MAXIMUM LIKELIHOOD ESTIMATION ON DETERMINANTAL VARIETIES

ADAM MERBERG

1. Introduction

Given a statistical model $M \in \mathbb{R}^I$ and a vector of counts $u \in \mathbb{N}^I$ for some finite index set I, the problem of maximum likelihood estimation is to find the element $P \in M$ maximizing the likelihood of observing the data u, that is, maximizing the function

$$L(P) \;=\; \prod_{i \in I} p_i^{u_i}$$

over the model M. We consider, in particular, problems relating to maximum likelihood estimation on the determinantal varieties of $m \times n$ matrices of rank at most r.

2. Maximum Likelihood Degree

We consider the problem of the maximum likelihood degree of the algebraic statistical models given by determinantal varieties. Before doing so, we define the maximum likelihood degree. We define the likelihood locus Z_u of a data vector u to be the set of all regular points of $V(P) \setminus V\left(\left(\prod_{i \in I} p_i\right) \cdot \sum_{i \in I} p_i\right)$ such that the gradient of $L(p)$ lies in the tangent space of $V(P)$ at p. There is a Zariski dense open subset $\mathcal{V} \subset \mathbb{R}^k$ such that for each $u \in \mathcal{V}$, the likelihood locus is a finite set, and the number of points in this set is independent of the choice of $u \in \mathcal{V}$. We define the maximum likelihood degree (ML degree) as $\#Z_u$ for any $u \in \mathcal{V}$.

The maximum likelihood degree can be computed using Algorithm 2.2.9 of [1], but this algorithm will not run to completion for more complicated models on typical computers. The saturation step of this algorithm is particularly resource-intensive. However, by omitting this saturation step, we are able to compute the following upper bounds on the maximum likelihood degree of the model of $m \times n$ probability matrices of rank at most r:

(r, m, n)	upper bound on ML degree
$(2, 3, 5)$	59
$(2, 3, 6)$	123

We remark that the bounds are computable in these cases because the number of 2×2 minors of 3×5 and 3×6 matrices is relatively small.

3. Generalizing the 100 Swiss Francs Conjecture

In [2], the authors prove the following conjecture of Sturmfels.

100 Swiss Francs Conjecture. *The maximum value of the likelihood function*

$$L(P) \;=\; \prod_{i=1}^{4} p_{ii}^4 \times \prod_{i\neq j} p_{ij}^2$$

over the set of all 4×4 probability matrices $P=(p_{ij})$ of rank at most 2 is attained for

$$P \;=\; \frac{1}{40}\begin{pmatrix} 3 & 3 & 2 & 2 \\ 3 & 3 & 2 & 2 \\ 2 & 2 & 3 & 3 \\ 2 & 2 & 3 & 3 \end{pmatrix}.$$

The also conjecture the following generalization of this result:

Conjecture 1. *For given $0<t<s$ where t,s are two integers, among the set of all non-negative 4×4 matrices whose rank is at most 2 and whose entries sum to 1, the matrix*

$$P = \frac{1}{4s+12t}\begin{pmatrix} \frac{s+t}{2} & \frac{s+t}{2} & t & t \\ \frac{s+t}{2} & \frac{s+t}{2} & t & t \\ t & t & \frac{s+t}{2} & \frac{s+t}{2} \\ t & t & \frac{s+t}{2} & \frac{s+t}{2} \end{pmatrix}$$

is a global maximum for the likelihood function

$$L_{s,t}(P) = \prod_{i=1}^{4} p_{ii}^s \times \prod_{i\neq j} p_{ij}^t.$$

We show that the method of [2] extends to the following case of this generalization:

Theorem 1. *Let M be the unique positive root of the polynomial*

$$p(x) \;=\; 9x^7+23x^6-182x^5-150x^4+2525x^3+1907x^2-12640x-16068.$$

(Note $M\approx 2.761157$). The conclusion of Conjecture 1 holds provided that $1<s/t<M$.

The proof uses the same methods as [2], so we primarily provide details where they diverge from that source.

As in [2], we can scale our probability matrices by $4^2=16$, so that the sum of the entries is 16. Evidently, this scaling will not affect the maximization problem. The row and column sums of an optimal solution must all be 4, so using the singular value decomposition theorem, can then write a rank 2 matrix P as

$$P = \begin{pmatrix} 1+a_1b_1 & 1+a_2b_1 & 1+a_3b_1 & 1+a_4b_1 \\ 1+a_1b_2 & 1+a_2b_2 & 1+a_3b_2 & 1+a_4b_2 \\ 1+a_1b_3 & 1+a_2b_3 & 1+a_3b_3 & 1+a_4b_3 \\ 1+a_1b_4 & 1+a_2b_4 & 1+a_3b_4 & 1+a_4b_4 \end{pmatrix}.$$

For convenience of notation, we let $\alpha = s/t-1$. We find the following analog of Lemma 2 in [2].

Lemma 1. *A global maximum of $L_{s,t}$ for any s,t satisfies*

$$\sum_{j=1}^{4} \frac{b_j}{1+a_ib_j} + \frac{\alpha b_i}{1+a_ib_i} = 0, \quad 1 \le i \le 4 \tag{1}$$

and

$$\sum_{i=1}^{4} \frac{a_i}{1+a_ib_j} + \frac{\alpha a_j}{1+a_jb_j} = 0, \quad 1 \le j \le 4. \tag{2}$$

The following is an analog of Corollary 3.

Corollary 1. *A matrix P maximizing $L_{s,t}$ satisfies*

$$\sum_{j=1}^{4} \frac{1}{1+a_ib_j} + \frac{\alpha}{1+a_ib_i} = 4+\alpha, \quad 1 \le i \le 4 \tag{3}$$

and

$$\sum_{i=1}^{4} \frac{1}{1+a_ib_j} + \frac{\alpha}{1+a_jb_j} = 4+\alpha, \quad 1 \le j \le 4. \tag{4}$$

The next results, which are Lemmas 4 and 5 in [2] also holds more generally with the same proof.

Lemma 2. *If P maximizes $L_{s,t}$, then the following are true for every i:*

(1) *$a_i = 0$ if and only if $b_i = 0$, and*
(2) *$a_i > 0$ if and only if $b_i > 0$.*

Lemma 3. *If P maximizes $L_{s,t}$, then the following are true for every i:*

(1) *$a_i = a_j$ if and only if $b_i = b_j$, and*
(2) *$a_i > a_j$ if and only if $b_i > b_j$.*

As in [2], we may make certain simplifying assumptions.

Assumption 1. *We can assume that some P maximizing $L_{s,t}$ satisfies*

(1) *$a_1 \ge a_2 \ge a_3 \ge a_4$ and $b_1 \ge b_2 \ge b_3 \ge b_4$,*
(2) *$a_1 = b_1 \ge 0$, and*
(3) *$a_1 \ge a_2 \ge 0$.*

We are now able to introduce an analog to Lemma 7. It is only now that we need to restrict our choice of s, t.

Lemma 4. *If P maximizes $L_{s,t}$ with $s/t < M+1$, then $a_2 = b_2$.*

While the method of proof is the same as [2], we present some details here because the computations are somewhat more complicated.

Proof. If either of a_2 or b_2 is 0, then $a_2 = 0 = b_2$. We can thus assume that $a_2 > 0$ and since $a_1 \geq a_2$, it follows that $a_1 > 0$ also.

Applying the first equation of Corollary 1 with $i = 1$, we find

$$\frac{2}{1+a_1^2} + \frac{1}{1+a_2a_1} + \frac{1}{1+a_3a_1} + \frac{1}{1+a_4a_1} = 4+\alpha.$$

Since the rows and columns of P sum to 4, we also have

$$a_1^2 + a_2a_1 + a_3a_1 + a_4a_1 \;=\; 0.$$

Defining

$$f_1(x,y) = \frac{2-x-y}{4+\alpha-\frac{1+\alpha}{1+x^2}-\frac{1}{1+y}} + x + y - 1,$$

it follows that

$$a_3a_a \cdot a_4 \cdot a_1 = f_1(a_1^2, a_1a_2).$$

We similarly find (using Corollary 1 with $i = 2$) that

$$a_3b_2 \cdot a_4b_2 = f_1(a_2b_2, a_1b_2).$$

Since a_1 and b_2 are nonzero, it follows that

$$\frac{f_1(a_1^2, a_1a_2)}{a_1^2} \;=\; \frac{f_1(a_2b_2, a_1b_2)}{b_2^2}.$$

Normalizing allows us to find a polynomial $f_2(x,y,z)$ such that

$$f_2(a_1, a_2, b_2) \;=\; 0.$$

Similarly applying the second equation in Corollary 1 with $j = 1$ and $j = 2$, we find that $f_2(a_1, b_2, a_2) = 0$. Thus

$$f_2(a_1, a_2, b_2) - f_2(a_1, a_2, b_2) \;=\; 0.$$

The left hand side factors, yielding

$$(a_2 - b_2)f_3(a_1, a_2, b_2) = 0$$

where

$$\begin{aligned} f_3(a_1,a_2,b_2) \;=\; & a_2^2((2(\alpha+3)(\alpha+4))a_1^4b_2^2 + (2(\alpha+4)(\alpha+2))a_1^3b_2 \\ & +(-\alpha^2-\alpha+8)a_1^2b_2^2 + (-(\alpha+3)(\alpha-2))a_1b_2 + (-2\alpha-6)b_2^2) \\ & +a_2((\alpha^2-\alpha+8)a_1^4b_2 + (2\alpha^2+12\alpha+16)a_1^3b_2^2 + (-\alpha^2-\alpha+6)a_1^3 \\ & +(2\alpha^2+6\alpha+12)a_1^2b_2 + (-\alpha^2-\alpha+6)a_1b_2^2 - 6\alpha a_1 + (-\alpha^2+\alpha)b_2) \\ & +(2\alpha-6)a_1^4 + (-\alpha^2-\alpha+6)a_1^3b_2 + (-\alpha^2-\alpha)a_1^2 - 6\alpha a_1b_2 - 4\alpha \end{aligned}$$

Using the bounds in Lemma 5, we find that

$$\begin{aligned} f_3(a_1,a_2,b_2) \;<\; & 2(\alpha+3)/(\alpha+4)^3\alpha^4+2/(\alpha+4)^2(\alpha+2)\alpha^3 \\ & +(-\alpha^2-\alpha+8)\alpha^2/(\alpha+4)^2 \\ & -(\alpha+3)(\alpha-2)(a_2b_2)\alpha/(\alpha+4)+(-2\alpha-6)(a_2b_2)^2 \\ & +3(-\alpha^2-\alpha+8)\alpha^3/(\alpha+3)/(\alpha+4)^2 \\ & +(2\alpha^2+12\alpha+16)\alpha^3/(\alpha+4)^3-3(\alpha-2)\alpha^2/(\alpha+4) \\ & +(2\alpha^2+6\alpha+12)\alpha^2/(\alpha+4)^2+(-\alpha^2-\alpha+6)\alpha/(\alpha+4)(a_2b_2) \\ & +(-\alpha^2+\alpha)(a_2b_2)+3(-\alpha^2-\alpha+6)\alpha^2/(\alpha+3)/(\alpha+4)-4\alpha \end{aligned}$$

This bound is quadratic in a_2b_2 and is negative so long as its discriminant (after normalizing),

$$\alpha\cdot p(\alpha+1) \;=\; 9\alpha^7+86\alpha^6+145\alpha^5-400\alpha^4+880\alpha^3+7296\alpha^2-2560\alpha-24576$$

is negative, which is the case so long as $\alpha+1<M$, or equivalently $s/t<M$. □

Finally, we give an analog of Lemma 8 of [2].

Lemma 5. *Suppose that P maximizes $L_{s,t}$. Then*

(1) $a_1^2 \le \frac{2\alpha}{3+\alpha}$
(2) *If* $\alpha < \frac{3+\sqrt{105}}{2}$ *then* $a_1a_2 \le \frac{\alpha}{4+\alpha}$, *and*
(3) *If* $\alpha < \frac{3+\sqrt{105}}{2}$ *then* $a_1b_2 \le \frac{\alpha}{4+\alpha}$.

Proof. The proof of the first statement is entirely parallel to the case $\alpha=1$ as given in [2], and is omitted.

The proof of the second statement is almost the same as for the case $\alpha=1$, but we include the proof to explain the role of the upper bound for α here. For $i=1,2,3,4$, define $A_i=1+a_1a_i$ and suppose that $A_2 = 1+a_1a_2 > 1+\frac{\alpha}{4+\alpha} = \frac{2\alpha+4}{4+\alpha}$ Define

$$g(x) \;=\; \frac{1+\alpha}{A_1}+\frac{1}{x}+\frac{4}{4-A_1-x}.$$

Then

$$g'(x) \;=\; -\frac{1}{x^2}+\frac{4}{(4-A_1-A_2)^2}$$

Since $a_1 \ge a_2 > 0$, we have $A_1 \ge A_2 \ge 1$, whence $4-A_1-x \le 2$ for all $x>1$, so $g'(x)>0$ for $x>1$. Thus g is increasing between $\frac{2\alpha+4}{4+\alpha}$ and A_2, whence

$$\frac{1+\alpha}{A_1}+\frac{1}{A_2}+\frac{4}{4-A_1-A_2} > \frac{1+\alpha}{A_1}+\frac{4+\alpha}{4+2\alpha}+\frac{4}{\frac{12+2\alpha}{4+\alpha}-A_1}$$

Since

$$\frac{1}{A_3}+\frac{1}{A_4} \ge \frac{4}{A_3+A_4} \;=\; \frac{4}{4-A_1-A_2},$$

we see that

$$\frac{1+\alpha}{A_1}+\frac{1}{A_2}+\frac{4}{4-A_1-A_2} = \frac{1+\alpha}{A_1}+\frac{1}{A_2}+\frac{4}{A_3+A_4} \leq \frac{1+\alpha}{A_1}+\frac{1}{A_2}+\frac{1}{A_3}+\frac{1}{A_4} = 4+\alpha,$$

whence

$$4+\alpha > \frac{1+\alpha}{A_1} + \frac{4+\alpha}{4+2\alpha} + \frac{4}{\frac{12+2\alpha}{4+\alpha} - A_1}.$$

Here, the condition that $\alpha < \frac{3+\sqrt{105}}{2}$ implies that the denominator on the rightmost fraction is positive, whence we can find that the solution set for A_1 in the inequality is $(-\infty, 0) \cup \left(\frac{2(\alpha^2+7\alpha+6)}{2\alpha^2+11\alpha+12}, \frac{4+2\alpha}{4+\alpha}\right) \cup \left(\frac{12+2\alpha}{4+\alpha}\right)$. Since $A_1 > 0$ and $A_1 = 1 + a_1^2 \leq \frac{3+3\alpha}{3+\alpha}$, it follows that we must have $A_1 \in \left(\frac{2(\alpha^2+7\alpha+6)}{2\alpha^2+11\alpha+12}, \frac{4+2\alpha}{4+\alpha}\right)$. But this contradicts $A_1 \geq A_2$, so the stated upper bound must hold.

The third inequality is proven by letting $A_i = 1 + a_1 b_i$. The same proof will work, as in [2]. $\square$

The following two results are proven exactly as in the case $s/t = 2$ as described in [2], so we omit the proofs.

Proposition 1. *$a_3 = b_3$ and $a_4 = b_4$.*

Proposition 2. *$a_i a_j > 0$ if and only if $a_i = a_j$.*

Thus, the problem is reduced to comparing likelihood functions for different sign patterns for the a_i. From the order assumptions, we need only consider four different sign patterns: $(+,+,+,-)$, $(+,+,-,-)$, $(+,+,0,-)$, and $(+,0,0,-)$. For the first of these patterns, we have $a_1 = a_2 = a_3 > 0$, which by the equations in Corollary 1 gives us $a_1 = a_2 = a_3 = \sqrt{\frac{s-t}{3s+9t}}$, which gives rise to the matrix

$$P_1 = \frac{4}{3s+9t} \begin{pmatrix} s+2t & s+2t & s+2t & 3t \\ s+2t & s+2t & s+2t & 3t \\ s+2t & s+2t & s+2t & 3t \\ 3t & 3t & 3t & 3s. \end{pmatrix}$$

For the second sign pattern, we find $a_1 = a_2 = \sqrt{\frac{s-t}{s+3t}}$ and then

$$P_2 = \frac{1}{s+3t} \begin{pmatrix} 2s+2t & 2s+2t & 4t & 4t \\ 2s+2t & 2s+2t & 4t & 4t \\ 4t & 4t & 2s+2t & 2s+2t \\ 4t & 4t & 2s+2t & 2s+2t \end{pmatrix}$$

We show that $L_{s,t}(P_2) > L_{s,t}(P_1)$. Consider the ratio of likelihood functions as a function of s and t:

$$\begin{aligned} R(s,t) := \frac{L_{s,t}(P_2)}{L_{s,t}(P_1)} &= \frac{\frac{1}{(s+3t)^{4s+12t}} \cdot (4t)^{8t}\,(2s+2t)^{4s+4t}}{\left(\frac{4}{3s+9t}\right)^{4s+12t} \cdot (s+2t)^{3s+6t}\,(3t)^{6t} \cdot (3s)^s} \\ &= \frac{3^{4s+12t} \cdot (4t)^{8t}\,(2s+2t)^{4s+4t}}{4^{4s+12t} \cdot (s+2t)^{3s+6t}\,(3t)^{6t} \cdot (3s)^s} \\ &= \frac{3^{3s+6t} \cdot t^{2t}\,(2s+2t)^{4s+4t}}{4^{4s+4t} \cdot (s+2t)^{3s+6t} \cdot s^s} \end{aligned}$$

Evidently, $R(t,t) = 1$ for any t. We compute the partial derivative with respect to s.

$$\begin{aligned} \frac{\partial R}{\partial s}(s,t) &= 3^{3s+6t} \cdot 16^{-s-t} \cdot t^{2t}(s+t)^{4s+4t} s^{-s}(s+2t)^{-3s-6t} \\ &\quad \cdot(\log(27/16) - \log(s) + 4\log(s+t) - 3\log(s+2t)) \\ &= 3^{3s+6t} \cdot 16^{-s-t} \cdot t^{2t}(s+t)^{4s+4t}(s^{-s}(s+2t)^{-3s+-6t}) \\ &\quad \cdot \log\left(\frac{27s(s+2t)^3}{16(s+t)^4}\right). \end{aligned}$$

All but the last factor are positive for any s and t, and it is an exercise in elementary calculus to show that the last factor in the last line is positive for $s > t$. Thus for fixed t, R is increasing with respect to s, from which it follows that $R(s,t) > 1$ for $s > t$. Thus $L_{s,t}(P_2) > L_{s,t}(P_1)$.

For the other two sign patterns, we proceed similarly, and it follows that P_2 is a global maximum of $L_{s,t}$, which is what we needed.

Although we are only able to use this method to prove the conjecture with $1 < s/t < M$, we expect that the result is true more generally. Computations for integer values $3 \leq s/t \leq 5000$ using the EM algorithm (c.f. [3]) did not find any counterexamples.

References

[1] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on Algebraic Statistics*. Birkhäuser Verlag, Basel, 2008.

[2] Shuhong Gao, Guangran Jiang, and Mingfu Zhu. Solving the 100 Swiss Francs Problem. 2008.

[3] Lior Pachter and Bernd Sturmfels. *Algebraic statistics for computational biology*. Cambridge Univ. Press, New York, 2005.