CONDITIONAL EQUI-CONCENTRATION OF TYPES, EXTENDED GIBBS CONDITIONING PRINCIPLE: PROOF DETAILS

M. GRENDAR

Abstract. Previously sketched proofs of Conditional Equi-concentration of Types on I-projections (ICET) and of Extended Gibbs Conditioning Principle (EGCP) are given in (some) detail.

1. Terminology and notation

Though the terminology and notation is the same as used at the previous works (cf. [2], [3]), it will be for completness recalled here.

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 $X \triangleq \{x_1, x_2, \dots, x_m\}$ 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 $\{X\}_{l=1}^{n}$ is the pmf $\nu^n \in P(X)$ whose i-th element ν_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 ν^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 Π 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 $A \subseteq \Pi$.

μ-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. [1]). Alternatively, the μ-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 Π_n then it is just the type ν^n.

2. 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 ϵ-ball - defined by the metric d - which is centered at $a \in P(X)$.

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

Date: October 3, 2004.

ICET. *Let* X *be a finite set. Let* Π *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* Π *in the ball* $B(\hat{p}^j, \epsilon)$. *Let* $n \to \infty$. *Then*

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

Proof of ICET relays upon the following Lemma, which states a standard inequality for ratio of probabilities:

Lemma. *Let* ν^n, $\dot{\nu}^n$ *be two types from* Π_n. *Then*

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

Also, MaxProb/MaxEnt Theorem - a result stating an asymptotic identity of μ-projections and I-projections - plays a role:

MaxProb/MaxEnt. [1] *Let* X *be a finite set. Let* M_n *be set of all* μ*-projections of* q *on* Π_n. *Let* I *be set of all* I*-projections of* q *on* Π. *For* $n \to \infty$, $M_n = I$.

Proof. *(of* I*CET)*
Clearly:

$$\pi(\nu^n \in B(\epsilon, \hat{p}^j) | \nu^n \in \Pi; q \mapsto \nu^n) \leq \frac{\sum_{\nu^n \in B} \pi(\nu^n; q)}{\sum_{\nu^n \in \Pi} \pi(\nu^n; q)} \tag{2}$$

$B_n(\epsilon, \hat{p}^j) \triangleq B(\epsilon, \hat{p}^j) \cap \Pi_n$.

Without loss of generality, let there be unique I-projection $\hat{p}^n_B$ of q on the ball $B_n(\hat{p}^j, \epsilon)$. (Sequence of the I-projections on Π_n converges to a proper I-projection of q on Π. To an I-projection on Π which is not proper, no sequence of I_n-projections converges.) Also, without loss of generality let there be k I-projections $\hat{p}^j_{\Pi_n}$, $j = 1, 2, \dots, k$ of q on Π_n.

Let $A \triangleq B_n \backslash \{\hat{p}^n_B\}$, $B \triangleq \Pi_n \backslash \{\hat{p}^j_{\Pi_n}\}, j \neq 1$, $C \triangleq \Pi_n \backslash B$.

Then the RHS of (1) can be rewritten as:

$$\frac{\pi(\hat{p}^n_B)}{\pi(\hat{p}^1_{\Pi_n})} \frac{1 + \frac{\sum_{\nu^n \in A} \pi(\nu^n)}{\pi(\hat{p}^n_B)}}{1 + \frac{\sum_{\nu^n \in B} \pi(\nu^n)}{\pi(\hat{p}^1_{\Pi_n})} + \frac{\sum_{\nu^n \in C} \pi(\nu^n)}{\pi(\hat{p}^1_{\Pi_n})}} \tag{3}$$

By MaxProb/MaxEnt Thm I-projections have for $n \to \infty$ the same and supremal value of $\pi(\cdot)$. This implies that $\pi(\hat{p}^n_B)/\pi(\hat{p}^1_{\Pi_n})$ converges to 1 (the case of $0/0$ limit is excluded by the supremity of $\pi(\cdot)$). The same argument implies that the first ratio in the denominator converges to $k - 1$. The Lemma implies that the ratio in the nominator as well as the second ratio in the denominator converge to zero. ☐

3. Extended Gibbs Conditioning Principle

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

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

Proof. Clearly:

(5)
$$\pi(X_1 = x_1, \dots, X_t = x_t | \nu^n \in \Pi; q \mapsto \nu^n) = \frac{\sum_{\nu^n \in \Pi} \pi(X_1 = x_1, \dots, X_t = x_t, \nu^n)}{\sum_{\nu^n \in \Pi} \pi(\nu^n; q)}$$

Let, in addition to partitioning used in proof of ICET, $D \triangleq \cup_{j=1}^{k}\{\hat{p}^{j}_{\Pi_n}\}$.

Then the RHS of (5) can be rewritten as:

(6)
$$\frac{\sum_{\nu^n \in D} \pi(X_1 = x_1, \dots, X_t = x_t, \nu^n) + \sum_{\nu^n \in \Pi_n \setminus D} \pi(X_1 = x_1, \dots, X_t = x_t, \nu^n)}{\pi(\hat{p}^1_{\Pi_n})(1 + \frac{\sum_{\nu^n \in B} \pi(\nu^n)}{\pi(\hat{p}^1_{\Pi_n})} + \frac{\sum_{\nu^n \in C} \pi(\nu^n)}{\pi(\hat{p}^1_{\Pi_n})})}$$

MaxProb/MaxEnt Thm implies that the first ratio in the denominator converges to $k-1$. By the Lemma, the second ratio in the denominator of (6) converges to zero as n goes to infinity. The second term in the nominator as well goes to zero as $n \to \infty$ (to see this, express the joint probability $\pi(X_1 = x_1, \dots, X_t = x_t, \nu^n)$ as $\pi(X_1 = x_1, \dots, X_t = x_t | \nu^n)\pi(\nu^n)$ and employ the Lemma).

Then, MaxProb/MaxEnt Thm implies that for $n \to \infty$ the RHS of (6) becomes $1/k \sum_{j=1}^{k} \pi(X_1 = x_1, \dots, X_t = x_t | \hat{p}^j)$. Finally, invoke Csiszár's 'urn argument' (cf. [?]) to conclude that the asymptotic form of the RHS of (6) is $1/k \sum_{j=1}^{k} \prod_{l=1}^{t} \hat{p}^{j}_{X_l}$. □

References

[1] Grendár, M., "Asymptotic identity of μ-projections and I-projections," to appear at Acta U. Belii Ser. Math., 2004.

[2] Grendár, M., "Conditional Equi-concentration of Types," to appear at *Focus on Probability Theory.*

[3] Grendár, M., "Gibbs conditioning extended, Boltzmann conditioning introduced," to appear at *Bayesian inference and Maximum Entropy methods in Science and Engineering.*

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