PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 131, Number 8, Pages 2597–2609
S 0002-9939(02)06785-0
Article electronically published on November 27, 2002

CONDITIONAL WEAK LAWS IN BANACH SPACES

ANA MEDA

(Communicated by Claudia M. Neuhauser)

Abstract. Let $(B, \|\cdot\|)$ be a separable Banach space. Let $Y, Y_1, Y_2, \ldots$ be centered i.i.d. random vectors taking values on B with law μ, $\mu(\cdot) = P(Y \in \cdot)$, and let $S_n = \sum_{i=1}^n Y_i$. Under suitable conditions it is shown for every open and convex set $0 \notin D \subset B$ that $P\left(\|\frac{S_n}{n} - v_d\| > \varepsilon \middle| \frac{S_n}{n} \in D\right)$ converges to zero (exponentially), where v_d is the dominating point of D. As applications we give a different conditional weak law of large numbers, and prove a limiting aposteriori structure to a specific Gibbs twisted measure (in the direction determined solely by the same dominating point).

1. Introduction

Conditional laws related to dominating points have been studied in recent years in the context of large deviations. For i.i.d. in $\mathbb{R}^d$ see, for example, [19], [12] and [13], [6]. For general sequences of the Gartner-Ellis type and for the Markov case in $\mathbb{R}^d$ see [14], [15], [16]; for the i.i.d. case in Banach spaces see [4]. Some authors refine a method developed by Csiszár (1984) [3], that deals directly with the aposteriori structure (also known as the Gibbs conditioning principle) without proving conditional laws of large numbers. See also [5] and [14], and references within, on the study of conditional laws and the Gibbs conditioning principle that do not deal directly with dominating points.

The main result of this work, Theorem 1, is a Nummelin conditional weak law of large numbers in Banach spaces that Nummelin (1987) proved first on $\mathbb{R}^d$ (see [19]). The limits are identified in terms of the dominating point. Then we apply a method developed by Lehtonen and Nummelin ([12] and [13]) to prove another conditional weak law of large numbers for different functions (Theorem 2), and a limiting aposteriori structure (Theorem 3).

The last result, Theorem 3, is implied by the results in [4] by Dembo and Kuelbs, but our approach is interesting because in some situations, as in the Markov case (which is not the subject of this paper), our treatment allows us to obtain the most general results to date. For a detailed account one can look at [16], Chapters 1 and 5 of [14] and the bibliography therein. We were able to see in [14], p. 69, that under a Harris condition and using the large deviations obtained for

Received by the editors July 16, 2000 and, in revised form, March 21, 2002.
2000 *Mathematics Subject Classification.* Primary 60F10, 60B10, 60F05, 60G50.
Key words and phrases. Conditional laws, dominating point, large deviations, Banach spaces.
The author was supported in part by Grant PAPIIT-DGAPA IN115799 of UNAM, and the final version was written while holding a Postdoctoral position at IMP, México.

Markov chains by Iscoe, Ney and Nummelin in [10], the k^{th} order empirical measure of $X_0, X_1, \ldots$ (defined as $\widetilde{P}_n^{(k)}(A) = 1/n \sum_{i=0}^{n-1} \delta_{(X_{i+1}, X_{i+2}, \ldots, X_{i+k})}(A)$) satisfies $\mathbb{E}_x \left(\widetilde{P}_n^{(k)} \,\middle|\, \frac{U_n}{n} \in C; X_n \in A \right) \xrightarrow{b} \rho^{(k)}$ for every k, where

$$\rho^{(k)}(\cdot) = \int_{(\cdot)} \pi^*(dx_1) \prod_{i=1}^{k-1} q^*(x_i, dx_{i+1}),$$

and q^* is a stochastic transition kernel with invariant measure π^*. This means that the aposteriori distribution *bounded* converges in a Césaro sense to a Markov process

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{P}_x \left(X_{i+1}^{i+k} \in \cdot \,\middle|\, \frac{U_n}{n} \in C; X_n \in A \right) \xrightarrow{b} \rho^{(k)}(\cdot).$$

The organization of the work is as follows: In Section 2 we state our results, in Section 3 we summarize some well-known results we use later. Sections 4, 5, and 6 are dedicated to proving Theorems 1, 2, and 3, respectively, except for some lemmas whose proofs we present in Section 7.

2. Hypothesis and statement of results

Let $(B, \|\cdot\|)$ be a separable Banach space. Let $Y, Y_1, Y_2, \ldots$ be i.i.d. random vectors taking values on B with law μ and underlying probability measure P, i.e. $\mu(\cdot) = P(Y \in \cdot)$. S_n denotes the sum $S_n = \sum_{i=1}^n Y_i$. B^* will denote the dual space of B, and for a set $U \subseteq B$ whenever we refer to its interior (U^o), closure ($\overline{U}$), boundary (∂U), etc., we use the norm topology.

We state the following assumption because it simplifies the notation and avoids degenerate situations.

Hypothesis 1. *Assume μ is not concentrated on a point and it has mean zero.*

Hypothesis 2. *Assume that for every $t > 0$,*

$$\int_B e^{t\|x\|} d\mu(x) < \infty.$$

Remark 1. Hypothesis 2 is satisfied for example if $\mathrm{supp}(\mu)$ is bounded, and this happens when μ is the law of a bounded random vector. This assumption used to be standard in the study of probability in Banach spaces. Kuelbs, however (see [11]), was able to construct a dominating point assuming that Hypothesis 2 holds only for *some* $t > 0$ instead of for *every* t. His result allows the proof of sharp asymptotics for the Large Deviations Principle (LDP) with minimal assumptions. In the discussion below we keep Hypothesis 2 because it gives us exponential tightness.

Let λ be, for $\zeta \in B^*$, the Laplace transform of μ

$$\lambda(\zeta) = \int_B e^{\zeta(x)} d\mu(x).$$

We consider $\Lambda(\zeta) = \log \lambda(\zeta)$ and Λ^* the convex conjugate of λ defined for all $x \in B$:

$$\Lambda^*(x) = \sup_{\{\zeta \in B^*\}} \left\{ \zeta(x) - \log \lambda(\zeta) \right\}, \tag{2.1}$$

and for a set U,

$$\Lambda^*(U) = \inf_{\{x \in U\}} \{\Lambda^*(x)\}.$$

Hypothesis 3. *Let D be an open convex subset of B with $0 \notin D$ and $D \cap \{x : \Lambda^*(x) < \infty\} \neq \varnothing$.*

Remark 2. We assume D to be open and convex in order to have a dominating point for D, and that is guaranteed by Theorem 1 below, in which the definition of dominating point is also stated. The rest of the assumption is to avoid degeneracies: $0 \notin D$ avoids the case in which the mean of μ is in D. In that case the conditioning set has probability one and therefore it has no effect on the limit, so we would not have a conditional law of large numbers. This was noted, among others, in [2].

Einmahl and Kuelbs (1996) in [9] proved the existence of a unique dominating point (with supporting hyperplane) for every open convex set in separable Banach spaces, plus asymptotic results. The following is part of their Theorem 1:

Proposition 1. *Under Hypotheses 1, 2 and 3 there is a unique point (called the dominating point of D) that depends only on D and Λ. It is denoted v_D and has the following properties:*

$$v_D \in \partial D,$$

$$\Lambda^*(v_D) = \Lambda^*(D) = \Lambda^*(\overline{D}).$$

For some $\zeta = \zeta_D \in B^*$ that we call the supporting hyperplane,

$$D \subseteq \{x : \zeta_D(x) \geq \zeta_D(v_D)\}$$

and

$$\Lambda^*(v_D) = \zeta_D(v_D) - \Lambda(\zeta_D). \tag{2.2}$$

The dominating point has a representation in terms of a Bochner integral

$$v_D = \int_B x \exp\{\zeta_D(x) - \Lambda(\zeta_D)\} d\mu(x). \tag{2.3}$$

Remark 3. The point ζ_D is called a *supporting hyperplane* only to shorten notation, since it should be the *functional that generates the supporting hyperplane.* Formula (2.2) means that it is at $\zeta = \zeta_D$ that the expression $\zeta(v_D) - \Lambda(\zeta)$ reaches its maximum.

Remark 4. The last integral of the Theorem above is the equivalent to the statement about $\nabla\Lambda$ for the $\mathbb{R}^d$ case, where dominating points were first defined by Ney. (See [17], [18], and also [9], [11], and [7].) In the notation of (2.3), v_D is the mean with respect to the twisted measure

$$d\mu_\zeta(x) = \frac{e^{\zeta(x)}}{e^{\Lambda(\zeta)}} \, d\mu(x) \tag{2.4}$$

in the direction $\zeta = \zeta_D$ of its supporting hyperplane.

Definition 1. Let $\mathbb{P}_1, \mathbb{P}_2, \ldots$ be a sequence of probability measures. We say that *Y_n converges exponentially to Y with respect to $\mathbb{P}_n$* if for every $\epsilon > 0$ there is a constant $a > 0$ such that

$$\mathbb{P}_n (\|Y_n - Y\| > \epsilon) < e^{-an}$$

for n sufficiently large. In this case we write

$$Y_n \xrightarrow{\exp} Y \qquad \text{with respect to } \mathbb{P}_n.$$

Now we are ready to state a conditional weak law of large numbers. The conditioning implies that the observations are away from the mean ($0 \notin D$, in Hypothesis 3), and our conclusion is that the dominating point turns out to be the *new mean* in the sense that $\frac{S_n}{n}$ converges to it.

Theorem 1 (Conditional weak law of large numbers)**.** *Assume Hypotheses 1, 2 and 3 above. Then* $\frac{S_n}{n}$ *converges exponentially to* v_D *with respect to the conditional probabilities*

$$\mathbb{P}_n = P\left(\cdot \middle| \frac{S_n}{n} \in D\right).$$

As applications we shall prove analogous theorems to those in [13] by Lehtonen and Nummelin. The first one (Theorem 2 below) states exponential convergence of sums of functions of random elements $X, X_1, \ldots$ when the converging and the conditioning functions are different. In order to state those theorems we need more notation.

Let B_i be separable Banach spaces with norms $\|\cdot\|_i$ and dual spaces B_i^*, for $i = 1, 2$.

Consider i.i.d. random elements $X, X_1, X_2, \ldots$ with underlying probability space $(\Omega, \mathcal{F}, P)$ taking values on a measurable space $(S, \mathcal{S})$ (an *intermediate* space, as will be seen), and with law μ. Take *bounded* measurable functions $g : S \to B_1$ and $u : S \to B_2$.

For a bounded measurable function $f : S \to B$ call M_f a bound for $\|f\|$ and μ^f will be the probability measure (on $(B, \mathbb{B}_B)$, the space B with its Borel sets) induced by $f(X)$. Let λ_f be the Laplace transform

$$\lambda_f(\zeta) = \int_B \exp^{\zeta(x)} d\mu^f(x) = \int_S \exp^{\zeta(f(x))} d\mu(x), \qquad \zeta \in B^*. \tag{2.5}$$

As before, we consider $\Lambda_f(\zeta) = \log \lambda_f(\zeta)$ and Λ_f^* the convex conjugate of λ_f.

We will slightly modify our hypotheses to accommodate our new situation.

Hypothesis 1′. *Assume that the laws* $\mu^u(\cdot) = \mathcal{L}(u(X)) = P(u(X) \in \cdot)$ *and* $\mu^g(\cdot) = \mathcal{L}(g(X))$ *are not concentrated on a point and* $\mu^u = \mathcal{L}(u(X))$ *has mean zero.*

Hypothesis 3′. *Let* D *be an open convex subset of* B_2, *with* $0 \notin D$ *and* $D \cap \{y : {\Lambda^*}_u(y) < \infty\} \neq \varnothing$.

Remark 5. We will see that the Hypotheses above and Theorem 1 give that (D, Λ_u) has a dominating point (on B_2). The dominating point of D will be denoted v_D with ζ_D its supporting hyperplane and μ_{ζ_D} the twisted measure in the direction of ζ_D as defined in (2.4). The only hypothesis we are left to verify for the existence of v_D is the analog of Hypothesis 2, but that is a consequence of the boundedness of the function u, as will be seen in (5.1).

Theorem 2 (Conditioned law for different functions)**.** *Let* g *and* u *be bounded measurable functions taking values on different Banach spaces. Let* $G_n = \sum_{i=1}^n g(X_i)$ *and* $U_n = \sum_{i=1}^n u(X_i)$. *Assume Hypotheses 1′ and 3′ with all the notation described above. Then*

$$\frac{G_n}{n} \xrightarrow{\exp} \int_S g(x) d\mu_{\zeta_D}(x),$$

with respect to the conditional probabilities

$$\mathbb{P}_n = P\left(\cdot \;\middle|\; \frac{U_n}{n} \in D \right).$$

Remark 6. The integral above is a Bochner integral. It is because of the specific form of the limit in Theorem 2 above that we will be able to prove an aposteriori limit law that applies when the function g in Theorem 2 above takes values on $\mathbb{R}$.

Definition 2. For measures $\mu, \mu_n, n = 1, 2, \ldots$, write $\mu_n \xrightarrow{b} \mu$ if

$$\lim_{n\to\infty} \int_S f(x)d\mu_n(x) = \int_S f(x)d\mu(x) \tag{2.6}$$

for all bounded measurable $f : S \to \mathbb{R}^d$. In this case we say μ_n *bounded converges* to μ.

Remark 7. Note that bounded convergence is convergence in the τ-topology. Note also that this convergence is the same if defined for functions $f : S \to \mathbb{R}$ (instead of $\mathbb{R}^d$, as we did in [16]).

Theorem 3. *Under the same hypotheses as in Theorem 2,*

$$P\left(X_1 \in \cdot \;\middle|\; \frac{U_n}{n} \in D \right) \xrightarrow{b} \mu_{\zeta_D}(\cdot).$$

Remark 8. This theorem is an aposteriori law. With the same method we are able to prove a weak convergence version of "quasi-independence", which means that for every fixed k,

$$P\left((X_1, X_2, \cdots, X_k) \in \cdot \;\middle|\; \frac{U_n}{n} \in D \right) \xrightarrow{b} \mu^k_{\zeta_D}(\cdot). \tag{2.7}$$

Regarding quasi-independence we must say there are better results than formula (2.7): Csiszár (1984) in [3] proved a similar aposteriori law but in a stronger form, which in particular implies convergence in total variation norm. Later, in 1998, Dembo and Kuelbs [4] proved a very interesting (and almost surprising) generalization of Csiszár's result, which says that k is allowed to grow with n (up to some order).

3. Summary of known results

Donsker and Varadhan (1976) [8], proved that Λ^* is a *good* rate function and the LDP holds with rate function Λ^*.

Theorem 4. *Under Hypothesis 2, for every closed set $F \subseteq B$*

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log P\left(\frac{S_n}{n} \in F \right) \le -\Lambda^*(F),$$

and for every open set $G \subseteq B$

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \in G \right) \ge -\Lambda^*(G).$$

Combining Theorems 4 and 1 we obtain that under Hypotheses 1, 2, and 3,

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \in D \right) = -\Lambda^*(v_D). \tag{3.1}$$

The next theorem is part of Lemma (2.2) by de Acosta in [1] (1985).

Theorem 5. *Assume Hypothesis 2. Then for every $a > 0$ there is a compact set $K = K_a \subseteq B$ and a constant $N = N_a$ such that*

$$P\left(\frac{S_n}{n} \notin K_a\right) \leq e^{-na} \qquad \text{for every } n \geq N.$$

The following lemma was stated by Lehtonen and Nummelin [13] (a proof can be found in [14]). To state the result we need a definition.

Definition 3. Let $\widetilde{\mu}_n(\cdot, \omega)$ with $\omega \in (\Omega, \mathcal{F})$ and $n = 1, 2, \ldots$ be random measures on $(S, \mathcal{S})$. Let μ be a nonrandom measure and let $\mathbb{P}_n, \quad n = 1, 2, \ldots,$ be probability measures on $(\Omega, \mathcal{F})$. Say that $\widetilde{\mu}_n$ converges exponentially to μ with respect to $\mathbb{P}_n$, and write

$$\widetilde{\mu}_n \xrightarrow{\exp} \mu \qquad \text{with respect to } \mathbb{P}_n$$

if for all bounded measurable $f : S \to \mathbb{R}^d$,

$$\int_S f(x) \widetilde{\mu}_n(dx, \omega) \xrightarrow{\exp} \int_S f(x) d\mu(x) \qquad \text{with respect to } \mathbb{P}_n.$$

Remark 9. d in the definition of f above need not be specified.

Lemma 1. *Let $\widetilde{\mu}_n$ be a sequence of random probability measures and let μ be a nonrandom probability measure. If $\widetilde{\mu}_n \xrightarrow{\exp} \mu$ with respect to $\mathbb{P}_n$, then*

$$E_{\mathbb{P}_n}(\widetilde{\mu}_n) \xrightarrow{b} \mu.$$

4. Proof of Theorem 1

Let $\epsilon > 0$. Define $D^{(\epsilon)} = D \cap \{x : \|x - v_D\| > \epsilon\}$.

For each $\zeta \in B^*$ and $\omega > 0$ let

$$H(\zeta, \omega) = \{x : \zeta(x) - \Lambda(\zeta) > \Lambda^*(v_D) + \omega\}$$

and

$$H^D(\zeta, \omega) = D \cap H(\zeta, \omega).$$

Lemma 2. *With the notation above, if $\zeta \in B^*$ and $\omega > 0$,*

$$\Lambda^*(\overline{D}) < \Lambda^*\left(\overline{H^D(\zeta, \omega)}\right) \tag{4.1}$$

and

$$\overline{D^{(\epsilon)}} \subseteq \overline{D} \backslash \{v_D\} \subseteq \bigcup_{\{\zeta \in B^*\}} \bigcup_{\{\omega > 0\}} H(\zeta, \omega). \tag{4.2}$$

Let us write, for a measurable set K_a,

$$\begin{aligned} P\left(\left\|\frac{S_n}{n} - v_D\right\| > \epsilon \,\Big|\, \frac{S_n}{n} \in D\right) &\leq \frac{P\left(\left\|\frac{S_n}{n} - v_D\right\| > \epsilon; \frac{S_n}{n} \in D \cap K_a\right)}{P\left(\frac{S_n}{n} \in D\right)} \\ &+ \frac{P\left(\left\|\frac{S_n}{n} - v_D\right\| > \epsilon; \frac{S_n}{n} \in D \cap \ B \backslash K_a\right)}{P\left(\frac{S_n}{n} \in D\right)} \\ &= \mathbf{I} + \mathbf{II} \end{aligned} \tag{4.3}$$

where it is quite general but will be a compact set, as in Theorem 5.

Let K_a be a compact set as in Theorem 5. Then $\overline{D^{(\epsilon)}} \cap K_a$ is also compact, and by (4.2) in Lemma 2 the sets $H(\zeta, \omega)$ are an open cover. Let $\bigcup_{i=1}^m H(\zeta_i, \omega_i)$ be a finite subcover, and rename $H_i = H(\zeta_i, \omega_i)$ for $i = 1, 2, \ldots, m$.

Lemma 3. *We have*

$$D^{(\epsilon)} \cap K_a \subseteq \bigcup_{i=1}^m (H_i \cap D) = \bigcup_{i=1}^m H^D(\zeta_i, \omega_i). \tag{4.4}$$

It only remains to bound both terms on the right-hand side of inequality (4.3).

$$\begin{aligned}
\overline{\lim_{n\to\infty}} \frac{1}{n} \log(\ \mathbf{I}\ ) &\leq \overline{\lim_{n\to\infty}} \frac{1}{n} \log P\left(\frac{S_n}{n} \in D^{(\epsilon)} \cap K_a\right) \\
&\quad - \underline{\lim}_{n\to\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \in D\right) \\
&\leq \overline{\lim_{n\to\infty}} \frac{1}{n} \log P\left(\frac{S_n}{n} \in \bigcup_{i=1}^m H^D(\zeta_i, \omega_i)\right) - (-\Lambda^*(D)) \\
&\leq \overline{\lim_{n\to\infty}} \frac{1}{n} \log P\left(\frac{S_n}{n} \in \bigcup_{i=1}^m \overline{H^D(\zeta_i, \omega_i)}\right) + \Lambda^*(v_D) \\
&\leq -\Lambda^*\left(\bigcup_{i=1}^m \overline{H^D(\zeta_i, \omega_i)}\right) + \Lambda^*(v_D) \\
&= -\min_{\{1\leq i\leq m\}} \left\{\Lambda^*\left(\overline{H^D(\zeta_i, \omega_i)}\right)\right\} + \Lambda^*(v_D) \\
&= \max_{\{1\leq i\leq m\}} \left\{-\Lambda^*\left(\overline{H^D(\zeta_i, \omega_i)}\right)\right\} + \Lambda^*(v_D) < 0, \qquad (4.5)
\end{aligned}$$

by Lemma 3, (3.1), Theorem 4 and Lemma 2. We also have, for similar reasons plus Theorem 5,

$$\begin{aligned}
\overline{\lim_{n\to\infty}} \frac{1}{n} \log(\ \mathbf{II}\ ) &\leq \overline{\lim_{n\to\infty}} \frac{1}{n} \log P\left(\|\frac{S_n}{n} - v_D\| > \epsilon; \frac{S_n}{n} \in D \cap B\backslash K_a\right) \\
&\quad - \underline{\lim}_{n\to\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \in D\right) \\
&\leq \overline{\lim_{n\to\infty}} \frac{1}{n} \log P\left(\frac{S_n}{n} \in B\backslash K_a\right) + \Lambda^*(v_D) \\
&\leq -a + \Lambda^*(v_D) < 0. \qquad (4.6)
\end{aligned}$$

The last inequality (4.6) holds if a is sufficiently large; a was free until now. This, plus the previous inequality (4.5), yields Theorem 1.

5. Proof of Theorem 2

The proof of Theorem 2 consists of three parts: First, we verify the existence of a dominating point for D. Second, in Lemma 4 we use an auxiliary function $f = (g, u) : S \to B = B_1 \times B_2$ and apply the conditional law in Theorem 1 to $\frac{F_n}{n}$ where $F_n = \sum_{i=1}^n f(X_i)$, with conditioning set $(B_1 \times D)$, dominating point $v_{B_1\times D} = v_{B_1\times D}(\Lambda_f, B_1 \times D)$ and supporting hyperplane $\zeta_{B_1\times D}$. Finally, in Lemma 5 we identify the limit in terms of $v_D = v_D(\Lambda_u, D)$, i.e., we will prove that the twisted measure $\mu_{\zeta_{B_1\times D}}$ determined by $v_{B_1\times D}$ *does not* depend on g at all.

For the existence of v_D we only need to verify one condition (Hypothesis 2) to apply Theorem 1. Let $t > 0$. The function $\exp^{t\|u(x)\|_2}$ is bounded (uniformly on S), so

$$\int_{B_2} \exp^{t\|x_2\|_2} d\mu^u(x) = \int_S \exp^{t\|u(x)\|_2} d\mu(x) \le e^{tM_u} < \infty. \tag{5.1}$$

Hypothesis 2 holds and this part is done.

To check the convergence of $\frac{F_n}{n}$ we construct the product space $B = B_1 \times B_2$ in the natural way. For $x \in B$ the sub-index i of x represents the coordinate of x that belongs to B_i for $i = 1, 2$ or equivalently $x = (x_1, x_2)$. The norm of x is $\|x\| = \|x_1\|_1 + \|x_2\|_2$ and with this structure B is a Banach space. We use exactly the same notation for elements ζ of B^*:

$$\zeta(x) = \zeta(x_1, x_2) = \zeta(x_1, 0) + \zeta(0, x_2) = \zeta_1(x_1) + \zeta_2(x_2), \qquad \zeta_i \in B_i^*,$$

and every functional $\zeta_i \in B_i^*$ induces an element of B^*, e.g. $\zeta_2(x_2) = (0, \zeta_2)(x_1, x_2)$. Abusing the notation, we may omit the sub-index for $\lambda = \lambda_f, \Lambda = \Lambda_f$ and $\Lambda^* = {\Lambda^*}_f$, whereas when referring to u we will not omit it. Note that

$$\Lambda((0, \zeta)) = \Lambda_u(\zeta) \tag{5.2}$$

follows from the definitions of Λ^* and ${\Lambda^*}_u$, and if $\mu^f = \mathcal{L}\left(f(X)\right)$,

$$\begin{aligned} \Lambda((0,\zeta)) &= \log \int_B \exp\{(0,\zeta)(x)\} d\mu^f \\ &= \log \int_S \exp\{(0,\zeta)(f(x))\} d\mu(x) \\ &= \log \int_S \exp\{\zeta(u(x))\} d\mu(x) \\ &= \Lambda_u(\zeta). \end{aligned}$$

Consequently the corresponding twisting as in (2.4) is

$$\begin{aligned} d\mu_{(0,\zeta_2)}(x) &= \frac{e^{(0,\zeta_2)(x)}}{e^{\Lambda((0,\zeta_2))}} d\mu(x) \\ &= \frac{e^{\zeta_2(x)}}{e^{\Lambda_u(\zeta_2)}}\, d\mu(x) \\ &= d\mu_{\zeta_2}(x). \end{aligned} \tag{5.3}$$

Lemma 4. *$f(X), f(X_1), f(X_2), \ldots$ and $B_1 \times D$ with Hypotheses 1′ and 3′ satisfy Hypotheses 1, 2 and 3 of Theorem 1.*

With Lemma 4 we can proceed to apply Theorem 1 to get $v_{B_1\times D}$, $\zeta_{B_1\times D}$ and $\mu_{\zeta_{B_1\times D}}$ such that

$$\frac{F_n}{n} \xrightarrow{\exp} v_{B_1\times D} = \int_B x\, d\mu^f_{\zeta_{B_1\times D}}(x) = \int_S f(x) d\mu_{\zeta_{B_1\times D}}(x) \tag{5.4}$$

with respect to

$$\mathbb{P}_n = P\left(\cdot \;\middle|\; \frac{F_n}{n} \in (B_1 \times D)\right).$$

We can simplify (5.4) as follows: The limiting functions

$$\begin{aligned}\left\|\frac{F_n}{n}-v_{B_1\times D}\right\| &= \left\|\frac{G_n}{n}-(v_{B_1\times D})_1\right\|_1+\left\|\frac{U_n}{n}-(v_{B_1\times D})_2\right\|_2\\ &\geq \left\|\frac{G_n}{n}-(v_{B_1\times D})_1\right\|_1.\end{aligned}$$

The conditioning sets

$$\left\{\frac{F_n}{n}\in(B_1\times D)\right\}=\left\{\frac{G_n}{n}\in B_1,\ \frac{U_n}{n}\in D\right\}=\left\{\frac{U_n}{n}\in D\right\}.$$

Hence, we have

$$\frac{G_n}{n}\xrightarrow{\exp}(v_{B_1\times D})_1 \tag{5.5}$$

with respect to

$$\mathbb{P}_n=P\left(\cdot\middle|\ \frac{U_n}{n}\in D\right).$$

For the identification of the limit it only remains to use the next lemma to have the statement of the theorem:

Lemma 5.

$$(v_{B_1\times D})_1=\int_S g(x)d\mu_{\zeta_D}(x). \tag{5.6}$$

Remark 10. We already had by Lemma 4 (but it is not enough),

$$(v_{B_1\times D})_1=\int_S g(x)d\mu_{\zeta_{B_1\times D}}(x).$$

6. The aposteriori structure

In this part we shall prove Theorem 3. The proof uses the empirical measures

$$\widehat{P}_n(\cdot)=\frac{1}{n}\sum_{i=1}^n\delta_{X_i}(\cdot).$$

Recall

$$\frac{F_n}{n}=\frac{1}{n}\sum_{i=1}^n f(X_i)=\int f(x)d\widehat{P}_n(x). \tag{6.1}$$

Now, by Theorem 2

$$\frac{G_n}{n}\xrightarrow{\exp}\int_S g(x)d\mu_{\zeta_D}(x)\qquad\text{with respect to }\ \mathbb{P}_n=P\left(\cdot\middle|\ \frac{U_n}{n}\in D\right).$$

Theorem 2 holds for every bounded function $g:S\to\mathbb{R}$, so we have by definition of exponential convergence of measures (Definition 3) and observation (6.1) that

$$\widehat{P}_n\xrightarrow{\exp}\mu_{\zeta_D}\qquad\text{with respect to }\mathbb{P}_n=P\left(\cdot\middle|\ \frac{U_n}{n}\in D\right).$$

Apply Lemma 1 to get $E_{\mathbb{P}_n}\widehat{P}_n\xrightarrow{b}\mu_{\zeta_D}$, but this formula says

$$\frac{1}{n}\ \sum_{i=1}^n P\left(X_i\in\cdot\ \middle|\ \frac{U_n}{n}\in D\right)\xrightarrow{b}\mu_{\zeta_D}(\cdot).$$

Independence gives, since each term above is identical,

$$P\left(X_1 \in \cdot \;\Big|\; \frac{U_n}{n} \in D\right) \xrightarrow{b} \mu_{\zeta_D}(\cdot),$$

and we are done.

7. Proof of the lemmas

Proof of Lemma 2. Note that for all $\zeta \in B^*$ and $\omega > 0$ both $H(\zeta,\omega)$ and $H^D(\zeta,\omega)$ are open and convex, so by uniqueness of the dominating point, for all $x \in \overline{D}\backslash\{v_D\}$, $\Lambda^*(x) - \Lambda^*(v_D) > 0$. Hence

$$\begin{aligned} \overline{D}\backslash\{v_D\} &\subseteq \bigcup_{\{\omega>0\}} \{x : \Lambda^*(x) > \Lambda^*(v_D) + \omega\} \\ &= \bigcup_{\{\omega>0\}} \bigcup_{\{\zeta\in B^*\}} H(\zeta,\omega). \end{aligned} \tag{7.1}$$

The last equality in (7.1) holds because $\Lambda^*(x)$ is a supremum. Set $\zeta \in B^*$ and $\omega > 0$. Let us see that

$$v_D \notin \overline{H^D(\zeta,\omega)}. \tag{7.2}$$

By definition of Λ^*, $\zeta(v_D) - \Lambda(\zeta) \le \Lambda^*(v_D)$, so for all $\omega > 0$, $\zeta(v_D) < \Lambda^*(v_D) + \Lambda(\zeta) + \omega$. Since ζ is continuous and the right-hand side of the last equation above is just a constant, there is an open neighborhood of v_D, say U, such that for every $u \in U$, $\zeta(u) < \Lambda^*(v_D) + \Lambda(\zeta) + \omega$. This means $U \subseteq B\backslash H^D(\zeta,\omega)$, so (7.2) holds. If $H^D(\zeta,\omega) \cap \mathbb{D}_{\Lambda^*} = \varnothing$, then the claim (4.1) is satisfied trivially. But, if $H^D(\zeta,\omega) \cap \mathbb{D}_{\Lambda^*} \neq \varnothing$, then Hypothesis 3 is satisfied and $H^D(\zeta,\omega)$ has a dominating point $v_{H^D(\zeta,\omega)}$. Summarizing, we have $v_{H^D(\zeta,\omega)} \in \overline{H^D(\zeta,\omega)}$ so $v_{H^D(\zeta,\omega)} \neq v_D$ by (7.2). But again $v_{H^D(\zeta,\omega)} \in \overline{H^D(\zeta,\omega)} \subseteq \overline{D}$. Because of the uniqueness of the dominating point v_D, we conclude $\Lambda^*(v_D) < \Lambda^*(v_{H^D(\zeta,\omega)})$, and as we claimed

$$\Lambda^*(\overline{D}) = \Lambda^*(v_D) < \Lambda^*(v_{H^D(\zeta,\omega)}) = \Lambda^*(\overline{H^D(\zeta,\omega)}).$$

Note that by (7.1), $D^{(\epsilon)} \subseteq \bigcup_{\{\omega>0\}} \bigcup_{\{\zeta\in B^*\}} H(\zeta,\omega)$. But even more is true: Not only $D^{(\epsilon)}$ but its closure is contained in that union because

$$\overline{D^{(\epsilon)}} \subseteq \overline{D} \cap \{\|x - v_D\| \ge \epsilon\} \subseteq \overline{D}\backslash\{v_D\}.$$

Now, by (7.1),

$$\overline{D^{(\epsilon)}} \subseteq \overline{D}\backslash\{v_D\} \subseteq \bigcup_{\{\zeta\in B^*\}} \bigcup_{\{\omega>0\}} H(\zeta,\omega). \tag{7.3}$$

Proof of Lemma 3. Formula (4.4) follows from

$$\begin{aligned} D^{(\epsilon)} \cap K_a \subseteq \overline{D^{(\epsilon)}} \cap K_a &\subseteq \bigcup_{i=1}^{m} H_i \\ D^{(\epsilon)} \cap K_a = D^{(\epsilon)} \cap K_a \cap D &\subseteq \left(\bigcup_{i=1}^{m} H_i\right) \cap D \\ &= \bigcup_{i=1}^{m} (H_i \cap D) = \bigcup_{i=1}^{m} \left(H^D(\zeta_i,\omega_i)\right). \end{aligned}$$

Proof of Lemma 4. First note that $f(X)$ is bounded because g and u are bounded. It is not concentrated on a point because of Hypothesis 1′. Also, since D does not contain the mean of $u(X)$, the set $B_1 \times D$ does not contain the mean of $f(X)$, and that is enough to have an equivalent to Hypothesis 1. (Another way to deal with this part is to assume the law of $g(X)$ is centered in Hypothesis 1′.) Exactly as in (5.1), for every $t > 0$, $\exp^{t\|f(x)\|}$ is bounded (uniformly on S) so

$$\int_B \exp^{t\|x\|} d\mu^f(x) = \int_S \exp^{t\|f(x)\|} d\mu(x) < \infty, \tag{7.4}$$

and Hypothesis 2 holds for f. D is open and convex in B_2, so $(B_1 \times D)$ is open and convex in B; $0 \notin (B_1 \times D)$ because $0 \notin D$. To satisfy Hypothesis 3 it only remains to see $(B_1 \times D) \cap \mathbb{D}_{\Lambda^*} \neq \varnothing$.

Let $\pi : B_1 \times B_2 \to B_2$ be the standard projection $\pi(x_1, x_2) = x_2$. It is continuous, and $\pi(f(x)) = u(x)$. Then the contraction principle (as in [5], p. 110) and Hypothesis 3′ imply

$$\begin{aligned} \Lambda^*(B_1 \times D) &= \inf_{\{d \in D\}} \Lambda^*(B_1 \times \{d\}) \\ &= \inf_{\{d \in D\}} \inf_{\{x : \pi(x) = d\}} \Lambda^*(x) \\ &= \inf_{\{d \in D\}} {\Lambda^*}_u(d) = {\Lambda^*}_u(D) < \infty, \end{aligned} \tag{7.5}$$

where the contraction principle allows us to get from Λ^* to ${\Lambda^*}_u$.

Now (7.5) says Hypothesis 3 holds, and we are done.

Proof of Lemma 5. To prove this Lemma we shall prove something which easily (using (5.3)) implies (5.6):

$$v_{B_1 \times D} = \int_S f(x) d\mu_{(0,\zeta_D)}(x). \tag{7.6}$$

Call $x^* = \int_S f(x) d\mu_{(0,\zeta_D)}(x)$. We will see that the right-hand side above, x^*, satisfies the properties of the dominating point for $(B_1 \times D, \Lambda^*)$ and then we shall invoke its uniqueness. Let us see that $x^* \in \partial(B_1 \times D)$, or equivalently, $(x^*)_2 \in \partial D$. But we know even more: $(x^*)_2 = v_D$ because

$$\begin{aligned} (x^*)_2 &= \left(\int_S f(x) d\mu_{(0,\zeta_D)}(x)\right)_2 \\ &= \left(\int_S f(x) \exp\{(0,\zeta_D)(f(x)) - \Lambda((0,\zeta_D))\} d\mu(x)\right)_2 \\ &= \left(\int_S f(x) \exp\{\zeta_D(u(x)) - \Lambda_u(\zeta_D)\} d\mu(x)\right)_2 \\ &= \int_S u(x) \exp\{\zeta_D(u(x)) - \Lambda_u(\zeta_D)\} d\mu(x) \\ &= \int_S u(x) d\mu_{\zeta_D}(x) = v_D. \end{aligned} \tag{7.7}$$

We used (5.2), and the last equality above is (2.3) of Theorem 1. By the definition of dominating point $v_D \in \partial D$, so $x^* \in \partial(B_1 \times D)$.

Now we want to prove that Λ^* achieves its infimum $\Lambda^*(B_1 \times D)$ at x^*. Thus far we know $x^* \in (\overline{B_1 \times D})$ so $\Lambda^*(x^*) \geq \Lambda^*(v_{B_1 \times D})$. It remains to prove the opposite inequality. By Theorem 5.2 in Donsker and Varadhan ([8]),

$$\Lambda^*(y) = \inf \left\{ D(\nu\|\mu) \quad | \quad \nu \text{ is a probability in } (S,\mathcal{S}); \right.$$
$$\left. \text{and } \int_S f(x) d\nu(x) = y \right\}. \tag{7.8}$$

In the special case of ν being a twisted measure $\nu = \mu_\zeta$, its entropy is easy to calculate

$$\begin{aligned} D(\mu_\zeta\|\mu) &= \int_S \log\left(\frac{d\mu_\zeta}{d\mu}\right) d\mu_\zeta \\ &= \int_S (\zeta(f(x)) - \Lambda_f(\zeta))\, d\mu_\zeta(x). \end{aligned} \tag{7.9}$$

Apply (7.8) to the definition of x^* with the particular $\zeta = (0, \zeta_D)$ in (7.9) to obtain

$$\Lambda^*\left(\int_S f d\mu_{(0,\zeta_D)}\right) \le \int_S (\zeta_D(u(x)) - \Lambda_u(\zeta_D))\, d\mu_{\zeta_D}(x). \tag{7.10}$$

We are almost done. We will identify the right-hand side of the equation above with ${\Lambda^*}_u(v_D)$. Assuming this last step to be true we are done because the contraction principle assured us in (7.5) that ${\Lambda^*}_u(v_D) = \Lambda^*(B_1 \times D)$.

To check the last step combine (2.2) and (2.3) in Theorem 1 to get

$$\begin{aligned} {\Lambda^*}_u(v_D) &= \zeta_D(v_D) - \Lambda_u(\zeta_D) \\ &= \zeta_D\left(\int_{B_2} x \exp\{\zeta_D(x) - \Lambda_u(\zeta_D)\} d\mu^u(x)\right) - \Lambda_u(\zeta_D) \\ &= \int_S \zeta_D(u(x)) \exp\{\zeta_D(u(x)) - \Lambda_u(\zeta_D)\} d\mu(x) - \Lambda_u(\zeta_D) \\ &= \int_S (\zeta_D(u(x)) - \Lambda_u(\zeta_D))\, d\mu_{\zeta_D}(x). \end{aligned} \tag{7.11}$$

Finally we note that (7.11) above is the right-hand side of (7.10), so we are done.

Acknowledgments

The author is deeply indebted to Peter Ney and Jim Kuelbs for very helpful discussions on the subject, and thanks the referee for her clarifying suggestions.

References

1. De Acosta, A. (1985). *On large deviations of sums of independent random vectors.* In *Probability in Banach Spaces V*. Lecture Notes in Math. Springer. (1153) 1–14. MR **87f:**60035
2. Van Campenhout, J. M., and Cover, T. M. (1981). Maximum entropy and conditional probability. *IEEE Trans. Inform. Theory.* (IT–27) 483–489. MR **83h:**94008
3. Csiszár, I. (1984). Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Probab.* (12) 768–793. MR **86h:**60065
4. Dembo, A., and Kuelbs, J. (1998). Refined Gibbs conditioning principle for certain infinite dimensional statistics. *Studia Sci. Math. Hungar.* (34) 107–126. MR **2000b:**60068
5. Dembo, A., and Zeitouni, O. (1998). *Large Deviation Techniques and Applications.* 2nd ed. Springer, New York. MR **99d:**60030
6. Dembo, A., and Zeitouni, O. (1996). Refinements of the Gibbs conditioning principle. *Probab. Theory Related Fields.* (104) 1–14. MR **97k:**60078
7. Dinwoodie, I. H. (1992). Mesures dominantes et Théorème de Sanov. *Ann. Inst. H. Poincaré Probab. Statist.* (28) 365–373. MR **93h:**60038
8. Donsker, M. D., and Varadhan, S. R. S. (1976). Asymptotic evaluation of certain Markov processes expectations for large time III. *Comm. Pure Appl. Math.* (29) 389–461. MR **55:**1492

9. Einmahl, U., and Kuelbs, J. (1996). Dominating points and large deviations for random vectors. *Probab. Theory Related Fields.* (105) 529–543. MR **97k:**60008
10. Iscoe, I., Ney, P., and Nummelin, E. (1985). Large deviations for uniformly recurrent Markov additive processes. *Adv. Appl. Math.* (6) 373–412. MR **88b:**60077
11. Kuelbs, J. (2000). Large deviation probabilities and dominating points for open, convex sets: non-logarithmic behavior. *Ann. Probab.* (28) 1259–1279. MR **2001k:**60003
12. Lehtonen, T., and Nummelin, E. (1988). On the convergence of empirical distributions under partial observations. *Ann. Acad. Sci. Fenn. Math. Ser. A.I.* (13) 219–223. MR **90e:**60035
13. Lehtonen, T., and Nummelin, E. (1990). Level I theory of large deviations in the ideal gas. *Int'l J. Theor. Phys.* (29) 621–635. MR **91f:**82033
14. Meda, A. (1998). *Conditional Laws and Dominating Points.* University of Wisconsin Ph.D. Thesis.
15. Meda, A., and Ney, P. (1998). A conditioned law of large numbers for Markov additive chains. *Studia Sci. Math. Hungar.* (34) 305–316.
16. Meda, A., and Ney, P. (1999). The Gibbs conditioning principle for Markov chains, in *Perplexing Problems in Probability.* M. Bramson and R. Durret, Ed., Ser. Progress in Probability. Birkhäuser, Boston. (44) 385-398. MR **2001i:**60037
17. Ney, P. (1983). Dominating points and the asymptotics of large deviations on $\mathbb{R}^d$. *Ann. Probab.* (11) 158–167. MR **85b:**60026
18. Ney, P. (1984). Convexity and large deviations. *Ann. Probab.* (12) 903–906. MR **85j:**60046
19. Nummelin, E. (1987). A conditional weak law of large numbers. In *Proc. of Seminar on Stability Problems for Stoch. Models Suhumi, USSR.* Lecture Notes in Math. (1142) 259–262. MR **91f:**60057

Departmento de Matemáticas, Cub. 132, Facultad de Ciencias, UNAM, Circuito Exterior s/n, Ciudad Universitaria, Coyoacán 04510, México D. F., México

E-mail address: amg@hp.fciencias.unam.mx