Generalized BIC for Singular Models Factoring through Regular Models

Shaowei Lin

http://math.berkeley.edu/~shaowei/

Department of Mathematics, University of California, Berkeley

PhD student (Advisor: Bernd Sturmfels)

Abstract

The Bayesian Information Criterion (BIC) is an important tool for model selection, but its use is limited only to regular models. Recently, Watanabe [5] generalized the BIC to singular models using ideas from algebraic geometry. In this paper, we compute this generalized BIC for singular models factoring through regular models, relating it to the real log canonical threshold of polynomial fiber ideals.

Bayesian Information Criterion

The BIC is a score used for model selection.

$$BIC = -\log L_0 + \frac{d}{2} \log N$$

L_0: maximum likelihood, d: dimension, N: sample size.

However, it applies only for ***regular*** models.

Regular and Singular Models

A model $\mathcal{M}(U)$ is ***regular*** if it is identifiable

$$p(x|u, \mathcal{M}) = p(x|u', \mathcal{M}) \ \forall x \quad \Rightarrow \quad u = u'$$

and the Fisher information matrix $I(u)$ is positive definite $\forall u$.

$$I_{jk}(u) = \int \frac{\partial \log p(x|u)}{\partial u_j} \frac{\partial \log p(x|u)}{\partial u_k} p(x|u) dx$$

Otherwise, the model is ***singular***.

e.g. most exponential families are regular,
while most hidden variable models are singular.

Factored Models

$\mathcal{M}_2(\Omega)$ ***factors*** through $\mathcal{M}_1(U)$ if for some function $u(\omega)$,

$$p(x|\omega, \mathcal{M}_2) = p(x|u(\omega), \mathcal{M}_1).$$

e.g. parametrized multivariate Gaussian models $\mathcal{N}(0, \Sigma(\omega))$.

Generalized BIC

In 2001, Watanabe showed that for singular models $\mathcal{M}(\Omega)$, the BIC generalizes to

$$-\log L_0 + \lambda \log N - (\theta - 1) \log \log N,$$

where λ is the smallest pole of the ***zeta function***

$$\zeta(z) = \int_\Omega K(\omega)^{-z} \varphi(\omega) d\omega, \quad z \in \mathbb{C}$$

and θ its multiplicity. Here, $\varphi(\omega)$ is a prior on Ω, and $K(\omega)$ the Kullback-Leibler distance to the true distribution.

We compute (λ, θ) by monomializing $K(\omega)$.
(a.k.a. ***resolution of singularities***)

Key Idea: Fiber Ideals

Even if $\mathcal{M}(\Omega)$ is parametrized by simple polynomials, the Kullback function $K(\omega)$ is non-polynomial and difficult to monomialize.

We show that if $\mathcal{M}(\Omega)$ factors through a regular model, we can define a polynomial ***fiber ideal*** and compute (λ, θ) by monomializing this ideal.

Real Log Canonical Thresholds

Given an ideal $I = \langle f_1, \ldots, f_r \rangle$ generated by polynomials, define the ***real log canonical threshold*** (λ, θ) to be the smallest pole λ and multiplicity θ of the zeta function

$$\zeta(z) = \int (f_1(\omega)^2 + \cdots + f_r(\omega)^2)^{-z/2} \, d\omega, \quad z \in \mathbb{C}.$$

The RLCT is independent of the choice of generators $f_1, \ldots, f_r$, and can be computed by monomializing I.

For monomial ideals, we find RLCTs using a geometric-combinatorial tool involving Newton polyhedra.

Main Result

Let $M_2(\Omega)$ factor through a regular model $M_1(U)$ via $u(\omega)$.
Given N i.i.d. samples, let their M.L.E. in $\mathcal{M}_1$ be $\hat{u} \in U$.
Let $Z(N)$ be the marginal likelihood of the data given $\mathcal{M}_2$.

Theorem (L.)

If $\hat{u} \in u(\Omega)$, then asymptotically,

$$-\log Z(N) \approx -\log L_0 + \lambda \log N - (\theta - 1) \log\log N$$

where $(2\lambda, \theta)$ is the RLCT of the ***fiber ideal***

$$I = \langle u_1(\omega) - \hat{u}_1, \ldots, u_d(\omega) - \hat{u}_d \rangle.$$

Example: Classical BIC

Let $\mathcal{M}(U)$ be a regular model and $\hat{u}$ the M.L.E. of the data. Then, $(2\lambda, \theta)$ is the RLCT of the fiber ideal

$$I = \langle u_1 - \hat{u}_1, \ldots, u_d - \hat{u}_d \rangle$$

which is just $(d, 1)$.

Example: Discrete Models

Let $\mathcal{M}(\Omega)$ be a model on discrete space $\{1, 2, \ldots, k\}$ with polynomial state probabilities $p_1(\omega), \ldots, p_k(\omega)$. Let $q_1, \ldots, q_k$ be relative frequencies of the data.

If $q \in p(\Omega)$, then $(2\lambda, \theta)$ is the RLCT of the fiber ideal

$$I = \langle p_1(\omega) - q_1, \ldots, p_k(\omega) - q_k \rangle.$$

Example: Gaussian Models

Let $\mathcal{N}(\mu(\Omega), \Sigma(\Omega))$ be a multivariate Gaussian model with sample mean $\hat{\mu}$ and covariance matrix $\hat{\Sigma}$.

If $\hat{\mu} \in \mu(\Omega)$ and $\hat{\Sigma} \in \Sigma(\Omega)$, then $(2\lambda, \theta)$ is the RLCT of the fiber ideal

$$I = \langle \mu(\omega) - \hat{\mu}, \Sigma(\omega) - \hat{\Sigma} \rangle.$$

Acknowledgements

Many thanks go to Mathias Drton and Sumio Watanabe for enlightening discussions and key ideas.

References

1. V. I. Arnol'd, S. M. Guseĭn-Zade and A. N. Varchenko: *Singularities of Differentiable Maps*, Vol. II, Birkhäuser, Boston, 1985.
2. H. Hironaka: Resolution of singularities of an algebraic variety over a field of characteristic zero I, II, *Ann. of Math.* (2) **79** (1964) 109–203.
3. S. Lin: Asymptotic Approximation of Marginal Likelihood Integrals, preprint `arXiv:1003.5338` (2010).
4. S. Watanabe: Algebraic analysis for nonidentifiable learning machines, *Neural Computation* **13** (2001) 899–933.
5. S. Watanabe: *Algebraic Geometry and Statistical Learning Theory*, Cambridge Monographs on Applied and Computational Mathematics **25**, Cambridge University Press, Cambridge, 2009.
6. G. Schwarz: Estimating the Dimension of a Model, *Annals of Statistics* **6** (1978) 461–464.