Maximization of the information divergence from multinomial distributions

Jozef Jurček
Charles University in Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

SUMMARY
Explicit solution of the problem of maximization of information divergence from the family of multinomial distributions is presented. General problem of maximization of information divergence from an exponential family has emerged in probabilistic models for evolution and learning in neural networks, based on infomax principles. The maximizers admit interpretation as stochastic systems with high complexity w.r.t. exponential family.

1 Introduction

PROBLEM “Find all empirical distributions of data, which lies farthest from the model, when modelling by the multinomial family; in the sense of information divergence and the method of maximum likelihood.”

- Exponential family
  \[ E_{\mu,f} = \{ Q_{\mu,f,\theta} \sim (e^{(\theta,f(z))} \mu(z)) \}_{z \in Z} : \theta \in \mathbb{R}^d \]
  * \( \mu \) nonzero reference measure on a finite set \( Z \)
  * \( f : Z \to \mathbb{R}^d \) the directional statistics

- Divergence of a pm \( P \) (on \( Z \)) from \( \nu \) (on \( Z \))
  \[ D(P||\nu) = \sum_{z \in Z} \frac{P(z)}{\nu(z)} s(P) \subseteq s(\nu), \]
  * \( s(\cdot) \) the support, from now on, let \( s(\mu) = Z \)

- Divergence of a pm \( P \) from exponential family \( E = E_{\mu,f} \)
  \[ D(P||E) = \inf_{Q \in E} D(P||Q) = \min_{Q \in E} D(P||Q) \]

THEOREM 1 There exist unique r-projection (generalized MLE) \( P^E \) arg min \( D(P||Q) \).

For \( P \) empirical distribution, s.t. \( P^E \in E \), \( P^E \) is the MLE for data with exponential distribution \( P \). Details in [2].

- Multinomial family \( N \) indep. trials, each with \( n \) outcomes
  \[ M = \left\{ Q(z) = \left( \frac{N}{z} \right) \left( \sum_{j=1}^{n} p_j^{z} \right) \right\} : p \in \mathcal{P}(1:n) \]
  * \( \mathcal{P}(1:n) \) all pm's on \( [1:n] = [n] = \{1, \ldots, n\} \)
  * \( Z = \{ z \in [0:N]^n : \sum_{j=1}^{n} z_j = N \} \)
  * \( M = E_{\mu,f} \) with \( f(z) = z, \mu(z) = \binom{N}{z} \)

PROBLEM Calculate \( \sup_{P \in \mathcal{P}(Z)} D(P||M) \) and find all maximizers \( \arg \sup_{P \in \mathcal{P}(Z)} D(P||M) \).

Generalization of [3].

2 Preliminaries

- \( \pi : [1:N] \to [1:N] \) set of all permutations \( [1:N]! \)
  \( x \in X = [1:n]^N, x^\pi = (x_{\pi(1)}, \ldots, x_{\pi(N)})^\top, Q^\pi(x) = Q(x^\pi) \)

- Exchangeable distributions' family
  \[ E := \{ P \in \mathcal{P}(X) : P(x) = P(x^\pi), x \in X; \forall \pi \in [1:N]! \} \]

- 1-factorizable distributions' family
  \[ F := \{ Q \in \mathcal{P}(X) : Q(x) = \prod_{i=1}^n Q_i(x_i), x \in X \}, Q_i \) marginal

- \( X^* := \{ x \in X : \forall j \in [1:n] : \{ i \in [1:N] : x_i = j \} = z_j \} \)

3 Result

- \( e_{\pi} = e^{\odot j} = (0, \ldots, 0, 1, 0, \ldots, 0, 1, 0, \ldots, 0, \ldots, 0)^\top, e_{\pi,j} = \delta_{2\pi,j} \)
  \( e_{\pi,j} = (0, \ldots, 0, 1_k, 0, \ldots, 0, 1_l, 0, \ldots, 0, \ldots, 0)^\top, e_{k,l} = \delta_{2k,l}, k < l \)

COROLLARY 4 (Maximization of multi-information)
\[ \arg \max_{P \in \mathcal{P}(X)} D(P||M) = h^{-1} \left( \mathcal{M} \cap \arg \sup_{P \in \mathcal{P}(X)} D(P||F) \right) \]

If \( N \geq 2 \)
\[ = \left\{ P_n = \frac{1}{n} \left( \sum_{j \in \{1:n\}, \pi \} e_{\pi,j} \right) : \pi \in \{1:n\}, \forall j, k \in [n] \right\} \]

If \( N > 2 \)
\[ = \left\{ P_n = \frac{1}{n} \left( \sum_{j \in \{1:n\}, \pi \} e_{\pi,j} \right) : \pi \in \{1:n\}, \forall j, k \in [n] \right\} \]

For every maximizer \( P_n = P_i^n \) on \( z \in Z \).

(1) More general situation and essentially simpler proof than in [3]
\( \Leftarrow \) Application of Theorem 1, Lemma 2 & Theorem 3

(2) \( D(P^n||M_k), \) where \( M_k = h^{-1}(\mathcal{M} \cap \mathcal{F}_k) \) and \( \mathcal{F}_k, \) the k-factorizable

4 Example
\[ [\mathcal{P}(Z) \to \mathcal{F} \subseteq \mathcal{P}(X); N = n = 2.] \]

References