Chipsplitting Games: A Combinatorial Approach to Classifying One-Dimensional Discrete Statistical Models with Rational Maximum Likelihood Estimator

by

Viet-Duc Nguyen

A thesis submitted in partial satisfaction of the

requirements for the degree of

Master of Science

in

Mathematics

in the

Mathematics Department

of

Technische Universität Berlin

Supervisors:

Professor Carlos Améndola
Professor Christian Haase

Winter 2024

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit eigenständig ohne Hilfe Dritter und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe. Alle Stellen die den benutzten Quellen und Hilfsmitteln unverändert oder sinngemäß entnommen sind, habe ich als solche kenntlich gemacht.

Sofern generative KI-Tools verwendet wurden, habe ich Produktnamen, Hersteller, die jeweils verwendete Softwareversion und die jeweiligen Einsatzzwecke (z.B. sprachliche Überprüfung und Verbesserung der Texte, systematische Recherche) benannt. Ich verantworte die Auswahl, die Übernahme und sämtliche Ergebnisse des von mir verwendeten KI-generierten Outputs vollumfänglich selbst.

Die Satzung zur Sicherung guter wissenschaftlicher Praxis an der TU Berlin vom 8. März 2017 habe ich zur Kenntnis genommen.

Ich erkläre weiterhin, dass ich die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt habe.

Berlin, den

Zusammenfassung in deutscher Sprache

Die vorliegende Ausarbeitung setzt die Forschung von Arthur Bik und Orlando Marigliano zur Klassifizierung ein-dimensionaler diskreter statistischer Modelle mit rationalen Maximum Likelihood Schätzern unter Verwendung fundamentaler Modelle fort. Wir erzielen bedeutende Fortschritte beim Beweis zur endlichen Anzahl der fundamentalen Modelle im Wahrscheinlichkeitssimplex Δ_5. Zudem bestimmen wir die Anzahl der fundamentalen Modelle im Simplex Δ_6 mit einem maximalen Grad von 11.

Abstract

Chipsplitting Games: A Combinatorial Approach to Classifying One-Dimensional Discrete Statistical Models with Rational Maximum Likelihood Estimator

by

Viet-Duc Nguyen

Technische Universität Berlin

This paper continues the research of Arthur Bik and Orlando Marigliano on the classification of one-dimensional discrete statistical models with rational maximum likelihood estimators using fundamental models. We present a missing proof of an algorithm from their work. Furthermore, we make significant progress in proving the finite number of fundamental models in the probability simplex Δ_5. We also determine the number of fundamental models in the simplex Δ_6 with a maximum degree of 11.

Contents

Chapter 1

Introduction

In statistics we come across various collections of probability distributions, such as the normal distribution, Poisson distribution, and binomial distribution. These distributions are used to model random variables in applications, and are referred to as *statistical models*. Precisely, a statistical model is just a set of probability distributions. If the set contains only discrete distributions, we call it a *discrete statistical model*. In this case, discrete statistical models are just subsets of the probability simplex $\Delta_n \coloneqq \{p \in \mathbb{R}^{n+1} \mid \sum p_i = 1\}$.

A discrete distribution $p \in \mathcal{M} \subset \Delta_n$ from a discrete statistical model encapsulates the probabilities of observing the states $0, \ldots, n$, i.e. if $X \in \{0, \ldots, n\}$ is a discrete random variable, then the state $X = i$ occurs with probability p_i for all $i = 0, \ldots, n$. Say we have a binomial random variable X with $n+1$ states, then $p_i = \binom{n}{i}\theta^i(1-\theta)^{n-i}$ computes the probability of observing i successes in n trials with success probability $\theta \in [0,1]$. The set $\mathcal{M}$ of all probability distributions of that form, i.e. $\mathcal{M} = \left\{(\binom{n}{i}\theta^i(1-\theta)^{n-i})_{i=0}^n \mid \theta \in [0,1]\right\}$, is our first example of a discrete statistical model, and is known as the *binomial model*.

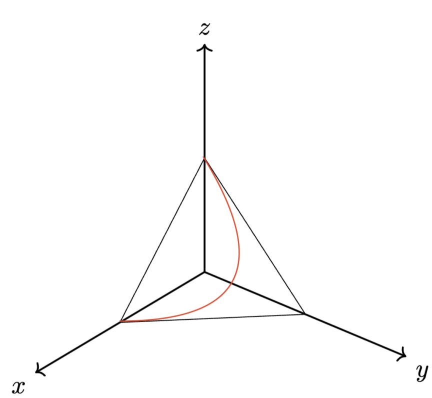

Figure 1.1: This figure shows the probability simplex Δ_2 with the binomial model (red curve). Every point on the curve is a binomial distribution.

Given a statistical model $\mathcal{M} \subset \Delta_n$ and data $u \in \mathbb{N}^{n+1}$, a typical problem in statistics is to find a distribution from a statistical model that best describes the data. "Best" can mean a lot of things, but in *maximum likelihood estimation* it means finding the distribution that maximizes the probability of observing the data; the map $\Phi : \Delta_n \to \mathcal{M}, u \mapsto \hat{p}$ that assigns the data u to a distribution $\hat{p} \in \mathcal{M}$ from the statistical model is called the *maximum likelihood estimator (MLE)*. This map is characterized by the property that $\hat{p}$ maximizes the log-likelihood function $\ell(p) = \sum u_i \log p_i$ for all $p \in \mathcal{M}$.

We focus on **one-dimensional discrete statistical models with rational MLE**. These are models $\mathcal{M}$ satisfying

- $\mathcal{M} = \text{image}(p)$ for some rational map $p = (p_0, \ldots, p_n) : I \to \Delta_n$ where p_i is rational, $I \subset \mathbb{R}$ is a union of closed intervals and $p(\partial I) \subset \partial \Delta_n$,
- all the $n+1$ coordinates of the maximum likelihood estimator Φ are rational functions in the data u.

There are two intriguing questions to ask about statistical models with rational MLE: the first one is about which *form* do the maximum likelihood estimators take; the second one is more concerned with the *classification* of the statistical models, i.e. can we divide these models into easier to understand classes? An answer to the first question was given by June Huh. He showed that if Φ is rational, then each of its coordinates is an alternative product of linear forms with numerator and denominator of the same degree, see [3, 2]. For the second question, Arthur Bik and Orlando Marigliano classified all one-dimensional discrete statistical models with rational MLE using *fundamental models* [1].

This thesis continues the work of Arthur Bik and Orlando Marigliano. In the first half, we present their classification results on how fundamental models serve as the building blocks of one-dimensional discrete models with rational MLE. In the second half, we establish and extend their finding that there are only finitely many fundamental models within the probability simplices Δ_n for $n \leq 4$. Due to the complexity of the problem, the cases $n \geq 5$ were left open. For the first time, we make significant progress for $n = 5$. Other contributions of this thesis include a new result on the number of fundamental models in Δ_6 with a maximum degree of 11, and a missing proof of a key algorithm from their work.

Chapter 2

Classification with Fundamental Models

In this chapter we present the classification of one-dimensional discrete statistical models with rational maximum likelihood estimator (MLE) using fundamental models. The classification is due to Arthur Bik and Orlando Marigliano [1].

> **Problem statement:** Can we find a class of easy to understand models that serve as building blocks for all one-dimensional discrete statistical models with rational MLE?

The answer to this question are *reduced* and *fundamental models*.

2.1 Parametrization

It turns out that one-dimensional discrete statistical models with rational MLE admit the following parametrization.

Proposition 2.1. *Let $\mathcal{M}$ be a one-dimensional discrete statistical models with rational maximum likelihood estimator. Then, there exists a map of the form*

$$p : [0,1] \to \Delta_n, \quad \theta \mapsto (w_k \theta^{i_k}(1-\theta)^{j_k})_{k=0}^n$$
$$i_k, j_k \in \mathbb{Z}_{\geq 0},\ w_k \in \mathbb{R}_{>0} \quad \forall k = 0, \ldots, n$$

such that $\mathcal{M} = \mathrm{image}(p)$.

We introduce some notation to simplify the proof of Proposition 2.1. Let $\mathcal{M} \subset \Delta_n$ be a one-dimensional discrete statistical model parametrized by rational functions $p_0 = \frac{g_0}{h_0}, \ldots, p_n = \frac{g_n}{h_n}$. Define b to be the least common multiple of $h_0, \ldots, h_n$ and $a_i \coloneqq bp_i$. Since $\sum p_k = 1$, we can multiply by b to obtain $\sum a_k = b$. We see that the polynomials $a_0, \ldots, a_n, b$

determine the statistical model $\mathcal{M}$, and have no common factors. The log-likelihood function is then given by

$$\begin{aligned}\ell(p) &= \sum u_i \log p_i \\ &= \sum u_i \log \frac{a_i}{b} \\ &= \sum u_i \log a_i - \sum u_i \log b.\end{aligned}$$

To find the maximum likelihood estimator, we need find all critical points of the log-likelihood function. This is equivalent to finding the roots of the gradient of the log-likelihood function

$$\ell(p(\theta))' = \sum u_k \frac{a_k'}{a_k} - \sum u_k \frac{b'}{b} = 0. \tag{2.1}$$

These equations are called the *score equations* in algebraic statistics, and the number of complex solutions to these equations for general data $u \in \mathbb{C}^{n+1}$ is called the *maximum likelihood degree* of the statistical model. This ML degree has an important meaning in algebraic statistics, as it determines the complexity of the model. We have the following relationship between the ML estimator and the ML degree.

Proposition 2.2. *Having rational maximum likelihood estimator can be expressed equivalently by saying that the maximum likelihood degree of the statistical model is one.*

Proof. Refer to [2] for a proof. □

To prove Proposition 2.1, we need the following lemma.

Lemma 2.3. *If $\mathcal{M}$ has rational MLE, then there are exactly two distinct complex linear factors in $a_0, \ldots, a_n$, and b.*

Proof. We prove the lemma in three steps:

- Let f be the product of all distinct complex linear factors in $a_0, \ldots, a_n, b$. If we multiply the score equations (2.1) by f, we get

$$f \cdot \ell(p(\theta))' = \sum u_k f \frac{a_k'}{a_k} - \sum u_k f \frac{b'}{b} = 0.$$

  Note that every linear factor of a_k with multiplicity m occurs in a_k' with multiplicity $m-1$; thus every summand of $\frac{a_k'}{a_k}$ is of the form $\frac{\lambda}{(x-\xi)}$, where $\lambda \in \mathbb{R}$ and $x - \xi$ is some linear factor of a_k; hence $f \cdot \frac{\lambda}{(x-\xi)}$ is of degree $\deg(f) - 1$, and therefore $f \cdot \ell(p(\theta))'$ is of degree $\deg(f) - 1$.

- We claim that the roots of $\ell(p(\theta))'$ are the same as the roots of $f \cdot \ell(p(\theta))'$. Assume we have shown this claim. By Proposition 2.2 the ML degree is one. So, $\ell(p(\theta))'$ has one root. Thus, $f \cdot \ell(p(\theta))'$ has one root, and therefore $f \cdot \ell(p(\theta))'$ is of degree one. This implies that $\deg(f) = 2$ with the previous step. Thus, there are exactly two distinct complex linear factors in $a_0, \ldots, a_n$, and b.

- It remains to show that the roots stay the same. Clearly, every root of $\ell(p(\theta))'$ is a root of $f \cdot \ell(p(\theta))'$. Conversely, we want to show that no new roots are introduced when multiplying by f, i.e. roots of f are not roots of $f \cdot \ell(p(\theta))'$. To do so, we rewrite

$$f \cdot \ell(p(\theta))' = \sum_{k=0}^{n} u_k f \frac{a_k'}{a_k} - \sum_{k=0}^{n} u_k f \frac{b'}{b} = \sum_{k=0}^{n+1} v_k f \frac{c_k'}{c_k}$$
$$v_k = u_k,\ c_k = a_k \quad \text{for } k = 0, \ldots, n,$$
$$v_{n+1} = -\sum_{k=0}^{n} u_k,\ c_{n+1} = b.$$

Let q be a complex linear factor of f. We define polynomials $r_0, \ldots, r_{n+1}$ and r such that $c_k = q^{l_k} r_k$, $f = qr$, and $r_0, \ldots, r_{n+1}, r$ do not have q as a factor. Then, we have for $k = 0, \ldots, n+1$ that

$$f \frac{c_k'}{c_k} = qr \cdot \frac{l_k q^{l_k - 1} q' r_k + q^{l_k} r_k'}{q^{l_k} r_k} = qr \frac{l_k q'}{q} + qr \frac{r_k'}{r_k} \equiv r l_k q' \pmod{q}.$$

Thus, we obtain

$$f \cdot \ell(p(\theta))' \equiv rq' \sum_{k=0}^{n+1} v_k l_k \equiv rq' \sum_{k=0}^{n} v_k (l_k - l_{n+1}) \pmod{q}.$$

Note that by definition of l_k, a value of $l_k = 0$ means that q is not a factor of c_k. By definition of f, at least one $l_k > 0$. On the other hand, not all l_k can be positive since $a_0, \ldots, a_n, b$ share no common factors. Hence, not all $l_k - l_{n+1} = 0$ vanish. Hence, for generic data u we assume $\sum_{k=0}^{n} v_k (l_k - l_{n+1}) \neq 0$. This with $q'r \not\equiv 0 \pmod{q}$ implies that q is not a complex linear factor of $f \cdot \ell(p(\theta))'$. We showed that the roots of f are not roots of $f \cdot \ell(p(\theta))'$.

□

Equipped with the lemma, we can now prove Proposition 2.1.

Proof. We want to show the following parametrization of $\mathcal{M}$:

$$p : [0,1] \to \Delta_n, \quad \theta \mapsto (w_k \theta^{i_k} (1-\theta)^{j_k})_{k=0}^{n}$$

First, we show that I is a single closed real interval and not a union of closed intervals. For the sake of contradiction assume that $I = \bigcup_k I_k$ is a union of closed disjoint intervals. By definition of $\mathcal{M}$ we know that $p(\partial I) \subset \partial\Delta_n$. Thus, there exist $\theta_1, \theta_2 \in \partial I_0$ and $\theta_3, \theta_4 \in \partial I_1$ with $p_i(\theta_1) = p_i(\theta_2) = 0$ and $p_j(\theta_3) = p_j(\theta_4) = 0$ for some $i, j = 0, \ldots, n$. Note that θ_1, θ_2 are roots of $\frac{a_i}{b}$ and θ_3, θ_4 are roots of $\frac{a_j}{b}$. By Lemma 2.3 exactly two distinct complex linear factors occur in $a_0, \ldots, a_n, b$. Hence, $\theta_3 = \theta_1$ or $\theta_3 = \theta_2$. Contradiction for I_0 and I_1 are disjoint.

The previous argument shows that $I = [\alpha, \beta]$ is a real single closed interval. Thus, the roots of $a_0, \ldots, a_n, b$ are real and take values in $\partial I = \{\alpha, \beta\}$. By a suitable parametrization, we can assume without loss of generality that $I = [0, 1]$. We can now write the polynomials $a_0, \ldots, a_n, b$ as

$$a_k(\theta) = w_k \theta^{i_k} (1-\theta)^{j_k}$$
$$b(\theta) = w \theta^{i} (1-\theta)^{j}$$

with $w_k, w \in \mathbb{R}_{>0}$, and $i_k, j_k, i, j \in \mathbb{Z}_{\geq 0}$ for all $k = 0, \ldots, n$. Since $a_0, \ldots, a_n, b$ share no common factors, there exists some $i_k = 0$ if $i > 0$; however this would contradict $0 < w_k \leq a_0(0) + \cdots + a_n(0) = b(0) = 0$. So $i = 0$. Similarly, $j = 0$. Finally, we divide p by w to obtain $b \equiv 1$. $\square$

Corollary 2.4. *Any one-dimensional discrete statistical models with rational MLE can be represented by* $(w_k, i_k, j_k)_{k=0}^n$ *for* $w_k \in \mathbb{R}_{>0}$ *and* $i_k, j_k \in \mathbb{Z}_{\geq 0}$.

From now on, we only consider one-dimensional discrete statistical models with rational MLE; we call them *models* for short.

Definition 2.5. The degree $\deg(\mathcal{M})$ of a model $\mathcal{M}$ represented by $(w_k, i_k, j_k)_{k=0}^n$ is defined as $\max\{i_k + j_k : k = 0, \ldots, n\}$.

Example 2.6. The sequence $((1,0,2),(2,1,1),(1,2,0))$ represents the binomial model with two trials. It has degree two. Its parametrization is given by $\theta \mapsto ((1-\theta)^2, 2\theta(1-\theta), \theta^2)$. Also see Figure 1.1 for a visualization of the binomial model within the probability simplex Δ_2.

Note that we view the sequences $((1,0,2),(2,1,1),(1,2,0))$ or $((2,1,1),(1,0,2),(1,2,0))$ as the same model as $((2,1,1),(1,2,0),(1,0,2))$ since the order of the coordinates does not matter.

Definition 2.7. Let $\mathcal{M}$ be a model represented by $(w_k, i_k, j_k)_{k=0}^n$. The set of exponent pairs $(i_k, j_k)_{k=0}^n$ is called the support of $\mathcal{M}$, denoted by $\mathrm{supp}(\mathcal{M})$.

This was our first step towards understanding the structure of models. The next step is to introduce the concept of reduced models.

2.2 Reduced Models

Models in this section refer to one-dimensional discrete statistical models with rational MLE.

Definition 2.8. We call a model represented by $(w_k, i_k, j_k)_{k=0}^n$ *reduced* if $(i_k, j_k) \neq \mathbf{0}$ for all $k = 0, \dots n$, and $(i_k, j_k) \neq (i_l, j_l)$ for all $k \neq l$.

Due to $(i_k, j_k) \neq (i_l, j_l)$, we can use functions to represent reduced models.

Remark 2.9. A reduced model $\mathcal{M}$ represented by $(w_k, i_k, j_k)_{k=0}^n$ can also be identified by a function $f : \mathbb{Z}^2 \to \mathbb{R}_{\geq 0}, (i, j) \mapsto w$, where $w = w_k$ if $(i_k, j_k) = (i, j)$ and $w = 0$ otherwise. The support of f is the set of all pairs (i, j) with $f(i, j) > 0$. It coincides with the support of $\mathcal{M}$.

Reduced models are our first building blocks for the classification of models. This statement is justified by the following two propositions. They show that every non-reduced model can be transformed into a reduced model by a sequence of linear embeddings.

Proposition 2.10. *Let $n \in \mathbb{N}_{>0}$. Let $\mathcal{M}$ be a model represented by $(w_k, i_k, j_k)_{k=0}^n$. If $(i_l, j_l) = \mathbf{0}$ for some index l, then there exist a model $\mathcal{M}'$, $\lambda \in [0, 1]$ and $k = 0, \dots, n$ such that*

$$\mathcal{M} = \Psi_{\lambda,k}(\mathcal{M}'),$$

where $\Psi_{\lambda,k} : \Delta_{n-1} \to \Delta_n$ *is defined as* $p_i \mapsto \begin{cases} \lambda p_i & \text{if } k \neq i, \\ 1 - \lambda & \text{if } k = i. \end{cases}$

Proof. Let $(i_l, j_l) = \mathbf{0}$ for some index l. If $w_l = 1$, then $w_m = 0$ for all $m \neq l$; this contradicts $w_m > 0$ by Proposition 2.1. Set $\lambda = 1 - w_l > 0$ and $k = l$. Define the model $\mathcal{M}'$ represented by

$$\left(\frac{w_h}{1 - w_l}, i_h, j_h \right)_{h=0, h \neq l}^n .$$

Then, $\mathcal{M} = \Psi_{\lambda,k}(\mathcal{M}')$. $\square$

Proposition 2.11. *Let $n \in \mathbb{N}_{>0}$. Let $\mathcal{M}$ be model represented by $(w_k, i_k, j_k)_{k=0}^n$. If $(i_m, j_m) = (i_l, j_l)$ for $m \neq l$, then there exist a model $\mathcal{M}'$, $\lambda \in [0, 1]$ and $k, h = 0, \dots, n$ such that*

$$\mathcal{M} = \Psi_{\lambda,k,h}(\mathcal{M}'),$$

where $\Psi_{\lambda,k,h} : \Delta_{n-1} \to \Delta_n$ *is defined as* $p_i \mapsto \begin{cases} p_i & \text{if } i \notin \{k, h\}, \\ \lambda p_k & \text{if } k = i, \\ (1 - \lambda) p_k & \text{if } h = i. \end{cases}$

Proof. Define $\lambda = \frac{w_m}{w_m+w_l}$, $k = m$, and $h = l$. Define the model $\mathcal{M}'$ represented by

$$(w_g + \delta_{gm} w_l, i_g, j_g)_{g=0,g\neq l}^{n} .$$

Then, $\mathcal{M} = \Psi_{\lambda,k}(\mathcal{M}')$. □

By repeatedly applying the two propositions, we can transform any model into a reduced model.

Corollary 2.12. *If Δ_n contains a model of degree d, then there also exists a reduced model of degree d in Δ_m for some $m \leq n$.*

2.3 Fundamental Models

As before, models refer to one-dimensional discrete statistical models with rational MLE. The main building blocks for the classification of models are *fundamental models*; we will see that reduced models come from fundamental models.

Definition 2.13. We call a model represented by $(w_k, i_k, j_k)_{k=0}^n$ *fundamental* if it is reduced and the equation $p_0 + \ldots p_n \equiv 1$ for given $(i_k, j_k)_{k=0}^n$ uniquely determines the weights $(w_k)_{k=0}^n$.

Example 2.14. The binomial model with two trials is fundamental. Given $(i_0, j_0) = (0, 2)$, $(i_1, j_1) = (1, 1)$, and $(i_2, j_2) = (2, 0)$, the equation $p_0+p_1+p_2 = w_0\theta^2+w_1\theta(1-\theta)+w_2(1-\theta)^2 \equiv 1$ uniquely determines the weights $w_0 = 1, w_1 = 2, w_2 = 1$. To see this observe that this equation is equivalent to $w_0\theta^2 + w_1\theta - w_1\theta^2 + w_2 - w_2 2\theta + w_2\theta^2 = 1$ which is equivalent to solving $w_2 - 1 + \theta(w_1 - 2w_2) + \theta^2(w_0 - w_1 + w_2) = 0$ for all $\theta \in \mathbb{R}$.

Example 2.15. Consider the probability simplex Δ_0. It only contains the model 1 which is fundamental.

Example 2.16. Now, consider the probability simplex Δ_1. It only contains the models $\theta \mapsto (\theta, 1-\theta)$ and $\theta \mapsto (1-\theta, \theta)$ which are equivalent. They are fundamental.

We will see that fundamental models like the ones above are building blocks for all reduced models by *composition*.

Definition 2.17. Let $\mathcal{M}$ and $\mathcal{M}'$ be reduced models which are represented by functions $f, g : \mathbb{Z}^2 \to \mathbb{R}_{\geq 0}$, see Remark 2.9. Let $\mu \in (0, 1)$. The *composite* $\mathcal{M} *_\mu \mathcal{M}'$ of $\mathcal{M}$ and $\mathcal{M}'$ is the reduced model represented by the function

$$(i, j) \mapsto \mu f(i, j) + (1 - \mu)g(i, j).$$

We are about to show that every reduced model is the composite of finitely many fundamental models.

Proposition 2.18. *Let $\mathcal{M}$ be a reduced model. Then $\mathcal{M}$ is the composite of finitely many fundamental models.*

Proof. For Δ_0 and Δ_1 we know that they only contain fundamental models, see Examples 2.15 and 2.16.

Assume we are given Δ_n with $n \geq 2$, and let $\mathcal{M}$ be a model that is not fundamental. We aim to show that $\mathcal{M}$ can be expressed as a composite of two models, $\mathcal{M}'$ and $\mathcal{M}''$, whose supports are proper subsets of $\operatorname{supp}(\mathcal{M})$. Assume this is indeed the case. Then, by applying the same argument to $\mathcal{M}'$ and $\mathcal{M}''$, we can recursively decompose each non-fundamental model into models with smaller supports. Since $\operatorname{supp}(\mathcal{M})$ is finite, this recursive decomposition must eventually terminate, yielding a decomposition of $\mathcal{M}$ into fundamental models. Thus, we have shown that any reduced model is the composite of a finite number of fundamental models.

Let us prove that $\mathcal{M}$ is the composite of two models whose supports are proper subsets of $\operatorname{supp}(\mathcal{M})$. Since $\mathcal{M}$ is not fundamental, the equation $p_0 + \cdots + p_n = 1$ has distinct solutions $\mathbf{w}, \mathbf{w}' \in \mathbb{R}^{n+1}_{>0}$. Define $\mathbf{v} \coloneqq \mathbf{w} - \mathbf{w}' \neq \mathbf{0}$. Then,

$$\sum_{k=0}^{n} v_k \theta^{i_k} (1-\theta)^{j_k} = 0 \quad \forall \theta \in (0,1).$$

Observe that there exist strictly positive and negative coefficients v_k. Define

$$\begin{aligned} \lambda &\coloneqq \min \left\{ \frac{w_k}{|v_k|} : k = 0, \ldots, n,\ v_k < 0 \right\}, \\ u_k &\coloneqq w_k + \lambda v_k \quad \text{for } k = 0, \ldots, n, \\ S_1 &\coloneqq \left\{ (i_k, j_k) : k = 0, \ldots, n,\ u_k \neq 0 \right\}. \end{aligned}$$

Note that $\lambda > 0$ since all the coefficients w_k are strictly positive by definition. Also observe that $u_k \geq 0$ if $v_k \geq 0$. Moreover, by definition $\frac{w_k}{|v_k|} \geq \lambda$ for all $k \geq 0$. Hence, if $v_k < 0$, we also have $\frac{u_k}{v_k} = \frac{w_k}{v_k} + \lambda \leq 0$. Multiplying by $v_k < 0$ we obtain $u_k \geq 0$. All in all, we have $u_k \geq 0$ for all $k = 0, \ldots, n$. Moreover, $u_k = 0$ if and only if $v_k < 0$ and $\lambda = \frac{w_k}{|v_k|}$. This shows that $S_1 \subsetneq \operatorname{supp}(\mathcal{M})$. Since $u_0 + \ldots u_n = 1$, we have found a reduced model $\mathcal{M}'$ represented by $(u_k, i_k, j_k)_{(i_k, j_k) \in S_1}$.

For the second model, we define

$$\begin{aligned} \mu &\coloneqq \min \left\{ \frac{w_k}{u_k} : k = 0, \ldots, n,\ u_k \neq 0 \right\}, \\ t_k &\coloneqq \frac{w_k - \mu u_k}{1 - \mu} \quad \text{for } k = 0, \ldots, n, \\ S_2 &\coloneqq \left\{ (i_k, j_k) : k = 0, \ldots, n,\ t_k \neq 0 \right\}. \end{aligned}$$

Similarly, $\mu > 0$. We have $\mu < 1$ because some v_k is positive implying $u_k > w_k$. By definition, we have $t_k \geq 0$, and $t_k = 0$ if and only if $u_k \neq 0$ and $\mu = \frac{w_k}{u_k}$. This shows that $S_2 \subsetneq \operatorname{supp}(\mathcal{M})$

and $S_1 \cup S_2 = \text{supp}(\mathcal{M})$. Since $t_0 + \cdots + t_n = 1$, we have found a reduced model $\mathcal{M}''$ represented by $(t_k, i_k, j_k)_{(i_k,j_k)\in S_2}$.

Finally, we see that $w_k = \mu u_k + (1-\mu)t_k$. This shows that $\mathcal{M} = \mathcal{M}' *_\mu \mathcal{M}''$. $\square$

By applying the previous proposition with Corollary 2.12, we obtain the following corollary.

Corollary 2.19. *If Δ_n contains a non-fundamental model of degree d, then there exists a fundamental model of degree d in Δ_m for some $m < n$.*

Example 2.20. For the two-dimensional probability simplex Δ_2, we can classify all models. Again, models refer to one-dimensional discrete statistical models with rational MLE. Note that the model $\mathcal{M}$ parametrized by $\theta \mapsto (\theta, 1-\theta)$ satisfies $\mathcal{M} *_\mu \mathcal{M} = \mathcal{M}$ for all μ. Since Δ_1 only contains the model $\theta \mapsto (\theta, 1-\theta)$, we can conclude that Δ_2 only contains fundamental models or models that are not reduced.

To find all the fundamental models in Δ_2, we need to check for all sets $S = \{(i_k, j_k)\}_{k=0}^2 \subset \mathbb{Z}^2_{>0}$ of size three if the equation $p_0 + p_1 + p_2 = \sum_{k=0}^2 w_k \theta^{i_k}(1-\theta)^{j_k} = 1$ has a unique solution (w_0, w_1, w_2). As we can see, a priori infinitely many sets S need to be checked. However, as we will see in the next section, only those sets S with $\max\{i + j : (i,j) \in S\} \le 2n - 1 = 3$ need to be considered. Clearly, this reduces the number of sets S to be checked to a finite number.

We compute that only the following supports uniquely determine the weights (w_0, w_1, w_2):

$$\{(0,3),(1,1),(3,0)\}, \{(0,2),(1,1),(2,0)\}, \{(0,1),(1,1),(2,0)\}, \{(0,2),(1,0),(1,1)\}.$$

They correspond to the fundamental models $((1-\theta)^3, 3\theta(1-\theta), \theta^3)$, $((1-\theta)^2, 2\theta(1-\theta), \theta^2)$, $(1-\theta, \theta(1-\theta), \theta^2)$, and $((1-\theta)^2, \theta, \theta(1-\theta))$. The last model is equivalent to the second last model by a parametrization $\theta \mapsto 1-\theta$ and permutation of the coordinates.

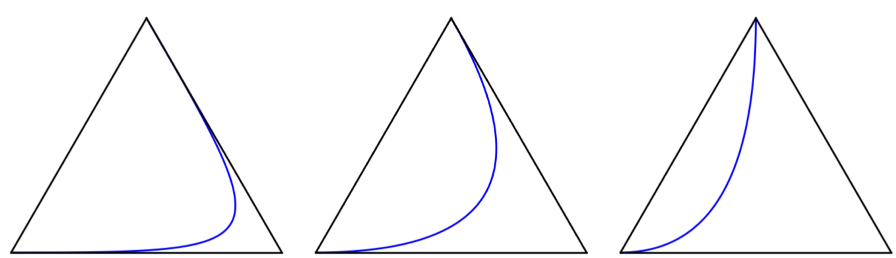

Figure 2.1: From left to right the illustration depicts the models parametrized $((1-\theta)^3, 3\theta(1-\theta), \theta^3)$, $((1-\theta)^2, 2\theta(1-\theta), \theta^2)$, $(1-\theta, \theta(1-\theta), \theta^2)$, and $((1-\theta)^2, \theta, \theta(1-\theta))$. The illustration is taken from [1].

We just computed all fundamental models of degree three or less in Δ_2. We will see shortly that these are all models in the probability simplex Δ_2. Of course, Δ_2 contains non-reduced models, too. These are models that come from linear embeddings $\Psi_{\lambda,k}$ and $\Psi_{\lambda,k,h}$, see Proposition 2.10 and Proposition 2.11. There are infinitely many of them, and for $\lambda = \frac{1}{3}$ we obtain the models $\theta \mapsto (\frac{2}{3}\theta, \frac{1}{3}, \frac{2}{3}(1-\theta))$ and $\theta \mapsto (1-\theta, \frac{1}{3}\theta, \frac{2}{3}\theta)$.

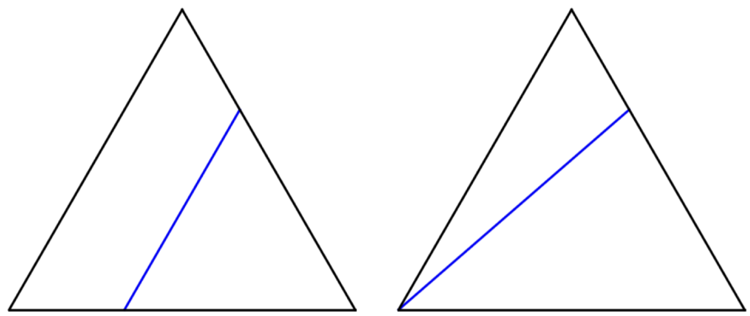

Figure 2.2: This illustration depicts two non-reduced models in Δ_2 for $\lambda = \frac{1}{3}$. They are parametrized by $\theta \mapsto (\frac{2}{3}\theta, \frac{1}{3}, \frac{2}{3}(1-\theta))$ and $\theta \mapsto (1-\theta, \frac{1}{3}\theta, \frac{2}{3}\theta)$. All other non-reduced models can be obtained by varying λ. The illustration is taken from [1].

Let us summarize the results of this section. It is the first part of our classification theorem.

Theorem 2.21. *Every one-dimensional discrete statistical model with rational MLE in Δ_n is the image of a reduced model in Δ_m under a linear embedding $\Delta_m \to \Delta_n$ for some $m \leq n$.*

Moreover, every reduced model $\mathcal{M} \subset \Delta$ can be written as a composite of finitely many fundamental models

$$\mathcal{M} = \mathcal{M}_1 *_{\mu_1} (\cdots *_{\mu_{m-2}} (\mathcal{M}_{m-1} *_{\mu_{m-1}} \mathcal{M}_m))$$

for some $m < n$ and $\mu_1, \ldots, \mu_m \in (0,1)$.

Proof. See Proposition 2.18, Proposition 2.10, and Proposition 2.11. □

2.4 On the Finiteness of Fundamental Models

We have established the first part of our classification theorem, namely that fundamental models are building blocks for all models. The second part is showing that there are only finitely many fundamental models in Δ_n given $n \in \mathbb{N}$. Artuhr Bik and Orlando Marigliano proved that there are only finitely many fundamental models in Δ_n for $n \leq 4$ [1]. We will later make significant progress towards proving the case $n = 5$. For $n \geq 6$ no attempt has been made yet to the best of our knowledge.

Arthur Bik and Orlando Marigliano first proved the following proposition.

Proposition 2.22. *Let $\mathcal{M}$ be a one dimensional discrete statistical model with rational MLE in Δ_n. For $n \leq 4$ we have* $\deg(\mathcal{M}) \leq 2n - 1$.

Given Proposition 2.22 it is easy to show the second part of our classification.

Theorem 2.23. *There are only finitely many fundamental models in Δ_n for all $n \leq 4$.*

Proof. By Proposition 2.22 we know that the degree of a fundamental model is at most $2n-1$. Since the number of supports of a fundamental model of degree $2n-1$ is finite, there are only finitely many fundamental models in Δ_n for all $n \leq 4$. $\square$

We will now spend the rest of this thesis on proving Proposition 2.22. The idea is to use the building blocks of fundamental models that we have established so far. Namely, it suffices to show the proposition for fundamental models.

Proposition 2.24. *Let $N \in \mathbb{N}$. If* $\deg(\mathcal{M}) \leq 2n - 1$ *for all fundamental models in Δ_n and $n \leq N$, then* $\deg(\mathcal{M}') \leq 2n - 1$ *holds for all models in Δ_n.*

Proof. Let $N \in \mathbb{N}$ and $n \leq N$. Assume there is some non-fundamental model $\mathcal{M}'$ in Δ_n of degree greater than $2n - 1$. By Corollary 2.19 there exists a fundamental model $\mathcal{M}$ in Δ_m for some $m < n$ of degree greater than $2m - 1$. This contradicts the assumption that the degree of fundamental models is at most $2n' - 1$ for all $n' \leq N$. $\square$

Counting all *fundamental* models in Δ_n for $n \leq 4$ is our guiding objective. As a first step, we introduce a combinatorial game that aids in counting fundamental models. We know that every reduced model can be represented by the sequence of triples $(w_k, i_k, j_k)_{k=0}^n$, where $w_k \in \mathbb{R}{>}\ 0$ and $i_k, j_k \in \mathbb{Z}_{\geq 0}$. The model can be visualized in a directed graph with vertices in $\mathbb{Z}^2$, where we can place values w_k on vertices (i_k, j_k). Each vertex (i, j) is connected by directed edges to $(i + 1, j)$ and $(i, j + 1)$.

Surprisingly, we can derive a combinatorial game from this graph by defining a specific set of rules. This game, called the *chipsplitting game*, will be rigorously introduced in the next chapter. After that, we will explore the game's properties and show how it can be used to count fundamental models in Δ_n for $n \leq 4$.

$$
\begin{tikzcd}
. \arrow[r] & . \arrow[r] & . \arrow[r] & . \\
1 \arrow[r] \arrow[u] & . \arrow[r] \arrow[u] & . \arrow[r] \arrow[u] & . \arrow[u] \\
. \arrow[r] \arrow[u] & 2 \arrow[r] \arrow[u] & . \arrow[r] \arrow[u] & . \arrow[u] \\
. \arrow[r] \arrow[u] & . \arrow[r] \arrow[u] & 1 \arrow[r] \arrow[u] & . \arrow[u]
\end{tikzcd}
$$

Figure 2.3: The binomial model with two trials visualized in a directed graph with vertices in $\{0, 1, 2, 3\}^2$.

Chapter 3

Chipsplitting Games

The notion of a chipsplitting game was introduced by [1] as a combinatorial approach to classifying one-dimensional discrete statistical models with rational maximum likelihood estimator. It was inspired by *chipfiring games* and for a subset of chipfiring games, the chipsplitting game is equivalent to the chipfiring game. We refer to [4] for a comprehensive introduction to chipfiring games.

Let us define the notion of a chipsplitting game.

Definition 3.1. Let (V, E) be a directed graph without loops.

1. A *chip configuration* is a vector $\mathbf{w} = (w_v)_{v\in V} \in \mathbb{Z}^V$ such that there are only finitely many nonzero components w_k.

2. The *initial configuration* is the chip configuration $\mathbf{0} \in \mathbb{Z}^V$.

3. A *splitting move* at $u \in V$ maps a chip configuration $\mathbf{w}$ to some chip configuration $\mathbf{w}'$ defined by

$$w'_v \coloneqq \begin{cases} w_v - 1 & \text{if } v = u, \\ w_v + 1 & \text{if } (u, v) \in E \\ w_v & \text{otherwise.} \end{cases}$$

4. An *unsplitting move* at $u \in V$ maps $\mathbf{w}'$ back to $\mathbf{w}$.

5. A *chipsplitting game* is a finite sequence of splitting and unsplitting moves.

6. An *outcome of a chipsplitting game* is the chip configuration obtained from applying the sequence of splitting and unsplitting moves defined by the game at the initial configuration.

7. Any outcome of a chipsplitting game is called an *outcome*.

Proposition 3.2. *The order of the moves in a chipsplitting game does not affect the outcome.*

Proof. This follows from commutativity of addition. □

Note that all moves are reversible. Thus, we obtain the following corollary with Proposition 3.2.

Corollary 3.3. *Let* $\mathbf{w}$ *be an outcome. Then, there exists a chipsplitting game whose outcome is* $\mathbf{w}$ *and where at no point both a splitting and an unsplitting move are applied at the same vertex.*

Games that satisfy the condition in the corollary are called *reduced*. We will only consider reduced games in this thesis for simplicity. The map

$$\begin{aligned}\{\text{reduced games on } (V,E)\} / \sim \quad &\to \quad \{g : V' \to \mathbb{Z} : \#\{p \in V' : g(p) \neq 0\} < \infty\}\\ f &\mapsto (p \mapsto \text{number of moves at } p \text{ in game } f)\end{aligned}$$

is a bijection, where $V' \subset V$ is the subset of vertices with at least one outgoing edge. The equivalence relation $\sim$ is defined by $f \sim g$ if f and g are the same up to reordering. Unsplitting moves are counted negatively by $p \mapsto$ number of moves at p in game f. Using the map above we identify a chipsplitting game with its corresponding function $V' \to \mathbb{Z}$. For every outcome $\mathbf{w} = (w_v)_{v \in V}$ we have

$$w_v = -f(v) + \sum_{u \in V', (u,v) \in E} f(u),$$

where we define $f(v) = 0$ for $v \notin V$.

Now, we define the directed graphs that we will consider in this thesis. For $d \in \mathbb{N} \cup \{\infty\}$ we write

$$\begin{aligned}V_d &\coloneqq \left\{(i,j) \in \mathbb{Z}^2_{\geq 0} \mid i + j \leq d\right\},\\ E_d &\coloneqq \{(v, v+e) \mid v \in V_{d-1}, e \in \{(1,0),(0,1)\}\}.\end{aligned}$$

Definition 3.4. The degree $\deg(\mathbf{v})$ of a vertex $\mathbf{v} = (i,j)$ is defined as $i + j$.

Example 3.5. A chip configuration $\mathbf{w} = (w_{i,j})_{(i,j) \in V_d} \in \mathbb{Z}^{V_d}$ can be illustrated as a triangle of numbers where $w_{i,j}$ is placed at the position (i,j) in the triangle. For example, $w_{2,4} = 4$ means that the value 4 is placed in the second column and fourth row of the triangle. The following is an example of a sequence of chip configurations for $d = 3$:

```
.           .           .           1           1           1
. .         . .         1 .         . 1         . 1         . .
. . .       1 . .       . 2 .       . 2 .       . 2 1       . 3 .
0 . . .     -1 1 . .    -1 . 1 .    -1 . 1 .    -1 . . 1    -1 . . 1
```

When $w_{i,j} = 0$, we omit the value in the triangle and write a dot instead. The sequence above starts with the initial configuration and then applies a splitting move at the vertex $(0,0), (1,0), (0,1), (0,2)$ and $(2,0)$. Finally, we apply an unsplitting move at the vertex $(1,1)$ to obtain the final configuration.

Figure 2.4 is represented as the third configuration of the triangle above.

Definition 3.6. Let $\mathbf{w} = (w_{i,j})_{(i,j)\in V_d}$ be a chip configuration.

1. The *positive support* of $\mathbf{w}$ is defined as $\operatorname{supp}^+(\mathbf{w}) \coloneqq \{(i,j) \in V_d \mid w_{i,j} > 0\}$.
2. The *negative support* of $\mathbf{w}$ is defined as $\operatorname{supp}^-(\mathbf{w}) \coloneqq \{(i,j) \in V_d \mid w_{i,j} < 0\}$.
3. The *support* of $\mathbf{w}$ is defined as the union of the positive and negative support.
4. The *degree* of $\mathbf{w}$ is defined as $\deg(\mathbf{w}) \coloneqq \max\{i+j \mid (i,j) \in \operatorname{supp}(\mathbf{w})\}$.
5. We say $\mathbf{w}$ is *valid* if its negative support is empty or only contains $(0,0)$.

Bibliography

[1] Arthur Bik and Orlando Marigliano. "Classifying one-dimensional discrete models with maximum likelihood degree one". In: *arXiv preprint arXiv:2205.09547* (2022).

[2] Eliana Duarte, Orlando Marigliano, and Bernd Sturmfels. "Discrete statistical models with rational maximum likelihood estimator". In: (2021).

[3] June Huh. "Varieties with maximum likelihood degree one". In: *arXiv preprint arXiv:1301.2732* (2013).

[4] Caroline J Klivans. *The mathematics of chip-firing*. Chapman and Hall/CRC, 2018.