Annales Academiæ Scientiarum Fennicæ
Series A. I. Mathematica
Volumen 13, 1988, 219–223

ON THE CONVERGENCE OF EMPIRICAL DISTRIBUTIONS UNDER PARTIAL OBSERVATIONS

T. Lehtonen and E. Nummelin

1. Introduction

Starting from the classical results of Maxwell, Boltzmann and Gibbs the application of combinatorics and probability theory into statistical mechanics and thermodynamics has a long history. Today the advancement of the theory of large deviations seems to provide a solid basis for a general and unified theory of large particle systems. Today this theory is far from complete but much work is in progress. To pick up a few important contributions we would like to mention the work of Lanford and Ruelle, especially the paper [7]. The monograph [9] by Martin–Löf is also worthwhile. Many recent developments are treated by Ellis in the book [4], which also contains a large bibliography. Simultaneously to the expansion of the theory a growing number of applications has emerged in many fields where large systems are in consideration. An important area of application is mathematical statistics. For various aspects of this connection see e.g. Bahadur and Zabell [1], Barndorff–Nielsen [2], Groeneboom [5], Kester and Kallenberg [6] and the references therein.

In this paper we consider a statistical application. Our main theme is the conditional convergence of empirical distributions in the case of i.i.d. random elements. We assume that the values (states) of these elements remain unknown but a macroscopic observation is made about the mean of an $\mathbf{R}^d$-valued function of the state. We are interested in the empirical distribution of the random elements as well as in the probability distribution of a single element given this observation. It turns out that under certain regularity conditions they both converge to a common limiting distribution (see below). Similar, but weaker earlier results about the distribution of a single random element were obtained by Lanford [7] and van Campenhout and Cover [12]. Our paper is also greatly influenced by the book of Ellis. However, Ellis does not formulate a general approach but derives results for the special case called the "discrete" ideal gas. As compared with the book by Ellis and Csiszar's advanced article [3] the derivation of the results is here elementary in the sense that we do not use the abstract "level II" theory of large deviations. Finally it should be noted that this paper is intimately connected to [8], where we give a complete treatment with an application to physical systems, especially to the (continuous) ideal gas. In the present paper we omit the physical framework and expose the statistical content of the results.

doi:10.5186/aasfm.1988.1313

2. Preliminaries

Let $(E, \mathcal{E})$ be a measurable space. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. E-valued random elements having the common distribution $\mathbf{P}X_1^{-1} = P$. Let $u : E \to \mathbf{R}^d$ be a Borel-measurable mapping and denote $U_i = u(X_i)$. Denote $\overline{U}_{(n)} = n^{-1}\sum_{i=1}^n U_i$.

Let ϵ_x denote the unit mass at $x \in E$. Then the *empirical distribution* of the sample $X_1, X_2, \ldots, X_n$ may be written as

$$\hat{P}_n(dx) = n^{-1}\sum_{i=1}^n \epsilon_{X_i}(dx).$$

$\hat{P}_n$ is a random probability measure on $(E, \mathcal{E})$.

We assume that $X_1, X_2, \ldots$ remain unknown as such, but instead, the mean $\overline{U}_{(n)}$ is observed. We are interested in the convergence, as $n \to \infty$, of $\hat{P}_n$ given that $\overline{U}_{(n)} \in C$, where $C \subset \mathbf{R}^d$ is a convex Borel set. To state our results we still need some more terminology and notation.

Consider the random variables $U_1, U_2, \ldots$. Denote by Z the Laplace transform $Z(t) = \mathbf{E}\, e^{<t, U_1>} = \int e^{<t, u(x)>} P(dx)$ and let $\mathcal{D} = \{t \in \mathbf{R}^d;\ Z(t) < \infty\}$. We suppose that $\mathcal{D}$ is open.

For $t \in \mathcal{D}$, denote by P_t the conjugate distribution

$$P_t(dx) = Z(t)^{-1} e^{<t, u(x)>} P(dx).$$

Note that $P_0 = P$. Denote the expectation with respect to P_t by $\mathbf{E}_t$ and let $m(t) = \mathbf{E}_t U_1$.

Let S be the closed convex hull of the support of the distribution of U_1 and let $\mathrm{ri}\,(S)$ denote the relative interior of S. We suppose that $\mathrm{ri}\,(S)$ is nonempty.

Let (W_n) be a sequence of $\mathbf{R}^d$-valued random variables and let $(\mathbf{P}_n)$ be a sequence of probability measures on the underlying sample space. Following [4] it is said that the *random variables W_n converge, as $n \to \infty$, to a random variable W_∞ exponentially with respect to the probabilities $\mathbf{P}_n$*, if for each $\epsilon > 0$ there exists a constant $I_\epsilon > 0$ so that

$$\mathbf{P}_n(|W_n - W_\infty| > \epsilon) < e^{-I_\epsilon n} \quad \text{eventually.}$$

This is denoted by $W_n \underset{\exp}{\longrightarrow} W_\infty$ (with respect to $(\mathbf{P}_n)$).

Let (π_n) be a sequence of random probability measures and let P be a fixed probability measure on $(E, \mathcal{E})$. We say that the *random probabilities π_n converge to P exponentially with respect to the probabilities $\mathbf{P}_n$*, if for all bounded measurable functions $f : E \to R$ the random variables $\int f d\pi_n$ converge to the limit $\int f dP$ exponentially with respect to $\mathbf{P}_n$.

Finally let $P, P_1, P_2, \ldots,$ be (nonrandom) probability measures on $(E, \mathcal{E})$. We say that the probabilities P_n *b-converge* to P, if for all bounded measurable functions $f : E \to R$ the integrals $\int f dP_n$ converge to $\int f dP$.

Below we consider the *Kullback–Leibler information* h defined by

$$h(t) = K(P_t, P) = E_t \log(dP_t/dP), \quad t \in \mathcal{D}.$$

Note that

$$h(t) = < t, E_t U_1 > - \log\ Z(t),$$

where

$$E_t U_1 = m(t) = (d/dt)(\log\ Z(t)).$$

3. Convergence results

A basic starting point for our results is the following wellknown lemma from convex analysis (see e.g. [2] and [10]). Let C^o (resp. $\overline{C}$) denote the interior (resp. the closure) of a set C.

Lemma. *Let $C \subset \mathbf{R}^d$ be a convex Borel set so that $C^o \cap \mathrm{ri}\,(S) \neq \emptyset$ and $\mathbf{E}U_1 \notin \overline{C}$. Then there exists a unique point $t = t_C \in m^{-1}(\overline{C})$ so that*

$$h(t_C) = \inf_{t: m(t) \in C} h(t)$$

and actually

$$m(t_C) \in \partial C \cap \mathrm{ri}\,(S).$$

It was shown in [11] that the following conditional weak law of large numbers holds true: Under the assumptions of the lemma

$$\overline{U}_{(n)} \underset{\mathbf{exp}}{\longrightarrow} m(t_C)$$

with respect to the conditional probabilities $\mathbf{P}_{n,C} = \mathbf{P}(\cdot \mid \overline{U}_{(n)} \in C)$.

Let $g : E \to \mathbf{R}^k$ be a bounded, Borel-measurable mapping and denote $G_i = g(X_i)$. Denote $\overline{G}_{(n)} = n^{-1} \sum_{i=1}^n G_i$. Applying the conditional weak law of large numbers in the case of partial observations we may derive the following result (see [8]).

Theorem. (Conditional convergence of sample means under partial observations.) *Under the assumptions of the lemma*

$$\overline{G}_{(n)} \underset{\mathbf{exp}}{\longrightarrow} \int g(x) P_{t_C}(dx) \quad \text{as} \quad n \to \infty$$

with respect to the conditional probabilities $\mathbf{P}_{n,C}$.

Our main results follow as easy corollaries from this theorem. First, considering a bounded, real valued function g note that $\int g d\hat{P}_n = \overline{G}_{(n)}$. Consequently we have

Corollary 1. (Conditional convergence of empirical distributions under partial observations.) *Under the assumptions of the lemma the empirical distributions* $\hat{P}_n$ *converge, as* $n \to \infty$, *to the conjugate distribution* P_{t_C} *exponentially with respect to the conditional probabilities* $\mathbf{P}_{n,C}$.

Second, observe that the conditional convergence of the empirical distributions implies the b-convergence of the probabilities

$$\int \hat{P}_n(\omega, \cdot)\mathbf{P}_{n,C}(d\omega) = P(X_1 \in \cdot \mid \overline{U}_{(n)} \in C)$$

to the same limit. Hence we obtain finally:

Corollary 2. *Under the assumptions of the lemma the conditional distributions* $P(X_1 \in \cdot \mid \overline{U}_{(n)} \in C)$ b*-converge, as* $n \to \infty$, *to the conjugate distribution* P_{t_C}.

References

[1] BAHADUR, R.R., and S.L. ZABELL: Large deviations of the sample mean in general vector spaces. - Ann. Probab. 7, 1979, 587–621.

[2] BARNDORFF-NIELSEN, O.: Information and exponential families in statistical theory. - Wiley, Chichester, 1978.

[3] CSISZAR, I.: Sanov property, generalized I-projection and a conditional limit theorem. - Ann. Probab. 12, 1984, 786–793.

[4] ELLIS, R.S.: Entropy, large deviations and statistical mechanics. - Springer-Verlag, New York, 1985.

[5] GROENEBOOM, P.: Large deviations and asymptotic efficiences. - Mathematical Centre Tracts 118, Amsterdam, 1980.

[6] KESTER, A.D.M., and W.C.M. KALLENBERG: Large deviations of estimators. - Ann. Statist. 14, 1986, 648–664.

[7] LANFORD, O.E.: Entropy and equilibrium states in classical statistical mechanics. - In: Statistical Mechanics and Mathematical Problems. Lecture Notes in Physics 20, 1–113. Springer-Verlag, Berlin, 1973.

[8] LEHTONEN, T., and E. NUMMELIN: The probability law of the ideal gas. – An approach via level I theory of large deviations. - Preprint, 1988.

[9] MARTIN-LÖF, A.: Statistical mechanics and the foundations of thermodynamics. - Lecture Notes in Physics 101. Springer-Verlag, Berlin, 1979.

[10] NEY, P.: Dominating points and asymptotics of large deviations for random walk on $\mathbf{R}^d$. - Ann. Probab. 11, 1983, 158–167.

[11] NUMMELIN, E.: A conditional weak law of large numbers. - To appear in: Proceedings of the International Seminar on Stability Problems for Stochastic Processes, Sukhumi, U.S.S.R., October 5–10, 1987.

[12] van Campenhout, J.M., and T.M. Cover: Maximum entropy and conditional probability. - IEEE Trans. Inform. Theory, IT–27, 1981, 483–489.

Helsinki School of Economics
Runeberginkatu 14–16
SF–00100 Helsinki
Finland

University of Helsinki
Department of Mathematics
Hallituskatu 15
SF–00100 Helsinki
Finland