Asymptotic behaviour of the marginal likelihood integral for general Markov models

Piotr Zwiernik
IPAM $\rightarrow$ 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: \( p : \Theta \rightarrow \Delta_{|\mathcal{X}|-1}, \mathcal{M} = p(\Theta), p \) polynomial map

- the true distribution of \( X \): \( q \in \Delta_{|\mathcal{X}|-1} \)
Asymptotics of the marginal likelihood

- random sample: \( X^{(N)} = (X^1, \ldots, X^N) \)

- marginal likelihood: \( Z_N = \int_{\Theta} L(\theta; X^{(N)}) \psi(\theta) d\theta \)

- 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 \( \Theta \) and the Jacobian of \( p \) has full rank then, as \( N \to \infty \),
  \[
  \mathbb{E}F_N = NS + \frac{d}{2} \log N + O(1).
  \]
Kullback-Leibler distance: \( K(\theta) = \sum_{i=1}^{m} q_i \log \frac{q_i}{p_i(\theta)} \)
- \( K(\theta) \geq 0 \) on \( \Theta \)
- \( K(\theta) = 0 \) only if \( p(\theta) = q \)

Zeta function on \( \mathbb{C} \):
\[
\zeta(z) = \int_{\Theta} K(\theta)^{-z} \psi(\theta) d\theta
\]
- real log-canonical threshold: \( \text{rlct}_{\Theta}(K; \psi) \) is the smallest pole of \( \zeta \)
- its multiplicity: \( \text{mult}_{\Theta}(K; \psi) \)

**Theorem**
- With some compactness assumptions, as \( N \to \infty \) then
\[
\mathbb{E} F_N = NS + \text{rlct}_{\Theta}(K; \psi) \log N + (\text{mult}_{\Theta}(K; \psi) - 1) \log \log N + O(1)
\]
Asymptotics and $q$-fibers

- $\text{RLCT}_\Theta(K) = \min_{\theta \in \Theta_0} \text{RLCT}_{\Theta_0}(K)$
  - important distinction: if $W_0$ neighbourhood of $\theta_0$ in $\mathbb{R}^d$ then $\text{RLCT}_{W_0}(K) = \text{RLCT}_{\Theta_0}(K)$
- $\text{RLCT}_\Theta(K(\theta)) = \text{RLCT}_{\Theta}(\sum_{x \in X} (p_x(\theta) - q_x)^2)$
- $q$-fiber: $\hat{\Theta} = p^{-1}(q)$
- $\theta_0 \notin \hat{\Theta} \implies \text{rlct}_{\theta_0}(\sum_{x \in X} (p_x(\theta) - q_x)^2) = \infty$
The binary tripod tree model

- \( X_1 \perp \perp X_2 \perp \perp X_3 \mid H \) or
- a Bayesian network a tripod tree
- seven free parameters
- the codimension is zero
Central moment parametrization

\[ \mathcal{M}_T : \]

\[ \mu_{12} = \frac{1}{4} (1-s^2) \eta_1 \eta_2, \]

\[ \mu_{13} = \frac{1}{4} (1-s^2) \eta_1 \eta_3, \]

\[ \mu_{23} = \frac{1}{4} (1-s^2) \eta_2 \eta_3, \]

\[ \mu_{123} = \frac{1}{4} (1-s^2) s \eta_1 \eta_2 \eta_3 \]
Finite $q$-fibers

\[
\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23} = \left(\frac{1}{4}(1 - s^2)s\eta_1\eta_2\eta_3\right)^2 + 4\left(\frac{1}{4}(1 - s^2)\right)^3(\eta_1\eta_2\eta_3)^2
\]

\[
= \left(\frac{1}{4}(1 - s^2)\eta_1\eta_2\eta_3\right)^2(s^2 + 1 - s^2) = \left(\frac{1}{4}(1 - s^2)\eta_1\eta_2\eta_3\right)^2
\]

\[
\frac{\mu_{123}^2}{\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23}} = \frac{\left(\frac{1}{4}(1 - s^2)s\eta_1\eta_2\eta_3\right)^2}{\left(\frac{1}{4}(1 - s^2)\eta_1\eta_2\eta_3\right)^2} = s^2
\]

\[
\frac{\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23}}{\mu_{23}^2} = \frac{\left(\frac{1}{4}(1 - s^2)\eta_1\eta_2\eta_3\right)^2}{\left(\frac{1}{4}(1 - s^2)\eta_2\eta_3\right)^2} = \eta_1^2, \text{ and so on}
\]

\[
\text{well defined for } q \in \mathcal{M}_T \text{ such that } \mu_{12}\mu_{13}\mu_{23} \neq 0
\]
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)
Asymptotics for GMMs

The case of zero covariances

Sketch picture

\[ \mathcal{M}_T \]

\[ \Omega_T \]

Asymptotic behaviour of the marginal likelihood integral for general Markov models
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 $E F_N$ determined by zeros in $\Sigma$ (marginal independencies).
General formula

**Theorem [Z.]**

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

\[
\mathbb{EF}_N = NS + \frac{1 + 2|E| - 2l_2}{2} \log N + O(1).
\]

- For trivalent trees

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

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

\[
\Theta_T \xrightarrow{p} \Delta_{2^n - 1} \\
\Omega_T \xrightarrow{\psi_T} \mathcal{K}_T \\
f_{\omega\theta} \downarrow f_{\theta\omega} \quad f_{\kappa p} \quad f_{p\kappa} \downarrow
\]

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

\[
\text{RLCT}_{\Theta_T}(\sum_x (p_x(\theta) - q_x)^2) = \left( \frac{n}{2}, 0 \right) + \text{RLCT}_{\Omega_T}(\sum_{|I| \geq 2} (\kappa_I(\omega) - \hat{\kappa}_I)^2)
\]

Asymptotic behaviour of the marginal likelihood integral for general Markov models
Step 2: The main reduction

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

Asymptotic behaviour of the marginal likelihood integral for general Markov models
The deepest singularity

- let \( \hat{\kappa}_{ij} = 0 \) for all \( i, j \in [n] \) (then \( \hat{\kappa}_I = 0 \) for all \( I \subseteq [n] \))

- \( \text{RLCT}_{\omega_0} (\sum_I (\kappa_I(\omega) - \hat{\kappa}_I)^2) = \text{RLCT}_{\omega_0} (\sum_{i,j \in [n]} \kappa_{ij}^2(\omega)) \) for \( \omega_0 \in \hat{\Omega} \).

- the \( q \)-fiber is given by a union of affine subspaces with non-empty common intersection denoted by \( \hat{\Omega}_{\text{deep}} \)

- \( \text{RLCT}_{\omega_0}(K) \leq \text{RLCT}_{\omega}(K) \) for every \( \omega \in \Omega \) and \( \omega_0 \in \hat{\Omega}_{\text{deep}} \)
The monomial case

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

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

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

- e.g. (quartet)

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

Let $f(x) = \sum_\alpha c_\alpha x^{2\alpha}$ and $\Gamma_+ = \text{conv}(2\alpha : c_\alpha \neq 0) + \mathbb{R}_+^n$.

Then $\text{RLCT}_0(f) = (\frac{1}{t}, c)$, 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: $(x_e), (y_v)$
- distinguished facet: $\sum_{e \in E_{term}} x_e \geq 4$

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 $P \leq \frac{4}{n} 1$

idea: for $n = 4$

$2 \cdot (2, 0; 2, 2, 0, 0, 0)$ and $2 \cdot (0, 2; 0, 0, 0, 0, 2, 2)$ gives

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

Asymptotic behaviour of the marginal likelihood integral for general Markov models
for $n = 5$

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