The Annals of Probability
1985, Vol. 13, No. 1, 90–96

ON THE LOWER BOUND OF LARGE DEVIATION OF RANDOM WALKS

BY TZUU-SHUH CHIANG

Institute of Mathematics, Academia, Sinica

In this note, we prove for a large class of random walks on R^n that

$$\liminf_{n\to\infty}(1/n)\log P_x(L_n(\omega, \cdot) \in N) \geq -I(\mu)$$

where $L_n(\omega, \cdot)$ is the occupation measure, N is a weak neighborhood of μ and $I(\mu)$ is the usual Donsker–Varadhan functional. This generalizes a previous theorem of the author where the state space is assumed to be compact.

Section 1. Let $X_0, X_1, \cdots$ be a Markov process with state space $\mathscr{X}$ and Feller transition $p(x, dy)$. For each ω in the sample space, let $L_n(\omega, \cdot)$ be the average occupation measure, i.e., $\forall A \subseteq \mathscr{X}$, $L_n(\omega, A) = (1/n)\sum_{i=0}^{n-1}\chi_A(X_i(\omega))$. Let $\mathscr{M}$ be the space of all probability measures on $\mathscr{X}$ and $I(\mu)$ be the usual entropy functional defined as follows:

$$\forall\mu \in \mathscr{M},\quad I(\mu) = -\inf_{f\in\mathscr{B}(\mathscr{X})}\int \log\frac{(pf)(x)}{f(x)}\,\mu(dx) \tag{1.0}$$

where $\mathscr{B}(\mathscr{X})$ is the set of all positive continuous functions for each of which there exist a, b s.t. $0 < a \leq f(y) \leq b < \infty$ $\forall y \in \mathscr{X}$ and $(pf)(x) = \int f(y)p(x, dy)$.

In this note, we shall be concerned with the following type of estimate:

For any $\mu \in \mathscr{M}$, $x \in \mathscr{X}$ and weakly open neighborhood G of μ,

$$\liminf_{n\to\infty}(1/n)\log P_x(\omega: L_n(\omega, \cdot) \in G) \geq -I(\mu). \tag{1.1}$$

The I-functional was first introduced by Donsker and Varadhan [2] and (1.1) was obtained under the hypothesis that for any x and $x' \in \mathscr{X}$, the resolvents $R(x, dy)$ $(= \sum_{n=1}^{\infty} P^n(x, dy)/2^n)$ and $R(x', dy)$ are equivalent. If $X_0, X_1, \cdots$ is a random walk on a *compact* group, it was later proved that (1.1) holds under a weaker condition: (see [1])

$$R(x, A) > 0 \quad \text{for every} \quad x \in \mathscr{X} \quad \text{and open set} \quad A \subset \mathscr{X}. \tag{1.2}$$

In Theorem (2.11), we will generalize this result to random walks on R^n (which is only a locally compact group) under the assumption (1.2). This is not a trivial generalization because all the estimates in [1] have to be sharpened and the vague topology on $\mathscr{M}$ has to be considered. Throughout the rest of this paper, $X_0, X_1, \cdots$ is assumed to be a random walk on R^n which satisfies (1.2).

Received May 1983; revised May 1984.
AMS 1980 *subject classifications.* Primary 60F10; secondary 60J05.
Key words and phrases. Random walks.

REMARK. We will actually assume that $P(x, A) > 0$ for every $x \in \mathscr{X}$ and open set A. This causes no loss of generality because of Theorem 3.3 [3].

Section 2. We begin with an easy lemma whose proof will be omitted.

LEMMA 2.1. *Let $\mathscr{X}$ be a locally compact Polish space and let μ be a probability on $\mathscr{X}$. Then for any $\varepsilon > 0$, there exists a compact set K and a weak neighborhood N of μ s.t. $\lambda(K) > 1 - \varepsilon$ for every $\lambda \in N$.*

The following lemma is due to Donsker and Varadhan [3]. We include it here only for completeness.

LEMMA 2.2 *If $I(\mu) < \infty$, then there exists a transition function $\bar{p}(x, dy) \ll p(x, dy)$ for almost every x with respect to μ s.t. μ is an invariant measure for $\bar{p}(x, dy)$ and*

$$I(\mu) = \int \log \frac{\bar{p}(x, dy)}{p(x, dy)} \bar{p}(x, dy)\mu(dx)$$

where $\bar{p}(x, dy)/p(x, dy)$ is the Radon–Nikodym derivative of $\bar{p}(x, dy)$ with respect to $p(x, dy)$.

For any $G \in \mathscr{M}$, we define $E_k^n(G) = \{\omega\colon (1/k) \sum_{i=0}^{k-1} \delta_{X_{n+i}(\omega)} \in G\}$. Also, we call a probability measure λ indecomposable with respect to a transition $\pi(x, dy)$ if λ is invariant relative to $\pi(x, dy)$ and $\int_A \pi(x, A^c)\lambda(dx) > 0$ for every A with $0 < \lambda(A) < 1$. λ is said finitely decomposable if it is invariant relative to $\pi(x, dy)$ and there are finitely many mutually disjoint sets $A_1, \cdots, A_n$ s.t. $\int_{A_j} \pi(x, A_i)\lambda(dx) = 0$ if $i \neq j$ and $\lambda\,|_{A_i}$ (λ restricted to A_i) is indecomposable with respect to $\pi(x, dy)$.

LEMMA 2.3. *Let $\mu \in \mathscr{M}$ with $I(\mu) < \infty$ and let N be a weak neighborhood of μ. If μ is indecomposable with respect to the transition $\bar{p}(x, dy)$ in Lemma (2.2), then for each $\varepsilon > 0$, compact set K there exists a compact set $A_\varepsilon \subseteq \mathscr{X}$ s.t. for every $x \in K$,*

$$\liminf_{n\to\infty}(1/n)\log P_x(L_{n+m}(\omega, \cdot) \in N, X_{n+m-1} \in A_\varepsilon \text{ for some } m = 0, \cdots, [\varepsilon n] + 1) \geq -I(\mu) + O(\varepsilon).$$

The convergence is uniform for $x \in K$. (As usual, $[\varepsilon n]$ denotes the integral part of εn).

PROOF. By Lemma 2.1, for each $\varepsilon > 0$, we can choose A_ε and N' such that $\mu \in N' \subseteq N$ and $\lambda(A_\varepsilon) > 1 - \varepsilon/2$ if $\lambda \in N'$. Let $S_n = \inf\{k \geq 0\colon X_{n-1+k} \in A_\varepsilon\}$.

Now,

$$\begin{aligned}&P_x(E^0_{k+n}(N),\ X_{k+n-1} \in A_\varepsilon,\ X_m \notin A_\varepsilon \text{ for } n-1 \le m < k+n-1)\\ &\qquad = P_x(E^0_{k+n}(N),\ S_n = k) = \int_{E^0_{k+n}\cap\{S_n=k\}} dP_x\\ &\qquad \ge \int_{E^0_{k+n}\cap\{S_n=k\}} \prod_{i=1}^{n+k-1} \frac{p(x_{i-1}, dx_i)}{\bar p(x_{i-1}, dx_i)}\, d\bar P_x\\ &\qquad = \int_{E^0_{k+n}\cap\{S_n=k\}} \exp\left(-\sum_{i=1}^{n+k-1} \log \frac{\bar p(x_{i-1}, dx_i)}{p(x_{i-1}, dx_i)}\right) d\bar P_x.\end{aligned}$$

Here, $\bar P_x$ denotes the Markov process with starting point x and transition $\bar p(x, dy)$. Let $h(x, y) = (\bar p(x, dy))/(p(x, dy))$ and let

$$F_{m,\varepsilon} = \{\omega\colon (1/(m-1)) \textstyle\sum_{i=1}^{m-1} \log h(x_{i-1}, x_i) \le I(\mu) + \varepsilon\}.$$

Then,

$$\begin{aligned}&\sum_{k=0}^{[\varepsilon n]+1} P_x(E^0_{k+n}(N),\ S_n = k)\\ &\quad \ge \sum_{k=0}^{[\varepsilon n]+1} \bar P_x(E^0_{k+n}(N) \cap \{S_n = k\} \cap F_{n+k,\varepsilon}) \exp(-(k+n-1)(I(\mu)+\varepsilon))\\ &\quad \ge \exp((-([\varepsilon n]+n))(I(\mu)+\varepsilon))\bar P_x(\cup_{k=0}^{[\varepsilon n]+1} (E^0_{k+n}(N) \cap \{S_n = k\} \cap F_{n+k,\varepsilon}))\\ &\quad \ge \exp((-([\varepsilon n]+n))(I(\mu)+\varepsilon))\\ &\qquad\qquad \cdot \bar P_x((\cap_{k=0}^{[\varepsilon n]+1} (E^0_{k+n}(N) \cap F_{n+k,\varepsilon})) \cap (\cup_{k=0}^{[\varepsilon n]+1} (S_n = k)))\\ &\quad \ge \exp((-([\varepsilon n]+n))(I(\mu)+\varepsilon))\bar P_x(\cap_{k=0}^{\infty} (E^0_{k+n}(N') \cap F_{n+k,\varepsilon}))\\ &\qquad\qquad\qquad\qquad \text{for almost every } x - \mu(dx).\end{aligned}$$

The last inequality is true because $S_n(\omega) \le 1 + [\varepsilon n]$ if $\omega \in \cap_{k=0}^{[\varepsilon n]+1} E^0_{k+n}(N')$. Since $\bar P_\mu$ is ergodic by the assumption that μ is indecomposable with respect to $\bar p(x, dy)$, we have $\bar P_x(\cap_{k=0}^{\infty} (E^0_{k+n}(N') \cap F_{n+k,\varepsilon})) \to 1$ for a.e. $x - \mu(dx)$. Thus

$$\begin{aligned}&\liminf_{n\to\infty}(1/n)\log P_x(L_{n+m}(\omega, \cdot) \in N,\\ &\qquad\qquad X_{n+m-1} \in A_\varepsilon \text{ for some } m = 0, 1, \cdots, [\varepsilon n]+1)\\ &\quad = \liminf_{n\to\infty}(1/n)\log \sum_{k=0}^{[\varepsilon n]+1} P_x(E^0_{k+n}(N),\ X_{k+n-1} \in A_\varepsilon,\\ &\qquad\qquad X_m \notin A \text{ for } n-1 \le m < k+n-1)\\ &\quad \ge -(I(\mu)+\varepsilon)(1+\varepsilon) = -I(\mu) + O(\varepsilon)\end{aligned}$$

for a.e. $x - \mu(dx)$. We will omit the proof for general starting point x since it is the same as that in Lemma 2.5 [1].

The same A_ε works for all x in a compact set and uniform convergence is an easy consequence of the fact that we are dealing with random walks.

The following lemma is the analogue of Lemma 2.4 [1], where compactness of $\mathscr{X}$ was assumed.

LEMMA 2.4. *Let $\mu = \sum_{i=1}^{m} \lambda_i \mu^i$ be a convex combination of μ^i's with $\mu^i \in \mathscr{M}$ and $I(\mu^i) < \infty$ for $i = 1, 2, \cdots, m$. Let $\bar{p}_i(x, dy)$ be the transition functions in Lemma 2.2 corresponding to μ^i for each $i = 1, 2, \cdots, m$. If μ^i is indecomposable with respect to $\bar{p}_i(x, dy)$ for each i, then, for each $x \in \mathscr{X}$,*

$$\liminf_{n\to\infty}(1/n)\log P_x(L_n(\omega, \cdot) \in N) \geq -\sum_{i=1}^{m} \lambda_i I(\mu^i)$$

where N is an arbitrary weak neighborhood of μ.

PROOF. For simplicity, we only prove the case that $m = 2$. The same proof works for general m.

For $\varepsilon > 0$, choose N_1 and N_2, weak neighborhoods of μ^1 and μ^2 respectively such that $t_1N_1 + t_2N_2 \subseteq N$ if $|t_i - \lambda_i| < \varepsilon$ for $i = 1, 2$. Let $N_i' \subseteq N_i$ be weak neighborhoods of μ^i's s.t. $\lambda \in N_i$ if $\| \lambda - \lambda' \| < \varepsilon$ for some $\lambda' \in N_i'$. Also, by Lemma 2.3, we can choose a compact set A_ε s.t. when n is large enough,

$$P_x(L_{n+m}(\omega, \cdot) \in N_1', X_{n+m-1} \in A_\varepsilon \text{ for some } 0 \leq m \leq [\varepsilon n] + 1) \geq \exp n(-I(\mu^1) + O(\varepsilon))$$

and

$$P_y(L_n(\omega, \cdot) \in N_2') \geq \exp n(-I(\mu^2) + O(\varepsilon)), \quad y \in A_\varepsilon.$$

For any positive n, let ℓ_1 be the integer s.t. $\ell_1/n \leq \lambda_1 < (\ell_1 + 1)/n$ and let $T_1 = \inf\{k \geq \ell_1 : X_{k-1} \in A_\varepsilon\}$. Now, for large n,

$$\begin{aligned}
&P_x(L_n(\omega, \cdot) \in N) \\
&\quad \geq P_x(E_{T_1}^0(N_1') \cap E_{n-T_1}^{T_1}(N_2'), \ell_1 \leq T_1 \leq \ell_1 + [\varepsilon\ell_1] + 1) \\
&\quad = \sum_{k=\ell_1}^{\ell_1+[\varepsilon\ell_1]+1} P_x(E_k^0(N_1') \cap E_{n-k}^k(N_2'), T_1 = k) \\
&\quad = \sum_{k=\ell_1}^{\ell_1+[\varepsilon\ell_1]+1} E(P_x(E_k^0(N_1') \cap E_{n-k}^k(N_2'), T_1 = k \mid X_0, X_1, \cdots, X_{k=1})) \\
&\quad \geq \sum_{k=\ell_1}^{\ell_1+[\varepsilon\ell_1]+1} P_x(E_k^0(N_1'), T_1 = k)\exp(n - k)(-I(\mu^2) + O(\tilde{\varepsilon})) \\
&\quad = P_x(L_{\ell_1+m}(\omega, \cdot) \in N_1', X_{\ell_1+m-1} \in A_\varepsilon \text{ for some } m = 0, \cdots, [\varepsilon\ell_1] + 1) \\
&\qquad \cdot \exp(n - \ell_1)(-I(\mu^2) + O(\varepsilon)) \\
&\quad \geq \exp \ell_1(-I(\mu^1) + O(\varepsilon))\exp(n - \ell_1)(-I(\mu^2) + O(\varepsilon)).
\end{aligned}$$

Thus:

$$\liminf_{n\to\infty}(1/n)\log P_x(L_n(\omega, \cdot) \in N) \geq \lambda_1(-I(\mu^1) + O(\varepsilon)) + \lambda_2(-I(\mu^2) + O(\varepsilon)).$$

We then complete the proof by letting $\varepsilon \to 0$.

COROLLARY 2.5. *If μ^i is finitely decomposable with respect to $\bar{p}_i(x, dy)$ for each $i = 1, 2, \cdots, m$, then*

$$\liminf_{n\to\infty}(1/n)\log P_x(L_n(\omega, \cdot) \in N) \geq -\sum_{i=1}^{m} \lambda_i I(\mu^i).$$

PROOF. For each k, there exist $A_j^k \subseteq \mathscr{X}$, $j = 1, 2, \cdots, n_k$ such that $\mu^k = \sum_{j=1}^{n_k} \mu^k(A_j^k)\mu_{A_j^k}$ and $\mu_{A_j^k}^k$'s are indecomposable with respect to $\bar{p}_k(x, dy)$. By Corollary 2.2 [1], $I(\mu^k) = \sum_{j=1}^{n_k} \mu^k(A_j^k)I(\mu_{A_j^k}^k)$. Since $\mu = \sum_{i=1}^m \lambda_i(\sum_{j=1}^{n_i} \mu^i(A_j^i)\mu_{A_j^i}^i)$, we have, by Lemma 2.5,

$$\liminf_{n\to\infty}(1/n)\log P_x(L_n(\omega, \cdot) \in N) \geq -\sum_{i=1}^m \lambda_i(\sum_{j=1}^{n_i} \mu^i(A_j^i)I(\mu_{A_j^i}^i)) = -\sum_{i=1}^n \lambda_i I(\mu^i).$$

This completes the proof.

For a sequence of sub-probability measures μ_n, we say μ_n converges to μ weakly if $\int f\, d\mu_n \to \int f\, d\mu$ for every continuous bounded function f and μ_n converges to μ vaguely, if $\int f\, d\mu_n \to \int f\, d\mu$ for every continuous function which vanishes at ∞. We use $\mu_n \to_{\mathrm{w}} \mu$ and $\mu_n \to_{\mathrm{v}} \mu$ respectively to express such convergences. Throughout this paper, $\mathscr{M}$ will denote the space of probability measures and M the space of all sub-probability measures. We now proceed to prove the analogue of Lemma 2.9 [1]. First, we state the following lemma without proof.

LEMMA 2.6. *Let μ_n be a sequence of sub-probability measures in M and let μ be a probability measure. Then $\mu_n \to_{\mathrm{v}} \mu$ if and only if $\mu_n \to_{\mathrm{w}} \mu$.*

If we extend the domain of I-functional from $\mathscr{M}$ to M with the obvious extension, we have the following.

LEMMA 2.7. *$I(\mu)$ is lower semi-continuous on M with the vague topology.*

PROOF. Since $p(x, dy)$ is a Feller transition function and $\mathscr{X}$ is a σ-compact space, we can write

$$I(\mu) = \sup_{f\in B^{\mathrm{c.c.}}(\mathscr{X})} \int \log \frac{f(x)}{(pf)(x)}\, \mu(dx)$$

where $B^{\mathrm{c.c.}}(\mathscr{X})$ consists of bounded positive continuous functions which are constants outside some compact set. (See [4]). Now, $pf(x) = \int f(y)p(x, dy) = \int f(y + x)p(dy)$, thus

$$\lim_{x\to\infty} |(pf)(x) - f(x)| = \lim_{x\to\infty} \int f(x + y) - f(x)p(dy) = 0.$$

Therefore $((pf)(x))/(f(x)) \to 1$ as $x \to \infty$, i.e., $\log((pf)(x))/(f(x))$ vanishes at ∞. This implies that $I(\mu)$ is lower semi-continuous with the vague topology.

Let $M_\ell = \{\mu \in M: I(\mu) \leq \ell\}$ and $\mathscr{M}_\ell = \{\mu \in \mathscr{M}: I(\mu) \leq \ell\}$.

LEMMA 2.8. *M_ℓ is convex and vaguely compact.*

PROOF. M_ℓ is vaguely closed because I is lower semi-continuous in the vague topology. Then the compactness follows from the fact that M is compact. Convexity of $I(\mu)$ is trivial.

LEMMA 2.9. *If μ ($\mu \neq 0$) is an extreme point of M_ℓ then $\mu/\|\mu\|$ is an extreme point of $\mathscr{M}_{\ell/\|\mu\|}$.*

PROOF. Trivial.

THEOREM 2.10. *Let $X_0, X_1, \cdots, X_n, \cdots$ be a random walk satisfying* (1.2). *Then for any $x \in \mathscr{X}$ and $\mu \in \mathscr{M}$,*

$$\liminf_{n\to\infty}(1/n)\log P_x(L_n(\omega, \cdot) \in N) \geq -I(\mu)$$

where $N \subseteq \mathscr{M}$ is an arbitrary weak neighborhood of μ.

PROOF. Let $I(\mu) = \ell$. Since M_ℓ is vaguely compact, there exists a sequence $\mu_i \in M_\ell$ such that $\mu_i \to_v \mu$ and $\mu_i = \sum_{k=1}^{k_i} \lambda_{i,k}\mu_{i,k}$ where $\mu_{i,k}$'s are extreme points of M_ℓ and $\lambda_{i,k} > 0$ with $\sum_{k=1}^{k_i} \lambda_{i,k} = 1$. Let $A(i, \varepsilon) = \{k: \|\mu_{i,k}\| \geq 1 - \varepsilon\}$. By Lemma 2.7, we have $\mu_i \to_w \mu$. Thus for any small $\varepsilon > 0$, $\sum_{k\in A(i,\varepsilon)} \lambda_{i,k} > 1 - \varepsilon$ when $i \to \infty$. If we let $N' \subseteq N$ be a weakly open neighborhood of μ such that $\lambda \in N$ if $\|\lambda - \lambda'\| \leq \varepsilon + \varepsilon/(1-\varepsilon)$ for some $\lambda' \in N'$ and let $\mu_{i,\varepsilon} = \sum_{k\in A(i,\varepsilon)} \lambda^\varepsilon_{i,k}\mu'_{i,k}$ where

$$\lambda^\varepsilon_{i,k} = \frac{\lambda_{i,k}}{\sum_{k\in A(i,\varepsilon)} \lambda_{i,k}} \quad \text{and} \quad \mu'_{i,k} = \frac{\mu_{i,k}}{\|\mu_{i,k}\|}.$$

It is easy to see that $\|\mu_{i,\varepsilon} - \mu_i\| \leq \varepsilon + \varepsilon/(1-\varepsilon)$, and therefore $\mu_{i,\varepsilon} \in N$ if $\mu_i \in N'$. Since each $\mu'_{i,k}$ is an extreme point of $\mathscr{M}_{\ell/\|\mu_{i,k}\|}$, it is finitely decomposable with respect to $\bar{p}_i(x, dy)$ by Lemma 2.9 [1]. By Corollary 2.5, we thus have:

$$\liminf_{n\to\infty}(1/n)\log P_x(L_n(\omega, \cdot) \in N) \geq \textstyle\sum_{k\in A(i,\varepsilon)} \lambda_{i,k} I(\mu'_{i,k})$$

when i is large enough. But

$$\begin{aligned}\textstyle\sum_{k\in A(i,\varepsilon)} \lambda^\varepsilon_{i,k} I(\mu'_{i,k}) &= \sum_{k\in A(i,\varepsilon)} \frac{\lambda_{i,k}}{\sum_{k\in A(i,\varepsilon)} \lambda_{i,k}} I\left(\frac{\mu_{i,k}}{\|\mu_{i,k}\|}\right)\\ &\leq \frac{1}{(1-\varepsilon)^2}\sum_{k\in A(i,\varepsilon)} \lambda_{i,k} I(\mu_{i,k}) \leq \frac{\ell}{(1-\varepsilon)^2},\end{aligned}$$

thus

$$\liminf_{n\to\infty} \frac{1}{n}\log P_x(L_n(\omega, \cdot) \in N) \geq \frac{-\ell}{(1-\varepsilon)^2} = -\frac{I(\mu)}{(1-\varepsilon)^2}.$$

We complete the proof by letting $\varepsilon \to 0$.

REFERENCES

[1] CHIANG, T.-S. (1982). Large deviations of some Markov processes on compact metric spaces. *Z. Wahrsch. verw. Gebiete* **61** 271–281.

[2] DONSKER, M. D. and VARADHAN, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time—I. *Comm. Pure Appl. Math.* **28** 1–47.

[3] DONSKER, M. D. and VARADHAN, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time—III. *Comm. Pure Appl. Math.* **29** 389–461.

[4] JAIN, N. (1982). A Donsker–Varadhan type invariance principle. *Z. Wahrsch. verw. Gebiete* **59** 117–138.

INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIPEI, TAIWAN