Maximum Entropy in the framework of Algebraic Statistics: A First Step

Ambedkar Dukkipati
ambedkar.dukkipati@gmail.com

Abstract. Algebraic statistics is a recently evolving field, where one would treat statistical models as algebraic objects and thereby use tools from computational commutative algebra and algebraic geometry in the analysis and computation of statistical models. In this approach, calculation of parameters of statistical models amounts to solving set of polynomial equations in several variables, for which one can use celebrated Gröbner basis theory. Owing to the important role of information theory in statistics, this paper as a first step, explores the possibility of describing maximum and minimum entropy (ME) models in the framework of algebraic statistics. We show that ME-models are toric models (a class of algebraic statistical models) when the constraint functions (that provide the information about the underlying random variable) are integer valued functions, and maximum entropy distributions can be calculated by solving set of (Laurent) polynomial equations when expected values of constraint functions are supplied as sample means.

1. Introduction

Algebra has always played an important role in statistics, a classical example being linear algebra. There are also many other instances of applying algebraic tools in statistics (e.g [1, 2]). But, treating statistical models as algebraic objects, and thereby using tools of computational commutative algebra and algebraic geometry in the analysis of statistical models is very recent and has led to the still evolving field of *algebraic statistics*.

The use of computational algebra and algebraic geometry in statistics was initiated in the work of Diaconis and Sturmfels [3] on exact hypothesis tests of conditional independence in contingency tables, and in the work of Pistone et al. [4] in experimental design. The term 'Algebraic Statistics' was first coined in the monograph by Pistone et al. [4] and appeared recently in the title of the book by Pachter and Sturmfels [5].

To extract the underlying algebraic structures in discrete statistical models, algebraic statistics treat statistical models as affine varieties. (An affine variety is the set of all solutions to family of polynomial equations.) Parametric statistical models are described in terms of a polynomial (or rational) mapping from a set of parameters to distributions. One can show that many statistical models, for example independence models, Bernouli random variable etc. (see [5] for more examples), can be given this algebraic formulation, and these are referred to as algebraic statistical models.

Statistical model which are studied mainly in algebraic statistics are exponential models, which have a long history [6]. These models have been studied in algebraic statistics under the name 'toric' models. Toric models are algebraic statistical models and the term 'toric' comes from important algebraic objects known as 'toric ideals' in computational algebra. In this view of very established role of information theory in statistics [7, 8] this paper attempts to describe maximum entropy models in algebraic statistical framework.

In particular we show that maximum entropy models (also minimum relative-entropy models) are indeed toric models, when the functions that provide the information about the underlying random variable in the form of expected valued are integer valued. We also show that when the information is available in the form of sample means, by modifying maximum entropy prescriptions calculating model parameters amounts to solving set of polynomial equations.

A note on the results presented in this paper: we will not present the details on Gröbner basis theory and related concepts to solve the polynomial equations due to space constraint; we refer reader to text books on computational algebra and Grobner basis theory [9, 10]. For computations on toric ideals which play important role in analysis of exponential models maximum and minimum entropy models (as we are going to see in this paper), one can refer to [11, 12, 13].

We organize our paper as follows. In § 2 we give basic notions of algebra and introduce notation. We also introduce basic concepts of algebraic statistics. § 3 describes maximum entropy (ME) prescriptions in algebraic statistical framework by introducing important algebraic objects called toric ideals. In § 4 we show how one can transform the problem of calculating ME distributions to solving set of polynomial equations.

2. Algebraic Statistical Models

2.1. Basic notions of Algebra

Through out this paper k represents a field. A monomial in n indeterminates $x_1, \ldots, x_n$ is a power product of the form $x_1^{\alpha_1} \ldots x_n^{\alpha_n}$, where all the exponents are nonnegative integers, i.e. $\alpha_i \in \mathbb{Z}_{\geq 0}$, $i = 1, \ldots n$. One can simplify the notation for monomial as follows: let $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$. Then by using multi-index notation we set

$$x^\alpha = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$$

with the understanding that $x = (x_1, \ldots, x_n)$. When $\alpha = (0, \ldots, 0) \in \mathbb{Z}^n_{\geq 0}$, note that $x^\alpha = 1$. Once the order of the indeterminates are fixed, monomial $x_1^{\alpha_1} \ldots x_n^{\alpha_n} = x^\alpha$ is identified by $(\alpha_1, \ldots, \alpha_n)$. Hence, set of all monomials in indeterminates $x_1, \ldots, x_n$ can be represented by $\mathbb{Z}^n_{\geq 0}$. Theory of monomials is central to the celebrated Gröbner basis theory in computational algebra which provides tools for solving set of polynomial equations and related problems in algebraic geometry [14]. Monomial theory itself plays important role in algebraic statistics in the representation of exponential models where probabilities are expressed in terms of power products [15].

A polynomial f in $x_1, \ldots, x_n$ with coefficients in k is a finite linear combination of monomials and can be written in the form of

$$f = \sum_{\alpha \in \Lambda_f} a_\alpha x^\alpha \ ,$$

where $\Lambda_f \subset \mathbb{Z}^n_{\geq 0}$ is a finite set and $a_\alpha \in k$. The collection of all polynomials in the indeterminates $x_1, \ldots, x_n$ is the set $k[x_1, \ldots, x_n]$ and it has structure not only of a vector space but also of a ring, and the ring structure plays main role in computational algebra and algebraic geometry. The ring $k[x_1, \ldots, x_n]$ is called the ring of polynomials in n indeterminates.

A subset $\mathfrak{a} \subset k[x_1, \ldots, x_n]$ is said to be ideal if it satisfies: (i) $0 \in \mathfrak{a}$ (ii) $f, g \in \mathfrak{a}$, then $f + g \in \mathfrak{a}$ (iii) $f \in \mathfrak{a}$ and $h \in k[x_1, \ldots, x_n]$ and then $hf \in \mathfrak{a}$. A set $V \subset k^n$ is said to be affine variety if there exists $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$ such that

$$V = \{(c_1, \ldots c_n) \in k^n : f_i(c_1, \ldots c_n) = 0, 1 \leq i \leq s\} \ .$$

We use the notation $\mathcal{V}(f_1, \ldots, f_s) = V$.

2.2. Algebraic Statistical Model

At the very core of the field of algebraic statistics lies the notion of an ‘algebraic statistical model’. While this notion has the potential of serving as a unifying theme for algebraic statistics, there is no unified definition of an algebraic statistical model [16]. Here, we adopt the appropriate definition of statistical model from [5, 16]. For a recent elaborate discussion on formal definition of algebraic statistical models one can refer to [16].

Let X be a discrete random variable taking finitely many values from the set $[m] = \{1, 2, \dots m\}$. A probability distribution p of X is naturally represented as a vector $p = (p_1, \dots, p_m) \in \mathbb{R}^m$ if we fix the order on $[m]$. Then set of all probability mass functions (pmfs) of X is called probability simplex

$$\Delta_{m-1} = \{p = (p_1, \dots, p_m) \in \mathbb{R}^m_{\geq 0} : \sum_{i=1}^{m} p_i = 1\} \ . \tag{1}$$

The index $m-1$ indicates the dimension of the simplex Δ_{m-1}. A statistical model $\mathcal{M}$ is any subset of Δ_{m-1}. A statistical model $\mathcal{M} \subseteq \Delta_{m-1}$ is said to be algebraic if $\exists f_1, \dots, f_s \in k[p_1, \dots, p_m]$ such that

$$\mathcal{M} = \mathcal{V}(f_1, \dots, f_s) \cap \Delta_{m-1} \ .$$

Now we move on to parametric statistical models and their algebraic formulations.

Let $\Theta \subseteq \mathbb{R}^d$ be a parametric space and $\kappa : \Theta \to \Delta_{m-1}$ be a map. The image $\kappa(\Theta)$ is called parametric statistical model. Given a statistical model $\mathcal{M} \subseteq \Delta_{m-1}$, by parameterization of $\mathcal{M}$ we mean, identifying a set $\Theta \subseteq \mathbb{R}^d$ and a function $\kappa : \Theta \to \Delta_{m-1}$ such that $\mathcal{M} = \kappa(\Theta)$. To describe more general statistical models in algebraic framework we need following notion of *semi-algebraic set*.

Definition 2.1. *A set $\Theta \subseteq \mathbb{R}^d$ is called semi-algebraic set, if there are two finite collection of polynomials $F \subset k[x_1, \dots, x_d]$ and $G \subset k[x_1, \dots, x_d]$ such that*

$$\Theta = \{\theta \in \mathbb{R}^d : f(\theta) = 0, \forall f \in F \text{ and } g(\theta) \geq 0, g \in G\} \ .$$

Now we have following definition of parametric algebraic statistical model.

Definition 2.2. *Let Δ_{m-1} be a probability simplex and $\Theta \subset \mathbb{R}^d$ be a semi-algebraic set. Let $\kappa : \mathbb{R}^d \to \mathbb{R}^m$ be a rational function (a rational function is a quotient of two polynomials) such that $\kappa(\Theta) \subseteq \Delta_{m-1}$. Then the image $\mathcal{M} = \kappa(\Theta)$ is a parametric algebraic statistical model.*

Conversely, a parametric statistical model $\mathcal{M} = \kappa(\Theta) \subseteq \Delta_{n-1}$ is said to be algebraic if Θ is semi-algebraic set and κ is a rational function. From now on we refer to 'parametric algebraic statistical models' as 'algebraic statistical models'.

In this paper we consider following special case of algebraic statistical models (cf. [5, pp 7]). Consider a map

$$\begin{aligned} \kappa : \Theta(\subseteq \mathbb{R}^d) \quad &\to \mathbb{R}^m \\ \kappa : \theta = (\theta_1, \dots, \theta_d) &\mapsto (\kappa_1(\theta), \dots, \kappa_m(\theta)) \end{aligned} \tag{2}$$

where $\kappa_i \in k[\theta_1, \dots, \theta_d]$ and $\forall \theta \in \Theta$ we have $\kappa_i(\theta) \geq 0$, $i = 1, \dots, m$. Also we post the condition that $\sum_{i=1}^{m} \kappa_i(\theta) = 1$. Under these conditions $\kappa(\Theta)$ is indeed an algebraic statistical model (Definition 2.2) since $\kappa(\Theta) \subset \Delta_{m-1}$, κ is a polynomial function and Θ is a semi-algebraic set ($H = \{\sum_{i=1}^{m} f_i - 1\}$ and $G = \{f_i : i = 1, \dots, m\}$ in the Definition 2.1).

Some statistical models are naturally given by a polynomial map κ (2) for which the condition $\sum_{i=1}^{m}\kappa_i(\theta) = 1$ does not hold. If this is the case one can consider following algebraic statistical model:

$$\kappa : \theta = (\theta_1, \ldots, \theta_d) \mapsto \frac{1}{\sum_{i=1}^{m}\kappa_i(\theta)}\left(\kappa_1(\theta), \ldots, \kappa_m(\theta)\right) \quad , \tag{3}$$

assuming that remaining conditions that have been specified for the model (2) are valid here too. The only difference is that instead of κ being a polynomial map, we have it as a rational map.

3. ME in algebraic statistical setup

3.1. Toric Models

In the algebraic description of exponential models monomials and binomials play a fundamental role. The study of relations of power products lead to the theory of toric ideals in the commutative algebra [17]. Here we describe basic notion of toric ideal that are relevant to representation and computation of discrete exponential models; for more details on theory and computation of toric ideals one can refer to [17, 18, 12].

Before we give the definition of toric ideal, we describe the notion of Laurent polynomial. If we allow negative exponents in a polynomial i.e., polynomial of the form $f = \sum_{\alpha\in\Lambda_f} a_\alpha x^\alpha$ where $\alpha \in \mathbb{Z}^n$ it is known as Laurent polynomial ($\Lambda_f \subset \mathbb{Z}^n_{\geq 0}$ is finite). Set of all Laurent polynomials in the indeterminates $x_1, \ldots, x_n$ is denoted by $k[x_1^{\pm}, \ldots, x_n^{\pm}]$ and it has a structure of a ring.

Now we define the toric ideal.

Definition 3.1. *Let $A = [a_{ij}] \in \mathbb{Z}^{d\times n}$ be a matrix with rank d. Consider the ring homeomorphism*

$$\begin{aligned} \hat{\pi} &: k[x_1, \ldots, x_n] \to k[\theta_1^{\pm}, \ldots, \theta_d^{\pm}] \\ \hat{\pi} &: x_j \qquad\qquad\quad \mapsto \theta_1^{a_{1j}} \ldots \theta_1^{a_{dj}} \end{aligned} \tag{4}$$

The toric ideal $\mathfrak{a}_A$ of A is defined as the kernel of the map $\hat{\pi}$, i.e., $\mathfrak{a}_A = ker\hat{\pi}$.

The mapping $\hat{\pi}$ can be viewed as "parameterization" and which can be explained by the following description of $\hat{\pi}$. Consider a map

$$\begin{aligned} \pi &: \mathbb{Z}^n_{\geq 0} \qquad\qquad\quad \to \mathbb{Z}^d \\ \pi &: u = (u_1, \ldots, u_n) \mapsto Au. \end{aligned} \tag{5}$$

The map π lifts to the ring homomorphism $\hat{\pi}$ in the sense of action of $\hat{\pi}$ on $x^u = x_1^{u_1} \ldots x_n^{u_n} \in k[x_1, \ldots, x_n]$. That is

$$\begin{aligned} \hat{\pi}(x^u) &= \hat{\pi}(x_1^{u_1}, \ldots, x_n^{u_n}) \\ &= \left(\prod_{i=1}^{d}\theta_i^{a_{i1}}\right)^{u_1} \ldots \left(\prod_{i=1}^{d}\theta_i^{a_{in}}\right)^{u_n} \\ &= \prod_{i=1}^{d}\theta_i^{\sum_{j=1}^{n} a_{ij}u_j} = \theta^{Au} \quad . \end{aligned} \tag{6}$$

Toric ideal theory plays an important role in applications of computational algebraic geometry like integer programming etc [17]. Note that in the algebraic descriptions of exponential models and their maximum likelihood estimates only non-negative cases of toric ideals (and hence toric models) is considered i.e., the matrix $A = [a_{ij}]$ in Definition 3.1 is assumed to be nonnegative and the map (4) is specified as $\hat{\pi} : k[x_1, \ldots, x_n] \to k[\theta_1, \ldots, \theta_d]$ (see [5]). As described later in this paper, in the algebraic descriptions of maximum entropy models one has to deal with the Laurent polynomials and hence one has to include the negative case in the definitions of toric ideals and toric models. This poses no problem because toric ideal theory in commutative algebra naturally includes the negative case (as in Definition 3.1) and Gröbner bases theory can be extended to Laurent polynomial ring [19]. The concept of toric ideals let to the description of exponential models under the name toric models in algebraic statistics.

Definition 3.2. *Let $A \in \mathbb{Z}_{\geq 0}^{d \times m}$ be a matrix such that the vector $(1, \ldots, 1) \in \mathbb{Z}_{\geq 0}^m$ is in the row span of A. Let $h \in \mathbb{R}_{>0}^m$ be a vector of positive real numbers. Let $\Theta = \mathbb{R}_{>0}^m$ and let $\kappa^{A,h}$ be the rational parameterization*

$$\begin{aligned} &\kappa^{A,h} : \Theta \to \mathbb{R}^m \\ &\kappa_j^{A,h} : \theta \ \mapsto Z(\theta)^{-1} h_j \prod_{i=1}^{d} \theta_i^{a_{ij}} \ , \end{aligned} \tag{7}$$

where $\theta = (\theta_1, \ldots, \theta_d)$ and $Z(\theta)$ is the appropriate normalizing constant. The toric model is the parametric algebraic statistical model

$$\mathcal{M}_{A,h} \triangleq \kappa^{A,h}(\Theta) \ . \tag{8}$$

Independence models, exponential models, Markov chains and Hidden Markov chains can be given an algebraic statistical description by means of toric models [5]. We keep positivity of A in the Definition 3.2 as a matter of convention.

3.2. ME in terms of Toric Models

Let X be a random variable taking values from the set $[m] = \{1, 2, \ldots m\}$. The only information we know about the pmf $p = (p_1, \ldots, p_m)$ of X is in the form of expected values of the functions $t_i : [m] \to \mathbb{R}$, $i = 1, \ldots, d$ (we refer these functions as constraint functions). We threfore have

$$\sum_{j=1}^{m} t_i(j) p_j = T_i \ \ , i = 1, \ldots d \ , \tag{9}$$

where T_i, $i = 1, \ldots, d$, are assumed to be known. In an information theoretic approach to statistics one would choose the pmf $p \in \Delta_{m-1}$ that maximize the Shannon entropy functional

$$S(p) = -\sum_{j=1}^{m} p_j \ln p_j \tag{10}$$

with respect to the constraints (9). This is also known as Jayens maximum entropy model.

The corresponding Lagrangian can be written as

$$\Xi(p,\xi) \equiv S(p) - \xi_0 \left(\sum_{j=1}^{m} p_j - 1 \right) - \sum_{i=1}^{d} \xi_d \left(\sum_{j=1}^{m} t_i(j) p_j - T_i \right)$$

Holding $\xi = (\xi_1, \ldots, \xi_d)$ fixed, the unconstrained maximum of Lagrangian $\Xi(p,\xi)$ over all $p \in \Delta_{m-1}$ is given by an exponential family [20]

$$p_j(\xi) = Z(\xi)^{-1} \exp \left(- \sum_{i=1}^{d} \xi_i t_i(j) \right) \ , j = 1, \ldots, m, \tag{11}$$

where $Z(\xi)$ is normalizing constant given by

$$Z(\xi) = \sum_{j=1}^{m} \exp \left(- \sum_{i=1}^{d} \xi_i t_i(j) \right) \ . \tag{12}$$

For various values of $\xi \in \mathbb{R}^d$, the family (11) is known as *maximum entropy model*.

Now assume that constraint functions t_i, $i = 1, \ldots, d$ are integer valued functions i.e., $t_i : [m] \to \mathbb{Z}_{\geq 0}$, $i = 1, \ldots$ (we assume positivity as a matter of convention). With this assumption it is easy to show that maximum entropy model is indeed a toric model.

Proposition 3.3. *The maximum entropy model (11) is a toric model provided that the constraint functions t_i, $i = 1, \ldots, d$ are integer valued.*

Proof. Set $\ln \theta_i = \xi_i$, $i = 1, \ldots, d$. Now (11) gives us

$$\begin{aligned} p_j &= Z(\theta)^{-1} \exp \left(- \sum_{i=1}^{d} t_i(j) \ln \theta_i \right) \\ &= Z(\theta)^{-1} \prod_{i=1}^{d} \theta_i^{t_i(j)} \ . \end{aligned} \tag{13}$$

By defining matrix $A = [t_i(j)] \in \mathbb{Z}^{d \times n}$ and setting $h = (\frac{1}{m}, \ldots, \frac{1}{m})$ we have rational parameterization as in (7). $\square$

Note that we allowed only integer valued functions in the above ME-model, which is necessary for algebraic descriptions of the same. Here we also mention that in the above proof by assuming $h \in \Delta_{m-1}$, we can imply that minimum I-divergence models are defined as

$$p_j = \widehat{Z}(\xi)^{-1} h_j \exp \left(- \sum_{i=1}^{d} \xi_i T_i(j) \right) \ , \ j = 1, \ldots, m, \tag{14}$$

(with appropriate normalizing constant $\widehat{Z}(\xi)$) which is indeed a toric model.

Once the specification of statistical model is done, the task is to calculate the model parameters with available information. In this case the available information is in the form of expected valued of functions t_i, $i = 1, \ldots d$. To calculate a ME-model, the Lagrange parameters ξ_i, $i = 1, \ldots, d$ are determined using the constrains (9).

4. Calculation of ME distributions via solving Polynomial equations

4.1. Calculation of model parameters

One can show that the Lagrange parameters in ME-model (11) can be estimated by solving following set of partial differential equations [21]

$$\frac{\partial}{\partial \xi_i} \ln Z(\xi) = T_i \ \ , i = 1, \ldots, d, \tag{15}$$

which has no explicit analytical solution. In literature there are several methods of estimating ME-models. One of the important method is Darroch and Ratcliff's genralized iterative scaling algorithm [22], which has a geometric interpretation [23].

Here we follow the method of *dual* optimization problem. By using Kuhn-Tucker theorem we calculate Lagrange parameters ξ_i, $i = 1, \ldots, d$ in (11) by optimizing dual of $\Xi(p, \xi)$. That is the task is to find ξ which maximizes

$$\Psi(\xi) \equiv \Xi(p^{(\xi)}, \xi) \ \ . \tag{16}$$

Note that $\Psi(\xi)$ is nothing but entropy of ME-distribution (11). We have

$$\Psi(\xi) = \ln Z + \sum_{i=1}^{d} \xi_i T_i \ \ . \tag{17}$$

This can be written as

$$\begin{aligned} \Psi(\xi) =& \ \ln \sum_{j=1}^{m} \exp\left(- \sum_{j=1}^{d} \xi_i t_i(j) \right) + \sum_{i=1}^{d} \xi_i T_i \\ =& \ \ln \sum_{j=1}^{m} \exp\left(\xi_i (T_i - t_i(j))\right) \ \ . \end{aligned} \tag{18}$$

Now maximizing $\Psi(\xi)$ is equivalent to maximizing

$$\Psi'(\xi) = \sum_{j=1}^{m} \exp\left(\xi_i (T_i - t_i(j))\right) \ \ . \tag{19}$$

By introducing $\xi_i = \ln \theta_i$, $i = 1, \ldots, d$ we have

$$\Psi'(\theta) = \sum_{j=1}^{m} \prod_{i=1}^{d} \theta_i^{T_i - t_i(j)} \ \ . \tag{20}$$

The solution is given by solving the following set of equations

$$\frac{\partial \Psi'}{\partial \theta_j} = 0 \ \ , j = 1, \ldots d. \tag{21}$$

Unfortunately $\frac{\partial \Psi}{\partial \theta_j} \in k[\theta_1^{\pm}, \ldots, \theta_d^{\pm}]$ only if $T_i \in \mathbb{Z}$. Next, we consider the case where the expected values are available as sample means.

4.2. The case of sample means

In most practical problems the information in the form of expected values is available via sample or empirical means. That is, given a sequence of observations $O_1, \ldots, O_N$ the sample means $\widetilde{T}_i$, $i = 1, \ldots, d$, with respect to the functions t_i, $i = 1, \ldots, d$ are given by $\widetilde{T}_i = \frac{1}{N}\sum_{l=1}^{N} t_i(O_l)$, $i = 1, \ldots, d$,

$$\widetilde{T}_i = \frac{1}{N}\sum_{l=1}^{N} t_i(O_l), i = 1, \ldots, d, \tag{22}$$

and the hypothesis is $T_i \approx \widetilde{T}_i$. That is

$$\sum_{j=1}^{m} p_j t_i(j) \approx \frac{1}{N}\sum_{l=1}^{N} t_i(O_l) \quad , i = 1, \ldots, d. \tag{23}$$

Now we show that, by choosing alternate Lagrangian we can transform the parameter estimation of ME-model to solving set of polynomial (Laurent) equations.

Proposition 4.1. *Given the hypothesis (23) the problem of estimating the ME-model amounts to solving set of Laurent polynomial equations.*

Proof. To retain the integer valued exponents in our final solution we consider the constrains of the form

$$N\sum_{j=1}^{m} t_i(j)p_j = \widetilde{R}_i \quad , \quad i = 1, \ldots d \quad , \tag{24}$$

where $\widetilde{R}_i = \sum_{l=1}^{N} t_i(O_l)$ denotes the sample sum. In this case Lagrangian is

$$\begin{aligned} \widetilde{\Xi}(p, \xi) \equiv S(p) - \xi_0 \left(\sum_{j=1}^{m} p_j - 1 \right) \\ - \sum_{i=1}^{d} \widetilde{\xi}_d N \left(\sum_{j=1}^{m} p_j t_i(j) - \widetilde{R}_i \right) \quad . \end{aligned} \tag{25}$$

This results in the ME-distribution

$$p_j(\xi) = \widetilde{Z}(\xi)^{-1} \exp\left(-N\sum_{i=1}^{d} \widetilde{\xi}_i t_i(j) \right) \quad , \quad j = 1, \ldots, m, \tag{26}$$

where $Z(\xi)$ is normalizing constant given by

$$\widetilde{Z}(\xi) = \sum_{j=1}^{m} \exp\left(-N\sum_{i=1}^{d} \widetilde{\xi}_i t_i(j) \right) \quad . \tag{27}$$

To calculate the parameters we maximize the dual $\widetilde{\Psi}(\widetilde{\xi})$ of $\widetilde{\Xi}(p, \xi)$. That is we maximize the functional

$$\widetilde{\Psi}(\widetilde{\xi}) = \ln \widetilde{Z} + \sum_{i=1}^{d} \widetilde{\xi}_i \widetilde{R}_i \quad . \tag{28}$$

It is equivalent to optimizing the functional

$$\widetilde{\Psi}'(\widetilde{\xi}) = \sum_{j=1}^{m} \exp\left(\sum_{i=1}^{d} \widetilde{\xi}_i \widetilde{R}_i - N \sum_{i=1}^{d} \widetilde{\xi}_i t_i(j)\right)$$

By setting $\ln \widetilde{\theta}_i = \widetilde{\xi}$ we have

$$\tilde{\Psi}'(\widetilde{\theta}) = \sum_{j=1}^{m} \prod_{i=1}^{d} \widetilde{\theta}_i^{(\widetilde{R}_i - N t_i(j))} \tag{29}$$

The solution is given by solving the following set of equations

$$\frac{\partial \widetilde{\Psi}'}{\partial \widetilde{\theta}_j} = 0 \ \ , j = 1, \ldots d. \tag{30}$$

We have

$$\frac{\partial \widetilde{\Psi}'}{\partial \widetilde{\theta}_j} \in k[\widetilde{\theta}_1^{\pm}, \ldots, \widetilde{\theta}_d^{\pm}] \ \ , i = 1, \ldots, d. \tag{31}$$

□

In algebraic statistics, algebraic descriptions are used to analyze the maximum likelihood estimates of exponential models [5]. In the view that maximum likelihood and maximum entropy are related,it will be interesting to compare these two methods from algebraic statistical point of view.

5. Conclusion and Directions for Future research

In this paper we attempted to describe maximum (and hence minimum) entropy model in algebraic statistical framework. We showed that maximum entropy models are toric models when the constraint functions are assumed to be integer valued functions. We demonstrated that when the information is in the form of empirical means the calculation of ME-models can be transformed to solving set of Laurent polynomial equations. Work on computational algebraic algorithms for estimating ME-models are in progress. We hope that this will also shed light on interesting algebraic structures in information theoretic statistics.

Acknowledgments

I would like to thank Prof. Roberto Battiti for providing facilities during the formulation of this work.

References

[1] P. Diaconis. *Group Representations in Probability and Statistics*, volume 11 of *Institute of Mathematical Statistics–Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1988.

[2] M. A. G. Viana and D. S. P. Richards, editors. *Algebraic Methods in Statistics and Probability*, volume 287 of *Contemporary Mathematics*. American Mathematical Society, Providence, RI, 2001.

[3] P. Diaconis and B. Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Annals of Statistics*, 26:363–397, 1998.

[4] G. Pistone, E. Riccomagno, and H. Wynn. *Algebraic Statistics. Computational Commutative Algebra in Statistics*. Chapman and Hall, New York, 2001.

[5] L. Pachter and B. Sturmbfels. *Algebraic Statistics and Computational Biology*. Cambridge University Press, Cambridge, 2005.

[6] O. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester, 2001.

[7] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

[8] I. Csiszár and P. Shields. *Information Theory and Statistics: A Tutorial*, volume 1 of *Foundations and Trends in Communications and Information Theory*. Now Publications, 2004.

[9] W. W. Adams and P. Loustaunau. *An Introduction to Gröbner Bases*, volume 3 of *Graduate Studies in Mathematics*. American Mathematical Society, 1994.

[10] D. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms*. Springer, New York, 2nd edition, 1991.

[11] L. Pottier. Gröbner bases of toric ideals. `Rapport de recherche` 2224, INRIA, Sophia Antipolis, 1997.

[12] A. M. Bigatti, R. La Scala, and L. Robbiano. Computing toric ideals. *Journal of Symbolic Computation*, 27:351–365, 1999.

[13] B. Huber and R. R. Thomas. Computing Gröbner fans of toric ideals. *Experimental Mathematics*, 9(3):321–331, 2000.

[14] Bud Mishra and Chee Yap. Notes on gröbner bases. *Information Sciences*, 48(3):219–252, 1989.

[15] F. Rapallo. Toric statistical models: Parametric and binomial representations. *Annals of the Institute of Statistical Mathematics*, 2006. In press.

[16] M. Drton and S. Sullivant. Algebraic statistical models. *(To be published in) Statistical Sinica*, 2006.

[17] B. Sturmbfels. *Gröbner Bases and Convex Polytopes*, volume 8 of *University Lecture Series*. AMS Press, Providence, RI, 1996.

[18] A. M. Bigatti and L. Robbiano. Toric ideals. *Mat. Contemp.*, 21:1–25, 2001.

[19] F. Pauer and A. Unterkircher. Gröbner bases for ideals in Laurent polynomial rings and their applications to systems of differential equations. *Applicable Algebra in Engineering, Communication and Computing*, 9:271–291, 1999.

[20] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[21] E. T. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, sec-4(3):227–241, 1968.

[22] J. N. Darroch and D. Ratcliff. Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, 43(5):1470–1480, 1972.

[23] Imre Csiszár. A geometric interpreation of Darroch and Ratcliff's generalized iterative scaling. *The Annals of Statistics*, 17(3):1409–1413, 1989.