Stochastic Processes and their Applications 25 (1987) 95-108
North-Holland

MARKOV PROCESS LARGE DEVIATIONS IN τ-TOPOLOGY

Erwin BOLTHAUSEN

Technische Universität Berlin, Fachbereich Mathematik, D-1000 Berlin 12, FR Germany

Received 18 February 1986
Revised 27 February 1987

The results of Donsker and Varadhan on the probability of large deviations for empirical measures (or occupation measures) of uniformly ergodic Markov processes are extended. Usually the large deviation results are formulated in the weak topology on the set of probability measures. We extend this to the topology which is generated by the integrals over bounded measurable functions.

AMS Subject Classifications: 60F10, 60J05, 60J10.

Markov process * large deviations * empirical measures * τ-topology

1. Introduction and notations

In recent years there has been a growing interest in large deviations for empirical measures, especially for Markov processes (see e.g. [2, 3, 6]). Usually, the results are stated for the weak topology on the space of probability measures and Feller properties are assumed. It is shown here that parts of the results can be extended by taking the finer τ-topology which is generated by the integrals over bounded measurable functions and no Feller properties are needed. In the i.i.d. cases this has been done by Groeneboom, Oosterhoff and Ruymgaart [5]. The present paper is an extension of their methods to some Markovian situations. The τ-topology appears to be the natural one at least for uniformly ergodic Markov processes. Results for topologies finer than the weak one have also been obtained by Gärtner [4].

We fix some notations:

If $(S, \mathcal{S})$ is an arbitrary measurable space, we write $\Delta(S)$ or Δ_S for the set of probability measures on S. The τ-topology on $\Delta(S)$ is induced by the mappings $\Delta(S) \ni \mu \mapsto \mu(A), A \in \mathcal{S}$. It is clear that, for any $f \in b\mathcal{S}$ (the set of bounded measurable functions), $\int f \, d\mu$ is τ-continuous in μ. These mappings also generate a σ-field $\mathcal{D}_S$. In general, $\mathcal{D}_S$ is not the Borel-field of $\Delta(S)$. If, however, S is a Polish space then

Part of this work has been done during a stay at the "Forschungsinstitut für Mathematik" in Zürich. I would like to thank the members of the institute and especially Hans Föller for the kind hospitality.

$\mathscr{D}_S$ is the Borel-field of the weak topology on $\Delta(S)$. If μ, $\nu \in \Delta(S)$ then the relative entropy of ν with respect to μ is defined by

$$h(\nu|\mu) = \begin{cases} \int \left(\log \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\nu & \text{if } \nu \ll \mu, \\ \infty & \text{otherwise.} \end{cases}$$

Let E be a fixed Polish space, the state space of the Markov chain, with Borel-field $\mathscr{E}$ and let P be a Markov transition kernel on $(E, \mathscr{E})$. We write P^n for the n-th iterate of P. Let $(E', \mathscr{E}')$ be a second Polish space, k a fixed natural number ≥ 2 and K be a Markov kernel from E^k to E'. We write $\Omega = E^{\mathbb{N}_0} \times E'^{\mathbb{N}}$ with the product σ-field $\mathscr{A}$ ($\mathbb{N}_0 = \mathbb{N} \cup \{0\}$). X_n, $n \in \mathbb{N}_0$, are the projections $\Omega \to E$ and ξ_n, $n \in \mathbb{N}$, the projections $\Omega \to E'$. We also write $\tilde{E} = E^k \times E'$ with product σ-field $\tilde{\mathscr{E}}$. If $\nu \in \Delta(E)$ then we define the probability measures $\mathbb{P}_\nu$ on $(\Omega, \mathscr{A})$ by

$$\begin{aligned} &\mathbb{P}_\nu(A_0 \times A_1 \times \cdots \times A_n \times E \times \cdots \times B_1 \times \cdots \times B_{n-k+2} \times E' \times \cdots) \\ &= \int \nu(\mathrm{d}x_0) P(x_0, \mathrm{d}x_1) \cdots P(x_{n-1}, \mathrm{d}x_n) K((x_0, \ldots, x_{k-1}), \mathrm{d}y_1) \\ &\quad \times K((x_1, \ldots, x_k), \mathrm{d}y_2) \cdots K((x_{n-k+1}, \ldots, x_n), \mathrm{d}y_{n-k+2}) \\ &\quad \times 1_{A_0}(x_0) 1_{A_1}(x_1) \cdots 1_{A_n}(x_n) 1_{B_1}(y_1) \cdots 1_{B_{n-k-2}}(y_{n-k+2}) \end{aligned} \tag{1.1}$$

$A_i \in \mathscr{E}$, $B_i \in \mathscr{E}'$. We write $\mathbb{P}_x$ for $\mathbb{P}_{\delta_x}$. $X_0, X_1, X_2, \ldots$ is the usual Markov chain with transition kernel P and starting measure ν. Of course, $\xi_1, \xi_2, \ldots$ is in general not Markovian but $\tilde{X}_n = (X_{n-k+1}, \ldots, X_n, \xi_{n-k+2})$ is a Markov chain with state space $\tilde{E}$ and transition kernel $\tilde{P}$ given by

$$\begin{aligned} &\tilde{P}((x_1, \ldots, x_k, \alpha), A_1 \times \cdots \times A_k \times B) \\ &= 1_{A_1 \times \cdots \times A_{k-1}}(x_2, \ldots, x_k) \int_{A_k} P(x_k, \mathrm{d}y) K((x_2, x_3, \ldots, x_k, y), B). \end{aligned} \tag{1.2}$$

Our main assumption is the following recurrence condition:

(1.3) There is a stationary probability measure π on $(E, \mathscr{E})$ for P such that P has a transition density p with respect to π which is bounded and bounded away from 0.

Let $L_n : \Omega \to \Delta(E')$ be the empirical measure of the ξ-chain, i.e.

$$L_n(\omega) = \frac{1}{n} \sum_{j=1}^{n} \delta_{\xi_j(\omega)}.$$

L_n is clearly $\mathscr{A} - \mathscr{D}_{E'}$-measurable. We describe now the correct Donsker–Varadhan entropy which governs the large deviation behavior of L_n. First some notation.

If $\mu \in \Delta(\tilde{E})$ then we write $\bar{\mu}$ for the marginal measure on E^k, $\mu_1 \in \Delta(E^{k-1})$ for the marginal of $\bar{\mu}$ on the first $k-1$ components and $\mu_2 \in \Delta(E^{k-1})$ for the marginal on the last $k-1$ components. We also write $\Delta_0(\tilde{E})=\{\mu\in\Delta(\tilde{E}): \mu_1=\mu_2\}$. If $k\geq 2$ then the Donsker–Varadhan entropy function $\tilde{I}:\Delta(\tilde{E})\to[0,\infty]$ is defined by

$$\tilde{I}(\mu)=\begin{cases} h(\mu\,|\,\mu_1\otimes P\otimes K) & \text{if } \mu\in\Delta_0(\tilde{E}),\\ \infty & \text{otherwise.}\end{cases}$$

Here

$$(\mu_1\otimes P\otimes K)(A\times B\times C)=\int_{A\times B}\mu_1(\mathrm{d}x)P(x,\mathrm{d}y)K((x,y),C).$$

If $\mu\in\Delta(E')$ then we define

$$I(\mu)=\inf\{\tilde{I}(\nu): \nu_3=\mu\}, \text{ where } \nu_3 \text{ is the marginal on } E'.$$

If $A\subset\Delta(E')$ then $I(A)=\inf\{I(\mu): \mu\in A\}$.

(1.4) Remark. We always assume $k\geq 2$. This is no restriction because a Markov kernel K from E to E' can trivially be extended to a kernel from $E\times E$ to E', just by letting it be independent of the first factor.

Our main result is the following.

(1.5) Theorem. *If* (1.3) *is satisfied and A is a measurable subset of* $\Delta(E')$ *then, for all* $\nu\in\Delta(E)$,

(i) $\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\nu(L_n\in A)\geq -I(\mathrm{int}_\tau A)$,

(ii) $\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\nu(L_n\in A)\leq -I(\mathrm{cl}_\tau A)$

This theorem especially applies to the situation where $k=2$, $E=E'$ and $K((x,y),A)=1_A(y)$. This gives a τ-topology version of the discrete time results of [2]. Indeed, in this case $I(\mu)=\inf\{h(\nu\,|\,\mu_1\otimes P): \nu\in\Delta(E^2): \nu_1=\nu_2=\mu\}$ which via theorem (2.1) of [3] is precisely the Donsker–Varadhan entropy as it has been defined in [2]:

$$I(\mu)=-\inf_u\int\log\frac{Pu}{u}\,\mathrm{d}\mu$$

where the infinimum is taken over all measurable positive functions with $\log u\in b\mathscr{E}$. The theorem also applies to continuous time processes:

Let $\hat{X}_t$, $t\geq 0$, be a time homogenous Markov process which is defined on a probability spaces $(\hat{\Omega},\hat{\mathscr{A}},\hat{\mathbb{P}}_x)$, $x\in E$, and which has state space E starting measure δ_x and transition kernels $\hat{P}_t$, $t\geq 0$, i.e. $\hat{\mathbb{P}}_x(\hat{X}_0\in A)=1_A(x)$ and $\hat{P}_t(x,A)$ is a regular conditional probability for $\hat{\mathbb{P}}(\hat{X}_{t+s}\in A\,|\,\hat{X}_s=x)$ for all $s,t\geq 0$. Furthermore, we assume

(1.6) $\hat{X}$ has paths in $D[0,\infty)$ the space of functions $[0,\infty)\to E$ which are right continuous and have left limits.

(1.7) $\hat{P}_1$ satisfies (1.3).

We now take $E' = D[0, 1]$, $P = \hat{P}_1$, $k = 2$ and $K((x, y), A)$ a regular version of the conditional distribution $\hat{\mathbb{P}}_x(\hat{X}|_{[0,1]} \in A \mid \hat{X}_1 = y)$. Here $\hat{X}|_{[0,1]}$ is the $\hat{X}$ path restricted on the time interval $[0, 1]$.

Let $\Psi : \Delta(D[0, 1]) \to \Delta(E)$ be defined by

$$\Psi(\mu)(A) = \int_{D[0,1]} \int_0^1 1_A(f(s))\, \mathrm{d}s\, \mu(\mathrm{d}f).$$

Ψ is measurable and τ-continuous.

Constructing $\mathbb{P}_x$ and L_n as above, we have that the $\mathbb{P}_x$-distribution of $\Psi(L_n)$ is the same as the $\hat{\mathbb{P}}_x$-distribution of

$$L_n = \frac{1}{n} \int_0^n \delta_{\hat{X}_s}\, \mathrm{d}s.$$

Therefore, we immediately obtain

(1.8) Theorem. *Under the conditions stated above one has, for $x \in E$ and measurable subsets $A \subseteq \Delta(E)$,*

(i) $\liminf_{n\to\infty} \frac{1}{n} \log \hat{\mathbb{P}}_x(L_n \in A) \geq -\hat{I}(\mathrm{int}_\tau A)$,

(ii) $\limsup_{n\to\infty} \frac{1}{n} \log \hat{\mathbb{P}}_x(L_n \in A) \leq -\hat{I}(\mathrm{cl}_\tau A)$.

Here $\hat{I}(B) = \inf\{\hat{I}(\mu) \colon \mu \in B\}$, $\hat{I}(\mu) = \inf\{I(\nu) \colon \Psi(\nu) = \mu\}$.

(1.9) Remark. In [2] Donsker and Varadhan proved similar statements assuming a Feller property where int A and cl A have to be taken in the weak topology. Their entropy functional is given by

$$I_{\mathrm{DV}}(\mu) = -\inf_u \int \frac{Lu}{u}\, \mathrm{d}\mu.$$

Here L is the infinitesimal generator and the infimum runs over nonnegative functions in the domain of the generator which are bounded and bounded away from 0. But then it is clear that $I_{\mathrm{DV}} = \hat{I}$. So (1.8) gives a τ-topology version of their results. This especially applies to diffusions on compacta.

Sometimes it is useful to have the statements of (1.5) (or (1.8)) for random elements in $\Delta(E')$ which differ slightly from L_n. So let L'_n be measurable mappings $\Omega \to \Delta(E')$. The following result will be useful in a forthcoming paper on maximum entropy principles.

(1.10) Proposition. *If, for some sequence of positive real numbers $\varepsilon_n \to 0$, $\| L_n(\omega) - L'_n(\omega) \|_v \leq \varepsilon_n$ for all ω where $\| \quad \|_v$ denotes the total variation norm then the statements* (i) *and* (ii) *of* (1.5) *remain true for L'_n.*

The proof of (1.10) is straightforward and omitted.

We will give the proof of theorm (1.5) for $k=2$. This simplifies somewhat the notation. The extension to cover the case of arbitrary k is straightforward.

2. Properties of the entropy

Let $\tilde{\mathscr{V}}=\{e^g: g\in b\tilde{\mathscr{E}}\}$.

(2.1) Lemma. *If $\mu\in\Delta_0(\tilde{E})$ then $\tilde{I}(\mu)=-\inf_{u\in\tilde{\mathscr{V}}}\int\log\tilde{P}u/u\,\mathrm{d}\mu$.*

Proof. From Theorem (2.1) of [3], we have

$$-\inf_{u\in\tilde{\mathscr{V}}}\int\log\frac{\tilde{P}u}{u}\,\mathrm{d}\mu=\inf\{h(Q\,|\,\mu\otimes\tilde{P}):\ Q\in\Delta(\tilde{E}^2),\ Q_1=Q_2=\mu\}\tag{2.2}$$

(Q_i are the two marginals of Q on $\tilde{E}$).

Let $\mu\in\Delta(\tilde{E})$ with $\tilde{I}(\mu)<\infty$ (especially $\mu\in\Delta_0(\tilde{E})$ and $\mu\ll\mu_1\otimes P\otimes K$) we construct $Q^*\in\Delta(\tilde{E}^2)$ minimizing the right side of (2.2).

Let $\mu(A\,|\,p_1=x)$ be a regular conditional distribution for μ conditional on the projection p_1 on the first factor E of $\tilde{E}$. We define $Q^*\in\Delta(\tilde{E}^2)$ by

$$Q^*(A\times B)=\int_A\mu(\mathrm{d}\alpha)\mu(B\,|\,p_1=p_2(\alpha)),$$

where p_2 is the projection on the second factor. Q^* is well defined as $\mu\in\Delta_0$. $Q^*\ll\mu\otimes\tilde{P}$, $Q_1^*=Q_2^*=\mu$ and $\rho(x,y)=(\mathrm{d}\mu/\mathrm{d}(\mu_1\otimes P\otimes K))(y)$ is a version of $(\mathrm{d}Q^*/\mathrm{d}(\mu\otimes\tilde{P}))(x,y)$. By Theorem (3.1) of [1], Q^* minimizes the right side of (2.2). But

$$h(Q^*\,|\,\mu\otimes\tilde{P})=h(\mu\,|\,\mu_1\otimes P\otimes K).$$

It remains to show that if the right side of (2.2) is finite then $\tilde{I}(\mu)<\infty$. Indeed, in this case, there exists $Q\in\Delta(\tilde{E}^2)$ with $Q_1=Q_2=\mu$ and $h(Q\,|\,\mu\otimes\tilde{P})<\infty$. Therefore, $Q\ll\mu\otimes\tilde{P}$ and this implies $\mu_1=\mu_2$. The second marginal of $\mu\otimes\tilde{P}$ on $\tilde{E}$ is then just $\mu_1\otimes P\otimes K$. Therefore, $\infty>h(Q\,|\,\mu\otimes\tilde{P})\geq h(\mu\,|\,\mu_1\otimes P\otimes K)$. So the lemma is proved. □

(2.3) Corollary. *$\tilde{I}:\Delta(\tilde{E})\to[0,\infty]$ is convex and τ-lower semicontinuous.* □

(2.4) Lemma. *For any positive real number r, $\{\mu\in\Delta(\tilde{E}):\ \tilde{I}(\mu)\leq r\}$ is τ-compact.*

Proof. Let $\tilde{P}^2$ be the iterate of $\tilde{P}$ and, for $\mu\in\Delta(\tilde{E})$,

$$\begin{aligned}\tilde{I}^{(2)}(\mu)&=-\inf_{u\in\tilde{\mathscr{V}}}\int\log\frac{\tilde{P}^2u}{u}\,\mathrm{d}\mu\\&=-\inf_{u\in\tilde{\mathscr{V}}}\left[\int\log\frac{\tilde{P}^2u}{\tilde{P}u}\,\mathrm{d}\mu+\int\log\frac{\tilde{P}u}{u}\,\mathrm{d}\mu\right]\leq 2\tilde{I}(\mu).\end{aligned}$$

The stationary measure for $\tilde{P}$ is $\tilde{\pi} = \pi \otimes P \otimes K$ and $\tilde{P}^2$ has a transition density $\tilde{p}^{(2)}$ with respect to $\tilde{\pi}$ which is bounded and bounded away from 0, say

$$c \leqslant \tilde{p}^{(2)}(x, y) \leqslant 1/c, \quad c > 0, \qquad x, y \in \tilde{E}.$$

So

$$\begin{aligned} I^{(2)}(\mu) &= \inf\{h(Q \mid \mu \otimes \tilde{P}^2): Q \in \Delta(\tilde{E}^2), Q_1 = Q_2 = \mu\} \\ &\geqslant \log c + \inf\{h(Q\mu \otimes \tilde{\pi}): Q \in \Delta(\tilde{E}^2), Q_1 = Q_2 = \mu\} \\ &= \log c + h(\mu \mid \tilde{\pi}). \end{aligned}$$

Therefore, $\{\mu: \tilde{I}(\mu) \leqslant r\} \subset \{\mu: h(\mu \mid \tilde{\pi}) \leqslant 2r - \log c\}$. By Lemma 2.3 of [5] the right side is τ-compact. From (2.3) one sees that $\{\mu: \tilde{I}(\mu) \leqslant r\}$ is τ-closed. So it is compact. □

Now let $f \in b\mathscr{E}'$ ($f: E' \to \mathbb{R}$). From Corollary (2.3) and (2.4) one immediately sees that there exists $\mu^* \in \Delta_0(\tilde{E})$ with

$$\int f \, \mathrm{d}\mu_3^* - \tilde{I}(\mu^*) = \sup\left\{\int f \, \mathrm{d}\mu_3 - \tilde{I}(\mu): \mu \in \Delta(\tilde{E})\right\}. \tag{2.5}$$

Our next task is to give some information about the structure of μ^*:

(2.6) Proposition. *For any* μ^* *satisfying* (2.5) *there exists* $g \in b\mathscr{E}$ *with*

$$\mathrm{d}\mu^*/\mathrm{d}(\mu_1^* \otimes P \otimes K)(x_1, x_2, \alpha) = \text{const.} \exp(f(\alpha) + g(x_1) - g(x_2)) \quad \text{a.s.}$$

$((x_1, x_2, \alpha) \in E \times E \times E' = \tilde{E})$.

We need some preparations.

(2.7) Lemma. μ_1^* *is equivalent to* π.

Proof. Ley $\mu^\varepsilon = (1-\varepsilon)\mu^* + \varepsilon(\pi \otimes P \otimes K)$ ($\varepsilon \in [0, 1]$). Then $\mu^\varepsilon \ll \mu_1^\varepsilon \otimes P \otimes K$ and $q^\varepsilon = \mathrm{d}\mu^\varepsilon/\mathrm{d}(\mu_1^\varepsilon \otimes P \otimes K)$ is positive for $\varepsilon > 0$. An easy computation shows that $\lambda(\varepsilon) = h(\mu^\varepsilon \mid \mu_1^\varepsilon \otimes P \otimes K)$ is differentiable on $(0, 1)$ and

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \lambda(\varepsilon) &= \frac{1}{1-\varepsilon}\left[\int \log q_\varepsilon d(\pi \otimes P \otimes K) - h(\mu^\varepsilon \mid \mu_1^\varepsilon \otimes P \otimes K)\right] \\ &\leqslant \frac{1}{1-\varepsilon} \int \log q_\varepsilon d(\pi \otimes P \otimes K). \end{aligned} \tag{2.8}$$

We first show that

$$\limsup_{\varepsilon \downarrow 0} \int \log^+ q_\varepsilon \mathrm{d}(\pi \otimes P \otimes K) < \infty. \tag{2.9}$$

To see this, let $\Lambda_\varepsilon(x)=\{(y,\alpha)\in E\times E': q_\varepsilon(x,y,\alpha)\geq 1\}$, $x\in E$. Then

$$\int \log^+ q^\varepsilon \,\mathrm{d}(\pi\otimes P\otimes K)$$

$$=\int_E \pi(\mathrm{d}x)(P\otimes K)(x,\Lambda_\varepsilon(x))$$

$$\times\int_{\Lambda_\varepsilon(x)} \frac{P(x,\mathrm{d}y)K((x,y),\mathrm{d}\alpha)}{(P\otimes K)(x,\Lambda_\varepsilon(x))}\log^+ q^\varepsilon(x,y,\alpha)$$

$$\leq \int_E \pi(\mathrm{d}x)(P\otimes K)(x,\Lambda_\varepsilon(x))\log^+\left(\int \frac{P(x,\mathrm{d}y)K((x,y),\mathrm{d}\alpha)q^\varepsilon(x,y,\alpha)}{(P\otimes K)(x,\Lambda_\varepsilon(x))}\right).$$

But $\int (P\otimes K)(x,\mathrm{d}(y,\alpha))q^\varepsilon(x,y,\alpha)=1$, so

$$\int \log^+ q^\varepsilon \,\mathrm{d}(\pi\otimes P\otimes K)\leq -\int_E \pi(\mathrm{d}x)(P\otimes K)(x,\Lambda_\varepsilon(x))$$

$$\times\log((P\otimes K)(x,\Lambda_\varepsilon(x)))$$

$$\leq 1/\mathrm{e}.$$

So (2.9) is proved.

As $\mu^*\ll\mu_1^*\otimes P\otimes K$ and $\mu_1^*=\mu_2^*$ it follows that $\mu_1^*\ll\pi$. Assume now that $\mu_1^*\not\sim\pi$. Then there exists $A\in\mathscr{E}$ with $\pi(A)>0$, $\mu_1^*(A)=0$. We then have $\mu^*(E\times A\times E')=0$ but $(\mu_1^*\otimes P\otimes K)(E\times A\times E')>0$. It follows that $q^\varepsilon(x,y,\alpha)\to 0$ on a set of positive $\pi\otimes P\otimes K$ measure. Therefore, from Fatou's lemma,

$$\liminf_{\varepsilon\downarrow 0}\int \log^- q^\varepsilon \,\mathrm{d}(\pi\otimes P\otimes K)=\infty.$$

Together with (2.9) and (2.8) this proves that if $\mu_1^*\not\sim\pi$ then

$$\limsup_{\varepsilon\downarrow 0}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\lambda(\varepsilon)=-\infty.$$

But this clearly contradicts the maximality property of μ^*. □

Proof of (2.6). Let $g\in b\tilde{\mathscr{E}}$ satisfy $\int g\,\mathrm{d}\mu^*=0$ and $\int(\varphi\circ p_1-\varphi\circ p_2)g\,\mathrm{d}\mu^*=0$ for all $\varphi\in b\mathscr{E}$ where p_i, $i=1,2$, are the projections $\tilde{E}\to E$. If $|\varepsilon|$ is small enough then $\mathrm{d}\mu^\varepsilon=(1+\varepsilon g)\,\mathrm{d}\mu$ defines an element in $\Delta_0(\tilde{E})$.

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}h(\mu^\varepsilon\,|\,\mu_1^\varepsilon\otimes P\otimes K)|_{\varepsilon=0}=\int\log\frac{\mathrm{d}\mu^*}{\mathrm{d}(\mu_1^*\otimes P\otimes K)}g\,\mathrm{d}\mu^*.$$

Therefore, we see that

$$\int\left(\log\frac{\mathrm{d}\mu^*}{\mathrm{d}(\mu_1^*\otimes P\otimes K)}-f\circ p_3\right)g\,\mathrm{d}\mu^*=0\quad\text{for all such } g.$$

Here p_3 is the projection $\tilde{E} \to E'$. The Hahn-Banach theorem implies that

$$\log \frac{\mathrm{d}\mu^*}{\mathrm{d}(\mu_1^* \otimes P \otimes K)} - f \circ p_3$$

is in the $L_1(\mu^*)$ closure $\mathscr{G}$ of $\{1, \varphi \circ p_1 - \varphi \circ p_2 \colon \varphi \in b\mathscr{E}\}$. Let $\bar{\mu}^*$ be the marginal of μ^* on $E \times E$ and let $\mathscr{G}_0$ be the $L_1(\bar{\mu}^*)$ span of the functions $\varphi(x_1) - \varphi(x_2)$, $\varphi \in b\mathscr{E}$. $\mathscr{G}$ is just the span of 1 and $\mathscr{G}_0 \circ \bar{p}$ where $\bar{p} \colon \tilde{E} \to E \times E$ is the projection.

As $h(\mu^* | \mu_1^* \otimes P \otimes K) < \infty$ we have $h(\bar{\mu}^* | \mu_1^* \otimes P) < \infty$ and so $h(\bar{\mu}^* | \mu_1^* \otimes \pi) < \infty$. But $h(\nu | \mu_1^* \otimes \pi)$ is minimized over ν with $\nu_1 = \nu_2 = \mu_1^*$ by $\nu = \mu_1^* \otimes \mu_1^*$. From Theorem 2.2 of [1] it follows that $\mu_1^* \otimes \mu_1^* \ll \bar{\mu}^*$. On the other hand, from $h(\mu_1^* \otimes \mu_1^* | \mu_1^* \otimes \pi) < \infty$ we conclude that $h((\mu_1^* \otimes \mu_1^*) \otimes K | \mu_1^* \otimes P \otimes K) < \infty$ and so $\mu^* \ll (\mu_1^* \otimes \mu_1^*) \otimes K$ and therefore $\bar{\mu}^* \ll \mu_1^* \otimes \mu_1^*$. Therefore $\mu_1^* \otimes \mu_1^* \sim \bar{\mu}^*$. But by (2.7) $\pi \sim \mu_1^*$ and therefore all thinkable a.e.-notions coincide and the null sets do no longer bother us.

Any element $v \in \mathscr{G}_0$ is an a.e. limit of elements of the form $\varphi(x_1) - \varphi(x_2)$. As a.e. refers also to a product measure, it is easy to see that v is of the form, too. Therefore,

$$\mathrm{d}\mu^* / \mathrm{d}(\mu_1^* \otimes P \otimes K) = \text{const.} \exp(f \circ p_3 + \varphi \circ p_1 - \varphi \circ p_2).$$

By integrating over $P \otimes K$ one sees that φ is bounded. □

Let $g \colon E' \to \mathbb{R}^d$ be a bounded measurable function. If $x \in \mathbb{R}^d$, we write

$$s(x) = \inf\left\{ I(\mu) \colon \mu \in \Delta(E'), \int g \, \mathrm{d}\mu = x \right\}.$$

(2.10) Lemma. *s is convex and lower semicontinuous.*

Proof. Convexity is clear. Let $x_n \to x \in \mathbb{R}^d$ and $\liminf_{n \to \infty} s(x_n) < \infty$. By selecting a subsequence, we may assume that $\lim_{n \to \infty} s(x_n)$ exists and is $< \infty$. By using the semicontinous property of $\tilde{I}$ and the fact that $\{\mu \colon \tilde{I}(\mu) \leq r\}$ is τ-compact for $0 < r < \infty$ there exist $\mu_n \in \Delta(\tilde{E})$ with

$$\tilde{I}(\mu_n) = s(x_n), \qquad \int g \, \mathrm{d}\mu_{n,3} = x_n.$$

We can assume that μ_n converges to a $\mu \in \Delta(\tilde{E})$ and $\int g \, \mathrm{d}\mu_3 = x$. By the semicontinuity of $\tilde{I}$ the lemma follows. □.

Let

$$S_n = \sum_{j=1}^{n} g(\xi_j).$$

As a consequence of (2.6), we obtain the following result.

(2.11) Corollary. *If $z \in \mathbb{R}^d$, then*

$$\lim_{n\to\infty} (1/n) \log \mathbb{E}_\nu(e^{\langle z, S_n\rangle}) = \sup_x(\langle z, x\rangle - s(x)).$$

($\langle \cdot, \cdot\rangle$ is the Euclidean inner product.

Proof. We write

$$f(e') = \langle z, g(e')\rangle, \quad e' \in E'.$$

Let μ^* be an element of $\Delta(\tilde{E})$ which maximizes $\int f \,\mathrm{d}\mu_3^* - \tilde{I}(\mu^*)$. According to (2.6) there exists $\varphi \in b\mathscr{E}$ with

$$\mathrm{d}\mu^*/\mathrm{d}(\mu_1^* \otimes P \otimes K) = C \exp(\langle g\circ p_3, z\rangle + \varphi\circ p_1 - \varphi\circ p_2).$$

Then

$$\begin{aligned}
&\mathbb{E}_\nu(\exp(\langle z, S_n\rangle)) \\
&\quad = \int \nu(\mathrm{d}x_0) P(x_0, \mathrm{d}x_1) K((x_0, x_1), \mathrm{d}\alpha_1) \cdots \\
&\qquad \times P(x_{n-1}, \mathrm{d}x_n) K((x_{n-1}, x_n), \mathrm{d}\alpha_n) \exp\Big(\sum_{j=1}^n \langle z, g(\alpha_j)\rangle\Big) \\
&\quad = C^{-n} \int \nu(\mathrm{d}x_0) P(x_0, \mathrm{d}x_1) \cdots K((x_{n-1}, x_n), \mathrm{d}\alpha_n) \\
&\qquad \times \exp(\varphi(x_n) - \varphi(x_0)) \prod_{j=1}^n [C \exp(\langle z, g(\alpha_j)\rangle - \varphi(x_j) + \varphi(x_{j-1}))].
\end{aligned}$$

Therefore,

$$C^{-n} \,\mathrm{e}^{-2\|\varphi\|_\infty} \leq \mathbb{E}_\nu(\mathrm{e}^{\langle z, S_n\rangle}) \leq C^{-n} \,\mathrm{e}^{2\|\varphi\|_\infty}.$$

Here $\|\varphi\|_\infty$ is the sup-norm of φ.

It follows that

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}_\nu(\mathrm{e}^{\langle z, S_n\rangle}) &= -\log C \\
&= \int \langle z, g\rangle \,\mathrm{d}\mu_3^* - \tilde{I}(\mu^*) \\
&= \sup_x(\langle z, x\rangle - s(x)). \qquad \square
\end{aligned}$$

3. Proof of Theorem (1.5)

Let g, s, S_n be defined as in the last section.

(3.1.) Proposition. *If $A \subset \mathbb{R}^d$ is measurable, then*

(i) $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu(S_n/n \in A) \leq -s(\mathrm{cl}(A))$,

(ii) $\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu(S_n/n \in A) \geq -s(\mathrm{int}(A)$.

Here $s(B) = \inf_{x\in B} s(x)$ and int *and* cl *refer to the interior and closure in the usual topology.*

Proof. (i) Follows from (2.11) and Lemma 1.1 of [4]. If

$$G(z) = \sup(\langle z, x\rangle - s(x))$$

satisfies a differentiability property then we could apply the Lemma 1.2 of [4] to prove (ii) of our proposition. To show that G is differentiable does not appear to be easy. Instead of doing this we establish (ii) directly by using the subadditivity technique which has been introduced by Lanford in the theory of large deviations. This technique has also been used by Stroock in [6].

If U is an open subset of $\mathbb{R}^d$ and $\varepsilon > 0$, we write U^ε for the ε-neighborhood of U and U_ε for $\{x \in U: y \in U$ holds for all y with $|y-x| < \varepsilon\}$. Clearly $(U^\varepsilon)_\varepsilon \supset U$.

Let $c > 0$ be such that $c \leq p(x, y) \leq 1/c$ for all $x, y \in E$. If $A \subset \mathbb{R}^d$ we write $a_n(A) = \log \mathbb{P}_\nu(S_n/n \in A)$. Let now $U \subset \mathbb{R}^d$ be open and convex. If $n, m \in \mathbb{N}$ then

$$\begin{aligned} a_{n+m}(U) &\geq \log \mathbb{P}_\nu(S_{n+m+1}/(n+m) \in U_{(n+m)^{-1}}) \\ &\geq \log \mathbb{P}_\nu\left(\frac{n}{n+m}\frac{1}{n}\sum_{j=1}^{n} f(\xi_j) + \frac{m}{n+m}\frac{1}{m}\sum_{j=n+2}^{n+m+1} f(\xi_j) \in U_{2(n+m)^{-1}}\right) \\ &\geq \log c + a_n(U_{2(n+m)^{-1}}) + a_m(U_{2(n+m)^{-1}}). \end{aligned} \tag{3.2}$$

Replacing U by $U^{2(n+m)^{-1}}$ we have

$$a_n(U) + a_m(U) \leq a_{n+m}(U^{2(n+m)^{-1}}) + \log\left(\frac{1}{c}\right). \tag{3.3}$$

We apply this to $n = m = 2^k$, $k \in \mathbb{N}$, and obtain

$$2^{-k} a_{2^k}(U) \leq 2^{-k-1} a_{2^{k+1}}(U^{2^{-k}}) + 2^{-k-1} \log\left(\frac{1}{c}\right)$$

Iterating this inequality, we obtain

$$\begin{aligned} &2^{-k-l} a_{2^{k+l}}(U^{(2^{-k+1}-2^{-k-l+1})}) - 2^{-k-l} \log\left(\frac{1}{c}\right) \\ &\quad\leq 2^{-k-l-1} a_{2^{k+l+1}}(U^{(2^{-k+1}-2^{-k-l})}) - 2^{-k-l-1} \log\left(\frac{1}{c}\right). \end{aligned}$$

Therefore

$$\lambda_k(U)=\lim_{l\to\infty} 2^{-k-l} a_{2^{k+l}}(U^{(2^{-k+1}-2^{-k-l+1})})$$

exists and furthermore

$$2^{-k-l} a_{2^{k+l}}(U^{(2^{-k+1}-2^{-k-l+1})}) \le \lambda_k(U)+2^{-k-l}\log\left(\frac{1}{c}\right) \tag{3.4}$$

Clearly $\lambda_k(U) \ge \lambda_{k+1}(U)$. So we define

$$\lambda(U)=\lim_k \lambda_k(U)$$

and

$$\bar{s}(x)=-\inf\{\lambda(U)\colon U\ni x,\ U \text{ open convex}\}.$$

$\bar{s}(x)\ge 0$ and $\bar{s}$ is lower semicontinuous by the construction. We claim that $\bar{s}$ is convex.

Let

$$x=\tfrac{1}{2}x_1+\tfrac{1}{2}x_2,\quad x_i\in\mathbb{R}^d$$

and U be an arbitrary convex open neighborhood of x. Given $\varepsilon>0$, we choose a $\delta>0$, convex neighborhoods V_i of x_i and k such that $(\frac{1}{2}V_1^\delta+\frac{1}{2}V_2^\delta)\subset U$, $|\lambda_k(V_i)+\bar{s}(x_i)|<\varepsilon$. If l is large enough then

$$|2^{-k-l}a_{2^{k+l}}(V_i^{(2^{-k+1}-2^{-k-l+1})})+\bar{s}(x_i)|<2\varepsilon.$$

An argument similar to the one leading to (3.3) gives

$$\begin{aligned}2^{-k-l-1}a_{2^{k+l+1}}(U) &\ge 2^{-k-l-1}a_{2^{k+l+1}}(\tfrac{1}{2}V_1^\delta+\tfrac{1}{2}V_2^\delta)\\ &\ge \sum_{i=1}^{2} 2^{-k-l-1}a_{l+k}((V_i^\delta)_{2^{-k-l}})-\frac{\log c}{2^{k+l+1}}.\end{aligned}$$

If k, l are large enough $(V_i^\delta)_{2^{-k-l}}\supset V_i^{(2^{-k+1}-2^{-k-l-1})}$. Letting $l\to\infty$ we conclude that

$$\lambda_k(U)\ge\tfrac{1}{2}\lambda_k(V_1)+\tfrac{1}{2}\lambda_k(V_2)\ge -\tfrac{1}{2}\bar{s}(x_1)-\tfrac{1}{2}\bar{s}(x_2)-4\varepsilon.$$

Therefore $\bar{s}(x)\le\frac{1}{2}\bar{s}(x_1)+\frac{1}{2}\bar{s}(x_2)$.

Let A be open and $x\in A$. We choose $\varepsilon>0$ and an open convex neighborhood U of x in A with $U^{3\varepsilon}\subset A$. For large enough k we have

$$\begin{aligned}\lambda_k(U)&=\lim_{l\to\infty}2^{-k-l}a_{2^{k+l}}(U^{(2^{-k+1}-2^{-k-l+1})})\\ &\le\liminf_{m\to\infty}2^{-m}a_{2^m}(U^\varepsilon).\end{aligned}$$

Therefore

$$-\bar{s}(x)\le\liminf_{m\to\infty}2^{-m}a_{2^m}(U^\varepsilon).$$

If $r\in\mathbb{N}$, let

$$N_r=\left\{\sum_{i=1}^{r}a_i2^{k+i-1}\colon a_i\in\{0,1\},\ k\in\mathbb{N}\right\}.$$

A repeated application of (3.2) yields that, for any $r \in \mathbb{N}$,

$$-\bar{s}(x) \leq \liminf_{\substack{n\to\infty \\ n\in N_r}} n^{-1} a_n(U^{2\varepsilon}).$$

Using the boundedness of g we obtain

$$\liminf_{n\to\infty} n^{-1} a_n(U^{3\varepsilon}) \geq \liminf_{\substack{n\to\infty \\ n\in N_r}} n^{-1} a_n(U^{2\varepsilon}).$$

Therefore,

$$\liminf_{n\to\infty} n^{-1} a_n(A) \geq -\bar{s}(A). \tag{3.5}$$

It remains to prove that $s = \bar{s}$.

Let now A be compact and $\varepsilon > 0$ be given. If $x \in A$ we choose a convex open neighborhood U_x of x with $-\bar{s}(x) \geq \lambda(U_x) - \varepsilon$ if $\bar{s}(x) < \infty$ and $\lambda(U_x) \leq -1/\varepsilon$ if $\bar{s}(x) = \infty$. Then $A \subset \bigcup_{j=1}^m U_{x_j}$ for suitable points $x_1, \ldots, x_m$ and from (3.7)

$$\limsup_m 2^{-m} a_{2^m}(A) \leq \max_{1\leq j\leq m} \lambda_k(U_{x_j}) + 2^{-k-1} \log\left(\frac{1}{c}\right)$$

for all k, so

$$\limsup_m 2^{-m} a_{2^m}(A) \leq \max_{1\leq j\leq m} \lambda(U_{x_j}).$$

Keeping in mind the definition of U_x and letting $\varepsilon \to 0$ one concludes that

(3.6) For A compact,

$$\limsup_m 2^{-m} a_{2^m}(A) \leq -\bar{s}(A).$$

Using (3.5) and (3.6), a standard argument gives, for $z \in \mathbb{R}^d$,

$$\lim_{m\to\infty} 2^{-m} \log \mathbb{E}_\nu(\exp(\langle z, S_{2^m}\rangle)) = \sup_{x\in\mathbb{R}^d}((\langle x, z\rangle - \bar{s}(x)).$$

Together with (2.11) and the inversion formula for Legendre transforms (see e.g. [6, Theorem (7.15)]) this proves $s = \bar{s}$. □

Proof of (1.5). (i) We consider subdivisions of E' into finitely many disjoint measurable sets. The set of such subdivisions is denoted by Σ. Σ is a directed set under the ordering $\gamma > \gamma'$ ($\gamma, \gamma' \in \Sigma$) if an only if γ is a refinement of γ'. If $\gamma \in \Sigma$ and Γ is a measurable subset of $\Delta(E')$ then we set $\Gamma_\gamma = \{\mu$: there exists $\nu \in \Gamma$ with $\nu(A_i) = \mu(A_i)$ for all $A_i \in \gamma\}$. Clearly $\Gamma_\gamma \supset \Gamma$. The set $\{(\nu(A_1), \ldots, \nu(A_d)) : \nu \in \Gamma\}$ where $\{A_1, \ldots, A_d\} = \gamma$ is a subset of $\mathbb{R}^d$. Its closure is denoted by F_γ.

We now apply the proposition (3.2) to $f(x) = (1_{A_1}(x), \ldots, 1_{A_d}(x))$ and obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu(L_n \in \Gamma) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu\left(\frac{1}{n}\sum_{j=1}^n f(\xi_j) \in F_\gamma\right)$$

$$\leq s(F_\gamma) = -I(\mathrm{cl}_\tau \Gamma_\gamma).$$

This holds for all $\gamma \in \Sigma$. Therefore

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu(L_n \in \Gamma) \leqslant -\sup_{\gamma\in\Sigma} I(\mathrm{cl}_\tau \Gamma_\gamma).$$

It remains to show that

$$\sup_{\gamma\in\Sigma} I(\mathrm{cl}_\tau \Gamma_\gamma) = I(\mathrm{cl}_\tau \Gamma).$$

If this were false then there would exist $r < I(\mathrm{cl}\, \Gamma)$ and $I(\mathrm{cl}(\Gamma_\gamma)) \leqslant r$ for all $\gamma \in \Sigma$. To any $\gamma \in \Sigma$ we can choose $\nu_\gamma \in \mathrm{cl}(\Gamma_\gamma)$ with $I(\nu_\gamma) \leqslant r$. By the τ-compactness of $\{\nu: I(\nu) \leqslant r\}$ there exists a cluster point ν of the net $\{\nu_\gamma: \gamma \in \Sigma\}$. By the semi-continuity property of I we have $I(\nu) \leqslant r$. If we show that $\nu \in \mathrm{cl}_\tau(\Gamma)$ then this would lead to a contradiction. If $\nu \notin \mathrm{cl}_\tau(\Gamma)$ then there is a neighborhood $U(\nu)$ of the form $\{\mu: |\mu(A_i) - \nu(A_i)| < \delta,\ i = 1, \ldots, d\}$ where $\gamma = \{A_1, \ldots, A_d\} \in \Sigma$ and $U(\nu) \cap \Gamma = \varnothing$. But we find a $\gamma' > \gamma$ with $\nu_{\gamma'} \in U(\nu)$ and $\nu_{\gamma'} \in \mathrm{cl}(\Gamma_{\gamma'}) \subset \mathrm{cl}(\Gamma_\gamma)$. This leads to $U(\nu) \cap \Gamma \neq \varnothing$, a contradiction.

(ii) The statement is void if $I(\mathrm{int}_\tau \Gamma) = \infty$. We, therefore, assume $I(\mathrm{int}_\tau \Gamma) < \infty$. For given $\varepsilon > 0$ we choose $\nu \in \mathrm{int}_\tau \Gamma$ with $I(\nu) \leqslant I(\mathrm{int}_\tau \Gamma) + \varepsilon$ and then a neighborhood $U(\nu)$ of ν, $U(\nu) \subset \mathrm{int}_\tau \Gamma$, of the form

$$U(\nu) = \{\mu: |\mu(A_i) - \nu(A_i)| < \delta,\ i = 1, \ldots, d\} \quad \text{where} \quad A_i \in \mathscr{E}'.$$

We again put

$$f = (1_{A_1}, \ldots, 1_{A_d}) \quad \text{and} \quad V = \{x \in \mathbb{R}^d: |x_i - \nu(A_i)| < \delta,\ i = 1, \ldots, d\}.$$

From (3.2) (ii) we obtain

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu(L_n \in \Gamma) &\geqslant \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\nu\left(\frac{1}{n}\sum_{j=1}^n f(\xi_j) \in V\right) \\ &\geqslant -h(V) = -I(U(\nu)) \\ &\geqslant -I(\nu) \geqslant -I(\mathrm{int}_\tau \Gamma) - \varepsilon.\end{aligned}$$

As $\varepsilon > 0$ is arbitrary, this proves the theorem. □

Acknowledgement

I would like to thank the referee for suggesting some improvements of the original manuscript.

References

[1] I. Csiszar, I-divergence geometry of probability distributions and minimization problems, Ann. Prob. 3 (1975) 146–158.

[2] M. Donsker and S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time I, Comm. Pure Appl. Math. 28 (1975) 1-47.
[3] M. Donsker and S.R.S. Varadhan, Asymptotic evalution of certain Markov process expectations for large time III, Comm. Pure Appl. Math. 29 (1976) 389-461.
[4] J. Gärtner, On large deviations from the invariant measure, Th. Prob. Appl. 22 (1977) 24-39.
[5] P. Groeneboom, J. Oosterhoff and F.H. Ruymgaart, Large deviation theorems for empirical probability measures. Ann. Prob. 7 (1979) 553-586.
[6] D.W. Stroock, An introduction to the theory of large deviations (Springer Verlag, Berlin, 1984).