



ALGEBRAIC STATISTICS & COMPUTATIONAL TECHNIQUES IN TORIC SURFACE

Dr. Manish Aghi

Extension Lecturer, Department of Mathematics, Govt. College, Nagina(Nuh)

Abstract

For different types of toric Fano models, we investigate the maximum likelihood estimation issue. To begin, we investigate the maximum probability degree for all Gorenstein toric Fano variations in dimension 2. We prove that, with the exception of the quintic del Pezzo surface with two ordinary double points, the ML degree equals the surface degree, and we provide explicit formulae for computing the maximum likelihood estimate in closed form whenever the ML degree is less than 5. We then use A-discriminants and interse to inquire into the causes of the ML degree decline. In this article, we survey the state of knowledge about Markov chains as toric statistical models. In it, you'll find a variety of things, including some history, findings from the most up-to-date literature, findings from your own research, and ongoing projects. In this study, we'll look at several optional side steps along the road to algebraic statistics. It provides a decoupled basis for algebraic models with the goal of enhancing the statistical inference instruments necessary for analysing genomes. In addition, this study provides an in-depth look of Markov models and graphical models.

Keywords: Graphical, Markov, Statistical, Toric

INTRODUCTION

Maximum likelihood estimation (MLE) is a basic computational problem in statistics and a common method for estimating parameters. The maximum likelihood estimate is the most probable configuration of model parameters given a collection of observed data and the model of interest. The calculation of maximum likelihood estimates for algebraic statistical models has been made easier with the development of algebraic approaches. Number of complex critical points of the likelihood function over the Zariski closure of the model is the maximum likelihood degree (ML degree) of an algebraic statistical model. An evaluation of a model's maximum likelihood estimation problem's complexity is made. In order to find the local maxima in the probability simplex, an algebraic approach is described for calculating all critical points of the likelihood function. As an upper constraint on the ML degree of special complete intersections, the same paper derives an explicit formula for the ML degree of a projective variety that is a generic complete intersection. In addition, when the divisor corresponding to the rational function is a normal crossings divisor, a geometric characterization of the ML degree of a smooth variety is provided. Relaxing the rigorous smoothness condition and permitting certain moderate singularities allowed the authors of the same study to construct an explicit combinatorial formula for the ML degree of a toric variety. Algebraic geometry and statistical motivation are both discussed in [29], which provides a primer on the geometry underlying the MLE for algebraic statistical models for discrete data.

Statistically, toric varieties are analogous to log-linear ones. Log-linear models have been extensively utilised in statistics and other fields, such as natural language processing, to analyse cross-classified data in multidimensional contingency tables since the pioneering publications by L.A. Goodman in the 1970s. Any toric variety has an upper constraint on its ML degree equal to its degree. Several intriguing groups of toric variants involving the ML degree decrease may be found among toric Fano varieties. Maximum likelihood estimate for Gorenstein toric Fano variety in dimension 2, the Veronese (2, 2) in various scalings, and toric Fano varieties connected to 3-valent phylogenetic trees are the primary areas of research.

STATISTICAL MODELS

Statisticians play a dominant role in data analysis to build statistical inferences. In the field of biology genome sequences are used to interpret data. The DNA chain is composed with nucleic acids Adenine (A_d), Cytosine (C_y), Guanine (G_u) & Thymine (T_y) to the benchmark in biology. Genome sequences (A_d, C_y, G_u, T_y) are used to unveil some true facts about the blueprint for life, their functionality, structure as well as evolution.

The genome sequences processed by DiaNA are known as DNA sequences. DNA sequences can be generated by random process with some assumptions. The statistical model is nothing but the representation of a family of probability distributions to generate DNA sequences. In this chapter parametric statistical models with finite dimension is illustrated, which belongs to same family of probability distributions. The estimation of DNA's parameters is said to be learning in the computer science. These sequences also use tetrahedral dice with tetrahedron shape and having four faces (A_d, C_y, G_u, T_y). When this tetrahedral fair dicerolls then any of the four letters (A_d, C_y, G_u, T_y) will appear with equal probability $1/4$. But in case of a loaded tetrahedral die with different probabilities having sum is equal to unity.

DNA causes each letter in the DiaNA sequence separately with the process described later in this section. Firstly, select dice-1 with probability β_1 , second dice with β_2 & last one with probability $1 - \beta_1 - \beta_2$ [2]. The above defined probabilities β_1 and β_2 are not known to us, but we are having information of letters (A_d, C_y, G_u, T_y) existing on dice face. In the application area of biology, the dice-1 respective to DiaNA which is $A_d + C_y$ rich. The dice-2 respective to DiaNA which is $A_d + C_y$ poor and the dice-3 is known to be a fair die because of its equal probabilities [3].

Let us suppose we have DiaNA sequence of length $L = 48$ shown in example. Since all the 48 characters was initiated independently, that the possible multiplication of the probabilities of these letters:

$$L = PA_d PC_y PG_u PA_d PT_y PC_y \dots PG_u = PA_d^{11} . PC_y^{13} . PG_u^{15} . PT_y^9$$

This above equality is the possible function of DNA's model for information in example. As we know that the parameters β_1 and β_2 are not known to us we describe:

$$L(\beta_1, \beta_2) = PA_d(\beta_1, \beta_2)^{11} . PC_y(\beta_1, \beta_2)^{13} . PG_u(\beta_1, \beta_2)^{15} . PT_y(\beta_1, \beta_2)^9 .$$

This possible function is defined as a real-valued function over the triangle

$$\Phi = \{(\beta_1, \beta_2) \in R^2 : \beta_1, \beta_2 > 0 \& \beta_1 + \beta_2 < 1\}$$

In the pattern of maximum possibility we obtain to maximize $L(\beta_1, \beta_2)$ over the triangle Φ . It is identical but quite appropriate to maximize the log-possible function

$$\begin{aligned} l(\beta_1, \beta_2) &= \log\{L(\beta_1, \beta_2)\} \\ &= 11 \log\{PA_d(\beta_1, \beta_2)\} + 13 \log\{PC_y(\beta_1, \beta_2)\} \\ &+ 15 \log\{PG_u(\beta_1, \beta_2)\} + 9 \log\{PT_y(\beta_1, \beta_2)\} \end{aligned}$$

This optimization problem’s result can be calculated in closed form, by putting the two partial derivatives of the log-possible function equal to zero in the following manner:

$$\frac{\partial l}{\partial \beta_1} = \frac{11}{PA_d} \cdot \frac{PA_d}{\partial \beta_1} + \frac{13}{PC_y} \cdot \frac{PC_y}{\partial \beta_1} + \frac{15}{PG_u} \cdot \frac{PG_u}{\partial \beta_1} + \frac{9}{PT_y} \cdot \frac{PT_y}{\partial \beta_1} = 0$$

$$\frac{\partial l}{\partial \beta_2} = \frac{11}{PA_d} \cdot \frac{PA_d}{\partial \beta_2} + \frac{13}{PC_y} \cdot \frac{PC_y}{\partial \beta_2} + \frac{15}{PG_u} \cdot \frac{PG_u}{\partial \beta_2} + \frac{9}{PT_y} \cdot \frac{PT_y}{\partial \beta_2} = 0$$

These above equalities are nothing but a rational function in (β_1, β_2) . By clearing denominators and by putting some algebraic methods of Grobner fundamentals it can convert these equations into the equalities.

TORIC MODELS

The second family of models with well-behaved likelihood functions are known as the toric models or we can call it as exponential model families. Let us assume matrix $B = (b_{ij})$ having +ve integer $q \times n$ matrix with the phenomena that sum of each column is similar:

$$\sum_{i=1}^q b_{i1} = \sum_{i=1}^q b_{i2} = \dots = \sum_{i=1}^q b_{in} \tag{4}$$

The j^{th} column vector b_j of the matrix B demonstrates the algebraic expression

$$\beta^{b_j} = \prod_{i=1}^q \beta_i^{b_{ij}} \text{ for } j \text{ vary from } 1 \text{ to } n.$$

As per eq. 4 we can notice that these algebraic equations have the equal degree. Since, the image of the orthant $\phi = R_{>0}^q$ of the given toric model with matrix B under the map shown below

$$S : R^q \rightarrow R^n, \phi \mapsto \frac{1}{\sum_{j=1}^n \beta^{b_j}} \cdot (\beta^{b_1}, \beta^{b_2}, \dots, \beta^{b_n}) \tag{5}$$

It is observe that we can scale the variable vector without varying the image: $S(\beta) = S(\lambda \cdot \beta)$. Thus, the dimension of this toric model $S(R_{>0}^q)$ is at most $q-1$. Furthermore, the dimension of $S(R_{>0}^q)$ is equal to rank of matrix B minus one. The denominator polynomial $\sum_{j=1}^n \beta^{b_j}$ is called as the partition function. So, the map S has been illustrated with positive constants $\sigma_1, \dots, \sigma_n > 0$ and modified according to eq. shown below:

$$S : R^q \rightarrow R^n, \phi \mapsto \frac{1}{\sum_{j=1}^n \sigma_j \beta^{b_j}} \cdot (\sigma_1 \beta^{b_1}, \sigma_2 \beta^{b_2}, \dots, \sigma_n \beta^{b_n}) \tag{6}$$

In a statistical toric model, the probabilities log is nothing but the linear functions with the log parameters β_j . That is why; statisticians try to treat some toric models as log-linear models.

Now, optimum estimation for this linear toric model described in eq.5 demonstrates the following optimization problem solution:

$$\text{Maximize } g_1^{\alpha_1} \dots g_n^{\alpha_n} \text{ subject to } (g_1, \dots, g_n) \in S(R_{>0}^q).$$

This optimization problem can also be written as:

$$\text{Maximize } \beta^{B\alpha} \text{ subject to } \beta \in R_{>0}^q \text{ and } \sum_{j=1}^n \beta^{b_j} = 1 \tag{6}$$

Although, we are dealing with multi-index notation for algebraic eq. in $\beta = (\beta_1, \dots, \beta_q)$:

$$\beta^{B\alpha} = \prod_{i=1}^q \prod_{j=1}^n \beta_i^{\sigma_{ij}\alpha_j} = \prod_{i=1}^q \beta_i^{\sigma_{i1}\alpha_1 + \sigma_{i2}\alpha_2 + \dots + \sigma_{in}\alpha_n} \text{ and } \beta^{b_j} = \prod_{i=1}^q \beta_i^{\sigma_{ij}}$$

Thus, $z = B\alpha$ is sufficient for statistical approach & our optimization problem eq. 6 is

$$\text{Maximize } \beta^z \text{ subject to } \beta \in R_{>0}^q \text{ and } \sum_{j=1}^n \beta^{b_j} = 1 \tag{7}$$

In the toric markov chains fix an alphabet Σ with y letters, and we also fix a positive integer I . Now, express a toric model whose statistical space is nothing but the group of all words of length I and described as Σ^I . This toric markov chains model is parameterized by the group Φ of real positive integer $y \times y$ matrices. It means the no. of elements is $e = y^2$ and the no. of states is given by $s=y^I$ [6].

Each toric model having elements as well as states is demonstrated by an $e \times s$ matrix B with real integer values as in shown in previous section. The $e \times s$ matrix which describes the toric Markov model will be illustrated by $B_{y,I}$. The rows & column of the given matrix are indexed by Σ^2 and Σ^I respectively. The parameter of the given matrix $B_{y,I}$ in the row & column indexed by the pair $\psi_1\psi_2 \in \Sigma^2$ and $\xi_1\xi_2\dots\xi_I \in \Sigma^I$ respectively is nothing but the repetition of the pair inside the given word.

It can be quite easy to demonstrate the toric Markov chain model as a general toric model defined with the help of the matrix $B_{y,I}$. For a benchmark example suppose that the words of length $I = 4$ over the binary alphabet $\Sigma = \{0, 1\}$, thus, $y = 2$, $e = 4$ and $s = 16$. The matrix $B_{2,4}$ explained in this section having 4×16 matrix as shown below:

$$\begin{array}{cccccccccccccccc}
 & 0000 & 0001 & 0010 & 0011 & 0100 & 0101 & 0110 & 0111 & 1000 & 1001 & 1010 & 1011 & 1100 & 1101 & 1110 & 1111 \\
 \begin{array}{l} 00 \\ 01 \\ 10 \\ 11 \end{array} & \left(\begin{array}{cccccccccccccccc}
 3 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\
 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 2 & 1 & 1 & 1 & 1 & 0 \\
 0 & 0 & 0 & 1 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 3
 \end{array} \right)
 \end{array}$$

Now, it can be written $R^{2 \times 2}$ for the space of 2×2 matrices as shown below:

$$\beta = \begin{pmatrix} \beta_{00} & \beta_{01} \\ \beta_{10} & \beta_{11} \end{pmatrix}$$

The element space Φ belongs to $R^{2 \times 2}$ exists of each and every matrices β whose four elements are positive real integers. The toric Markov chain model of length I is equals to four for the binary alphabet (y is equals to two) is the image of $\phi = R_{>0}^{2 \times 2}$ under the monomial map

$$b_{2,4} : R^{2 \times 2} \rightarrow R^{16}, \beta \mapsto \frac{1}{\sum_{ijkl} p_{ijkl}} \cdot (p_{0000}, p_{0001}, \dots, p_{1111})$$

Where $p_{i_1 i_2 i_3 i_4} = \beta_{i_1 i_2} \cdot \beta_{i_2 i_3} \cdot \beta_{i_3 i_4}$ for every $i_1 i_2 i_3 i_4 \in \{0,1\}^4$.

The map $b_{y,I}$ is expressed analogously for bigger alphabets as well as huge sequences. The toric Markov chain model $b_{2,4}(\phi)$ is a 3-D object inside the 15-dimensional simplex Y which contains of complete set of probability distributions over the specified state space $\{0,1\}^4$.

The simplest model invariant is described using the equation 15 and the remaining linear invariants derive from the repetitive column phenomena of the matrix $B_{2,4}$.

$$p_{0110} = p_{1011} = p_{1101} \text{ and } p_{0010} = p_{0100} = p_{1001} \quad (16)$$

The relations defined above state that $B_{2,4}$ is nothing but a configuration of only twelve different points. Moreover, there exists four relations which denotes the location of the four non-vertices.

EXPECTATION MAXIMIZATION

In the previous defined section we have gone through the linear as well as toric models which follow the conditions that the possible function has not more than one local maximum. Sometimes, this statement does not suit some other algebraic statistical models of computational biology. Moreover, with the help of suitable example a model is described with multiple local maxima. Expectation-Maximization (EM) is a mathematical approach for optimizing the possible function for the models which does not belongs to either linear or toric category. This approach is best suited to analysis practical problems. Although, it must be focused that expectation maximization is not ensured to obtain a global maximum.

Let us establish EM approach for the family of algebraic statistical models. Assume, $B = b_{ij}(\beta)$ be an $q \times n$ matrix of polynomials similarly shown in toric model in the unknown elements $\beta = (\beta_1, \dots, \beta_q)$. It is supposed that the addition of all elements of matrix B equals to unity and there occurs an open subset $\Phi \subset \mathbb{R}^q$ of acceptable elements such that $b_{ij}(\beta)$ greater than zero for each value of β belongs to Φ . We also acknowledge the matrix B with the polynomial map $B: \mathbb{R}^q \rightarrow \mathbb{R}^{q \times n}$ whose coordinates are the $b_{ij}(\beta)$. Here, $\mathbb{R}^{q \times n}$ represents the qn -dimensional real vector space consisting of every $q \times n$ matrices. So, this matrix can be defined as a hidden model or the complete data model [5].

Now, one can consider a hidden model B such that it consist an easy and reliable algorithmic approach for solving the optimum possible problem. This data model B could be exists in a class of either linear model or a toric model, thus it can be easily said that the possible function has not more than one local maximum in Φ , So this global maximum can be achieved effectively in a smart manner with the help of convex optimization techniques. For a special case of toric models; likewise, the independence model or Markov models, there are simple explicit mathematical approaches to solve the optimum possible estimates.

We are having the linear map which receives a $q \times n$ matrix to its vector of row additions

$$\eta: \mathbb{R}^{q \times n} \rightarrow \mathbb{R}^q, H = (h_{ij}) \mapsto \left(\sum_{j=1}^n h_{1j}, \sum_{j=1}^n h_{2j}, \dots, \sum_{j=1}^n h_{qj} \right)$$

The observed model is nothing but the formation $b = \eta \circ B$ of the hidden model B as well as the marginalization map η . The observed model can be demonstrated as follows:

$$b: \mathbb{R}^d \rightarrow \mathbb{R}^q, \beta \mapsto \left(\sum_{j=1}^n b_{1j}(\beta), \sum_{j=1}^n b_{2j}(\beta), \dots, \sum_{j=1}^n b_{qj}(\beta) \right)$$

Thus, $b_i(\beta) = \sum_{j=1}^q b_{ij}(\beta)$. This model b is called as partial data model. Let us say we are having a vector $\nu = (\nu_1, \nu_2, \dots, \nu_q) \in N^q$ of data for the observed model b . Our main emphasis is to optimize the possible function for the given data w.r.t. the observed model:

$$\text{Maximize } F_{\text{observed}}(\beta) = b_1(\beta)^{\nu_1} b_2(\beta)^{\nu_2} \dots b_q(\beta)^{\nu_q} \text{ subject to } \beta \in \phi$$

This is a quite complex problem, in case of finding multiple local solutions. The above problem would be easily analysed with the help of the hidden model B instead:

$$\text{Maximize } F_{\text{hidden}}(\beta) = b_{11}(\beta)^{\nu_{11}} b_{21}(\beta)^{\nu_{21}} \dots b_{qn}(\beta)^{\nu_{qn}} \text{ subject to } \beta \in \phi$$

The problem is that we do not have any idea about the hidden data, the matrix $F = (\nu_{ij}) \in N^{q \times n}$. As we are having the idea of row sums about the matrix F which is equal to the data given by relation, $\eta(F) = \nu$. The basic formation of the EM fundamentals is described as follows. Firstly, start with assumption what the parameter vector β might be. Secondly make an estimation of given value β as well as the hidden data F .

This latter step is known as the expectation step. It is observed that the expected values are not integers for the hidden data vector. Further, one can find out the optimal solution with the help of efficient and reliable method which we supposed to be available for the hidden model B . This step is described as the maximization step/M-step. It can also be assumed that β^* be the maximum output found in this step. Further this is exchanged with the previous parameter say β by the present one and improved parameter say β^* , and we process the steps in such a manner $E \rightarrow M \rightarrow E \rightarrow M \rightarrow E \rightarrow M \rightarrow \dots$ until we fulfilled the requirement.

In reality, what requires to be demonstrated is that the possible function keep on continuing increases during this given process and that the parameter sequence β which converges to a local maximum. However, the formal relation of Expectation-Maximization algorithm is described below:

$$l_{\text{observed}}(\beta) := \log(L_{\text{observed}}(\beta)) \text{ and } l_{\text{hidden}}(\beta) := \log(L_{\text{hidden}}(\beta))$$

TREE MODEL

Markov chains as well as hidden Markov models are nothing but the branches of tree models, or we can call it as a family of models which we are going to illustrate later. First of all, we define the completely observed tree model, so that hidden tree model can easily be derived from this concept. These specified models are having relation with each other in a manner that the hidden Markov model is made up of with the fully observed Markov model [9].

Let us consider R , which is said to be a rooted tree with l leaves. We can express $S(R)$ as a set of all nodes of R . This defined set contains the root, which is represented by p & the leaves are indexed by this relation $[l] = \{1, 2, \dots, l\}$. The set $D(R)$ of edges of R is nothing but a subset of $S(R) \times S(R)$. Each edge is projected far from the given root p . We define the term “ ek ” for edges (e, k) belong to $D(R)$. All nodes n belongs to $S(R)$ demonstrates a random variable which contains values in a finite alphabet Σ_n . Now the defined tree models are parameterized by a set of different matrices β^{ek} , corresponding to every edge (e, k) belongs to $D(R)$. The rows of the mentioned matrix β^{ek} are represented by Σ_e , likewise the columns are denoted as Σ_k . Although, we did not focus on non-negative matrices with unity sum.

GRAPHICAL MODELS

The statistical models in earlier sections are basic fundamental branches of Graphical models. A certain algebraic statistical models for joint probability distributions of l random variables Y_1, Y_2, Y_l are said to be discrete graphical models which can be explained in two possible different ways: [10]

- By a parameterization $h: \mathbb{R}^q \rightarrow \mathbb{R}^n$
- By a set of conditional independence statements

Our main emphasis over Hamersley-Clifford Theorem, which explains conditional independence fundamentals achieved from various graphs. These graphs that belong to graphical models are playing very important role in developing effective inference algorithms. Let us consider that each random variable Y_n receives its values in a finite alphabet Σ_n .

CONCLUSION

In this chapter, we have discussed few auxiliary steps through the algebraic statistics path. In this chapter various model like linear, toric as well as markov model has been illustrated. Further, Expectation-Maximization (EM) is expressed for optimizing the possible function for the models which does not belongs to either linear or toric category. This approach is best suited to analysis practical problems. Although, it must be focused that expectation maximization is not ensured to obtain a global maximum. Moreover, Markov models and graphical models are revealed in details.

REFERENCES

1. Corey Harris and Martin Helmer. Segre class computation and practical applications. *Math. Comp.*, 89(321):465–491, 2020.
2. Leonor Godinho, Frederik von Heymann, and Silvia Sabatini. 12, 24 and beyond. *Adv. Math.*, 319:472–521, 2017.
3. S. Fiorini: f0; 1 2g-cuts and the straight requesting issue: surfaces that characterize aspects, *SIAM J. Discrete Math.* 20 (2006), 893_912.
4. D. Geiger, C. Docile and B. Sturmfels: On the toric variable based math of graphical models, *Ann. Statist.* 34 (2006), 1463-1492.
5. G. Hommel, F. Bretz and W. Maurer: Powerful alternate routes for various testing systems with uncommon reference to gatekeeping procedures, *Statist. Med.* 26 (2007) 4063-4073.
6. Katsabekis and A. Thoma: Parametrizations of toric assortments over any field, *J. Variable based math* 308 (2007) 751_763.
7. V. Reiner, F. Saliola and V. Welker: Spectra of symmetrized shu_ing administrators, composition, 2010.
8. J. D. Hauenstein, J. Rodriguez, and B. Sturmfels, Maximum likelihood for matrices with rank constraints, *Journal of Algebraic Statistics*, 5 (2014), 18-38.
9. Seth Sullivant. Algebraic statistics, volume 194 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2018.
10. Carlos Améndola, Nathan Bliss, Isaac Burke, Courtney R. Gibbons, Martin Helmer, Serkan Hoşten, Evan D. Nash, Jose Israel Rodriguez, and Daniel Smolkin. The maximum likelihood degree of toric varieties. *J. Symbolic Comput.*, 92:222–242, 2019.