Asymptotic Approximation of Marginal Likelihood Integrals

Shaowei Lin

10 Dec 2008

Abstract

We study the asymptotics of marginal likelihood integrals for discrete models using resolution of singularities from algebraic geometry, a method introduced recently by Sumio Watanabe. We briefly describe the statistical and mathematical foundations of this method, and explore how Newton diagrams and toric modifications help solve the problem. The approximations are then compared with exact computations of the integrals.

1 Introduction

Evaluation of marginal likelihood integrals is central to Bayesian statistics. Unfortunately, these integrals are generally difficult to compute. They are often estimated using general techniques such as Markov Chain Monte Carlo (MCMC) methods. For certain specific models, approximation formulas are also available. In this project, we hope to find approximation formulas for a large class of discrete statistical models, namely mixtures of independence models. It extends the work in [4] where efficient *exact* algorithms for evaluating integrals with small sample sizes in this class of models were proposed. We refer the reader to [4] for definitions of independence models and their mixtures. In the algebraic geometric context, these mixtures are secant varieties of Segre-Veronese varieties.

We now describe the problem at hand. Let $\mathcal{M}$ be a statistical model on a finite discrete space $[k] = \{1, 2, \ldots, k\}$ parametrized by a real polynomial map $p : \Omega \to \Delta_{k-1}$ where Ω is a compact subset of $\mathbb{R}^d$. Let $q \in \Delta_{k-1}$ be a point in the model with non-zero entries. Consider a sample of size n drawn from the distribution q, and let $u = (u_i)$ be the vector of counts for this sample. We want to estimate the marginal likelihood integral

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

2 Statistical Background

For $\omega \in \Omega$, define the Kullback-Leibler information

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

and the log likelihood ratio function

$$K_n(\omega) = \sum_{i=1}^{k} \frac{u_i}{n} \log \frac{q_i}{p_i(\omega)}$$

where $u = (u_i)$ is the summary for n identically distributed random variables under the model. Note that $K_n(\omega)$ is a random variable that depends on the data u, while $K(\omega) = E[K_n(\omega)]$ is deterministic. One integral of interest is the normalized stochastic complexity

$$F_n = -\log \int_\Omega e^{-nK_n(\omega)} d\omega = \log(\prod_{i=1}^{k} q_i^{u_i}) - \log Z_n \tag{1}$$

where $Z_n = Z_n(u)$ is the marginal likelihood integral. F_n and Z_n are random variables because of their dependence on the random data u. Now, consider a deterministic version of stochastic complexity

$$f(n) = -\log \int_\Omega e^{-nK(\omega)} d\omega. \tag{2}$$

Here the random variable $K_n(\omega)$ in (1) is replaced by the deterministic $K(\omega)$. In general, it is not true that $f(n) = E[F_n]$, but Watanabe showed that they have similar asymptotic properties [5].

Theorem 2.1. *The normalized stochastic complexity satisfies*

$$F_n = \lambda \log n - (m-1)\log\log n + O_p(1),$$

where $O_p(1)$ is bounded in probability and λ, m come from the asymptotics

$$f(n) = \lambda \log n - (m-1)\log\log n + O(1).$$

Furthermore, if $J(z)$ is the zeta function

$$J(z) = \int_\Omega K(w)^z d\omega,$$

then $(-\lambda)$ is the largest pole of $J(z)$ and m its multiplicity.

From (1) and Theorem 2.1, we conclude that

$$E[\log Z_n] = n\sum_{i=1}^{k} q_i \log q_i - \lambda \log n + (m-1)\log\log n + O(1). \qquad (3)$$

Therefore, to estimate marginal likelihood integrals, it is extremely useful to study the zeta function $J(z)$ of the model.

3 Relation to Algebraic Geometry

We begin with the following notations.

Definition 3.1. For a compact set $\Omega \subset \mathbb{R}^d$ with standard Lebesgue measure, a function $K : \Omega \to \mathbb{R}_{\geq 0}$ and $\delta > 0$, define

$$\begin{aligned} \Omega_{K\leq\delta} &= \{\omega \in \Omega : K(\omega) \leq \delta\}, \\ \Omega_{K>\delta} &= \{\omega \in \Omega : K(\omega) > \delta\}. \end{aligned}$$

Also, define the *complexity* of K over Ω to be

$$f(n, \Omega, K) = -\log \int_\Omega e^{-nK(\omega)} d\omega$$

and the *zeta function* of K to be the analytic continuation of

$$J(z) = \int_\Omega K(w)^z d\omega, \quad z \in \mathbb{C}$$

to the entire complex plane.

Definition 3.2. Given functions $f_1, f_2 : \mathbb{N} \to \mathbb{R}$, we say that f_2 is asymptotically larger than f_1 if $f_2(n) - f_1(n)$ is positive and unbounded for large n. We write $f_2 > f_1$. In this case, if the functions are of the form

$$\begin{aligned} f_1(n) &= \lambda_1 \log n - (m_1 - 1) \log\log n + O(1), \\ f_2(n) &= \lambda_2 \log n - (m_2 - 1) \log\log n + O(1), \end{aligned}$$

then $\lambda_1 < \lambda_2$, or $\lambda_1 = \lambda_2$ and $m_1 > m_2$. We write $(\lambda_2, m_2) > (\lambda_1, m_1)$. This gives a total ordering on pairs $(\lambda, m) \in \mathbb{Q}\times\mathbb{N}$. If $f_2(n) = f_1(n) + O(1)$, we say that the functions are asymptotically similar and write $f_2 \sim f_1$. Here the pairs satistfy $(\lambda_2, m_2) = (\lambda_1, m_1)$.

The main idea in attacking our problem is to simplify the form of $K(w)$. The following theorem and corollary is useful for this purpose.

Theorem 3.3. *Suppose that K_1, K_2 satisfy*

$$0 \le cK_1(\omega) \le K_2(\omega)$$

for all $\omega \in \Omega$ and some constant $c > 0$. Then,

$$f(\,\cdot\,, \Omega, K_2) \ge f(\,\cdot\,, \Omega, K_1).$$

Proof. Compare the zeta functions corresponding to K_1 and to K_2. □

Corollary 3.4. *Suppose there exists positive constants c_1, c_2 such that*

$$c_1 K_2(\omega) \le K_1(\omega) \le c_2 K_2(\omega)$$

for all $\omega \in \Omega$. Then,

$$f(\,\cdot\,, \Omega, K_1) \;\sim\; f(\,\cdot\,, \Omega, K_2).$$

In the next theorem, we show that we can replace $K(\omega)$ with a function $Q(\omega)$ quadratic in the $p_i(\omega)$, and shrink the domain to integration to a local neighborhood of the variety $\mathcal{V}(Q) = \{\omega \in \Omega : Q(\omega) = 0\}$. The theorem hints that the asymptotic coefficients λ and m are invariants of $\mathcal{V}(Q)$. In fact, λ is known as the *real log-canonical threshold*.

Theorem 3.5. *For all $\epsilon > 0$, the complexity $f(n)$ of the model is asymptotically similar to $f(n, \Omega_{Q\le\epsilon}, Q)$ where*

$$Q(\omega) = \|\, p(\omega) - q \,\|^2 = \sum_{i=1}^{k} (p_i(\omega) - q_i)^2.$$

Proof. First, rewrite the Kullback information as

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

where $l(x) = -\log x + x - 1$. Now, given $\delta > 0$, suppose $K(\omega) \leq \delta$. Then,

$$\begin{aligned} l(\frac{p_i}{q_i}) &= \frac{1}{k}\sum_{i=1}^{k} l(\frac{p_i}{q_i}) \\ &< \frac{1}{k}\sum_{i=1}^{k} q_i l(\frac{p_i}{q_i}) \\ &= \frac{1}{k}K(\omega) \leq \delta/k. \end{aligned}$$

Since $l(x)$ is convex and attains the minimum 0 at $x = 1$, there exists non-zero constants c_1, c_2 such that

$$c_1(x-1)^2 \leq l(x) \leq c_2(x-1)^2$$

for all x satisfying $l(x) < \delta/k$. Thus, if ω satisfies $K(\omega) \leq \delta$,

$$c_1\sum_{i=1}^{k} q_i(\frac{p_i}{q_i} - 1)^2 \leq K(\omega) \leq c_2\sum_{i=1}^{k} q_i(\frac{p_i}{q_i} - 1)^2.$$

Since the q_i are non-zero and bounded, we have

$$c_3 Q(\omega) \leq K(\omega) \leq c_4 Q(\omega) \tag{4}$$

where $c_3 = c_1 \min_i(1/q_i)$ and $c_4 = c_2 \max_i(1/q_i)$. Hence, by Corollary 3.4,

$$f(\,\cdot\,, \Omega_{K\leq\delta}, Q) \quad\sim\quad f(\,\cdot\,, \Omega_{K\leq\delta}, K). \tag{5}$$

Next, we write the stochastic complexity as

$$f(n) = -\log[\int_{\Omega_{K\leq\delta}} e^{-nK(\omega)}d\omega + \int_{\Omega_{K>\delta}} e^{-nK(\omega)}d\omega].$$

The first integral is bounded below by

$$I_1(n) = \int_{\Omega_{K\leq\delta}} e^{-nK(\omega)}d\omega \geq \int_{\Omega_{K\leq\delta}} e^{-n\delta}d\omega = c_5 e^{-n\delta}$$

while the second integral is bounded above by

$$I_2(n) = \int_{\Omega_{K>\delta}} e^{-nK(\omega)} d\omega \leq \int_{\Omega_{K>\delta}} e^{-n\delta} d\omega = c_6 e^{-n\delta}$$

where the constants c_5, c_6 are measures of the subsets $\Omega_{K\leq\delta}$ and $\Omega_{K>\delta}$ respectively. By the regularity condition, $c_5 \neq 0$, so

$$I_1(n) \leq I_1(n) + I_2(n) \leq I_1(n) + \frac{c_6}{c_5} I_1(n) = (1 + \frac{c_6}{c_5}) I_1(n).$$

Hence, by taking logarithms, we observe that

$$f(\,\cdot\,) \quad \sim \quad -\log I_1(\,\cdot\,).$$

In other words,

$$f(\,\cdot\,, \Omega, K) \quad \sim \quad f(\,\cdot\,, \Omega_{K\leq\delta}, K). \tag{6}$$

We need a few more analytic arguments. Let $\delta' = \delta/c_4$. Then, by (4),

$$\Omega_{Q\leq\delta'} \subset \Omega_{K\leq\delta}$$

and thus,

$$\Omega_{Q\leq\delta'} = (\Omega_{K\leq\delta})_{Q\leq\delta'}.$$

Therefore, the argument which proved (6) also shows that

$$f(\,\cdot\,, \Omega_{K\leq\delta}, Q) \quad \sim \quad f(\,\cdot\,, \Omega_{Q\leq\delta'}, Q). \tag{7}$$

Similarly,

$$f(\,\cdot\,, \Omega, Q) \quad \sim \quad f(\,\cdot\,, \Omega_{Q\leq\delta'}, Q), \tag{8}$$

$$f(\,\cdot\,, \Omega, Q) \quad \sim \quad f(\,\cdot\,, \Omega_{Q\leq\epsilon}, Q). \tag{9}$$

Combining (5-9) completes the proof. □

This theorem shifts our focus from the analytic Kullback information $K(\omega)$ to the polynomial

$$Q(\omega) = \sum_{i=1}^{k} (p_i(\omega) - q_i)^2.$$

This allows us to use tools from algebraic geometry to solve our problem.

4 Resolution of Singularities

In the previous section, we reduced our problem to finding the asymptotics of the minus-log-integral of $e^{-nQ(\omega)}$ in a neighborhood $\Omega_{Q\leq\epsilon}$ of $\mathcal{V}(Q)$ where

$$Q(\omega) = \sum_{i=1}^{k}(p_i(\omega) - q_i)^2.$$

Since Ω is compact, the neighborhood $\Omega_{Q\leq\epsilon}$ is also compact, so we can cover it with *finitely* many open neighborhoods

$$W_x = \{\omega \in \Omega : |\omega - x| < \delta\}$$

where $x \in \mathcal{V}(Q)(x)$ and $\delta > 0$ is fixed over all x. Let $\{\phi_x\}$ be a partition of unity induced by this cover. Then, the zeta function may be rewritten as

$$\begin{aligned} J(z) &= \int_{\Omega_{Q\leq\epsilon}} Q(w)^z d\omega \\ &= \sum_x \int_{W_x} Q(w)^z \phi_x(\omega) d\omega \end{aligned}$$

Since this sum is finite, the pair (λ, m) for $J(z)$ recording its largest pole $(-\lambda)$ and multiplicity m is the smallest of the pairs (λ_x, m_x) for each

$$J_x(z) = \int_{W_x} Q(w)^z d\omega. \tag{10}$$

Furthermore, δ can be as small as we like, as the next lemma shows.

Lemma 4.1. For every $\delta > 0$, there exists some $\epsilon > 0$ such that

$$\Omega_{Q\leq\epsilon} \quad \subset \quad \bigcup_{x\in\mathcal{V}(Q)} B_x(\delta).$$

Proof. Suppose on the contrary that there exists a sequence $\omega_n \in \Omega$ such that $Q(\omega_n) \leq 1/n$ and $|\omega_n - x| > \delta$ for all $x \in \mathcal{V}(Q)$. Since Ω is compact, $\{\omega_n\}$ has a convergent subsequence with limit ω. Since Q is continuous, $Q(\omega_n) \leq 1/n$ implies that $Q(\omega) = 0$ but $|\omega_n - x| > \delta$ for all $x \in \mathcal{V}(Q)$, a contradiction. □

Now, to find the poles of $J_x(z)$, a particular change of variables known as a *resolution of singularities* comes in useful. In 1964, Hironaka proved that such resolutions always exists. The following theorem of Atiyah is a special case of Hironaka's original result [5], and it shows that a *local* resolution of singularities with desirable properties exists.

Theorem 4.2. *Let $\Omega \subset \mathbb{R}^d$ be a neighborhood of the origin, and $K : \Omega \to \mathbb{R}$ a non-constant real analytic function satisfying $K(0) = 0$. Then, there exists a triple $(\mathcal{M}, W, g)$ where*

1. *$W \subset \Omega$ is an open neighborhood of the origin,*
2. *$\mathcal{M}$ is a d-dimensional real analytic manifold,*
3. *$g : \mathcal{M} \to W$ is a real analytic map*

satisfying the following properties.

1. *g is proper, i.e. the inverse image of any compact set is compact.*
2. *g is a real analytic isomorphism between $\mathcal{M} \setminus \mathcal{M}^0$ and $W \setminus W^0$ where $W^0 = \{\omega \in W : K(\omega) = 0\}$ and $\mathcal{M}^0 = \{\mu \in \mathcal{M} : K(g(\mu)) = 0\}$.*
3. *For any $P \in \mathcal{M}^0$, there exists a local chart $\mathcal{M}_P$ with coordinates $\mu = (\mu_1, \mu_2, \ldots \mu_d)$ such that P is the origin and*

$$K(g(\mu)) = c\mu_1^{\sigma_1}\mu_2^{\sigma_2} \cdots \mu_d^{\sigma_d} = c\mu^{\sigma}$$

where $\sigma_1, \sigma_2, \ldots, \sigma_d$ are non-negative integers and the constant c equals 1 or -1. Furthermore, the Jacobian determinant equals

$$|g'(\mu)| = h(u)\mu_1^{\tau_1}\mu_2^{\tau_2} \cdots \mu_d^{\tau_d} = h(\mu)\mu^{\tau}$$

where $h(\mu) \neq 0$ is a real analytic function and $\tau_1, \tau_2, \ldots, \tau_d$ are non-negative integers.

Coming back to $J_x(z)$, Atiyah's theorem implies that there is a d-dimensional manifold $\mathcal{M}$ and a real analytic map $g : \mathcal{M} \to W$ satisfying the above properties on charts $\{\mathcal{M}_P\}$ forming a finite cover of $\mathcal{M}$. A similar partition of unity argument shows that the pair (λ, m) for $J_x(z)$ is the smallest of the pairs (λ_P, m_P) for each

$$\begin{aligned} J_P(z) &= \int_{\mathcal{M}_P} K(g(\mu))^z |g'(\mu)| d\mu \\ &= \int_{\mathcal{M}_P} \mu^{2z\sigma+\tau} h(\mu) d\mu \end{aligned}$$

From this equation, it follows that largest pole $(-\lambda_P)$ is determined by

$$\lambda_P = \min_{1 \le j \le d} \frac{\tau_j + 1}{2\sigma_j}$$

and its multiplicity m_P is the number of arguments j that attain this minimum. Thus, in summary, our original problem is solved if we can find a local resolution of singularities at each point $x \in \mathcal{V}(Q)$.

In practice, such local resolution maps g are often difficult to find. However, simple algorithms exist when the polynomial $Q(\omega)$ is *non-degenerate* [8]. This will be described in the next section.

5 Newton Diagrams and Toric Modifications

We begin with some useful notations from [1, 9]. Given a polynomial

$$Q(\omega) = \sum_{\alpha} c_\alpha \omega^\alpha,$$

where $\omega = (\omega_1, \ldots, \omega_d)$ and each $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$, define its *Newton polyhedron* $\Gamma_+(Q)$ to be the convex hull in $\mathbb{R}^d$ of the set

$$\{\alpha + \alpha' : c_\alpha \neq 0, \alpha' \in \mathbb{R}^d_{\ge 0}\}.$$

A subset $\gamma \subset \Gamma_+(Q)$ is a *face* if there exists some $\beta \in \mathbb{R}^d_{\ge 0}$ such that

$$\gamma = \{\alpha \in \Gamma_+(Q) : \langle \alpha, \beta \rangle \le \langle \alpha', \beta \rangle \text{ for all } \alpha' \in \Gamma_+(Q)\}.$$

Dually, the *normal cone* at γ is the set of all $\beta \in \mathbb{R}^d_{\ge 0}$ satisfying the above condition. Note that the union of all the normal cones gives a partition of the orthant $R^d_{\ge 0}$. We call this partition the *normal fan*. The union of all the compact faces of $\Gamma_+(Q)$ is the *Newton diagram* $\Gamma(Q)$. For each compact face γ of $\Gamma_+(Q)$, define the face polynomial

$$Q_\gamma(\omega) = \sum_{\alpha \in \gamma} c_\alpha \omega^\alpha.$$

The *principal part* of Q is the polynomial

$$Q_0(\omega) = \sum_{\alpha \in \Gamma(Q)} c_\alpha \omega^\alpha.$$

We say that Q is *non-degenerate* if

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

for all compact faces γ of $\Gamma_+(Q)$. Otherwise, Q is *degenerate*.

Non-degeneracy greatly facilitates the resolution of singularities. Indeed, when Q is non-degenerate, there exists a local resolution $g : \mathcal{M} \to W$ around the origin with chart maps $g_P : \mathcal{M}_\mathcal{P} \to W$ given by monomial mappings

$$\begin{array}{rcl} \omega_1 & = & \mu_1^{\beta_{11}} \mu_2^{\beta_{12}} \cdots \mu_d^{\beta_{1d}} \\ \omega_2 & = & \mu_1^{\beta_{21}} \mu_2^{\beta_{22}} \cdots \mu_d^{\beta_{2d}} \\ & \vdots & \\ \omega_d & = & \mu_1^{\beta_{d1}} \mu_2^{\beta_{d2}} \cdots \mu_d^{\beta_{dd}} \end{array} \tag{11}$$

which we write as $\omega = \mu^\beta$, $\beta = (\beta_{ij})$. Furthermore, each chart map satisfies $\det(\beta) = \pm 1$. We call such a chart map a *toric modification*.

We now describe how to find the chart maps from the Newton polyhedron. We say an integer vector $\beta \in \mathbb{Z}^d$ is *primitive* if its coordinates are co-prime. A cone spanned by integer vectors $\beta_1, \ldots, \beta_r$ is *regular* if there exists integer vectors $\beta_{r+1}, \ldots, \beta_d$ such that $\det(\beta) = \pm 1, \beta = (\beta_j)$ and each β_j is primitive. A fan (union of disjoint cones) F' is regular if all its cones are regular, and is a *subdivision* of another fan F if they partition the same set and every cone of F' is contained in some cone of F. Given the normal fan F of $\Gamma_+(Q)$, we may add primitive vectors to get a regular subdivision F'. Then, each maximal cone of F' spanned by some $\{\beta_1, \ldots, \beta_d\}$ describes a toric modification $\omega = \mu^\beta$ where the matrix β has columns β_j. Together, they define a local resolution around the origin as was desired.

Suppose now that upon substitution of (11) into $Q(\omega)$, we have

$$Q(\mu^\beta) = \mu_1^{2\sigma_1}\mu_2^{2\sigma_2}\cdots\mu_d^{2\sigma_d}a(\mu)$$

for some polynomial $a(\mu) > 0$ for all μ. Here the exponents of the μ_i are even because $Q(\omega)$ is non-negative. Define β_{+j} to be the sum of all β_{ij} in the j-th column. Then, the largest pole $(-\lambda)$ of the zeta function is given by

$$\lambda = \min_{1\leq j\leq d} \frac{\beta_{+j}}{2\sigma_j} \tag{12}$$

and its multiplicity m is the number of j at which the minimum is attained. One can show that geometrically, $1/\lambda$ is the smallest real number such that $(1/\lambda)\mathbf{1} \in \Gamma_+(Q)$ where $\mathbf{1}$ is the vector of ones, and m is the number of codimension-one faces meeting $(1/\lambda)\mathbf{1}$.

Regardless of the non-degeneracy of Q, the next proposition shows that it suffices to study its principal part.

Proposition 5.1. *Locally at the origin, the complexity of Q is asymptotically similar to the complexity of Q_0.*

Proof. There exists a sufficiently small neighborhood W of the orgin and positive constants c_1, c_2 such that

$$c_1 Q_0(\omega) \leq Q(\omega) \leq c_2 Q_0(\omega) \quad \text{for all } \omega \in W$$

where Q_0 is the principal part of Q. By Corollary 3.4, the result follows. □

If the principal part Q_0 is degenerate, we can try a change of variable on Q_0, take the principal part of the resulting polynomial and hope that it is non-degenerate. Repeatedly taking principal parts and changing variables might allow us to find a local resolution in degenerate situations.

6 Example

Consider the model that is the two-mixture of the independence model with two ternary random variables. It consists of 3×3 singular matrices $(p_{ij})_{i,j=0}^2$ whose entries sum to one. The defining equations are

$$p_{ij}(\sigma,\theta,\rho) = \sigma_0\theta_i^{(1)}\theta_j^{(2)} + \sigma_1\rho_i^{(1)}\rho_j^{(2)}.$$

where $\sigma \in \Delta_1$ and $\theta^{(1)}, \theta^{(2)}, \rho^{(1)}, \rho^{(2)} \in \Delta_2$. Let Ω be the parameter space $\Delta_1 \times \Delta_2^4$. In the notation of [4], this corresponds to the two-mixture of the independence model with $k = 2$, $s_1 = s_2 = 1$ and $t_1 = t_2 = 2$. Assume that the true distribution that we are sampling from is the uniform distribution $q_{ij} = \frac{1}{9}$, and let $u = (u_{ij})$ be the frequency counts of a sample of size n. We want to estimate the marginal likelihood integral

$$Z_n(u) = \int_\Omega \prod_{i,j} p_{ij}(\sigma, \theta, \rho)^{u_{ij}} d\sigma d\theta d\rho.$$

For this example, we study the zeta function of

$$Q(\sigma, \theta, \rho) = \sum_{i,j=0}^{2} (\sigma_0 \theta_i^{(1)} \theta_j^{(2)} + \sigma_1 \rho_i^{(1)} \rho_j^{(2)} - \frac{1}{9})^2 = \sum_{i,j=0}^{2} f_{ij}^2$$

6.1 Toric Modification

We compute the asymptotics of the complexity of Q around the singular point x with coordinates

$$\begin{aligned} \sigma &= (0,1), \\ \theta^{(1)} = \theta^{(2)} &= (0,0,1), \\ \rho^{(1)} = \rho^{(2)} &= (1/3, 1/3, 1/3). \end{aligned}$$

We do a change of variable to bring the origin to this point.

$$\begin{aligned} \sigma &= (a, 1-a), \\ \theta^{(1)} &= (b_0, b_1, 1 - b_0 - b_1), \\ \theta^{(2)} &= (c_0, c_1, 1 - c_0 - c_1), \\ \rho^{(1)} &= (1/3 + d_0, 1/3 + d_1, 1/3 + d_2), \quad d_2 = -d_0 - d_1, \\ \rho^{(2)} &= (1/3 + e_0, 1/3 + e_1, 1/3 + e_2), \quad e_2 = -e_0 - e_1. \end{aligned}$$

Thus, in terms of the new variables we have

$$Q(a, b, c, d, e) = \\ \tfrac{1}{81} \textstyle\sum_{i,j=0}^{2} [a(9b_i c_j - 9d_i e_j - 3d_i - 3e_j - 1) + 9d_i e_j + 3d_i + 3e_j]^2.$$

The principal part of Q is

$$Q'(a,b,c,d,e) = \tfrac{1}{9}[6(d_1d_2 + e_1e_2 + d_1^2 + d_2^2 + e_1^2 + e_2^2) \\ -6a(d_1 + d_2 + e_1 + e_2) + 8a^2].$$

From this, one can check that Q is non-degenerate. Here we use `Singular` to resolve the variety $\mathcal{V}(Q')$ locally at the origin. This gives us the five charts below, which correspond to the maximal cones of a regular subdivision of the normal fan of the Newton polyhedron of Q. In what follows, H denotes the determinant of the Jacobian of the change of variable.

1. Chart 1:

$$\begin{array}{rclrcl} a &=& e_2a' & H &=& e_2^4 \\ d_1 &=& e_2d_1' & 9Q' &=& 2e_2^2(3+4a'^2-3a'+3d_1'^2+3d_2'^2+3d_1'd_2' \\ d_2 &=& e_2d_2' & & & -3a'd_1'-3a'd_2'-3a'e_1'+3e_1'^2+3e_1') \\ e_1 &=& e_2e_1' & (\lambda,m) &=& (\frac{5}{2},1) \end{array}$$

2. Chart 2:

$$\begin{array}{rclrcl} a &=& e_1a' & H &=& e_1^4 \\ d_1 &=& e_1d_1' & 9Q' &=& 2e_1^2(3+4a'^2-3a'+3d_1'^2+3d_2'^2+3d_1'd_2' \\ d_2 &=& e_1d_2' & & & -3a'd_1'-3a'd_2'-3a'e_2'+3e_2'^2+3e_2') \\ e_2 &=& e_1e_2' & (\lambda,m) &=& (\frac{5}{2},1) \end{array}$$

3. Chart 3:

$$\begin{array}{rclrcl} a &=& d_2a' & H &=& d_2^4 \\ d_1 &=& d_2d_1' & 9Q' &=& 2d_2^2(3+4a'^2-3a'+3e_1'^2+3e_2'^2+3e_1'e_2' \\ e_1 &=& d_2e_1' & & & -3a'e_1'-3a'e_2'-3a'd_1'+3d_1'^2+3d_1') \\ e_2 &=& d_2e_2' & (\lambda,m) &=& (\frac{5}{2},1) \end{array}$$

4. Chart 4:

$$\begin{array}{rclrcl} a &=& d_1a' & H &=& d_1^4 \\ d_2 &=& d_1d_2' & 9Q' &=& 2d_1^2(3+4a'^2-3a'+3e_1'^2+3e_2'^2+3e_1'e_2' \\ e_1 &=& d_1e_1' & & & -3a'e_1'-3a'e_2'-3a'd_2'+3d_2'^2+3d_2') \\ e_2 &=& d_1e_2' & (\lambda,m) &=& (\frac{5}{2},1) \end{array}$$

5. Chart 5:

$$\begin{array}{rclrcl} d_1 &=& ad_1' & H &=& a^4 \\ d_2 &=& ad_2' & 9Q' &=& 2a^2(4-3d_1-3d_2-3e_1-3e_2 \\ e_1 &=& ae_1' & && +3d_1^2+3d_1d_2+3d_2^2+3e_1^2+3e_1e_2+3e_2^2) \\ e_2 &=& ae_2' & (\lambda, m) &=& (\frac{5}{2}, 1) \end{array}$$

Note that the charts come with blowing up $\mathcal{V}(Q')$ at the subvariety

$$\mathcal{V}(a, d_1, d_2, e_1, e_2).$$

Also, in each case, we can use (12) to derive the value of λ and m. Therefore, the asymptotics at this singular point x is given by $\lambda_x = \frac{5}{2}, m_x = 1$.

6.2 Non-toric modification

We compute the asymptotics around the singular point

$$\begin{array}{rcl} \sigma &=& (1/2, 1/2), \\ \theta^{(1)} = \theta^{(2)} &=& (1/3, 1/3, 1/3), \\ \rho^{(1)} = \rho^{(2)} &=& (1/3, 1/3, 1/3). \end{array}$$

We do a change of variable to bring the origin to this point.

$$\begin{array}{rcll} \sigma &=& (1/2+a, 1/2-a), & \\ \theta^{(1)} &=& (1/3+b_0, 1/3+b_1, 1/3+b_2), & b_2 = -b_0 - b_1, \\ \theta^{(2)} &=& (1/3+c_0, 1/3+c_1, 1/3+c_2), & c_2 = -c_0 - c_1, \\ \rho^{(1)} &=& (1/3+d_0, 1/3+d_1, 1/3+d_2), & d_2 = -d_0 - d_1, \\ \rho^{(2)} &=& (1/3+e_0, 1/3+e_1, 1/3+e_2), & e_2 = -e_0 - e_1. \end{array}$$

Thus, in terms of the new variables we have

$$\begin{array}{r} Q(a,b,c,d,e) = \frac{1}{36}\sum_{i,j=0}^{2}[2a(b_i+c_j-d_i-e_j+3b_ic_j-3d_ie_j) \\ +(b_i+c_j+d_i+e_j+3b_ic_j+3d_ie_j)]^2 \end{array}$$

The principal part of Q is

$$\begin{array}{rcl} Q'(a,b,c,d,e) &=& \frac{1}{6}[b_1^2+b_2^2+c_1^2+c_2^2+d_1^2+d_2^2+e_1^2+e_2^2 \\ && +b_1b_2+c_1c_2+d_1d_2+e_1e_2 \\ && +c_1e_2+c_2e_1+b_1d_2+b_2d_1 \\ && +2(b_1d_1+b_2d_2+c_1e_1+c_2e_2)]. \end{array}$$

The Newton diagram of this polynomial has vertices corresponding to the monomials $b_1^2, b_2^2, c_1^2, c_2^2, d_1^2, d_2^2, e_1^2$, and e_2^2. The face with vertices b_1^2 and d_1^2 has the face polynomial $(b_1 + d_1)^2$ which is degenerate. This suggests applying the following change of variable to Q'.

$$\begin{array}{rclcrcl} b_1 &=& b'_1, & & b_2 &=& b'_2 \\ c_1 &=& c'_2, & & c_2 &=& c'_2 \\ d_1 &=& d'_1 - b'_1, & & d_2 &=& d'_2 - b'_2 \\ e_1 &=& e'_1 - c'_1, & & e_2 &=& e'_2 - c'_2 \end{array}$$

The Jacobian determinant of this change of variable is 1. The new polynomial (after removing the "primes" in the notation) is

$$Q(a, b, c, d, e) = d_1^2 + d_1 d_2 + d_2^2 + e_1^2 + e_1 e_2 + e_2^2$$

This polynomial is non-degenerate. Using `Singular`, we found the following resolution of singularities.

1. Chart 1:

$$\begin{array}{rclcrcl} d_1 &=& e_2 d'_1 & & H &=& e_2^3 \\ d_2 &=& e_2 d'_2 & & Q' &=& e_2^2(d_1'^2 + d'_1 d'_2 + d_2'^2 + e_1'^2 + e'_1 + 1) \\ e_1 &=& e_2 e'_1 & & (\lambda, m) &=& (2, 1) \end{array}$$

2. Chart 2:

$$\begin{array}{rclcrcl} d_1 &=& e_1 d'_1 & & H &=& e_1^3 \\ d_2 &=& e_1 d'_2 & & Q' &=& e_1^2(d_1'^2 + d'_1 d'_2 + d_2'^2 + e_2'^2 + e'_2 + 1) \\ e_2 &=& e_1 e'_2 & & (\lambda, m) &=& (2, 1) \end{array}$$

3. Chart 3:

$$\begin{array}{rclcrcl} d_1 &=& d_2 d'_1 & & H &=& d_2^3 \\ e_1 &=& d_2 e'_1 & & Q' &=& d_2^2(e_1'^2 + e'_1 e'_2 + e_2'^2 + d_1'^2 + d'_1 + 1) \\ e_2 &=& d_2 e'_2 & & (\lambda, m) &=& (2, 1) \end{array}$$

4. Chart 4:

$$\begin{array}{rclcrcl} d_2 &=& d_1 d'_2 & & H &=& d_1^3 \\ e_1 &=& d_1 e'_1 & & Q' &=& d_1^2(e_1'^2 + e'_1 e'_2 + e_2'^2 + d_2'^2 + d'_2 + 1) \\ e_2 &=& d_1 e'_2 & & (\lambda, m) &=& (2, 1) \end{array}$$

Therefore, the asymptotics at this point x is given by $(\lambda_x, m_x) = (2, 1)$.

6.3 Other Singularities

We were not able to show that the asymptotics of the complexity of Q at all other points on the variety $\mathcal{V}(Q)$ are at least that given by $(\lambda, m) = (2, 1)$, but we conjecture that asymptotically,

$$\mathrm{E}[-\log Z_n] = n \log 9 + 2 \log n + O(1).$$

In general, we hope to find an algorithm which takes an arbitrary mixture of independence models as defined in [4] and computes the asymptotic coefficients (λ, m) for the model.

7 Comparison with Exact Integrals

In this section, we consider the *Cheating Coin Flipper* example [2]. It is the two-mixture of the independence model of four identically distributed binary random variables. The defining equations are

$$p_i(\sigma, \theta, \rho) = \binom{4}{i}(\sigma_0 \theta_0^i \theta_1^{4-i} + \sigma_1 \rho_0^i \rho_1^{4-i}), \qquad \text{for } i = 0, 1, 2, 3, 4,$$

where $\sigma, \theta, \rho \in \Delta_1$. In the notation of [3], this corresponds to the two-mixture of the independence model with $k = 1$, $s_1 = 4$ and $t_1 = 1$. We assume that the true distribution (q_i) comes from parameters $\sigma = (1, 0), \theta = \rho = (\frac{1}{2}, \frac{1}{2})$, and choose samples u of size n where n is a multiple of 16 and

$$u_i = \frac{n}{16}\binom{4}{i} = nq_i.$$

We want to compute the marginal likelihood integral

$$Z_n(u) = \int_{\Delta_1^3} p_0^{u_0} p_1^{u_1} p_2^{u_2} p_3^{u_3} p_4^{u_4} \; d\sigma d\theta d\rho.$$

According to [9], we have the asymptotics

$$\mathrm{E}[-\log Z_n(u)] = -n \sum_{i=0}^{4} q_i \log q_i + \frac{3}{4} \log n + O(1). \tag{13}$$

Table 1: Comparison of exact computations with asymptotics

n	$F_{16+n} - F_n$	$g(n)$
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

We use the methods of [4] to compute $Z_n(u)$ exactly and compare its values with the above asymptotics. Recall the stochastic complexity

$$F_n = n \sum_{i=0}^{4} q_i \log q_i - \log Z_n(u).$$

By (13), we should expect

$$F_{16+n} - F_n \;\approx\; \frac{3}{4}(\log(16+n) - \log n) = g(n).$$

Indeed, a comparison is shown in Table 1 and the approximation is reasonably accurate. Meanwhile, the *Bayesian Information Criterion* (BIC) predicts

$$F_{16+n} - F_n \;\approx\; \frac{3}{2}(\log(16+n) - \log n) = g(n),$$

which will be off by a factor of 2 from the actual values.

Acknowledgements. I am grateful to Mathias Drton, Philipp Rostalski, Bernd Sturmfels and, in particular, Sumio Watanabe for many useful conversations and suggestions.

References

[1] V. I. Arnold, V. V. Goryunov, O. V. Lyashko and V. A. Vasil'ev: *Singularity Theory I*, Springer, 1998.

[2] M. Drton, B. Sturmfels and S. Sullivant: *Lectures on Algebraic Statistics*, Springer, 2009.

[3] D. Geiger and D. Rusakov: Asymptotic model selection for naive Bayesian networks, *Journal of Machine Learning Research* **6** (2005) 1–35.

[4] S. Lin, B. Sturmfels and Z. Xu: Marginal Likelihood Integrals for Mixtures of Independence Models, `arXiv:0805.3602`, May 2008, submitted to *Journal of Machine Learning Research*.

[5] S. Watanabe: *Algebraic Geometry and Statistical Learning Theory*, to be published.

[6] S. Watanabe: Algebraic Analysis for Nonidentifiable Learning Machines, *Neural Computation* **13** (2001) 899–933.

[7] K.Yamazaki and S.Watanabe: Singularities in mixture models and upper bounds of stochastic complexity, *International Journal of Neural Networks* **16** (2003) 1029–1038.

[8] K. Yamazaki, M. Aoyagi and S. Watanabe: Stochastic Complexity and Newton Diagram, *International Symposium on Information Theory and its Applications*, Parma, Italy, (2004) CD-ROM.

[9] K. Yamazaki and S. Watanabe: Newton Diagram and Stochastic Complexity in Mixture of Binomial Distributions, *Algorithmic Learning Theory* (2004) 350–364.