenon generalizes to the 3-way independence model when the mass concentrates on the (1, 1, 1) cell. To this end, we close with an explicit description of Markov basis moves for the 3-way model of independence. These are the moves analogous to the moves in the 2-way table model that can be used to explore the fiber uniformly, under the "tame" regime before the phase transition occurs.

The Markov moves $m = (m_{ijk})$ for the independence model on the $n_1 \times n_2 \times n_3$ table $Y = (Y_{ijk})$ have entries $m_{i_1 j_1 k_1} = 1$ and $m_{i_2 j_2 k_2} = 1$ with one of

$$\begin{cases} m_{i_1 j_2 k_2} = -1, \\ m_{i_2 j_1 k_2} = -1, \end{cases} \quad \begin{cases} m_{i_1 j_2 k_2} = -1, \\ m_{i_2 j_2 k_1} = -1, \end{cases} \quad \text{or} \quad \begin{cases} m_{i_2 j_1 k_2} = -1, \\ m_{i_2 j_2 k_1} = -1, \end{cases}$$

where all other entries are zero, for each $i_1, i_2 \in [n_1]$, $j_1, j_2 \in [n_2]$, and $k_1, k_2 \in [n_3]$, where $i_1 \neq i_2$, $j_1 \neq j_2$, and $k_1 \neq k_2$. There are a total of $\frac{3}{2} n_1 n_2 n_3 (n_1 - 1)(n_2 - 1)(n_3 - 1)$ of these moves. Fixing $i_1 j_1 k_1 = (1, 1, 1)$ leaves $3(n_1 - 1)(n_2 - 1)(n_3 - 1)$ moves that apply to the entry Y_{111} in the table. Then, the ratio of applicable moves to nonapplicable moves is

$$\frac{2}{n_1 n_2 n_3} \to 0 \quad \text{as } n_1, n_2, n_3 \to \infty.$$

Simply the number of available moves does not reveal directly any phase transition phenomena, but this is likely the case because Markov moves connect all possible fibers. In forthcoming work we will study what moves are necessary and/or inapplicable for the particular margins are of the form (Bn^2, n^2, n^2).

Acknowledgements

This research was performed while the authors were visiting the Institute for Mathematical and Statistical Innovation, which is supported by the National Science Foundation grant DMS-1929348. The research of Gu is partially supported by National Science Foundation grant DMS-2210796. Gross and Curiel were supported by the National Science Foundation grant DMS-1945584. The research by Rodriguez is partially supported by the Alfred P. Sloan Fellowship. The Office of the Vice Chancellor for Research and Graduate Education at UW-Madison with funding from the Wisconsin Alumni Research Foundation partially supports Deshpande and Rodriguez. Petrović is partially supported by the Simons Foundation Collaboration Grant for Mathematicians #854770 and DOE/SC award #1010629.

References

[1] Y. Alexandr and A. Heaton, "Logarithmic Voronoi cells", *Algebr. Stat.* **12**:1 (2021), 75–95.

[2] E. S. Allman, H. Baños, R. Evans, S. Hoşten, K. Kubjas, D. Lemke, J. A. Rhodes, and P. Zwiernik, "Maximum likelihood estimation of the latent class model through model boundary decomposition", *J. Algebr. Stat.* **10**:1 (2019), 51–84.

[3] E. S. Allman, J. A. Rhodes, B. Sturmfels, and P. Zwiernik, "Tensors of nonnegative rank two", *Linear Algebra Appl.* **473** (2015), 37–53.

[4] F. Almendra-Hernández, J. A. De Loera, and S. Petrović, "Markov bases: a 25 year update", *J. Amer. Statist. Assoc.* **119**:546 (2024), 1671–1686.

[5] M. Bakenhus, E. Gross, M. Hill, V. Karwa, H. Lyu, and S. Petrović, "Sharp phase transition in 3-way contingency tables and Markov bases", In preparation.

[6] A. Barvinok, "Asymptotic estimates for the number of contingency tables, integer flows, and volumes of transportation polytopes", *Int. Math. Res. Not.* **2009**:2 (2009), 348–385.

[7] A. Barvinok, "What does a random contingency table look like?", *Combin. Probab. Comput.* **19**:4 (2010), 517–539.

[8] A. Barvinok and J. A. Hartigan, "An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums", *Trans. Amer. Math. Soc.* **364**:8 (2012), 4323–4368.

[9] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete multivariate analysis*: *theory and practice*, Springer, 2007.

[10] P. Brändén, J. Leake, and I. Pak, "Lower bounds for contingency tables via Lorentzian polynomials", *Israel J. Math.* **253**:1 (2023), 43–90.

[11] E. R. Canfield and B. D. McKay, "Asymptotic enumeration of integer matrices with large equal row and column sums", *Combinatorica* **30**:6 (2010), 655–680.

[12] Y. Chen, P. Diaconis, S. P. Holmes, and J. S. Liu, "Sequential Monte Carlo methods for statistical analysis of tables", *J. Amer. Statist. Assoc.* **100**:469 (2005), 109–120.

[13] H. A. Chipman, E. I. George, and R. E. McCulloch, "BART: Bayesian additive regression trees", *Ann. Appl. Stat.* **4**:1 (2010), 266–298.

[14] E. F. Connor and D. Simberloff, "The assembly of species communities: chance or competition?", *Ecology* **60**:6 (1979), 1132–1140.

[15] J. A. De Loera and S. Onn, "Markov bases of three-way tables are arbitrarily complicated", *J. Symbolic Comput.* **41**:2 (2006), 173–181.

[16] P. Diaconis and B. Efron, "Testing for independence in a two-way table: new interpretations of the chi-square statistic", *Ann. Statist.* **13**:3 (1985), 845–913.

[17] S. Dittmer, H. Lyu, and I. Pak, "Phase transition in random contingency tables with non-uniform margins", *Trans. Amer. Math. Soc.* **373**:12 (2020), 8313–8338.

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

[19] A. Dobra and S. Sullivant, "A divide-and-conquer algorithm for generating Markov bases of multi-way tables", *Comput. Statist.* **19**:3 (2004), 347–366.

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

[21] S. E. Fienberg, *The analysis of cross-classified categorical data*, 2nd ed., Springer, 2007.

[22] S. E. Fienberg, P. Hersh, A. Rinaldo, and Y. Zhou, "Maximum likelihood estimation in latent class models for contingency table data", pp. 27–62 in *Algebraic and geometric methods in statistics*, edited by P. Gibilisco et al., Cambridge University Press, 2010.

[23] I. J. Good, *Probability and the Weighing of Evidence*, Griffin, London, 1950.

[24] N. J. Gotelli, "Null model analysis of species co-occurrence patterns", *Ecology* **81**:9 (2000), 2606–2621.

[25] E. Gross and J. I. Rodriguez, "Maximum likelihood geometry in the presence of data zeros", pp. 232–239 in *ISSAC* 2014 — *Proceedings of the* 39*th International Symposium on Symbolic and Algebraic Computation*, edited by K. Nabeshima, ACM, New York, 2014.

[26] Y. Gu, "Generic identifiability of the DINA model and blessing of latent dependence", *Psychometrika* **88**:1 (2023), 117–131.

[27] Y. Gu, "Blessing of dependence: identifiability and geometry of discrete models with multiple binary latent variables", 2024. To appear in *Bernoulli*. arXiv 2203.04403

[28] Y. Gu and D. B. Dunson, "Bayesian pyramids: identifiable multilayer discrete latent structure models for discrete data", *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85**:2 (2023), 399–426.

[29] Y. Gu and G. Xu, "Partial identifiability of restricted latent class models", *Ann. Statist.* **48**:4 (2020), 2082–2107.

[30] J. Hauenstein, J. I. Rodriguez, and B. Sturmfels, "Maximum likelihood for matrices with rank constraints", *J. Algebr. Stat.* **5**:1 (2014), 18–38.

[31] V. Karwa, D. Pati, S. Petrović, L. Solus, N. Alexeev, M. Raič, D. Wilburne, R. Williams, and B. Yan, "Monte Carlo goodness-of-fit tests for degree corrected and related stochastic blockmodels", *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86**:1 (2024), 90–121.

[32] J. Kim and V. Ročková, "On mixing rates for Bayesian CART", 2023. arXiv 2306.00126

[33] K. Kubjas, E. Robeva, and B. Sturmfels, "Fixed points EM algorithm and nonnegative rank boundaries", *Ann. Statist.* **43**:1 (2015), 422–461.

[34] J. M. Landsberg, *Tensors: geometry and applications*, Graduate Studies in Mathematics **128**, American Mathematical Society, Providence, RI, 2012.

[35] H. Lyu and I. Pak, "On the number of contingency tables and the independence heuristic", *Bull. Lond. Math. Soc.* **54**:1 (2022), 242–255.

[36] M. T. Pratola, "Efficient Metropolis–Hastings proposal mechanisms for Bayesian regression tree models", *Bayesian Anal.* **11**:3 (2016), 885–911.

[37] C. Raicu, "3×3 minors of catalecticants", *Math. Res. Lett.* **20**:4 (2013), 745–756.

[38] J. I. Rodriguez and B. Wang, "The maximum likelihood degree of mixtures of independence models", *SIAM J. Appl. Algebra Geom.* **1**:1 (2017), 484–506.

[39] O. Ronen, T. Saarinen, Y. S. Tan, J. Duncan, and B. Yu, "A mixing time lower bound for a simplified version of BART", 2022. arXiv 2210.09352v1

[40] A. Seigal and G. Montúfar, "Mixtures and products in two graphical models", *J. Algebr. Stat.* **9**:1 (2018), 1–20.

[41] T. A. B. Snijders, "Enumeration and simulation methods for 0-1 matrices with given marginals", *Psychometrika* **56**:3 (1991), 397–417.

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

Received 2024-02-20. Revised 2024-06-19. Accepted 2024-07-30.

YULIA ALEXANDR: yulia@math.berkeley.edu
Department of Mathematics, University of California, Los Angeles, CA, United States

MILES BAKENHUS: mbakenhus@hawk.iit.edu
Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL, United States

MAIZE CURIEL: curielm@hawaii.edu
Department of Mathematics, University of Hawai'i at Mānoa, Honolulu, HI, United States

SAMEER K. DESHPANDE: sameer.deshpande@wisc.edu
Department of Statistics, University of Wisconsin, Madison, WI, United States

ELIZABETH GROSS: egross@hawaii.edu
Department of Mathematics, University of Hawai'i at Mānoa, Honolulu, HI, United States

YUQI GU: yuqi.gu@columbia.edu
Department of Statistics, Columbia University, New York, NY, United States

MAX HILL: max.hill1@ucr.edu
Department of Mathematics, University of California, Riverside, CA, United States

JOSEPH JOHNSON: josjohn@kth.se
Institutionen för Matematik, KTH Royal Institute of Technology, Stockholm, Sweden

BRYSON KAGY: bgkagy@ncsu.edu
Department of Mathematics, North Carolina State University, Raleigh, NC, United States

VISHESH KARWA: vishesh@temple.edu
Department of Statistics, Operations and Data Science, Temple University, Philadelphia, PA, United States

JIAYI LI: jiayi.li@g.ucla.edu
Department of Statistics and Data Science, University of California, Los Angeles, CA, United States

HANBAEK LYU: hlyu36@wisc.edu
Department of Mathematics, University of Wisconsin, Madison, WI, United States

SONJA PETROVIĆ: sonja.petrovic@iit.edu
Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL, United States

JOSE ISRAEL RODRIGUEZ: jose@math.wisc.edu
Department of Mathematics, University of Wisconsin, Madison, WI, United States

msp.org/astat

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

The subscription price for 2024 is US $235/year for the electronic version, and $290/year (+$15, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

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.

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

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

Algebraic Statistics

2024 15 : 2

Special Issue

IMSI "Apprenticeship Week" Projects