Asymptotic Approximation of Marginal Likelihood Integrals

Shaowei Lin (UC Berkeley)

shaowei@math.berkeley.edu

26 Aug 2010

Leibniz Universität, Hannover, Germany

A Statistical Example

132 Schizophrenic Patients

- 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**

- Proposed two statistical models to explain the data.

132 Schizophrenic Patients

- Model 1: Independence Model

	$2 \leq Y < 10$	$10 \leq Y < 20$	$20 \leq Y$
Regularly	$a_1 b_1$	$a_1 b_2$	$a_1 b_3$
Rarely	$a_2 b_1$	$a_2 b_2$	$a_2 b_3$
Never	$a_3 b_1$	$a_3 b_2$	$a_3 b_3$

132 Schizophrenic Patients

- Model 1: Independence Model

	$2 \leq Y < 10$	$10 \leq Y < 20$	$20 \leq Y$
Regularly	$a_1 b_1$	$a_1 b_2$	$a_1 b_3$
Rarely	$a_2 b_1$	$a_2 b_2$	$a_2 b_3$
Never	$a_3 b_1$	$a_3 b_2$	$a_3 b_3$

- Model 2: Hidden Variable Model

	$2 \leq Y < 10$	$10 \leq Y < 20$	$20 \leq Y$
Regularly	$ta_1 b_1 + (1-t)c_1 d_1$	$ta_1 b_2 + (1-t)c_1 d_2$	$ta_1 b_3 + (1-t)c_1 d_3$
Rarely	$ta_2 b_1 + (1-t)c_2 d_1$	$ta_2 b_2 + (1-t)c_2 d_2$	$ta_2 b_3 + (1-t)c_2 d_3$
Never	$ta_3 b_1 + (1-t)c_3 d_1$	$ta_3 b_2 + (1-t)c_3 d_2$	$ta_3 b_3 + (1-t)c_3 d_3$

Marginal Likelihood Integrals

- In Bayesian statistics, models are selected by comparing *marginal likelihood integrals*.

$$Z = \int_\Omega \prod_i p_i(\omega)^{U_i}\, \varphi(\omega) d\omega$$

U_i the data, Ω parameter space
$p_i(\omega)$ functions parametrizing the model
$\varphi(\omega)$ prior belief about parameter space

Marginal Likelihood Integrals

- e.g. Independence Model

	$2 \leq Y < 10$	$10 \leq Y < 20$	$20 \leq Y$	*Totals*
Reg.	$a_1 b_1$ (43)	$a_1 b_2$ (16)	$a_1 b_3$ (3)	(62)
Rarely	$a_2 b_1$ (6)	$a_2 b_2$ (11)	$a_2 b_3$ (10)	(27)
Never	$a_3 b_1$ (9)	$a_3 b_2$ (18)	$a_3 b_3$ (16)	(43)
Totals	(58)	(45)	(29)	

$$Z_1 \quad = \quad \int_{\Delta_2} \int_{\Delta_2} a_1^{62} a_2^{27} a_3^{43} b_1^{58} b_2^{45} b_3^{29} \, da \, db$$

$$a = (a_1, a_2, a_3), \; b = (b_1, b_2, b_3)$$

$$\Delta_2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_i \geq 0, \textstyle\sum_i x_i = 1\}$$

Asymptotic Approximation

- **IMPORTANT PROBLEM IN STATISTICS**

Compute $$Z(n) = \int_{\Omega} \prod_{i=1}^{k} p_i(\omega)^{nq_i}\, \varphi(\omega) d\omega$$

where n sample size, $q_i = U_i/n$ relative frequencies.

Assumptions:

Ω compact and semianalytic

i.e. $\Omega = \{\omega \in \mathbb{R}^d : g_1 \geq 0, \ldots, g_l \geq 0\}$, g_i real analytic on $\mathbb{R}^d$

φ nearly analytic

i.e. $\varphi = \varphi_s \varphi_a$, φ_s positive and smooth, φ_a real analytic on Ω

p_i positive real analytic functions on Ω summing to 1

q true distribution lying in $p(\Omega)$

Asymptotic Approximation

- **IMPORTANT PROBLEM IN STATISTICS**

  Compute $$Z(n) = \int_\Omega \prod_{i=1}^{k} p_i(\omega)^{nq_i}\, \varphi(\omega) d\omega$$

- L.-Sturmfels-Xu(2008) gave efficient algorithms for computing $Z(n)$ *exactly* for small samples n.

Asymptotic Approximation

- **IMPORTANT PROBLEM IN STATISTICS**

  Compute $\quad Z(n) = \int_\Omega \prod_{i=1}^{k} p_i(\omega)^{nq_i}\, \varphi(\omega) d\omega$

- L.-Sturmfels-Xu(2008) gave efficient algorithms for computing $Z(n)$ *exactly* for small samples n.
- *Asymptotically*, as $n \to \infty$,

$$Z(n) \;\approx\; (\textstyle\prod_{i=1}^{k} q_i^{q_i})^n \cdot C n^{-\lambda} (\log n)^{\theta - 1}$$

In this talk, we discuss how to compute (λ, θ).

In machine learning, λ is called the *learning coefficient* of the statistical model and θ its *multiplicity*.

Statistical Learning Theory and Singularity Theory

Statistical Learning Theory

In statistics, it is customary to write our marginal likelihood integral as the Laplace integral

$$\begin{aligned} Z(n) &= \int_\Omega \prod_{i=1}^{k} p_i(\omega)^{nq_i}\, \varphi(\omega)d\omega \\ &= \left(\prod_{i=1}^{k} q_i^{q_i}\right)^n \int_\Omega e^{-nK(\omega)}\, \varphi(\omega)d\omega \end{aligned}$$

where $\prod_{i=1}^{k} q_i^{q_i}$ is the maximum likelihood and $K(\omega)$ is the Kullback-Leibler distance

$$K(\omega) = \sum_{i=1}^{k} q_i \log \frac{q_i}{p_i(\omega)}.$$

Statistical Learning Theory

Theorem (Watanabe)

The Laplace integrals defined by functions

$$\begin{aligned} K(\omega) &= \textstyle\sum_{i=1}^{k} q_i \log \frac{q_i}{p_i(\omega)} \\ Q(\omega) &= \textstyle\sum_{i=1}^{k} \big(p_i(\omega) - q_i\big)^2 \end{aligned}$$

are asymptotically the same up to a constant, i.e.

$$\begin{aligned} \int_\Omega e^{-nK(\omega)} |\varphi(\omega)| d\omega &\approx C_1 n^{-\lambda} (\log n)^{\theta-1} \\ \int_\Omega e^{-nQ(\omega)} |\varphi(\omega)| d\omega &\approx C_2 n^{-\lambda} (\log n)^{\theta-1} \end{aligned}$$

for some constants $\lambda, \theta, C_1, C_2$.

Singularity Theory

Theorem (Arnold-Gusein·Zade-Varchenko)

Let f be a real analytic function on Ω with $f(\omega^*) = 0$ for some $\omega^* \in \Omega$. If we have asymptotics

$$Z(n) = \int_\Omega e^{-n|f(\omega)|}|\varphi(\omega)|d\omega \;\approx\; Cn^{-\lambda}(\log n)^{\theta-1},$$

then λ is the smallest pole of the zeta function

$$\zeta(z) = \int_\Omega |f(\omega)|^{-z}|\varphi(\omega)|d\omega, \quad z \in \mathbb{C}$$

and θ is the multiplicity of this pole.

Singularity Theory

Theorem (Arnold-Gusein·Zade-Varchenko)

Let f be a real analytic function on Ω with $f(\omega^*) = 0$ for some $\omega^* \in \Omega$. Then, asymptotically, we have

$$Z(n) = \int_\Omega e^{-n|f(\omega)|}|\varphi(\omega)|d\omega \;\approx\; \sum_\alpha \sum_{i=0}^{d-1} c_{i,\alpha}\, n^{-\alpha}(\log n)^i.$$

The α in this expansion range over positive rational numbers which are poles of

$$\zeta(z) = \int_{\Omega_\delta} \left|f(\omega)\right|^{-z}|\varphi(\omega)|\, d\omega$$

where $\Omega_\delta = \{\omega \in \Omega : |f(\omega)| < \delta\}$ for any $\delta > 0$. The coefficients $c_{i,\alpha}$ satisfy

$$c_{i,\alpha} = \frac{(-1)^{i+1}}{i!} \sum_{j=i}^{d-1} \frac{\Gamma^{(j-i)}(\alpha)}{(j-i)!}\, d_{j+1,\alpha}$$

where $d_{j,\alpha}$ is the coefficient of $(z-\alpha)^{-(j)}$ in the Laurent expansion of $\zeta(z)$.

Example: Monomial Functions

- Let $f = \omega_1^{\kappa_1} \cdots \omega_d^{\kappa_d}$ and $\varphi = \omega_1^{\tau_1} \cdots \omega_d^{\tau_d}$.

$$\int_{[0,\varepsilon]^d} e^{-n\omega^\kappa} \omega^\tau \, d\omega = C n^{-\lambda} (\log n)^{\theta - 1}$$

Example: Monomial Functions

- Let $f = \omega_1^{\kappa_1} \cdots \omega_d^{\kappa_d}$ and $\varphi = \omega_1^{\tau_1} \cdots \omega_d^{\tau_d}$.

$$\int_{[0,\varepsilon]^d} e^{-n\omega^\kappa} \omega^\tau d\omega = Cn^{-\lambda}(\log n)^{\theta-1}$$

- To find (λ, θ), we study the zeta function

$$\int_{[0,\varepsilon]^d} \omega^{-\kappa z+\tau} d\omega = \Big[\frac{\omega_1^{-\kappa_1 z+\tau_1+1}}{-\kappa_1 z + \tau_1 + 1}\Big]_0^\varepsilon \cdots \Big[\frac{\omega_d^{-\kappa_d z+\tau_d+1}}{-\kappa_d z + \tau_d + 1}\Big]_0^\varepsilon$$

Example: Monomial Functions

- Let $f = \omega_1^{\kappa_1} \cdots \omega_d^{\kappa_d}$ and $\varphi = \omega_1^{\tau_1} \cdots \omega_d^{\tau_d}$.

$$\int_{[0,\varepsilon]^d} e^{-n\omega^\kappa} \omega^\tau \, d\omega = C n^{-\lambda} (\log n)^{\theta-1}$$

- To find (λ, θ), we study the zeta function

$$\int_{[0,\varepsilon]^d} \omega^{-\kappa z+\tau} d\omega = \left[\frac{\omega_1^{-\kappa_1 z+\tau_1+1}}{-\kappa_1 z + \tau_1 + 1} \right]_0^\varepsilon \cdots \left[\frac{\omega_d^{-\kappa_d z+\tau_d+1}}{-\kappa_d z + \tau_d + 1} \right]_0^\varepsilon$$

- Thus, $\lambda = \min_i \left\{ \frac{\tau_i + 1}{\kappa_i} \right\}$, $\theta = \# \min_i \left\{ \frac{\tau_i + 1}{\kappa_i} \right\}$

  where $\# \min S$ is the number of times
  the minimum is attained in a set S.

Resolution of Singularities

Theorem (Hironaka)

Let f be a real analytic function at the origin with $f(0) = 0$.

Then, there exists a manifold M, a neighborhood W of the origin and a proper real analytic map $\rho : M \to W$ such that

- ρ is an isomorphism on $M \setminus (f \circ \rho)^{-1}(0)$
- $f \circ \rho$ and $|\rho'|$ are monomial functions locally at each $y \in (f \circ \rho)^{-1}(0)$

Resolution of Singularities

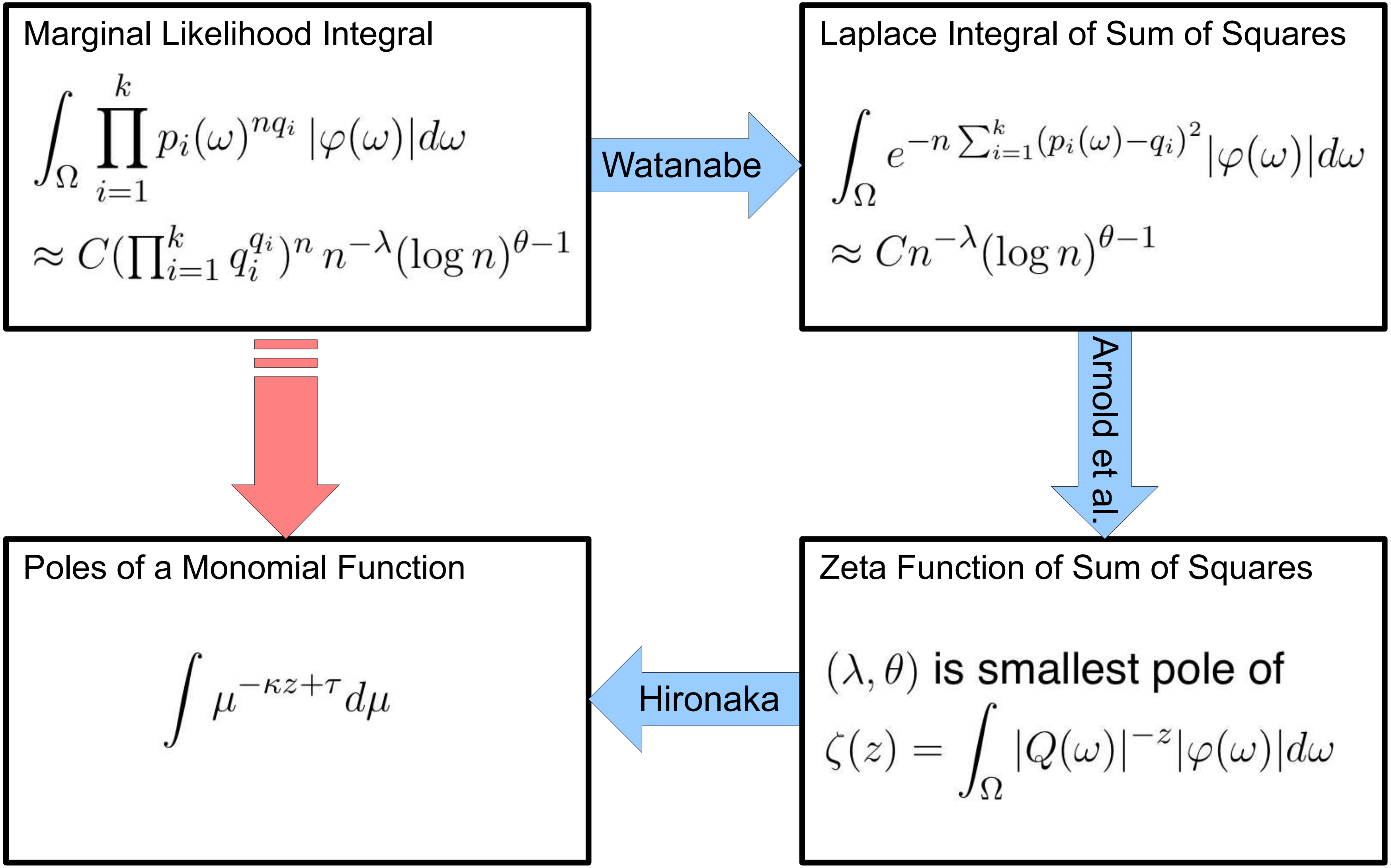
Theorem (Hironaka)

Let f be a real analytic function at the origin with $f(0) = 0$.

Then, there exists a manifold M, a neighborhood W of the origin and a proper real analytic map $\rho : M \to W$ such that

- ρ is an isomorphism on $M \setminus (f \circ \rho)^{-1}(0)$
- $f \circ \rho$ and $|\rho'|$ are monomial functions locally at each $y \in (f \circ \rho)^{-1}(0)$

Thus, we can find the poles of the zeta function of any f, provided we have a resolution of singularities for f.
Finding resolutions is generally a hard problem.


Marginal Likelihood Integral
$\int_\Omega \prod_{i=1}^{k} p_i(\omega)^{nq_i} \, |\varphi(\omega)| d\omega$
$\approx C(\prod_{i=1}^{k} q_i^{q_i})^n \, n^{-\lambda}(\log n)^{\theta-1}$
Watanabe
Laplace Integral of Sum of Squares
$\int_\Omega e^{-n\sum_{i=1}^{k}(p_i(\omega)-q_i)^2} |\varphi(\omega)| d\omega$
$\approx Cn^{-\lambda}(\log n)^{\theta-1}$
Arnold et al.
Zeta Function of Sum of Squares
(λ, θ) is smallest pole of
$\zeta(z) = \int_\Omega |Q(\omega)|^{-z} |\varphi(\omega)| d\omega$
Hironaka
Poles of a Monomial Function
$\int \mu^{-\kappa z+\tau} d\mu$

Real Log Canonical Thresholds

Real Log Canonical Thresholds

- $\Omega \subset \mathbb{R}^d$ compact semianalytic subset
  $\mathcal{A}_\Omega$ ring of real analytic functions on Ω
  $I = \langle f_1, \ldots, f_r \rangle \subset A_\Omega$, φ nearly analytic

Real Log Canonical Thresholds

- $\Omega \subset \mathbb{R}^d$ compact semianalytic subset
  $\mathcal{A}_\Omega$ ring of real analytic functions on Ω
  $I = \langle f_1, \ldots, f_r \rangle \subset A_\Omega$, φ nearly analytic
- Consider the zeta function

$$\zeta(z) = \int_\Omega \left(f_1(\omega)^2 + \cdots + f_r(\omega)^2\right)^{-z/2} |\varphi(\omega)| d\omega$$

Real Log Canonical Thresholds

- $\Omega \subset \mathbb{R}^d$ compact semianalytic subset
  $\mathcal{A}_\Omega$ ring of real analytic functions on Ω
  $I = \langle f_1, \ldots, f_r \rangle \subset A_\Omega$, φ nearly analytic
- Consider the zeta function

$$\zeta(z) = \int_\Omega \left(f_1(\omega)^2 + \cdots + f_r(\omega)^2\right)^{-z/2} |\varphi(\omega)| d\omega$$

- Define $\mathrm{RLCT}_\Omega(I;\varphi) = (\lambda, \theta)$ where λ is the smallest pole of $\zeta(z)$ and θ its multiplicity.
  If $\zeta(z)$ does not have any poles, set $(\lambda, \theta) = (\infty, \infty)$.

  Call λ the *real log canonical threshold* of $(I;\varphi)$ on Ω.

Fundamental Properties

- RLCT's are independent of the choice of generators $f_1, \ldots, f_r$ for the ideal I.

Fundamental Properties

- RLCT's are independent of the choice of generators $f_1, \ldots, f_r$ for the ideal I.
- RLCT's are positive rational numbers.

Fundamental Properties

- RLCT's are independent of the choice of generators $f_1, \ldots, f_r$ for the ideal I.

- RLCT's are positive rational numbers.

- RLCT's are local in nature.

$$\mathrm{RLCT}_{\Omega}(I;\varphi) = \min_{x \in \mathcal{V}(I)} \mathrm{RLCT}_{\Omega_x}(I;\varphi)$$

  where each Ω_x is a sufficiently small nbhd of x in Ω and $(\lambda_1, \theta_1) < (\lambda_2, \theta_2)$ if $\lambda_1 < \lambda_2$, or $\lambda_1 = \lambda_2$ and $\theta_1 > \theta_2$.

Fundamental Properties

- RLCT's are independent of the choice of generators $f_1, \dots, f_r$ for the ideal I.

- RLCT's are positive rational numbers.

- RLCT's are local in nature.

$$\mathrm{RLCT}_\Omega(I;\varphi) = \min_{x \in \mathcal{V}(I)} \mathrm{RLCT}_{\Omega_x}(I;\varphi)$$

  where each Ω_x is a sufficiently small nbhd of x in Ω and $(\lambda_1, \theta_1) < (\lambda_2, \theta_2)$ if $\lambda_1 < \lambda_2$, or $\lambda_1 = \lambda_2$ and $\theta_1 > \theta_2$.

- RLCT's depend on the boundary structure of Ω_x.

Fundamental Properties

- Formula for disjoint variables

$$\mathrm{RLCT}_{X_0\times Y_0}(I_x{+}I_y;\varphi_x\varphi_y) \quad = \quad (\lambda_x+\lambda_y,\theta_x+\theta_y-1)$$

$$\mathrm{RLCT}_{X_0\times Y_0}(I_xI_y;\varphi_x\varphi_y) \quad = \quad \begin{cases} (\lambda_x,\theta_x) & \text{if } \lambda_x<\lambda_y \\ (\lambda_y,\theta_y) & \text{if } \lambda_x>\lambda_y \\ (\lambda_x,\theta_x+\theta_y) & \text{if } \lambda_x=\lambda_y \end{cases}$$

Fundamental Properties

- Formula for disjoint variables

$$\begin{aligned}\mathrm{RLCT}_{X_0\times Y_0}(I_x+I_y;\varphi_x\varphi_y) &= (\lambda_x+\lambda_y,\theta_x+\theta_y-1)\\ \mathrm{RLCT}_{X_0\times Y_0}(I_xI_y;\varphi_x\varphi_y) &= \begin{cases}(\lambda_x,\theta_x) & \text{if } \lambda_x<\lambda_y\\ (\lambda_y,\theta_y) & \text{if } \lambda_x>\lambda_y\\ (\lambda_x,\theta_x+\theta_y) & \text{if } \lambda_x=\lambda_y\end{cases}\end{aligned}$$

- Formula for change of variables

$$\mathrm{RLCT}_{\Omega_0}(I;\varphi)=\min_{y\in\rho^{-1}(0)}\mathrm{RLCT}_{\rho^{-1}(\Omega_0)_y}(\rho^*I;(\varphi\circ\rho)|\rho'|)$$

Newton Polyhedra

Newton Polyhedra

- $\omega_1, \ldots, \omega_d$ local coordinates at the origin
  I an ideal of real analytic functions at the origin

  Given power series $f = \sum_\alpha c_\alpha \omega^\alpha$, define $[\omega^\alpha]f := c_\alpha$

Newton Polyhedra

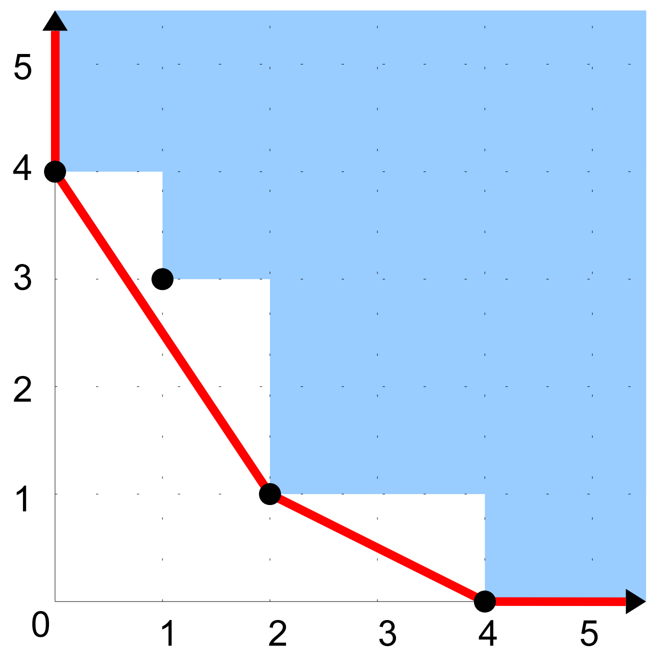
- $\omega_1, \ldots, \omega_d$ local coordinates at the origin
  I an ideal of real analytic functions at the origin

  Given power series $f = \sum_\alpha c_\alpha \omega^\alpha$, define $[\omega^\alpha]f := c_\alpha$

- The *Newton polyhedron* of I is the convex hull

$$\mathcal{P}(I) = \mathrm{conv}\{\alpha : f \in I, [\omega^\alpha]f \neq 0\}$$

$$I = \langle x^4 + x^2y + xy^3 + y^4 \rangle$$
$$J = \langle x^4, x^2y, xy^3, y^4 \rangle$$



Newton Polyhedra

- $\omega_1, \ldots, \omega_d$ local coordinates at the origin
  I an ideal of real analytic functions at the origin

  Given power series $f = \sum_\alpha c_\alpha \omega^\alpha$, define $[\omega^\alpha]f := c_\alpha$

- The *Newton polyhedron* of I is the convex hull

$$\mathcal{P}(I) = \mathrm{conv}\{\alpha : f \in I, [\omega^\alpha]f \neq 0\}$$

- $\tau = (\tau_1, \ldots, \tau_d)$ vector of non-negative integers

  The *distance* l_τ is the smallest t such that

$$t \cdot (\tau_1 + 1, \ldots, \tau_d + 1) \in \mathcal{P}(I)$$

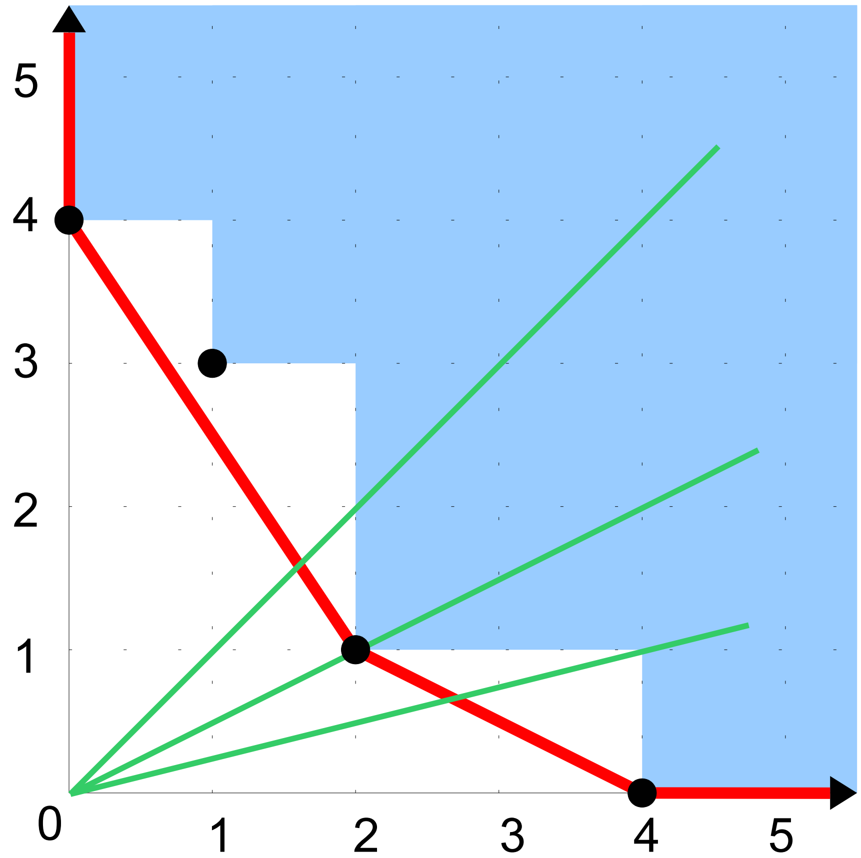
  The *multiplicity* θ_τ is the codimension
  of the face of $\mathcal{P}(I)$ at this intersection.

Example: Newton Polyhedra

$$I = \langle x^4 + x^2y + xy^3 + y^4 \rangle$$
$$J = \langle x^4, x^2y, xy^3, y^4 \rangle$$

Both I,J have the same Newton polyhedron.

$$l_{(0,0)} = \frac{8}{5}, \theta_{(0,0)} = 1$$
$$l_{(1,0)} = 1, \theta_{(1,0)} = 2$$
$$l_{(3,0)} = \frac{2}{3}, \theta_{(3,0)} = 1$$



Relation to RLCT

Theorem (L.)

Suppose the origin is not on the boundary of Ω.

Then, when φ is a monomial function ω^τ,

$$\mathrm{RLCT}_{\Omega_0}(I;\omega^\tau) \leq (1/l_\tau, \theta_\tau).$$

Equality holds when I is a monomial ideal.

Relation to RLCT

Theorem (L.)

Suppose the origin is not on the boundary of Ω.

Then, when φ is a monomial function ω^τ,

$$\mathrm{RLCT}_{\Omega_0}(I;\omega^\tau) \leq (1/l_\tau, \theta_\tau).$$

Equality holds when I is a monomial ideal.

Remark

Equality also holds for ideals which are *nondegenerate* (a term due to Varchenko).

Back to Schizophrenic Patients

Learning Coefficients

$P = (p_{ij})$, $S_i = \{$rank i matrices$\}$
$S_{21} = \{p_{11} = 0;\ p_{12}, p_{21}, p_{22}$ non-zero; up to perm$\} \subset S_2$
$S_{22} = \{p_{11} = p_{22} = 0;\ p_{12}, p_{21}$ non-zero; up to perm$\} \subset S_2$

Theorem (L.)

The learning coefficient (λ, θ) of the model is

$$(\lambda, \theta) = \begin{cases} (5/2, 1) & \text{if } P \in S_1, \\ (7/2, 1) & \text{if } P \in S_2 \setminus (S_{21} \cup S_{22}), \\ (4, 1) & \text{if } P \in S_{21} \setminus S_{22}, \\ (9/2, 1) & \text{if } P \in S_{22}. \end{cases}$$

Learning Coefficients

Proof

Four basic techniques:

1. Changing generators for the ideal
2. Change of variables formula
3. Disjoint variables formula
4. Newton polyhedra method

(systematically peeling an onion)

Complete First Term Asymptotics

Theorem (L.)

If f is nondegenerate, $f(0) = 0$ and $f \geq 0$, then

$$\int_{[0,1]^d} e^{-nf(\omega)} \omega^\tau d\omega \;\approx\; C\, n^{-\lambda} (\log n)^\theta$$

where $(\lambda, \theta) = (1/l_\tau, \theta_\tau)$ and

$$C = \frac{\Gamma(\lambda)}{(\theta - 1)!} \sum_{v \in \mathcal{F}} \prod_{i=1}^{\theta} (v\beta)_i^{-1} \int_{[0,1]^{d-\theta}} g_v(\bar{\mu}) d\bar{\mu}.$$

$\bar{\mu} = (\mu_{\theta+1}, \ldots, \mu_d)$, $\bar{e} = e_{\theta+1} + \ldots + e_d$,
$g_v(\bar{\mu})$ is $f(\mu^v)^{-\lambda} \mu^{v\tau + \alpha - \bar{e}}$ evaluated at $(0, \ldots, 0, \bar{\mu})$,
$\mathcal{F}$ is a unimodular refinement of the normal fan of $\mathcal{P}(f)$,
$\beta(v)$ is the vertex of $\mathcal{P}(f)$ corresponding to v,
$\alpha(v)$ is the vector of row sums of v.

Complete First Term Asymptotics

Proposition (L.)

If $\langle f_1(\omega), \ldots, f_k(\omega) \rangle$ is a monomial ideal,
then $f = \sum_i f_i(\omega)^2$ is nondegenerate.

Proposition (L.)

Let $q_1, \ldots, q_k$ be positive real numbers and
$p_1(\omega), \ldots, p_k(\omega)$ be positive real analytic functions
such that $p_i(\omega), q_i \geq 0$ and $\sum_i p_i(\omega) = \sum_i q_i = 1$.

If $\langle p_1(\omega) - q_i, \ldots, p_k(\omega) - q_k \rangle$ is a monomial ideal,
then $f = \sum_i q_i \log \frac{q_i}{p_i(\omega)}$ is nondegenerate.

Take Home

1. Compute asymptotics using *zeta functions*.

2. When computing learning coefficients,
   work with RLCT of *ideals* not *functions*.

3. Newton polyhedra methods can be extended to
   work with monomial *amplitude functions*.

Open Questions:

1. The RLCT over Ω is the minimum of RLCT's at $x \in \Omega$.
   How do we identify points with the minimum RLCT?

2. Is there a way to extend Newton polyhedra methods
   to cases where the origin is on the boundary of Ω?

Thank you for your kind attention :)

"Asymptotic Approximation of Marginal Likelihood Integrals"

Shaowei Lin

`http://arxiv.org/abs/1003.5338`

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

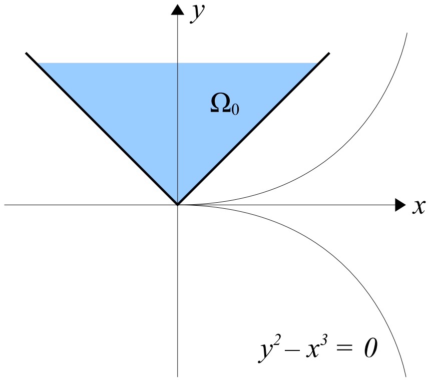
References

1. V. I. Arnol'd, S. M. Guseĭn-Zade and A. N. Varchenko: *Singularities of Differentiable Maps*, Vol. II, Birkhäuser, Boston, 1985.
2. M. Evans, Z. Gilula and I. Guttman: Latent class analysis of two-way contingency tables by Bayesian methods, *Biometrika* **76** (1989) 557–563.
3. 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.
4. S. Lin, B. Sturmfels and Z. Xu: Marginal likelihood integrals for mixtures of independence models, *J. Mach. Learn. Res.* **10** (2009) 1611–1631.
5. S. Lin: Asymptotic Approximation of Marginal Likelihood Integrals, in preparation.
6. S. Watanabe: *Algebraic Geometry and Statistical Learning Theory*, Cambridge Monographs on Applied and Computational Mathematics **25**, Cambridge University Press, Cambridge, 2009.

Example: Boundary Structure

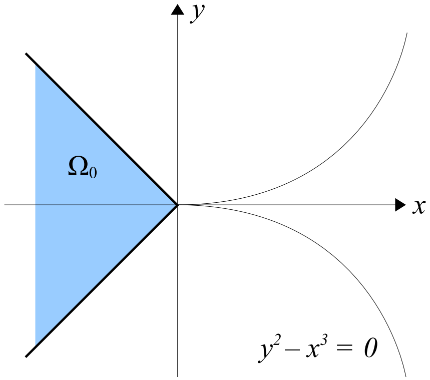
Let $I = \langle y^2 - x^3 \rangle$ and $\varphi = 1$.

- Case 1: $\Omega_0 = \{(x, y) \in \mathbb{R}^2 : 0 \le y \le \varepsilon, -y \le x \le y\}$



$$\mathrm{RLCT}_{\Omega_0}(I; \varphi) = (1, 1)$$

- Case 2: $\Omega_0 = \{(x, y) \in \mathbb{R}^2 : -\varepsilon \le x \le 0, x \le y \le -x\}$



$$\mathrm{RLCT}_{\Omega_0}(I; \varphi) = (\tfrac{5}{6}, 1)$$

Back

Disjoint Variables

- Suppose we have disjoint sets of variables

$$x = (x_1, \ldots, x_m) \qquad y = (y_1, \ldots, y_n)$$

$$I_x = \langle f_1(x), \ldots, f_r(x) \rangle \qquad I_y = \langle g_1(y), \ldots, g_s(y) \rangle$$

$$(\lambda_x, \theta_x) = \mathrm{RLCT}_{X_0}(I_x; \varphi_x) \quad (\lambda_y, \theta_y) = \mathrm{RLCT}_{Y_0}(I_y; \varphi_y)$$

- Recall $I_x + I_y = \langle f_i, g_j$ for all $i, j \rangle$, $I_x I_y = \langle f_i g_j$ for all $i, j \rangle$

Proposition

$$\mathrm{RLCT}_{X_0 \times Y_0}(I_x + I_y; \varphi_x \varphi_y) = (\lambda_x + \lambda_y, \theta_x + \theta_y - 1)$$

$$\mathrm{RLCT}_{X_0 \times Y_0}(I_x I_y; \varphi_x \varphi_y) = \begin{cases} (\lambda_x, \theta_x) & \text{if } \lambda_x < \lambda_y \\ (\lambda_y, \theta_y) & \text{if } \lambda_x > \lambda_y \\ (\lambda_x, \theta_x + \theta_y) & \text{if } \lambda_x = \lambda_y \end{cases}$$

Back

Change of Variables

- $I = \langle f_1, \ldots, f_r \rangle$
- ρ change of variables outside $\mathcal{V}(I)$

  i.e. $\rho : M \to W$ is a proper real analytic map from a manifold M to a neighborhood W of the origin that is an isomorphism on $M \setminus \rho^{-1}(\mathcal{V}(I))$
- $\rho^* I = \langle f_1 \circ \rho, \ldots, f_r \circ \rho \rangle$, $\mathcal{M} = \rho^{-1}(\Omega_0)$

Proposition

$$\mathrm{RLCT}_{\Omega_0}(I; \varphi) = \min_{y \in \rho^{-1}(0)} \mathrm{RLCT}_{\mathcal{M}_y}(\rho^* I; (\varphi \circ \rho)|\rho'|)$$

Back

Computation

Recall $p_{ij}(t,a,b,c,d) = ta_ib_j + (1-t)c_jd_j$.
Consider $t^* = \frac{1}{2}$ and $a^* = b^* = c^* = d^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.
Denote $\omega = (t,a,b,c,d)$ and $\omega^* = (t^*, a^*, b^*, c^*, d^*)$.

Let $I = \langle p_{ij}(\omega + \omega^*) - p_{ij}(\omega^*)\rangle$ and $\varphi = 1$.
We want to find $\mathrm{RLCT}_{\Omega_{\omega^*}}(I;\varphi)$.

Note that ω^* is not on the boundary of Ω.

Computation

Now, $\varphi = 1$ and I is generated by

$p_{ij}(\omega + \omega^*) - p_{ij}(\omega^*)$ for all $i, j \in \{1, 2, 3\}$

Computation

Now, $\varphi = 1$ and I is generated by

$p_{ij}(\omega + \omega^*) - p_{ij}(\omega^*)$ for all $i, j \in \{1, 2, 3\}$

Note that

$$p_{i1} + p_{i2} + p_{i3} = ta_i + tc_i =: p_{i0}$$
$$p_{1j} + p_{2j} + p_{3j} = tb_j + td_j =: p_{0j}$$

Let $g_{ij}(\omega)$ denote $p_{ij}(\omega + \omega^*) - p_{ij}(\omega^*)$.

Computation

Now, $\varphi = 1$ and I is generated by

$g_{ij}(\omega)$ for all $i, j \in \{0, 1, 2\}$

Computation

Now, $\varphi = 1$ and I is generated by

$g_{ij}(\omega)$ for all $i, j \in \{0, 1, 2\}$

For $i, j \in \{1, 2\}$, we replace $g_{ij}(\omega)$ with

$$g_{ij}(\omega) - (d_j + d_j^*)g_{i0}(\omega) - (a_i + a_i^*)g_{0j}(\omega)$$

Computation

Now, $\varphi = 1$ and I is generated by

$$g_{01}(\omega)$$
$$g_{02}(\omega)$$
$$g_{10}(\omega)$$
$$g_{20}(\omega)$$
$$g_{11}(\omega) - (d_1 + d_1^*)g_{10}(\omega) - (a_1 + a_1^*)g_{01}(\omega)$$
$$g_{12}(\omega) - (d_2 + d_2^*)g_{10}(\omega) - (a_1 + a_1^*)g_{02}(\omega)$$
$$g_{21}(\omega) - (d_1 + d_1^*)g_{20}(\omega) - (a_2 + a_2^*)g_{01}(\omega)$$
$$g_{22}(\omega) - (d_2 + d_2^*)g_{20}(\omega) - (a_2 + a_2^*)g_{02}(\omega)$$

Computation

Now, $\varphi = 1$ and I is generated by

$$g_{01}(\omega)$$
$$g_{02}(\omega)$$
$$g_{10}(\omega)$$
$$g_{20}(\omega)$$
$$g_{11}(\omega) - (d_1 + d_1^*)g_{10}(\omega) - (a_1 + a_1^*)g_{01}(\omega)$$
$$g_{12}(\omega) - (d_2 + d_2^*)g_{10}(\omega) - (a_1 + a_1^*)g_{02}(\omega)$$
$$g_{21}(\omega) - (d_1 + d_1^*)g_{20}(\omega) - (a_2 + a_2^*)g_{01}(\omega)$$
$$g_{22}(\omega) - (d_2 + d_2^*)g_{20}(\omega) - (a_2 + a_2^*)g_{02}(\omega)$$

Expanding these polynomials, we get...

Computation

Now, $\varphi = 1$ and I is generated by

$$c_1(\tfrac{1}{2} - t) + a_1(t + \tfrac{1}{2})$$
$$c_2(\tfrac{1}{2} - t) + a_2(t + \tfrac{1}{2})$$
$$d_1(\tfrac{1}{2} - t) + b_1(t + \tfrac{1}{2})$$
$$d_2(\tfrac{1}{2} - t) + b_2(t + \tfrac{1}{2})$$
$$a_1 d_1$$
$$a_1 d_2$$
$$a_2 d_1$$
$$a_2 d_2$$

Computation

Now, $\varphi = 1$ and I is generated by

$$c_1(\tfrac{1}{2} - t) + a_1(t + \tfrac{1}{2})$$
$$c_2(\tfrac{1}{2} - t) + a_2(t + \tfrac{1}{2})$$
$$d_1(\tfrac{1}{2} - t) + b_1(t + \tfrac{1}{2})$$
$$d_2(\tfrac{1}{2} - t) + b_2(t + \tfrac{1}{2})$$
$$a_1 d_1$$
$$a_1 d_2$$
$$a_2 d_1$$
$$a_2 d_2$$

Substitute $b_i = \frac{b_i' - d_i(\frac{1}{2} - t)}{t + \frac{1}{2}}$, $c_i = \frac{c_i' - a_i(t + \frac{1}{2})}{\frac{1}{2} - t}$.

The Jacobian determinant of this change of variable is 16.

Computation

Now, $\varphi = 16$ and I is generated by

$$c_1', c_2', b_1', b_2', a_1d_1, a_1d_2, a_2d_1, a_2d_2$$

Computation

Now, $\varphi = 16$ and I is generated by

$$c_1', c_2', b_1', b_2', a_1 d_1, a_1 d_2, a_2 d_1, a_2 d_2$$

This is a monomial ideal so we may use the Newton polyhedra method to compute its RLCT.

Alternatively, we can apply the formula for disjoint variables.

$$I \;=\; \langle c_1' \rangle + \langle c_2' \rangle + \langle b_1' \rangle + \langle b_2' \rangle + \Big(\langle a_1 \rangle + \langle a_2 \rangle\Big)\Big(\langle d_1 \rangle + \langle d_2 \rangle\Big)$$

Computation

Now, $\varphi = 16$ and I is generated by

$$c_1', c_2', b_1', b_2', a_1d_1, a_1d_2, a_2d_1, a_2d_2$$

This is a monomial ideal so we may use the Newton polyhedra method to compute its RLCT.

Alternatively, we can apply the formula for disjoint variables.

$$I \;=\; \langle c_1' \rangle + \langle c_2' \rangle + \langle b_1' \rangle + \langle b_2' \rangle + \Big(\langle a_1 \rangle + \langle a_2 \rangle \Big)\Big(\langle d_1 \rangle + \langle d_2 \rangle \Big)$$

Conclusion: $\mathrm{RLCT}_{\Omega_{\omega^*}}(I; \varphi) = (6, 2)$

Back