SINGULAR LEARNING THEORY

Part II: Real Log Canonical Thresholds

Shaowei Lin
(Institute for Infocomm Research, Singapore)

21-25 May 2013
Motivic Invariants and Singularities Thematic Program
Notre Dame University

Integral Asymptotics

A Coin Toss Integral

For large N, approximate

$$Z(N) \;=\; \int_{[0,1]^2} (1 - x^2y^2)^{N/2}\, dxdy.$$

- Write $Z(N)$ as $\int e^{-Nf(x,y)}dxdy$ where

$$f(x,y) = -\frac{1}{2}\log(1 - x^2y^2).$$

- Can we use the Gaussian integral

$$\int_{\mathbb{R}^d} e^{-\frac{N}{2}(\omega_1^2+\cdots+\omega_d^2)}d\omega = \left(\frac{2\pi}{N}\right)^{d/2}$$

by finding a suitable change of coordinates for x, y?

Laplace Approximation

Ω small nbhd of origin, $f : \Omega \to \mathbb{R}$ analytic function with unique minimum $f(0)$ at origin, $\partial^2 f$ Hessian of f. If $\det \partial^2 f(0) \neq 0$,

$$Z(N) = \int_\Omega e^{-Nf(\omega)} d\omega \approx e^{-Nf(0)} \cdot \sqrt{\frac{(2\pi)^d}{\det \partial^2 f(0)}} \cdot N^{-d/2}.$$

- e.g. Bayesian Information Criterion (BIC)

$$-\log Z(N) \approx \left(-\sum_{i=1}^{N} \log q^*(X_i) \right) + \frac{d}{2} \log N$$

- e.g. Stirling's approximation

$$N! = N^{N+1} \int_0^\infty e^{-N(x-\log x)}\, dx \approx N^{N+1} e^{-N} \sqrt{\frac{2\pi}{N}}$$

However, we cannot apply the Laplace approximation to our example because $\det \partial^2 f(0) = 0$.

Real Log Canonical Threshold

Asymptotic theory (Arnol'd·Guseĭn-Zade·Varchenko, 1985) states that for a Laplace integral,

$$Z(N) = \int_\Omega e^{-Nf(\omega)}\varphi(\omega)d\omega \approx e^{-Nf^*} \cdot CN^{-\lambda}(\log N)^{\theta-1}$$

asymptotically as $N \to \infty$ for some positive constants $C \in \mathbb{R}, \lambda \in \mathbb{Q}, \theta \in \mathbb{Z}$ and where $f^* = \min_{\omega\in\Omega} f(\omega)$.

The pair (λ, θ) is the *real log canonical threshold* of $f(\omega)$ with respect to the measure $\varphi(\omega)d\omega$.

Upper bound (trivial) $\lambda \leq \frac{d}{2}$

Upper bound (Watanabe) $\lambda \leq \frac{1}{2}($ codim of minimum locus of f $)$

Geometry of the Integral

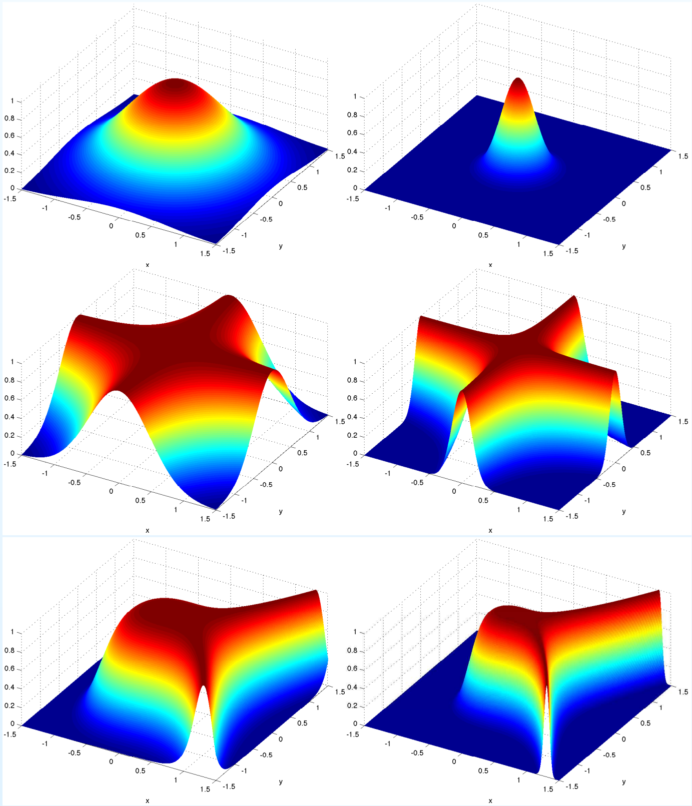
$$Z(N) = \int_{\Omega} e^{-Nf(\omega)} \varphi(\omega) d\omega \approx e^{-Nf^*} \cdot CN^{-\lambda} (\log N)^{\theta - 1}$$

Many integrals in statistics, physics and information theory can be written in the form above. As $N \to \infty$, the asymptotic behavior of the integral depends on the *minimum locus* of $f(\omega)$.

$$f(x, y) = x^2 + y^2$$
$$(\lambda, \theta) = (1, 1)$$

$$f(x, y) = (xy)^2$$
$$(\lambda, \theta) = (\tfrac{1}{2}, 2)$$

$$f(x, y) = (y^2 - x^3)^2$$
$$(\lambda, \theta) = (\tfrac{5}{12}, 1)$$



Monomial Functions

Notation: $\omega^\kappa = \omega_1^{\kappa_1} \cdots \omega_d^{\kappa_d}$.

Asymptotic theory of Arnol'd, Guseĭn-Zade and Varchenko (1974).

Theorem (AGV). Given $\kappa, \tau \in \mathbb{Z}^d_{\geq 0}$,

$$Z(N) \;=\; \int_{\mathbb{R}^d_{\geq 0}} e^{-N\omega^\kappa} \omega^\tau d\omega \;\approx\; CN^{-\lambda} (\log N)^{\theta - 1}$$

where C is a constant,

$$\lambda = \min_i \frac{\tau_i + 1}{\kappa_i},$$

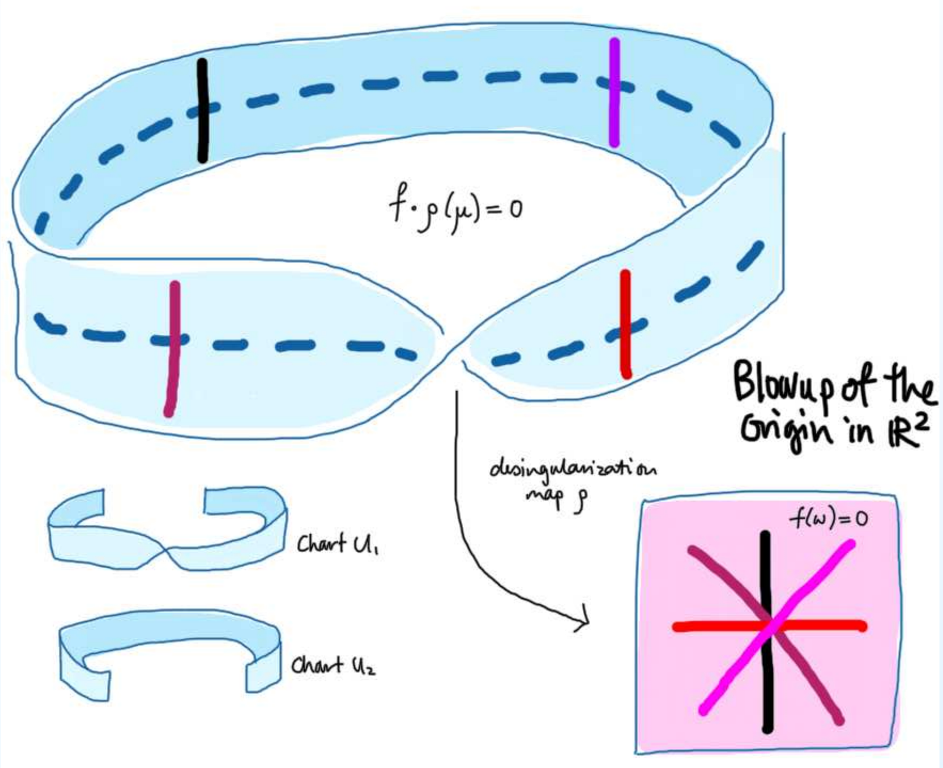
$\theta =$ number of times minimum is attained.

Proof idea: Zeta functions $\zeta(z)$ and state density functions $v(t)$.

$$\zeta(z) = \int_\Omega |f(\omega)|^{-z} \varphi(\omega) d\omega, \quad v(t) = \frac{d}{dt} \int_{|f(\omega)|<t} \varphi(\omega) d\omega.$$

Desingularizations

To resolve a singularity is to find a change of variables so that after the transformation, the singularities are "nice" intersections.



A famous deep result of Hironaka (1964) says that every variety has a resolution of singularities (also known as a *desingularization*).

Desingularizations

Let $\Omega \subset \mathbb{R}^d$ and $f : \Omega \to \mathbb{R}$ real analytic function.

- We say $\rho : U \to \Omega$ *desingularizes* f if

  1. U is a d-dimensional real analytic manifold covered by coordinate patches $U_1, \ldots, U_s$ ($\simeq$ subsets of $\mathbb{R}^d$).

  2. ρ is a proper real analytic map that is an isomorphism onto the subset $\{\omega \in \Omega : f(\omega) \neq 0\}$.

  3. For each restriction $\rho : U_i \to \Omega$,

  $$f \circ \rho(\mu) = a(\mu)\mu^{\kappa}, \quad \det \partial\rho(\mu) = b(\mu)\mu^{\tau}$$

  where $a(\mu)$ and $b(\mu)$ are nonzero on U_i.

- The preimage $\{\mu : f \circ \rho(\mu) = 0\}$ of the variety $\{\omega : f(\omega) = 0\}$ has **simple normal crossings**.
  This preimage is also called the **transform**.

Algorithm for Computing RLCTs

- We know how to find RLCTs of *monomial functions* (AGV, 1985).

$$\int_\Omega e^{-Na(\mu)\mu^\kappa} b(\mu)\mu^\tau d\mu \approx CN^{-\lambda}(\log N)^{\theta-1}$$

  where $\lambda = \min_i \frac{\tau_i+1}{\kappa_i}$, $\theta = |\{i : \frac{\tau_i+1}{\kappa_i} = \lambda\}|$.

- To compute the RLCT of any function $f(\omega)$:
  1. Find minimum f^* of f over Ω.
  2. Find a desingularization ρ for $f - f^*$.
  3. Use AGV Theorem to find (λ_i, θ_i) on each patch U_i.
  4. $\lambda = \min\{\lambda_i\}$, $\theta = \max\{\theta_i : \lambda_i = \lambda\}$.

- The difficult part is finding a desingularization, e.g (Bravo·Encinas·Villamayor, 2005).

Higher Order Asymptotics

If we are able to desingularize $f(x,y) = -\frac{1}{2}\log(1-x^2y^2)$, the higher order asymptotics of $Z(N)$ can also be derived.

$$\begin{aligned}
&\sqrt{\frac{\pi}{8}}N^{-\frac{1}{2}}\log N && -\sqrt{\frac{\pi}{8}}\left(\frac{1}{\log 2} - 2\log 2 - \gamma\right)N^{-\frac{1}{2}} \\
&-\frac{1}{4}N^{-1}\log N && +\frac{1}{4}\left(\frac{1}{\log 2} + 1 - \gamma\right)N^{-1} \\
&-\frac{\sqrt{2\pi}}{128}N^{-\frac{3}{2}}\log N && +\frac{\sqrt{2\pi}}{128}\left(\frac{1}{\log 2} - 2\log 2 - \frac{10}{3} - \gamma\right)N^{-\frac{3}{2}} \\
& && -\frac{1}{24}N^{-2} \quad + \cdots
\end{aligned}$$

Euler-Mascheroni constant

$$\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \log n\right) \approx 0.5772156649.$$

Singular Learning

Sumio Watanabe

Integral Asymptotics

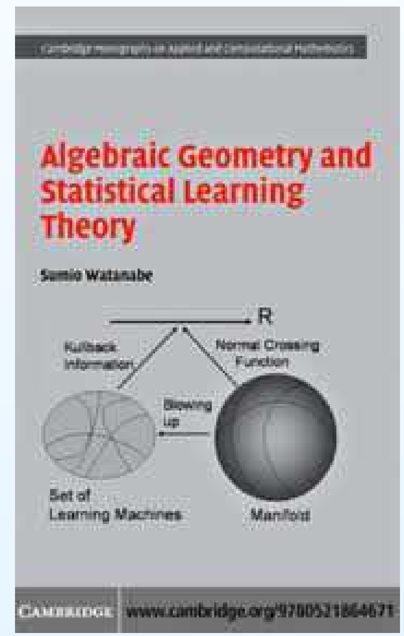
Singular Learning

- Sumio Watanabe
- Statistical Model
- Learning Coefficient
- Geometry
- Standard Form
- Fiber Ideals
- Examples

RLCTs

Computations

Many models used in machine learning are *singular*
e.g. normal mixtures, neural networks, hidden markov models,
but their asymptotic behavior is poorly understood.



In 1998, Sumio Watanabe discovered how to solve this problem
using Hironaka's theorem on the resolution of singularities.
Algebraic geometry is essential in the analysis of singular models.

Statistical Model

X random variable with state space $\mathbb{R}^k$

Δ space of probability distributions on $\mathbb{R}^k$

$\mathcal{M} \subset \Delta$ statistical model

Ω parameter space

$p(x|\omega)$ distribution at $\omega \in \Omega$

$\varphi(\omega)d\omega$ prior distribution on Ω

$X_1, \ldots, X_N$ sample of X

$q \in \mathcal{M}$ *true distribution* of X

Log likelihood ratio
$$K_N(\omega) = \frac{1}{N}\sum_{i=1}^{N} \log \frac{q(X_i)}{p(X_i|\omega)}$$

Kullback-Leibler function
$$K(\omega) = \int_{\mathbb{R}^k} q(x) \log \frac{q(x)}{p(x|\omega)} dx$$

Likelihood integral
$$Z_N = \int_{\Omega} \prod_{i=1}^{N} p(X_i|\omega)\, \varphi(\omega) d\omega$$

Learning Coefficient

Define *empirical entropy* $S_N = -\frac{1}{N}\sum_{i=1}^{N} \log q(X_i)$.
Then, we can rewrite the likelihood integral as

$$Z_N = e^{-NS_N} \int_\Omega e^{-NK_N(\omega)}\varphi(\omega)d\omega.$$

Convergence of stochastic complexity (Watanabe)

The *stochastic complexity* has the asymptotic expansion

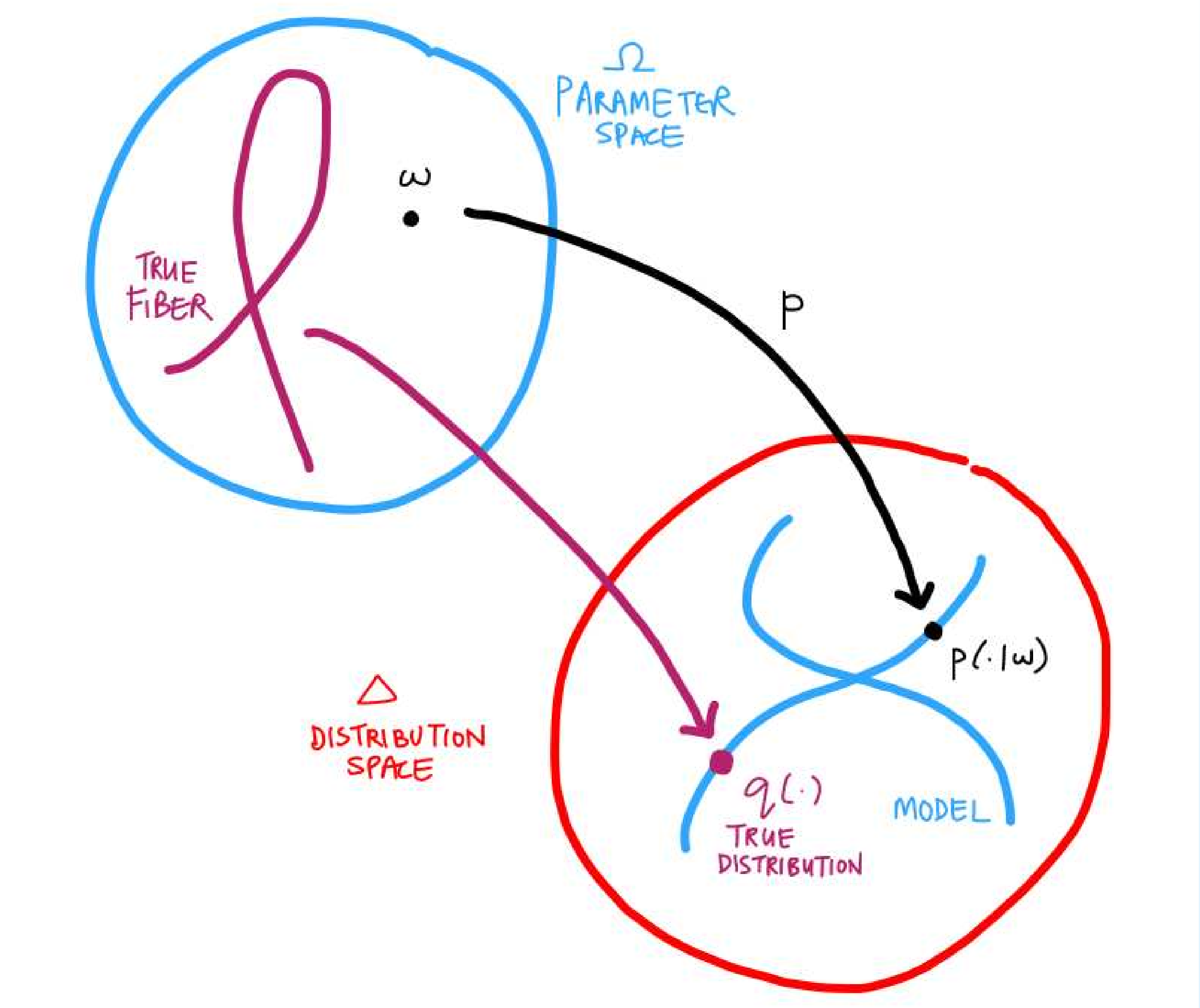
$$-\log Z_N = NS_N + \lambda_q \log N - (\theta_q - 1)\log\log N + F_N^R$$

where F_N^R converges in law to a random variable. Moreover, λ_q, θ_q are asymptotic coefficients of the deterministic integral

$$Z(N) = \int_\Omega e^{-NK(\omega)}\varphi(\omega)d\omega \;\approx\; CN^{-\lambda_q}(\log N)^{\theta_q - 1}.$$

Think of this as *generalized BIC* for singular models.
λ_q, θ_q *learning coefficient* (and its *order*) of the model $\mathcal{M}$ at q.

Geometry of Singular Models



Standard Form of Log Likelihood Ratio

Define *log likelihood ratio*. Note that its expectation is $K(\omega)$.

$$K_N(\omega) = \frac{1}{N}\sum_{i=1}^N \log \frac{q(X_i)}{p(X_i|\omega)}.$$

Standard Form of Log Likelihood Ratio (Watanabe)

Suppose $\rho : \mathscr{M} \to \Omega$ desingularizes $K(\omega)$. Then,

$$K_N \circ \rho(\mu) = \mu^{2\kappa} - \frac{1}{\sqrt{N}}\mu^{\kappa}\xi_N(\mu)$$

where $\xi_N(\mu)$ converges in law to a Gaussian process on $\mathscr{M}$.

Think of this as *generalized CLT* for singular models.
Classical central limit theorem (CLT):

$$\text{sample mean} = \frac{1}{N}\sum_{i=1}^N X_i = \mu + \frac{1}{\sqrt{N}}\sigma\xi_N$$

where ξ_N converges in law to standard normal distribution.

Fiber Ideals

How do we desingularize $K(\omega) = \int_{\mathcal{X}} q(x) \log \frac{q(x)}{p(x|\omega)} dx$?

- Algorithms (e.g. Bravo-Encinas-Villamayor) intractable
- Many models parametrized by *polynomials*. Exploit this?

Regularly parametrized functions

- A function $f : \Omega \to \mathbb{R}$ is *regularly parametrized* if it factors

$$\Omega \xrightarrow{u} U \xrightarrow{g} \mathbb{R}$$

  where $U \subset \mathbb{R}^k$ nbhd of origin, u is polynomial, g has unique minimum $g(0) = 0$ at the origin and $\det \partial^2 g(0) \neq 0$.
- For such functions, define *fiber ideal*

$$I = \langle u_1(\omega), \ldots, u_k(\omega) \rangle \subset \mathbb{R}[\omega_1, \ldots, \omega_d].$$

  The variety $\mathcal{V}(I)$ is the fiber $f^{-1}(0)$.

Equivalence (Watanabe) RLCT of f = RLCT of $u_1^2 + \cdots + u_k^2$.

Examples of Fiber Ideals

$$f : \Omega \xrightarrow{u} U \xrightarrow{g} \mathbb{R}$$

- **Laplace Approximation**. When f is the sum-of-squares
$$f = \omega_1^2 + \ldots + \omega_d^2,$$
we let g be f and u be the identity map. The fiber ideal is
$$I = \langle \omega_1, \ldots, \omega_d \rangle.$$

- **Coin Toss Integral**. In one of our earlier examples
$$f = -\frac{1}{2}\log(1 - x^2y^2),$$
let $u(x, y) = xy, g(u) = -\frac{1}{2}\log(1 - u^2)$. The fiber ideal is
$$I = \langle xy \rangle.$$

Examples of Fiber Ideals

- **Discrete Models**. Given true distribution $\hat{p} \in \mathcal{M}$ and state probabilities $p(i|\omega)$, the Kullback-Leibler distance $K(\omega)$ factors

$$K : \Omega \xrightarrow{p} \Delta_{k-1} \xrightarrow{g} \mathbb{R}$$

where

$$g(p) = \sum_{i=1}^{k} \hat{p}(i) \log \frac{\hat{p}(i)}{p(i)}$$

and $\det \partial^2 g$ is nonzero at $\hat{p}$. The fiber ideal is

$$I_{\hat{p}} = \langle p(1|\omega) - \hat{p}(1), \ldots, p(k|\omega) - \hat{p}(k) \rangle.$$

- **Gaussian models**. Given true distribution $\mathcal{N}(\hat{\mu}, \hat{\Sigma})$ and model distributions $\mathcal{N}(\mu(\omega), \Sigma(\omega))$, the Kullback-Leibler function $K(\omega)$ is also regularly parametrized. The fiber ideal is

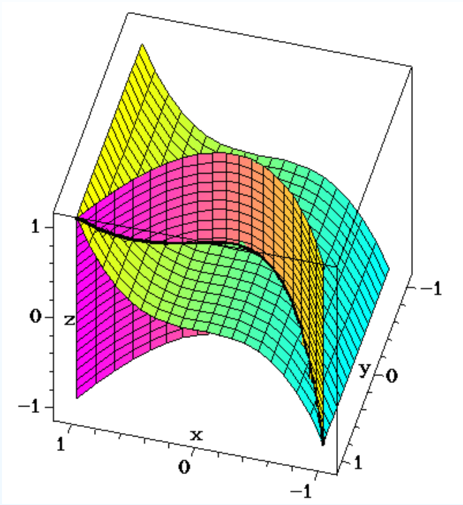
$$I_{\hat{\mu},\hat{\Sigma}} = \langle \mu_i(\omega) - \hat{\mu}_i, \Sigma_{ij}(\omega) - \hat{\Sigma}_{ij} \rangle_{ij}$$

Real Log Canonical Thresholds

Ideals and Varieties

Ideal $\langle y - x^2, z - x^3 \rangle$
set of polynomials generated by $y - x^2$ and $z - x^3$ via addition and polynomial-scaling

Variety $\mathcal{V}(y - x^2, z - x^3)$
set of points where polynomials in the ideal evaluate to zero



In linear algebra, we solve linear equations by computing a *row echelon form* using *Gaussian elimination*.
In algebraic geometry, we solve polynomial equations by computing a *Gröbner basis* using *Buchberger's algorithm*.

Textbook: "Ideals, Varieties, and Algorithms," Cox-Little-O'Shea.
Software: `Macaulay2`, `Singular`, `Maple`, etc.

Real Log Canonical Thresholds of Ideals

Given ideal $I = \langle f_1(\omega), \ldots, f_k(\omega) \rangle \subset \mathbb{R}[\omega_1, \ldots, \omega_d]$,
polynomial $\varphi(\omega)$, semialgebraic $\Omega \subset \mathbb{R}^d$.

The *real log canonical threshold* (λ, θ) of I at $x \in \Omega$ satisfies

$$\int_{\Omega_x} e^{-N(f_1^2+\cdots+f_k^2)} \varphi(\omega) d\omega \;\approx\; CN^{-\lambda/2} (\log N)^{\theta-1}$$

for suff small nbhd Ω_x of x in Ω. Denote $(\lambda, \theta) = \mathrm{RLCT}_{\Omega_x}(I; \varphi)$.

Properties

- Definition is independent of choice of generators $f_1, \ldots, f_k$.
- λ positive *rational* number, θ positive *integer*.
- Depends on structure of boundary $\partial\Omega$ if $x \in \partial\Omega$.
- Order the (λ, θ) by the value of $N^\lambda (\log N)^{-\theta}$ for large N.

Discrete and Gaussian Models

- *Discrete models* with state probabilities $p(i|\omega)$.
  Fiber ideal at a true distribution $\hat{p}$
  $$I_{\hat{p}} = \langle p(i|\omega) - \hat{p}(i) \rangle_i$$
- *Gaussian models* with mean $\mu(\omega)$ and covariance $\Sigma(\omega)$.
  Fiber ideal at a true distribution $\mathcal{N}(\hat{\mu}, \hat{\Sigma})$
  $$I_{\hat{\mu},\hat{\Sigma}} = \langle \mu_i(\omega) - \hat{\mu}_i, \Sigma_{ij}(\omega) - \hat{\Sigma}_{ij} \rangle_{ij}$$

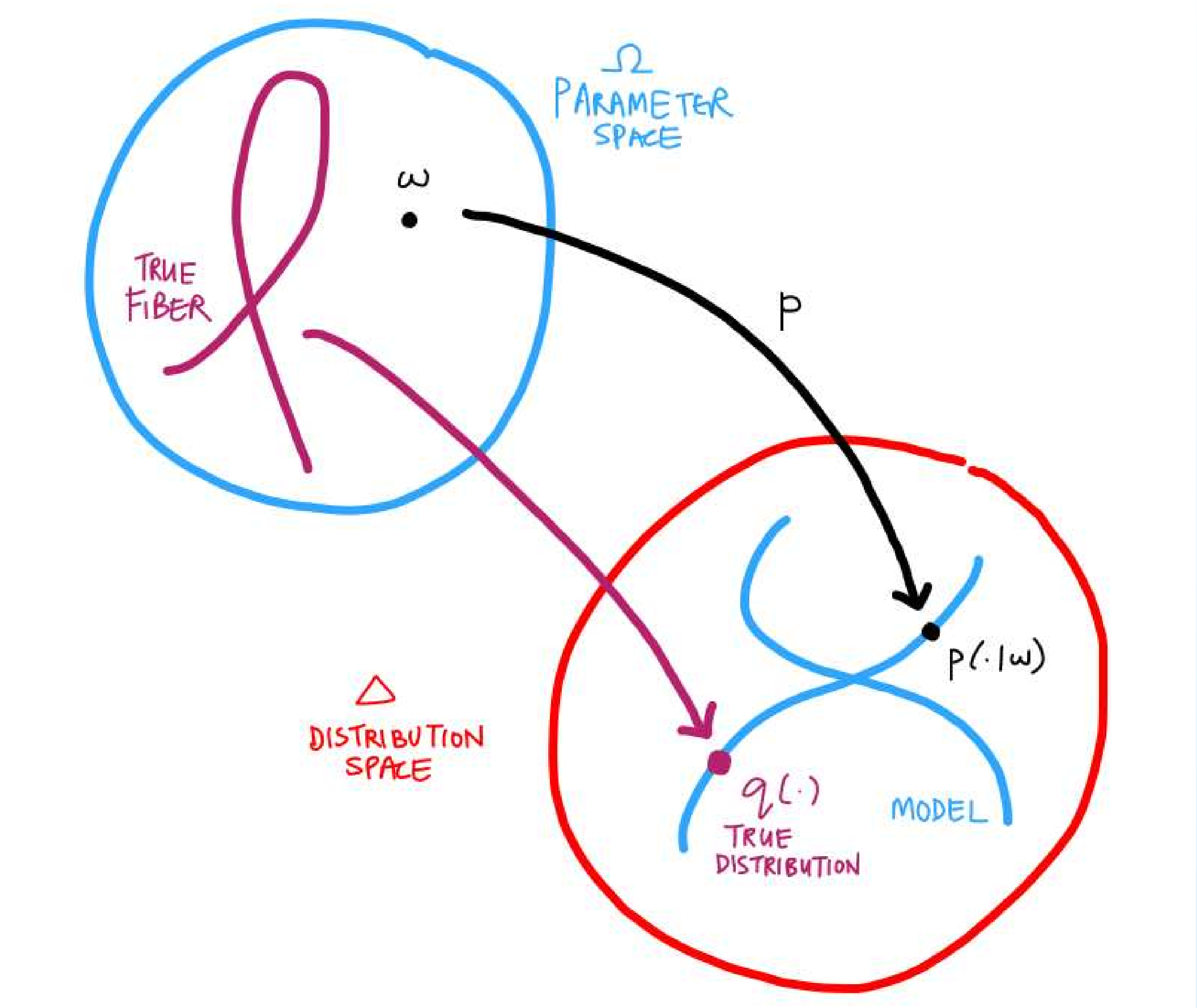
Learning coefficients and RLCTs of fiber ideals (L.)

If the true distribution q is in the model,
then the learning coefficient (λ_q, θ_q) is given by

$$(2\lambda_q, \theta_q) = \min_{x \in \mathcal{V}(I_q)} \mathrm{RLCT}_{\Omega_x}(I_q; \varphi)$$

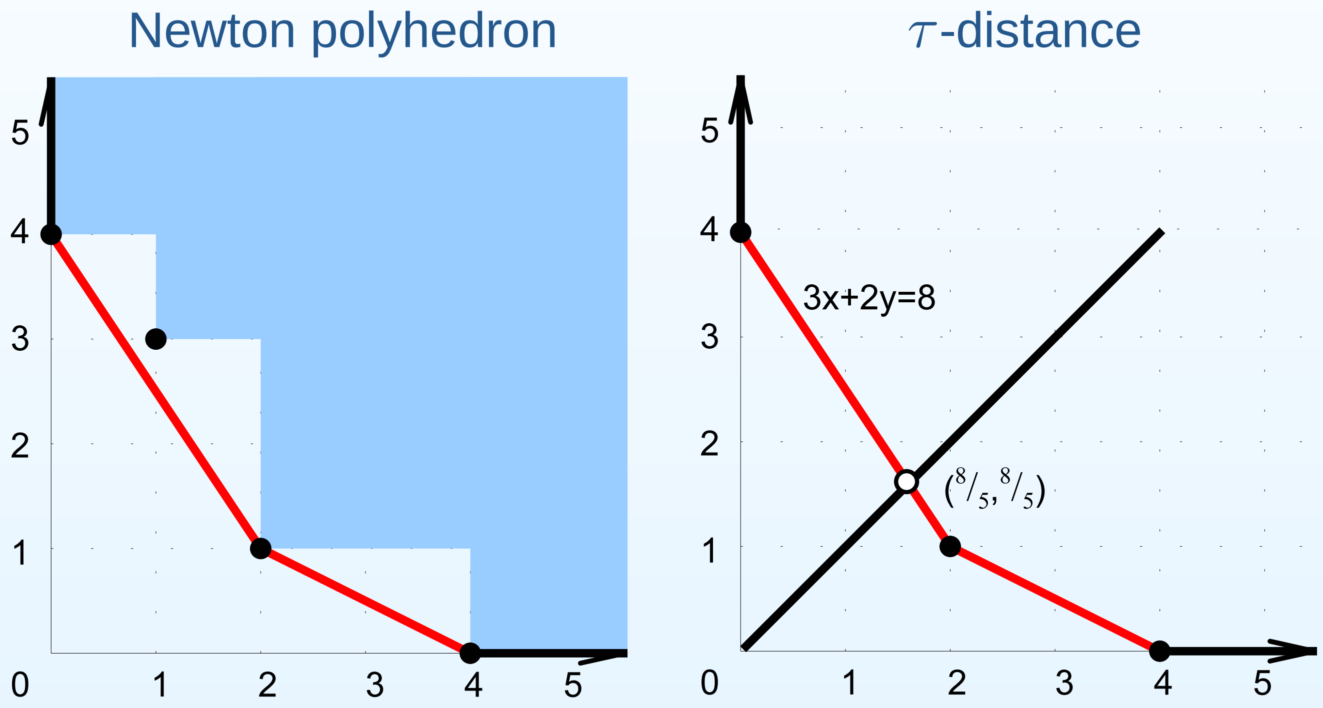
where I_q is the fiber ideal at q and $\mathcal{V}(I_q) \subset \Omega$ is the fiber over q.

Geometry of Singular Models



Distance and Multiplicity

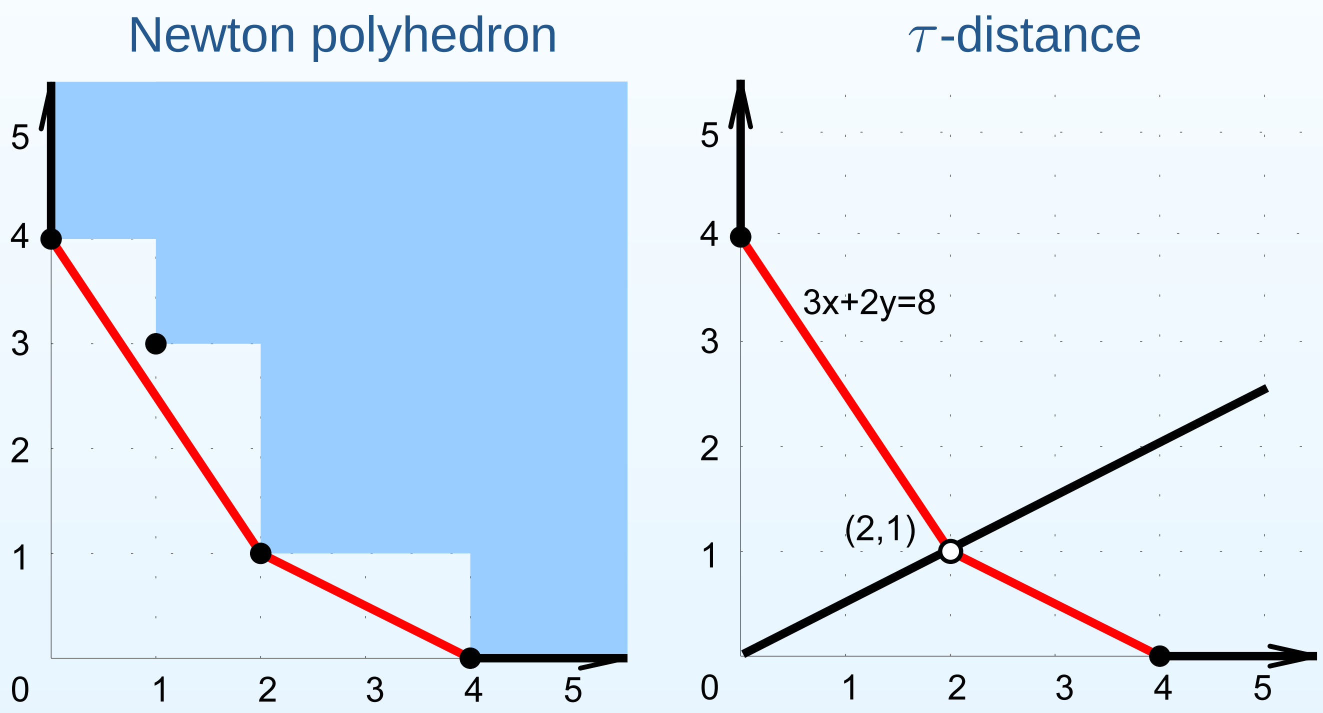
e.g. Let $I = \langle x^4, x^2y, xy^3, y^4\rangle$ and $\tau = (1,1)$.



The τ-distance is $l_\tau = 8/5$ and the multiplicity is $\theta_\tau = 1$.

Distance and Multiplicity

e.g. Let $I = \langle x^4, x^2y, xy^3, y^4\rangle$ and $\tau = (2,1)$.



The τ-distance is $l_\tau = 1$ and the multiplicity is $\theta_\tau = 2$.

Upper Bounds

Given an ideal $I \subset \mathbb{R}[\omega_1, \ldots, \omega_d]$,

1. Plot $\alpha \in \mathbb{R}^d$ for each monomial ω^α appearing in some $f \in I$.
2. Take the convex hull $\mathcal{P}(I)$ of all plotted points.

This convex hull $\mathcal{P}(I)$ is the *Newton polyhedron* of I.

Given a vector $\tau \in \mathbb{Z}^d_{\geq 0}$, define

1. *τ-distance* $l_\tau = \min\{t : t\tau \in \mathcal{P}(I)\}$.
2. *multiplicity* θ_τ = codim of face of $\mathcal{P}(I)$ at this intersection.

Upper bound and equality for RLCT (L.)

If l_τ is the τ-distance of $\mathcal{P}(I)$ and θ_τ is its multiplicity, then

$$\mathrm{RLCT}_{\Omega_x}(I; \omega^{\tau-1}) \leq (1/l_\tau, \theta_\tau).$$

Equality occurs when I is a monomial ideal.

Integral Asymptotics

Bayesian Information Criterion (BIC)

When the model is regular, the fiber ideal is $I = \langle \omega_1, \ldots, \omega_d \rangle$.
Using Newton polyhedra, $\mathrm{RLCT}(I) = (d, 1)$ (exercise).
By Watanabe's theorem, the likelihood integral Z_n is asymptotically

$$-\log Z_N \ \approx \ N S_N + \frac{d}{2} \log N.$$

Coin Toss Integral

$$Z(N) \ = \ \int_{[0,1]^2} (1 - x^2 y^2)^{N/2} \, dx dy.$$

Earlier, we saw that the fiber ideal for this integral is $I = \langle xy \rangle$.
Using Newton polyhedra, $\mathrm{RLCT}(I) = (1, 2)$ (exercise).
Therefore, for some $C > 0$, the integral $Z(N)$ is asymptotically

$$Z(N) \ \approx \ C N^{-1/2} \log N$$

`Macaulay2`
Computations

132 Schizophrenic Patients (Evans·Gilula·Guttman)

Naïve Bayes network with 2 ternary variables, 2 hidden states.
Model parametrized in $\omega = (t, a_1, a_2, \ldots, d_3)$ by

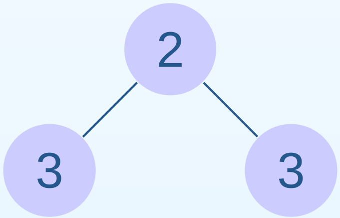
$$p = \begin{pmatrix} ta_1b_1 + (1-t)c_1d_1 & ta_1b_2 + (1-t)c_1d_2 & ta_1b_3 + (1-t)c_1d_3 \\ ta_2b_1 + (1-t)c_2d_1 & ta_2b_2 + (1-t)c_2d_2 & ta_2b_3 + (1-t)c_2d_3 \\ ta_3b_1 + (1-t)c_3d_1 & ta_3b_2 + (1-t)c_3d_2 & ta_3b_3 + (1-t)c_3d_3 \end{pmatrix}.$$

Assume true distribution $\hat{p}_{ij} = \frac{1}{9}$ for all i, j.

Compute RLCT of fiber ideal

$$I = \langle p_{11}(\omega) - \hat{p}, \ldots, p_{33}(\omega) - \hat{p} \rangle$$

at the point $\hat{w} = (\frac{1}{2}, \frac{1}{3}, \frac{1}{3}, \ldots, \frac{1}{3}) \in \mathcal{V}(I)$.



Computations using our library `asymptotics.m2` show that

$$\mathrm{RLCT}_{\hat{\omega}}(I; 1) = (6, 2).$$

All other learning coefficients can be computed in this fashion.

Model Definition

```
Macaulay2, version 1.4
with packages: ConwayPolynomials, Elimination,
               IntegralClosure, LLLBases,
               PrimaryDecomposition, ReesAlgebra,
               TangentCone

i1 : load "asymptotics.m2";
i2 : R = QQ[t,a1,a2,b1,b2,c1,c2,d1,d2];
i3 : A = matrix {{a1,a2,1-a1-a2}};
i4 : B = matrix {{b1,b2,1-b1-b2}};
i5 : C = matrix {{c1,c2,1-c1-c2}};
i6 : D = matrix {{d1,d2,1-d1-d2}};
i7 : P = t*(transpose A)*B + (1-t)*(transpose C)*D;
             3       3
o7 : Matrix R  <--- R
```

Fiber Ideal

Maps for shifting the origin to $\hat{\omega}$ and evaluating a polynomial at $\hat{\omega}$.

```
i8 : shift = map(R,R,{t+1/2,a1+1/3,a2+1/3,b1+1/3,b2+1/3,
                            c1+1/3,c2+1/3,d1+1/3,d2+1/3});
i9 : eval = map(R,R,{1/2,1/3,1/3,1/3,1/3,
                         1/3,1/3,1/3,1/3});
```

The true distribution.

```
i10 : eval P
o10 = {-1} | 1/9 1/9 1/9 |
      {-1} | 1/9 1/9 1/9 |
      {-1} | 1/9 1/9 1/9 |
```

The fiber ideal.

```
i11 : I = ideal (shift P - eval P);
o11 : Ideal of R
```

Gröbner Basis

Gröbner basis of the fiber ideal.

```
i12 : I = ideal gens gb I
o12 = ideal (a2*d2, a1*d2, b2*d1 - b1*d2, a2*d1, a1*d1,
      b2*c2, b1*c2, b2*c1, b1*c1, a2*c1 - a1*c2,
      2t*b2 - 2t*d2 + b2 + d2, 2t*b1 - 2t*d1 + b1 + d1,
      2t*a2 - 2t*c2 + a2 + c2, 2t*a1 - 2t*c1 + a1 + c1,
      2t*c2*d2 - c2*d2, 2t*c1*d2 - c1*d2,
      2t*c2*d1 - c2*d1, 2t*c1*d1 - c1*d1)
```

Preliminary upper bound of the RLCT.

```
i13 : RLCT(I,1)
[RLCT] Warning: Output RLCT is an upper bound.

o13 = (8, 1)
```

To compute the RLCT, we transform I into a monomial ideal.

Gröbner Basis

Gröbner basis of the fiber ideal.

```
i12 : I = ideal gens gb I
o12 = ideal (a2*d2, a1*d2, b2*d1 - b1*d2, a2*d1, a1*d1,
      b2*c2, b1*c2, b2*c1, b1*c1, a2*c1 - a1*c2,
      2t*b2 - 2t*d2 + b2 + d2, 2t*b1 - 2t*d1 + b1 + d1,
      2t*a2 - 2t*c2 + a2 + c2, 2t*a1 - 2t*c1 + a1 + c1,
      2t*c2*d2 - c2*d2, 2t*c1*d2 - c1*d2,
      2t*c2*d1 - c2*d1, 2t*c1*d1 - c1*d1)
```

The red generator prevents I from being a monomial ideal. Replace it with new indeterminate β_2 via the change of variable

$$b_2 = \frac{\beta_2 - (1-2t)d_2}{1+2t}$$

which is a real-analytic isomorphism near the origin.

We can also accomplish this by introducing a new polynomial $-\beta_2 + 2tb_2 - 2td_2 + b_2 + d_2$ to the ideal and eliminating b_2.

Monomialization

Perform similar transformations to a_1, a_2, b_1, b_2.

```
i14 : R1 = QQ[t,a1,a2,b1,b2,c1,c2,d1,d2,bb1,bb2,cc1,cc2];
i15 : liftR1 = map(R1,R,{t,a1,a2,b1,b2,c1,c2,d1,d2});
i16 : I1 = (liftR1 I) + ideal(
           -bb2 + 2*t*b2 - 2*t*d2 + b2 + d2,
           -bb1 + 2*t*b1 - 2*t*d1 + b1 + d1,
           -cc2 + 2*t*a2 - 2*t*c2 + a2 + c2,
           -cc1 + 2*t*a1 - 2*t*c1 + a1 + c1);
i17 : I1 = eliminate({c1,c2,b1,b2},I1)
o17 = ideal (cc2, cc1, bb2, bb1,
             a2*d2, a1*d2, a2*d1, a1*d1)
```

Finally, we have a monomial ideal so we can compute its RLCT.

```
i18 : RLCT(I1,1)
o18 = (6, 2)
```

Automation

This analysis can be automated somewhat using the following algorithms from `asymptotics.m2`.

```
i21 : I1 = simplifyRegularParameters I
o21 = ideal (a1, a2, b1, b2,
             2t*c2*d2 - c2*d2, 2t*c1*d2 - c1*d2,
             2t*c2*d1 - c2*d1, 2t*c1*d1 - c1*d1)

i22 : removeUnitComponents I1
o22 = ideal (b2, b1, a2, a1, c2*d2, c1*d2, c2*d1, c1*d1)
```

For more information about this `Macaulay2` library:

`http://math.berkeley.edu/~shaowei/rlct.html`

"Algebraic Methods for Evaluating Integrals in Bayesian Statistics"

`http://math.berkeley.edu/~shaowei/swthesis.pdf`

(PhD dissertation, May 2011)

References

1. V. I. Arnol'd, S. M. Guseĭn-Zade and A. N. Varchenko: *Singularities of Differentiable Maps*, Vol. II, Birkhäuser, Boston, 1985.
2. A. Bravo, S. Encinas and O. Villamayor: A simplified proof of desingularisation and applications, *Rev. Math. Iberoamericana* **21** (2005) 349–458.
3. D. A. Cox, J. B. Little, and D. O'Shea: *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, Springer-Verlag, New York, 1997.
4. M. Evans, Z. Gilula and I. Guttman: Latent class analysis of two-way contingency tables by Bayesian methods, *Biometrika* **76** (1989) 557–563.
5. H. Hironaka: Resolution of singularities of an algebraic variety over a field of characteristic zero I, II, *Ann. of Math.* (2) **79** (1964) 109–203.
6. S. Lin, B. Sturmfels and Z. Xu: Marginal likelihood integrals for mixtures of independence models, *J. Mach. Learn. Res.* **10** (2009) 1611–1631.
7. S. Lin: Algebraic methods for evaluating integrals in Bayesian statistics, PhD dissertation, Dept. Mathematics, UC Berkeley (2011).
8. S. Watanabe: *Algebraic Geometry and Statistical Learning Theory*, Cambridge Monographs on Applied and Computational Mathematics **25**, Cambridge University Press, Cambridge, 2009.

Supplementary Material

Coin Toss Integral

The integral $Z(N)$ with $f(x,y) = -\frac{1}{2}\log(1-x^2y^2)$ comes from the coin toss model parametrized by

$$p_1(\omega,t) = \frac{1}{2}t + (1-t)\omega$$
$$p_2(\omega,t) = \frac{1}{2}t + (1-t)(1-\omega)$$

where the Kullback-Leibler function at the distribution (q_1, q_2)

$$K(\omega,t) = q_1 \log\frac{q_1}{p_1(\omega,t)} + q_2 \log\frac{q_2}{p_2(\omega,t)}.$$

The function $f(x,y)$ comes from $K(x,y)$ at $q_1 = q_2 = 1/2$ and substituting $\omega = (1+x)/2, t = 1-y$.

Nondegenerate Ideals

Let $[\omega^\alpha]f$ denote coefficient of monomial ω^α in polynomial f.

Given $\gamma \subset \mathbb{R}^d$ and poly f, define *face poly* $f_\gamma = \sum_{\alpha\in\gamma}([\omega^\alpha]f)\omega^\alpha$.
Given $\gamma \subset \mathbb{R}^d$ and ideal I, define *face ideal* $I_\gamma = \langle f_\gamma : f \in I\rangle$.

We say I is *sos-nondegenerate* if for all compact faces $\gamma \subset \mathcal{P}(I)$, the real variety $\mathcal{V}(I_\gamma)$ does not intersect the torus $(\mathbb{R}^*)^d$.

Remark sos = sum-of-squares. Saia has similar notion of nondegeneracy for ideals of *complex* formal power series.

Proposition (L.) If $I = \langle f_1, \ldots, f_r\rangle$ and γ is a compact face of the Newton polyhedron $\mathcal{P}(I)$, then $I_\gamma = \langle f_{1\gamma}, \ldots, f_{r\gamma}\rangle$.

Proposition (L.) $\mathrm{RLCT}(I;\omega^{\tau-1}) = (1/l_\tau, \theta_\tau)$ if I is sos-ndg.

Proposition (Zwiernik) Monomial ideals are sos-ndg.

Toric Blowups

Let $\mathcal{F}$ be a *smooth polyhedral fan* supported on the orthant $\mathbb{R}^d_{\geq 0}$.
[smooth: each cone is generated by a subset of some basis of $\mathbb{Z}^d$]

Recall that we can associate to $\mathcal{F}$, a *toric variety* $\mathbb{P}(\mathcal{F})$ covered by open affines $U_\sigma \simeq \mathbb{R}^d$, one for each maximal cone σ of $\mathcal{F}$.

We also have a *blowup map* $\rho_\mathcal{F} : \mathbb{P}(\mathcal{F}) \to \mathbb{R}^d$ described by monomial maps $\rho_{\mathcal{F},\sigma} : U_\sigma \to \mathbb{R}^d$, $\mu \mapsto \mu^\nu$, on the open affines.
[The columns of the matrix ν are minimal generators of the maximal cone σ, and $(\mu^\nu)_i = \mu^{\nu_i}$ where ν_i is the ith row of ν.]

Proposition (L.):
Given a fiber ideal I, let $\mathcal{F}$ be a *smooth refinement* of the normal fan of the Newton polyhedron $\mathcal{P}(I)$. If I is sos-nondegenerate, then the toric blowup $\rho_\mathcal{F} : \mathbb{P}(\mathcal{F}) \to \mathbb{R}^d$ desingularizes f.

Strategy for Regularly Parametrized Functions

Given a regularly parametrized function $f = g \circ u : \Omega \to \mathbb{R}$, we want to *exploit the polynomiality* in u in desingularizing f.
Let $I = \langle u_1, \ldots, u_k \rangle$ be the polynomial fiber ideal.
Given $\rho : M \to \Omega$, define *pullback* $\rho^* I = \langle u_1 \circ \rho, \ldots, u_k \circ \rho \rangle$.

1. **Monomialization** (polynomial):
   Find a map $\rho : M \to \Omega$ which *monomializes* I,
   i.e. $\rho^* I$ is a monomial ideal in each patch of M.
   Use algorithm of Bravo-Encinas-Villamayor.

2. **Principalization** (combinatorial):
   Find a map $\eta : \mathscr{M} \to M$ which *principalizes* $J = \rho^* I$,
   i.e. $\eta^* J$ is generated by one monomial in each patch of $\mathscr{M}$.
   Use toric blowups or Goward's principalization map.

Theorem (L.) The composition $\rho \circ \eta$ desingularizes f.

132 Schizophrenic Patients: The Model

Evans-Gilula-Guttman(1989) studied schizophrenic patients for connections between recovery time (in years Y) and frequency of visits by relatives.

	$2 \leq Y < 10$	$10 \leq Y < 20$	$20 \leq Y$	*Totals*
Regularly	43	16	3	*62*
Rarely	6	11	10	*27*
Never	9	18	16	*43*
Totals	*58*	*45*	*29*	**132**

They wanted to find out if the data can be explained by a *naïve Bayesian network* with two hidden states (e.g. male and female).

132 Schizophrenic Patients: Exact Integral

Model parametrized by $(t, a, b, c, d) \in \Delta_1 \times \Delta_2 \times \Delta_2 \times \Delta_2 \times \Delta_2$.

	$2 \le Y < 10$	$10 \le Y < 20$	$20 \le Y$
Regularly	$ta_1b_1 + (1-t)c_1d_1$	$ta_1b_2 + (1-t)c_1d_2$	$ta_1b_3 + (1-t)c_1d_3$
Rarely	$ta_2b_1 + (1-t)c_2d_1$	$ta_2b_2 + (1-t)c_2d_2$	$ta_2b_3 + (1-t)c_2d_3$
Never	$ta_3b_1 + (1-t)c_3d_1$	$ta_3b_2 + (1-t)c_3d_2$	$ta_3b_3 + (1-t)c_3d_3$

We compute the *marginal likelihood* of this model, given the above data and a uniform prior on the parameter space.

Lin-Sturmfels-Xu(2009) computed this integral *exactly*.
It is the rational number with numerator

$$\begin{array}{l}278019488531063389120643600324989329103876140805\\285242839582092569357265886675322845874097528033\\99493069713103633199906939405711180837568853737\end{array}$$

and denominator

$$\begin{array}{l}12288402873591935400678094796599848745442833177572204\\50448819979286456995185542195946815073112429169997801\\33503900169921912167352239204153786645029153951176422\\43298328046163472261962028461650432024356339706541132\\34375318471880274818667657423749120000000000000000.\end{array}$$

132 Schizophrenic Patients: Maximum Likelihood

We want to approximate the integral using asymptotic methods.

The EM algorithm gives us the *maximum likelihood distribution*

$$q = \frac{1}{132}\begin{pmatrix} 43.002 & 15.998 & 3.000 \\ 5.980 & 11.123 & 9.897 \\ 9.019 & 17.879 & 16.102 \end{pmatrix}.$$

Compare this distribution with the data

$$\begin{pmatrix} 43 & 16 & 3 \\ 6 & 11 & 10 \\ 9 & 18 & 16 \end{pmatrix}.$$

Use ML distribution as *true distribution* for our approximations.

132 Schizophrenic Patients: Asymptotic Approximation

Recall that stochastic complexity $= -\log$ (marginal likelihood).

- The BIC approximates the stochastic complexity as

$$NS_N + \frac{9}{2}\log N.$$

- By computing the RLCT of the fiber ideal, our approximation is

$$NS_N + \frac{7}{2}\log N.$$

- Summary:

	Stochastic Complexity
Exact	273.1911759
BIC	278.3558034
RLCT	275.9144024

132 Schizophrenic Patients: Learning Coefficients

$$Z_N = \int_\Omega \prod_{i,j} p_{ij}(\omega)^{U_{ij}} \varphi(\omega) d\omega$$

Using Watanabe's *Singular Learning Theory*,

$$-\log Z_N \approx -\sum_{i,j} U_{ij} \log q_{ij} + \lambda_q \log N - (\theta_q - 1) \log\log N$$

where the *learning coefficient* (λ_q, θ_q) is given by

$$(\lambda_q, \theta_q) = \begin{cases} (5/2, 1) & \text{if } \operatorname{rank} q = 1, \\ (7/2, 1) & \text{if } \operatorname{rank} q = 2,\ q \notin \left[\begin{smallmatrix} 0 & \times \\ \times & \times \end{smallmatrix}\right] \cup \left[\begin{smallmatrix} 0 & \times \\ \times & 0 \end{smallmatrix}\right], \\ (4, 1) & \text{if } \operatorname{rank} q = 2,\ q \in \left[\begin{smallmatrix} 0 & \times \\ \times & \times \end{smallmatrix}\right] \setminus \left[\begin{smallmatrix} 0 & \times \\ \times & 0 \end{smallmatrix}\right], \\ (9/2, 1) & \text{if } \operatorname{rank} q = 2,\ q \in \left[\begin{smallmatrix} 0 & \times \\ \times & 0 \end{smallmatrix}\right]. \end{cases}$$

Here, $q \in \left[\begin{smallmatrix} 0 & \times \\ \times & \times \end{smallmatrix}\right]$ if for some $i, j,\ q_{ii} = 0$ and $q_{ij}\, q_{ji}\, q_{jj} \neq 0$,
$q \in \left[\begin{smallmatrix} 0 & \times \\ \times & 0 \end{smallmatrix}\right]$ if for some $i, j,\ q_{ii} = q_{jj} = 0$ and $q_{ij}\, q_{ji} \neq 0$.