

Maximum Likelihood Estimation from a Tropical and a Bernstein–Sato Perspective

Anna-Laura Sattelberger (MPI MiS, Leipzig)

Onlineseminar on Applied Algebra and Analysis
TU Braunschweig and Osnabrück University

February 12, 2021



“ D -modules, Statistics, and Tropical Geometry
don't have anything in common.”

~~“ D -modules, Statistics, and Tropical Geometry
don't have anything in common.”~~

Yes, they do!

What is this talk about?

[vdVS21] Robin van der Veer¹ and A.-L. S.: [Maximum Likelihood Estimation from a Tropical and a Bernstein–Sato Perspective](#). arXiv:2101.03570, 2021.

Connecting three fields of research

- ▶ Bernstein–Sato Theory
- ▶ Likelihood Geometry
- ▶ Tropical Geometry

Providing new tools for...

- ▶ Algebraic Statistics
- ▶ High Energy Physics: scattering amplitudes [ST20]

¹KU Leuven, Belgium

Maximum Likelihood Estimation

Discrete statistical experiment: Flip a biased coin. If it shows *head*, flip again.

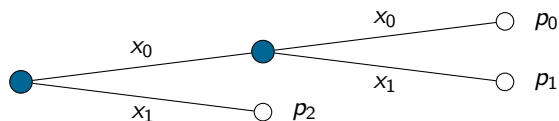


Figure: Staged tree modeling the discrete statistical experiment [DMS21]

- ▶ (s_0, s_1, s_2) count of outcome $\rightsquigarrow 2s_0 + 2s_1 + s_2$ coin tosses
- ▶ Maximum Likelihood Estimate (MLE):

$$\Psi(s_0, s_1, s_2) = \left(\frac{(2s_0 + s_1)^2}{(2s_0 + 2s_1 + s_2)^2}, \frac{(2s_0 + s_1)(s_1 + s_2)}{(2s_0 + 2s_1 + s_2)^2}, \frac{s_1 + s_2}{2s_0 + 2s_1 + s_2} \right)$$

- ▶ Parametrization of the model: $\Delta_1 \rightarrow \Delta_2$, $(x_0, x_1) \mapsto (x_0^2, x_0x_1, x_1)$, where $x_0, x_1 > 0$, $x_0 + x_1 = 1$
- ▶ Implicitization: $\mathcal{M} := V(p_0p_2 - (p_0 + p_1)p_1) \subseteq \mathbb{P}^2$

Bernstein–Sato ideals

$D = \mathbb{C}[x_1, \dots, x_n] \langle \partial_1, \dots, \partial_n \rangle$ the **Weyl algebra**, $[\partial_i, x_i] = \partial_i x_i - x_i \partial_i = 1$
 $F = (f_1, \dots, f_p) \in \mathbb{C}[x_1, \dots, x_n]^p$ a tuple of polynomials

Definition

The **Bernstein–Sato ideal** of F is the ideal B_F in $\mathbb{C}[s_1, \dots, s_p]$ of polynomials b for which there exists $P \in D[s_1, \dots, s_p]$ such that

$$P \bullet \left(f_1^{s_1+1} \cdots f_p^{s_p+1} \right) = b \cdot f_1^{s_1} \cdots f_p^{s_p}.$$

Example: $F = (x^2, x(1-x), 1-x)$

Observed in [SS19, Example 3.1]:²

$$B_F = \left\langle \prod_{k=1}^3 (2s_0 + s_1 + k) \cdot \prod_{\ell=1}^2 (s_1 + s_2 + \ell) \right\rangle \triangleleft \mathbb{C}[s_0, s_1, s_2].$$

²computed with the library `dmod.lib` in Singular

How to compute the MLE in practice?

In **holonomic** case: **Holonomic Gradient Method** [NNN⁺11]

- ▶ numerical evaluation of holonomic functions
- ▶ keeping track of gradient by Pfaffian system
- ▶ Holonomic Gradient Descent: minimization method based on HGM
- ▶ applied to Fisher distribution of rotation data in [SST⁺13]
 - ▶ generalized to compact Lie groups other than $SO(n)$ in [ALSS20]
- ▶ applied to distribution of largest eigenvalue of Wishart matrix in [HN⁺TT13]
 - ▶ further study of Muirhead's D -ideal in [GLS20]

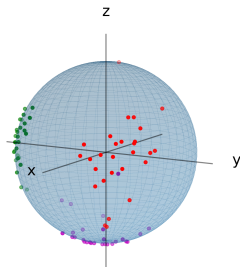


Figure: A dataset from a study in vectorcardiography [DLM74]

Likelihood Geometry [HS14]

\mathbb{P}^2 homogeneous coordinates $(p_0 : p_1 : p_2)$

Identification of the following two:

- 1 real points in \mathbb{P}^2 with $\text{sign}(p_0) = \text{sign}(p_1) = \text{sign}(p_2)$
- 2 $\Delta_2 = \{(p_0, p_1, p_2) \in \mathbb{R}^3 \mid p_0, p_1, p_2 > 0 \text{ and } p_0 + p_1 + p_2 = 1\}$

The likelihood function ℓ

- ▶ given $(s_0, s_1, s_2) \in \mathbb{N}_{>0}^3$, $\ell_{s_0, s_1, s_2}(p_0, p_1, p_2) := \frac{p_0^{s_0} p_1^{s_1} p_2^{s_2}}{(p_0 + p_1 + p_2)^{s_0 + s_1 + s_2}}$.
- ▶ ℓ regular function on $\mathbb{P}^2 \setminus \mathcal{H}$ with $\mathcal{H} := \{p \in \mathbb{P}^2 \mid p_0 p_1 p_2 (p_0 + p_1 + p_2) = 0\}$
- ▶ critical points of ℓ : zeros of $\text{dlog } \ell$, the **logarithmic differential** of ℓ

Statistics	Algebraic Geometry
statistical model	smooth curve \mathcal{M} in \mathbb{P}^2
parameters of the model	point in $\mathcal{M} \cap \Delta_2$
MLE problem	maximizing ℓ over $\mathcal{M} \cap \Delta_2$

Two important entities

- ▶ Likelihood locus the set of critical points of ℓ_{u_0, u_1, u_2} over $\mathcal{M} \cap \Delta_2$
- ▶ ML degree cardinality of the likelihood locus for a general data vector

Critical slopes

X	smooth variety
F	a tuple of nowhere-vanishing regular functions (f_1, \dots, f_p) on X
$X \hookrightarrow Y$	a smooth compactification with boundary $E = Y \setminus X$
f^α	$f_1^{\alpha_1} \cdots f_p^{\alpha_p}$
E_1, \dots, E_q	the irreducible components of E

Some definitions

H_{E_i}	the hyperplane $\{\text{ord}_{E_i}(f_1)s_1 + \cdots + \text{ord}_{E_i}(f_p)s_p = 0\} \subseteq \mathbb{P}^{p-1}$
C_F	the critical locus of $F: \{(x, \alpha) \mid \text{dlog } f^\alpha(x) = 0\} \subseteq X \times \mathbb{P}^{p-1}$, with $\text{dlog } f^\alpha = \sum_{i=1}^p \alpha_i \frac{df_i}{f_i} \in \Gamma(X, \Omega_X^1)$ the logarithmic differential of f^α
S_F	the critical slopes of $F: \pi_2(\overline{C_F}^{Y \times \mathbb{P}^{p-1}} \cap \pi_1^{-1}(E)) \subseteq \mathbb{P}^{p-1}$, π_1, π_2 the projections from $Y \times \mathbb{P}^{p-1}$

Refined asymptotic behavior or critical points

- X a smooth variety
- F a tuple (f_1, \dots, f_p) of nowhere vanishing regular functions on X
- Y a compactification of X
- π_1 the projection from $Y \times \mathbb{P}^{p-1}$ to the first factor
- Δ the formal disc $\text{Spec } \mathbb{C}[[t]]$ around 0
- Δ° the punctured formal disc $\text{Spec } \mathbb{C}((t))$

Definition ($Q_{F,\alpha}$)

Let $\alpha \in \mathbb{P}^{p-1}$. An integer vector $v \in \mathbb{Z}^p$ is in $Q_{F,\alpha}$ if $v = (\text{ord}_t(\gamma^*(\pi_1^*F)))$ for some $\gamma: \Delta \rightarrow Y \times \mathbb{P}^{p-1}$ such that $\gamma(\Delta^\circ) \in C_F$ and $\gamma(0) \in Y \setminus X \times \{\alpha\}$.

Then: $S_F = \{\alpha \in \mathbb{P}^{p-1} \mid Q_{F,\alpha} \neq \emptyset\}$.

Very affine varieties

Definition

A **very affine variety** is a closed subvariety $X \hookrightarrow (\mathbb{C}^*)^p$ of the algebraic p -torus.

Examples

- ▶ $\mathbb{P}^2 \setminus \mathcal{H} \hookrightarrow (\mathbb{C}^*)^3$, $(p_0 : p_1 : p_2) \mapsto \left(\frac{p_0}{p_0+p_1+p_2}, \frac{p_1}{p_0+p_1+p_2}, \frac{p_2}{p_0+p_1+p_2} \right)$
- ▶ $\mathcal{M} \setminus \mathcal{H}$
- ▶ complements of essential hyperplane arrangements

Theorem ([FK00], [Huh14])

For X smooth, very affine of dimension d : $d_{\text{ML}}(X) = (-1)^d \chi(X)$.

Tropical varieties

$X \subseteq (\mathbb{C}^*)^p$ very affine variety defined by $I \triangleleft \mathbb{C}[t_1^{\pm 1}, \dots, t_p^{\pm 1}]$

Fundamental Theorem of Tropical Geometry [MS15, Thm. 3.2.3]

The **tropical variety** of X is $\text{trop}(X) := \{w \in \mathbb{R}^p \mid \text{in}_w(I) \neq \langle 1 \rangle\}$.

Definition

A ray τ is **rigid** if any small perturbation of the ray changes the initial ideal of I w.r.t. the primitive generator $v_\tau \in \mathbb{Z}^p$ of τ .

Tropical compactifications

$X \hookrightarrow (\mathbb{C}^*)^p$ a very affine variety
 $\Sigma \subseteq \mathbb{R}^p$ a fan
 \mathbb{T}^Σ toric variety of Σ

Definition ([Hac07])

X is **schön** iff there exists a fan structure Σ on $\text{Trop}(X)$ s.t. the closure X^Σ of X in \mathbb{T}^Σ is proper, smooth and $X^\Sigma \setminus X$ is a simple normal crossing divisor.

Theorem ([LQ11])

If X is schön, any fan supported on $\text{Trop}(X)$ can be refined to Σ s.t. X^Σ is a smooth SNC compactification of X .

Such X^Σ is a **tropical compactification** of X .

Codimension-one components of S_F

- X a schön very affine variety
- Σ a fan supported on $\text{Trop}(X)$
- X^Σ tropical compactification of X
- \mathcal{O}_τ torus orbit in the toric variety \mathbb{T}^Σ
- E_i irreducible boundary component
- E_i° $E_i \setminus \cup_{j \neq i} (E_j \cap E_i)$

[vdVS21, Theorem 2.7]

Assume $\chi(E_i^\circ) \neq 0$ for all i . For every E_i , the hyperplane H_{E_i} is contained in S_F . Those are the only codimension-one components of S_F .

[vdVS21, Proposition 2.12]

Assume $X^\Sigma \cap \mathcal{O}_\tau$ is connected for all $\tau \in \Sigma$. Then the rigid rays in $\text{Trop}(X)$ are in bijection with the codimension-one components of S_F .

Bernstein–Sato ideals

Y a smooth algebraic variety

G a tuple of regular functions (g_1, \dots, g_p) on Y

Definition

The **Bernstein–Sato ideal** of G is the ideal B_G in $\mathbb{C}[s_1, \dots, s_p]$ of polynomials b for which there exists a global algebraic linear partial differential operator $P \in \Gamma(X, \mathcal{D}_Y[s_1, \dots, s_p])$ such that

$$P \bullet \left(g_1^{s_1+1} \cdots g_p^{s_p+1} \right) = b \cdot g_1^{s_1} \cdots g_p^{s_p}.$$

- ▶ $V(B_G) \subseteq \mathbb{C}^p$ the **Bernstein–Sato variety** of G
- ▶ codimension-one components of $V(B_G)$ are affine hyperplanes

Bernstein–Sato slopes BS_G

- Y smooth closed subvariety of \mathbb{C}^p
- G the tuple of coordinate functions on \mathbb{C}^p restricted to Y
- BS_G the affine hyperplanes of $V(B_G)$ translated to the origin
- X the very affine variety $Y \cap (\mathbb{C}^*)^p$
- F the tuple of coordinate functions restricted to X

[Mai16, Résultat 6]

Let $W_G = \left\{ \left(\sum_{i=1}^p \alpha_i \frac{dg_i}{g_i}(x), \alpha \right) \mid x \in X, \alpha \in \mathbb{C}^p \right\} \subseteq T^*X \times \mathbb{C}^p$. Then

$$BS_G = \pi_2 \left(\overline{W_G}^{T^*Y \times \mathbb{C}^p} \cap V(\pi_1^*(\pi^*(g_1 \cdots g_p))) \right),$$

with π_1, π_2 the projections from $T^*Y \times \mathbb{C}^p$ to the first and second component, $\pi: T^*Y \rightarrow Y$ the natural map.

Linking $Q_{F,\alpha}$, BS_G , and $\text{Trop}(X)$

- Y smooth closed subvariety of \mathbb{C}^P
- G the tuple of coordinate functions restricted to Y
- X the very affine variety $Y \cap (\mathbb{C}^*)^P$
- F the tuple of coordinate functions restricted to X

[vdVS21, Theorem 3.3]

Let $\alpha \in \mathbb{P}^{P-1}$ and $L_\alpha \subseteq \mathbb{C}^P$ the line through the origin corresponding to α . If $Q_{F,\alpha} \cap \mathbb{Z}_{\geq 0}^P \neq \emptyset$, then $L_\alpha \subseteq BS_G$.

[vdVS21, Theorem 3.4]

Assume X is schön and $X^\Sigma \cap \mathcal{O}_\tau$ is connected for all $\tau \in \Sigma$. Then the irreducible components of $S_F \cap \mathbb{P}(BS_G)$ are exactly the hyperplanes τ^\perp for $\tau \subset \mathbb{R}_{\geq 0}^P$ rigid.

Illustration at the flipping the coin example

- X the very affine variety $V(p_0 p_2 - (p_0 + p_1) p_1) \setminus \mathcal{H} \subseteq (\mathbb{C}^*)^3$
- F the tuple of coordinate functions restricted to X
- \overline{X} the closure of X in \mathbb{P}^3

Curve

$$\gamma: t \mapsto \left(\frac{2t^2}{(2t+1)^2}, \frac{2t}{(2t+1)^2}, \frac{1}{(2t+1)^2}, (t:0:1) \right) \in \overline{X} \times \mathbb{P}^2$$

- ▶ $\lim_{t \rightarrow 0} \gamma(t) \in \overline{X} \setminus X \times \{(0:0:1)\}$
- ▶ $v = (2, 1, 0) \in Q_{F, (0:0:1)} \rightsquigarrow Q_{F, (0:0:1)} \cap \mathbb{Z}_{\geq 0}^3 \neq \emptyset$
- ▶ indeed: $\mathbb{R} \cdot (0, 0, 1)$ contained in Bernstein–Sato slopes

Maximum likelihood degree one

- X schön very affine variety with $d_{\text{ML}}(X) = 1$
- Ψ the maximum likelihood estimator
- Σ a fan supported on $\text{Trop}(X)$
- v_τ primitive generator of the ray τ
- \mathcal{O}_τ torus orbit in \mathbb{T}^Σ arising from τ

[vdVS21, Proposition 2.14]

Assume $X^\Sigma \cap \mathcal{O}_\tau$ is connected for all $\tau \in \Sigma$. For τ rigid, let $g_\tau = 0$ be a defining equation of τ^\perp . Then there exist $c_1, \dots, c_p \in \mathbb{C}$ such that

$$t_i \circ \Psi = c_i \cdot \prod_{\tau \text{ rigid}} g_\tau^{(v_\tau)_i}.$$

Moreover:

$$\sum_{\tau \text{ rigid}} v_\tau = 0.$$

Revisiting the coin example

Implicit representation of the statistical model: smooth curve \mathcal{M} in \mathbb{P}^2 defined by

$$f = \det \begin{pmatrix} p_0 & p_1 \\ p_0 + p_1 & p_2 \end{pmatrix} = p_0 p_2 - (p_0 + p_1) p_1.$$

- ▶ X the very affine variety $\mathcal{M} \setminus \{p_0 p_1 p_2 (p_0 + p_1 + p_2) = 0\}$
- ▶ rays in the tropical variety of X are the rows of³

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \\ -2 & -2 & -1 \end{pmatrix} =: \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

- ▶ codimension-one part of S_F : $V(2s_0 + s_1) \cup V(s_1 + s_2) \cup V(2s_0 + 2s_1 + s_2)$
- ▶ Bernstein–Sato ideal of the tuple $(x^2, x(1-x), 1-x)$ on \mathbb{C} :

$$\left\langle \prod_{k=1}^3 (2s_0 + s_1 + k) \cdot \prod_{\ell=1}^2 (s_1 + s_2 + \ell) \right\rangle \triangleleft \mathbb{C}[s_0, s_1, s_2]$$

³computed with Gfan

Thank you very much for your attention!

References I

- [ALSS20] Michael F. Adamer, Andras L. Lőrincz, A.-L. S., and Bernd Sturmfels.
Algebraic analysis of rotation data.
Algebraic Statistics, 11:189–211, 2020.
- [BMM02] Joël Briançon, Philippe Maisonobe, and Michel Merle.
Éventails associés à des fonctions analytiques.
Tr. Mat. Inst. Steklova, 238:70–80, 2002.
- [Bud15] Nero Budur.
Bernstein–Sato ideals and local systems.
Ann. Inst. Fourier (Grenoble), 65(2):549–603, 2015.
- [BvWZ19] Nero Budur, Robin van der Veer, Lei Wu, and Peng Zhou.
Zero loci of Bernstein–Sato ideals.
To appear in *Invent. Math.* arXiv:1907.04010, 2019.
- [DGPS20] Wolfram Decker, Gert-Martin Greuel, Gerhard Pfister, and Hans Schönemann.
SINGULAR 4-1-3—A computer algebra system for polynomial computations.
<http://www.singular.uni-kl.de>, 2020.
- [DLM74] Thomas D. Downs, Jerome Liebman, and Wilma Mackay.
Statistical methods for vectorcardiogram orientations.
Proc. XIth Int. Symp. Vectorcardiography, pages 216–222, 1974.
- [DMS21] Eliana Duarte, Orlando Marigliano, and Bernd Sturmfels.
Discrete Statistical Models with Rational Maximum Likelihood Estimates.
Bernoulli, 27(1):135–154, 02 2021.
- [FK00] Joseph Franecki and Mikhail Kapranov.
The Gauss map and a noncompact Riemann–Roch formula for constructible sheaves on semiabelian varieties.
Duke Math. J., 104(1):171–180, 2000.
- [GLS20] Paul Görlach, Christian Lehn, and A.-L. S.
Algebraic analysis of the hypergeometric function ${}_1F_1$ of a matrix argument.
Beiträge zur Algebra und Geometrie, November 2020.

References II

- [Hac07] Paul Hacking.
The homology of tropical varieties.
Collect. Math., 59:263–273, 2007.
- [HNNT13] Hiroki Hashiguchi, Yasuhide Numata, Nobuki Takayama, and Akimichi Takemura.
The holonomic gradient method for the distribution function of the largest root of a Wishart matrix.
J. Multivariate Anal., 117:296–312, 2013.
- [HS14] June Huh and Bernd Sturmfels.
Likelihood geometry.
In *Combinatorial algebraic geometry*, volume 2108 of *Lecture notes in mathematics*, pages 63–117. Springer, New York, 2014.
- [Huh14] June Huh.
Varieties with maximum likelihood degree one.
J. Algebr. Stat., 5:1–17, 2014.
- [Jen] Anders N. Jensen.
Gfan, a software system for Gröbner fans and tropical varieties.
Available at <http://home.imf.au.dk/jensen/software/gfan/gfan.html>.
- [LMM] Viktor Levandovskyy and Jorge Martín-Morales.
dmodLib: A Singular:PLural library for algorithms for algebraic D -modules.
https://www.singular.uni-kl.de/Manual/4-2-0/sing_537.
- [LQ11] Mark Luxton and Zhenhua Qu.
Some results on tropical compactifications.
Trans. Amer. Math. Soc., 363(9):4853–4876, 2011.
- [Mai16] Philippe Maisonobe.
Filtration relative, l'idéal de Bernstein et ses pentes.
hal-01285562v2, 2016.
- [MS15] Diane Maclagan and Bernd Sturmfels.
Introduction to tropical geometry, volume 161 of *Graduate studies in mathematics*.
American Mathematical Society, Providence, R.I., 2015.

References III

- [NNN⁺11] Hiromasa Nakayama, Kenta Nishiyama, Masayuki Noro, Katsuyoshi Ohara, Tomonari Sei, Nobuki Takayama, and Akimichi Takemura. Holonomic gradient descent and its application to the Fisher–Bingham integral. *Adv. in Appl. Math.*, 47(3):639–658, 2011.
- [SS19] A.-L. S. and Bernd Sturmfels. *D-modules and holonomic functions*. arXiv:1910.01395 [math.AG], 2019.
- [SST⁺13] Tomonari Sei, Hiroki Shibata, Akimichi Takemura, Katsuyoshi Ohara, and Nobuki Takayama. Properties and applications of Fisher distribution on the rotation group. *J. Multivariate Anal.*, 116:440–455, 2013.
- [ST20] Bernd Sturmfels and Simon Telen. Likelihood equations and scattering amplitudes. arXiv:2012.05041 [math.AG], 2020.
- [vdVS21] Robin van der Veer and A.-L. S. Maximum likelihood estimation from a tropical and a Bernstein–Sato perspective. arXiv:2101.03570 [math.AG], 2021.