Chin. Ann. of Math.
12B:3 (1991), 348—357.

LARGE DEVIATION FOR EMPIRICAL FIELD OF A SYMMETRIC MEASURE**

WU LIMING (吴黎明)*

Abstract

This paper discusses the large deviation for the empirical field of a symmetric measure. The lower bound of the large deviation is obtained by extending the classcal Shannon-McMillan theorem. The upper bound is established by means of Legendre transformation and the minimax theorem.

§ 1. Introduction

Let F be a finite set and introduce the compact product space $\Omega=F^{\mathbf{N}}$. Let $P(\Omega)$ denote the compact space of probability measures on Ω, and $P_s(\Omega)$ denote the subspace of all elements of $P(\Omega)$ which are invariant under all shifts $\theta_i:\Omega\to\Omega$, $i\in\mathbf{N}$, where $(\theta_i\omega)(j)=\omega(i+j)$.

Let us consider a symmetric measure μ on Ω, i.e, an element of $P(\Omega)$ which is invariant under all permutations. By the well known de Finetti theorem, μ is a mixture of homogenous product measures, i.e., has a representation

$$\mu=\int_{P(F)} m(d\rho)\rho^{\mathbf{N}},$$

where m is a probability measure on the space $P(F)$ of all probability measures on F.

Now we define the empirical field R_n as the random element of $P(\Omega)$ given by

$$R_n(\omega)=\frac{1}{n}\sum_{i=0}^{n-1}\delta_{\theta_i\omega},$$

where δ_ω is the Dirac measure at ω. The distribution $\mu\circ R_n^{-1}$ of R_n under μ belongs to $P(\Omega)$. Our principal result is that the sequence

$$\mu\circ R_n^{-1},\ n=1,\ 2,\cdots$$

satisfies a large deviation principle with the rate function given by $I(\nu)=\inf_{\rho\in\operatorname{supp}(m)} h(\nu,\ \rho^N)$, where $h(\nu,\ \rho^N)$ is the Donsker-Varadhan's specific relative entropy of ν with respect to ρ^N. More explicitely, we show in Section 3 that for open subsets

Manuscript received October 3, 1989.
* Department of Mathematics, Wuhan University, Wuhan, Hubei. China.
** This research is supported by the National Postdoctoral Science Fund.

A of $P(\Omega)$

$$\underline{\lim_{n}} \frac{1}{n} \log \mu[R_n \in A] \geqslant - \inf_{\nu \in A \cap P_s(\Omega)} I(\nu)$$

and in Section 4 that for closed subsets A of $P(\Omega)$

$$\overline{\lim_{n}} \frac{1}{n} \log \mu[R_n \in A] \geqslant - \inf_{\nu \in A \cap P_s(\Omega)} I(\nu).$$

By the contraction principle [4], we obtain in particular a large deviation principle for the empirical distributions

$$D_n(\omega) = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{\omega(i)}$$

on the state space F, and the resulting rate function $I_0(Q)$, $Q \in P(F)$, is given by

$$I_0(Q) = \inf_{\substack{\nu \in P_s(\Omega) \\ \pi\nu = Q}} I(\nu)$$

$$= \inf_{\rho \in \mathrm{supp}(m)} \inf_{\substack{\nu \in P_s(\Omega) \\ \pi\nu = Q}} h(\nu, \rho^N)$$

$$= \inf_{\rho \in \mathrm{supp}(m)} H(Q|\rho),$$

where $\pi : \omega \to \omega(0)$ and $H(Q|\rho)$ is the relative entropy of Q with respect to ρ, i. e.

$$H(Q|\rho) = \sum_{a \in F} Q(a) \log Q(a)/\rho(a).$$

In the classic case, i.e, $m = \delta_\rho$ for some $\rho \in P(F)$, the large deviation principle for the empirical distributions $\mathscr{D}_n$ is the well-known Sanov theorem and that for R_n is also the classic one, obtained by Donsker, Varadhan [4]. When $m = \sum_{k=1}^{l} \delta_{\rho_k}$, $\rho_k \in P(F)$, $l \in N$, our results stated above are the direct consequences of the previous classical results. But in the general case, since we have not an uniform large deviation estimation for $\rho^N \circ R_n^{-1}$ for all $\rho \in P(F)$, our results are not so direct. In fact, De Acosta has given in [1] a rough upper bound of large deviation for the empirical distribution of symmetric measure by means of Legendre transformation. Our results improve essentially his results.

§ 2. Preliminaries

The purpose of this section is to introduce notations and summarize some known results which will be need ed below.

The notations introduced in Section 1 are reserved.

We denote by $\mathscr{F}_n$ the σ-field on $\Omega = F^N$ generated by the projections $\omega \to \omega(i)$, $i = 0, 1, \cdots, n-1$: and by $\mathscr{D}_n$ the σ-field on Ω generated by the empirical distribution $\mathscr{D}_n$ and the projections $\omega \to \omega(i)$, $i = n, n+1, \cdots$ (clearly $\mathscr{D}_n \supseteq \mathscr{D}_{n+1}$). Let $\mathscr{J}$ be the σ-field of all subsets of Ω which are invariant under all shifts θ_i, $i \in N$, and let

$\mathscr{D}_\infty=\bigcap_{n=0}^{\infty}\mathscr{D}_n.$

For a symmetric measure μ on Ω, the regular conditional probability measure given on $\mathscr{D}_\infty$ is known to be (de Finetti):

$$\mu(\cdot\,|\mathscr{D}_\infty)(\omega)=\rho(\omega)^N \quad \text{a.s.}(\mu), \tag{2.1}$$

where $\rho\in P(F)$. Now denote by $m(d\rho)$ the distribution of random function $\omega\to\rho(\omega)$ under μ. By (2.1) we have for all $B\in\mathscr{F}=\bigcup_{n=0}^{\infty}\mathscr{F}_n$

$$\mu(B)=\int_\Omega \rho(\omega)^N(B)\mu(d\omega)=\int_{P(F)}\rho^N(B)m(d\rho). \tag{2.2}$$

From (2.1) and the ergodic decomposition of μ, we get also

$$\mathscr{J}^{(\mu)}=\mathscr{D}_\infty.$$

For $\nu\in P_s(\Omega)$ and $n>0$, we denote by

$$H_n(\nu):=-\sum_{\xi\in F^n}\nu[\omega:\omega(k)=\xi(k),k=0,1,\cdots,n-1],$$
$$\log\nu(\omega:\omega(k)=\xi(k),\ k=0,1,\cdots,n-1) \tag{2.3}$$

the entropy of ν restricted to $\mathscr{F}_n$. The specific entropy of ν is defined as

$$h(\nu):=\lim_n\frac{1}{n}H_n(\nu)=\inf_n\frac{1}{n}H_n(\nu) \tag{2.4}$$

because $H_n(\nu)$ is subadditive in n. More precisely, the Shannon McMillan Theorem states that

$$\lim_n\frac{1}{n}\log\nu[\omega:\omega(k)=\xi(k),\ k=0,1,\cdots,n-1]=-h(\nu^\xi) \tag{2.5}$$

holds for ν-a.e $\xi\in\Omega$, and in $L^1(\nu)$, where $\nu^\xi=\nu(\cdot\,|J)(\xi)$, ν-a.s, is the ergodic decomposition of ν. We remark that $h(\cdot)$ is concave, upper semi-continuous on $P_s(\Omega)$.

Now let $\rho\in P(F)$ and $\nu\in P_s(\Omega)$. Then

$$H_n(\nu,\ \rho^N)=\int\log\frac{d\nu|\mathscr{F}n}{d\rho^N|\mathscr{F}n}d\nu\geqslant 0 \tag{2.6}$$

defines the relative entropy of $\nu|\mathscr{F}_n$ with respect to $\rho^N|\mathscr{F}_n$, and the specific relative entropy is given by

$$h(\nu,\ \rho^N)=\lim_n\frac{1}{n}H_n(\nu,\ \rho^N)=\lim_n\frac{1}{n}\,[H_n(\nu)-n\sum_{a\in F}\nu(\omega(0)=a)\cdot\log\rho(a)]$$
$$=-h(\nu)-\sum_{a\in F}\nu(\omega(0)=a)\log\rho(a) \tag{2.7}$$

with convention: $\log 0=-\infty$, $0\log 0=0$.

We need a remark on the space $P(\Omega)$.

Since Ω is a compact space, $P(\Omega)$ is compact too. We shall use the following explicit metric. Choose a sequence (φ_n) of continuous functions on Ω satisfying the following conditions:

① the linear span of (φ_n) is dense $C_b(\Omega)$;

② $\|\varphi_n\| \leqslant 1$, $n=1,2,\cdots$;

③ for each n, there exists $k(n) \in \mathbf{N}$ so that

$$\varphi_n \in \mathscr{F}_{k(n)}.$$

Now define the metric d on $P(\Omega)$ by

$$d(\nu, \mu) = \sum_{n=1}^{\infty} 2^{-n} \left| \int \varphi_n d\nu - \int \varphi_n d\mu \right| \tag{2.8}$$

for $\nu, \mu \in P(\Omega)$.

§ 3. Lower Bound

The following theorem will be proved in this section.

Theorem 3.1. *Let μ be a symmetric measure on $\Omega = F^{\mathbf{N}}$, and let $\mu = \int_{P(F)} \rho^{\mathbf{N}} m(d\rho)$ be the de Finetti decomposition of μ. Then for any open subset G of $P(\Omega)$*

$$\lim_n \frac{1}{n} \log \mu(R_n \in G) \geqslant - \inf_{\nu \in G \cap P_s(\Omega)} I(\nu) \tag{3.1}$$

where

$$I(\nu) = \inf_{\rho \in \mathrm{supp}(m)} h(\nu, \rho^{\mathbf{N}}).$$

The main idea to get the lower bound estimation above, as indicated by [4], [3], [2], is to give an extension of Shannon-McMillan-Breiman Theorem. As a result of independent interest, the following generalization of S-M-B Theorem is stated as

Theorem 3.2. *Let $\nu \in P_s(\Omega)$ and μ given in Theorem 1. Then we have*

$$\frac{1}{n} \log \frac{\nu[\omega:\omega(k)=\xi(k),\ k=0,1,\vdots,n-1]}{\mu[\omega;\omega(k)=\xi(k),\ k=0,1,\cdots,n-1]} \to \inf_{\rho \in \mathrm{supp}(m)} h(\nu^{\xi}, \rho^{N}) = I(\nu^{\xi}) \tag{3.2}$$

for ν-a.e. $\xi \in \Omega$, where $\nu^{\xi} = (\cdot | \mathscr{J})(\xi)$, ν-a.s, is the ergodic decomposition of ν.

Proof of Theorem 3.2 Set $A_n = [\omega \in \Omega: \omega(k) = \xi(k),\ k=0,1,\cdots,n-1]$. Obviously,

$$\rho^N(A_n) = \prod_{a \in F} \rho(a)^{n \cdot D_n(\xi)(a)} \tag{3.3}$$

(with convention; $0^r = 0$ for $r > 0$ and $0^0 = 1$).

Now, the first term in (3.2) can be writen as

$$\frac{1}{n} \log \nu(A_n) - \frac{1}{n} \log \mu(A_n) = \frac{1}{n} \log \nu(A_N) - \frac{1}{n} \log \int_{P(F)} \rho^N(A_n) m(d\rho)$$

$$= \frac{1}{n} \log \nu(A_n) - \frac{1}{n} \log \int_{P(F)} \left(\prod_{a \in F} \rho(a)^{D_n(\xi, a)} \right)^n m(d\rho). \tag{3.4}$$

Letting n approach infinity, we have

$$\frac{1}{n} \log \nu(A_n) \to -h(\nu^{\xi})\ \ \nu.\ \text{a.s.} \tag{3.5}$$

by (2.5) and

$$D_n(\xi, a) \to \nu^{\xi}(\omega(0) = a)\ \ \nu.\ \text{a.s.} \tag{3.6}$$

by Birkhoff ergodic theorem.

Now, we see clearly that the key to the proof of Theorem 3.2 is

$$\Big[\int_{P(F)}\big(\prod_{a\in F}\rho(a)^{D_n(\xi,a)}\big)^n m(d\rho)\Big]^{1/n}\longrightarrow \sup_{\rho\in\mathrm{supp}(m)}\prod_{a\in F}\rho(a)^{\nu^\xi(\omega(0)=a)},\ \nu\text{—a.s.} \tag{3.7}$$

In order to prove (3.7), we remark firstly for $a\in F$,

$$\nu^\xi(\omega(0)=a)=0\Rightarrow D_n(\xi,a)=0 \text{ for all } n,\ \nu.\ \text{a.s.} \tag{3.8}$$

Indeed, setting $A:=\{\xi\in\Omega:\nu^\xi(\omega(0)=a)=0\}\in\mathscr{J}$, we get

$$\int_A D_n(\xi,a)d\nu=\int_A E_\nu[D_n(\xi,a)\,|\,\mathscr{J}]d\nu=\int_A E_\nu(1_{[\omega(0)=a]}\,|\,\mathscr{J})d\nu$$

$$=\int_A \nu^\xi(\omega(0)=a)d\nu=0.$$

Now, fix $\xi\in\Omega$ such that(3.6) and (3.8) hold for all $a\in F$. An elementary estimation gives us

$$\Big|\prod_{a\in F}\rho(a)^{D_n(\xi,a)}-\prod_{a\in F}\rho(a)^{\nu^\xi(\omega(0)=a)}\Big|\leqslant-\sum_{a;\,\nu^\xi(\omega(0)=a)>0}\rho(a)^{\nu^\xi(\omega(0)=a)-\varepsilon}$$

$$\cdot\log\rho(a)\cdot|D_n(\xi,a)-\nu^\xi(\omega(0)=a)| \tag{3.9}$$

for all n with $|D_n(\xi,a)-\nu^\xi(\omega(0)=a)|\leqslant\varepsilon$, where $\varepsilon>0$ is chosen smaller than $\min(\nu^\xi(\omega(0)=a):\nu^\xi(\omega(0)=a)>0)$.

Therefore,

$$\prod_{a\in F}\rho(a)^{D_n(\xi,a)}\longrightarrow\prod_{a\in F}\rho(a)^{\nu^\xi(\omega(0)=a)}$$

uniformely for $\rho\in P(F)$, and (3.7) can be easily deduced from the triangular inequality and the following fact that

$$\|\cdot\|_n\to\|\cdot\|_\infty \text{ on any probability space.}$$

Finally, combined with (3.5) and (3.7), the formula (3.4) implies that the left term in (3.2) tends ν—a.s to

$$-h(\nu^\xi)-\log\sup_{\rho\in\mathrm{supp}(m)}\prod_{a\in F}\rho(a)^{\nu^\xi(\omega(0)=a)}$$

$$=\inf_{\rho\in\mathrm{supp}(m)}\Big[-h(\nu^\xi)-\sum_{a\in F}\nu^\xi(\omega(0)=a)\log\rho(a)\Big]$$

$$=\inf_{\rho\in\mathrm{supp}(m)}h(\nu^\xi,\rho^N) \qquad (by (2.7)).$$

In order to obtain our lower bound we follow [2], [3].

The following lemma is taken from [2], [3].

Lemma 3.3. *Let $\nu\in P_s(\Omega)$. Then there exists a sequence (ν_n) of ergodic measures converging to ν such that*

$$\lim_n h(\nu_n)=h(\nu), \tag{3.10}$$

$$\mathrm{supp}(\pi(\nu_n))=\mathrm{supp}(\pi(\nu)). \tag{3.11}$$

We remark that (3.11) was not stated in [2], [3], but is provided by their proof.

(3.10) and (3.11) imply

$$I(\nu_n)\to I(\nu).$$

Proof of Theorim 3.1 Let $\nu\in P_s(\Omega)$ and U be an open neighborhood of ν. It will suffice to show

$$\lim \frac{1}{n}\log\mu(R_n\in U)\geqslant -I(\nu). \tag{3.13}$$

1) Assume now that ν is ergodic. In view of(2.8), there is a neighbourhood of ν

$$U_0=\{\nu'|f_k d\nu'-f_k d\nu|<\varepsilon \text{ for } k=1,2,\cdots,m\}U$$

where $f_1,\cdots,f_m$ are F_p-measurable for some $p\geqslant 1$.

Since ν-a.s R_n converges weakly to ν,

$$\lim_n \nu[R_n\in U_0)=1.$$

Assume $I(\nu)<\infty$, as otherwise there is nothing to prove, by Theorem 3.2, $\nu|\mathscr{F}_{n+p}\ll\mu|\mathscr{F}_{n+p}$, and for $g_n=\dfrac{d\nu|\mathscr{F}_{n+p}}{d\mu|\mathscr{F}_{n+p}}$, we have

$$\frac{1}{n}\log g_n\xrightarrow{a.s} I(\nu), \text{ a.s.}$$

Now note that the set

$$[R_n\in U_0]=\bigcap_{k=1}^{m}\left\{\left|\frac{1}{n}\sum_{i=0}^{n-1}f_k\circ\theta_i-\int f_k d\nu\right|<\varepsilon\right\}$$

belongs to $\mathscr{F}_{n+p}$. This allows us to write

$$\mu(R_n\in U)\geqslant\mu\left(R_n\in U_0,\ \frac{1}{n}\log g_n\leqslant I(\nu)+\varepsilon,\ g_n>0\right)$$

$$\geqslant\exp(-n(I(\nu)+\varepsilon))\nu(R_n\in U_0,\ \frac{1}{n}\log g_n\leqslant I(\nu)+\varepsilon)$$

and since the second factor in the last number approaches one, we obtain (3.13).

2) Now let $\nu\in P_s(\Omega)$. Approximate ν by ergodic ν_n with $I(\nu_n)\to I(\nu)$ as ensured by Lemma 3.4. and (3.12). Then the left side of (3.13) is greater or equal to $-I(\nu_n)$ for all n big enough and then we get (3.13).

§4. Upper Bound

In this section, we shall prove

Thorem 4.1. *For any closed subset C of $P(\Omega)$, we have*

$$\overline{\lim_{n\to\infty}}\ n^{-1}\log\mu(R_n\in C)\leqslant-\inf_{\nu\in C\cap P_s(\Omega)}I(\nu) \tag{4.1}$$

where μ and $I(\cdot)$ are given in Theorem 3.1.

In order to obtain the upper bound (4.1) we follow [4] and introduce a stationary modification of R_n. For each non-negative integer n define $\pi_n:\Omega\to\Omega$ by

$$\pi_n(\omega)\,(i+nj)=\omega(i),\ j\in N,\ i=0,1,\cdots,n-1.$$

and

$$R_n^s(\omega)=R_n\circ\pi_n.$$

Clearly $R_n^s(\omega)\in P_s(\Omega)$.

Lemma 4.2. *For any $\mathscr{F}_k$-measurable function f, we have*

$$\left|\int f dR_n - \int f dR_n^s\right| \leqslant \frac{2k}{n}\|f\|_\infty. \tag{4.2}$$

Proof Note that

$$\int_\Omega f dR_n^s = \frac{1}{n}\sum_{i=0}^{n-1} f\circ\theta_i\circ\pi_n,$$

$$\int_\Omega f dR_n = \frac{1}{n}\sum_{i=0}^{n-1} f\circ\theta_i$$

and $f\circ\theta_i\circ\pi_n = f\circ\theta_i$ *for* $i=0,1,\cdots, n-k$.

The following lemma is the key to our Theorem 4.1.

Lemma 4.3. *For any measurable subset A of $P_s(\Omega)$,*

$$\varlimsup_{n\to\infty} n^{-1}\log\mu(R_n^s\in A) \leqslant -\sup_{k\geqslant 1}\sup_{f\in\mathscr{F}_k}\inf_{\nu\in A}\inf_{\rho\in\mathrm{supp}(m)} k^{-1}\left(\int f d\nu - \log\int e^f d\rho^N\right),$$

where "$f\in\mathscr{F}_k$" signifies "f is $\mathscr{F}_k$-measurable"

Proof For any $k>0$ and F_k-measurable function f, we have by the Markov's inequality and (4.2)

$$\mu(R_n^s\in A) \leqslant \exp\left(-\frac{n}{k}\inf_{\nu\in A}\int f d\nu\right)\int\exp\left(\frac{n}{k}\int f dR_n^s\right)d\mu$$

$$\leqslant \exp\left(-\frac{n}{k}\inf_{\nu\in A}\int f d\nu + 2\|f\|_\infty\right)\cdot\int\exp\left(\frac{n}{k}\int f dR_n\right)d\mu. \tag{4.3}$$

Letus estimate the last factor in (4.3).

Let

$$g_j = \sum_{i=0}^{[\frac{n}{k}]-1} f\circ\theta_{ki+j} \quad j=0,1,\cdots,k-1.$$

Then

$$\left|\frac{n}{k}\int f dR_n - \frac{1}{k}\sum_{j=0}^{k-1} g_j\right| \leqslant \|f\|_\infty$$

which implies

$$\int\exp\left(\frac{n}{k}\int f dR_n\right)d\mu \leqslant e^{\|f\|}\int\exp\left(\frac{1}{k}\sum_{j=0}^{k-1} g_j\right)d\mu \leqslant \exp(\|f\|_\infty)\cdot\frac{1}{k}\sum_{j=0}^{k-1}\int_\Omega \exp(g_j)d\mu \tag{4.4}$$

by Jensen's inequality.

Noting that $\mu = \int \rho^N m(d\rho)$ and that for $l\geqslant k$, $f\circ\theta_i$, $f\circ\theta_{i+l}$ is independent under the probability measure ρ^N, we get

$$\int_\Omega \exp(g_j)d\mu = \int\left[\int\exp\left(\sum_{i=0}^{[\frac{n}{k}]-1} f\circ\theta_{ki+j}\right)d\rho^N\right]m(d\rho) = \int\left(\int e^f d\rho^N\right)^{[\frac{n}{k}]} m(d\rho),$$

where it follows by (4.3) and (4.4) that

$$\mu(R_n^s\in A) \leqslant \exp\left(-\frac{n}{k}\inf_{\nu\in A}\int f d\nu + 3\|f\|_\infty\right)\cdot\int\left(\int e^f d\rho^N\right)^{[\frac{n}{k}]} m(d\rho).$$

Therefore we get

$$\overline{\lim_{n\to\infty}} n^{-1}\log\mu(R_n^s\in A)\leqslant -\frac{1}{k}\inf_{\nu\in A}f d\nu+\frac{1}{k}\log\sup_{\rho\in\mathrm{supp}(m)}\int e^f d\rho^N$$
$$= -k^{-1}\inf_{\nu\in A}\inf_{\rho\in\mathrm{supp}(m)}\left(\int f d\nu-\log\int e^f d\rho^N\right)$$

for any $k\geqslant 1$ and any F_k-measurable function f.

Proof of Theorem 4.1 We establish firstly (4.1) with R_n^s replacing R_n. Its proof is divided into four steps.

Step 1. Assume firstly supp(m) and C are convex.

Because the function

$$(f,(\rho,\nu))\to\int f d\nu-\log\int\rho(e^f)d^N$$

is continnous, concave on f and convex on (ρ,ν), we can apply the minimax theorem and get subsequently

$$\sup_{f\in F_k}\inf_{\nu\in C}\inf_{\rho\in\mathrm{supp}(m)}\left(\int f d\nu-\log\int e^f d\rho^N\right)=\inf_{\nu\in C}\inf_{\rho\in\mathrm{supp}(m)}\sup_{f\in F_k}\left(\int f d\nu-\log\int e^f d\rho^N\right),$$

which is equal to

$$\inf_{\nu\in C}\inf_{\rho\in\mathrm{upp}(m)}H_k(\nu,\rho^N)$$

by the well known variational principle of the entropy (see e.g. [4]).

Therefore, we get by Lemma 4.3

$$\overline{\lim_{n\to\infty}} n^{-1}\log\mu(R_n^s\in C)\leqslant -\sup_k\inf_{\nu\in C}\inf_{\rho\in\mathrm{supp}(m)}k^{-1}H_k(\nu,\rho^N). \tag{4.5}$$

Since the function $k^{-1}H_k(\nu,\rho^N)$ is l.s.c. on the compact space $P_s(\Omega)-P(F)$ for each $k\geqslant 1$, and

$$\lim_{k\to\infty}k^{-1}H_k(\nu,\rho^N)=h(\nu,\rho^N)=\sup_k k^{-1}H_k(\nu,\rho^N),$$

we can deduce in an elementary way

$$\lim_{k\to\infty}\inf_{\nu\in C}\inf_{\rho\in\mathrm{supp}(m)}k^{-1}H_k(\nu,\rho^N)=\inf_{\nu\in C}\inf_{\rho\in\mathrm{supp}(m)}h(\nu,\rho^N). \tag{4.6}$$

Then we have by (4.5)and (4.6)

$$\overline{\lim_{n\to\infty}} n^{-1}\log\mu(R_n^s\in C)\leqslant -\inf_{\nu\in C}\inf_{\rho\in\rho\in\mathrm{supp}(m)}h(\nu,\rho^N)=-\inf_{\nu\in C}I(\nu). \tag{4.7}$$

Step 2. Assume that supp(m) is convex, and C is an arbitrary closed subset of $P_s(\Omega)$.

Since C is closed, and then compact, we can choose for each $\varepsilon>0$ a finite number of couvex subsets $C_j(\varepsilon)$ such that

$$C\subseteq C_\varepsilon=\bigcup_j C_j(\varepsilon)$$

and

$$\inf_{\nu\in C_\varepsilon}I(\nu)\geqslant\inf_{\nu\in C}I(\nu)-\varepsilon$$

(by the lower semi continuity of the function $(\nu,\rho)\to h(\nu,\rho^N)$.

We get finally from (4.7)

$$\overline{\lim_{n\to\infty}} n^{-1}\log \nu(R_n^s\in C)\leqslant -\inf_{\nu\in C_\varepsilon} I(\nu)\leqslant -\inf_{\nu\in C} I(\nu)+\varepsilon.$$

Step 3. Assume only that

$$\text{supp}(m)=\bigcup_j B_j \quad (\textit{finite union})$$

where B_j is closed, convex in $P(F)$.

In this case, let

$$\mu_j=\frac{1}{m(B_j)}\int_{B_j}\rho^N\, m(d\rho)$$

which satisfies the assumption of Step 2.

Evidently, one has

$$\mu\leqslant\sum_j m(B_j)\mu_j$$

which implies

$$\overline{\lim_{n\to\infty}} n^{-1}\log\mu(R_n^s\in C)\leqslant\sup_j \overline{\lim_{n\to\infty}} n^{-1}\log\mu_j(R_n^s\in C)\leqslant\sup_j(-\inf_{\nu\in C}\inf_{\rho\in B_j} h(\nu,\rho^N))$$
$$=-\inf_{\nu\in C}\inf_{\rho\in\text{supp}(m)} h(\nu,\rho^N).$$

Step 4. In general case, supp(m), being closed in $P(F)$, is compact. Then for any $\varepsilon>0$, we can choose a probability measure m^ε on F such that

$$\begin{cases}\text{supp}(m^\varepsilon) \textit{ is the union of a finite number of the closed convex subsets of } P(F)\\ (1-\varepsilon)m\leqslant m^\varepsilon,\\ \bigcap_{\varepsilon>0}\text{supp}(m^\varepsilon)=\text{supp}(m).\end{cases}$$

Letting $\mu_\varepsilon=\int\rho^N m^\varepsilon(d\rho)$.

We have therefore by Step 3

$$\overline{\lim_{n\to\infty}} n^{-1}\log\mu(R_s^s\in C)\leqslant\inf_{\varepsilon>0}\overline{\lim_{n\to\infty}} n^{-1}\log\mu_\varepsilon(R_n^s\in C)$$
$$\leqslant\inf_{\varepsilon>0}(-\inf_{\rho\in\text{supp}(m_\varepsilon)}\inf_{\nu\in C} h(\nu,\rho^N))=-\inf_{\rho\in\text{supp}(m)}\inf_{\nu\in C} h(\nu,\rho^N)=-\inf_{\nu\in C} I(\nu) \quad (4.1)'$$

by the lower semi-continuity of function $(\nu,\rho)\to h(\nu,\rho^N)$

It remains to show (4.1) for R_n in place of R_n^s. In fact, Orey has shown in [3] that the large deviation principle for $\mu\circ R_n^{-1}$ or $\mu\circ(R_n^s)^{-1}$ is equivalent. But for the completeness, we give here the demonstration.

Metrize $P(\Omega)$ as in (2.8). Then by Lemma 4.2 the distance between R_n and R_n^s converges to zero uniformly in ω as $n\to\infty$. Hence if $\varepsilon>0$ and C_ε is a closed ε-neighbourhood of C, there exists $n_0(\varepsilon)$ such that for any $n\geqslant n_0(\varepsilon)$

$$[R_n\in C]\subseteq[R_n^s\in C_\varepsilon]=[R_n^s\in C_\varepsilon]\cap P_s(\Omega).$$

Thus (4.1)' implies:

$$\overline{\lim_{n\to\infty}} n^{-1}\log\mu(R_n\in C)\leqslant\overline{\lim_{n\to\infty}} n^{-1}\log\mu(R_n^s\in C_\varepsilon)=-\inf_{\nu\in C_\varepsilon\cap P_s(\Omega)}\inf_{\rho\in\text{supp}(m)} h(\nu,\rho^N).$$

Now let $\varepsilon\downarrow 0$ and use the fact that $h(\nu,\rho^N)$ is a lowersemi continuous function to obtain (4.1),

Remarks. 1) Theoerm 4.1 still holds when F is a general compact space. The proof is exactly the same as above after the change of "$f \in \mathscr{F}_k$" by "f is $\mathscr{F}_k$-measurable and continuous".

2) When one considers the large deviation for the empirical distributions $\mu \circ D_n^{-1}$, Theorems 3.1, 4.1, and the contraction principle say that $I_0(Q) = \inf_{\rho \in \operatorname{supp}(m)} H(Q|\rho)$ is the rate function.

De Acosta has given in [1] an upper bound of the large deviation for $\mu \circ D_n^{-1}$ as following

$$I^D(Q) = \sup_{g \in C_b(F)} \inf_{\rho \in \operatorname{supp}(m)} \left(\int_F g dQ - \log \int_F e^g d\rho \right)$$

which coincides with $I_0(Q)$ in the case that $\operatorname{supp}(m)$ is convex by the minimax theorem. But in general, $I^D(Q) < I_0(Q)$. The difference between I_0 and I^D can be shown by the following example:

$$F = \{0, 1\}, \ \mu = \frac{1}{2} \delta_0^N + \frac{1}{2} \delta_1^N,$$

where $\delta_\cdot$ is the Dirac measure. In this simple case, we can give easily

$$I_0(Q) = \begin{cases} 0 & \text{if } Q = \delta_0 \text{ or } \delta_1, \\ +\infty & \text{otherwise}, \end{cases}$$

$$I^D(Q) = \sup_{g \in C_b(F)} (Q(0)g(0) + Q(1)g(1) - g(0) \vee g(1)) \equiv 0.$$

In general, $I^D(\cdot)$ is always a convex function, but $I_0(\cdot)$ is not. So Theorems 3.1 and 4.1 extend and improve much the results obtained in [1].

References

[1] de Acosta, A. Upper bounds for large deviations of dependent random variables, *Z. Wahr, verw. Gebiete*, **69**, (1985), 551—565.

[2] Föllmer, H. & Orey, S., Large deviations for the empirical field of a Gibbs measure. Preprint (1986).

[3] Orey, S. Large deviations in ergodic theory, Sem. on Probability, Birkhauser, (Ed. Cinlar), 195—247 (1984)

[4] Varadhan, S. R. S., Large deviation and applications SIAM, Philadelphia (1984).