Maximum likelihood estimators for discrete exponential families and random graphs

Tomasz Skalski

Wrocław University of Science and Technology

Université d'Angers

(joint research with K. Bogdan and M. Bosy)

Statistical Learning Seminars, 15^{TH} October 2020

Discrete exponential families – Notation

- $\mathcal{X} = \{x_1, \ldots, x_K\}$ – finite state space, $K = |\mathcal{X}|$
- $\mu : \mathcal{X} \to (0, \infty)$ – weight function
- $\mathcal{B} \subset \mathbb{R}^{\mathcal{X}}$ – linear space of functions ($\phi = \mathbb{1} \in \mathcal{B}$)
- $\mathcal{B}_+ = \{\phi \in \mathcal{B} : \phi \geq 0\}$ – subclass (cone) of non-negative functions
- $Z(\phi) = \sum_{x \in \mathcal{X}} e^{\phi(x)} \mu(x)$ – normalising constant (partition function)
- $p = e(\phi) = \frac{e^{\phi}}{Z(\phi)}$ – exponential density
- $e(\mathcal{B}) = \{p = e(\phi) : \phi \in \mathcal{B}\}$ – exponential family

MLE

Definition

Let $x_1, \ldots, x_n$ *be a sample from the finite set* $\mathcal{X}$ *and let* $\phi \in \mathcal{B}$. *The likelihood function of* $p = e(\phi)$ *is defined as:*

$$L_p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i).$$

log-likelihood function: $\ell_p(x_1, \ldots, x_n) = \log L_p(x_1, \ldots, x_n)$.

Definition

The $\hat{p} \in e(\mathcal{B})$ *is called the maximum likelihood estimator (MLE), if*

$$L_{\hat{p}}(x_1, \ldots, x_n) = \sup_{p \in e(\mathcal{B})} L_p(x_1, \ldots, x_n).$$

Existence of MLE

History

- *O. Barndorff-Nielsen (1978) – criterion of existence of MLE for the exponential families in terms of convex geometry*
- *S. J. Haberman (1974) – criterion of existence of MLE in hierarchical log-linear models*
- *K. Bogdan, M. Bogdan (2000) – criterion of existence of for exponential families of continuous functions on* $[0, 1]$ *in terms of sets of uniqueness.*
- *N. Eriksson, S. E. Fienberg, A. Rinaldo, S. Sullivant (2006) – interpretation of the criterion in terms of polyhedral geometry*
- *A. Rinaldo, S. E. Fienberg, Y. Zhou (2009) – application to exponential models of random graphs (ERGM).*
- *K. Bogdan, M. Bosy, TS (2019+, this talk) – criterion of existence of MLE in discrete exponential families in terms of sets of uniqueness.*

Existence of MLE

Maximization of likelihood is fundamental in estimation, model selection and testing. In many procedures it is important to know if MLE actually exists for given data $x_1 \ldots, x_n$ and the linear space of exponents.

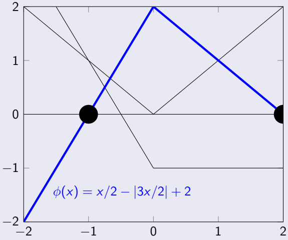
Sets of uniqueness

Definition

We say that $U \subset \mathcal{X}$ *is a set of uniqueness for* $\mathcal{B}$*, if* $\phi \equiv 0$ *is the only function in* $\mathcal{B}$ *such that* $\phi(U) = 0$*.*

Example

Let $\mathcal{X} = \{-2, -1, 0, 1, 2\}$*. Let* $\mathcal{B}$ *denote the class of all the real functions on* $\mathcal{X}$ *that are linear (affine) both on* $\{-2, -1, 0\}$ *and on* $\{0, 1, 2\}$*.*



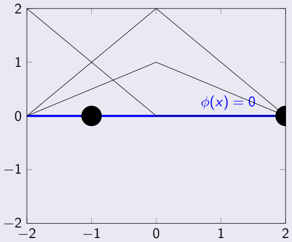
Then the set $\{-1, 1, 2\}$ *is of uniqueness for* $\mathcal{B}$*, but the set* $\{-1, 2\}$ *is not.*

Definition

U is a set of uniqueness for $\mathcal{B}_+$, if $[\phi \in \mathcal{B}_+,\ \phi(U) = 0\,] \Rightarrow [\,\phi \equiv 0\,]$.

Example

Again, let $\mathcal{X} = \{-2, -1, 0, 1, 2\}$ and let $\mathcal{B}$ be the class of all the real functions on $\mathcal{X}$ that are linear (affine) on $\{-2, -1, 0\}$ and on $\{0, 1, 2\}$.



Then the set $\{-1, 2\}$ is of uniqueness for $\mathcal{B}_+$.

Existence of MLE – main criterion

Theorem (K. Bogdan, M. Bosy, TS (2019+))

The maximum likelihood estimator for $e(\mathcal{B})$ and $x_1, \ldots, x_n \in \mathcal{X}$ exists if and only if $\{x_1, \ldots, x_n\}$ is a set of uniqueness for $\mathcal{B}_+$.

Proof.

($\Rightarrow$) If $\{x_1, \ldots, x_n\}$ is not of uniqueness for $\mathcal{B}_+$, we may subtract from every candidate for MLE ϕ a non-negative function ψ vanishing on $\{x_1, \ldots, x_n\}$, so $\psi - \phi = \psi$ on $\{x_1, \ldots, x_n\}$. Thus $Z(\psi - \phi) < Z(\psi)$ and the resulting likelihood is increased. □

Existence of MLE – main criterion

Theorem (K. Bogdan, M. Bosy, TS (2019+))

The maximum likelihood estimator for $e(\mathcal{B})$ and $x_1, \ldots, x_n \in \mathcal{X}$ exists if and only if $\{x_1, \ldots, x_n\}$ is a set of uniqueness for $\mathcal{B}_+$.

Proof.

($\Leftarrow$) We introduce a special seminorm related to given set of uniqueness on $\mathcal{B}$ and compare it with an oscillation seminorm. □

Applications

There are two types of application we propose:

- Conditions for the existence of MLE for specific exponential families
- Probability bounds for MLE for i.i.d. samples

For the i.i.d. random variables $X_1, X_2, \ldots$ valued in $\mathcal{X}$ it will be useful to define the following (random) time:

$$\nu_{uniq} = \inf\{n \geq 1 : \{X_1, \ldots, X_n\} \text{ is a set of uniqueness for } \mathcal{B}_+\}$$

Threshold functions

Definition (Threshold)

A function $n^* = n^*(K)$ *is a threshold of the size of the sample* $\mathbb{X} = (X_1, \ldots, X_n)$ *for a given (monotone) property* $\mathscr{P}$ *if*

$$\lim_{K\to\infty} \mathbb{P}(\mathbb{X} \in \mathscr{P}) = \begin{cases} 0 & \text{if } n(K)/n^*(K) \to 0, \qquad K \to \infty, \\ 1 & \text{if } n(K)/n^*(K) \to \infty, \qquad K \to \infty. \end{cases}$$

Definition (Sharp threshold)

A function $n^* = n^*(K)$ *is a sharp threshold of the size of the sample* $\mathbb{X} = (X_1, \ldots, X_n)$ *for a given (monotone) property* $\mathscr{P}$ *if for every* $\varepsilon > 0$

$$\lim_{K\to\infty} \mathbb{P}(\mathbb{X} \in \mathscr{P}) = \begin{cases} 0 & \text{if } n(K)/n^*(K) < 1 - \varepsilon, \\ 1 & \text{if } n(K)/n^*(K) > 1 + \varepsilon. \end{cases}$$

Threshold functions

Example

Examples of monotone properties:

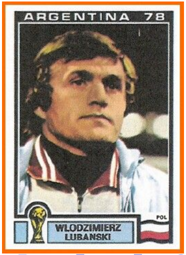
- $\mathbb{X} = (X_1, \ldots, X_n) \supset \mathcal{X}$,
- $|\{X_1, \ldots, X_n\}| \geq 3$,
- $\exists \;\; 1 \leq i, j \leq n : \quad X_i + X_j = K + 1$,
- $\ldots$

Let $\mathcal{B} = \mathbb{R}^{\mathcal{X}}$. As $\mathcal{X}$ is the only set of uniqueness for $\mathcal{B}_+$, we observe that

Lemma

MLE for $e(\mathbb{R}^{\mathcal{X}})$ and $x_1, \ldots, x_n$ exists if and only if $\{x_1, \ldots, x_n\} = \mathcal{X}$.

Then the existence of MLE for $\{x_1, \ldots, x_n\}$ is a reformulation of the Coupon Collector Problem.



Applications – $\mathbb{R}^{\mathcal{X}}$

Corollary

Let $\mathcal{B} = \mathbb{R}^{\mathcal{X}}$ *and* $K = |\mathcal{X}|$. *Let* $X_1, X_2, \ldots$ *be independent random variables, each with uniform distribution on* $\mathcal{X}$. *Then, for every* $c \in \mathbb{R}$,

$$\lim_{K \to \infty} (\nu_{uniq} < K \log K + Kc) = e^{-e^{-c}}.$$

In particular, $n^*(K) = K \log K$ *is a sharp threshold of the sample size for the existence of MLE for* $e(\mathcal{X})$.

Applications – Rademacher functions

For $k \in \mathbb{N}$ consider the discrete hypercube $\mathcal{X} = Q_k = \{-1, 1\}^k$. Let $K = |\mathcal{X}| = 2^k$.
For $j = 1, \ldots, k$ we define Rademacher functions:

$$r_j(\chi) = \chi_j, \quad \chi = (\chi_1, \ldots, \chi_k) \in Q_k.$$

Denote $r_0(\chi) = 1$.

Applications – Rademacher functions

Theorem (K. Bogdan, M. Bosy, TS (2019+))

Let $\mathcal{B}^k = Lin\{r_0, r_1, \ldots, r_k\}$. MLE for $e(\mathcal{B}^k)$ and $x_1, \ldots, x_n \in Q_k$ exists if and only if for all $j = 1, \ldots, k$ we have $\{r_j(x_1), \ldots, r_j(x_n)\} = \{-1, 1\}$.

In other words, the condition above is satisfied if and only if $\{x_1, \ldots, x_n\}$ intersects with every half-cube of Q_k,
e.g. $\{x_1 = (-1, -1, \ldots, -1), x_2 = (1, 1, \ldots, 1)\}$.

Theorem (K. Bogdan, M. Bosy, TS (2019+))

Let $k \in \mathbb{N}$, $n(k) = \log_2 k + b + o(1)$. Let $X_1, \ldots, X_{n(k)}$ be independent random variables, each with uniform distribution on Q_k. Then

$$\lim_{k\to\infty} \mathbb{P}(\{X_1, \ldots, X_{n(k)}\} \textit{ is a set of uniqueness for } \mathcal{B}_+) = \\ = \exp\{-2^{1-b}\}.$$

and $n^*(K) = \log_2 k = \log_2 \log_2 K$ is a sharp threshold of the sample size for the existence of MLE for $e(\mathcal{B}^k)$ and i.i.d. uniform samples on Q_k.

We consider simple undirected graphs containing no loops or multiple edges. Let N and m denote the number of vertices and edges of the graph. Let $\mathcal{G}_N$ denote the family of all the graphs with N vertices.
For graphs $G = (V, E_1)$, $H = (V, E_2)$ we let, as usual,

$$G \cup H := (V, E_1 \cup E_2), \qquad G \cap H := (V, E_1 \cap E_2).$$

Also, by $G \subset H$ we mean $E_1 \subset E_2$.
We define $\chi_{r,s}(G) = 1 - 2\mathbb{1}_G(r, s)$ and consider the following linear space

$$\mathcal{B}^{\mathcal{G}_N} = \mathrm{Lin}\left\{ 1, \chi_{r,s}(G) : 1 \leq r < s \leq N \right\}.$$

Consider coefficients $c \in \mathbb{R}^{\binom{V}{2}}$, indexed by the edges of the complete graph K_N, and the following exponential family:

$$e(\mathcal{B}^{\mathcal{G}_N}) = \left\{ p_c := e^{\phi_c - \psi(\phi_c)} : c \in \mathbb{R}^{\binom{V}{2}} \right\},$$

where

$$\phi_c(G) = \sum_{(r,s)\in\binom{V}{2}} c_{r,s}\chi_{r,s}(G), \qquad \psi(\phi_c) = \log \sum_{G\in\mathcal{G}_N} e^{\phi_c(G)},$$

and $G \in \mathcal{G}_N$.

Observation

Fix $c \in \mathbb{R}^{\binom{V}{2}}$. In the random graph $\mathbb{G}$ sampled from $p_c \in e(\mathcal{B}^{\mathcal{G}_N})$, each edge (r, s) appears independently with probability

$$p_{r,s} = \frac{e^{c_{r,s}}}{1 + e^{c_{r,s}}}.$$

Applications – ERGM

Theorem (K. Bogdan, M. Bosy, TS (2019+))

MLE for $e(\mathcal{B}^{\mathcal{G}_N})$ *and* $G_1, \ldots, G_n \in \mathcal{G}_N$ *exists if and only if*

$$\bigcup_{i=1}^{n} G_i = K_N \qquad \text{and} \qquad \bigcap_{i=1}^{n} G_i = \overline{K_N}.$$

Lemma (K. Bogdan, M. Bosy, TS (2019+))

Let $\{\mathbb{G}_1, \ldots, \mathbb{G}_n\}$ *be independent random graphs from* $p_c \in e(\mathcal{B}^{\mathcal{G}_N})$*. Then the probability of the existence of MLE for* $e(\mathcal{B}^{\mathcal{G}_N})$ *equals*

$$\prod_{1 \le r < s \le N} \left(1 - p_{r,s}^n - (1 - p_{r,s})^n\right).$$

In particular, $n^*(N) = \log N$ *is a threshold of the sample size* n *for the existence of MLE for* $e(\mathcal{B}^{\mathcal{G}_N})$*.*

Let $k \in \mathbb{N}, 1 \leq q \leq k$, and $\mathcal{B}_q^k = \mathrm{Lin}\{w_S : S \subset \{1, \ldots, k\} \text{ and } |S| \leq q\}$, where $w_S(x) = \prod_{i \in S} r_i(x), x \in Q_k, S \subset \{1, \ldots, k\}$, are the Walsh functions.

Observation

$\mathcal{B}_q^k$ is the linear space spanned by indicator functions of the sub-cubes of Q_k, obtained by fixing q out of k coordinates.

- $q = 1$: Rademacher functions (already discussed)
- $q = 2$: The Ising model

$\mathcal{B}^k_{k-1}$ corresponds to indicators of edges of Q_k. Consider the following partition: $Q_k = \mathcal{E} \cup \mathcal{O}$:

Definition

- $\mathcal{E} := \{\chi \in Q_k : \chi \textit{ has even number of positive coordinates}\}$
- $\mathcal{O} := \{\chi \in Q_k : \chi \textit{ has odd number of positive coordinates}\}$

Theorem (K. Bogdan, M. Bosy, TS (2019+))

MLE exists for $e(\mathcal{B}^k_{k-1})$ *and* $x_1, \ldots, x_n \in Q_k$ *if and only if* $\mathcal{E} \subset \{x_1, \ldots, x_n\}$ *or* $\mathcal{O} \subset \{x_1, \ldots, x_n\}$.

References

Barndorff-Nielsen, O. (1978)
"'Information and exponential families in statistical theory"', John Wiley & Sons Ltd., Chichester. Wiley Series in Probability and Mathematical Statistics.

Bogdan, K., Bogdan, M. (2000)
"'On existence of maximum likelihood estimators in exponential families"', Statistics, 34(2):137–149.

Bogdan, K., Bosy, M., Skalski, T. (2019+)
"'Maximum likelihood estimation for discrete exponential families and random graphs"', https://arxiv.org/abs/1911.13143

Erdős, P., Rényi, A. (1959)
"'On random graphs. I."', Publ. Math. Debrecen, 6:290–297.

Erdős, P., Rényi, A. (1961)
"'On a classical problem of probability theory"', Magyar Tud. Akad. Mat. Kutató Int. Közl., 6:215-220.

Eriksson, N., Fienberg, S. E., Rinaldo, A., Sullivant, S. (2006)
"'Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models"', J. Symbolic Comput., 41(2):222–233.

Haberman, S. J. (1974)
"'The analysis of frequency data"', The University of Chicago Press, Chicago, Ill.-London. Statistical Research Monographs, Vol. IV.

Laplace, P.-S. (1812)
"'Théorie analytique des probabilités"', pp. 194–195.

Rinaldo, A., Fienberg, S. E., Zhou, Y. (2009)
"'On the geometry of discrete exponential families with application to exponential random graph models"', Electron. J. Stat., 3:446–484

Tack för att ni kom idag!