

LARGE DEVIATIONS FOR EXCHANGEABLE OBSERVATIONS WITH APPLICATIONS

JINWEN CHEN

Received 7 January 2003

We first prove some large deviation results for a mixture of i.i.d. random variables. Compared with most of the known results in the literature, our results are built on relaxing some restrictive conditions that may not be easy to be checked in certain typical cases. The main feature in our main results is that we require little knowledge of (continuity of) the component measures and/or of the compactness of the support of the mixing measure. Instead, we pose certain moment conditions, which may be more practical in applications. We then use the large deviation approach to study the problem of estimating the component and the mixing measures.

2000 Mathematics Subject Classification: 60F10, 60G09, 62G05, 62G20.

1. Introduction. The interests in the theory of exchangeable random variables come from both its connection with that of i.i.d. random variables and its remarkable background in many applied areas such as statistics, population genetics, and so forth. From the de Finetti theorem we know that an infinitely exchangeable sequence of random variables taking values in a Polish space can be represented as a mixture of i.i.d. random variables. With the large deviation principle (LDP) for i.i.d. sequence and its many important applications in mind, it is natural to consider the LDP for exchangeable sequences. This topic was discussed in several papers; see, for example, [2, 3, 4, 7, 8] and some other references cited therein. In these papers, to obtain a *full* LDP for an exchangeable sequence, either certain continuity of the component measures or compactness of the support of the mixing measure appearing in the de Finetti representation is assumed. For example, in [7], under a condition named exponential continuity, it has been proved that if the support of the mixing measure is compact, then the corresponding sequence of exchangeable random variables satisfies a full LDP. Of course, under these conditions, LDP for a mixture of more general large deviation systems can also be derived (cf. [7]). However, we notice that such kinds of assumptions are a little too restrictive for applications. In many interesting situations, the supports of the mixing measures are not compact. We do not even know the component and/or the mixing measures, and therefore we do not know how to check the above-mentioned conditions. For these reasons we first consider, in [Section 1](#), LDP for more general exchangeable random sequences, with little knowledge of their mixing structure. The main purpose is to cover some substantially new and interesting situations. In particular, our results show that under a certain exponential tightness condition, the empirical measures of any exchangeable random field satisfy a full LDP.

Our investigation on the LDP turns out to have some significant implications to the problem of making inference on the component and/or the mixing measures which are usually unknown or contain some unknown key parameters. Suppose we have a sample, (ξ_1, \dots, ξ_n) , of exchangeable observations. A typical way to generate such a sample is as follows: first choose a distribution P_θ at random from a family $\{P_\theta, \theta \in \Theta\}$ according to certain law μ , where Θ is a space of parameters; then sample from this distribution. Thus it is easily seen that such a sample is typically suitable for making inference on the component distribution P_θ . Suppose a realization of this sample is obtained, one can, by using standard statistical methods, make inference or estimate on the distribution P_θ from which this sample comes. However, from the point of view of Bayesian statistics, it is important to make inference or estimate on the mixing measure or the prior distribution in Bayesian terminology, μ . From such a sample given above, it is not hard to find unbiased, even uniformly minimum variance unbiased estimators of some functionals of μ that play key roles in statistics. But, by the LLN for exchangeable random variables, such estimators are usually not consistent ones. Actually it is hard to find consistent estimators from a single sample. To construct a consistent estimator, a number of such samples are usually needed. From this point of view, in Section 3 we propose estimators for μ or its functionals and use a large deviation approach to show their consistency.

In Section 2 we formulate the main large deviation results followed by some examples. Section 3 is devoted to the proof of these results. In Section 4 we study the estimate problem.

2. Large deviations. We consider a mixture of general i.i.d. random fields and start with some notations. First, for a topological space Y , denote by \mathcal{B}_Y the Borel σ -field; $M_1(Y)$ is the space of all probability measures on (Y, \mathcal{B}_Y) . Now let X be a locally convex Hausdorff topological vector space with the conjugate X' and let $X_0 \subset X$ be a separable metric space in the relative topology. Let (Ω, \mathcal{F}) be a given measurable space, Θ a Hausdorff topological space, and $\{P_\theta, \theta \in \Theta\}$ a family of probability measures on (Ω, \mathcal{F}) satisfying that for each $A \in \mathcal{F}$, $\theta \rightarrow P_\theta(A)$ is measurable on $(\Theta, \mathcal{B}_\Theta)$. Let $\{\xi_t, t \in \mathbb{Z}^d\}$ ($d \geq 1$) be a family of X_0 -valued random variables on (Ω, \mathcal{F}) which are i.i.d. under each P_θ . Let μ be a Radon probability measure on $(\Theta, \mathcal{B}_\Theta)$ and define

$$P = \int P_\theta \mu(d\theta). \tag{2.1}$$

Finally, let $\{\Lambda_n, n \geq 1\}$ be a sequence of finite subsets of \mathbb{Z}^d satisfying $\Lambda_n \uparrow \mathbb{Z}^d$. Denote by $|\Lambda_n|$ the cardinality of Λ_n . For each $n \geq 1$, define

$$\bar{\xi}_n = \frac{1}{|\Lambda_n|} \sum_{t \in \Lambda_n} \xi_t \tag{2.2}$$

and let $P_\theta^n = P_\theta \circ \bar{\xi}_n^{-1}$, $P^n = P \circ \bar{\xi}_n^{-1}$ which are probability measures on (X_0, \mathcal{B}_{X_0}) . The aim of this section is to study the LDP of $\{P^n, n \geq 1\}$. The main technique we will use is the asymptotic value method provided in [1]. So we need the following exponential tightness assumption:

(H) for every $a > 0$, there is a compact set $C_a \subset X_0$, such that

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log P^n(C_a^c) \leq -a. \tag{2.3}$$

Our main result is the following.

THEOREM 2.1. *Let P^n be defined as above and let (H) hold. If for each $\lambda \in X'$,*

$$M_\theta(\lambda) = \int e^{\langle \lambda, \xi_0 \rangle} dP_\theta < \infty, \quad \mu\text{-a.s.}, \tag{2.4}$$

then $\{P^n, n \geq 1\}$ satisfies an LDP with a good rate function, that is, there is a function $I : X_0 \rightarrow [0, \infty]$ such that for every $A \in \mathcal{B}_{X_0}$,

$$-\inf_{x \in A^0} I(x) \leq \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log P^n(A) \leq \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log P^n(A) \leq -\inf_{x \in \bar{A}} I(x) \tag{2.5}$$

and I has compact level sets, that is, for any $a > 0$, $\{x : I(x) \leq a\}$ is compact in X_0 , where A^0 and \bar{A} are the interior and closure of A , respectively. In particular, if for all $\lambda \in X'$,

$$M(\lambda) \equiv \int e^{\langle \lambda, \xi_0 \rangle} dP < \infty, \tag{2.6}$$

then the above conclusions hold.

REMARK 2.2. Assumption (H) is not a restrictive one. Indeed, in our case, to obtain a full LD result with a good rate function, that is, with I having compact level sets, (H) is a necessary condition. If X_0 is compact, (H) is trivial. Some practically used sufficient condition for (H) to hold can be found in [4, 5, 6]. In the case where $X = \mathbb{R}^d$, $\sup_\theta M_\theta(\lambda) < \infty$ for some $\lambda > 0$ is sufficient for (H). In the case of a more general Banach space, see [Example 2.6](#).

REMARK 2.3. As a special case, let $X = M(E)$ be the space of all finite signed measures on some Polish space E , equipped with the weak topology, $X_0 = M_1(E)$, and let $\{\eta_t, t \in \mathbb{Z}^d\}$ be a family of E -valued random variables on (Ω, \mathcal{F}) that are i.i.d. under each P_θ , $\xi_t = \delta_{\eta_t}$ is the Dirac measure on η_t . Then under (H), $\{P((1/|\Lambda_n|) \sum_{t \in \Lambda_n} \delta_{\eta_t} \in \cdot), n \geq 1\}$ satisfies an LDP, since in this case (2.4) is automatically satisfied. In particular, if X_0 is compact, then all the conditions in [Theorem 2.1](#) are satisfied.

The following are some typical examples that satisfy the conditions of [Theorem 2.1](#).

EXAMPLE 2.4. Let $X = X_0 = \mathbb{R}$ and P_θ the Gaussian distribution with mean α and variance σ^2 , and $\theta = (\alpha, \sigma^2) \in \Theta \subset \mathbb{R} \times \mathbb{R}_+$, where \mathbb{R}_+ is the space of strictly positive real numbers. Then (2.4) is satisfied for any $\lambda \in \mathbb{R}$. In particular, if Θ is a bounded set, then (H) is also satisfied. If P_θ is a Poisson distribution with mean $\theta \in \Theta \subset \mathbb{R}_+$, then we have the same conclusions.

EXAMPLE 2.5. Let P_θ be the Bernoulli distribution with parameter $\theta \in (0, 1)$. Then all the conditions in [Theorem 2.1](#) are satisfied.

In the cases where P_θ is exponential or geometric, [\(2.4\)](#) is not satisfied. These cases will be treated in the next theorem.

EXAMPLE 2.6. The third application is to Banach-valued sequences. In this case, $X_0 = B$ is assumed to be a separable Banach space with norm $\|\cdot\|$ and $\{\xi_n, n \geq 1\}$ is an X_0 -valued sequence. $S_n = \sum_{i=1}^n \xi_i$. Suppose (i) there is a μ -null set $\Theta_0 \subset \Theta$ such that the family $\{P_\theta, \theta \in \Theta_0^c\}$ is tight, and (ii) $\| \int e^{\alpha \|\xi_1\|} dP_\theta \|_{L^\infty(\mu)} < \infty$, for all $\alpha > 0$. Note that de Acosta [\[4\]](#) proved that the family $\{P(n^{-1}S_n \in \cdot), n \geq 1\}$ is exponentially tight. Thus our [Theorem 2.1](#) and its proof imply that this family satisfies a full LDP with the good rate function given by

$$I(x) = \sup_{\lambda \in B'} [\langle \lambda, x \rangle - L(\lambda)], \tag{2.7}$$

where $L(\lambda) = \log \int \exp(\langle \lambda, \xi_1 \rangle) dP_\theta \|_{L^\infty(\mu)}$. This complements [\[4, Theorem 5.1\]](#) by providing the accompanying large deviation lower bounds, and thus the full LDP.

As pointed out in [\[7\]](#), some simple examples do not satisfy the moment condition [\(2.4\)](#) or [\(2.6\)](#). For example, let ξ_0 have exponential distribution with parameter $\theta > \theta_0 \geq 0$ under P_θ [\[7, Example 3.3\]](#). Then it is easy to check that [\(2.4\)](#) does not hold. For such cases, it has been proved in [\[7\]](#) that if Θ is compact and $\{P_\theta^n, n \geq 1, \theta \in \Theta\}$ is exponentially continuous (to be stated below), then under (H) $\{P^n, n \geq 1\}$ also satisfies an LDP. Based on this result, we can treat the case where Θ is not compact. To this end we first recall that the family $\{P_\theta^n, n \geq 1, \theta \in \Theta\}$ is said to be *exponentially continuous* if whenever $\theta_n \rightarrow \theta$, the family $\{P_{\theta_n}^n, n \geq 1\}$ satisfies an LDP. We have the following.

THEOREM 2.7. *Assume that (H) holds and that $\{P_\theta^n, n \geq 1, \theta \in \Theta\}$ is exponentially continuous. If for each $\lambda \in X'$, there exist $\delta > 0$ and a compact set $\Theta_\delta \subset \Theta$ with $\mu(\Theta_\delta) < 1$, such that*

$$\int e^{(1+\delta)\langle \lambda, \xi_0 \rangle} dP_\theta < \infty \quad \text{for } \mu\text{-almost all } \theta \in \Theta_\delta^c, \tag{2.8}$$

then $\{P^n, n \geq 1\}$ satisfies an LDP with a good rate function.

It is easy to check that [\(2.8\)](#) is satisfied in both the exponential and the geometric cases.

Using the same argument, when Θ is countable, we can prove the following.

THEOREM 2.8. *Let Θ be countable, and for each $n \geq 1$, let μ_n be a probability measure on $(\Theta, \mathcal{B}_\Theta)$ satisfying $\lim_{n \rightarrow \infty} |\Lambda_n|^{-1} \log \mu_n(\theta) = 0$ for each θ . If the family $\{P_\theta^n, n \geq 1, \theta \in \Theta\}$ satisfies that (1) for each θ , $\{P_\theta^n, n \geq 1\}$ satisfies an LDP; (2) for any $\lambda \in X'$, there exist $\delta > 0$ and a finite set $\Theta_\delta \subset \Theta$ such that [\(2.8\)](#) holds; and (3) the sequence defined by*

$$P^n = \sum_{\theta \in \Theta} P_\theta^n \mu_n(\theta), \quad n \geq 1, \tag{2.9}$$

is exponentially tight, then $\{P^n, n \geq 1\}$ satisfies an LDP.

For all the above results, X' can be replaced by any of its dense linear subspaces. In the above theorems, the LD rate function should be

$$I(x) = \sup_{f \in C_b(X_0)} [f(x) - L(f)], \tag{2.10}$$

where

$$L(f) = \lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int e^{|\Lambda_n|f} dP^n, \tag{2.11}$$

which exists for $f \in C_b(X_0)$; see Section 3. At the same time, under the conditions of any of these theorems, for each $\theta \in \Theta$, $\{P_\theta^n, n \geq 1\}$ satisfies an LDP with the rate function given by

$$I_\theta(x) = \sup_{f \in C_b(X_0)} [f(x) - L_\theta(f)], \tag{2.12}$$

which has compact level sets, where $L_\theta(f)$ is defined similarly to $L(f)$ with P^n replaced by P_θ^n . Our proofs of the theorems (Section 3) show that for some $\Theta_0 \subset \Theta$ with $\mu(\Theta_0) = 1$, $L(f) = \sup_{\theta \in \Theta_0} L_\theta(f)$, which suggests that one may consider whether we have

$$I = \inf_{\theta \in \Theta_0} I_\theta. \tag{2.13}$$

This turns out to be a minimax problem. Equation (2.13) need not be true in general. There are various conditions for a minimax theorem to hold; see [9] or [10]. Usually three types of conditions for $\Lambda(\theta, f) \equiv f(x) - L_\theta(f)$ for fixed $x \in X_0$ are involved: (i) compactness or connectivity of the domains of the two arguments of Λ ; and (ii) convexity of Λ ; (iii) certain continuity of Λ . Simple examples show that only two of these conditions are insufficient for (2.13). From the results mentioned in [9, 10], we can see that if one has compactness, then the solutions are satisfactory. The situation is very complicated in the noncompact cases. Here we will not give further discussion.

3. Proof of the large deviations. To use the asymptotic value method [1] for a measurable function f on (X_0, \mathcal{B}_{X_0}) , define

$$L_n(f) = |\Lambda_n|^{-1} \log \int e^{|\Lambda_n|f} dP^n, \quad n \geq 1, \tag{3.1}$$

and for a subset $\Theta_0 \subset \Theta$, define

$$L_{\Theta_0, n}(f) = |\Lambda_n|^{-1} \log \int_{\Theta_0} \mu(d\theta) \int e^{|\Lambda_n|f} dP_\theta^n, \quad n \geq 1. \tag{3.2}$$

The main task is to prove that for all $f \in C_b(X_0)$, the limit $\lim_{n \rightarrow \infty} L_n(f)$ exists. To this end, we will use linear functionals to approximate bounded continuous functions. Set

$$\mathcal{A} = \left\{ \min_{1 \leq i \leq m} (\langle \lambda_i, \cdot \rangle + c_i), \lambda_i \in X', c_i \in \mathbf{R}, m \geq 1 \right\}. \tag{3.3}$$

The following is the key lemma for our proofs.

LEMMA 3.1. *Let Θ_0 be a measurable subset of Θ with $\mu(\Theta_0) > 0$, $\lambda_i \in X'$, $1 \leq i \leq m$. If there exists $\delta > 0$ such that (2.8) holds for every λ_i and $\theta \in \Theta_0$, then for any $c_i \in \mathbf{R}$, $1 \leq i \leq m$, and $M > 0$, with $g = \min_{1 \leq i \leq m} (\langle \lambda_i, \cdot \rangle + c_i)$, the limit*

$$L_{\Theta_0}(g \wedge M) = \lim_{n \rightarrow \infty} L_{\Theta_0, n}(g \wedge M) \tag{3.4}$$

exists and is finite, where $a \wedge b = \min(a, b)$.

To prove this lemma, we need the following result which was proved in [1] (see [1, Lemma L.6.1 and its proof]).

LEMMA 3.2. *For fixed $n \geq 1$ and $\theta \in \Theta$. Let f_1, \dots, f_n be measurable and such that*

(1) *for any linear combination $f = \sum_{i=1}^n \alpha_i f_i$, the limit*

$$L_{\theta}(f) \equiv \lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int e^{|\Lambda_n| f} dP_{\theta} \tag{3.5}$$

exists;

(2) *$(d/dt)L_{\theta}(tf + (1 - t)g)|_{t=0}$ exists for each pair of convex combinations $f = \sum_{i=1}^n \alpha_i f_i$ and $g = \sum_{i=1}^n \beta_i f_i$.*

Then $L_{\theta}(\min_{1 \leq i \leq n} f_i)$ exists and there is a convex combination $f_0 = \sum_{i=1}^n \alpha_i^0 f_i$ such that

$$L_{\theta}\left(\min_{1 \leq i \leq n} f_i\right) = \inf \left\{ L_{\theta}\left(\sum_{i=1}^n \alpha_i f_i\right), \alpha_i \geq 0, \sum_{i=1}^n \alpha_i = 1 \right\} = L_{\theta}(f_0). \tag{3.6}$$

PROOF OF LEMMA 3.1. Write $g_i = \langle \lambda_i, \cdot \rangle + c_i$. Then under our assumptions, for each $\theta \in \Theta_0$, the limits

$$\begin{aligned} L_{\theta}\left(\sum_{i=1}^m \alpha_i g_i + \alpha_{m+1} M\right) &= \lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int e^{|\Lambda_n|(\sum_{i=1}^m \alpha_i g_i + \alpha_{m+1} M)} dP_{\theta}^n \\ &= \sum_{i=1}^m \alpha_i c_i + \alpha_{m+1} M + \log \int e^{\sum_{i=1}^m \alpha_i \langle \lambda_i, \xi_0 \rangle} dP_{\theta} \end{aligned} \tag{3.7}$$

are finite in some open convex set A_1 containing

$$A = \left\{ (\alpha_1, \dots, \alpha_{m+1}) \in \mathbf{R}^{m+1} : \alpha_i \geq 0, \sum_{i=1}^{m+1} \alpha_i = 1 \right\} \tag{3.8}$$

and for any $(\alpha_1, \dots, \alpha_{m+1})$ and (y_1, \dots, y_{m+1}) in A , if we set $f = \sum_{i=1}^m \alpha_i g_i + \alpha_{m+1} M$ and $h = \sum_{i=1}^m y_i g_i + \alpha_{m+1} M$, then $(d/dt)L_{\theta}((1 - t)f + th)|_{t=0}$ exists. Hence from Lemma 3.2 we know that there exists $g_0^{\theta} = \sum_{i=1}^m \alpha_i g_i + \alpha_{m+1} M$ with $(\alpha_1, \dots, \alpha_{m+1}) \in A$, such that $g_0^{\theta} \wedge g \wedge M = g \wedge M$ and

$$h(\theta) \equiv L_{\theta}(g \wedge M) = \lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int e^{|\Lambda_n|(g \wedge M)} dP_{\theta}^n = L_{\theta}(g_0^{\theta}). \tag{3.9}$$

$\theta \rightarrow h(\theta)$ is measurable on $(\Theta_0, \mathcal{B}_{\Theta_0})$. Since μ is a Radon measure, by Luzin's theorem, for every $k \geq 1$, there is a closed subset Θ_k of Θ_0 with $\mu(\Theta_k^c) < 1/k$, such that h is

continuous on Θ_k , and we can choose Θ_k to be increasing. Let $\Theta_\infty = \lim_{k \rightarrow \infty} \Theta_k$. Then $\mu(\Theta_\infty) = \mu(\Theta_0)$. Define $\Theta_{k,\mu} = \text{Supp}(\mu|_{\Theta_k})$, where $\mu|_{\Theta_k}$ is the restriction of μ to Θ_k . Then $\Theta_{k,\mu} \uparrow$ and $\mu(\Theta_{k,\mu}) = \mu(\Theta_k)$. We will show that

$$L_{\Theta_0}(g \wedge M) = \lim_{k \rightarrow \infty} \sup_{\theta \in \Theta_{k,\mu}} L_\theta(g \wedge M) = \sup_{\theta \in \bigcup_{k=1}^\infty \Theta_{k,\mu}} h(\theta). \tag{3.10}$$

First notice that $g \wedge M = g_0^\theta \wedge g \wedge M \leq g_0^\theta$ for each $\theta \in \Theta_0$, thus, by (2.4),

$$\begin{aligned} \limsup_{n \rightarrow \infty} L_{\Theta_0,n}(g \wedge M) &\leq \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int_{\Theta_0} \mu(d\theta) \int e^{|\Lambda_n|g_0^\theta} dP_\theta^n \\ &= \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int_{\Theta_0} \mu(d\theta) \int e^{|\Lambda_n|L_\theta(g_0^\theta)} dP_\theta^n \\ &= \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int_{\Theta_0} e^{|\Lambda_n|h(\theta)} \mu(d\theta) \\ &= \limsup_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \frac{1}{|\Lambda_n|} \int_{\Theta_k} e^{|\Lambda_n|h(\theta)} \mu(d\theta) \leq \lim_{k \rightarrow \infty} \sup_{\theta \in \Theta_{k,\mu}} h(\theta). \end{aligned} \tag{3.11}$$

Notice that under (H) the sequence $\{\int_{\Theta_0} P_\theta^n \mu(d\theta), n \geq 1\}$ is also exponentially tight; by the above inequality and [1, Lemma 5.1] we know that $l = \sup_{\theta \in \bigcup_{k=1}^\infty \Theta_{k,\mu}} h(\theta)$ is finite. Therefore for any $\varepsilon > 0$, we can choose $\theta_0 \in \bigcup_{k=1}^\infty \Theta_{k,\mu}$ such that

$$h(\theta_0) = L_{\theta_0}(g \wedge M) > l - \frac{\varepsilon}{2}. \tag{3.12}$$

Suppose $\theta_0 \in \Theta_{k,\mu}$. Then by the continuity of h on Θ_k , we can choose a neighborhood V_{θ_0} of θ_0 , such that for any $\theta \in V_{\theta_0} \cap \Theta_k$,

$$h(\theta) > h(\theta_0) - \frac{\varepsilon}{2} > l - \varepsilon. \tag{3.13}$$

Thus by Jensen's inequality, [1, Lemma 5.1], and Fatou's lemma, we see that for sufficiently large N ,

$$\begin{aligned} \liminf_{n \rightarrow \infty} L_{\Theta_0,n}(g \wedge M) &= \liminf_{n \rightarrow \infty} L_{\Theta_0,n}((g \wedge M) \vee (-N)) \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int_{V_{\theta_0} \cap \Theta_k} \mu(d\theta) \int e^{|\Lambda_n|[(g \wedge M) \vee (-N)]} dP_\theta^n \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{\mu(V_{\theta_0} \cap \Theta_k)} \frac{1}{|\Lambda_n|} \int_{V_{\theta_0} \cap \Theta_k} \left(\log \int e^{|\Lambda_n|[(g \wedge M) \vee (-N)]} dP_\theta^n \right) \mu(d\theta) \\ &\geq \frac{1}{\mu(V_{\theta_0} \cap \Theta_k)} \int_{V_{\theta_0} \cap \Theta_k} \left(\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int e^{|\Lambda_n|(g \wedge M)} dP_\theta^n \right) \mu(d\theta) \\ &= \frac{1}{\mu(V_{\theta_0} \cap \Theta_k)} \int_{V_{\theta_0} \cap \Theta_k} h(\theta) \mu(d\theta) > l - \varepsilon. \end{aligned} \tag{3.14}$$

Hence $\liminf_{n \rightarrow \infty} L_{\Theta_0,n}(g \wedge M) \geq l$. Combining this with (2.8), we obtain (2.6), proving the lemma. □

PROOF OF THEOREM 2.1. Let $g = \min_{1 \leq i \leq m} (\langle \lambda_i, \cdot \rangle + c_i) \in \mathcal{A}$ and $M > 0$. Then for any $\delta > 0$, by assumption, there is $\Theta_0 \subset \Theta$ with $\mu(\Theta_0) = 1$ such that (2.8) holds for all $\theta \in \Theta_0$. Thus, by Lemma 3.1, the limit $L(g \wedge M) = \lim_{n \rightarrow \infty} L_n(g \wedge M) = L_{\Theta_0}(g \wedge M)$ exists and is finite. Therefore from the proof of [1, Theorem T.1.3], we see that for all $f \in C_b(X_0)$, the limit

$$L(f) = \lim_{n \rightarrow \infty} L_n(f) \tag{3.15}$$

exists and is finite. Hence by [1, Theorem T.1.2], $\{P^n, n \geq 1\}$ satisfies an LDP with the good rate function given by

$$I(x) = \sup_{f \in C_b(X_0)} [f(x) - L(f)]. \tag{3.16}$$

□

PROOF OF THEOREM 2.7. From the proof of Theorem 2.1, we know that it suffices to show that for all $g \in \mathcal{A}$ and $M > 0$, the limit $L(g \wedge M)$ exists and is finite. To do this, let $g_i = \langle \lambda_i, \cdot \rangle + c_i$ and $g = g_1 \wedge \dots \wedge g_m$. Then under the assumptions of the theorem, there exist $\delta > 0$ and a compact set $\Theta_\delta \subset \Theta$ with $\mu(\Theta_\delta) < 1$, and a $\Theta_0 \subset \Theta_\delta^c$ with $\mu(\Theta_0) = \mu(\Theta_\delta^c)$ such that (2.8) holds for all $\theta \in \Theta_0$. Then the lemma implies that the limit $L_{\Theta_\delta^c}(g \wedge M) = L_{\Theta_0}(g \wedge M)$ exists and is finite. By the exponential continuity of $\{P_\theta^n, n \geq 1, \theta \in \Theta\}$ and [4, Theorems 2.1 and 2.2], we know that the sequence $\{\int_{\Theta_\delta} P_\theta^n \mu(d\theta), n \geq 1\}$ satisfies an LDP. Hence for any $f \in C_b(X_0)$, the limit

$$L_{\Theta_\delta}(f) = \lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int_{\Theta_\delta} \mu(d\theta) \int e^{|\Lambda_n|f} dP_\theta^n \tag{3.17}$$

exists and is finite. Therefore $L_{\Theta_\delta}(g \wedge M) \vee (-N)$ exists and is finite for any $N > 0$. Notice that $\{\int_{\Theta_\delta} P_\theta^n \mu(d\theta), n \geq 1\}$ is also exponentially tight; again by [1, Lemma 5.1] we know that for sufficiently large N ,

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \left[\log \int_{\Theta_\delta} \mu(d\theta) \int e^{|\Lambda_n|(g \wedge M) \vee (-N)} dP_\theta^n - \log \int_{\Theta_\delta} \mu(d\theta) \int e^{|\Lambda_n|(g \wedge M)} dP_\theta^n \right] = 0. \tag{3.18}$$

This means that the limit $L_{\Theta_\delta}(g \wedge M)$ exists and is finite. Thus

$$\begin{aligned} L(g \wedge M) &= \lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \left[\int_{\Theta_\delta} \mu(d\theta) \int e^{|\Lambda_n|(g \wedge M)} dP_\theta^n + \int_{\Theta_0} \mu(d\theta) \int e^{|\Lambda_n|(g \wedge M)} dP_\theta^n \right] \\ &= (L_{\Theta_\delta}(g \wedge M)) \vee (L_{\Theta_0}(g \wedge M)) \end{aligned} \tag{3.19}$$

exists and is finite. □

Finally we prove Theorem 2.8. Let $g_i = \langle \lambda_i, \cdot \rangle + c_i, 1 \leq i \leq m, M > 0, g = g_1 \wedge \dots \wedge g_m$. Then by our assumptions and the proof of the lemma, there is a finite set $\Theta_\delta \subset \Theta$, such that for any $\theta \in \Theta_\delta^c$, the limit

$$L_\theta(g \wedge M) = \lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int e^{|\Lambda_n|(g \wedge M)} dP_\theta^n \tag{3.20}$$

exists and is finite, and there is a $g_0^\theta = \sum_{i=1}^m \alpha_i g_i + \alpha_{m+1} M$ with $\alpha_i \geq 0$, $\sum_{i=1}^{m+1} \alpha_i = 1$, such that $L_\theta(g \wedge M) = L_\theta(g_0^\theta)$, $g_0^\theta \wedge g \wedge M = g \wedge M$. Thus for each $\theta \in \Theta_\delta^c$,

$$\liminf_{n \rightarrow \infty} L_{\Theta_\delta^c, n} \geq \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \mu_n(\theta) \int e^{|\Lambda_n|(g \wedge M)} dP_\theta^n = L_\theta(g \wedge M); \tag{3.21}$$

that is,

$$\liminf_{n \rightarrow \infty} L_{\Theta_\delta^c, n}(g \wedge M) \geq \sup_{\theta \in \Theta_\delta^c} L_\theta(g \wedge M). \tag{3.22}$$

On the other hand,

$$\begin{aligned} \limsup_{n \rightarrow \infty} L_{\Theta_\delta^c, n}(g \wedge M) &= \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \sum_{\theta \in \Theta_\delta^c} \mu_n(\theta) \int e^{|\Lambda_n|(g_0^\theta \wedge g \wedge M)} dP_\theta^n \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \sum_{\theta \in \Theta_\delta^c} \mu_n(\theta) \int e^{|\Lambda_n|g_0^\theta} dP_\theta^n \\ &= \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \sum_{\theta \in \Theta_\delta^c} \mu_n(\theta) e^{|\Lambda_n|L_\theta(g \wedge M)} \leq \sup_{\theta \in \Theta_\delta^c} L_\theta(g \wedge M). \end{aligned} \tag{3.23}$$

Hence the limit

$$L_{\Theta_\delta^c}(g \wedge M) = \lim_{n \rightarrow \infty} L_{\Theta_\delta^c, n}(g \wedge M) = \sup_{\theta \in \Theta_\delta^c} L_\theta(g \wedge M) \tag{3.24}$$

exists and is finite. For $\theta \in \Theta_\delta$, since $\{P_\theta^n, n \geq 1\}$ is exponentially tight and satisfies an LDP, by [1, lemma 5.1], for large N ,

$$L_\theta(g \wedge M) = \lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log \int e^{|\Lambda_n|(g \wedge M)} dP_\theta^n \tag{3.25}$$

exists and is equal to $L_\theta((g \wedge M) \vee (-N))$, hence finite. Therefore, the limit

$$L_{\Theta_\delta}(g \wedge M) = \lim_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \sum_{\theta \in \Theta_\delta} \mu_n(\theta) \int e^{|\Lambda_n|(g \wedge M)} dP_\theta^n = \max_{\theta \in \Theta_\delta} L_\theta(g \wedge M) \tag{3.26}$$

exists and is finite and, furthermore,

$$L(g \wedge M) = (L_{\Theta_\delta}(g \wedge M)) \vee (L_{\Theta_\delta^c}(g \wedge M)) \tag{3.27}$$

exists and is finite, completing the proof.

4. Estimate of the mixing measure. Now we turn to the study of estimating the mixing measure μ or some of its functionals such as expectation and moments. We adopt almost all the setting given in Section 2, except that we replace the set of indices \mathbb{Z}^d with \mathbb{N} , that is, we are considering infinitely exchangeable observations $\{\xi_n, n \geq 1\}$,

simply to make the notations simple. Let $\mu_\theta = P_\theta \circ \xi_1^{-1}$ be the law of ξ_1 on X_0 under P_θ . Let T be the mapping from Θ to $M_1(X_0)$ with $T(\theta) = \mu_\theta$, where $M_1(X_0)$ is the space of probability measures on X_0 . Denote $T(\mu) = \mu \circ T^{-1}$. We consider the problem of estimating $T(\mu)$, instead of μ , since we can, if needed, make a certain change of the space of parameters so that the mixing measure is $T(\mu)$. Let $\{\xi_{i,j}, 1 \leq j \leq n\}, 1 \leq i \leq n$, be n independent copies of $\{\xi_i, 1 \leq i \leq n\}$. We still denote by P the underlying probability law. Define, for $n \geq 1$,

$$L_{n,i} = \frac{1}{n} \sum_{j=1}^n \delta_{\xi_{i,j}}, \quad 1 \leq i \leq n, \quad R_n = \frac{1}{n} \sum_{i=1}^n \delta_{L_{n,i}}. \tag{4.1}$$

The following result implies that R_n is a (strongly) consistent estimator of $T(\mu)$.

THEOREM 4.1. *Assume (H). Then $\lim_{n \rightarrow \infty} R_n = T(\mu)$ a.s. in the sense of weak convergence.*

REMARK 4.2. From this theorem it is easily seen that if $\phi \in C_b(M_1(X_0))$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \phi(L_{n,i}) = \int \phi dT(\mu) = \int \phi(\mu_\theta) \mu(d\theta) \quad \text{a.s.} \tag{4.2}$$

In particular, if $\phi \in C_b(X_0)$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i,j=1}^n \phi(\xi_{i,j}) = \int \mu_\theta(\phi) \mu(d\theta) \quad \text{a.s.} \tag{4.3}$$

This suggests consistent estimators for expectation or more general moments of μ .

Obviously, the above theorem is a consequence of the following large deviation result.

THEOREM 4.3. *Assume (H). Then $\{Q_n = P \circ R_n^{-1}, n \geq 1\}$ satisfies a full LDP on $M_1(X_0)$ endowed with the weak topology, with the good rate function J given by $J(R) = H(R, T(\mu))$, the relative entropy of R with respect to $T(\mu)$.*

The proof of [Theorem 4.3](#) is completed with the following three lemmas.

LEMMA 4.4. *Let $L_n = (1/n) \sum_{i=1}^n \delta_{\xi_i}$. Then $\lim_{n \rightarrow \infty} P \circ L_n^{-1} = T(\mu)$ weakly.*

PROOF. For each $\theta \in \Theta$, by applying Sanov’s theorem, it is not hard to prove that $\lim_{n \rightarrow \infty} P_\theta \circ L_n^{-1} = \delta_{\mu_\theta}$ weakly. Thus for any $f \in C_b(M_1(X_0))$,

$$\lim_{n \rightarrow \infty} \int f dP \circ L_n^{-1} = \lim_{n \rightarrow \infty} \int \mu(d\theta) \int f dP_\theta \circ L_n^{-1} = \int f(\mu_\theta) \mu(d\theta) = \int f dT(\mu), \tag{4.4}$$

proving the lemma. □

LEMMA 4.5. For any $f \in C_b(M_1(X_0))$, the limit

$$\Lambda(f) \equiv \lim_{n \rightarrow \infty} \frac{1}{n} \log \int e^{nf(v)} Q_n(dv) \tag{4.5}$$

exists and equals $\log \int e^f dT(\mu)$. Moreover, for any pair f and g in $C_b(M_1(X_0))$, $t \rightarrow \Lambda(tf + (1-t)g)$ is differentiable.

PROOF. The first assertion is a consequence of Lemma 4.4, and thus the second assertion easily follows. □

LEMMA 4.6. Assume (H). Then $\{Q_n, n \geq 1\}$ is exponentially tight on $M_1(X_0)$.

PROOF. For any $a > 0$, from (H) we know that there is a compact subset C_a of $M_1(X_0)$ such that

$$P(L_n \in C_a^c) \leq e^{-an} \quad \forall n \geq 1. \tag{4.6}$$

Now define

$$K_a = \left\{ R \in M_1(M_1(X_0)) : R(C_{ak}^c) \leq \frac{a}{k}, \forall k \geq 1 \right\}. \tag{4.7}$$

Then K_a is compact in $M_1(M_1(X_0))$ and, for each $1 \leq i \leq n$,

$$\begin{aligned} P(\delta_{L_{n,i}} \in K_a^c) &= P(\delta_{L_n} \in K_a^c) = P\left(\exists k \geq 1 \mid \delta_{L_n}(C_{ak}^c) \geq \frac{a}{k}\right) \\ &= P(\exists k \geq 1 \mid L_n \in C_{ak}^c) \leq \frac{e^{-an}}{1 - e^{-an}} \end{aligned} \tag{4.8}$$

by (4.6). Note that K_a is convex. We have

$$P(R_n \in K_a^c) \leq \sum_{i=1}^n P(\delta_{L_{n,i}} \in K_a^c) \leq \frac{ne^{-an}}{1 - e^{-an}}, \tag{4.9}$$

which implies the exponential tightness of $\{Q_n, n \geq 1\}$. □

Now the proof of Theorem 4.3 is completed by applying [1, Theorem T.1.4], with the rate function being given by

$$J(R) = \sup_{f \in C_b(M_1(X_0))} \left[\int f dR - \Lambda(f) \right] = \sup_{f \in C_b(M_1(X_0))} \left[\int f dR - \log \int e^f dT(\mu) \right] \tag{4.10}$$

which is just $H(R, T(\mu))$.

PROOF OF THEOREM 4.1. Let $d(\cdot, \cdot)$ be any metric on $M_1(M_1(X_0))$ that generates the weak topology. Then from Theorem 4.1 we know that for any $\epsilon > 0$, there is a constant $\alpha_\epsilon > 0$ such that

$$P(d(R_n, T(\mu)) \geq \epsilon) \leq e^{-\alpha_\epsilon n} \quad \forall n \geq 1. \tag{4.11}$$

Then a standard argument gives the desired conclusion, completing the proof of Theorem 4.1. □

ACKNOWLEDGMENT. This research was supported by the National Natural Science Foundation of China (NSFC 10371063).

REFERENCES

- [1] W. Bryc, *Large deviations by the asymptotic value method*, Diffusion Processes and Related Problems in Analysis, Vol. I (Evanston, Ill, 1989) (M. Pinsky, ed.), Progr. Probab., vol. 22, Birkhäuser Boston, Massachusetts, 1990, pp. 447-472.
- [2] T. Daras, *Large and moderate deviations for the empirical measures of an exchangeable sequence*, Statist. Probab. Lett. **36** (1997), no. 1, 91-100.
- [3] ———, *Trajectories of exchangeable sequences: large and moderate deviations results*, Statist. Probab. Lett. **39** (1998), no. 4, 289-304.
- [4] A. de Acosta, *Upper bounds for large deviations of dependent random vectors*, Z. Wahrsch. Verw. Gebiete **69** (1985), no. 4, 551-565.
- [5] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett Publishers, Massachusetts, 1993.
- [6] J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Pure and Applied Mathematics, vol. 137, Academic Press, Massachusetts, 1989.
- [7] I. H. Dinwoodie and S. L. Zabell, *Large deviations for exchangeable random vectors*, Ann. Probab. **20** (1992), no. 3, 1147-1166.
- [8] ———, *Large deviations for sequences of mixtures*, Statistics and Probability: A Raghu Raj Bahadur Festschrift (J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy, and B. L. S. Prakasa Rao, eds.), Wiley Eastern, New Delhi, 1993, pp. 171-190.
- [9] B. Ricceri and S. Simons (eds.), *Minimax Theory and Applications*, Nonconvex Optimization and Its Applications, vol. 26, Kluwer Academic Publishers, Dordrecht, 1998.
- [10] S. Simons, *Minimax theorems and their proofs*, Minimax and Applications (D.-Z. Du and P. M. Pardalos, eds.), Nonconvex Optim. Appl., vol. 4, Kluwer Academic Publishers, Dordrecht, 1995, pp. 1-23.

Jinwen Chen: Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China
E-mail address: jchen@math.tsinghua.edu.cn