Acta Univ. M. Belii
Math no. 10(2003), pp. 3–8

ASYMPTOTIC EQUIPROBABILITY OF I-PROJECTIONS

M. Grendár, Jr. and M. Grendár

Dedicated to Mar

Abstract. When there is only one I-projection $\hat{\boldsymbol{p}}$ of the probability distribution $\boldsymbol{q}$ on the set Π of types, the conditional probability $\pi(\boldsymbol{\nu} \in B(\hat{\boldsymbol{p}}, \epsilon)|\boldsymbol{\nu} \in \Pi)$ that if $\boldsymbol{q}$ generates a type $\boldsymbol{\nu} \in \Pi$ then the type is close to the I-projection, approaches unity, as size n of type-supporting sequences goes to the infinity. What is the probability $\pi(\boldsymbol{\nu} \in B(\cdot)|\boldsymbol{\nu} \in \Pi)$ when Π admits several I-projections? Asymptotic Equiprobability of I-projections Theorem states that for n growing beyond any limit the conditional probability (measure) becomes split among the I-projections equally.

1. Introduction

In subjects as different as Statistical Physics and Computer Tomography, it is not rare to encounter an applied problem which can be cast into the following generic form:

There is a set Π of empirical probability mass functions (types) which could be obtained from random samples of size n, drawn independently from a supposed probability distribution $\boldsymbol{q}$. The problem (from category of ill-posed inverse problems) lays in making a choice of specific type(s) from the set Π.

A result of the Method of Types, which was developed in the Information Theory (cf. [3]), justifies application of the I-divergence minimization method (or equivalently, maximization of relative entropy method, MaxEnt) for making the choice, when n tends to infinity and Π has certain properties. The result is usually known as Conditioned Weak Law of Large Numbers (CWLLN), or as 'Gibbs' conditioning principle' (in large deviations literature, see [4]). For a convenience it will be recalled here, after a brief survey of necessary terminology. Another probabilistic justification of MaxEnt which is not based on large deviations techniques was proposed at [5].

1.1 Gibbs' conditioning principle, entropy concentration.

Let $\mathcal{X}$ be a discrete finite set with m elements and let $\{X_i, i = 1, 2, \ldots, n\}$ be a sequence of size n of identically and independently drawn random variables taking values in $\mathcal{X}$.

2000 Mathematics Subject Classification. Primary 62B15; Secondary 94A15, 65J20.
Key words and phrases. Conditioned law of large numbers, multiple I-projections, non-convex set.
Received 22. 4. 2003; Accepted 20. 6. 2003

A type $\boldsymbol{\nu} \triangleq [n_1, n_2, \dots, n_m]/n$ is an empirical probability mass function which can be based on sequence $\{X_i, i = 1, 2, \dots, n\}$. Thus, n_i denotes number of occurrences of i-th element of $\mathcal{X}$ in the sequence.

Let $\mathcal{P}(\mathcal{X})$ be a set of all probability mass functions (pmf's) on $\mathcal{X}$.

Let the supposed source of the sequences (and hence also of types) be $\boldsymbol{q} \in \mathcal{P}(\mathcal{X})$.

Let $\pi(\boldsymbol{\nu})$ be the probability that $\boldsymbol{q}$ will generate type $\boldsymbol{\nu}$, ie. $\pi(\boldsymbol{\nu}) \triangleq \prod_{i\in\mathcal{X}} q_i^{n\nu_i} \cdot \Gamma(\boldsymbol{\nu})$, where $\Gamma(\boldsymbol{\nu}) \triangleq \frac{n!}{n_1!\, n_2! \dots n_m!}$ is the multiplicity (or Boltzmann's complexion).

Then, $\pi(\boldsymbol{\nu} \in \mathcal{A})$ denotes the probability that $\boldsymbol{q}$ will generate a type $\boldsymbol{\nu}$ which belongs to $\mathcal{A} \subseteq \Pi$, ie. $\pi(\boldsymbol{\nu} \in \mathcal{A}) = \sum_{\boldsymbol{\nu}\in\mathcal{A}} \pi(\boldsymbol{\nu})$. Finally, let $\pi(\boldsymbol{\nu} \in \mathcal{A} | \boldsymbol{\nu} \in \Pi)$ denote the conditional probability that if $\boldsymbol{q}$ generates type $\boldsymbol{\nu} \in \Pi$ then the type belongs to $\mathcal{A}$. It is assumed that the conditional probability exists.

I-projection $\hat{\boldsymbol{p}}$ of $\boldsymbol{q}$ on set $\Pi \subseteq \mathcal{P}(\mathcal{X})$ is such $\hat{\boldsymbol{p}} \in \Pi$ that $I(\hat{\boldsymbol{p}}\|\boldsymbol{q}) = \inf_{\boldsymbol{p}\in\Pi} I(\boldsymbol{p}\|\boldsymbol{q})$, where[1] $I(\boldsymbol{p}\|\boldsymbol{q}) \triangleq \sum_{\mathcal{X}} p_i \log \frac{p_i}{q_i}$ is the I-divergence.

CWLLN. *Let $\hat{\boldsymbol{p}}$ be unique I-projection of $\boldsymbol{q}$ on Π. Let $\boldsymbol{q} \notin \Pi$. Then for any $\epsilon > 0$*

$$\lim_{n\to\infty} \pi\left(|\nu_i - \hat{p}_i| > \epsilon \ \mid \boldsymbol{\nu} \in \Pi\right) = 0 \qquad i = 1, 2, \dots, m \tag{1}$$

Well-studied is the case of convex Π, which ensures existence of the unique I-projection (cf. [2], and [7], [8], [9] for further developments).

Without the assumption of uniqueness of the I-projection, a claim known as the Entropy Concentration Theorem (ECT), weaker than (1), can be still made (see [1]):

ECT. *Let $\Pi \subseteq \mathcal{P}(\mathcal{X})$ be nonempty. Let $\hat{I}$ be such that $\hat{I} \leq I(\boldsymbol{\nu}\|\boldsymbol{q})$ for any $\boldsymbol{\nu} \in \Pi$. Then for any $\epsilon > 0$*

$$\lim_{n\to\infty} \pi\left(\left|I(\boldsymbol{\nu}\|\boldsymbol{q}) - \hat{I}\right| < \epsilon \mid \boldsymbol{\nu} \in \Pi\right) = 1 \tag{2}$$

Assumption (of whatever form) which guarantees existence and uniqueness of the I-projection is crucial for coming from statement (2) to the stronger claim (1).

2. Multiple I-projections

Recently, physicists (see for instance [10] and literature cited therein) started to study problems of the generic form mentioned in the Introduction, which led into non-convex Π, with possibly multiple I-projections.

In this note, the following questions are addressed: Do the conditional concentration of types happen on the I-projections? If yes, do types concentrate on each of them? And, if yes, what is the proportion?

In order to investigate these questions, implications of Sanov's Theorem for the case of several I-projections will be studied, at first. Then, a heuristics which suggests answers to the questions will be made. Finally, the answers will be provided by Asymptotic Equiprobability of I-projections Theorem, which will be supported by a proof and illustrative examples.

[1] There, $\log 0 = -\infty$, $\log \frac{b}{0} = +\infty$, $0 \cdot (\pm\infty) = 0$, conventions are assumed. The definition of I-projection was adapted from [2]. Throughout the paper log denotes the natural logarithm.

2.1 Implications of Sanov's Theorem.

Sanov's Theorem (ST) is a fundamental tool for proving both ECT and CWLLN. It reads

$$\lim_{n\to\infty} \frac{\log \pi(\boldsymbol{\nu} \in \Pi)}{n} = -I(\hat{\boldsymbol{p}}||\boldsymbol{q}) \tag{3}$$

provided that Π is closure of its interior, and $\boldsymbol{q} \notin \Pi$. Thus, according to the ST

$$\pi(\boldsymbol{\nu} \in \Pi) \doteq e^{-nI(\hat{\boldsymbol{p}}||\boldsymbol{q})} \tag{4}$$

to first order in the exponent.

An application of the ST to $\pi(\boldsymbol{\nu} \in B|\boldsymbol{\nu} \in \Pi)$, where $B \subseteq \Pi$, gives

$$\lim_{n\to\infty} \frac{\log \pi(\boldsymbol{\nu} \in B|\boldsymbol{\nu} \in \Pi)}{n} = -(I(\hat{\boldsymbol{p}}_B||\boldsymbol{q}) - I(\hat{\boldsymbol{p}}_\Pi||\boldsymbol{q})) \tag{5}$$

where the subscripts B, Π denote the set on which $\boldsymbol{q}$ is I-projected. This can be also stated as

$$\pi(\boldsymbol{\nu} \in B|\boldsymbol{\nu} \in \Pi) \doteq e^{-n(I(\hat{\boldsymbol{p}}_B||\boldsymbol{q}) - I(\hat{\boldsymbol{p}}_\Pi||\boldsymbol{q}))} \tag{6}$$

to first order in the exponent.

Let now Π be such that there are k I-projections $[\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2, \dots, \hat{\boldsymbol{p}}_k]$, of $\boldsymbol{q}$ on Π. What does the ST and (5), (6) imply in this case?

Let $B(\hat{\boldsymbol{p}}_j, \epsilon)$ denote an open Euclidean ball of radius ϵ centered around j-th I-projection, which is assumed to be the only I-projection in there.

Let $C \triangleq \cup_{j=1,2,\dots,k} B(\hat{\boldsymbol{p}}_j, \epsilon)$. For types outside of C the RHS of (5) is negative and thus, fur such types (6) implies that $\pi(\boldsymbol{\nu} \notin C|\boldsymbol{\nu} \in \Pi)$ goes to zero exponentially fast. In turn, this implies that

$$\lim_{n\to\infty} \pi(\boldsymbol{\nu} \in C|\boldsymbol{\nu} \in \Pi) = 1 \tag{7}$$

Informally, for sufficiently large n it is virtually impossible to find a type which does not belong to union of the I-projection balls. So, now it remains only to find the proportion in which the probability is split among the different I-projections. The next crude heuristics indicates that the split should be equal.

2.2 Heuristics.

According to (6),

$$\pi(\boldsymbol{\nu} \in C|\boldsymbol{\nu} \in \Pi) \doteq e^{-n(I(\hat{\boldsymbol{p}}_C||\boldsymbol{q}) - I(\hat{\boldsymbol{p}}_\Pi||\boldsymbol{q}))} \tag{8}$$

to first order in the exponent. At the same time (6) implies

$$\pi(\boldsymbol{\nu} \in B(\hat{\boldsymbol{p}}_j, \epsilon)|\boldsymbol{\nu} \in \Pi) \doteq e^{-n(I(\hat{\boldsymbol{p}}_C||\boldsymbol{q}) - I(\hat{\boldsymbol{p}}_\Pi||\boldsymbol{q}))} \qquad j = 1, 2, \dots, k \tag{9}$$

since

$$I(\hat{\boldsymbol{p}}_{B_j}||\boldsymbol{q}) = I(\hat{\boldsymbol{p}}_C||\boldsymbol{q}) \qquad j = 1, 2, \dots, k \tag{10}$$

However,

$$\pi(\boldsymbol{\nu} \in C|\boldsymbol{\nu} \in \Pi) = \sum_{j=1}^{k} \pi(\boldsymbol{\nu} \in B(\hat{\boldsymbol{p}}_j, \epsilon)|\boldsymbol{\nu} \in \Pi) \tag{11}$$

which together with (8) and (9) implies that for sufficiently large n

$$\pi(\boldsymbol{\nu} \in C|\boldsymbol{\nu} \in \Pi) = k\, \pi(\boldsymbol{\nu} \in B(\hat{\boldsymbol{p}}_j, \epsilon)|\boldsymbol{\nu} \in \Pi) \qquad j = 1, 2, \dots, k \tag{12}$$

After recalling (7), (12) implies that the conditional measure should be split among the I-projections equally.

2.3 Asymptotic Equiprobability of I-projections.

AEI Theorem. *Let there be k I-projections of $\boldsymbol{q}$ on Π. Let $B(\hat{\boldsymbol{p}}_j, \epsilon)$ be open Euclidean ball of radius ϵ centered at j-th I-projection, which is assumed to be the only I-projection in there. Then*

$$\lim_{n\to\infty} \pi(\boldsymbol{\nu} \in B(\hat{\boldsymbol{p}}_j, \epsilon) | \boldsymbol{\nu} \in \Pi) = \frac{1}{k} \qquad j = 1, 2, \dots, k \tag{13}$$

Sketch of proof of the Theorem will employ MaxProb justification of MaxEnt (cf. [5]) and the ST.

Proof. Let $\hat{\boldsymbol{\nu}} \triangleq \arg\sup_{\boldsymbol{\nu}^\star \in \Pi} \pi(\boldsymbol{\nu} = \boldsymbol{\nu}^\star | \boldsymbol{\nu} \in \Pi)$. Let there be k such types, $\mathcal{N} \triangleq [\hat{\boldsymbol{\nu}}_1, \hat{\boldsymbol{\nu}}_2, \dots, \hat{\boldsymbol{\nu}}_k]$. Then, by the Theorem 1 of [5], $\hat{\boldsymbol{\nu}}_j \to \hat{\boldsymbol{p}}_j$, for any $j = 1, 2, \dots, k$, ie. the most probable type converges to the corresponding I-projection of $\boldsymbol{q}$ on Π. Since there are k types in $\mathcal{N}$ with equal value of the conditional probability π (which is the highest one), and each of the types from $\mathcal{N}$ converges to the corresponding I-projection, the last two facts imply that all the I-projections should be equally probable. Recall (7) to conclude that each of the conditional probabilities π should be equal to $1/k$ for n growing beyond any limit. □

Note that the number of I-projections k can be at most $\sum_{i=1}^m \binom{m}{i}$. Since m is assumed finite, k is finite as well.

The Asymptotic Equiprobability of I-projections (AEI) states that the probability that if $\boldsymbol{q}$ generated a type from Π then the type is close to the particular I-projection among k possible I-projections, approaches $\frac{1}{k}$ as n gets large. In other words, the conditional concentration of types happens on each of the I-projections with equal measure.

The AEI will be illustrated by the next two Examples.

2.4 Illustrative examples.

Example 1. Let $\Pi = \{\boldsymbol{p} : \sum_{i=1}^m p_i^\alpha - a = 0, \sum_{i=1}^m p_i - 1 = 0\}$, where $\alpha, a \in \mathbb{R}$. Note that the first constraint, known as frequency constraint, is non-linear in $\boldsymbol{p}$ and Π is for $|\alpha| > 1$ non-convex.

Let $\alpha = 2$, $m = 3$ and $a = 0.42$ (the value was obtained for $p = [0.5\ 0.4\ 0.1]$). Then there are three I-projections of uniform distribution $\boldsymbol{q} = [1/3\ 1/3\ 1/3]$ on Π: $\hat{\boldsymbol{p}}_1 = [0.5737\ 0.2131\ 0.2131]$, $\hat{\boldsymbol{p}}_2 = [0.2131\ 0.5737\ \ 0.2131]$ and $\hat{\boldsymbol{p}}_3 = [0.2131\ 0.2131\ 0.5737]$ (see [6]). Note that they form a group of permutations. As it will become clear later, it suffices to investigate convergence to say $\hat{\boldsymbol{p}}_1$.

For $n = 30$ there are only two groups of types in Π: G1 comprises [0.5666 0.2666 0.1666] and five other permutations; G2 consists of [0.5 0.4 0.1] and the other five permutations. So, together there are 12 types.

Value of the square of the Euclidean distance δ between $\boldsymbol{\nu}$ and $\hat{\boldsymbol{p}}_1$ attains its minimum $\delta_{G1} = 0.0051$ within G1 group for two types: [0.5666 0.2666 0.1666], [0.5666 0.1666 0.2666]. Within G2 the smallest $\delta_{G2} = 0.0532$ is attained by [0.5 0.4 0.1] and [0.5 0.1 0.4].

Thus, if $\epsilon = \epsilon_1$ is chosen so that the ball $B(\hat{\boldsymbol{p}}_1, \epsilon_1)$ contains only the two types from G1 (which at the same time guarantees that $\hat{\boldsymbol{p}}_1$ is the only I-projection in the ball), then $\pi(\boldsymbol{\nu} \in B(\hat{\boldsymbol{p}}_1, \epsilon_1) | \boldsymbol{\nu} \in \Pi) = 2 * 0.1152 = 0.2304$. In words: probability that if $\boldsymbol{q}$ generated a type from Π than the type falls into the ball containing only types which are closest to the I-projection is 0.2304. If $\epsilon = \epsilon_2$ is chosen so that also the two types from G2 are included in the ball and also so that it is the only I-projection in the ball (any $\epsilon_2 \in (\sqrt{0.0532}, \sqrt{0.1253})$ satisfies both the requirements), then $\pi(\boldsymbol{\nu} \in B(\hat{\boldsymbol{p}}_1, \epsilon_2)|\boldsymbol{\nu} \in \Pi) = \frac{1}{3}$.

For $n = 330$ there are four groups of types in Π: G1, G2 and a couple of new one: G3 consists of [0.4727 0.4333 0.0939] and all its permutations; G4 comprises [0.5727 0.2333 0.1939] and its permutations. Hence, the total number of types from Π which are supported by random sequences of size $n = 330$ is 24.

δ_{G3} for the two types from G3 which are closest to $\hat{\boldsymbol{p}}_1$ is 0.0729. The smallest $\delta_{G4} = 0.00077$ is attained by [0.5727 0.2333 0.1939] and by [0.5727 0.1939 0.2333]. Thus, clearly, the two types from G4 have the smallest Euclidean distance to $\hat{\boldsymbol{p}}_1$ among all types from Π which are based on samples of size $n = 330$. Again, setting ϵ such that the ball $B(\hat{\boldsymbol{p}}_1, \epsilon)$ contains only the two types which are closest to $\hat{\boldsymbol{p}}_1$ leads to the 0.261 value of the conditional probability. Note the important fact, that the probability has risen, as compared to the corresponding value 0.2304 for $n = 30$.

Moreover, if ϵ is set such that besides the two types from G4 also the second closest types (i.e. the two types from G1) are included in the ball then the conditional probability is indistinguishable from $\frac{1}{3}$. Hence, there is virtually no conditional chance of observing any of the remaining 4 types. The same holds for the types which concentrate around $\hat{\boldsymbol{p}}_2$ or $\hat{\boldsymbol{p}}_3$. Thus, in total, a half of the 24 types is almost impossible to observe.

So, this Example illustrates, that indeed, as n gets large, $\pi(\boldsymbol{\nu} \notin C|\boldsymbol{\nu} \in \Pi)$ goes to zero, and that the conditional probability of finding a type which is close (in the Euclidean distance) to one of the three I-projections goes to $\frac{1}{3}$.

Example 2. Let $\Pi = \Pi_1 \cup \Pi_2$, where $\Pi_j = \{\boldsymbol{p} : \sum_{i=1}^m p_i x_i = a_j; \sum_{i=1}^m p_i = 1\}, j = 1, 2$. Thus Π is union of two sets, each of whose is given by the classical moment consistency constraint. If $\boldsymbol{q}$ is chosen to be the uniform distribution, then there is no difficulty to find values a_1, a_2 such that there will be two different I-projections of the uniform $\boldsymbol{q}$ on Π with the same value of I-divergence (as well as of the Shannon's entropy). This is indeed true for any $a_1 = \mu + \Delta$, $a_2 = \mu - \Delta$, where $\mu \triangleq EX$ and $\Delta \in (0, (X_{\max} - X_{\min})/2)$, since then $\hat{\boldsymbol{p}}_1$ is just a permutation of $\hat{\boldsymbol{p}}_2$, and as such attains the same value of Shannon's entropy. To see that types which are based on random samples of size n from Π indeed concentrate on the I-projections with equal measure note, that for any n to each type in Π_1 corresponds a unique permutation of the type in Π_2. Thus, types in ϵ-ball with center at $\hat{\boldsymbol{p}}_1$ have the same conditional probabilities π as types in the ϵ-ball centered at $\hat{\boldsymbol{p}}_2$. This, together with convexity of both Π_j, for which the conditional concentration of types on the respective I-projection is well-established, directly implies that

$$\lim_{n\to\infty} \pi(\boldsymbol{\nu} \in B(\hat{\boldsymbol{p}}_j, \epsilon)|\boldsymbol{\nu} \in \Pi) = \frac{1}{2} \qquad j = 1, 2$$

The same reasoning could be made for arbitrary $\boldsymbol{q}$. Convexity of I-divergence guarantees that there exists a pair a_1, a_2 with the desired property. For general $\boldsymbol{q}$ the sought values of a_1, a_2 are not displaced around μ equally.

3. Concluding notes

Probabilistic justification of MaxEnt by CWLLN as well as several axiomatic and non-axiomatic foundations of MaxEnt require assumption of uniqueness of I-projection. As [10] indicates, problems which lead to non-convex Π with possibly non-unique I-projection appear in Physics. This facts prompted the presented study of convergence of types to multiple I-projections. In particular, we were interested in assessing the proportion by which types conditionally concentrate on each of the I-projections. As the AEI Theorem states, concentration of types happens on each of the I-projections with equal conditional measure. To support the sketched proof of the AEI Theorem, two examples were developed.

Informally, the AEI can be described using the Statistical Physics terminology, by stating that each of equilibrium points (I-projections) is asymptotically conditionally equally possible. Yet another informal formulation: If a random generator (probability distribution $\boldsymbol{q}$) is confined to produce types in Π then, as n gets large, the generator hides itself equally likely behind any of its I-projections on Π.

Asymptotic Equiprobability of I-projections enhances the Conditioned Weak Law of Large Numbers.

Aknowledgement

Hospitality of Banach Centre of the Institute of Mathematics of Polish Academy of Sciences, where a part of this study was performed as a part of the European Community Center of Excellence programme (package 'Information Theory and its Applications to Physics, Finance and Biology') is gratefully acknowledged. The work was also supported by the grant VEGA 1/7295/20 from the Scientific Grant Agency of the Slovak Republic. It is a pleasure to thank Brian R. La Cour for valuable discussions and an anonymous referee for careful reading and constructive criticism of the manuscript.

References

1. Cover, T. and Thomas, J., *Elements of Information Theory*, John Wiley and Sons, NY, 1991.
2. Csiszár. I., *Sanov Property, Generalized I-projection and a Conditional Limit Theorem*, Ann. Probab. **12** (1984), 768–793.
3. ———, *The Method of Types*, IEEE Trans. IT **44** (1998), 2505–2523.
4. Dembo, A and Zeitouni, O., *Large Deviations Techniques and Applications*, 2-nd ed., Springer, Application of Mathematics, vol. 38,, NY, 1998.
5. 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.
6. ———, *Maximum Entropy method: Shannon-Kullback vs. Rényi-Tsallis*, Advances of Mathematical Physics, NSP, NY, 2003 (to appear).
7. La Cour, B. R. and Schieve, W. C., *Macroscopic determinism in interacting systems using Large Deviations Theory*, Jour. Stat. Phys. **107 3/4** (2002), 729-755.
8. Leonard, Ch. and Najim, J., *An extension of Sanov's Theorem: Application to the Gibbs Conditioning Principle*, Bernoulli **8 (6)** (2002), 721–743.
9. Lewis, J. T., Pfister, C.-E. and Sullivan, W. G., *Entropy, concentration of probability and conditional theorems*, Markov Proc. Rel. Field. **1** (1995), 319–386.
10. Romera, E., Angulo, J. C. and Dehesa, J. S., *Reconstruction of a density from its entropic moments*, Bayesian Inference and Maximum Entropy Methods in Science and Engineering (R. L. Fry, ed.), AIP, Melville, NY, 2002, pp. 449–457.

Institute of Mathematics and Computer Science (of Matej Bel University and of Mathematical Institute of Slovak Academy of Sciences); Severná 5; SK-974 00 Banska Bystrica; Slovak Republic
Institute of Measurement Science of Slovak Academy of Sciences; Dúbravská cesta 9; SK-841 04 Bratislava; Slovak Republic
marian.grendar@savba.sk