

# GIBBS CONDITIONING EXTENDED, BOLTZMANN CONDITIONING INTRODUCED

M. GRENDAR

*To Mar, in memoriam*

ABSTRACT. Conditional Equi-concentration of Types on I-projections (ICET) and Extended Gibbs Conditioning Principle (EGCP) provide an extension of Conditioned Weak Law of Large Numbers and of Gibbs Conditioning Principle to the case of non-unique Relative Entropy Maximizing (REM) distribution (aka I-projection). ICET and EGCP give a probabilistic justification to REM under rather general conditions.  $\mu$ -projection variants of the results are introduced. They provide a probabilistic justification to Maximum Probability (MaxProb) method. 'REM/MaxEnt or MaxProb?' question is discussed briefly. Jeffreys Conditioning Principle is mentioned.

## 1. INTRODUCTION

Relative entropy maximization (REM/MaxEnt) is usually performed under moment consistency constraints. The constraints define a feasible set of probability distributions which is convex, closed and hence the relative entropy maximizing distribution (aka I-projection) is unique. For such sets Conditioned Weak Law of Large Numbers (CWLLN) is established and provides a probabilistic justification of REM/MaxEnt. Gibbs conditioning principle (GCP) - a stronger version of CWLLN - which is as well established for such sets, gives a further insight into the 'phenomenon' of conditional concentration of empirical measure on I-projections.

This work strives to develop extensions of CWLLN and GCP to the case of non-unique I-projection<sup>1</sup>. Proposed Conditional Equi-concentration of Types on I-projections (ICET) which extends CWLLN says, informally, that types (i.e., empirical distributions) conditionally concentrate on each of proper I-projections in equal measure. Extended Gibbs conditioning principle (EGCP) states, that in the case of multiple proper I-projections, probability of an outcome is given by equal-weight mixture of proper I-projection probabilities of the outcome.

A generalization (cf. [12]) of a result on convergence of maximum/supremum probability types ( $\mu$ -projections) to I-projections (cf. [8], Thm 1) directly permits to state either the well-established CWLLN, GCP or their extensions equivalently in terms of  $\mu$ -projections. The  $\mu$ -projection variants of the probabilistic laws allows for a deeper reading than their I-projection counterparts - since the  $\mu$ -laws express the asymptotic conditional behavior of types in terms of the asymptotically most probable types. They provide probabilistic justification to Maximum Probability (MaxProb) method.

---

*Date:* September 17, 2004 revision of August 12, 2004 revision of July 10, 2004 version.

*Key words and phrases.* multiple I-projections, Conditioned Weak Law of Large Numbers, Gibbs Conditioning Principle,  $\mu$ -projection, J-projection,  $\gamma$ -projection, Maximum Probability method, MaxProb/MaxEnt convergence.

This work was supported by VEGA 1/0264/03 grant. Valuable discussions with Brian R. La Cour, Alberto Solana-Ortega, Ondrej Šuch and Viktor Witkovský are gratefully acknowledged.

<sup>1</sup>For a motivation see [11], [16]. For an exploratory work in this direction see [9].

Though  $\mu$ -projections and I-projections are asymptotically identical, in the case of finite samples, they are in general different.

## 2. TERMINOLOGY AND NOTATION

Let  $\{X_i\}_{i=1}^n$  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  $X \triangleq \{x_1, x_2, \dots, x_m\}$  called support or alphabet. Hereafter  $X$  will be assumed finite. An element of  $X$  will be called outcome or letter. Let  $q_i$  denote the probability (measure) of  $i$ -th element of  $X$ ;  $q$  will be called source or generator. Let  $P(X)$  be a set of all probability mass functions (pmf's) on  $X$ .

A type (also called  $n$ -type, empirical measure, frequency distribution or occurrence vector) induced by a sequence  $\{X_i\}_{i=1}^n$  is the pmf  $\nu^n \in P(X)$  whose  $i$ -th element  $\nu_i^n$  is defined as:  $\nu_i^n \triangleq n_i/n$  where  $n_i \triangleq \sum_{l=1}^n I(X_l = x_i)$ , and  $I(\cdot)$  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  $\Pi \subseteq P(X)$ . Let  $P_n$  denote a subset of  $P(X)$  which consists of all  $n$ -types. Let  $\Pi_n = \Pi \cap P_n$ .

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

$\pi(\nu^n \in A | \nu^n \in B; q \mapsto \nu^n)$  will denote the conditional probability that if a type drawn from  $q \in P(X)$  belongs to  $B \subseteq \Pi$  then it belongs to  $A \subseteq \Pi$ .

## 3. CWLLN, GIBBS CONDITIONING

Conditioned Weak Law of Large Numbers (cf. [1], [15], [21], [20], [3], [6], [14]) in its standard form (cf. [2]) reads:

**CWLLN.** *Let  $X$  be a finite set. Let  $\Pi$  be closed, convex set which does not contain  $q$ . Let  $n \rightarrow \infty$ . Then for  $\epsilon > 0$*

$$\lim_{n \rightarrow \infty} \pi(|\nu_i^n - \hat{p}_i| < \epsilon | \nu^n \in \Pi; q \mapsto \nu^n) = 1 \quad \text{for } i = 1, 2, \dots, m.$$

CWLLN says that if types are confined to a closed, convex set  $\Pi$  then they asymptotically conditionally concentrate on the I-projection  $\hat{p}$  of the source of types  $q$  on the set  $\Pi$  (i.e., informally, on the probability distribution from  $\Pi$  which has the highest value of the relative entropy with respect to the source  $q$ ).

Gibbs conditioning principle (GCP) says, very informally, that if the source  $q$  is confined to produce sequences which lead to types in a convex, closed set  $\Pi$  then elements of any such sequence (of fixed length  $t$ ) behave asymptotically conditionally as if they were drawn identically and independently from the I-projection of  $q$  on  $\Pi$  - provided that the last is unique (among other things).

**GCP.** *Let  $X$  be a finite set. Let  $\Pi$  be closed, convex set which does not contain  $q$ . Let  $n \rightarrow \infty$ . Then for a fixed  $t$*

$$\lim_{n \rightarrow \infty} \pi(X_1 = x_1, \dots, X_t = x_t | \nu^n \in \Pi; q \mapsto \nu^n) = \prod_{l=1}^t \hat{p}_{x_l}.$$

GCP was developed at [3] under the name of conditional quasi-independence of outcomes. Later on, it was brought into more abstract form in large deviations literature, where it also obtained the GCP name (cf. [5], [18]). A simple proof of GCP can be found at [4]. GCP is proven also for continuous alphabet (cf. [13], [4], [5]).

## 4. THE CASE OF SEVERAL I-PROJECTIONS

What happens when  $\Pi$  admits multiple I-projections? Do the conditional concentration of types happens on them? If yes, do types concentrate on each of them? If yes, what is the proportion? How does GCP extend to the case of multiple I-projections?

**4.1. Conditional Equi-concentration of Types on I-projections.** Let  $d(a, b) \triangleq \sum_{i=1}^m |a_i - b_i|$  be the total variation metric (or any other equivalent metric) on the set of probability distributions  $P(X)$ . Let  $B(a, \epsilon)$  denote an  $\epsilon$ -ball - defined by the metric  $d$  - which is centered at  $a \in P(X)$ .

An I-projection  $\hat{p}$  of  $q$  on  $\Pi$  will be called proper if  $\hat{p}$  is not an isolated point of  $\Pi$ .

**ICET.** *Let  $X$  be a finite set. Let  $\Pi$  be such that it admits  $k$  proper I-projections  $\hat{p}^1, \hat{p}^2, \dots, \hat{p}^k$  of  $q$ . Let  $\epsilon > 0$  be such that for  $j = 1, 2, \dots, k$   $\hat{p}^j$  is the only proper I-projection of  $q$  on  $\Pi$  in the ball  $B(\hat{p}^j, \epsilon)$ . Let  $n \rightarrow \infty$ . Then*

$$(1) \quad \pi(\nu^n \in B(\epsilon, \hat{p}^j) | \nu^n \in \Pi; q \mapsto \nu^n) = 1/k \quad \text{for } j = 1, 2, \dots, k.$$

ICET<sup>2</sup> states that if a set  $\Pi$  admits several I-projections then the conditional measure is spread among the proper I-projections equally. In less formal words: if a random generator (i.e.,  $q$ ) is confined to produce types in  $\Pi$  then, as  $n$  gets large, the generator 'hides itself' equally likely behind any of its proper I-projections on  $\Pi$ . Yet in other (statistical physics) words: each of the equilibrium points (i.e., proper I-projections) is asymptotically conditionally equally probable. The conditional equi-concentration of types 'phenomenon' resembles Thermodynamic coexistence of phases (e.g., triple point of water, vapor and ice).

*Notes.* 1) On an I-projection  $\hat{p}$  which is not rational and at the same time it is an isolated point no conditional concentration of types happens. However, if the set  $\Pi$  is such that an I-projection  $\hat{p}$  of  $q$  on it is rational and at the same time it is an isolated point, then types can concentrate on it. 2) Since  $X$  is finite,  $k$  is finite.

Weak Law of Large Numbers is special - unconditional - case of CWLLN. CWLLN itself is just a special - unique proper I-projection - case of ICET.

Two illustrative examples of the Conditional Equi-concentration of Types on I-projections (ICET) can be found at the exploratory study [9]. There also Asymptotic Equiprobability of I-projections - a precursor to ICET - was formulated.

## 4.2. Extended Gibbs conditioning principle.

**EGCP.** *Let  $X$  be a finite set. Let  $\Pi$  be such that it admits  $k$  proper I-projections  $\hat{p}^1, \hat{p}^2, \dots, \hat{p}^k$  of  $q$  on  $\Pi$ . Then for a fixed  $t$ :*

$$(2) \quad \lim_{n \rightarrow \infty} \pi(X_1 = x_1, \dots, X_t = x_t | \nu^n \in \Pi; q \mapsto \nu^n) = 1/k \sum_{j=1}^k \prod_{l=1}^t \hat{p}_{x_l}^j.$$

EGCP, for  $t = 1$ , says that the conditional probability of a letter is asymptotically given by the equal-weight mixture of proper I-projection probabilities of the letter. For a general sequence, EGCP states that the conditional probability of a sequence is asymptotically equal to the mixture of joint probability distributions. Each of the  $k$  joint distributions is such as if the sequence was iid distributed according to a proper I-projection.

<sup>2</sup>See Appendix for a proof of ICET and EGCP.

### 5. $\mu$ -PROJECTIONS, MAXIMUM PROBABILITY METHOD

$\mu$ -projection  $\hat{\nu}^n$  of  $q$  on  $\Pi_n \neq \emptyset$  is defined as:  $\hat{\nu}^n \triangleq \arg \sup_{\nu^n \in \Pi_n} \pi(\nu^n; q)$ , where  $\pi(\nu^n; q) \triangleq \Gamma(\nu^n) \prod (q_i)^{n\nu_i^n}$ , (cf. [12]). Alternatively, the  $\mu$ -projection can be defined as  $\hat{\nu}^n \triangleq \arg \sup_{\nu^n \in \Pi_n} \pi(\nu^n | \nu^n \in \Pi_n; q)$ , where  $\pi(\nu^n | \nu^n \in \Pi_n; q)$  denotes the conditional probability that if an  $n$ -type belongs to  $\Pi_n$  then it is just the type  $\nu^n$ . Yet another equivalent definition - a bayesian one - of  $\mu$ -projection can be adapted from [10].

Concept of  $\mu$ -projection is associated with the Maximum Probability method (cf. [8]).

**5.1. Asymptotic identity of  $\mu$ -projections and I-projections.** At ([8], Thm 1 and its Corollary, aka MaxProb/MaxEnt Thm) it was shown that maximum probability type converges to I-projection; provided that  $\Pi$  is defined by a differentiable constraints. A more general result which states asymptotic identity of  $\mu$ -projections and I-projections was presented at [12]. It will be recalled here.

**MaxProb/MaxEnt.** [12] *Let  $X$  be a finite set. Let  $M_n$  be set of all  $\mu$ -projections of  $q$  on  $\Pi_n$ . Let  $I$  be set of all I-projections of  $q$  on  $\Pi$ . For  $n \rightarrow \infty$ ,  $M_n = I$ .*

Since  $\pi(\nu^n; q)$  is defined for  $\nu^n \in Q^m$ ,  $\mu$ -projection can be defined only for  $\Pi_n$  when  $n$  is finite. The Theorem permits to define a  $\mu$ -projection  $\hat{\nu}$  also on  $\Pi$ :  $\hat{\nu} \triangleq \arg \sup_{r \in \Pi} - \sum_{i=1}^m r_i \log \frac{r_i}{q_i}$ . Thus  $\mu$ -projections and I-projections on  $\Pi$  are undistinguishable.

It is worth highlighting that for a finite  $n$ ,  $\mu$ -projections and I-projections of  $q$  on  $\Pi_n$  are in general different. This explains why  $\mu$ -form of the probabilistic laws deserves to be stated separately of the I-form; though formally they are undistinguishable. Thus, MaxProb/MaxEnt Thm (in its new and to a smaller extent also in its old version) permits directly to state  $\mu$ -projection variants of CWLLN, GCP, ICET and EGCP:  $\mu$ CWLLN,  $\mu$ GCP,  $\mu$ CET and Boltzmann Conditioning Principle (BCP).

### 5.2. $\mu$ -form of CWLLN and GCP.

**$\mu$ CWLLN.** *Let  $X$  be a finite set. Let  $\Pi$  be closed, convex set which does not contain  $q$ . Let  $n \rightarrow \infty$ . Then for  $\epsilon > 0$*

$$\lim_{n \rightarrow \infty} \pi(|\nu_i^n - \hat{\nu}_i| < \epsilon | \nu^n \in \Pi; q \mapsto \nu^n) = 1 \quad \text{for } i = 1, 2, \dots, m.$$

Core of  $\mu$ CWLLN can be loosely expressed as: *types, when confined to a convex, closed set  $\Pi$ , conditionally concentrate on the asymptotically most probable type  $\hat{\nu}$ .* It is worth a comparison with the reading of the I-projection variant of CWLLN (see Sect. 3).

Similarly, to the GCP its  $\mu$ -variant exists:

**$\mu$ GCP.** *Let  $X$  be a finite set. Let  $\Pi$  be closed, convex set which does not contain  $q$ . Let  $n \rightarrow \infty$ . Then for a fixed  $t$*

$$\lim_{n \rightarrow \infty} \pi(X_1 = x_1, \dots, X_t = x_t | \nu^n \in \Pi; q \mapsto \nu^n) = \prod_{l=1}^t \hat{\nu}_{x_l}.$$

**5.3. Conditional Equi-concentration of Types on  $\mu$ -projections.** A  $\mu$ -projection  $\hat{\nu}$  of  $q$  on  $\Pi$  will be called proper if  $\hat{\nu}$  is not an isolated point of  $\Pi$ .

**$\mu$ CET.** *Let  $X$  be a finite set. Let there be  $k$  proper  $\mu$ -projections  $\hat{\nu}^1, \hat{\nu}^2, \dots, \hat{\nu}^k$  of  $q$  on  $\Pi$ . Let  $\epsilon > 0$  be such that for  $j = 1, 2, \dots, k$   $\hat{\nu}^j$  is the only proper  $\mu$ -projection of  $q$  on  $\Pi$  in the ball  $B(\hat{\nu}^j, \epsilon)$ . Let  $n \rightarrow \infty$ . Then*

$$(3) \quad \pi(\nu^n \in B(\epsilon, \hat{\nu}^j) | \nu^n \in \Pi; q \mapsto \nu^n) = 1/k \quad \text{for } j = 1, 2, \dots, k.$$

#### 5.4. Boltzmann conditioning principle.

**BCP.** *Let  $X$  be a finite set. Let there be  $k$  proper  $\mu$ -projections  $\hat{\nu}^1, \hat{\nu}^2, \dots, \hat{\nu}^k$  of  $\mathbf{q}$  on  $\Pi$ . Then for a fixed  $t$ :*

$$(4) \quad \lim_{n \rightarrow \infty} \pi(X_1 = x_1, \dots, X_t = x_t | \nu^n \in \Pi; \mathbf{q} \mapsto \nu^n) = 1/k \sum_{j=1}^k \prod_{l=1}^t \hat{\nu}_{x_l}^j.$$

**5.5. MaxEnt or MaxProb?**  $\mu$ -projections and I-projections are asymptotically indistinguishable (recall MaxProb/MaxEnt Thm, Sect. 5.1). In plain words: for  $n \rightarrow \infty$  REM/MaxEnt selects the same distribution(s) as MaxProb (in its more general form which instead of the maximum probable types selects supremum-probable  $\mu$ -projections). This result (in the older form, [8]) was at [8] *interpreted* as saying that REM/MaxEnt can be viewed as an asymptotic instance of the simple and self-evident Maximum Probability method.

Alternatively, [19] suggests to view REM/MaxEnt as a separate method and hence to read the MaxProb/MaxEnt Thm as claiming that REM/MaxEnt asymptotically coincides with MaxProb. If one adopts this interesting and legitimate view then it is necessary to face the fact that if  $n$  is finite, the two methods in general differ.

### 6. JEFFREYS CONDITIONING MENTIONED

Instead of Summary (which is already presented at Sect. 1), Conditional Equi-concentration of types on J-projections (JCET) and Jeffreys conditioning principle<sup>3</sup> (JCP) will be mentioned, in passing.

$\gamma$ -projection  $\tilde{\nu}^n$  of  $\mathbf{q} \in Q^m$  on  $\Pi_n$  is:  $\tilde{\nu}^n \triangleq \arg \sup_{\nu^n \in \Pi_n} \pi(\nu^n; \mathbf{q}) \pi(\mathbf{q}; \nu^n)$ . J-projection (or Jeffreys projection)  $\tilde{p}$  of  $\mathbf{q} \in Q^m$  on  $\Pi$  is  $\tilde{p} \triangleq \arg \inf_{p \in \Pi} \sum_{i=1}^m p_i \log \frac{p_i}{q_i} + q_i \log \frac{q_i}{p_i}$ .

Let  $\mathbf{q} \in Q^m$ .  $\pi(\nu^n \in A | \nu^n \in B; (\mathbf{q} \mapsto \nu^n) \wedge (\nu^n \mapsto \mathbf{q}))$  will denote the conditional probability that if a type - which was drawn from  $\mathbf{q} \in P(X)$  and was at the same time used as a source of the type  $\mathbf{q}$  - belongs to  $B \subseteq \Pi$  then it belongs to  $A \subseteq \Pi$ . A J-projection  $\tilde{p}$  of  $\mathbf{q}$  on  $\Pi$  will be called proper if it is not isolated point of  $\Pi$ .

**JCET.** *Let  $X$  be a finite set. Let  $\mathbf{q} \in Q^m$ . Let there be  $k$  proper J-projections  $\tilde{p}^1, \tilde{p}^2, \dots, \tilde{p}^k$  of  $\mathbf{q}$  on  $\Pi$ . Let  $\epsilon > 0$  be such that for  $j = 1, 2, \dots, k$   $\tilde{p}^j$  is the only proper J-projection of  $\mathbf{q}$  on  $\Pi$  in the ball  $B(\tilde{p}^j, \epsilon)$ . Let  $n_0$  be denominator of the smallest common divisor of  $q_1, q_2, \dots, q_m$ . Let  $n = un_0$ ,  $u \in N$ . Let  $u \rightarrow \infty$ . Then*

$$(5) \quad \pi(\nu^n \in B(\epsilon, \tilde{p}^j) | \nu^n \in \Pi; (\mathbf{q} \mapsto \nu^n) \wedge (\nu^n \mapsto \mathbf{q})) = 1/k \quad \text{for } j = 1, 2, \dots, k.$$

In words, types which were 'emitted' from  $\mathbf{q}$  and were at the same time used as a source of  $\mathbf{q}$ -types, conditionally equi-concentrate on J-projections of  $\mathbf{q}$  on  $\Pi$ .

J-projections and  $\gamma$ -projections asymptotically coincide (cf. [12], and [7] for an example). Hence, a  $\gamma$ -projection alternative of JCET is valid as well. It says that: types which were 'emitted' from  $\mathbf{q}$  and were at the same time used as a source of  $\mathbf{q}$ -types, conditionally equi-concentrate on those of them which have the highest/supremal value of  $\pi(\nu^n; \mathbf{q}) \pi(\mathbf{nq}; \nu^n)$ . - Similarly, JCP can be considered in its J- or  $\gamma$ -form.

$\mu$ -projection is based on the probability  $\pi(\nu^n; \mathbf{q})$ ; thus it can be viewed as a UNI-projection.  $\gamma$ -projection is based on  $\pi(\nu^n; \mathbf{q}) \pi(\mathbf{nq}; \nu^n)$ , thus it can be viewed as AND-projection. It is possible to consider also an OR-projection defined as  $\hat{\nu}^n \triangleq \arg \sup_{\nu^n \in \Pi_n} \pi(\nu^n; \mathbf{q}) + \pi(\mathbf{nq}; \nu^n)$ . However, there seems to be no obvious

<sup>3</sup>It should not be confused with Jeffrey principle of updating subjective probability.

analytic way how to define its asymptotic form. Despite that, it is possible to expect that OR-type of CWLLN/CET holds.

The  $\mu$ -,  $\gamma$ -, OR-projection CET can be summarized by a (bold) statement: types conditionally equi-concentrate on those which are asymptotically the most probable.

**Acknowledgments** Supported by VEGA 1/0264/03. It is a pleasure to thank Brian R. La Cour, Alberto Solana-Ortega, Ondrej Šuch and Viktor Witkovský for valuable discussions. Lapses are mine. The author is indebted to the Editors for patience.

## 7. APPENDIX

*A sketch of proof of ICET.*

$$(6) \quad \pi(\nu^n \in B(\epsilon, \hat{p}^j) | \nu^n \in \Pi; \mathbf{q} \mapsto \nu^n) \leq \frac{\sum_{\nu^n \in B} \pi(\nu^n; \mathbf{q})}{\sum_{\nu^n \in \Pi} \pi(\nu^n; \mathbf{q})}$$

$B_n(\epsilon, \hat{p}^j) \triangleq B(\epsilon, \hat{p}^j) \cap \Pi_n$ . Let there be  $k_{B,n}$  I-projections  $\hat{p}_{B,n}^1, \hat{p}_{B,n}^2, \dots, \hat{p}_{B,n}^{k_{B,n}}$  of  $\mathbf{q}$  on  $\bigcup_{j=1}^k B_n(\epsilon, \hat{p}^j)$ . Let  $k_{B,n}^j$  denote the number of I-projections of  $\mathbf{q}$  on  $B_n(\epsilon, \hat{p}^j)$ .  $\hat{p}_{B,n}^j$  will stand for any of such I-projections. By  $B \setminus k_{B,n}^j$  denote the set  $B_n(\epsilon, \hat{p}^j) \setminus \bigcup_{i=1}^{k_{B,n}^j} \{\hat{p}_{B,n}^i\}$ .

Similarly, let there be  $k_{\Pi,n}$  I-projections  $\hat{p}_{\Pi,n}^1, \hat{p}_{\Pi,n}^2, \dots, \hat{p}_{\Pi,n}^{k_{\Pi,n}}$  of  $\mathbf{q}$  on  $\Pi_n$ . Denote the set  $\Pi_n \setminus \bigcup_{i=1}^{k_{\Pi,n}} \{\hat{p}_{\Pi,n}^i\}$  as  $\Pi \setminus k_{\Pi,n}$ . The MaxProb/MaxEnt Thm implies that for  $n \rightarrow \infty$  the RHS of (6) can be written as:

$$(7) \quad \frac{\pi(\hat{p}_{B,n}^j; \mathbf{q}) \left( k_{B,n}^j + \frac{\sum_{\nu^n \in B \setminus k_{B,n}^j} \pi(\nu^n; \mathbf{q})}{\pi(\hat{p}_{B,n}^j; \mathbf{q})} \right)}{\pi(\hat{p}_{\Pi,n}; \mathbf{q}) \left( k_{\Pi,n} + \frac{\sum_{\nu^n \in \Pi \setminus k_{\Pi,n}} \pi(\nu^n; \mathbf{q})}{\pi(\hat{p}_{\Pi,n}; \mathbf{q})} \right)}$$

Recall a standard inequality:

**Lemma.** *Let  $\nu^n, \hat{\nu}^n$  be two types from  $\Pi_n$ . Then*

$$\frac{\pi(\nu^n; \mathbf{q})}{\pi(\hat{\nu}^n; \mathbf{q})} < \left( \frac{n}{m} \right)^m \prod_{i=1}^m \frac{\left( \frac{q_i}{\nu_i^n} \right)^{n \nu_i^n}}{\left( \frac{q_i}{\hat{\nu}_i^n} \right)^{n \hat{\nu}_i^n}}$$

The Lemma implies that the ratio in the nominator of (7) converges to zero as  $n \rightarrow \infty$ . The same implication holds for the ratio in the denominator.  $\hat{p}_{B,n}^j$  converges in the metric to  $\hat{p}^j$ , hence  $k_{B,n}^j$  converges to 1 as  $n \rightarrow \infty$ . Similarly,  $k_{\Pi,n}$  converges to  $k$  and  $\frac{\pi(\hat{p}_{B,n}^j; \mathbf{q})}{\pi(\hat{p}_{\Pi,n}; \mathbf{q})}$  converges to 1 as  $n$  goes to infinity. This taken together implies that the RHS of (6) converges to  $1/k$  as  $n \rightarrow \infty$ . The inequality (6) thus turns into equality.  $\square$

*A sketch of proof of EGCP.*

$$(8) \quad \pi(X_1 = x_1, \dots, X_t = x_t | \nu^n \in \Pi; \mathbf{q} \mapsto \nu^n) = \frac{\sum_{\nu^n \in \Pi} \pi(X_1 = x_1, \dots, X_t = x_t, \nu^n)}{\sum_{\nu^n \in \Pi} \pi(\nu^n; \mathbf{q})}$$

Partition  $\Pi_n$  into  $\Pi \setminus k_{\Pi,n}$  and the rest, which will be denoted by  $\bigcup \hat{p}_{\Pi,n}$ . The MaxProb/MaxEnt Thm implies that for  $n \rightarrow \infty$  the RHS of (8) can be written as:

$$(9) \quad \frac{\sum_{\nu^n \in \bigcup \hat{p}_{\Pi,n}} \pi(X_1 = x_1, \dots, X_t = x_t, \nu^n) + \sum_{\nu^n \in \Pi \setminus k_{\Pi,n}} \pi(X_1 = x_1, \dots, X_t = x_t, \nu^n)}{\pi(\hat{p}_{\Pi,n}; \mathbf{q}) \left( k_{\Pi,n} + \frac{\sum_{\nu^n \in \Pi \setminus k_{\Pi,n}} \pi(\nu^n; \mathbf{q})}{\pi(\hat{p}_{\Pi,n}; \mathbf{q})} \right)}$$

By the Lemma, the ratio in the denominator of (9) converges to zero as  $n$  goes to infinity. The second term in the nominator as well goes to zero as  $n \rightarrow \infty$  (to see this, express the joint probability  $\pi(X_1 = x_1, \dots, X_t = x_t, \nu^n)$  as  $\pi(X_1 = x_1, \dots, X_t = x_t | \nu^n) \pi(\nu^n; q)$  and employ the Lemma). Thus, for  $n \rightarrow \infty$  the RHS of (8) becomes  $1/k \sum_{j=1}^k \pi(X_1 = x_1, \dots, X_t = x_t | \beta^j)$ . Finally, invoke Csiszár's 'urn argument' (cf. [4]) to conclude that the asymptotic form of the RHS of (8) is  $1/k \sum_{j=1}^k \prod_{l=1}^t \hat{p}_{X_l}^j$ .  $\square$

## REFERENCES

- [1] Bártfai, P. "On a conditional limit theorem," *Progress in Statistics* **1**, 85-91, NH, Amsterdam, 1974.
- [2] Cover, T. and Thomas, J., *Elements of Information Theory*, Wiley, NY, 1991.
- [3] Csiszár, I., *Ann. Probab.*, **12**, 768-793, (1984).
- [4] Csiszár, I., *IEEE IT*, **44**, No. 6, 2505-2523, (1998).
- [5] Dembo, A. and Zeitouni O., *Large Deviations Techniques and Applications*, Springer, (NY), 1998.
- [6] Ellis, R. S., *Physica D*, 106-136, (1999).
- [7] Grendár, M., Jr. and Grendár, M., preprint, Elsevier Math Preprint Server MPS:Statistics/0107008.
- [8] Grendár, M., Jr. and Grendár, M., "What is the question that MaxEnt answers? A probabilistic interpretation", in AIP CP568, A. Mohammad-Djafari (ed.), pp. 83-94, AIP, Melville, 2001. Also at arXiv:math-ph/0009020.
- [9] Grendár, M., Jr. and Grendár, M., *Acta U. Belii Ser. Math.*, **10**, 3-8, (2003).
- [10] Grendár, M., Jr. and Grendár, M., "Maximum Probability and Maximum Entropy methods: Bayesian interpretation", in AIP CP707, G. Erickson and Y. Zhai (eds.), pp. 490-494, AIP, Melville, 2004. Also at arXiv:physics/0308005.
- [11] Grendár, M., Jr. and Grendár, M., "Maximum Entropy method with non-linear moment constraints: challenges", in AIP CP707, G. Erickson and Y. Zhai (eds.), pp. 97-109, AIP, Melville, 2004. Also at arXiv:physics/0308006.
- [12] Grendár, M., "Asymptotic identity of  $\mu$ -projections and I-projections, to appear at *Acta U. Belii*.
- [13] Groeneboom, P., Oosterhoff, J. and Ruymgaart, F. H., *Ann. Probab.*, **7**, 553-586, (1979).
- [14] Harremoes, P., "Information topologies with applications, to appear at *Bolyai Studies*.
- [15] Jaynes, E. T., *Papers on Probability, Statistics and Statistical Physics*, R. D. Rosenkratz (ed.), D. Reidel, Dordrecht, 1979.
- [16] La Cour, B. R. and Schieve, W. C., *Phys. Rev. E*, **62**, 7494-7496, (2002). Also at arXiv:condmat/0009216.
- [17] Lanford, O. E., "Entropy and equilibrium states in classical statistical mechanics," *Statistical Mechanics and Mathematical Problems*, Lecture Notes in Physics **20**, pp. 1-113, Springer, Berlin, 1973.
- [18] Leonard, Ch. and Najim, J., *Bernoulli*, **8**, 6, 721-743, (2002).
- [19] A. Solana-Ortega, personal communication, 2003.
- [20] van Campenhout, J. M. and Cover, T. M., *IEEE IT*, **27**, 483-489, (1981).
- [21] Vasicek, O. A., *Ann. Probab.*, **8**, 142-147, (1980).
- [22] Zabel, S., *Ann. Probab.*, **8**, 928-941, (1980).

## 8. CHANGES WRT THE ARXIV:MATH-PH/0407009, VERSION 3

Three major changes: 1) Definition of proper I-projection has been changed. 2) An argument preceding Eq. (7) at the proof of ICET (and similarly Eq. (9) at the proof of EGCP) is now correctly stated. 3) Abstract was rewritten to better reflect contents of paper.

This is the definitive form of the work. To appear at the Proceedings of Max-Ent'04 workshop.

INSTITUTE OF MATHEMATICS AND CS OF MATHEMATICAL INSTITUTE OF SLOVAK ACADEMY OF SCIENCES (SAS) AND OF MATEJ BEL UNIVERSITY, SEVERNA 5, 974 01 BANSKA BYSTRICA, SLOVAKIA AND INSTITUTE OF MEASUREMENT SCIENCE OF SAS, DUBRAVSKA CESTA 9, 841 04 BRATISLAVA, SLOVAKIA

*E-mail address:* marian.grendar@savba.sk