Journal of Multivariate Analysis 133 (2015) 61–76

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# Asymptotic normality in the maximum entropy models on graphs with an increasing number of parameters

Ting Yan [a,*], Yunpeng Zhao [b], Hong Qin [a]

[a] Department of Statistics, Central China Normal University, Wuhan, 430079, China
[b] Department of Statistics, George Mason University, VA 22030-4444, USA

ARTICLE INFO

*Article history:*
Received 9 August 2013
Available online 4 October 2014

*AMS subject classifications:*
62E20
62F12

*Keywords:*
Maximum entropy models
Maximum likelihood estimator
Asymptotic normality
Increasing number of parameters

ABSTRACT

Maximum entropy models, motivated by applications in neuron science, are natural generalizations of the $\beta$-model to weighted graphs. Similar to the $\beta$-model, each vertex in maximum entropy models is assigned a potential parameter, and the degree sequence is the natural sufficient statistic. Hillar and Wibisono (2013) have proved the consistency of the maximum likelihood estimators. In this paper, we further establish the asymptotic normality for any finite number of the maximum likelihood estimators in the maximum entropy models with three types of edge weights, when the total number of parameters goes to infinity. Simulation studies are provided to illustrate the asymptotic results.


## 1. Introduction

In neuron networks, neurons in one region of the brain may transmit a continuous signal using sequences of spikes to a second receiver region. The coincidence detectors in the second region capture the absolute difference in spike times between pairs of neurons projecting from the first region. There may be three possible types of timing differences: zero or nonzero indicator; countable number of possible values; any nonnegative real value. Exploring how the transmitted signal in the first region can be recovered by the second is a basic question in the analysis of neuron networks. Maximum entropy models provide a possible solution to this question for the above three possible weighted edges. For detailed explanations, see [11]; for their wide applications in the biological studies as well as other disciplines such as economics and physics, see [11,8,1,23,25] and references therein. Maximum entropy models (sometimes with different names) also appear in other fields of network analysis, e.g., community detection and social network analysis. For example, see [7,2,3,16,26,18].

In the maximum entropy models, the degree sequence is the exclusively natural sufficient statistics on the exponential family distributions and fully captures the information of an undirected graph. Its study primarily focuses on understanding the generating mechanisms of networks. When network edge takes dichotomous values (“0” or “1”), the maximum entropy model becomes the $\beta$-model (a name given by Chatterjee, Diaconis and Sly [7]), an undirected version of the $p_1$ model for directed graphs by Holland and Leinhardt [12]. Rinaldo, Petrović and Fienberg [18] derived necessary and sufficient conditions for the existence and uniqueness of the maximum likelihood estimate (MLE). As the number of parameters goes to infinity, Chatterjee, Diaconis and Sly [7] proved that the MLE is uniformly consistent; Yan and Xu [24] further derived its asymptotical normality. When the maximum entropy models involve the finite discrete, infinite discrete or continuous weighted edges, Hillar and Wibisono [11] have obtained the explicit conditions for the existence and uniqueness of the MLE and proved that the MLE is uniformly consistent as the number of parameters goes to infinity.

* Corresponding author.
*E-mail addresses:* tingyanty@gmail.com, tingyanty@mail.ccnu.edu.cn (T. Yan), yzhao15@gmu.edu (Y. Zhao), qinhong@mail.ccnu.edu.cn (H. Qin).

http://dx.doi.org/10.1016/j.jmva.2014.08.013
0047-259X/

Statistical interests are involved with not only the consistency of estimators but also its asymptotic distributions. The latter can be used to construct the confidence interval on parameters and performed the hypothesis testing. In the asymptotic framework considered in this paper, the number of network vertices goes to infinity and the number of parameter is identical to the dimension of networks (i.e., the number of vertices). Instead of studying a more complicated situation on linear combinations of all MLEs, we describe the central limit theorems for the MLEs through the asymptotic behavior of a finite number of the MLEs, although the total number of parameters goes to infinity. With this point, we aim to establish the asymptotic normality of the MLEs when edges take three types of weights as in [11]. A key step in our proofs applies a highly accurate approximate inverse of the Fisher information matrix by Yan and Xu [24].

The remainder of this article is organized as follows. In Section 2, we lay out the asymptotic distributions of the MLEs in the maximum entropy models with the finite discrete, infinite discrete and continuous weighted edges in Sections 2.1–2.3, respectively. Simulation studies are given in Section 3. Section 4 concludes with summary and discussion. All proofs are relegated to Appendices.

## 2. Asymptotic normalities

We first give a brief description on the maximum entropy models. Consider an undirected graph $\mathcal{G}_n$ with no self-loops on $n$ vertices labeled by "$1, \ldots, n$". Let $a_{ij}$ be the weight of edge $(i,j)$ taking values from the set $\Omega$, where $\Omega$ could be a finite discrete, infinite discrete or continuous set. Define $d_i = \sum_{j\neq i} a_{ij}$ as the degree of vertex $i$, and $\mathbf{d} = (d_1, \ldots, d_n)^T$ is the degree sequence of $\mathcal{G}_n$. Let $\mathcal{S}$ be a $\sigma$-algebra over the set $\Omega$ of all possible values of $a_{ij}$, $1 \le i < j \le n$. Assume there is a canonical $\sigma$-finite probability measure $\nu$ on $(\Omega, \mathcal{S})$. Let $\nu^{\binom{n}{2}}$ be the product measure on $\Omega^{\binom{n}{2}}$. The maximum entropy models assume that the density function of the symmetric adjacent matrix $\mathbf{a} = (a_{ij})_{i,j=1}^n$ with respective to $\nu^{\binom{n}{2}}$ has the exponential form with the degree sequence as natural sufficient statistics,[1] i.e.,

$$p_\theta(\mathbf{a}) = \exp\big(-\boldsymbol{\theta}^T\mathbf{d} - z(\boldsymbol{\theta})\big), \tag{1}$$

where $z(\boldsymbol{\theta})$ is the normalizing constant,

$$z(\boldsymbol{\theta}) = \log \int_{S^{\binom{n}{2}}} \exp\big(-\boldsymbol{\theta}^T\mathbf{d}\big)\ \nu^{\binom{n}{2}}(d\mathbf{a}) = \log \prod_{1\le i<j\le n} \int_S \exp\big(-(\theta_i+\theta_j)a_{ij}\big)\ \nu(da_{ij}),$$

and for fixed $n$, the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ belongs to the natural parameter space (p. 1, [5])

$$\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^n : z(\boldsymbol{\theta}) < \infty\}.$$

The parameters $\theta_1, \ldots, \theta_n$ can be interpreted as the strength of each vertex that determines how strongly the vertices are connected to each other. The probability distribution (1) implies that the edges $(i,j)$ for all $1 \le i < j \le n$ are mutually independent. Since the sample is one realization of a graph, the density function in (1) is also the likelihood function. We can see that the solution to $-\nabla z(\boldsymbol{\theta}) = \mathbf{d}$ is the maximum likelihood estimator (MLE) of $\theta$.

We now consider the asymptotic distributions of the MLEs as the number of parameters goes to infinity. Let $V_n = (v_{ij})_{i,j=1,\ldots,n}$ be the Fisher information matrix of the parameters $\theta_1, \ldots, \theta_n$. It can be written as

$$V_n = \frac{\partial^2 z(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}.$$

For three common types of weights as introduced in Section 1, $V_n$ is the diagonal dominant matrix with nonnegative entries. This property is crucially used in the proof of the central limit theorem on the MLE.

### 2.1. Finite discrete weights

When network edges take finite discrete weights, we assume $\Omega = \{0, 1, \ldots, q-1\}$ with $q$ a fixed integer. In this case, $\nu$ is the counting measure and the edge weights $a_{ij}$ are independent multinomial random variables with the probability:

$$P(a_{ij} = a) = \frac{e^{a(\theta_i+\theta_j)}}{\sum_{k=0}^{q-1} e^{k(\theta_i+\theta_j)}}, \quad a = 0, 1, \ldots, q-1.$$

This model is a direct generalization of the $\beta$-model that only consider the dichotomous edges. The normalizing constant is

$$z(\boldsymbol{\theta}) = \sum_{1\le i<j\le n} \log \sum_{a=0}^{q-1} e^{-(\theta_i+\theta_j)a},$$

[1] Following Hillar and Wibisono [11], we use $-\theta$ in the parameterization (1) instead of the classical $\theta$ since it will simplify the notations in the later presentation.