The Annals of Probability
2001, Vol. 29, No. 3, 1303–1328

STOCHASTIC SUB-ADDITIVITY APPROACH TO THE CONDITIONAL LARGE DEVIATION PRINCIPLE

BY ZHIYI CHI

University of Chicago

Given two Polish spaces A_X and A_Y, let $\rho : A_X \times A_Y \to \mathbb{R}^d$ be a bounded measurable function. Let $X = \{X_n : n \geq 1\}$ and $Y = \{Y_n : n \geq 1\}$ be two independent stationary processes on A_X^∞ and A_Y^∞, respectively. The article studies the large deviation principle (LDP) for $n^{-1}\sum_{k=1}^n \rho(X_k, Y_k)$, conditional on X. Based on a stochastic version of approximate sub-additivity, it is shown that if Y satisfies certain mixing condition, then for almost all random realization x of X, the laws of $n^{-1}\sum_{k=1}^n \rho(x_k, Y_k)$ satisfy the conditional LDP with a non-random convex rate function. Conditions for the rate function to be non-trivial (that is, not $0/\infty$ function) are also given.

1. Introduction. This article aims to establish the conditional large deviation principle (LDP) for the partial sums of $\mathbb{R}^d$–valued functions of general processes. Given two Polish spaces A_X and A_Y, that is, metrizable complete separable topological spaces, suppose $X = \{X_n; n \in \mathbb{Z}\}$ and $Y = \{Y_n; n \in \mathbb{Z}\}$ are two independent stationary processes taking values in $(A_X^{\mathbb{Z}}, \mathscr{F}_X)$ and $(A_Y^{\mathbb{Z}}, \mathscr{F}_Y)$, respectively. Let $P = \text{dist}(X)$ and $Q = \text{dist}(Y)$. For the process X, denote by $\sigma(X_i^j)$ the σ–field generated by $X_i^j = (X_i, \dots, X_j)$ and likewise for Y. Given a bounded measurable function $\rho : A_X \times A_Y \to \mathbb{R}^d$, we are interested in the LDP of

$$\rho_n(x_1^n, Y_1^n) = \frac{1}{n}\sum_{k=1}^n \rho(x_k, Y_k), \tag{1.1}$$

given a random realization $x = \{x_n; n \in \mathbb{Z}\}$ of X. Because x is fixed in the partial sum once it is chosen randomly, the conditional LDP of the partial sum sometimes is referred to as the "quenched" LDP.

Our interest in the quenched LDP for $\rho_n(X_1^n, Y_1^n)$ largely comes from the asymptotics of waiting times between stationary processes, which are important to data compression based on string matching [10, 9, 8, 11]. When ρ is a bounded real valued non-negative function, $\rho_n(x_1^n, y_1^n)$ is termed the distortion between $x_1^n \in A_X^n$ and $y_1^n \in A_Y^n$. In lossy data compression, $y = \{y_n; n \geq 1\}$ is taken as a randomly generated code book, and x_1^n is encoded as the smallest k such that the distortion between x_1^n and y_k^{n+k-1} is no more than a given value. It was established in several places [7, 11, 4] that if the distortion is restricted to be less than D, then using a code book randomly generated from Y, the

Received August 2000; revised January 2001.
AMS 2000 *subject classifications*. Primary 60F10; secondary 94A34.
Key words and phrases. Conditional large deviation principle, stochastic approximate sub-additivity, mixing conditions.

compression rate for the initial segment of length n of a random realization of X is asymptotically equal to the P-a.s. limit of

$$\frac{1}{n}\log Q\{Y : \rho_n(X_1^n, Y_1^n) \le D\}. \tag{1.2}$$

The P-a.s. limit of (1.2) was first studied in [11, 8], with A_X and A_Y finite, and Y either an i.i.d. process or an irreducible Markov chain. The large deviations approach to (1.2) was initiated by Dembo and Kontoyiannis [4]. Assuming A_X and A_Y to be general Polish spaces and Y an i.i.d. process, they proved the P–almost sure convergence of (1.2) to a limit in terms of D. The main gradient in their proof was the standard change of measure combined with the central limit theorem. With different methods, similar results were established in [12]. In [3], the P–almost sure convergence of (1.2) was generalized to the case where Y is ψ–mixing. The method there was to divide X into disjoint blocks, and, by ψ–mixing, treat the blocks as vector-valued independent random variables, making it possible to apply change of measure and the central limit theorem to establish the limit.

We will study the conditional LDP of $\rho_n(X_1^n, Y_1^n)$ under a more general mixing condition for Y, namely condition (S) (see Definition 1). The implication of condition (S) to LDP was first studied in [2]. We combine the asymptotic value method of Bryc (1990) and Gärtner-Ellis theorem to approach the LDP. Specifically, the asymptotic value method is used for the lower bound of the LDP, while Gärtner-Ellis theorem is used for the upper bound.

The key to the conditional LDP for $\rho_n(X_1^n, Y_1^n)$ by the above combined methods is a stochastic version of Hammersley's approximate sub-additivity [6]. The following result on "stochastic approximate sub-additivity" is the basis for the other results in this article.

THEOREM 1. *Let T be an ergodic measure-preserving transformation on a probability space $(\Omega, \mathscr{F}, P)$. Let $h_n : \Omega \to \mathbb{R}$, $n \ge 1$ be a sequence of measurable functions satisfying two conditions*:

(i) *There exists a non-decreasing sequence $\Delta(n) \ge 0$ with*

$$\sum_{k=1}^{\infty} \frac{\Delta(k)}{k(k+1)} < \infty, \tag{1.3}$$

such that

$$h_{n+m}(\omega) \le h_n(\omega) + h_m(T^n\omega) + \Delta(n+m), \qquad P\text{-a.s.} \tag{1.4}$$

(ii) $h_n \in L_1(P)$, $n \ge 1$ *and*

$$\lim_{n\to\infty} \frac{h_n(\omega) - h_n(T\omega)}{n} = 0, \qquad P\text{-a.s.} \tag{1.5}$$

Then $\lim_{n\to\infty} E[h_n(\omega)/n]$ *exists and*

$$\limsup_{n\to\infty} \frac{h_n(\omega)}{n} = \lim_{n\to\infty} E\left[\frac{h_n(\omega)}{n}\right], \qquad P\text{-a.s.} \tag{1.6}$$

where E is expectation under P.

Most part of the conditional LDP can be proved without assuming independence between processes. To see this, let $(\Omega, \mathscr{F}, \nu) = (A_X^{\mathbb{Z}} \times A_Y^{\mathbb{Z}}, \mathscr{F}_X \times \mathscr{F}_Y, P \times Q)$. Denoting by ω a generic element of Ω, then $\omega = (x, y)$, with $x = \{x_n : n \in \mathbb{Z}\} \in A_X^{\mathbb{Z}}$ and $y = \{y_n : n \in \mathbb{Z}\} \in A_Y^{\mathbb{Z}}$. Regard ρ as a function $\Omega \to \mathbb{R}^d$, such that $\rho(\omega) = \rho(x_1, y_1)$. Denote by T the shift operator, which maps ω to $\omega' = (x', y')$, with $x'_n = x_{n+1}$, $y'_n = y_{n+1}$, $n \in \mathbb{Z}$. T is not only measure-preserving, but also one-to-one, and T^{-1} is measurable as well. Also, regarding X as a measurable function $\Omega \to A_X^{\mathbb{Z}}$ such that $X(\omega) = x$, the σ-algebra generated by X, denoted $\sigma(X)$, is a sub-algebra of $\mathscr{F}$ and is closed under T and T^{-1}. Finally, T is ergodic on $(\Omega, \sigma(X), \nu)$.

To take into account the randomness involved in the conditional LDP, condition (S) used here is modified from the original one in [2] (see Definition 2). It still consists of two parts, one is condition (S_-), the other one condition (S_+). Using Theorem 1, we can prove the following statement which only needs (S_-).

THEOREM 2. *Let T be a one-to-one measure-preserving transformation on a probability space $(\Omega, \mathscr{F}, P)$, and assume T^{-1} is measurable. Given $\mathscr{B} \subset \mathscr{F}$, suppose for any $B \in \mathscr{B}$, $T(B), T^{-1}(B) \in \mathscr{B}$ and T is ergodic on $\mathscr{B}$. Let $\rho : \Omega \to \mathbb{R}^d$ be a bounded measurable function. Denote*

$$\rho_n(\omega) = \frac{1}{n} \sum_{k=1}^{n} \rho(T^k \omega).$$

If the process $\{\rho \circ T^n : n \in \mathbb{Z}\}$ satisfies condition (S_-) uniformly, conditional on $\mathscr{B}$, then for any continuous, bounded above, concave function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\Lambda_f = \lim_{n\to\infty} E\left[\frac{1}{n} \log E\left[e^{nf(\rho_n(\cdot))} | \mathscr{B}\right]\right] \tag{1.7}$$

exists and

$$\liminf_{n\to\infty} \frac{1}{n} \log E\left[e^{nf(\rho_n(\cdot))} | \mathscr{B}\right] = \Lambda_f, \qquad P\text{-a.s.} \tag{1.8}$$

Consequently, for $\lambda \in \mathbb{R}^d$,

$$\Lambda(\lambda) = \lim_{n\to\infty} E\left[\frac{1}{n} \log E\left[e^{n\langle \lambda, \rho_n(\cdot)\rangle} | \mathscr{B}\right]\right] \tag{1.9}$$

exists.

By Bryc's inverse Varadhan lemma, Theorem 2 suggests that if the laws of ρ_n conditional on $\mathscr{B}$ satisfy a conditional LDP, then the associated rate function should be $I(u) = \sup_{f \in C(\Gamma)}\{f(u) - \Lambda_f\}$. This however can not be proved by directly applying the asymptotic value method, because (1.8) only asserts the existence of lim inf of $n^{-1} \log E[e^{nf(\rho_n(\cdot))} | \mathscr{B}]$. Despite this, we can first establish a weaker result. Denote by $C_b(\mathbb{R}^d)$ the space of bounded continuous functions on $\mathbb{R}^d$.

THEOREM 3. *Let* $\mu_n(du|\omega) \stackrel{\Delta}{=} \mu_n(du \mid \mathscr{B})(\omega)$ *be the regular version of the conditional probability measure on* $\mathbb{R}^d$ *induced by* ρ_n. *If for any continuous, bounded above, concave function* $f : \mathbb{R}^d \to \mathbb{R}$, *the following limits exist*:

$$\Lambda_f = \lim_{n\to\infty} E\left[\frac{1}{n}\log\left\{\int e^{nf(u)}\mu_n(du|\omega)\right\}\right] \tag{1.10}$$

and

$$\liminf_{n\to\infty} \frac{1}{n}\log\left\{\int e^{nf(u)}\mu_n(du|\omega)\right\} = \Lambda_f, \qquad P\text{-a.s.}; \tag{1.11}$$

then there exists a sequence n_i *and* $Z \subset \Omega$ *with* $P(Z) = 0$, *such that for* $\omega \notin Z$ *and* $f \in C_b(\mathbb{R}^d)$,

$$\Lambda_f = \lim_{i\to\infty} \frac{1}{n_i}\log\left\{\int e^{n_i f(u)}\mu_{n_i} E(du|\omega)\right\} \tag{1.12}$$

exists, *directly leading to*

$$\Lambda_f = \lim_{i\to\infty} E\left[\frac{1}{n_i}\log\left\{\int e^{n_i f(u)}\mu_{n_i}(du|\omega)\right\}\right]. \tag{1.13}$$

Finally,

$$\liminf_{n\to\infty} \frac{1}{n}\log\left\{\int e^{nf(u)}\mu_n(du|\omega)\right\} \geq \Lambda_f, \qquad f \in C_b(\mathbb{R}^d),\ \omega \notin Z. \tag{1.14}$$

REMARK. Regular versions of the conditional probability measures $\mu_n(du|\omega)$ always exist if μ_n are defined on the Borel σ-algebra of $\mathbb{R}^d$ ([1], pages 77–80). Also, because ρ is bounded, the support of μ_n, $n \geq 1$, is uniformly bounded.

By Theorem 3, we can get a constant lower bound for the P–almost sure first order asymptotic of $\mu_n(G|\omega)$, with G open, hence proving the lower bound of the conditional LDP for ρ_n.

COROLLARY 1. *Let*

$$I(u) = \sup_{f\in C_b(\mathbb{R}^d)} \{f(u) - \Lambda_f\}. \tag{1.15}$$

If (1.14) *holds*, *then*

$$\liminf_{n\to\infty} \frac{1}{n}\log \mu_n(G|\omega) \geq -I(u), \qquad \omega \notin Z,\ G \subset \mathbb{R}^d \text{ open},\ u \in G. \tag{1.16}$$

Turning to the upper bound for the conditional LDP of ρ_n, because of the absence of a limit for (1.11), it is not clear how to modify the argument of the asymptotic value method to get a constant upper bound in terms of I. To get around this difficulty, we adopt the convexity argument of Gärtner-Ellis theorem, which requires condition (S_+).

THEOREM 4. *Fix T, $\mathscr{B}$ and ρ as in Theorem 2. Assume the process $\{\rho \circ T^n : n \in \mathbb{Z}\}$ satisfies condition (S_+) uniformly, conditional on $\mathscr{B}$. Then $\Lambda(\lambda)$ defined by (1.9) exists for all $\lambda \in \mathbb{R}^d$. In addition, there is a set $Z \subset \Omega$ with $P(Z) = 0$, such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log E\left[e^{n\langle\lambda,\rho_n(\cdot)\rangle}|\mathscr{B}\right](\omega) = \Lambda(\lambda), \qquad \omega \notin Z,\ \lambda \in \mathbb{R}^d. \tag{1.17}$$

The proof for Theorem 4 follows closely the one for Theorem 2 and also uses the stochastic approximate sub-additivity. By the argument of Gärtner-Ellis theorem we have:

COROLLARY 2. *With $\mu_n(du|\omega)$ defined as in Corollary 1, suppose there is $Z \subset \Omega$ with $P(Z) = 0$, such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log\left\{\int e^{n\langle\lambda,u\rangle}\mu_n(du|\omega)\right\} = \Lambda(\lambda), \qquad \lambda \in \mathbb{R}^d,\ \omega \notin Z.$$

Then for any compact set $F \subset \mathbb{R}^d$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F|\omega) \le -\inf_{u\in F}\{\Lambda^*(u)\}, \qquad \omega \notin Z, \tag{1.18}$$

where $\Lambda^*(u) = \sup_{\lambda\in\mathbb{R}^d}\{\langle\lambda, u\rangle - \Lambda(\lambda)\}$ is the Fenchel-Legendre transform of $\Lambda(\lambda)$.

Returning back to the original problem, that is, the quenched LDP for $\rho_n(X_1^n, Y_1^n)$, we see that since ρ is bounded, given random realization x of X, the laws for $\rho_n(x_1^n, Y_1^n)$, $n \ge 1$, are exponentially tight. Therefore, (1.18) holds for arbitrary closed set F. To complete the proof of the quenched LDP for ρ_n, we finally need to demonstrate $\Lambda^* = I$. By Varadhan's integral lemma, it is enough to show that I is convex. Because the convexity of I requires some extra work than the non-stochastic case, we present it as a theorem.

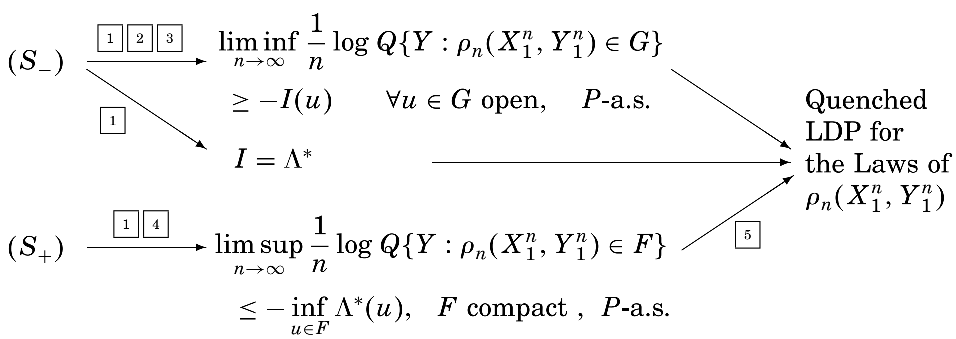
THEOREM 5. *Let $\mu_n(du|x)$, $x \in A_X^{\mathbb{Z}}$, be the conditional probability measures induced by $\rho_n(x_1^n, Y_1^n)$ on $\mathbb{R}^d$. If Y satisfies condition (S_-), then I defined by (1.15) is a convex good rate function, and hence $I = \Lambda^*$.*

Combining the above results, we get:

PROPOSITION 1. *Suppose X and Y are two independent stationary processes on $A_X^{\mathbb{Z}}$ and $A_Y^{\mathbb{Z}}$, respectively. Let X be ergodic and Y satisfy condition (S). Then given bounded measurable function $\rho : A_X \times A_Y \to \mathbb{R}^d$, for almost all random realization $x \in A_X^{\mathbb{Z}}$ of X, $\rho_n(x_1^n, Y_1^n)$ given by (1.1) satisfies the LDP with good rate function $I = \Lambda^*$ with*

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n} E_X\left[\log E_Y\left[e^{\lambda\rho_n(X_1^n,Y_1^n)}\right]\right]. \tag{1.19}$$

The following diagram summarizes the steps to prove the quenched LDP for $\rho_n(x_1^n, Y_1^n)$, for Y satisfying condition (S). Main ingredients of the proof are indicated by numbers in boxes.



1. Stochastic approximate sub-additivity; 2. Bryc's asymptotic value method; 3. Separability of $C(\Gamma)$, for any compact $\Gamma \subset \mathbb{R}^d$; 4. Gärtner-Ellis theorem; 5. Boundedness of $\rho \Rightarrow$ exponential tightness.

The rate function Λ^* in Proposition 1 can also be given in terms of relative entropy (Corollary 4, [7]; Proposition 1, [4]; Property 1, [12]). Also, following the argument in [2], the conditional LDP for the empirical measures of $\{\rho(X_n, Y_n) : n \in \mathbb{Z}\}$ could be proved without much difficulty.

Next we consider the functional property of the convex rate function I. As mentioned earlier, I is a good rate function. In order to see whether the quenched LDP gives any interesting information, we would like to investigate whether I is non-trivial, that is, not a $0/\infty$ function. When Y is i.i.d. and X is an arbitrary stationary process, it is easy to demonstrate that under some minimal conditions, I is non-trivial. To get such rate function when Y only satisfies condition (S_-), we shall consider the case where X be is an i.i.d. process. For simplicity, let ρ be $\mathbb{R}$–valued. It is not hard to see that if either $\Lambda'(0+) < \Lambda'(\infty) \triangleq \lim_{\lambda\to\infty} \Lambda'(\lambda+)$ or $\Lambda'(0-) > \Lambda'(-\infty) \triangleq \lim_{\lambda\to-\infty} \Lambda'(\lambda-)$, then $I = \Lambda^*$ is non-trivial.

To get the local property of Λ at 0, consider the "mean" process $\{\bar\rho(Y_n); n \ge 1\}$, where $\bar\rho(y) = E_X[\rho(X_1, y)]$, $y \in A_Y$. Because Y satisfies condition (S_-), so does $\{\bar\rho(Y_n); n \ge 1\}$, implying it satisfies the LDP with good rate function

$$\bar\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n} \log E_Y\left[\exp\left\{\lambda \sum_{i=1}^{n} \bar\rho(Y_i)\right\}\right] \tag{1.20}$$

It turns out that $\bar\Lambda \le \Lambda$ and the two functions have the same local property at 0 (Proposition 2). This leads to the following:

THEOREM 6. *Fix X, Y and ρ as in Proposition* 1. *Let X be an i.i.d. process and ρ be an $\mathbb{R}$–valued function. Denote $P_1 = \mathrm{dist}(X_1)$ and $Q_1 = \mathrm{dist}(Y_1)$. Suppose that Q_1–almost surely, the P_1–measure of the set of discontinuity points of $f_{Y_1}(x) \stackrel{\Delta}{=} \rho(x, Y_1)$ is* 0. *Define $\bar{\Lambda}$ by* (1.20) *and Λ by* (1.19). *If $\bar{\Lambda}^*$ is non-trivial, then Λ^* is also non-trivial.*

When $\bar{\Lambda}^*$ is trivial, we need to further exploit the assumption that X is i.i.d. Consider the case where the mean process $\{\bar{\rho}(Y_n); n \geq 1\}$ is zero. Then $\Lambda'(0) = 0$ and it is enough to check whether $\Lambda'(\infty) > 0$ or $\Lambda'(-\infty) < 0$. It can be shown that given $n \geq 1$ and $J \subset \{1, 2, \ldots, n\}$,

$$E_X\left[\log E_Y\left[\exp\left\{\sum_{i=1}^n \rho(X_i, Y_i)\right\}\right]\right] \geq E_X\left[\log E_Y\left[\exp\left\{\sum_{i\in J} \rho(X_i, Y_i)\right\}\right]\right].$$

The right hand side suggests that one may remove dependence structure from Y and uncover some of its independence structure. Indeed, if there is $J \subset \mathbb{N}$ with non-zero asymptotic density, that is, $\lim |J \cap \{1, 2, \ldots, n\}|/n > 0$, such that $Y_i, i \in J$ are only weakly dependent, then by the above inequality, it is possible to get non-trivial rate function for $\rho_n(X_1^n, Y_1^n)$. The existence of such J will be formulated as condition (A) (Definition 3). Since the condition does not require weak dependence for blocks of elements of Y it is not too restrictive. In particular, if Y satisfies condition (S_-) and A_Y is finite, then Y satisfies Condition (A). We thus can show:

THEOREM 7. *Given X, Y and ρ as in Theorem* 6, *suppose $\bar{\rho}(y) = E_X[\rho(X_1, y)]$, $y \in A_Y$, is a constant a. Assume A_Y is finite and, without loss of generality,* $\Pr\{Y_1 = y\} > 0$, $y \in A_Y$. *If Y satisfies condition (S_-) and*

$$E_X\left[\max_Y \rho(X_1, Y_1) - \min_Y \rho(X_1, Y_1)\right] > 0, \tag{1.21}$$

then $\Lambda^*(x)$, with $\Lambda(\lambda)$ given by (1.18), *is non-trivial.*

Besides Theorem 7, the case where $\bar{\rho}$ is not constant will also be considered in Section 7.

The remaining part of the article proceeds as follows. In Section 2 we prove Theorem 1. In Section 3 we prove Theorem 2, Theorem 3 and their implications on the lower bound of the LDP for ρ_n. In Section 4, we prove Theorem 4 and its implications on the upper bound of the quenched LDP for ρ_n. Convexity of the rate function of the LDP is proved in Section 5. Theorem 6 is proved in Section 6. In Section 7, we discuss condition (A) and prove Theorem 7.

2. Stochastic approximate sub-additivity. In this section, we prove Theorem 1. First note a simple property of $\Delta(n)$ satisfying (1.3):

$$\lim_{n\to\infty} \frac{\Delta(n)}{n} = 0. \tag{2.1}$$

Indeed, for any $n \geq 1$, since $\Delta(k)$ is non-decreasing and non-negative, by (1.3) of Theorem 1,

$$0 \leq \frac{\Delta(n)}{n} = \sum_{k=n}^{\infty} \frac{\Delta(n)}{k(k+1)} \leq \sum_{k=n}^{\infty} \frac{\Delta(k)}{k(k+1)} \to 0.$$

PROOF OF THEOREM 1. Since $h_n \in L_1(P)$ and T is measure-preserving, from (1.4),

$$E[h_{n+m}] \leq E[h_n] + E[h_m] + \Delta(n+m), \qquad m,\ n \in \mathbb{N}.$$

Then Hammersley's approximate sub-additivity lemma implies that $E[h_n/n]$ converges [2].

To show (1.6), there is $Z \subset \Omega$ with $P(Z) = 0$, such that (1.4) is satisfied by $\omega \notin Z$ and all $m, n \geq 1$. Enlarge Z to $\cup_{k=0}^{\infty} T^{-k} Z'$, with $Z' = Z \cup \{|h_k| = \infty$, for some $k \geq 1\}$, so that if $\omega \notin Z$, then $T^n \omega \notin Z$ and $|h_n(\omega)| < \infty$, $n \geq 1$. Given $k \geq N$, define a sequence $D(n)$ by

$$\begin{cases} D(1) = 0, \quad D(2^s) = 2D(2^{s-1}) + \Delta(2^s k), & s \geq 1, \\ D(2^l + r) = D(2^s) + D(r) + \Delta((2^l + r)k), & r = 1, \dots, 2^l - 1. \end{cases}$$

We show by induction

$$h_{nk}(\omega) \leq \sum_{j=0}^{n-1} h_k(T^{jk}\omega) + D(n), \qquad P\text{-a.s.} \tag{2.2}$$

When $n = 1$, (2.2) is obvious. Suppose (2.2) holds for 2^l, $l \leq m-1$. Let $n = 2^m$. By the induction hypothesis, for $\omega \notin Z$,

$$\begin{aligned} h_{nk}(\omega) &\leq h_{nk/2}(\omega) + h_{nk/2}(T^{nk/2}\omega) + \Delta(nk) \\ &\leq \sum_{j=0}^{2^{m-1}-1} h_k(T^{jk}\omega) + D(2^{m-1}) \\ &\quad + \sum_{j=0}^{2^{m-1}-1} h_k(T^{(n/2+j)k}\omega) + D(2^{m-1}) + \Delta(2^m k) \\ &= \sum_{j=0}^{2^m-1} h_k(T^{jk}\omega) + D(2^m). \end{aligned}$$

So (2.2) is proved when n is a dyadic integer.

On the other hand, assume (2.2) holds for all $n < 2^m$. If $2^m < n < 2^{m+1}$, then $n = 2^m + r$, with $1 \leq r \leq 2^m - 1$. By the induction hypothesis and what has been proved for dyadic integers,

$$\begin{aligned} h_{nk}(\omega) &\le h_{rk}(\omega) + h_{2^m k}(T^{rk}\omega) + \Delta(nk) \\ &\le \sum_{j=0}^{r-1} h_k(T^{jk}\omega) + D(r) + \sum_{j=0}^{2^m-1} h_k(T^{(r+j)k}\omega) + D(2^m) + \Delta(nk) \\ &= \sum_{j=0}^{n-1} h_k(T^{jk}\omega) + D(n), \end{aligned}$$

proving (2.2).

Given $\omega \notin Z$ and $s = 0, \dots, k-1$, by (2.2), $h_{nk}(T^s\omega) \le \sum_{j=0}^{n-1} h_k(T^{jk+s}\omega) + D(n)$ and hence

$$\begin{aligned} \frac{1}{k}\frac{1}{nk}\sum_{s=0}^{k-1} h_{nk}(T^s\omega) &\le \frac{1}{k}\frac{1}{nk}\sum_{j=0}^{n-1}\sum_{s=0}^{k-1} h_k(T^{jk+s}\omega) + \frac{D(n)}{nk} \\ &= \frac{1}{k}\frac{1}{nk}\sum_{j=0}^{nk-1} h_k(T^j\omega) + \frac{D(n)}{nk}. \end{aligned} \tag{2.3}$$

Let $n \to \infty$. Because $h_k \in L_1(P)$, by (1.5) and the assumption that T is ergodic,

$$\limsup_{n\to\infty} \frac{h_{nk}}{nk} \le \frac{E[h_k]}{k} + \limsup_{n\to\infty} \frac{D(n)}{nk}, \qquad P\text{-a.s.} \tag{2.4}$$

For $r = 1, \dots, k-1$,

$$h_{nk+r}(\omega) \le h_r(\omega) + h_{nk}(T^r\omega) + \Delta(nk+r).$$

Divide both sides by $nk+r$ and let $n \to \infty$. By (2.1), $(nk+r)^{-1}\Delta(nk+r) \to 0$. By (2.4) and the assumption that T is measure-preserving, it is easy to get

$$\limsup_{n\to\infty} \frac{h_{nk+r}}{nk+r} \le \frac{E[h_k(T^r\omega)]}{k} + \limsup_{n\to\infty}\frac{D(n)}{nk} = \frac{E[h_k]}{k} + \limsup_{n\to\infty}\frac{D(n)}{nk}.$$

Thus

$$\limsup_{n\to\infty} \frac{h_n}{n} \le \lim_{k\to\infty} \frac{E[h_k]}{k} + \liminf_{k\to\infty}\limsup_{n\to\infty} \frac{D(n)}{nk}, \qquad P\text{-a.s.}$$

We next show

$$\lim_{k\to\infty}\limsup_{n\to\infty} \frac{D(n)}{nk} = 0.$$

By the assumption, $\Delta(n)$ is non-decreasing and non-negative, then

$$\begin{aligned} \frac{D(2^n)}{2^n k} &= \sum_{j=0}^{n-1} \frac{\Delta(2^j k)}{2^j k} = 2\sum_{j=0}^{n-1} \Delta(2^j k) \sum_{i=2^j k}^{2^{j+1}k-1} \frac{1}{i(i+1)} \\ &\le 2\sum_{j=0}^{n-1}\sum_{i=2^j k}^{2^{j+1}k-1} \frac{\Delta(i)}{i(i+1)} \le 2\sum_{i=k}^{\infty} \frac{\Delta(i)}{i(i+1)}. \end{aligned}$$

It is easy to see that $D(r)$ is increasing. Then for $1 \le r \le 2^n - 1$, $(2^n + r)^{-1} D(2^n + r) \le 2^{-n} D(2^{n+1})$. It then follows that $[D(n)/nk] \le 4 \sum_{l=k}^{\infty} [\Delta(l)/l(l+1)]$. Let $n, k \to \infty$ to complete the proof.

Thus we have

$$\limsup_{n\to\infty} \frac{h_n}{n} \le \lim_{k\to\infty} \frac{E[h_k]}{k}, \qquad P\text{-a.s.} \tag{2.5}$$

Integrate both sides of (2.5). By Fatou's lemma,

$$\lim_{k\to\infty} \frac{E[h_k]}{k} \le E\left[\limsup_{n\to\infty} \frac{h_n}{n}\right] \le \lim_{k\to\infty} \frac{E[h_k]}{k},$$

which implies equality holds P–almost surely in (2.5), hence completing the proof. □

COROLLARY 3. *Let h_n satisfy all the conditions in Theorem* 1, *except* (1.5) *being replaced by*

$$\lim_{n\to\infty} \sup_{\omega\in\Omega} \left| \frac{h_n(\omega) - h_n(T\omega)}{n} \right| = 0. \tag{2.6}$$

Then for any $r > 0$,

$$\limsup_{n\to\infty} \frac{h_n(T^{\lfloor nr \rfloor}\omega)}{n} \le \lim_{n\to\infty} E\left[\frac{h_n(\omega)}{n}\right], \qquad P\text{-a.s.} \tag{2.7}$$

where $\lfloor t \rfloor$ is the largest integer $\le t$.

PROOF. Apply (2.3) to $T^{\lfloor rn \rfloor}\omega$ to get

$$\frac{1}{k}\frac{1}{nk} \sum_{s=\lfloor rn \rfloor}^{k+\lfloor rn \rfloor - 1} h_{nk}(T^s\omega) \le \frac{1}{k}\frac{1}{nk} \sum_{j=\lfloor rn \rfloor}^{nk-1+\lfloor rn \rfloor} h_k(T^j\omega) + \frac{D(n)}{nk}.$$

By the same argument for Theorem 1 and (2.6),

$$\limsup_{n\to\infty} [h_n(T^{\lfloor rn \rfloor}\omega)/n] \le [Eh_k(\omega)/k] + \limsup_{n\to\infty} [D(n)/nk], \qquad P\text{-a.s.}$$

The remaining part of the proof is identical to Theorem 1. □

3. Condition (S) and the lower bound of the conditional LDP. In this section, we apply Bryc's asymptotic value method to get a constant lower bound for the LDP of $\rho_n(X_1^n, Y_1^n)$ conditional on X, given that Y satisfies condition (S).

DEFINITION 1. Let Y be a stationary stochastic process defined on $(A_Y^\infty, \mathscr{F}_Y, Q)$. The process is said to satisfy condition (S) if for every $C > 0$, there is a non-decreasing sequence $\ell(k) \in \mathbb{N}$ satisfying (1.3), such that

$$(S_-): \sup\Big\{Q(A)Q(B) - e^{\ell(n)}Q(A \cap B):$$
$$A \in \sigma(Y_1^k), B \in \sigma\left(Y_{k+\ell(n)}^{k+h+\ell(n)}\right), k, h \in \mathbb{N}\Big\} \le e^{-Cn},$$
$$(S_+): \sup\Big\{Q(A \cap B) - e^{\ell(n)}Q(A)Q(B):$$
$$A \in \sigma(Y_1^k), B \in \sigma\left(Y_{k+\ell(n)}^{k+h+\ell(n)}\right), k, h \in \mathbb{N}\Big\} \le e^{-Cn}.$$

To get the conditional LDP without assuming independence between X and Y, we introduce the following version of condition (S).

DEFINITION 2. Let $U = \{U_n; n \in \mathbb{Z}\}$ be an $\mathbb{R}^d$–valued stationary process defined on probability space $(\Omega, \mathscr{F}, P)$. Given σ–algebra $\mathscr{B} \subset \mathscr{F}$, U is said to satisfy condition (S_-) [respectively, (S_+), (S)] uniformly, conditional on $\mathscr{B}$, if for every $C > 0$, there is a non-decreasing sequence $\ell(k) \in \mathbb{N}$ satisfying (1.3), such that condition (S_-) [respcetively, (S_+), (S)] in Definition 1 is satisfied P–almost surely, with Q replaced by $P(\cdot|\mathscr{B})$.

We shall need several inequalities from [2]. They are quoted below for completeness.

LEMMA 1. *Fix two σ–fields $\mathscr{F}$ and $\mathscr{G}$. If* $\sup\{P(A)P(B) - aP(A \cap B): A \in \mathscr{F}, B \in \mathscr{G}\} \le b$, *then for non-negative random variables* $W \in L^\infty(\mathscr{F})$, $Z \in L^\infty(\mathscr{G})$,

$$E(W)E(Z) - aE(WZ) \le b\|W\|_\infty \|Z\|_\infty.$$

If $\sup\{P(A \cap B) - aP(A)P(B): A \in \mathscr{F}, B \in \mathscr{G}\} \le b$, *then for non-negative random variables* $W \in L^\infty(\mathscr{F})$, $Z \in L^\infty(\mathscr{G})$,

$$E(WZ) - aE(W)E(Z) \le b\|W\|_\infty \|Z\|_\infty.$$

PROOF OF THEOREM 2. We follow the proof for Theorem 6.4.4 in [5]. Because ρ is bounded, we can assume for some $D > 0$, $\rho(\omega) \in B_D$, $\omega \in \Omega$, where B_D is the ball with center 0 and radius D. Given a continuous, bounded above, concave function f, there is $K > 0$, such that

$$|f(u) - f(v)| \le K|u - v|, \qquad -K \le f(u) \le K, \quad u, v \in B_D.$$

Fix $C = 2K + 2$ and a sequence $\ell(n)$ satisfying condition (S_-) corresponding to C. For brevity, in the remaining part of the proof, denote

$$N = n + m, \quad \ell = \ell(N), \quad S(i, j; \omega) = \sum_{k=i}^{j} \rho(T^k\omega).$$

Then

$$\left|S(1, N; \omega) - [S(1, n; \omega) + S(n+1+\ell, N+\ell; \omega)]\right| \\ \le \sum_{k=n+1}^{n+\ell} |\rho(T^k\omega)| + \sum_{k=N+1}^{N+\ell} |\rho(T^k\omega)| \le 2D\ell.$$

Let $u = N^{-1}S(1, N; \omega)$ and $v = N^{-1}[S(1, n; \omega) + S(n+1+\ell, N+\ell; \omega)]$. By $|f(u) - f(v)| \le K|u - v|$ and the concavity of f,

$$Nf\left(\frac{1}{N}S(1,N;\omega)\right) \ge Nf\left(\frac{1}{N}[S(1,n;\omega)+S(n+1+\ell,N+\ell;\omega)]\right) - 2KD\ell \\ \ge nf\left(\frac{1}{n}S(1,n;\omega)\right) + mf\left(\frac{1}{m}S(n+1+\ell,N+\ell;\omega)\right) - 2KD\ell.$$

Take exponential, then expectation conditional on $\mathscr{B}$ of both ends of the above formula. Since $P(\cdot|\mathscr{B})$ satisfies (S_-) almost surely, applying Lemma 1 to $Z = nf(n^{-1}S(1, n; \cdot))$ and $W = mf(m^{-1}S(n+1+\ell, N+\ell; \cdot))$ yields

$$\begin{aligned} &E[e^{Nf(\frac{1}{N}S(1,N;\cdot))}|\mathscr{B}] \\ &\ge e^{-(2KD+1)\ell}\left\{E[e^{nf(\frac{1}{n}S(1,n;\cdot))}|\mathscr{B}]E[e^{mf(\frac{1}{m}S(n+\ell+1,N+\ell;\cdot))}|\mathscr{B}] - e^{-CN}\right\}, \quad P\text{-a.s.} \end{aligned} \tag{3.1}$$

Similarly, by

$$mf\left(\frac{1}{m}S(n+\ell+1, N+\ell; \omega)\right) \ge mf\left(\frac{1}{m}S(n+1, N; \omega)\right) - 2KD\ell,$$

we get

$$E[e^{mf(\frac{1}{m}S(n+\ell+1,N+\ell;\cdot))}|\mathscr{B}] \ge e^{-2KD\ell}E[e^{mf(\frac{1}{m}S(n+1,N;\cdot))}|\mathscr{B}], \qquad P\text{-a.s.} \tag{3.2}$$

Combining (3.1) and (3.2) then gives

$$\begin{aligned} &E[e^{Nf(\frac{1}{N}S(1,N;\cdot))}|\mathscr{B}] \\ &\quad \ge e^{-(4KD+1)\ell}\left\{E[e^{nf(\frac{1}{n}S(1,n;\cdot))}|\mathscr{B}]E[e^{mf(\frac{1}{m}S(n+1,N;\cdot))}|\mathscr{B}] - e^{-CN+2KD\ell}\right\}, \quad P\text{-a.s.} \end{aligned}$$

Write $M = 4KD + 1$. Because $|f(u)| \le K$, the product of the two conditional expectations on the right hand side is larger than e^{-KN}. Whereas by $\ell/N \to 0$, $e^{-CN+2KD\ell} = e^{-(2K+2)N+2KD\ell} = o(e^{-2KN})$. Therefore, for N large enough,

$$\begin{aligned} E\left[e^{Nf(\frac{1}{N}S(1,N;\cdot))}|\mathscr{B}\right] \ge \tfrac{1}{2}\, e^{-M\ell}E\left[e^{nf(\frac{1}{n}S(1,n;\cdot))}|\mathscr{B}\right] \\ \times E\left[e^{mf(\frac{1}{m}S(n+1,N;\cdot))}|\mathscr{B}\right], \quad P\text{-a.s.} \end{aligned} \tag{3.3}$$

Given $g \in L_1(\Omega, P)$, $B \in \mathscr{B}$, since $T(B) \in \mathscr{B}$ and T and T^{-1} are measure-preserving,

$$\int_B E[g \circ T|\mathscr{B}] = \int_B g \circ T = \int_{T(B)} g = \int_{T(B)} E[g|\mathscr{B}] = \int_B E[g|\mathscr{B}] \circ T,$$

leading to $E[g\circ T|\mathscr{B}](\omega)=E[g|\mathscr{B}](T\omega)$, P-a.s. Therefore by $S(n+1, N;\omega)=S(1, m; T^n\omega)$, (3.3) yields

$$E\left[e^{Nf(\frac{1}{N}S(1,N;\cdot))}|\mathscr{B}\right](\omega)\ge \tfrac{1}{2}\,e^{-M\ell}E\left[e^{nf(\frac{1}{n}S(1,n;\cdot))}|\mathscr{B}\right](\omega)$$
$$\times E\left[e^{mf(\frac{1}{m}S(1,m;\cdot))}|\mathscr{B}\right](T^n\omega),\qquad P\text{-a.s.}$$

Take $a=\log 2$ and

$$h_n(\omega)=-\log E[e^{nf(\frac{1}{n}S(1,n;\cdot))}|\mathscr{B}](\omega). \tag{3.4}$$

Then it is easy to see h_n satisfies (1.5) and, letting $\Delta(n)=M\ell(n)+a$, for N large enough,

$$h_N(\omega)\le h_n(\omega)+h_m(T^n\omega)+\Delta(N),\qquad P\text{-a.s.}$$

Applying Theorem 1 to h_n then proves (1.8). □

We now turn to Theorem 3. For convenience, in the remaining part of this section as well as the following sections, we will use the following semi-standard notation,

$$\Lambda(f,\omega,n)=\frac{1}{n}\log\left\{\int e^{nf(u)}\mu_n(du|\omega)\right\}.$$

We need the following simple result.

LEMMA 2. *Let $\{f_n\}$ be a sequence of bounded measurable functions with $|f_n|\le M$, $n\ge 1$, for some constant. If for some constant a, $E_P[f_n]\to a$ and $\liminf_{n\to\infty} f_n=a$, P-a.s., then $f_n\xrightarrow{P} a$, as $n\to\infty$.*

PROOF. Letting $g_n=\inf\{f_k : k\ge n\}$, it is seen $|g_n|\le M$ and $\lim_{n\to\infty} g_n=a$, P-a.s. Then $E_P[f_n-g_n]\to 0$. Since $f_n-g_n\ge 0$, this implies $f_n-g_n\xrightarrow{P}0$, and hence $f_n\xrightarrow{P} a$. □

PROOF OF THEOREM 3. Because ρ is bounded, the support of μ_n, $n\ge 1$, is uniformly bounded as well. Suppose $\operatorname{supp}(\mu_n)\subset\Gamma$ for some convex compact set Γ. Then $C_b(\mathbb{R}^d)$ can be replaced by $C(\Gamma)$. Fix $\delta_k\downarrow 0$, $n^{(0)}=\{n_k^{(0)}\}\subset\mathbb{N}$ increasing, and $f\in C(\Gamma)$. By Lemmas 4.4.8 and 4.4.9 of [5], the set of all continuous, bounded above, concave functions on $\mathbb{R}^d$ is well-separating and there is a finite set $\mathscr{G}_1$ of such functions, such that $\sup_{u\in\Gamma}|f(u)-\max_{g\in\mathscr{G}_1} g(u)|\le\delta_1$, $\sup_{u\in\mathbb{R}^d} g(u)\le\sup_{u\in\Gamma} f(u)$, $g\in\mathscr{G}_1$. Let $h(u)=\max\{g(u): g\in\mathscr{G}_1\}$. Then

$$\max_{g\in\mathscr{G}_1}\{\Lambda(g,\omega,n)\}\le\Lambda(h,\omega,n)\le\frac{1}{n}\log\left\{\sum_{g\in\mathscr{G}_1}\int e^{ng(u)}\mu_n(du|\mathscr{B})(\omega)\right\}.$$

For each $g\in\mathscr{G}_1$, it is easy to see that $\Lambda(g,\cdot,n)$ are uniformly bounded functions on Ω. Then combining Lemma 2 and formulae (1.10) and (1.11), it is seen that there is $n^{(1)} = \{n_k^{(1)}\} \subset n^{(0)}$, such that $\lim_{k\to\infty} \Lambda(g, \omega, n_k^{(1)}) = \Lambda_g$, P-a.s., for $g \in \mathscr{G}_1$. Therefore,

$$\Lambda_h = \lim_{i\to\infty} \Lambda(h, \omega, n_i^{(1)}) = \max_{g\in\mathscr{G}_1}\{\Lambda_g\}, \qquad P\text{-a.s.} \tag{3.5}$$

It is easy to check that

$$|\Lambda(f, \omega, n) - \Lambda(g, \omega, n)| \le \|f - g\| \overset{\Delta}{=} \max_{x\in\Gamma} |f(x) - g(x)|, \ f, \ g \in C(\Gamma). \tag{3.6}$$

Because $\|f - h\| \le \delta$,

$$\begin{aligned}\limsup_{k\to\infty} \Lambda(f, \omega, n_k^{(1)}) &\le \delta_1 + \lim_{k\to\infty} \Lambda(h, \omega, n_k^{(1)})\\ &\le 2\delta_1 + \liminf_{k\to\infty} \Lambda(f, \omega, n_k^{(1)}), \qquad P\text{-a.s.}\end{aligned}$$

Repeating the above argument, we obtain nested sequences $n^{(0)} \supset n^{(1)} \supset \cdots$, $n^{(j)} = \{n_k^{(j)}\}$, such that $\limsup_{i\to\infty} \Lambda(f, \omega, n_i^{(j)}) \le 2\delta_j + \liminf_{i\to\infty} \Lambda(f, \omega, n_i^{(j)})$, P-a.s. Then by the diagonal argument, (1.12) holds for the sequence $n_k = n_k^{(k)}$.

To show that P–almost surely, (1.12)–(1.14) hold simultaneous for all $f \in C(\Gamma)$, note that by (3.6), $\{\Lambda(\cdot, \omega, n) : \omega \in \Omega, n \in \mathbb{N}\}$ as a family of functions on $C(\Gamma)$ is equi-continuous under the norm $\|\cdot\|$. In addition, what has been shown is that given any sequence $\{n_k\} \subset \mathbb{N}$, for $f \in C(\Gamma)$, there is a subsequence $\{n'_k\} \subset \{n_k\}$ such that (1.12) holds. Because $C(\Gamma)$ is separable, by the diagonal argument, there is a sequence $\{n_k\}$ such that with probability 1, (1.12) holds for all $f \in C(\Gamma)$.

To prove (1.14), given $f \in C(\Gamma)$ and $\delta > 0$, find a finite set $\mathscr{G}$ in the same way as to find $\mathscr{G}_1$ given δ_1. Let $h(u) = \max\{g(u) : g \in \mathscr{G}\}$. Given ω, from

$$\max\{\Lambda(g, \omega, n) : g \in \mathscr{G}_1\} \le \Lambda(h, \omega, n)$$

and

$$\max\left\{\liminf_{n\to\infty} a_n^{(i)} : 1 \le i \le K\right\} \le \liminf_{n\to\infty} \max\{a_n^{(i)} : 1 \le i \le K\}, a_n^{(i)} \in \mathbb{R},$$

it is not hard to show that

$$\max_{g\in\mathscr{G}}\{\Lambda_g\} = \max_{g\in\mathscr{G}}\left\{\liminf_{n\to\infty} \Lambda(g, \omega, n)\right\} \le \liminf_{n\to\infty} \Lambda(h, \omega, n), \qquad P\text{-a.s.}$$

On the other hand, by (3.5), $\Lambda_h = \max\{\Lambda_g : g \in \mathscr{G}\}$. Therefore, $\liminf_{n\to\infty} \Lambda(h, \omega, n) \ge \Lambda_h$. By $\|f - h\| \le \delta$, $\Lambda(h, \omega, n) \le \Lambda(f, \omega, n) + \delta$ and $\Lambda_h \ge \Lambda_f - \delta$, giving

$$\Lambda_f \le 2\delta + \liminf_{n\to\infty} \Lambda(f, \omega, n), \qquad P\text{-a.s.}$$

Because δ is arbitrary, (1.14) is proved for each fixed f. Now use the separability of $C(\Gamma)$, continuity of the map $f \to \Lambda_f$ on $C(\Gamma)$, and the equi-continuity of $\{\Lambda(\cdot, \omega, n) : \omega \in \Omega, n \in \mathbb{N}\}$ as a family of functions defined on $C(\Gamma)$, to complete the proof of (1.14). □

Now we can prove the lower bound for the quenched LDP in terms of I for $\rho_n(X_1^n, Y_1^n)$.

PROOF OF COROLLARY 1. This again follows the proof of Lemma 4.4.6 in [5]. First of all, I is lower semi-continuous, because it is the supremum of continuous functions. In addition, since the support of μ_n is in Γ, I is a good rate function. Given open set G and $u_0 \in G$, let $f : \mathbb{R}^d \to [0,1]$ be continuous, such that $f(u_0) = 1$ and $f(u) = 0$, for all $u \notin G$. Given $m > 0$, define $f_m(u) = m(f(u) - 1)$. Then for all $\omega \notin Z$,

$$\int_{\mathbb{R}^d} e^{nf_m(u)} \mu_n(du|\omega) \le e^{-nm} \mu_n(G^c|\omega) + \mu_n(G|\omega) \le e^{-mn} + \mu_n(G|\omega).$$

By $f_m \in C_b(\mathbb{R}^d)$ and $f_m(u_0) = 0$,

$$\begin{aligned}\max\left\{\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G|\omega), -m\right\} &\ge \liminf_{n\to\infty} \frac{1}{n} \log \left\{\int_{\mathbb{R}^d} e^{nf_m(u)} \mu_n(du|\omega)\right\} \\ &\ge \Lambda_{f_m} = -[f_m(u_0) - \Lambda_{f_m}] \\ &\ge -\sup_{f \in C_b(\mathbb{R}^d)} \{f(u_0) - \Lambda_f\} = -I(u_0).\end{aligned}$$

Finally, let $m \to \infty$ to complete the proof. □

From the proofs of Theorems 2 and 3, we can demonstrate the following two corollaries, which will be used in proving the convexity of the rate function I.

COROLLARY 4. *Under the same conditions as in Theorem 2, there is a sequence n_i and $Z \subset \Omega$, with $P(Z) = 0$, such that for $\omega \notin Z$, $f \in C_b(\mathbb{R}^d)$ and $r \in [0, 1)$,*

$$\lim_{i\to\infty} \frac{1}{(1-r)n_i} \log E\left[\exp\left\{(1-r)n_i f(\rho_{n_i - \lfloor rn_i \rfloor}(\cdot))\right\} | \mathscr{B}\right](T^{\lfloor rn_i \rfloor}\omega) = \Lambda_f. \tag{3.7}$$

In addition, for any $f \in C_b(\mathbb{R}^d)$,

$$\liminf_{n\to\infty} \frac{1}{(1-r)n} \log E\left[\exp\left\{(1-r)nf(\rho_{n-\lfloor rn \rfloor}(\cdot))\right\} | \mathscr{B}\right](T^{\lfloor rn \rfloor}\omega) \ge \Lambda_f. \tag{3.8}$$

PROOF. Given $r > 0$, and continuous, bounded above, concave function f, because (2.6) is satisfied by h_n defined by (3.4), applying Corollary 3 to $h_n(T^{\lfloor rn \rfloor}\omega)$ gives

$$\liminf_{n\to\infty} \frac{1}{(1-r)n} \log E\left[\exp\left\{(1-r)nf(\rho_{n-\lfloor rn \rfloor}(\cdot))\right\} | \mathscr{B}\right](T^{\lfloor rn \rfloor}\omega) = \Lambda_f, \qquad P\text{-a.s.}$$

Because the set of rational numbers is dense in $[0, 1)$, and $C_b(\mathbb{R}^d)$ is separable, by the same argument for Theorem 3 we can find a sequence n_i and a null set Z, such that for $\omega \notin Z$, (3.7) holds for any $f \in C_b(\mathbb{R}^d)$ and $r \in [0, 1)$. Inequality (3.8) follows from (3.7) and the argument for (1.14). □

COROLLARY 5. *Under the same conditions as in Theorem* 2, *there is a set* $Z \subset \Omega$ *with* $P(Z) = 0$, *such that*

$$\liminf_{n\to\infty} \frac{1}{(1-r)n} \Pr\left\{\rho_{n-\lfloor rn\rfloor}(\cdot) \in G|\omega\right\}$$
$$\geq -I(u),\ r \in [0, 1),\ u \in G \subset \mathbb{R}^d \text{ open}, \quad \omega \notin Z$$

This corollary can be proved in the same way as Theorem 3, and we omit the detail.

4. The upper bound of the conditional LDP.

PROOF OF THEOREM 4. The proof is almost identical to that of Theorem 2. Given $\lambda \in \mathbb{R}^d$, let $f(u) = \langle \lambda, u\rangle$. Then f is a continuous, convex function. Set constants K, D, C and functions $S(i, j; \omega)$ as in Theorem 2, and fix sequence ℓ correspondingly. Because $\{\rho \circ T^n : n \in \mathbb{Z}\}$ satisfies condition (S_+) uniformly, conditional on $\mathscr{B}$, by $S(1, N; \omega) \leq S(1, n; \omega) + S(n+1+\ell, N+\ell; \omega) + 2KD\ell$,

$$E[e^{S(1,N;\cdot)}|\omega] \leq e^{(2KD+1)\ell} \left\{E[e^{S(1,n;\cdot)}|\omega] + E[e^{S(n+\ell+1,N+\ell;\cdot)}|\omega] + e^{-CN}\right\}$$

and by $S(n+\ell+1, N+\ell; \omega) \leq S(n+1, N; \omega) + 2KD\ell$,

$$E[e^{S(n+\ell+1,N+\ell;\cdot)}|\omega] \leq e^{2KD\ell} E[e^{S(n+1,N;\cdot)}|\omega].$$

Then

$$E[e^{S(1,N;\cdot)}|\omega] \leq e^{(4KD+1)\ell} \left\{E[e^{S(1,n;\cdot)}|\omega] E[e^{S(n+1,N;\cdot)}|\mathscr{B}](\omega) + e^{-CN+2KD\ell}\right\},$$

which gives

$$E[e^{S(1,N;\cdot)}|\omega] \leq e^{(4KD+1)\ell} \left\{E[e^{S(1,n;\cdot)}|\omega] E[e^{S(1,m;\cdot)}|T^n\omega] + e^{-CN+2KD\ell}\right\}.$$

Letting $M = 4KD + 1$, for N large enough,

$$E[e^{S(1,N;\cdot)}|\omega] \leq 2e^{M\ell} E[e^{S(1,n;\cdot)}|\omega]\, E[e^{S(1,m;\cdot)}|T^n\omega].$$

Let $h_n(\omega) = \log E[e^{S(1,n;\cdot)}|\omega]$ and $\Delta(n) = M\ell(n) + \log 2$. Applying Theorem 1 shows that (1.17) holds for each $\lambda \in \mathbb{R}^d$, P–almost surely. Finally because $\mathbb{R}^d$ is separable, the same argument for Theorem 3 completes the proof. □

PROOF OF COROLLARY 2. For simplicity, given $\lambda \in \mathbb{R}^d$, $x \in \mathbb{R}^d$ and $n \geq 1$, write $\Lambda(\lambda, \omega, n) \triangleq \Lambda(f_\lambda, \omega, n)$, with $f_\lambda(u) = \langle \lambda, u\rangle$. Given $\omega \notin Z$, choose n_i such that

$$\lim_{i\to\infty} \frac{1}{n_i} \log \mu_{n_i}(F|\omega) = \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F|\omega). \tag{4.1}$$

By the equi-continuity of $\Lambda(\lambda, \omega, n)$ in λ, the separability of $\mathbb{R}^d$ and Lemma 2, there is a subsequence m_i of n_i such that $\Lambda(\lambda, \omega, m_i)$ converges for all λ. Let the limit be $\hat{\Lambda}(\lambda)$. By (1.18), $\hat{\Lambda}(\lambda) \leq \Lambda(\lambda)$. Since $\{\Lambda(\lambda, \omega, n) : \omega \in \Omega, n \in \mathbb{N}\}$ is an equi-continuous family of functions in λ, $\hat{\Lambda}(\lambda)$ is continuous.

Given a compact set F, apply Gärtner-Ellis theorem to $\mu_{m_i}(\cdot|\omega)$ and $\hat{\Lambda}(\lambda)$ to get

$$\lim_{i\to\infty} \frac{1}{m_i} \log \mu_{m_i}(F|\omega) \le -\inf_{u\in F}\{\hat{\Lambda}^*(u)\},$$

where $\hat{\Lambda}^*(u)$ is the Fenchel-Legendre transform of $\hat{\Lambda}(\lambda)$. Since $\hat{\Lambda}(\lambda) \le \Lambda(\lambda)$, then $\hat{\Lambda}^*(u) \ge \Lambda^*(\lambda)$, which combined with (4.1) implies (1.18). □

5. Convexity of the rate function *I*.

PROOF OF THEOREM 5. By Corollary 4, for some sequence $\{n_i : i \ge 1\}$ and $Z \subset \Omega$ with $P(Z) = 0$, such that for any $\omega \notin Z$ and $f \in C_b(\mathbb{R}^d)$, (3.7) holds for $r = 0, \frac{1}{2}$. Still denoting by μ_n be the probability measure on $\mathbb{R}^d$ induced by ρ_n, Bryc's inverse Varadhan lemma implies that for $\omega \notin Z$, $\mu_{n_i}(\cdot|\omega)$ and $\mu_{\lceil n_i/2\rceil}(\cdot|T^{\lfloor n_i/2\rfloor}\omega)$ satisfy the LDP with a good rate function $I(u)$. Applying Varadhan's integral lemma to μ_{n_i} then yields $\Lambda(\lambda) = \sup_{u\in\mathbb{R}^d}(\langle\lambda, u\rangle - I(u))$. By the duality between a lower semi-continuous convex function and its Fenchel-Legendre transform, in order to have $I(u) = \Lambda^*(u)$, it is enough that $I(u)$ be convex. Because ρ is bounded, (3.7) also holds for $n_i + 1$, and hence μ_{n_i+1} also satisfy the conditional LDP with the rate function I. Therefore, without loss of generality, assume all n_i are even numbers.

Since the collection of all balls $B(u, \delta)$ form a base for the topology of $\mathbb{R}^d$, for $\omega \notin Z$, $u \in \mathbb{R}^d$

$$\begin{aligned} -I(u) &= \inf_{\delta>0, v\in B(u,\delta)} \liminf_{i\to\infty} \frac{1}{n_i} \log \mu_{n_i}(B(v, \delta)|\omega) \\ &= \inf_{\delta>0, v\in B(u,\delta)} \limsup_{i\to\infty} \frac{1}{n_i} \log \mu_{n_i}(B(v, \delta)|\omega). \end{aligned} \tag{5.1}$$

Since $v \in B(u, \delta)$ implies $B(u, \varepsilon) \subset B(v, \delta)$ for all $\varepsilon > 0$ small enough, (5.1) can be rewritten as

$$-I(u) = \inf_{\delta>0} \liminf_{i\to\infty} \frac{1}{n_i} \log\mu_{n_i}(B(u,\delta)|\omega) = \inf_{\delta>0} \limsup_{i\to\infty} \frac{1}{n_i} \log\mu_{n_i}(B(u,\delta)|\omega). \tag{5.2}$$

Similarly, for $m_i = n_i/2$, by $\mu_{m_i}(B(u, \delta)|T^{m_i}\omega) = \Pr\{\rho_{m_i} \in B(u, \delta)|T^{m_i}\omega\}$,

$$-I(u) = \inf_{\delta>0} \liminf_{i\to\infty} \frac{1}{m_i} \log \mu_{m_i}(B(u, \delta)|T^{m_i}\omega). \tag{5.3}$$

On the other hand, Corollary 1 implies

$$-I(u) \le \inf_{\delta>0} \liminf_{i\to\infty} \frac{1}{m_i} \log \mu_{m_i}(B(u, \delta)|\omega). \tag{5.4}$$

Given u_1 and u_2 with $I(u_1), I(u_2) < \infty$, fix $M > \max\{I(u_i) : i = 1, 2\}$ large enough. Then given $\delta > 0$, for m_i large enough,

$$\mu_{m_i}(B(u_1, \delta/2)|\omega)\, \mu_{m_i}(B(u_2, \delta/2)|T^{m_i}\omega) \ge e^{-Mm_i}.$$

Fix a sequence $\ell(n)$ corresponding to $C = 2M$ for condition (S_-). Since ρ is bounded, as m is large enough,

$$\frac{1}{m}S(m+\ell(m), 2m+\ell(m)-1;\omega) \in B(u_2, 3\delta/4)$$
$$\Rightarrow \frac{1}{m}S(m+1, 2m;\omega) \in B(u_2, \delta),$$
$$\frac{1}{m}S(m+1, 2m;\omega) \in B(u_2, \delta/2)$$
$$\Rightarrow \frac{1}{m}S(m+\ell(m), 2m+\ell(m)-1;\omega) \in B(u_2, 3\delta/4)$$

and, letting $\bar{u} = (u_1+u_2)/2$,

$$\frac{1}{m}S(1, m;\omega) \in B(u_1, \delta),\ \frac{1}{m}S(m+1, 2m;\omega) \in B(u_2, \delta)$$
$$\Rightarrow \frac{1}{2m}S(1, 2m;\omega) \in B(\bar{u}, \delta).$$

Then by condition (S_-), for n_i large enough,

$$\begin{aligned}\mu_{n_i}(B(\bar{u},\delta)|\omega) &\geq e^{-\ell(n_i)}\left[\mu_{m_i}(B(u_1,\delta/2)|\omega)\,\mu_{m_i}(B(u_2,\delta/2)|T^{m_i}\omega) - e^{-2Mm_i}\right] \\ &\geq \tfrac{1}{2}\,e^{-\ell(n_i)}\mu_{m_i}(B(u_1,\delta/2)|\omega)\,\mu_{m_i}(B(u_2,\delta/2)|T^{m_i}\omega).\end{aligned}$$

Therefore, by approximate sub-additivity, for $\omega \notin Z$, we get

$$\begin{aligned}&\limsup_{i\to\infty}\frac{1}{n_i}\log\mu_{n_i}(B(\bar{u},\delta)|\omega) \\ &\quad\geq \frac{1}{2}\left[\liminf_{i\to\infty}\frac{1}{m_i}\mu_{m_i}(B(u_1,\delta/2)|\omega) + \liminf_{i\to\infty}\frac{1}{m_i}\mu_{m_i}(B(u_2,\delta/2)|T^{m_i}\omega)\right].\end{aligned}$$

Take $\inf_{\delta>0}$ on both sides and apply (5.2), (5.3) and (5.4) to get $-I(\bar{u}) \geq -(I(u_1)+I(u_2))/2$. The lower continuity of I then implies I is convex. □

6. More on the rate function for the quenched LDP. In this section and the next one, we shall establish some conditions for the rate function I of the LDP of $n^{-1}\sum_{k=1}^{n}\rho(X_k, Y_k)$ conditional on X to be non-trivial. We will consider the case where Y satisfies condition (S), and X is i.i.d. From now on we assume ρ is a $\mathbb{R}$–valued bounded measurable function such that Q–almost surely, the P–measure of the set of discontinuity points of $\rho(x, Y_1)$ is 0, where $P = \mathrm{dist}(X)$, $Q = \mathrm{dist}(Y)$ and the continuity is with respect to a metric on A_X. The following result will play an important role.

LEMMA 3. *Let A_X be a Polish space with metric d_X, and A_Y a measure space. Fix $f : A_X \times A_Y \to \mathbb{R}$ a bounded measurable function. Given random* *variables X and Y on A_X and A_Y, with probability measures P and Q, respectively, suppose that Q–almost surely, the P–measure of the set of discontinuity points of $f(x, Y)$ is 0. Then*

$$E_X\left[\log E_Y\left[e^{f(X,Y)}\right]\right] \geq \log E_Y\left[e^{E_X[f(X,Y)]}\right]. \tag{6.1}$$

PROOF. Suppose $|f| \leq D$. Given $\varepsilon > 0$, fix compact subset $K \subset A_X$, such that $P(K) > 1 - \varepsilon$. Given $\delta > 0$, let $\mathscr{I}(\delta) = \{I_1, \dots, I_n\}$ be a partition of K, such that (1) $d_X(x, x') < \delta$, $x, x' \in I_k$, $k = 1, \dots, n$ and (2) if $\delta' < \delta$ and $\mathscr{I}(\delta') = \{J_1, \dots, J_m\}$, then for each k, $J_k \subset I_i$ for some i. Letting $P_k = P(I_k)$, $h_k(y) = \inf\{f(x, y) : x \in I_k\}$, $y \in A_Y$, $k = 1, \dots, n$, then

$$\begin{aligned} E_X\left[\log E_Y\left[e^{f(X,Y)}\right]\right] &\geq -D\varepsilon + \sum_{k=1}^n P_k \log E_Y\left[e^{h_k(Y)}\right] \\ &= -D\varepsilon + \log \prod_{k=1}^n \left(E_Y\left[e^{h_k(Y)}\right]\right)^{P_k} \\ &\geq -D\varepsilon + \log\left(E_Y\left[e^{\sum_{k=1}^n P_k h_k(Y)}\right]\right) \end{aligned} \tag{6.2}$$

where the last inequality is due to Hölder's inequality. We have

$$\sum_{k=1}^n P_k h_k(Y) = E_X\left[\sum_{k=1}^n h_k(Y)\mathbf{1}_{I_k}(x)\right] = E_X[f_n(X, Y)].$$

Given $y \in A_Y$, if $x \in A_X$ is a point where $f(x, y)$ is continuous, then $f_n(x, y) \uparrow f(x, y)$ as $\delta \downarrow 0$. Therefore, for y such that the P–measure of the set of discontinuity points of $f(x, y)$ is 0, by dominated convergence theorem, as $\delta \downarrow 0$, $E[f_n(X, y)] \uparrow E[f(X, y)\mathbf{1}_K(X)]$. Because Q–almost surely, the set of discontinuity points of $f(x, Y)$ is 0, therefore, again by dominated convergence theorem,

$$\begin{aligned} \lim_{\delta\to 0} \log\left(E_Y\left[e^{\sum_{k=1}^n P_k h_k(Y)}\right]\right) &= \lim_{\delta\to 0} \log\left(E_Y\left[E_X[e^{f_n(X,Y)}]\right]\right) \\ &= \log\left(E_Y\left[e^{E_X[f(X,Y)\mathbf{1}_K(X)]}\right]\right) \\ &\geq \log\left(E_Y\left[e^{E_X[f(X,Y)] - DP(A_X\setminus K)}\right]\right) \\ &\geq \log\left(E_Y\left[e^{E_X[f(X,Y)]}\right]\right) - D\varepsilon. \end{aligned}$$

We thus get from (6.1) and the above limit

$$E_X[\log E_Y[e^{f(X,Y)}]] \geq -2D\varepsilon + \log(E_Y[e^{E_X[f(X,Y)]}]).$$

Letting $\varepsilon \to 0$ thus completes the proof. □

The following consequence of Lemma 3 will be useful.

COROLLARY 6. *Under the same condition of Lemma 3, given $n \geq 1$, suppose $X = (X_1, \dots, X_n)$, with X_i independent of each other. Then for any $S \subset \{1, \dots, n\}$,*

$$E_X\left[\log E_Y\left[\exp\left\{\sum_{i=1}^n f(X_i, Y_i)\right\}\right]\right] \geq E_X\left[\log E_Y\left[\exp\left\{\sum_{i\in S} E_X[f(X_i, Y_i)] + \sum_{i\notin S} f(X_i, Y_i)\right\}\right]\right].$$

PROPOSITION 2. *Let $X = \{X_n; n \geq 1\}$ and $Y = \{Y_n; n \geq 1\}$ be two independent stationary processes on the Polish spaces $(A_X^\infty, \mathscr{F}_X)$ and $(A_Y^\infty, \mathscr{F}_Y)$, respectively. Suppose X is an i.i.d. process. Given $\rho : A_X \times A_Y \to \mathbb{R}$ measurable and bounded, define*

$$\bar{\rho}(y) = E_X[\rho(X_1, y)]. \tag{6.3}$$

Define $\bar{\Lambda}$ by (1.20) *and*

$$\hat{\Lambda}(\lambda) = \limsup_{n\to\infty} \frac{1}{n} \log E_{(X,Y)}\left[\exp\left\{\lambda \sum_{i=1}^n \rho(X_i, Y_i)\right\}\right], \tag{6.4}$$

where $E_{(X,Y)}$ is the expectation with respect to the joint distribution of X and Y. Then

$$\bar{\Lambda}'(0-) = \hat{\Lambda}'(0-), \qquad \bar{\Lambda}'(0+) = \hat{\Lambda}'(0+). \tag{6.5}$$

PROOF. It is easy to see that both $\bar{\Lambda}$ and $\hat{\Lambda}$ are convex, and therefore their right and left derivatives at 0 exist. Define

$$\hat{\rho}(\lambda; y) = \log E_X\left[e^{\lambda\rho(X_1, y)}\right].$$

Then because X and Y are independent, and X is an i.i.d. process,

$$\begin{aligned} \log E_{(X,Y)}&\left[\exp\left\{\lambda \sum_{i=1}^n \rho(X_i, Y_i)\right\}\right] \\ &= \log E_Y\left[\prod_{i=1}^n E_X[\exp\{\lambda\rho(X_i, Y_i)\}]\right] \\ &= \log E_Y\left[\exp\left\{\sum_{i=1}^n \hat{\rho}(\lambda; Y_i)\right\}\right]. \end{aligned} \tag{6.6}$$

Suppose $|\rho| \leq D$. Then for any $y \in A_Y$, it is seen

$$0 \leq \hat{\rho}''(\lambda; y) = \frac{E_X[\rho^2(X_1, y)e^{\lambda\rho(X_1, y)}]}{E_X[e^{\lambda\rho(X_1, y)}]} - \left(\frac{E_X[\rho(X_1, y)e^{\lambda\rho(X_1, y)}]}{E_X[e^{\lambda\rho(X_1, y)}]}\right)^2 \leq D^2.$$

By the Taylor expansion,

$$\hat{\rho}(\lambda; y) = \hat{\rho}(0; y) + \lambda\hat{\rho}'(0; y) + \frac{\lambda^2}{2}\int_0^1 (1-t)^2 \hat{\rho}''(t\lambda; y)\, dt.$$

Then, from $\hat{\rho}(0; y) = 0$ and $\hat{\rho}'(0; y) = \bar{\rho}(y)$, it follows $\hat{\rho}(\lambda; y) \le \bar{\rho}(y)\lambda + D^2\lambda^2$. Then by (6.6),

$$\begin{aligned}\frac{1}{n}\log E_{(X,Y)}\left[\exp\left\{\lambda\sum_{i=1}^n \rho(X_i, Y_i)\right\}\right] &\le \frac{1}{n}\log E_Y\left[\exp\left\{\lambda\sum_{i=1}^n \bar{\rho}(Y_i) + nD^2\lambda^2\right\}\right]\\ &= D^2\lambda^2 + \frac{1}{n}\log E_Y\left[\exp\left\{\lambda\sum_{i=1}^n \bar{\rho}(Y_i)\right\}\right].\end{aligned}$$

Letting $n \to \infty$, we get $\hat{\Lambda}(\lambda) \le \bar{\Lambda}(\lambda) + D\lambda^2$. Since $\hat{\Lambda}(0) = \bar{\Lambda}(0) = 0$, this leads to $\hat{\Lambda}'(0+) \le \bar{\Lambda}'(0+)$ and $\hat{\Lambda}'(0-) \ge \bar{\Lambda}'(0-)$. On the other hand, as e^x is a convex function of x, Jensen's inequality gives

$$\begin{aligned}E_{(X,Y)}\left[\exp\left\{\lambda\sum_{i=1}^n \rho(X_i, Y_i)\right\}\right] &= E_Y\left[E_X\left[\exp\left\{\lambda\sum_{i=1}^n \rho(X_i, Y_i)\right\}\right]\right]\\ &\ge E_Y\left[\exp\left\{\lambda\sum_{i=1}^n E_X[\rho(X_i, Y_i)]\right\}\right]\\ &= E_Y\left[\exp\left\{\lambda\sum_{i=1}^n \bar{\rho}(Y_i)\right\}\right],\end{aligned}$$

leading to $\hat{\Lambda}(\lambda) \ge \bar{\Lambda}(\lambda)$, and hence $\hat{\Lambda}'(0+) \ge \bar{\Lambda}'(0+)$ and $\hat{\Lambda}'(0-) \le \bar{\Lambda}'(0-)$. □

PROOF OF THEOREM 6. By the assumption on ρ, we can apply Lemma 3 to get

$$E_X\left[\frac{1}{n}\log E_Y\left[\exp\{n\lambda\rho_n(X_1^n, Y_1^n)\}\right]\right] \ge \frac{1}{n}\log E_Y\left[\exp\left\{n\lambda\sum_{i=1}^n \bar{\rho}(Y_i)\right\}\right]$$

Since Y satisfies condition (S_-), when $n \to \infty$, both sides converge. Thus $\Lambda(\lambda) \ge \bar{\Lambda}(\lambda)$. Since $\Lambda(0) = \bar{\Lambda}(0)$, then,

$$\begin{aligned}&\Lambda'(0-) \le \bar{\Lambda}'(0-),\\ &\Lambda'(0+) \ge \bar{\Lambda}'(0+),\end{aligned} \tag{6.7}$$

On the other hand, Jensen's inequality yields

$$E_X\left[\frac{1}{n}\log E_Y\left[\exp\{n\lambda\rho_n(X_1^n, Y_1^n)\}\right]\right] \le \frac{1}{n}\log E_{(X,Y)}\left[\exp\left\{n\lambda\sum_{i=1}^n \rho(X_i, Y_i)\right\}\right]$$

Since X and Y are independent, and each one satisfies condition (S_-), the process (X, Y) also satisfies condition (S_-). Therefore, when $n \to \infty$, the right hand side converges. Then, by (6.4), $\Lambda(\lambda) \le \hat{\Lambda}(\lambda)$ and similar to (6.7), $\Lambda'(0-) \ge \hat{\Lambda}'(0-)$ and $\Lambda'(0+) \le \hat{\Lambda}'(0+)$. By Proposition 2, we get

$$\Lambda'(0-) = \bar{\Lambda}'(0-), \qquad \Lambda'(0+) = \bar{\Lambda}'(0+). \tag{6.8}$$

Because Λ and $\bar{\Lambda}$ are convex, $\Lambda(0) = \bar{\Lambda}(0) = 0$ and $\Lambda(\lambda) \ge \bar{\Lambda}(\lambda)$, it is seen $\Lambda'(\infty) \ge \bar{\Lambda}'(\infty)$ and $\Lambda'(-\infty) \le \bar{\Lambda}'(-\infty)$. Because $\bar{\Lambda}^*$ is non-trivial, either $\bar{\Lambda}'(-\infty) < \bar{\Lambda}'(0-)$ or $\bar{\Lambda}'(\infty) > \bar{\Lambda}(0)$. By (6.8), this implies that either $\Lambda'(-\infty) < \Lambda'(0-)$ or $\Lambda'(\infty) > \Lambda(0)$. Thus Λ^* is non-trivial. □

COROLLARY 7. *Suppose* $A_X = A_Y = \{0, 1\}$, *and let* ρ *be the Hamming distance* $\rho(x, y) = \mathbf{1}_{\{x=y\}}$. *If the rate function for the LDP of* $n^{-1}\sum_{j=1}^n Y_j$ *is non-trivial, then for any Bernouli process* X *with* $\Pr\{X_1 = 0\} = 1 - \Pr\{X_1 = 1\} = p \ne 1/2$, Λ^* *is non-trivial.*

PROOF. Let Λ_0^* be the rate function for the LDP of $n^{-1}\sum_{j=1}^n Y_j$. Letting $q = 1 - p$, it is seen

$$\bar{\Lambda}(\lambda) = q\lambda + \lim_{n\to\infty} \frac{1}{n} \log E_Y \left[\exp\left\{\lambda(p-q)\sum_{j=1}^n Y_j\right\}\right],$$

giving $\bar{\Lambda}(\lambda) = q\lambda + \Lambda_0((p-q)\lambda)$, leading to the conclusion. □

7. Condition (A) and non-triviality of the rate function. Next we consider the case where $\bar{\rho}(y)$ is a constant, in which case $\bar{\Lambda}(\lambda)$ can not be convex. Suppose S_n, $n = 1, 2, \ldots$ are finite subsets of $\mathbb{N}$. We say S_n have asymptotic density L if $|S_n| \to \infty$ and

$$L = \lim_{n\to\infty} [\,|S_n| / \operatorname{diam}(S_n)\,] \tag{7.1}$$

exists, where $|S_n|$ is the cardinality of S_n, and $\operatorname{diam}(S_n) = \max\{x : x \in S_n\} - \min\{x : x \in S_n\} + 1$.

DEFINITION 3. Given $Y = \{Y_n; n \ge 1\}$ with $Q = \operatorname{dist}(Y)$, denote by (A) the following condition.

(A) There are finite $S_n \subset \mathbb{N}$, $n = 1, 2, \ldots$, with asymptotic density $L > 0$, and a function $\ell(n) \ge 0$, satisfying

$$\limsup_{n\to\infty} [\ell(n)/n] < \infty$$

such that

$$Q\left(\bigcap_{i\in S_n} A_i\right) \ge e^{-\ell(\operatorname{diam}(S_n))} \prod_{i\in S_n} Q(A_i), \quad A_i \in \sigma(Y_i),\ i \in S_n,\ n = 1, 2, \ldots$$

□

LEMMA 4. *Given processes X and Y and function ρ as in Proposition 2, let $\bar{\rho}(y) = E_X[\rho(X, y)]$ and Λ given by (1.19) (assuming the limit exists). If Y satisfies condition (A), then*

$$\Lambda(\lambda) + \bar{\Lambda}(-\lambda) \geq -2 \limsup_{n\to\infty} \frac{\ell(n)}{n} + 2LE_X\left[\log E_Y\left[e^{\frac{\lambda}{2}(\rho(X_1,Y_1)-\bar{\rho}(Y_1))}\right]\right]. \tag{7.2}$$

PROOF. Suppose S_n are sets meeting condition (A). Without loss of generality, assume $\min\{x : x \in S_n\} = 1$. Letting $N_n = \operatorname{diam}(S_n)$, $S_n \subset [1, N_n]$. Then

$$\frac{1}{N_n} E_X\left[\log E_Y\left[\exp\left\{\lambda \sum_{i=1}^{N_n} \rho(X_i, Y_i)\right\}\right]\right]$$

$$\geq \frac{1}{N_n} E_X\left[\log E_Y\left[\exp\left\{\lambda \sum_{i\in[1,N_n]\setminus S_n} \bar{\rho}(Y_i) + \lambda \sum_{i\in S_n} \rho(X_i, Y_i)\right\}\right]\right] \tag{7.3}$$

$$\begin{aligned} &\geq -\frac{1}{N_n} E_X\left[\log E_Y\left[\exp\left\{-\lambda \sum_{i=1}^{N_n} \bar{\rho}(Y_i)\right\}\right]\right] \\ &\quad + \frac{2}{N_n} E_X\left[\log E_Y\left[\exp\left\{\frac{\lambda}{2} \sum_{i\in S_n} (\rho(X_i, Y_i) - \bar{\rho}(Y_i)\right\}\right]\right] \end{aligned} \tag{7.4}$$

$$\begin{aligned} &\geq -\frac{1}{N_n} E_X\left[\log E_Y\left[\exp\left\{\lambda \sum_{i=1}^{N_n} \bar{\rho}(Y_i)\right\}\right]\right] \\ &\quad + \frac{2}{N_n} E_X\Bigg[\log\Bigg\{\exp\{-\ell(N_n)\} \\ &\qquad \times \prod_{j\in S_n} E_Y\left[\exp\left\{\frac{\lambda}{2}(\rho(X_i, Y_i) - \bar{\rho}(Y_i))\right\}\right]\Bigg\}\Bigg] \end{aligned} \tag{7.5}$$

$$\begin{aligned} &= -\frac{1}{N_n} E_X\left[\log E_Y\left[\exp\left\{\lambda \sum_{i=1}^{N_n} \bar{\rho}(Y_i)\right\}\right]\right] \\ &\quad - \frac{2\ell(N_n)}{N_n} + \frac{2}{N_n} \sum_{j\in S_n} E_X\left[\log E_Y\left[\exp\left\{\frac{\lambda}{2}(\rho(X_i, Y_i) - \bar{\rho}(Y_i))\right\}\right]\right] \end{aligned}$$

where (7.3) is due to Corollary 6, (7.5) to Hölder inequality and (7.6) to condition (A). Let $n \to \infty$ to complete the proof. □

PROPOSITION 3. *Given processes X and Y and function ρ as in Proposition 2, suppose $\bar{\Lambda}^*$ is trivial. Then $\Lambda^*(x)$ is non-trivial if one of the following two* *inequalities holds*:

$$\bar{\Lambda}'(0+) - \bar{\Lambda}'(0-) < LE_X\left[\operatorname*{ess\,sup}_Y (\rho(X_1, Y_1) - \bar{\rho}(Y_1))\right], \tag{7.6}$$

$$\bar{\Lambda}'(0+) - \bar{\Lambda}'(0-) < LE_X\left[\operatorname*{ess\,sup}_Y (\bar{\rho}(Y_1) - \rho(X_1, Y_1))\right]. \tag{7.7}$$

PROOF. From (7.2),

$$\Lambda'(\infty) - \bar{\Lambda}'(-\infty) \geq a \stackrel{\Delta}{=} LE_X\left[\operatorname*{ess\,sup}_Y (\rho(X_1, Y_1) - \bar{\rho}(Y_1))\right],$$

$$-\Lambda'(-\infty) + \bar{\Lambda}'(\infty) \geq b \stackrel{\Delta}{=} LE_X\left[\operatorname*{ess\,sup}_Y (\bar{\rho}(Y_1) - \rho(X_1, Y_1))\right].$$

Because $\bar{\Lambda}^*$ is trivial, $\bar{\Lambda}'(-\infty) = \bar{\Lambda}'(0-)$ and $\bar{\Lambda}'(\infty) = \bar{\Lambda}'(0+)$. If Λ^* is also trivial, then by (6.8), $\Lambda^* = \bar{\Lambda}^*$, leading to $\bar{\Lambda}'(0+) - \bar{\Lambda}'(0-) \geq a$ and $-\bar{\Lambda}'(0-) + \bar{\Lambda}'(0+) \geq b$, contradicting either (7.6) or (7.7). □

PROOF OF THEOREM 7. Because $\bar{\rho}(y)$ is constant, $\bar{\Lambda}'(0+) = \bar{\Lambda}'(0-) = 0$. Therefore, condition (1.21) implies that either (7.6) or (7.7) is satisfied as long as L is positive. It is then enough to show Y satisfies condition (A).

Fix C such that

$$e^C \geq e^3 q_0^{-2}, \qquad q_0 = \min\{Q(y) : y \in A_Y\}.$$

Condition (S_-) implies that there is a non-decreasing sequence $\ell(n) \geq 0$ with $\sum_{n=1}^{\infty} \ell(n)/[n(n+1)] < \infty$, such that

$$Q(A \cap B) \geq 2e^{-\ell(n)} Q(A)Q(B) - e^{-Cn},$$
$$A \in \sigma(Y_j; j \leq h),\ B \in \sigma(Y_j; j \geq h + \ell(n)),\ n, h \geq 1.$$

By (2.1), $\ell(n)/n \to 0$. Fix k, such that for all $n \geq k$, $\ell(n) \leq n$. For $S = \{s\}$ and $n \in \mathbb{N}$, denote $S + n = \{s + n\}$. Define a sequence in $\mathbb{N}$ and a sequence of finite subsets of $\mathbb{N}$ as follows:

$$L_1 = 2 + k, \qquad L_n = 2L_{n-1} + \ell(L_{n-1}),$$

$$S_1 = \{1, 2 + k\}, \quad S'_{n-1} = S_{n-1} + L_{n-1} + \ell(L_{n-1}), \quad S_n = S_{n-1} \cup S'_{n-1}.$$

It is easy to check the following:

$$\operatorname{diam}(S_n) = L_n \geq 2^n, \quad S_n \cap S'_n = \varnothing, \quad \operatorname{dist}(S_n, S'_n) = \ell(L_n), \quad |S_n| = 2^n,$$

where $d(A, B)$, $A, B \subset \mathbb{N}$, denotes $\min\{|t - s| : t \in A, s \in B\}$. We prove by induction

$$Q\left(\bigcap_{j \in S_n} A_j\right) \geq e^{-L_n} \prod_{j \in S_n} Q(A_j), \qquad A_j \in \sigma(Y_j). \tag{7.8}$$

Once we show (7.8), we can complete the proof by choosing $\ell(n) = n$ in (A). When $n = 1$, since $e^{-\ell(k)} Q(A_1)\, Q(A_{k+2}) \ge e^{-\ell(k)} q_0^2$ and $e^{-Ck} \le e^{-3k} q_0^{2k} \le e^{-3k} q_0^2 \le e^{-\ell(k)} q_0^2$,

$$\text{LHS} = Q(A_1 \cap A_{k+2}) \ge 2e^{-\ell(k)} Q(A_1)\, Q(A_{k+1}) - e^{-Ck} \ge e^{-\ell(k)} Q(A_1)\, Q(A_{k+1}),$$

and thus by $\ell(k) \le \ell(k+2) \le k+2 = L_1$, (7.8) holds.

Suppose (7.8) holds for $j < n$. Then because $d(S_{n-1}, S'_{n-1}) = \ell(L_{n-1})$,

$$Q\left(\bigcap_{j \in S_n} A_j\right) \ge 2e^{-\ell(L_{n-1})} Q\left(\bigcap_{j \in S_{n-1}} A_j\right) Q\left(\bigcap_{j \in S'_{n-1}} A_j\right) - e^{-CL_{n-1}}.$$

By the induction hypothesis,

$$\begin{aligned} e^{-\ell(L_{n-1})} Q\left(\bigcap_{j \in S_{n-1}} A_j\right) Q\left(\bigcap_{j \in S'_{n-1}} A_j\right) &\ge e^{-\ell(L_{n-1})} e^{-2L_{n-1}} \prod_{j \in S_n} Q(A_j) \\ &\ge e^{-\ell(L_{n-1})} e^{-2L_{n-1}} q_0^{2^n} \ge e^{-3L_{n-1}} q_0^{2L_{n-1}} \\ &= e^{-CL_{n-1}}. \end{aligned}$$

Therefore,

$$Q\left(\bigcap_{j \in S_n} A_j\right) \ge e^{-\ell(L_{n-1})} e^{-2L_{n-1}} \prod_{j \in S_n} Q(A_j) = e^{-L_n} \prod_{j \in S_n} Q(A_j),$$

completing the induction for (7.8). To complete the proof, we need to show

$$\lim_{n \to \infty} \frac{|S_n|}{\operatorname{diam}(S_n)} > 0$$

exists. From

$$\frac{\operatorname{diam}(S_n)}{|S_n|} = \frac{L_n}{2^n} = \frac{2L_{n-1} + \ell(L_{n-1})}{2^n} = \frac{L_{n-1}}{2^{n-1}} + \frac{\ell(L_{n-1})}{2^n}$$

it follows

$$\frac{\operatorname{diam}(S_{n-1})}{|S_{n-1}|} \le \frac{\operatorname{diam}(S_n)}{|S_n|} \le \frac{\operatorname{diam}(S_{n-1})}{|S_{n-1}|} \left(1 + \frac{\ell(L_{n-1})}{2L_{n-1}}\right),$$

yielding the convergence of $\operatorname{diam}(S_n)/|S_n|$ and, since

$$\sum_{n=1}^{\infty} \frac{\ell(L_n)}{2L_n} \le \sum_{n=1}^{\infty} \ell(L_n) \sum_{i=L_n}^{L_{n+1}-1} \frac{1}{i(i+1)} \le \sum_{n=1}^{\infty} \sum_{i=L_n}^{L_{n+1}-1} \frac{\ell(i)}{i(i+1)} = \sum_{n=1}^{\infty} \frac{\ell(n)}{n(n+1)} < \infty,$$

there is $\lim_{n \to \infty} [\operatorname{diam}(S_n)/|S_n|] < \infty$, which completes the proof. □

Acknowledgments. The author would like to thank Amir Dembo and Ioannis Kontoyiannis for useful discussion.

REFERENCES

[1] BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, MA.

[2] BRYC, W. and DEMBO, A. (1996). Large deviations and strong mixing. *Ann. Inst. H. Poincaré Probab. Statist.* **32** 549–569.

[3] CHI, Z. (2001). The first order asymptotic of waiting times with distortion between stationary processes. *IEEE Trans. Inform. Theory* **47** 338–347.

[4] DEMBO, A. and KONTOYIANNIS, I. (1999). The asymptotics of waiting times between stationary processes, allowing distortion. *Ann. Appl. Probab.* **9** 413–429.

[5] DEMBO, A. and ZEITOUNI, O. (1992). *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston.

[6] HAMMERSLEY, J. M. (1962). Generalization of the fundamental theorem on sub-additive functions. *Proc. Cambridge Philos. Soc.* **58** 235–238.

[7] KONTOYIANNIS, I. (1998). Asymptotic recurrence and waiting times for stationary processes. *J. Theoret. Probab.* **11** 795–811.

[8] LUCZAK, T. and SZPANKOWSKI, W. (1997). A suboptimal lossy data compression based on approximate pattern matching. *IEEE Trans. Inform. Theory* **43** 1439–1451.

[9] SZPANKOWSKI, W (1993). Asymptotic properties of data compression and suffix trees. *IEEE Trans. Inform. Theory* **39** 1647–1659.

[10] WYNER, A. D. and ZIV, J. (1989). Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inform. Theory* **35** 1250–1258.

[11] YANG, E. and KIEFFER, J. C. (1998). On the performance of data compression algorithms based upon string matching. *IEEE Trans. Inform. Theory* **44** 47–65.

[12] YANG, E. and ZHANG, Z. (1999). On the redundancy of lossy source coding with abstract alphabets. *IEEE Trans. Inform. Theory* **45** 1092–1110.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637
E-MAIL: chi@galton.uchicago.edu