Asymptotic Approximation of Marginal Likelihood Integrals

Shaowei Lin

shaowei@math.berkeley.edu

17 Dec 2008, MSRI Algebraic Statistics Workshop

Introduction

Discrete Model
State space $[k] = \{1, 2, \ldots, k\}$
Compact parameter space $\Omega \subset \mathbb{R}^d$
Polynomial map $p = (p_i)$, $p : \Omega \to \Delta_{k-1}$

Example
Two-mixture of the independence model
with two ternary random variables

$$p_{ij}(a, b, c, d, e) = ab_i c_j + (1-a) d_i e_j, \quad i, j = 1, 2, 3$$

where $(a, 1-a) \in \Delta_1$, $(b_i), (c_j), (d_i), (e_j) \in \Delta_2$
and $\Omega = \Delta_1 \times \Delta_2 \times \Delta_2 \times \Delta_2 \times \Delta_2$.

Introduction

Let $u = (u_i)$ be a vector of counts for a sample of size n.

Marginal Likelihood Integral

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

Plays important role in various practical applications, especially model selection.

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

Introduction

Let $u = (u_i)$ be a vector of counts for a sample of size n.

Marginal Likelihood Integral

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

Plays important role in various practical applications, especially model selection.

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

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 largest pole of the *zeta function*

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

   for a sufficiently small neighborhood Ω_x of x.
   Let m_x be the multiplicity of this pole.
3. Define an ordering on $\mathbb{Q}_{>0} \times \mathbb{Z}_{>0}$:
   $(\lambda_2, m_2) > (\lambda_1, m_1) \;\Leftrightarrow\; \lambda_2 > \lambda_1$, or $\lambda_1 = \lambda_2$ and $m_2 < m_1$.

Asymptotic Approximation

Theorem (based on Watanabe 2001)

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

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

Remarks

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

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

Newton Diagrams

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

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

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

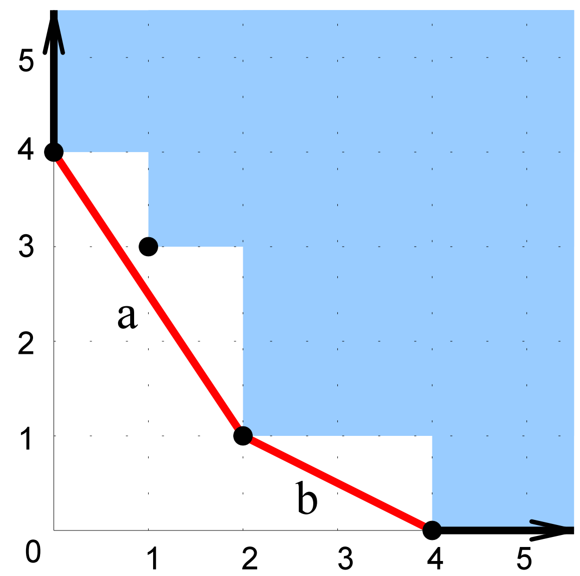
Example

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

$$Q_a(x, y) = x^2 y + y^4$$

$$Q_b(x, y) = x^4 + x^2 y$$

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



Non-degeneracy

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

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

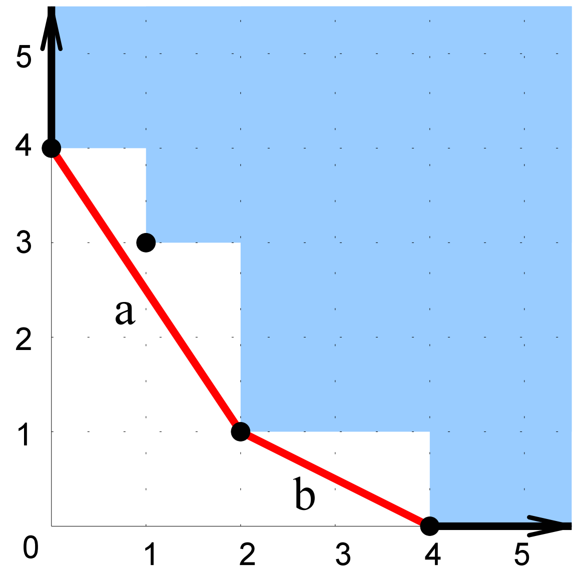
Example

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

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

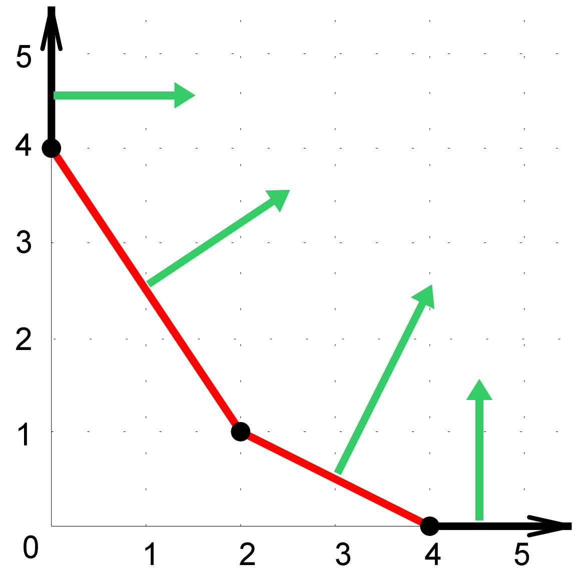
Proposition 1

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.



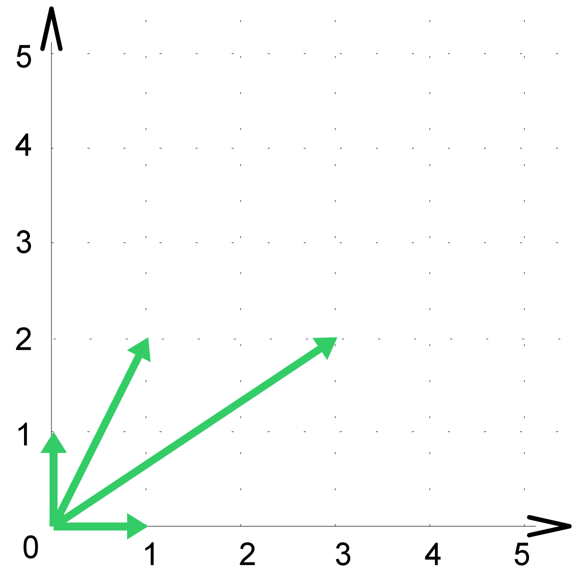
Proposition 1

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

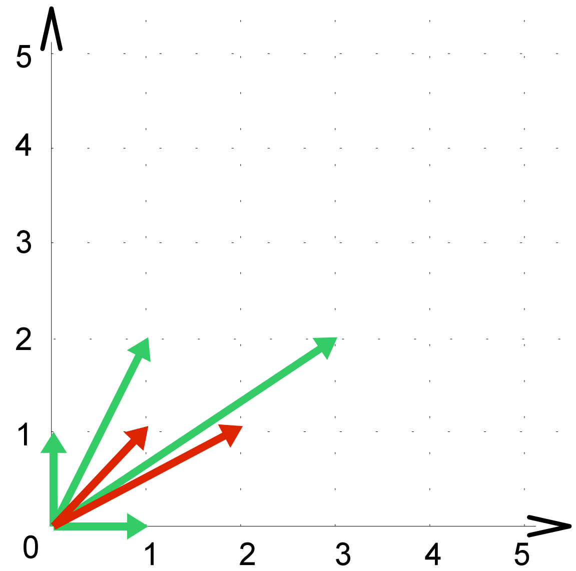
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.



Proposition 1

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

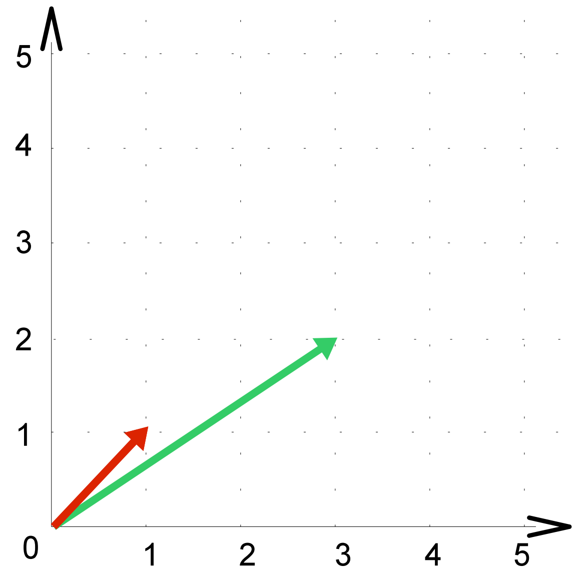
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.



Proposition 1

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

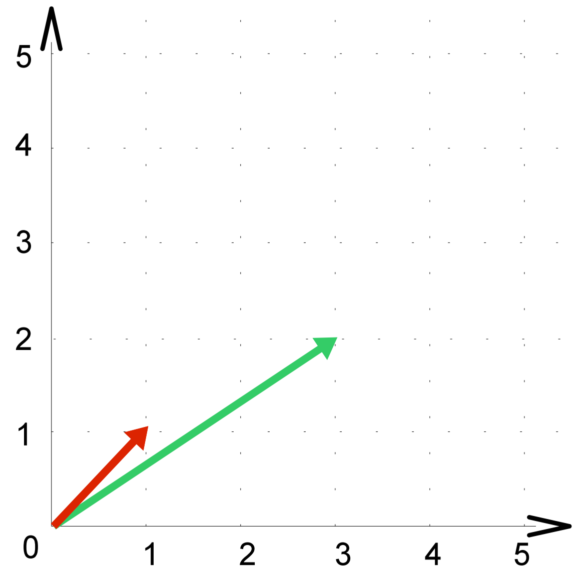
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.



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.

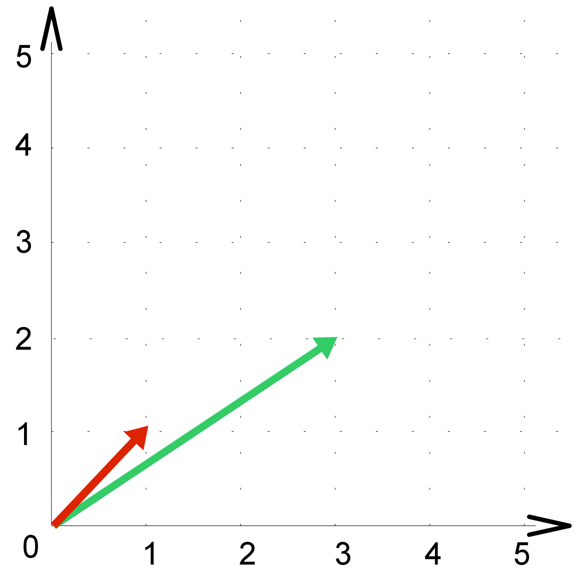
$$Q(x,y) \quad = \quad x^4 + x^2y + xy^3 + y^4$$

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



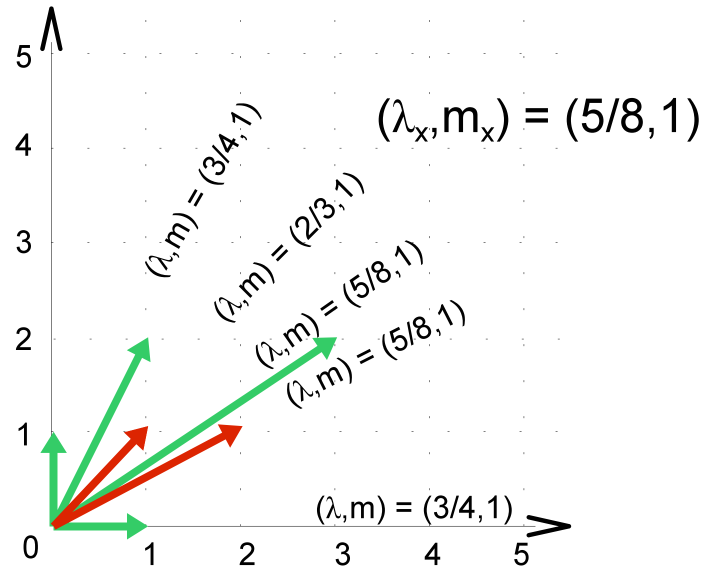
Proposition 1
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.

$$
\begin{aligned}
Q(x,y) &= x^4 + x^2y + xy^3 + y^4 \\
J_0(z) &= \int Q(x,y)^z \; dxdy \\
&= \int Q(u,v)^z \, u^4 v \; dudv \\
&= \int u^{8z+4} v^{3z+1} \\
&\quad\; (1+v+uv+u^4v)^z \; dudv \\
(\lambda, m) &= (\tfrac{5}{8},\; 1)
\end{aligned}
$$



Proposition 1

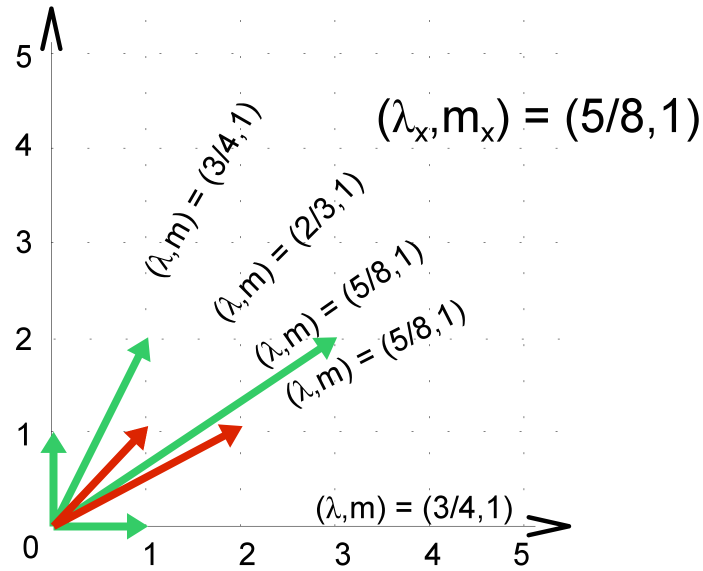
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.



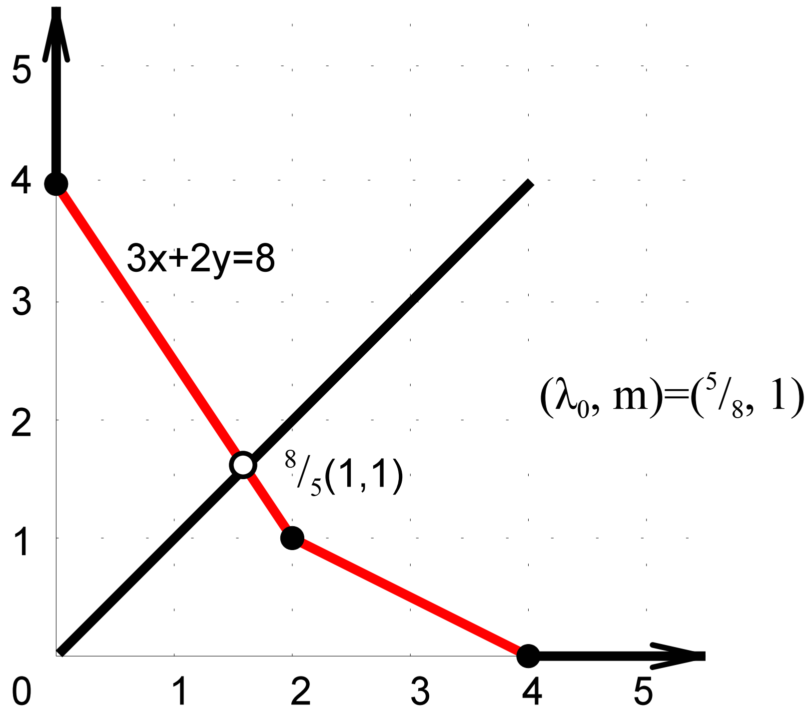
Proposition 1
Suppose $Q(\omega)$ is non-degenerate.
A regular subdivision of the normal fan of $\Gamma_+(Q)$ describes a local resolution of singularities at the origin via toric modifications.

The regular subdivision describes a resolution map $g : \mathcal{M} \to W$, where $\mathcal{M}$ is a manifold 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.
The RLCT λ_0 of $Q(\omega)$ comes from
the intersection $\frac{1}{\lambda_0}(1,1,\ldots,1) = \frac{1}{\lambda_0}\mathbf{1}$
of the diagonal $\{t\mathbf{1}\}$ with the boundary of $\Gamma_+(Q)$.



Proposition 3
Regardless of the degeneracy of $Q(\omega)$,
the RLCT and its multiplicity (λ_0, m_0) of $Q(\omega)$
is the same as that of the principal part $Q_\Gamma(\omega)$.

(Afternote: this proposition is false.
I will address this in a new paper.)

$$Q(\omega) = x^4 + x^2 y + x y^3 + y^4$$

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

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

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

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

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

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

From this, one can check that Q is non-degenerate.
Using the method of Newton diagrams, we derive

$$(\lambda_x, m_x) = (\frac{5}{2}, 1)$$

Example

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

Non-toric Modification

Suppose $a^* = \frac{1}{2}$, $b^* = c^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $d^* = e^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

Principal part $Q_\Gamma = \frac{1}{6}(b_1^2 + b_2^2 + c_1^2 + c_2^2 + d_1^2 + d_2^2 + e_1^2 + e_2^2$

$$+ b_1 b_2 + c_1 c_2 + d_1 d_2 + e_1 e_2 + c_1 e_2 + c_2 e_1 + b_1 d_2 + b_2 d_1)$$

$$+ \tfrac{1}{3}(b_1 d_1 + b_2 d_2 + c_1 e_1 + c_2 e_2)$$

From this, one can check that Q is degenerate.

We do a linear change of variable.

Example

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

Non-toric Modification

Suppose $a^* = \frac{1}{2}$, $b^* = c^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $d^* = e^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

Let $d_1 = d_1' - b_1'$, $d_2 = d_2' - b_2'$, $e_1 = e_1' - c_1'$, $e_2 = e_2' - c_2'$.

Then, $Q_\Gamma = d_1'^2 + d_1'd_2' + d_2'^2 + e_1'^2 + e_1'e_2' + e_2'^2$.

In the new variables, Q_Γ is non-degenerate.

Using the method of Newton diagrams, we derive

$$(\lambda_x, m_x) = (2, 1)$$

Example

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

Conjecture

The RLCT and its multiplicity of the model is $(2,1)$.

Moral of the story (Watanabe-Yamazaki 2004)

1. Take principal part.
2. If non-degenerate, apply Newton diagram method.
3. If degenerate, apply change of variable and go to Step 1.

Critical Question

For which singularities $x \in \mathrm{fiber}(q)$ do we compute (λ_x, m_x)?

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

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.