Asymptotic behaviour of the marginal likelihood integral for general Markov models

Piotr Zwiernik
IPAM $\longrightarrow$ TU Eindhoven
(*With special thanks to Shaowei Lin.*)

Singular Learning Theory,
AIM, Palo Alto, 13 Dec 2011

Outline of the talk

- The asymptotic behavior of the marginal likelihood integral and the real log-canonical threshold.
- The real log-canonical threshold for general Markov models.
    - q-fibers for (binary) general Markov models.
    - Polyhedral geometry for the deepest singularity.

Discrete algebraic statistical model

- random variable: $X \in \mathcal{X}$, $|\mathcal{X}| < \infty$, $p = (p_x)_{x \in \mathcal{X}}$
- probability simplex: $\Delta_{|\mathcal{X}|-1} = \{p \in \mathbb{R}^{\mathcal{X}} : \sum_{x \in \mathcal{X}} p_x = 1, p_x \geq 0\}$
- model: $p : \Theta \to \Delta_{|\mathcal{X}|-1}$, $\mathcal{M} = p(\Theta)$, p polynomial map
- the true distribution of X: $q \in \Delta_{|\mathcal{X}|-1}$

Asymptotics of the marginal likelihood

- random sample: $X^{(N)} = (X^1, \dots, X^N)$
- marginal likelihood: $Z_N = \int_\Theta L(\theta; X^{(N)})\psi(\theta)\mathrm{d}\theta$
- stochastic complexity: $F_N = -\log Z_N$; entropy: $S = -\sum_{x\in\mathcal{X}} q_x \log q_x$

Bayesian Information Criterion (BIC)

- If $p^{-1}(q) = \hat{\theta}$ lies in the interior of Θ and the Jacobian of p has full rank then, as $N \to \infty$,

$$\mathbb{E}F_N = NS + \frac{d}{2}\log N + O(1).$$

The general case (S. Watanabe)

- Kullback-Leibler distance: $K(\theta) = \sum_{i=1}^{m} q_i \log \frac{q_i}{p_i(\theta)}$
  - $K(\theta) \geq 0$ on Θ
  - $K(\theta) = 0$ only if $p(\theta) = q$
- zeta function on $\mathbb{C}$: $\zeta(z) = \int_\Theta K(\theta)^{-z}\psi(\theta)\mathrm{d}\theta$
  - **real log-canonical threshold**: $\mathrm{rlct}_\Theta(K;\psi)$ is the smallest pole of ζ
  - its **multiplicity**: $\mathrm{mult}_\Theta(K;\psi)$

Theorem

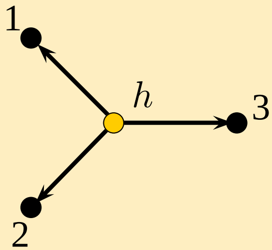
- With some compactness assumptions, as $N \to \infty$ then $\mathbb{E}F_N = NS + \mathrm{rlct}_\Theta(K;\psi)\log N + (\mathrm{mult}_\Theta(K;\psi) - 1)\log\log N + O(1)$

Asymptotics and q-fibers

- $\mathrm{RLCT}_{\Theta}(K) = \min_{\theta \in \Theta_0} \mathrm{RLCT}_{\Theta_0}(K)$
  - important distinction: if W_0 neighbourhood of θ_0 in $\mathbb{R}^d$ then $\mathrm{RLCT}_{W_0}(K) = \mathrm{RLCT}_{\theta_0}(K)$
- $\mathrm{RLCT}_{\Theta}(K(\theta)) = \mathrm{RLCT}_{\Theta}(\sum_{x \in \mathcal{X}} (p_x(\theta) - q_x)^2)$
- q-fiber: $\widehat{\Theta} = p^{-1}(q)$
- $\theta_0 \notin \widehat{\Theta} \implies \mathrm{rlct}_{\theta_0}(\sum_{x \in \mathcal{X}} (p_x(\theta) - q_x)^2) = \infty$

The binary tripod tree model

- $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3 | H$ or
- a Bayesian network a tripod tree
- seven free parameters
- the codimension is zero



Central moment parametrization

$$\mathcal{M}_T : \quad \begin{aligned} \mu_{12} &= \tfrac{1}{4}(1-s^2)\,\eta_1\eta_2, \\ \mu_{13} &= \tfrac{1}{4}(1-s^2)\,\eta_1\eta_3, \\ \mu_{23} &= \tfrac{1}{4}(1-s^2)\,\eta_2\eta_3, \\ \mu_{123} &= \tfrac{1}{4}(1-s^2)s\,\eta_1\eta_2\eta_3 \end{aligned}$$

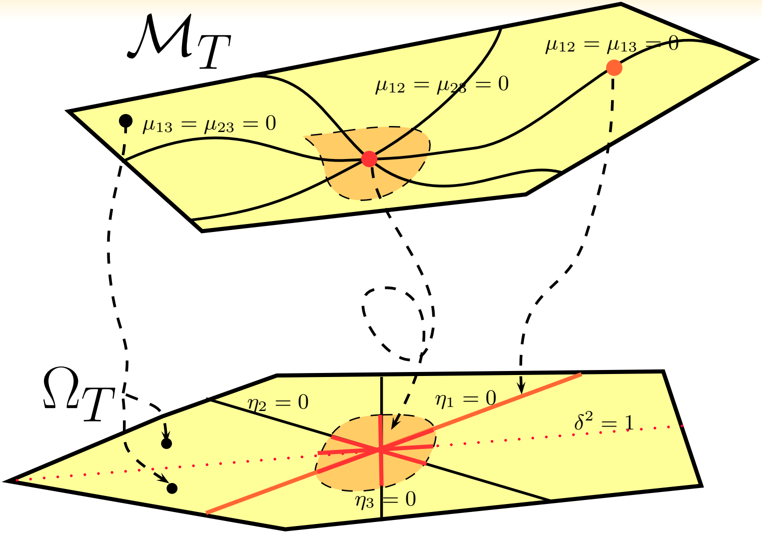
Finite q-fibers

- $$\boldsymbol{\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23}} = (\tfrac{1}{4}(1-s^2)s\eta_1\eta_2\eta_3)^2 + 4(\tfrac{1}{4}(1-s^2))^3(\eta_1\eta_2\eta_3)^2$$
  $$= (\tfrac{1}{4}(1-s^2)\eta_1\eta_2\eta_3)^2(s^2+1-s^2) = (\tfrac{1}{4}(1-s^2)\eta_1\eta_2\eta_3)^2$$
- $\frac{\mu_{123}^2}{\mu_{123}^2+4\mu_{12}\mu_{13}\mu_{23}} = \frac{(\frac{1}{4}(1-s^2)s\eta_1\eta_2\eta_3)^2}{(\frac{1}{4}(1-s^2)\eta_1\eta_2\eta_3)^2} = s^2$
- $\frac{\mu_{123}^2+4\mu_{12}\mu_{13}\mu_{23}}{\mu_{23}^2} = \frac{(\frac{1}{4}(1-s^2)\eta_1\eta_2\eta_3)^2}{(\frac{1}{4}(1-s^2)\eta_2\eta_3)^2} = \eta_1^2$, and so on
- well defined for $q \in \mathcal{M}_T$ such that $\mu_{12}\mu_{13}\mu_{23} \neq 0$

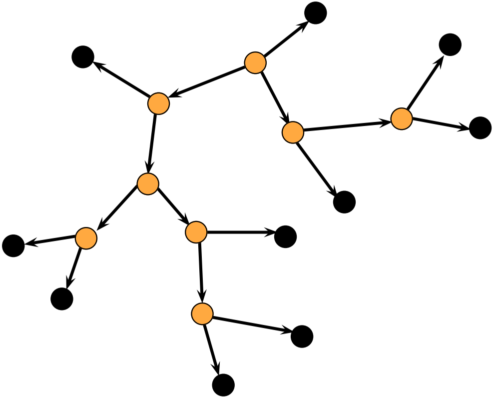
Submodels and singularities

(draw four models,
show that the fiber for $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3$ is a union of affine spaces)

Sketch picture



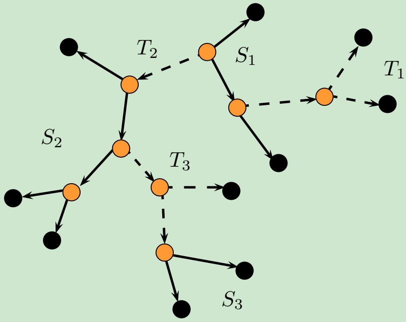
This generalizes



- $(X_1, \ldots, X_n) \in \{0, 1\}^n$ represented by leaves of T

The general case

- let $q \in \mathcal{M}_T$ and $\Sigma = [\mu_{ij}]_{i,j \in [n]}$ the covariance matrix
- the asymptotics of $\mathbb{E}F_N$ determined by zeros in Σ (marginal independencies).



General formula

Theorem [Z.]

- if $p^{-1}(q)$ is a manifold with corners (there are no degree zero nodes) then, as $N \to \infty$,

$$\mathbb{E}F_N = NS + \frac{1+2|E|-2l_2}{2}\log N + O(1).$$

- for trivalent trees

$$\mathbb{E}F_N = NS + \left(\frac{1+2|E|-2l_2}{2} - \frac{5l_0}{4}\right)\log N - c\log\log N + O(1),$$

where c is a nonnegative integer. Moreover $c = 0$ always if either both r is nondegenerate or if r and all its neighbors are degenerate.

Step 1: Reparametrization

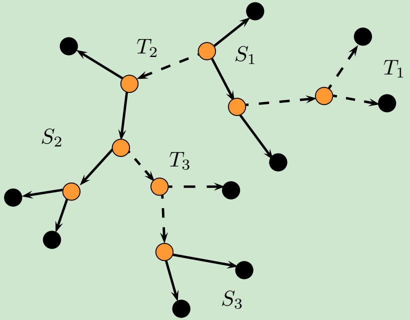
- tree cumulants: $\kappa = (\kappa_I)_{I \subseteq [n], I \neq \emptyset}$, $\omega = ((s_v)_{v \in V}, (\eta_e)_{e \in E})$

$$\begin{array}{ccc} \Theta_T & \xrightarrow{\quad p \quad} & \Delta_{2^n - 1} \\ f_{\omega\theta} \uparrow \downarrow f_{\theta\omega} & & f_{\kappa p} \uparrow \downarrow f_{p\kappa} \\ \Omega_T & \xrightarrow{\quad \psi_T \quad} & \mathcal{K}_T \end{array}$$

- $\kappa_I(\omega) = \frac{1}{4}(1 - s_{r(I)}^2) \prod_{v \in N(I)} s_v^{\deg(v)-2} \prod_{e \in E(I)} \eta_e$, for $I \subseteq [n], |I| \geq 2$.

$$\mathrm{RLCT}_{\Theta_T}\Big(\sum_x (p_x(\theta) - q_x)^2\Big) = \Big(\frac{n}{2}, 0\Big) + \mathrm{RLCT}_{\Omega_T}\Big(\sum_{|I| \geq 2} (\kappa_I(\omega) - \hat{\kappa}_I)^2\Big)$$

Step 2: The main reduction



- note: if T is trivalent then $T_1, \ldots, T_k$ are trivalent

The deepest singularity

- let $\hat{\kappa}_{ij} = 0$ for all $i, j \in [n]$ (then $\hat{\kappa}_I = 0$ for all $I \subseteq [n]$)
- $\mathrm{RLCT}_{\omega_0}(\sum_I (\kappa_I(\omega) - \hat{\kappa}_I)^2) = \mathrm{RLCT}_{\omega_0}(\sum_{i,j \in [n]} \kappa_{ij}^2(\omega))$ for $\omega_0 \in \widehat{\Omega}$.

- the q-fiber is given by a union of *affine subspaces* with non-empty common intersection denoted by $\widehat{\Omega}_{\mathrm{deep}}$
- $\mathrm{RLCT}_{\omega_0}(K) \leq \mathrm{RLCT}_{\omega}(K)$ for every $\omega \in \Omega$ and $\omega_0 \in \widehat{\Omega}_{\mathrm{deep}}$

The monomial case

- if $\omega_0 \in \widehat{\Omega}_{\text{deep}}$ then

$$\mathrm{RLCT}_{\omega_0}(\sum_{i,j\in[n]} \kappa_{ij}^2(\omega)) = \mathrm{RLCT}_0(\sum_{i,j\in[n]} m_{ij}^2(\omega)),$$

where $m_{ij}(\omega) = s_{r(ij)} \prod_{e\in E(ij)} \eta_e$

- e.g. (quartet)

$m_{12} = s_a\eta_{a1}\eta_{a2}$, $m_{13} = s_a\eta_{a1}\eta_{ab}\eta_{b3}$ and $m_{34} = s_b\eta_{b3}\eta_{b4}$

The Newton diagram method

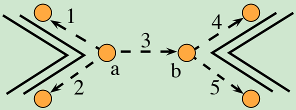
Theorem

- Let $f(x) = \sum_\alpha c_\alpha x^{2\alpha}$ and $\Gamma_+ = \mathrm{conv}(2\alpha : c_{\alpha \neq 0}) + \mathbb{R}^n_{\geq 0}$.
- Then $\mathrm{RLCT}_0(f) = (\frac{1}{t}, c)$, where and t be the smallest such that $(t, \ldots, t) \in \Gamma_+$ and c is the codimension of the face hit by $(t, \ldots, t)$.

rlct from the Newton diagram

- coordinates of the ambient space: $(x_e), (y_v)$
- distinguised facet: $\sum_{e \in E_{\text{term}}} x_e \geq 4$

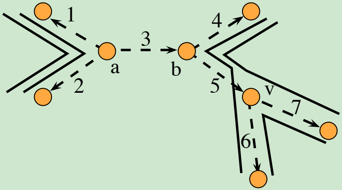
- show $\text{rlct}_0(\sum_{ij} m_{ij}^2) = \frac{n}{4}$ by:
  - note $t < \frac{4}{n}$ then $(t, \ldots, t) \notin \Gamma_+$
  - constructing a point $P \in \Gamma_+$ such that $P \leq \frac{4}{n}\mathbf{1}$
- idea: for $n = 4$



$2 \cdot (2, 0; 2, 2, 0, 0, 0)$ and $2 \cdot (0, 2; 0, 0, 0, 2, 2)$ gives
$P = \frac{1}{4}(4, 4; 4, 4, 0, 4, 4) \leq (1, 1; 1, 1, 1, 1, 1)$.

rlct from the Newton diagram 2

- for $n = 5$



gives $\frac{1}{5}(4,4,2;4,4,0,4,4,4,4) \leq \frac{4}{5}\mathbf{1}$.

Conclusions and remarks

- Understanding of q-fibers is essential
- In our case the degenerate cases corresponded to graphical submodels
- Tree remains important even for the Newton diagram method.

- generalization to Gaussian models