Asymptotic Approximation of Marginal Likelihood Integrals

Shaowei Lin

shaowei@math.berkeley.edu

21 Apr 2009, University of Pennsylvania

The Statistics

Occasionally Dishonest Coin-Tosser

- *The Deal*:
  Four coin tosses. If all are equal, you lose.

Occasionally Dishonest Coin-Tosser

- *The Deal*:
  Four coin tosses. If all are equal, you lose.
- *The Scam*:
  Two coins are involved, one fair and one biased.

Occasionally Dishonest Coin-Tosser

- *The Deal*:
  Four coin tosses. If all are equal, you lose.
- *The Scam*:
  Two coins are involved, one fair and one biased.
- *The Data*:

#Heads	0	1	2	3	4
#Occurrences	51	18	73	25	75

Occasionally Dishonest Coin-Tosser

- *The Deal*:
  Four coin tosses. If all are equal, you lose.
- *The Scam*:
  Two coins are involved, one fair and one biased.
- *The Data*:

#Heads	0	1	2	3	4
#Occurrences	51	18	73	25	75

- *The Burning Question*:
  How many coins did he use?

Occasionally Dishonest Coin-Tosser

- Model *One*:

Parameters	Coin:	$0 \leq \theta_h, \theta_t \leq 1,\ \theta_h + \theta_t = 1$
Prob(i heads)	p_i =	$\binom{4}{i}\theta_h^i \theta_t^{4-i}$
Likelihood of data U	$L_U(\theta)$ =	$z p_0^{51} p_1^{18} p_2^{73} p_3^{25} p_4^{75}$
	where z =	$242!/(51! \cdot 18! \cdot 73! \cdot 25! \cdot 75!)$

Occasionally Dishonest Coin-Tosser

- Model *One*:

Parameters	Coin:	$0 \leq \theta_h, \theta_t \leq 1,\ \theta_h + \theta_t = 1$
Prob(i heads)	p_i =	$\binom{4}{i}\theta_h^i \theta_t^{4-i}$
Likelihood of data U	$L_U(\theta)$ =	$zp_0^{51}p_1^{18}p_2^{73}p_3^{25}p_4^{75}$
	where z =	$242!/(51! \cdot 18! \cdot 73! \cdot 25! \cdot 75!)$

- Model *Two*:

Parameters	Coin 0:	$0 \leq \theta_h, \theta_t, \leq 1,\ \ \theta_h + \theta_t = 1$
	Coin 1:	$0 \leq \rho_h, \rho_t \leq 1,\ \ \rho_h + \rho_t = 1$
	Choice of coin:	$0 \leq \sigma_0, \sigma_1 \leq 1,\ \ \sigma_0 + \sigma_1 = 1$
Prob(i heads)	p_i =	$\binom{4}{i}(\sigma_0\theta_h^i\theta_t^{4-i} + \sigma_1\rho_h^i\rho_t^{4-i})$
Likelihood of data U	$L_U(\theta)$ =	$zp_0^{51}p_1^{18}p_2^{73}p_3^{25}p_4^{75}$

Statistical Models

- **Question:** How do we do model selection?

Statistical Models

- **Question:** How do we do model selection?
- **Method 1:** Maximum Likelihood

  Compare the maximum values of the likelihood functions.

$$\max_{\theta \in \Theta} L_U(\theta)$$

Statistical Models

- **Question:** How do we do model selection?
- **Method 1:** Maximum Likelihood

  Compare the maximum values of the likelihood functions.

$$\max_{\theta \in \Theta} L_U(\theta)$$

- **Method 2:** Marginal Likelihood

  Integrate the likelihood functions over the parameter space.

$$\int_{\Theta} L_U(\theta) d\theta$$

Statistical Models

Discrete Model
State space $[k] = \{1, 2, \ldots, k\}$
Compact parameter space $\Omega \subset \mathbb{R}^d$
Polynomial map $p = (p_i)$, $p : \Omega \to \Delta_{k-1}$
Vector of counts $u = (u_i)$ with sample size n

Marginal Likelihood Integral

$$Z_n(u) = \int_\Omega \prod_{i=1}^{k} p_i(\omega)^{u_i} d\omega$$

Previous Work
Computed $Z_n(u)$ exactly for small samples.
(L.-Sturmfels-Xu 2008)

Statistical Models

Discrete Model
State space $[k] = \{1, 2, \ldots, k\}$
Compact parameter space $\Omega \subset \mathbb{R}^d$
Polynomial map $p = (p_i)$, $p : \Omega \to \Delta_{k-1}$
Vector of counts $u = (u_i)$ with sample size n

Marginal Likelihood Integral

$$Z_n(u) = \int_\Omega \prod_{i=1}^{k} p_i(\omega)^{u_i} d\omega$$

Goal
Assume sample drawn from *true distribution* $q \in \operatorname{Im} p$.
Find asymptotics of $\mathrm{E}[\log Z_n(U)]$ as $n \to \infty$.

The Link

Asymptotic Approximation

Notations

1. Define $Q(\omega) = \|p(\omega) - q\|^2 = \sum_{i=1}^{k} (p_i(\omega) - q_i)^2$.
2. Given $x \in \text{fiber}(q) = \{\omega : p(\omega) = q\}$,
   let $\lambda_x \in \mathbb{Q}_+$ be the smallest pole of the *zeta function*

   $$J_x(z) = \int_{\Omega_x} Q(\omega)^{-2z} d\omega$$

   for a sufficiently small neighborhood Ω_x of x.
   Let $\theta_x \in \mathbb{Z}_+$ be the order of this pole.
3. Define an ordering on $\mathbb{Q}_+ \times \mathbb{Z}_+$:
   $(\lambda_2, \theta_2) > (\lambda_1, \theta_1) \; \Leftrightarrow \; \lambda_2 > \lambda_1$, or $\lambda_1 = \lambda_2$ and $\theta_2 < \theta_1$.

Asymptotic Approximation

Theorem (based on Watanabe 2001)

$$\mathrm{E}[\log Z_n] = n \sum_{i=1}^{k} q_i \log q_i - 2\lambda \log n + (\theta - 1) \log \log n + O(1)$$

where (λ, θ) is *smallest* among all (λ_x, θ_x), $x \in \mathrm{fiber}(q)$.

Remarks

1. λ is the *real log canonical threshold* (RLCT) of the Q.
2. The (λ_x, θ_x) can be found using *local* resolution of singularities, which exists by (Atiyah 1970).

The Log Canonical Threshold

Log Canonical Threshold

Given an ideal $I = \langle f_1, \ldots, f_k \rangle \subseteq \mathbb{C}[\omega_1, \ldots, \omega_d]$, the log canonical threshold $\mathrm{lct}_x(I)$ of I at a point $x \in \mathcal{V}(I) \subseteq \mathbb{C}^d$ is the smallest pole of the zeta function

$$J_x(z) = \int_{\Omega_x} \left(|f_1|^2 + \ldots + |f_k|^2 \right)^{-z} d\omega$$

for a sufficiently small neighborhood Ω_x of x. This pole is independent of the choice of generators f_i for I.

Alternative Formulations: using multiplier ideals, Bernstein-Sato polynomials.

Log Canonical Threshold

Given an ideal $I = \langle f_1, \ldots, f_k \rangle \subseteq \mathbb{R}[\omega_1, \ldots, \omega_d]$, the *real* log canonical threshold $\mathrm{rlct}_x(I)$ of I at a point $x \in \mathcal{V}(I) \subseteq \mathbb{R}^d$ is the smallest pole of the zeta function

$$J_x(z) = \int_{\Omega_x} \left(f_1^2 + \ldots + f_k^2\right)^{-z} d\omega$$

for a sufficiently small neighborhood Ω_x of x. This pole is independent of the choice of generators f_i for I.

Remark: For $Q(\omega) = \sum_{i=1}^{k} (p_i(\omega) - q_i)^2$,

$$2\,\mathrm{rlct}_x(Q) = \mathrm{rlct}_x(p_1 - q_1, \ldots, p_k - q_k).$$

Resolution of Singularities

Theorem (Hironaka, Atiyah 1970)

Suppose $\Omega \subset \mathbb{R}^d$ nbhd of origin, $f : \Omega \to \mathbb{R}$ polynomial, $f(0) = 0$.

Then, there exist $W \subset \Omega$ open nbhd of origin, $\mathcal{M}$ smooth variety of dim d and $g : \mathcal{M} \to W$ proper birational map with isomorphism on $\mathcal{M} \setminus (fg)^{-1}(0)$ such that:

For any $P \in (fg)^{-1}(0)$, $\exists$ local coords $\mu = (\mu_1, \mu_2, \ldots \mu_d)$ with $P = (0, \ldots, 0)$ and

$$f(g(\mu)) = \pm\mu_1^{\sigma_1}\mu_2^{\sigma_2}\cdots\mu_d^{\sigma_d} = \pm\mu^{\sigma}, \quad \sigma_1, \sigma_2, \ldots, \sigma_d \in \mathbb{Z}_{\geq 0}.$$

Furthermore, the Jacobian determinant of g equals

$$|g'(\mu)| = h(u)\mu_1^{\tau_1}\mu_2^{\tau_2}\cdots\mu_d^{\tau_d} = h(\mu)\mu^{\tau}, \quad \tau_1, \tau_2, \ldots, \tau_d \in \mathbb{Z}_{\geq 0}$$

where $h(\mu)$ is a non-vanishing rational function.

Resolution of Singularities

Using a partition of unity argument, we can show that $\operatorname{rlct}_0(f)$ is the smallest of the poles λ_P of the zeta function

$$J_P(z) = \int_{\mathcal{M}_P} f(g(\mu))^{-2z} |g'(\mu)| d\mu = \int_{\mathcal{M}_P} \mu^{-2z\sigma+\tau} h(\mu) d\mu$$

for a sufficiently small neighborhood $\mathcal{M}_P$ of P,
as P varies over $(fg)^{-1}(0)$.

From the above equation, we get $\lambda_P = \min_{1 \le j \le d} \frac{\tau_j + 1}{2\sigma_j}$.

BIG Problem: How to find a resolution of singularities?

Given $x \in \mathrm{fiber}(q)$, how do we compute (λ_x, θ_x)?

Newton Diagrams

Let $Q(\omega) = \sum_\alpha c_\alpha \omega^\alpha$ be a polynomial in d variables.

Newton polyhedron $\Gamma_+(Q)$: $\operatorname{conv}(\{\alpha + \alpha' : c_\alpha \neq 0, \alpha' \in \mathbb{R}^d_{\geq 0}\})$.

Face polynomial $Q_\gamma(\omega)$: $\sum_{\alpha \in \gamma} c_\alpha \omega^\alpha$, γ compact face.

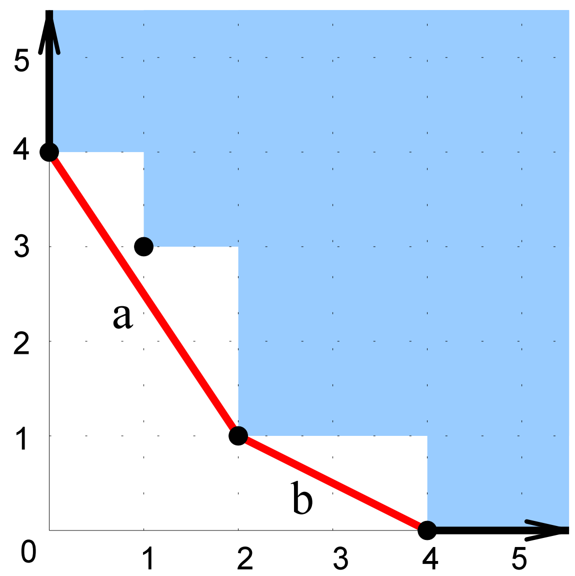
Newton diagram $\Gamma(Q)$: union of all compact faces.

Principal part $Q_\Gamma(\omega)$: $\sum_{\alpha \in \Gamma(Q)} c_\alpha \omega^\alpha$.

Example

$$Q(x,y) = x^4 + x^2y + xy^3 + y^4$$

$$Q_a(x,y) = x^2y + y^4$$
$$Q_b(x,y) = x^4 + x^2y$$
$$Q_\Gamma(x,y) = x^4 + x^2y + y^4$$



Non-degeneracy

$Q(\omega)$ is *non-degenerate* if for all compact faces γ,

$$\mathcal{V}\left(\frac{\partial Q_\gamma}{\partial \omega_1}, \frac{\partial Q_\gamma}{\partial \omega_2}, \ldots, \frac{\partial Q_\gamma}{\partial \omega_d}\right) \quad \subseteq \quad \mathcal{V}(\omega_1 \omega_2 \cdots \omega_d)$$

Example

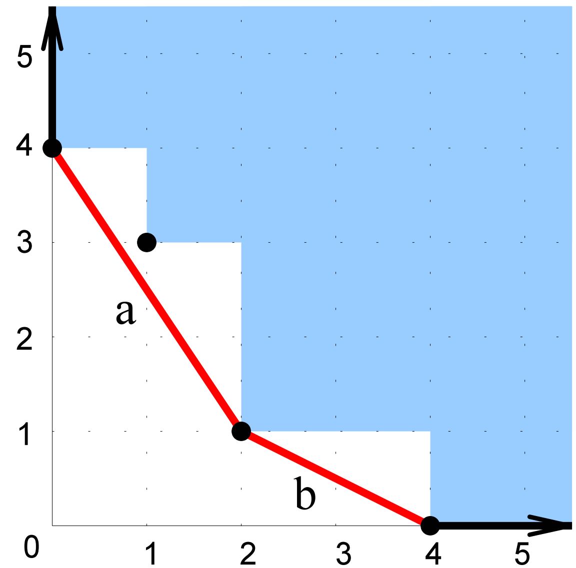
$Q(x,y) = x^4 + x^2 y + x y^3 + y^4$ is non-degenerate.
$Q(x,y) = (x+y)^2$ is degenerate.

We now present some tools which allow us to compute resolution of singularities using Newton diagrams.

Proposition 1

Suppose $Q(\omega)$ is non-degenerate.

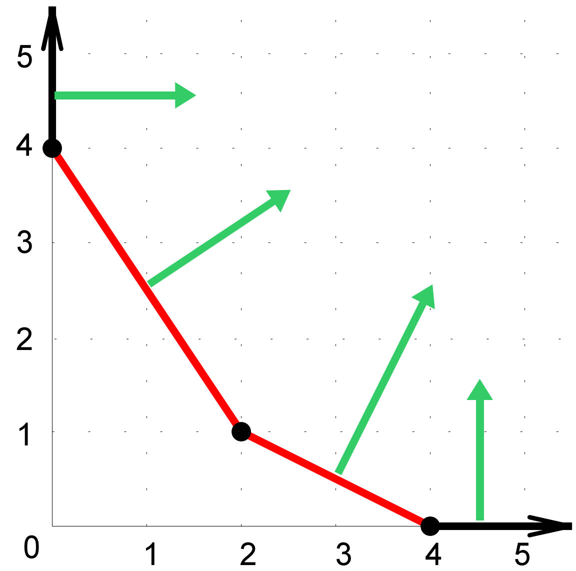
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.



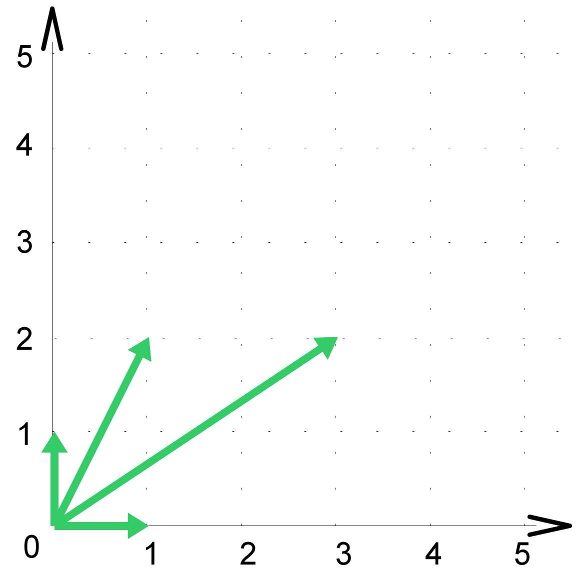
Proposition 1

Suppose $Q(\omega)$ is non-degenerate.

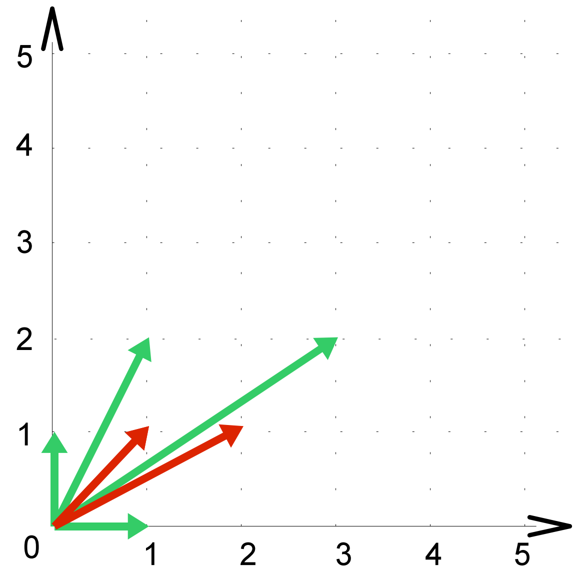
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.



Proposition 1
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.

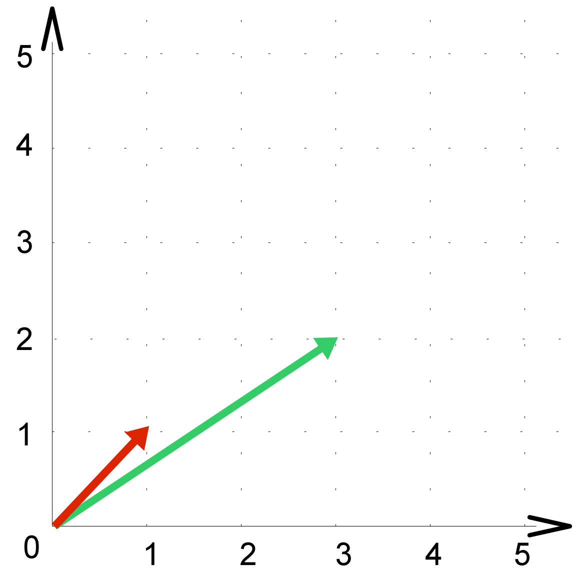


Proposition 1
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.



Proposition 1

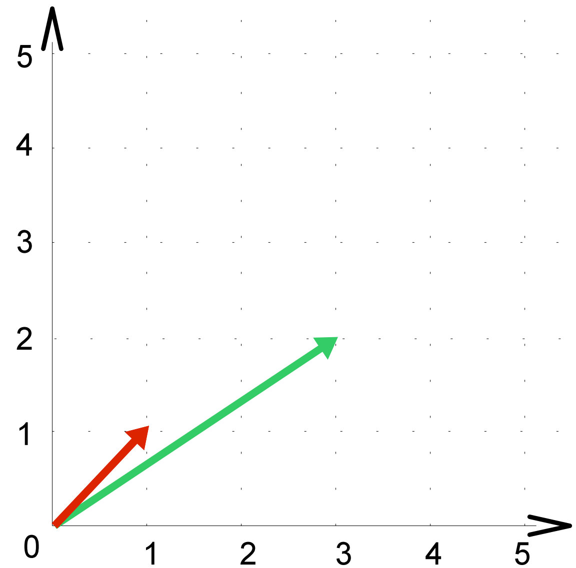
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.



Proposition 1
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.

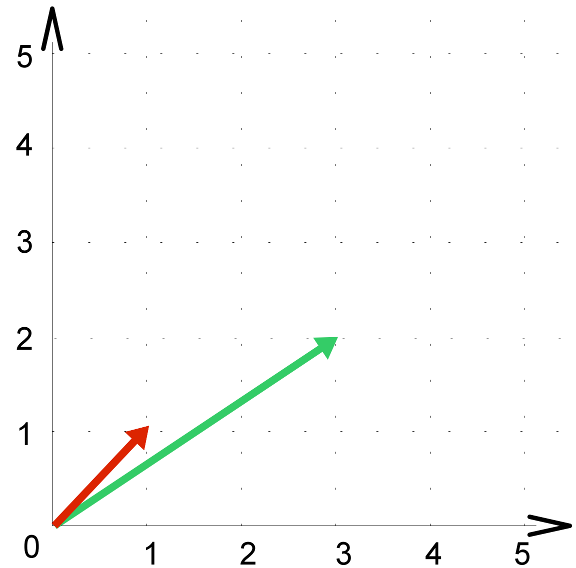
$$Q(x,y) = x^4 + x^2y + xy^3 + y^4$$

$$\begin{aligned} x &= u^3v^1 \\ y &= u^2v^1 \\ Q(u,v) &= u^8v^3(1+v+uv+u^4v) \end{aligned}$$

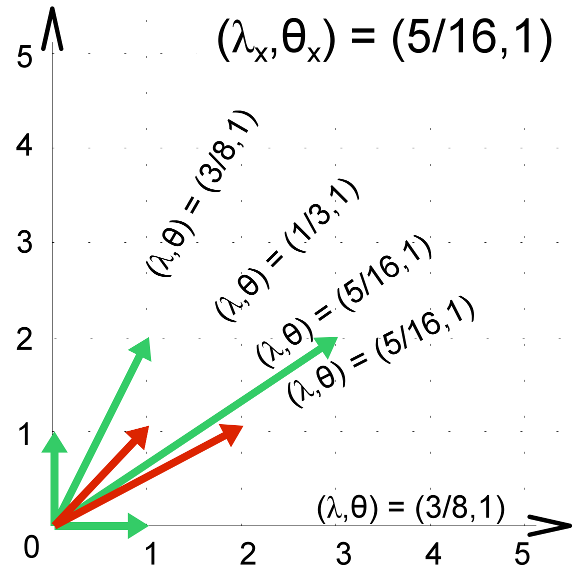


Proposition 1
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.

$$\begin{aligned}
Q(x,y) &= x^4 + x^2y + xy^3 + y^4 \\
J_0(z) &= \int Q(x,y)^{-2z}\ dxdy \\
&= \int Q(u,v)^{-2z}\, u^4 v\ dudv \\
&= \int u^{-16z+4} v^{-6z+1} \\
&\qquad (1+v+uv+u^4v)^{-2z}\ dudv \\
(\lambda,\theta) &= (\tfrac{5}{16},\ 1)
\end{aligned}$$



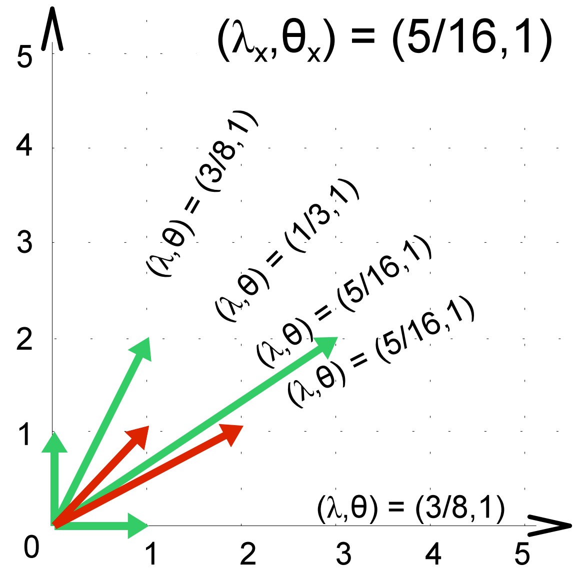
Proposition 1
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.



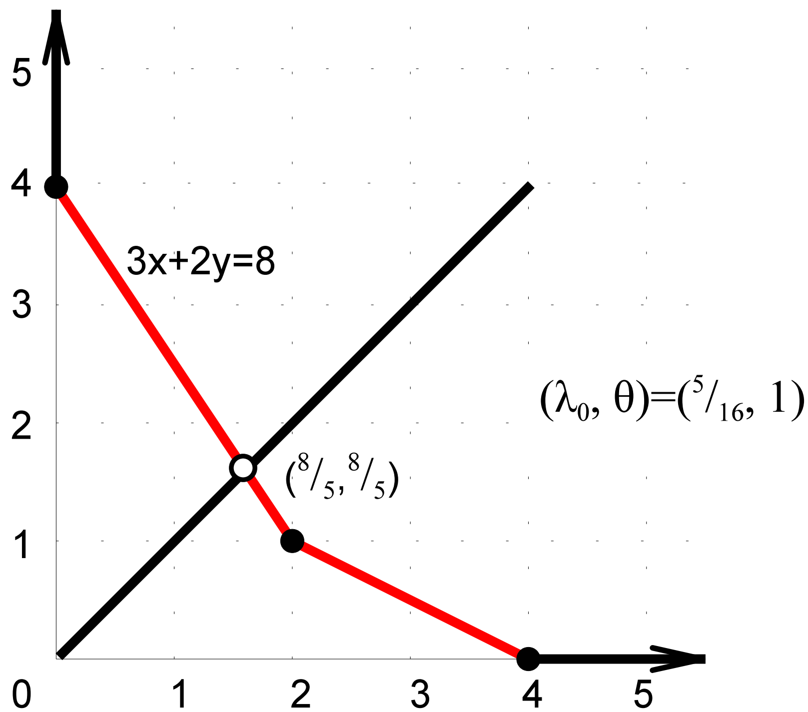
Proposition 1
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.

The regular subdivision describes a resolution map $g : \mathcal{M} \to W$, where $\mathcal{M}$ is a toric variety and W is an open nbhd of the origin.

In particular, each maximal cone gives a chart map $U \to W$ defined by monomials.



Proposition 2
Suppose $Q(\omega)$ is non-degenerate.
Let the intersection of the diagonal $\{(t,\ldots,t), t \in \mathbb{R}\}$ with the boundary of $\Gamma_+(Q)$ be $(\frac{1}{\lambda},\ldots,\frac{1}{\lambda})$.
Then, $\mathrm{rlct}_0(Q) = \lambda/2$.



Example

Discrete model

$$p_{ij}(a,b,c,d,e) = ab_ic_j + (1-a)d_ie_j, \quad i,j = 1,2,3$$

True distribution

$$q_{11} = q_{12} = \ldots = q_{33} = \frac{1}{9}$$

$$Q(a,b,c,d,e) = \sum_{i,j}(ab_ic_j + (1-a)d_ie_j - \tfrac{1}{9})^2$$

Pick $x = (a^*, b^*, c^*, d^*, e^*) \in \text{fiber}(q)$.

Shift the origin to x.

$$\begin{aligned} Q(a,b,c,d,e) = \textstyle\sum_{i,j}[&(a+a^*)(b_i+b_i^*)(c_j+c_j^*) \\ &+(1-a-a^*)(d_i+d_i^*)(e_j+e_j^*) - \tfrac{1}{9})]^2 \end{aligned}$$

Example

$Q = \sum_{i,j} [(a + a^*)(b_i + b_i^*)(c_j + c_j^*) + (1 - a - a^*)(d_i + d_i^*)(e_j + e_j^*) - \frac{1}{9}]^2$

Toric Modification

Suppose $a^* = 0$, $b^* = c^* = (0, 0, 1)$, $d^* = e^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

Principal part $Q_\Gamma = \frac{2}{3}(d_1 d_2 + e_1 e_2 + d_1^2 + d_2^2 + e_1^2 + e_2^2)$
$-\frac{2}{3} a(d_1 + d_2 + e_1 + e_2) + \frac{8}{9} a^2$

From this, one can check that Q is non-degenerate.

Using the method of Newton diagrams, we derive

$$(\lambda_x, \theta_x) = (\frac{5}{4}, 1)$$

Comparison with Exact Evaluation

Model $\quad p_i(\sigma,\theta,\rho) = \binom{4}{i}(\sigma_0\theta_0^i\theta_1^{4-i} + \sigma_1\rho_0^i\rho_1^{4-i})$, $\quad 0 \le i \le 4$

True Distribution $\quad q_i = \frac{1}{16}\binom{4}{i}$

Asymptotics $\quad \log Z_n = n\sum_i q_i \log q_i - \frac{3}{4}\log n + O(1)$

We compute $G(n) = 16\sum_i q_i \log q_i + \log Z_n - \log Z_{16+n}$.

We expect it to be close to $g(n) = \frac{3}{4}(\log(16+n) - \log n)$.

n	$G(n)$	$g(n)$
16	0.21027043	0.225772497
32	0.12553837	0.132068444
48	0.08977938	0.093704053
64	0.06993586	0.072682510
80	0.05729553	0.059385934
96	0.04853292	0.050210092
112	0.04209916	0.043493960

Open Questions

1. What is the relationship between complex log canonical thresholds and real log canonical thresholds? How does the radicality of the ideal affect the thresholds?

2. Which higher order terms in a polynomial can we discard so as not to affect the thresholds? (*Kollar’s Perturbation Theorem*?)

3. Is there a way to read off the (real) log canonical threshold from the tropicalization of the variety?

4. How do we identify the most complicated singularity (i.e. smallest log canonical threshold) on the variety? (*Whitney Stratification?*)

References

1. V. I. Arnold, V. V. Goryunov, O. V. Lyashko and V. A. Vasil'ev: *Singularity Theory I*, Springer, 1998.
2. S. Lin, B. Sturmfels and Z. Xu: Marginal Likelihood Integrals for Mixtures of Independence Models, `arXiv:0805.3602`, submitted to *Journal of Machine Learning Research*.
3. S. Watanabe: Algebraic Analysis for Nonidentifiable Learning Machines, *Neural Computation* **13** (2001) 899–933.
4. K. Yamazaki and S. Watanabe: Newton Diagram and Stochastic Complexity in Mixture of Binomial Distributions, *Algorithmic Learning Theory* (2004) 350–364.