

# ASYMPTOTIC IDENTITY OF $\mu$ -PROJECTIONS AND $I$ -PROJECTIONS

M. GRENDAR

*Dedicated to Mar*

ABSTRACT. Concept of  $\mu$ -projection, closely related to that of constrained mode of multinomial distribution, is introduced. Sets of  $\mu$ -projections and  $I$ -projections are shown to be asymptotically identical.

## INTRODUCTION

At [1], a convergence of constrained mode of a multinomial distribution for sample size  $n \rightarrow \infty$  to  $I$ -projection of  $q$  on the constraining set  $\mathcal{H}$ , was investigated and illustrated by a numeric examples. There the point-wise convergence was also proven, for  $\mathcal{H}$  defined by a differentiable constraint (cf. [1], Thm 1).

Here, a concept of  $\mu$ -projection, closely related to that of the constrained mode of the multinomial distribution, is introduced. For a general feasible set it is shown here that  $\mu$ -projections are asymptotically indistinguishable from  $I$ -projections.

## TERMINOLOGY AND NOTATION

Let  $\{X\}_{l=1}^n$  be a sequence of independently and identically distributed random variables with a common law (measure) on a measurable space. Let the measure be concentrated on  $m$  atoms from a set  $\mathcal{X} \triangleq \{x_1, x_2, \dots, x_m\}$  called support or alphabet. Let  $q_i$  denote the probability (measure) of  $i$ -th element of  $\mathcal{X}$ . Let  $\mathcal{P}(\mathcal{X})$  be a set of all probability mass functions (pmf's) on  $\mathcal{X}$ .

A type (also called  $n$ -type, empirical measure, frequency distribution or occurrence vector) induced by a sequence  $\{X\}_{l=1}^n$  is the pmf  $\nu^n \in \mathcal{P}(\mathcal{X})$  whose  $i$ -th element  $\nu_i^n$  is defined as:  $\nu_i^n \triangleq n_i/n$  where  $n_i \triangleq \sum_{l=1}^n I(X_l = x_i)$ ; there  $I(\cdot)$  is the characteristic function. Multiplicity  $\Gamma(\nu^n)$  of type  $\nu^n$  is:  $\Gamma(\nu^n) \triangleq n! / \prod_{i=1}^m n_i!$ .

Let  $\Pi \subseteq \mathcal{P}(\mathcal{X})$ . Let  $\mathcal{P}_n$  denote a subset of  $\mathcal{P}(\mathcal{X})$  which consists of all  $n$ -types. Let  $\Pi_n \triangleq \Pi \cap \mathcal{P}_n$ .

$\mu$ -projection  $\hat{\nu}^n$  of  $q$  on  $\Pi_n$  is defined as:  $\hat{\nu}^n \triangleq \arg \sup_{\nu^n \in \Pi_n} \pi(\nu^n; q)$ , where  $\pi(\nu^n; q) \triangleq \Gamma(\nu^n) \prod (q_i)^{n\nu_i^n}$ . Equivalently, for  $\Pi_n \neq \emptyset$   $\mu$ -projection  $\hat{\nu}^n$  of  $q$  on  $\Pi_n$  can be defined in terms of supremum of a conditional probability:  $\hat{\nu}^n \triangleq$

---

*2000 Mathematics Subject Classification.* Primary 62B15; Secondary 94A15, 65J20.

*Key words and phrases.* Information projection,  $\mu$ -projection,  $J$ -projection, Maximum Probability method, Maximum Entropy method, Jeffreys Entropy Maximization method.

Supported by VEGA 1/7295/20 grant. Valuable discussions with Ondrej Šuch are gratefully acknowledged. Anonymous referee is thanked for careful reading and inspiring comments. Lapses are mine.

Typeset by  $\mathcal{A}\mathcal{M}\mathcal{S}$ - $\text{\TeX}$

$\arg \sup_{\nu^n \in \Pi_n} \pi(\nu^n | \nu^n \in \Pi_n; q \mapsto \nu^n)$  where  $\pi(\nu^n | \nu^n \in \Pi_n; q \mapsto \nu^n)$  denotes the conditional probability that if  $q$  generated an  $n$ -type from  $\Pi_n$  then it is just the type  $\nu^n$ .

$I$ -projection  $\hat{p}$  of  $q$  on  $\Pi$  is  $\hat{p} \triangleq \arg \inf_{p \in \Pi} I(p|q)$ , where  $I(p|q) \triangleq \sum_{\mathcal{X}} p_i \log \frac{p_i}{q_i}$  where  $I(\cdot|\cdot)$  is Kullback-Leibler distance, information divergence or (minus) relative entropy.

#### ASYMPTOTIC IDENTITY OF $\mu$ -PROJECTIONS AND $I$ -PROJECTIONS

**Theorem 1.** Let  $\mathcal{X}$  be a finite set. Let  $\mathcal{M}_n$  be set of all  $\mu$ -projections of  $q$  on  $\Pi_n$ . Let  $\mathcal{I}$  be set of all  $I$ -projections of  $q$  on  $\Pi$ . For  $n \rightarrow \infty$ ,  $\mathcal{M}_n = \mathcal{I}$ .

*Proof.* Necessary and sufficient conditions for  $\hat{\nu}^n$  to be a  $\mu$ -projection of  $q$  on  $\Pi_n$  are: *a)*  $\pi(\hat{\nu}^n; q) \geq \pi(\nu^n; q)$ ,  $\forall \nu^n \in \Pi_n$ ; *b)* whenever  $\tilde{\nu}^n$  has the property *a)* then  $\pi(\hat{\nu}^n; q) \leq \pi(\tilde{\nu}^n; q)$ . The requirement *a)* can be equivalently stated as:

$$(1) \quad \left( \prod \frac{n_i!}{\hat{n}_i!} \right)^{1/n} \geq \left( \prod q_i^{n_i - \hat{n}_i} \right)^{1/n}$$

and *b)* similarly. Standard inequality  $(n/e)^n < n! < n(n/e)^n$  (valid for  $n > 6$ ) allows to bind the LHS of (1):

$$(2) \quad \frac{\prod (\nu_i^n)^{\nu_i^n}}{n^{m/n} \prod (\hat{\nu}_i^n)^{\hat{\nu}_i^n} (\prod \hat{\nu}_i^n)^{1/n}} < \text{LHS} < \frac{n^{m/n} \prod (\nu_i^n)^{\nu_i^n} (\prod \nu_i^n)^{1/n}}{\prod (\hat{\nu}_i^n)^{\hat{\nu}_i^n}}$$

and similar bounds can be stated in the case of the requirement *b)*<sup>1</sup>. Since  $m$  is by assumption finite, as  $n \rightarrow \infty$  the upper and lower bounds at (2) collapse into  $\prod (\nu_i^n)^{\nu_i^n} / \prod (\hat{\nu}_i^n)^{\hat{\nu}_i^n}$ . Consequently, the necessary and sufficient conditions *a)*, *b)* for  $\mu$ -projection turn as  $n \rightarrow \infty$  into (expressed in an equivalent log-form): *i)*  $\sum (\nu_i^n \log \nu_i^n - \hat{\nu}_i^n \log \hat{\nu}_i^n) \geq \sum (\nu_i^n - \hat{\nu}_i^n) \log q_i$  for all  $\nu^n \in \Pi_n$ ; and *ii)* whenever  $\tilde{\nu}^n$  has the property *i)* then  $\sum \hat{\nu}_i^n \log \hat{\nu}_i^n - \tilde{\nu}_i^n \log \tilde{\nu}_i^n \geq \sum (\hat{\nu}_i^n - \tilde{\nu}_i^n) \log q_i$ .

Necessary and sufficient conditions for  $\hat{p}$  to be an  $I$ -projection of  $q$  on  $\Pi$  are: *I)*  $\sum (p_i \log p_i - \hat{p}_i \log \hat{p}_i) \geq \sum (p_i - \hat{p}_i) \log q_i$  for all  $p \in \Pi$ ; and *II)* whenever  $\tilde{p}$  has the property *I)* then  $\sum (\hat{p}_i \log \hat{p}_i - \tilde{p}_i \log \tilde{p}_i) \geq \sum (\hat{p}_i - \tilde{p}_i) \log q_i$ .

A comparison of *i)*, *ii)* and *I)*, *II)* then completes the proof.  $\square$

*Note.* Since  $\pi(\nu^n; q)$  is defined for  $\nu^n \in \mathcal{Q}^m$ ,  $\mu$ -projection can be defined for  $\Pi_n$ , when  $n$  is finite, only. Theorem 1 makes possible to extend the definition by defining a  $\mu$ -projection of  $q$  on  $\Pi$  as follows:  $\hat{\nu} \triangleq \arg \sup_{r \in \Pi} - \sum_{\mathcal{X}} r_i \log \frac{r_i}{q_i}$ .

#### Asymptotic identity of $\gamma$ -projections and $J$ -projections.

$\gamma$  projection  $\tilde{\nu}^n$  of  $q \in \mathcal{Q}^m$  on  $\Pi_n$  is  $\tilde{\nu}^n \triangleq \arg \sup_{\nu^n \in \Pi_n} \pi(\nu^n; q) \pi(nq; \nu^n)$ .

$J$ -projection (or Jeffreys' projection)  $\tilde{p}$  of  $q \in \mathcal{Q}^m$  on  $\Pi$  is defined as:  $\tilde{p} \triangleq \arg \inf_{p \in \Pi} \sum_{\mathcal{X}} p_i \log \frac{p_i}{q_i} + q_i \log \frac{q_i}{p_i}$ .

**Theorem 2.** Let  $q \in \mathcal{Q}^m$ . Let  $\mathcal{X}$  be a finite set. Let  $\mathcal{G}_n$  be set of all  $\gamma$ -projections of  $q$  on  $\Pi_n$ . Let  $\mathcal{J}$  be set of all  $J$ -projections of  $q$  on  $\Pi$ . Let  $n_0$  be denominator of the smallest common divisor of  $q_1, q_2, \dots, q_m$ . Let  $n = un_0$ ,  $u \in \mathcal{N}$ . Let  $u \rightarrow \infty$ . Then  $\mathcal{G}_n = \mathcal{J}$ .

*Proof.* Along the same lines as proof of the Theorem 1.

<sup>1</sup>Note that if an  $i$ -th component of a type is zero then it does not change value of  $\pi(\nu^n; q)$ . Thus it is assumed that product operations at (1), (2) are performed on non-zero components only.

## COMMENTS

1) Theorem 1 which is intended to replace Thm 1 of [1] (a.k.a. MaxProb/MaxEnt Thm) shows that  $\mu$ -projections are asymptotically indistinguishable from  $I$ -projections. In other words, Maximum Probability (MaxProb, cf. [1]) and Relative Entropy Maximization method (REM/MaxEnt) methods, when applied to the Boltzmann-Jaynes inverse problem (cf. [2]), make asymptotically the same choice. However, for finite  $n$ ,  $I$ - and  $\mu$ -projections on  $\Pi_n$  are in general different. In light of this, the asymptotic identity of  $I$ - and  $\mu$ -projections can be viewed in two ways: either as saying that 1)  $I$ -projection is asymptotic form of  $\mu$ -projection (the view presented at [1]) or that 2)  $I$ - and  $\mu$ -projections asymptotically coincide. If one adopts the second view then for a finite  $n$  it is necessary to face a challenge of choosing between  $I$ - and  $\mu$ -projections.

2)  $\mu$ -projection is related to the probability  $\pi(\nu^n; q)$ , hence it can be viewed as a 'UNI'-projection.  $\gamma$ -projection is related to  $\pi(\nu^n; q) \pi(nq; \nu^n)$ , thus it can be viewed as an 'AND'-projection (cf. [5]). It is then natural to consider also an 'OR'-projection defined as  $\nu^n \triangleq \arg \sup_{\nu^n \in \Pi_n} \pi(\nu^n; q) + \pi(nq; \nu^n)$ . However there seems to be no analytic way how to define its asymptotic form.

3) In the same manner it can be proven that infimum (minimum) probability type(s)  $\arg \inf_{\nu^n \in \Pi_n} \pi(\nu^n; q)$  is asymptotically identical with infimum (minimum) relative entropy distribution(s)  $\arg \inf_{p \in \Pi} - \sum p_i \log(p_i/q_i)$ .

4) Conditioned Weak Law of Large Numbers (CWLLN, [3]) and Gibbs Conditioning Principle (GCP, [4]) which provide for a convex, closed set  $\Pi$  a probabilistic justification of REM thus thanks to the Theorem 1 justify under that conditions MaxProb, as well.

5) CWLLN and GCP hold when (among other things) the  $I$ -projection is unique. An extension of CWLLN to the case of multiple  $I$ -projections was explored at [6]. A proof of Conditional Equi-concentration of Types on  $I$ -projections (ICET) - which sharpens the Asymptotic Equiprobability of  $I$ -projections (cf. [6]) - will be given elsewhere [7]. Theorem 1 of the present paper makes possible to state directly also a Conditional Equi-concentration of Types on  $\mu$ -projections.

## REFERENCES

1. Grendár, M., Jr. and Grendár, M., *What is the question that MaxEnt answers? A probabilistic interpretation*, Bayesian inference and Maximum Entropy methods in Science and Engineering (A. Mohammad-Djafari, ed.), AIP, Melville, NY, 2001, pp. 83-94.
2. ———, *Maximum Entropy method with non-linear moment constraints: challenges*, Bayesian inference and Maximum Entropy methods in Science and Engineering (G. Erickson and Y. Zhai, eds.), AIP, Melville, NY, 2004, pp. 97-109.
3. Cover, T. and Thomas, J., *Elements of Information Theory*, Wiley, NY, 1991.
4. Dembo, A. and Zeitouni, O., *Large Deviations Techniques and Applications*, Springer, NY, 1998.
5. Grendár, M., Jr. and Grendár, M., *On the probabilistic rationale of  $I$ -divergence and  $J$ -divergence minimization*, arxiv/math.PR/0008037, 2000.
6. ———, *Asymptotic Equiprobability of  $I$ -projections*, Acta Univ. M. Belii **10** (2003), 3-8.
7. Grendár, M., *Conditional Equi-concentration of Types*, Focus on Probability Theory, NSP, NY, 2004 in print.

INSTITUTE OF MATHEMATICS AND CS (OF SLOVAK ACADEMY OF SCIENCES (SAS) AND OF MATEJ BEL UNIVERSITY); SEVERNA 5; SK-974 00 BANSKA BYSTRICA; SLOVAKIA INSTITUTE OF MEASUREMENT SCIENCE OF SAS; DUBRAVSKA CESTA 9; SK-841 04 BRATISLAVA; SLOVAKIA  
E-mail address: marian.grendar@savba.sk