Bayesian Integrals on Toric Varieties

Michael Borinsky, *Anna-Laura Sattelberger*, Bernd Sturmfels, Simon Telen

Hypergeometric School 2023
Kobe University

August 18, 2023

Aim

Computation of marginal likelihood integrals

$$\boxed{\int_{X_{>0}} p_0^{u_0} p_1^{u_1} \cdots p_m^{u_m} \ \Omega_X^{\text{prior}}}$$

for statistical models that are parameterized by a toric variety.

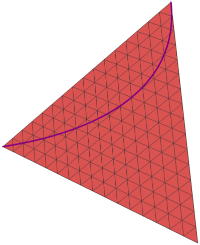
How?

Tropical sampling algorithms.

Outline

1. Toric varieties and statistical models
2. Toric varieties as probability spaces
3. Tropical sampling

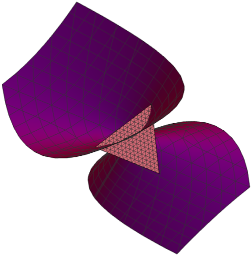
Michael Borinsky, Anna-Laura Sattelberger, Bernd Sturmfels, and Simon Telen. Bayesian Integrals on Toric Varieties. *SIAM J. Appl. Algebra Geom.*, 7:77–103, 2023.



Definition

A **discrete statistical model** taking $m+1$ states is a parameterized subset of the probability m-simplex

$$\Delta_m = \left\{ (p_0, \dots, p_m) \mid p_i \in (0,1),\ \sum_{i=0}^{m} p_i = 1 \right\}.$$



Definition.

An algebraic variety X is **toric** if it contains a dense algebraic torus whose action on itself extends to X.

Normal toric varieties...

... of dimension n are encoded by complete fans in $\mathbb{R}^n$.

Example: a coin model

$$\begin{array}{lll} X &= \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 & \text{homogeneous coordinates } (x_0 : x_1), (s_0 : s_1), (t_0 : t_1) \\ X &= X_\Sigma & \Sigma \text{ the inner normal fan of } [0,1]^3 \\ X_{>0} &\widehat{=} (0,1)^3 & \text{the positive part of } X \end{array}$$

Model: image of $X_{>0} \to \Delta_m$, $(x, s, t) \mapsto (p_\ell(x, s, t))_{\ell=0,\dots,m}$,
$x = x_0,\ x_1 = 1 - x,\ s = s_0,\ s_1 = 1 - s,\ t = t_0,\ t_1 = 1 - t$

$$p_\ell = \binom{m}{\ell} x\, s^\ell (1-s)^{m-\ell} + \binom{m}{\ell} (1-x) t^\ell (1-t)^{m-\ell}, \quad \ell = 0, 1, \dots, m.$$

probability for observing ℓ times head

Marginal likelihood integral

For uniform prior on $(0,1)^3$, data $u = (u_0, \dots, u_m)$, the **marginal likelihood integral** is

$$\mathcal{I}_u = \int_{X_{>0}} \underbrace{p_0^{u_0} \cdots p_m^{u_m}}_{= L_u \text{ likelihood fct.}} \underbrace{\Omega_X^{\text{unif}}}_{\text{prior distribution}} .$$

$u_+ = u_0 + \cdots + u_m$ many repetitions

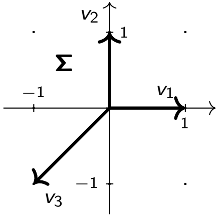
Example: complex projective plane

Σ the inner normal fan of Δ_2, $V = (v_1|v_2|v_3) = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$

$$X_\Sigma = \mathbb{P}^2_{\mathbb{C}} = \left(\mathbb{C}^3\right)^* / \mathbb{C}^* = \left(\mathbb{A}^3_{\mathbb{C}} \setminus \mathcal{V}(x_0, x_1, x_2)\right) / \mathbb{G}^1_m$$

Homogeneous coordinates: $(x_0 : x_1 : x_2)$ — Cox coordinates

Three affine charts: $\{x_i \neq 0\}$ — one for each maximal cone



In general

Σ a complete fan in $\mathbb{R}^n$ — e.g. the inner normal fan of a polytope P

- $V = (v_1| \cdots |v_k)$ — columns: primitive ray generators of the $\rho_i \in \Sigma(1)$
- $\mathrm{Cl}(X) = \mathbb{Z}^k / \operatorname{im}(V^\top)$ — divisor class group of X
- $G = \mathrm{Hom}(\mathrm{Cl}(X), \mathbb{C}^*)$ — the characters of $\mathrm{Cl}(X)$
- $S = \mathbb{C}[x_1, \ldots, x_k] = \bigoplus_{\gamma \in \mathrm{Cl}(X)} S_\gamma$ — Cox ring
- $B = \langle \prod_{\rho \notin \sigma} x_\rho \mid \sigma \in \Sigma(n) \rangle \subset S$ — the irrelevant ideal
- $X_\Sigma = \left(\mathbb{C}^k \setminus \mathcal{V}(B)\right) / G$ — the **toric variety** of Σ

Simon Telen. Introduction to Toric Geometry. Preprint arXiv:2203.01690, 2022.

Positive toric varieties

Setup

Σ	the inner normal fan of a polytope P
$X = X_\Sigma$	the toric variety of Σ
P°	the interior of P

Positive part of $X_\Sigma = \left(\mathbb{C}^k \setminus \mathcal{V}(B)\right)/G$

- $\pi\colon \mathbb{C}^k \setminus \mathcal{V}(B) \longrightarrow \left(\mathbb{C}^k \setminus \mathcal{V}(B)\right)/G$ the projection
- $\boxed{\pi\left(\mathbb{R}^k_{>0}\right) =: X_{>0}}$ the **positive part** of X_Σ $\qquad$ $X_{\geq 0}$ its Euclidean closure

Algebraic moment map

One identifies $X_{>0}$ and P° via the **moment map**

$$X_{>0} \overset{\phi}{\simeq} \mathbb{R}^n_{>0} \overset{\varphi}{\simeq} P^\circ,$$

with φ the affine moment map

$$\varphi(t) = \sum_{a\in\mathcal{V}(P)} \frac{c_a t^a}{q(t)} \cdot a\,, \qquad q = \sum_{a\in\mathcal{V}(P)} c_a t^a \in \mathbb{R}_{>0}[t_1^{\pm 1}, \ldots, t_n^{\pm 1}]\,.$$

$X_{>0}$ as probability space

Definition

The **canonical form** of $(X, X_{\geq 0})$ is the meromorphic differential n-form

$$\Omega_X = \sum_{I \subset \Sigma(1),\, |I|=n} \det(V_I) \bigwedge_{\rho \in I} \frac{\mathrm{d}x_\rho}{x_\rho}$$

on X. The pair $(X, X_{\geq 0})$ is a positive geometry.

Proposition

The pullback of $\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_n$ on P° under the moment map $X_{>0} \to P^\circ$ is a positive rational function r times Ω_X. We obtain $r(x)$ from $|\det|$ of the **toric Hessian** of $\log(q(t))$

$$H(t) = (\theta_i\theta_j \bullet \log(q(t)))_{i,j} \qquad \theta_i = t_i \partial_{t_i}$$

by replacing $t_1, \ldots, t_n$ with Laurent monomials in $x_1, \ldots, x_k$ given by the rows of V.

Observation: Scaled by a rational function $\frac{f}{g}$, Ω_X gives a probability measure on $X_{>0}$!

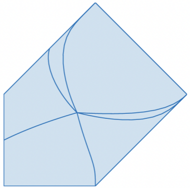
Integrals of interest: $\mathcal{I}_{f,g} = \int_{X_{>0}} \frac{f}{g}\, \Omega_X$ $\quad f, g \in S$ homogeneous of the same degree

Nima Arkani-Hamed, Yuntao Bai, and Thomas Lam. Positive Geometries and Canonical Forms. *J. High Energ. Phys.*, 39(2017), 2017.

Definition

The **tropical approximation** of $f \in \mathbb{C}[x_1, \ldots, x_k]$ is the piecewise monomial function

$$f^{\mathrm{tr}} \colon \mathbb{R}^k_{>0} \longrightarrow \mathbb{R}_{>0}, \quad x \mapsto \max_{\ell \in \mathrm{supp}(f)} x^{\ell}.$$



Proposition

Let $\mathcal{F}$ be a simplicial refinement of the normal fan of $\mathcal{N}(f) + \mathcal{N}(g)$. Then

$$\mathcal{I}_{f,g} = \int_{X_{>0}} \frac{f}{g}\,\Omega_X = \sum_{\sigma \in \mathcal{F}(n)} \int_{\mathrm{Exp}(\sigma)} \frac{f^{\mathrm{tr}}}{g^{\mathrm{tr}}} \underbrace{\frac{f \cdot g^{\mathrm{tr}}}{g \cdot f^{\mathrm{tr}}}}_{=:\, h,\ \text{positive and bounded on } X_{>0}} \Omega_X = \sum_{\sigma \in \mathcal{F}(n)} \mathcal{I}_\sigma \quad \textbf{sector integrals}$$

◇ $\mathrm{Exp} \colon \mathbb{R}^k / K \to X_{>0}, \ [y_1, \ldots, y_k] \mapsto \pi(e^y)$ ◇ parameterization $x^\sigma \colon [0,1]^n \to \mathrm{Exp}(\sigma)$

Tropical detour

Also the tropical integral $\mathcal{I}^{\mathrm{tr}}_{f,g} = \int_{X_{>0}} f^{\mathrm{tr}} / g^{\mathrm{tr}}\, \Omega_X$ decomposes as $\mathcal{I}^{\mathrm{tr}} = \sum_{\sigma \in \mathcal{F}(n)} \mathcal{I}^{\mathrm{tr}}_\sigma$. Each tropical sector integral $\mathcal{I}^{\mathrm{tr}}_\sigma$ is an integral over a monomial encoded by data of $\mathcal{F}$!

$$\mathcal{I}^{\mathrm{tr}}_\sigma = \int_{\mathrm{Exp}(\sigma)} x^{-(\nu_g - \nu_f)}\, \Omega_X$$

Toric data of f, g

Theorem

Suppose that the Newton polytope of g is n-dimensional and contains that of the numerators f in its relative interior. Then the integral $\int_{X_{>0}} f/g\ \Omega_X$ converges.

Proposition

Let $\mathcal{F}$ a simplicial refinement of $\mathcal{N}(f) + \mathcal{N}(g)$. Let σ be a cone of $\mathcal{F}$, ν_f and ν_g corresponding faces of $\mathcal{N}(f)$ and $\mathcal{N}(g)$. Then:

$$\boxed{\frac{f^{\mathrm{tr}(x)}}{g^{\mathrm{tr}}(x)} = x^{-(\nu_g - \nu_f)}} \quad \text{for all } x \in \mathbb{R}^k \text{ such that } \pi(x) \in \mathrm{Exp}(\sigma)\,.$$

Then

$$\mathcal{I}^{\mathrm{tr}} = \sum_{\sigma \in \mathcal{F}(n)} \mathcal{I}^{\mathrm{tr}}_\sigma \qquad \text{where} \qquad \mathcal{I}^{\mathrm{tr}}_\sigma = \int_{\mathrm{Exp}(\sigma)} \frac{f^{\mathrm{tr}}}{g^{\mathrm{tr}}} \Omega_X = \int_{\mathrm{Exp}(\sigma)} x^{-(\nu_g - \nu_f)}\, \Omega_X\,.$$

Write $\mathrm{im}(V^\top)$ as $\ker(W)$, $W = (w_1 | \cdots | w_n)$. The tropical sector integral is equal to

$$\boxed{\mathcal{I}^{\mathrm{tr}}_\sigma = \frac{\det(VW)}{\prod_{\ell=1}^n w_\ell \cdot (\nu_g - \nu_f)}}\,.$$

Sampling from $\left(X_{>0}, d_{f,g}^{(\mathrm{tr})}\right)$

$$\mu_{f,g} = \underbrace{\frac{1}{\mathcal{I}_{f,g}} \cdot \frac{f}{g}}_{\text{density } d_{f,g}} \Omega_X \quad \text{and} \quad \mu_{f,g}^{\mathrm{tr}} = \underbrace{\frac{1}{\mathcal{I}_{f,g}^{\mathrm{tr}}} \cdot \frac{f^{\mathrm{tr}}}{g^{\mathrm{tr}}}}_{\textbf{tropical}\text{ density } d_{f,g}^{\mathrm{tr}}} \Omega_X$$

are probability measures on $X_{>0}$!

Sampling from the tropical density

Input: $\mathcal{F}$, $\mathcal{I}_\sigma^{tr}$, *and* $\mathcal{I}^{tr}$.

Step 1. *Draw an n-dimensional cone* σ *from* $\mathcal{F}(n)$ *with probability* $\mathcal{I}_\sigma^{tr}/\mathcal{I}^{tr}$.
Step 2. *Draw a sample* q *from the unit hypercube* $[0,1]^n$ *using the uniform distribution.*
Step 3. *Compute* $x^\sigma(q) \in X_{>0}$.

Output: *The element* $x^\sigma(q) \in X_{>0}$*, a sample from* $\left(X_{>0}, d_{f,g}^{tr}\right)$.

Sampling from $d_{f,g}$ *via rejection sampling!*

Proposition

Let $x^{(1)}, \ldots, x^{(N)}$ be tropical samples from $X_{>0}$. Then

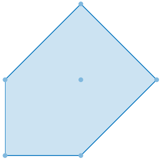
$$h(x) = \frac{f(x) \cdot g^{\mathrm{tr}}(x)}{g(x) \cdot f^{\mathrm{tr}}(x)}$$

$$\boxed{\mathcal{I}_{f,g} \approx \mathcal{I}_N = \frac{\mathcal{I}_{f,g}^{\mathrm{tr}}}{N} \cdot \sum_{i=1}^{N} h\left(x^{(i)}\right).}$$

Toric polytope models $c = (c_0, \ldots, c_m),\ c_i \in \mathbb{R}_{>0}$

- $Z = c_0 x^{a_0} + c_1 x^{a_1} + \cdots + c_m x^{a_m} \in S$ homogeneous of degree $\gamma \in \mathrm{Cl}(X)$
  a_i lattice points of P
- $p_i = c_i x^{a_i}/Z,\ i = 0, \ldots, m$, are positive on $X_{>0}$, $\sum_{i=0}^m p_i = 1$
  statistical model: image of resulting map $X_{>0} \to \Delta_m$

Bayes' factor for toric pentagon model



Prior: distribution $\mu_{f,g}$ arising from uniform distribution on P°
Data: $u = (u_0, \ldots, u_5) = (1, 2, 4, 8, 16, 32)$ $u_+ = \sum u_i = 63$

Competing models: toric models $\mathcal{M}_c$ for

$$c^{(1)} = (2, 3, 5, 7, 11, 13) \quad \text{and} \quad c^{(2)} = (32, 16, 8, 4, 2, 1)\,.$$

Marginal likelihood integrals:

$$\mathcal{I}_u^{(i)} = \int_{X_{>0}} \underbrace{L_u^{(i)}(x)}_{= p_0^{u_0} \cdots p_5^{u_5} \text{ likelihood function}} \mu_{f,g}, \qquad i = 1, 2\,.$$

Bayes' factor: $\boxed{K = \mathcal{I}_u^{(1)}/\mathcal{I}_u^{(2)} \approx 21.06.}$ $\mathcal{M}_{c^{(1)}}$ is a better fit for the data than $\mathcal{M}_{c^{(2)}}$!

Sampling from $(X_{>0}, d_{f,g})$

Setup

- d_1 and d_2 two densities on the same space with the same differential form
  e.g. on $(X_{>0}, \Omega_X)$
- suppose it is hard to sample from d_1, but easy to sample from d_2
- suppose there exists $C \geq 1$ such that $d_1(x)/d_2(x) \leq C$ for all x

Rejection sampling

Step 1. *Draw a sample* $x \in X$ *using* d_2*, and* $\xi \in [0, C]$ *with the uniform distribution.*
Step 2. *If* $\xi < d_1(x)/d_2(x)$*, accept* x *as a sample. Otherwise, reject* x*.*

Output: A sample from $d_2(x) \cdot d_1(x)/d_2(x)$, i.e., $d_1(x)$.

Proposition

Suppose that $f = \sum_{\ell \in \mathrm{supp}(f)} f_\ell x^\ell$ has positive coefficients. Set $C_1 = \min_{\ell \in \mathrm{supp}(f)} f_\ell$ and $C_2 = \sum_{\ell \in \mathrm{supp}(f)} f_\ell$. Then

$$\boxed{0 < C_1 \leq \frac{f(x)}{f^{\mathrm{tr}}(x)} \leq C_2 < \infty \qquad \text{for all} \quad x \in X_{>0}\,.}$$

Sampling from $d_{f,g}$ via rejection sampling with $d^{\mathrm{tr}}_{f,g}$!

Conclusion

In a nutshell

1. Statistical models parameterized by toric varieties occur naturally.
2. Positive toric varieties are probability spaces. positive geometries
3. Bayesian inference via tropical methods. $\int_{X_{>0}} L_u \, \Omega_X^{\text{prior}}, \quad \int_{X_{>0}} f/g \, \Omega_X$

Supplementary material

- code in `Julia` available at: https://mathrepo.mis.mpg.de/BayesianIntegrals
- painting inspired by the pentagon model: https://alsattelberger.de/painting/

Thank you for your attention!

Let $h(x) = \frac{f(x) \cdot g^{\mathrm{tr}}(x)}{g(x) \cdot f^{\mathrm{tr}}(x)}$. Then

$$M_1 \leq h(x) \leq M_2 \qquad \text{for all} \quad x \in X_{>0} \,,$$

where

$$M_1 = \frac{\min_{\ell \in \mathrm{supp}(f)} f_\ell}{\sum_{\ell \in \mathrm{supp}(g)} g_\ell} \qquad \text{and} \qquad M_2 = \frac{\sum_{\ell \in \mathrm{supp}(f)} f_\ell}{\min_{\ell \in \mathrm{supp}(g)} g_\ell} \,.$$

Proposition

Let $x^{(1)}, \ldots, x^{(N)}$ be tropical samples from $X_{>0}$. Then

$$\boxed{\mathcal{I}_{f,g} \approx \mathcal{I}_N = \frac{\mathcal{I}^{\mathrm{tr}}_{f,g}}{N} \cdot \sum_{i=1}^{N} h\left(x^{(i)}\right) \,.}$$

Proposition

The standard deviation of the approximation above satisfies

$$\boxed{\sqrt{\mathbb{E}\left[(\mathcal{I} - \mathcal{I}_N)^2\right]} \leq \mathcal{I}^{\mathrm{tr}} \cdot \sqrt{\frac{M_2^2 - M_1^2}{N}} \,.}$$

The Wachspress model

$P \subset \mathbb{R}^n$ a polytope, Σ its inner normal fan, $V = (v_1|\cdots|v_k)$

Inequality representation of P

$$\boxed{P \;=\; \{y \in \mathbb{R}^n \,|\, \langle v_i, y\rangle + \alpha_i \geq 0,\;\; i = 1, 2, \ldots, k\}}$$

with $\alpha_1, \ldots, \alpha_k \in \mathbb{Z}_{>0}$. The vertices q_I of P are indexed by cones $I \in \Sigma(n)$: the vertex $q_I \in \mathbb{Z}^n$ is the unique solution of $\langle v_i, y\rangle = -\alpha_i$ for $i \in I$.

Definitions

The **adjoint** of P is the polynomial in variables $y_1, \ldots, y_n$

$$\boxed{A \;=\; \sum_{I \in \Sigma(n)} |\det(\widetilde{V}_I)| \cdot \prod_{i \notin I} \left(1 + \frac{1}{\alpha_i}\langle v_i, y\rangle\right).}$$

The **Wachspress model** of P is the image of $P \to \Delta_m,\;\; y \mapsto (p_I(y))_{I \in \Sigma(n)}$ with

$$p_I(y) \;=\; \frac{|\det\,(\widetilde{V}_I)|}{A(y)} \cdot \prod_{I \in \Sigma(n)} \left(1 + \frac{1}{\alpha_i}\langle v_i, y\rangle\right).$$