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.”

Yes, they do!
What is this talk about?

[SvdV21] Robin van der Veer$^1$ 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]

$^1$KU Leuven, Belgium
Maximum Likelihood Estimation

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

$\begin{align*}
\psi(s_0, s_1, s_2) &= \left( \frac{(2s_0 + s_1)^2}{(2s_0 + 2s_1 + s_2)^2}, \frac{(2s_0 + s_1)(s_1 + s_2)}{(2s_0 + 2s_1 + s_2)^2}, \frac{s_1 + s_2}{2s_0 + 2s_1 + s_2} \right) \\
\text{Parametrization of the model: } &\Delta_1 \rightarrow \Delta_2, \ (x_0, x_1) \mapsto (x_0^2, x_0x_1, x_1), \text{ where } x_0, x_1 > 0, \ x_0 + x_1 = 1 \\
\text{Implicitization: } &\mathcal{M} := V(p_0p_2 - (p_0 + p_1)p_1) \subseteq \mathbb{P}^2
\end{align*}$

Figure: Staged tree modeling the discrete statistical experiment [DMS21]
Bernstein–Sato ideals

\[ D = \mathbb{C}[x_1, \ldots, x_n] \langle \partial_1, \ldots, \partial_n \rangle \]

\[ F = (f_1, \ldots, f_p) \in \mathbb{C}[x_1, \ldots, x_n]^p \]

the Weyl algebra, \([\partial_i, x_i] = \partial_i x_i - x_i \partial_i = 1\)
a tuple of polynomials

**Definition**

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

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

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

Observed in [SS19, Example 3.1].

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

\[ ^2 \text{computed with the library dmod.lib in Singular} \]
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 [HNTT13]
  - further study of Muirhead’s $D$-ideal in [GLS21]

**Figure:** A dataset from a study in vectorcardiography [DLM74]
\[
\mathbb{P}^2 \quad \text{homogeneous coordinates} \ (p_0 : p_1 : p_2)
\]

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 \( \ell \)

- given \( (s_0, s_1, s_2) \in \mathbb{N}_0^3 \), \( \ell_{s_0,s_1,s_2}(p_0, p_1, p_2) := \frac{p_0^{s_0} p_1^{s_1} p_2^{s_2}}{(p_0 + p_1 + p_2)^{s_0+s_1+s_2}} \).
- \( \ell \) 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 \( \ell \): zeros of \( \text{dlog} \, \ell \), the logarithmic differential of \( \ell \)
| 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 $\ell$ over $\mathcal{M} \cap \Delta_2$ |

Two important entities

- **Likelihood locus**
  - the set of critical points of $\ell_{u_0, u_1, u_2}$ 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 \((f_1, \ldots, f_p)\) on \( X \)

\( X \hookrightarrow Y \) a smooth compactification with boundary \( E = Y \setminus X \)

\( f^\alpha = f_1^{\alpha_1} \cdots f_p^{\alpha_p} \)

\( E_1, \ldots, E_q \) 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 \( \text{dlog} f^\alpha = \sum_{i=1}^{p} \alpha_i \frac{df_i}{f_i} \in \Gamma(X, \Omega^1_X) \) the logarithmic differential of \( f^\alpha \)

\( S_F \) the critical slopes of \( F \): \( \pi_2(C_F \cap \pi_1^{-1}(E)) \subseteq \mathbb{P}^{p-1} \), \( \pi_1, \pi_2 \) the projections from \( Y \times \mathbb{P}^{p-1} \)
Refined asymptotic behavior or critical points

\( X \)  a smooth variety
\( F \)  a tuple \((f_1, \ldots, f_p)\) of nowhere vanishing regular functions on \( X \)
\( Y \)  a compactification of \( X \)
\( \pi_1 \)  the projection from \( Y \times \mathbb{P}^{p-1} \) to the first factor
\( \Delta \)  the formal disc \( \text{Spec} \, \mathbb{C}[t] \) around 0
\( \Delta^\circ \)  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 \( \gamma: \Delta \to Y \times \mathbb{P}^{p-1} \) 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 $$I \triangleleft \mathbb{C}[t_1^{\pm 1}, \ldots, t_p^{\pm 1}]$$

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

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

Definition

A ray $$\tau$$ is rigid if any small perturbation of the ray changes the initial ideal of $$I$$ w.r.t. the primitive generator $$\nu_\tau \in \mathbb{Z}^P$$ of $$\tau$$.
Tropical compactifications

\[ X \hookrightarrow (\mathbb{C}^*)^p \quad \text{a very affine variety} \]
\[ \Sigma \subseteq \mathbb{R}^p \quad \text{a fan} \]
\[ \mathbb{T}^\Sigma \quad \text{toric variety of } \Sigma \]

Definition ([Hac07])

\( X \) is \textbf{schön} iff there exists a fan structure \( \Sigma \) on \( \text{Trop}(X) \) s.t. the closure \( X^\Sigma \) of \( X \) in \( \mathbb{T}^\Sigma \) 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 \( \Sigma \) s.t. \( X^\Sigma \) is a smooth SNC compactification of \( X \).

Such \( X^\Sigma \) is a \textbf{tropical compactification} of \( X \).
Codimension-one components of $S_F$

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

[SvdV21, 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 $S_F$.

[SvdV21, 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 \((g_1, \ldots, g_p)\) on \( Y \)

**Definition**

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

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

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

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

with $\pi_1, \pi_2$ the projections from $T^*Y \times \mathbb{C}^p$ to the first and second component, $\pi: T^*Y \to Y$ 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 $X$

[SvdV21, Theorem 3.3]

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

[SvdV21, Theorem 3.4]

Assume $X$ is schön and $X^\Sigma \cap O_\tau$ is connected for all $\tau \in \Sigma$. Then the irreducible components of $S_F \cap \mathbb{P}(BS_G)$ are exactly the hyperplanes $\mathbb{P}(\tau^\perp)$ for $\tau \subset \mathbb{R}^p_{\geq 0}$ rigid.
Illustration at the flipping the coin example

\[ X \text{ the very affine variety } V(p_0 p_2 - (p_0 + p_1)p_1) \setminus \mathcal{H} \subseteq (\mathbb{C}^*)^3 \]

\[ F \text{ the tuple of coordinate functions restricted to } X \]

\[ \overline{X} \text{ the closure of } X \text{ in } \mathbb{P}^3 \]

Curve

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

\[ \lim_{t \to 0} \gamma(t) \in \overline{X} \setminus X \times \{(0 : 0 : 1)\} \]

\[ \nu = (2, 1, 0) \in Q_{F,(0:0:1)} \sim Q_{F,(0:0:1)} \cap \mathbb{Z}_{\geq 0} \neq \emptyset \]

\[ \text{indeed: } \mathbb{R} \cdot (0, 0, 1) \text{ contained in Bernstein–Sato slopes} \]
Maximum likelihood degree one

\(X\) schön very affine variety with \(d_{\text{ML}}(X) = 1\)

\(\Psi\) the maximum likelihood estimator

\(\Sigma\) a fan supported on \(\text{Trop}(X)\)

\(v_\tau\) primitive generator of the ray \(\tau\)

\(O_\tau\) torus orbit in \(\mathbb{T}^{\Sigma}\) arising from \(\tau\)

[SvdV21, Proposition 2.14]

Assume \(X^\Sigma \cap O_\tau\) is connected for all \(\tau \in \Sigma\). For \(\tau\) rigid, let \(g_\tau = 0\) be a defining equation of \(\tau^\perp\). Then there exist \(c_1, \ldots, c_p \in \mathbb{C}\) 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 of

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

- 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 $\mathbb{C}$:

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

$^3$computed with Gfan
Thank you very much for your attention!
References

[ALSS20] Michael F. Adamer, Andras L. Lörincz, Anna-Laura Sattelberger, and Bernd Sturmfels. Algebraic Analysis of Rotation Data. *Algebraic Statistics*, 11(2):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.

[BvWZ] Nero Budur, Robin van der Veer, Lei Wu, and Peng Zhou. Zero loci of Bernstein–Sato ideals. arXiv:1907.04010 [math.AG], 2019. To appear in *Invent. Math*.

[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 Lieberman, and Wilma Mackay. Statistical methods for vectorcardiogram orientations. *Proc. Xlth 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.

[GLS21] Paul Görlach, Christian Lehn, and Anna-Laura Sattelberger. Algebraic Analysis of the Hypergeometric Function $\binom{1}{1}$ of a Matrix Argument. *Beitr. Algebra Geom.*, 62:397–423, 2021.
References II

[Hac07] Paul Hacking. The homology of tropical varieties. Collect. Math., 59:263–273, 2007.

[HNTT13] 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. dmod_lib: 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.
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.

Anna-Laura Sattelberger and Bernd Sturmfels.
D-Modules and Holonomic Functions.
arXiv:1910.01395 [math.AG], 2019.

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.

Bernd Sturmfels and Simon Telen.
Likelihood Equations and Scattering Amplitudes.
arXiv:2012.05041 [math.AG], 2020.

Anna-Laura Sattelberger and Robin van der Veer.
Maximum Likelihood Estimation from a Tropical and a Bernstein–Sato Perspective.
arXiv:2101.03570 [math.AG], 2021.