Large Deviations for Dynamical Systems

Richard Sharp, University of Warwick *

Abstract

We give a survey of large deviations results in dynamical systems. We discuss abstract Level 1 and Level results and their application to hyperbolic systems. We also discuss exponential and polynomial results for systems modelled by Young towers.

1 Introduction

1.1 The Ergodic Theorem

Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T : X \to X$ be an ergodic measure-preserving transformation. In this setting, the Birkhoff Ergodic Theorem says that, for every observable $f \in L^1(X, \mu)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \int f \, d\mu, \tag{1}$$

for μ-a.e. $x \in X$. This is the starting point for examining other statistical properties (for more restricted classes of observable). One direction is to ask whether one can normalise the sums by a sequence growing more slowly than n and obtain a limit with non-trivial distribution. Under appropriate conditions, this leads to the Central Limit Theorem (where the normalisation is by $\sqrt{n}$) and ultimately the Almost Sure Invariance Principle (where the sums $\sum_{j=0}^{n-1} f(T^j x)$ are approximated by a Brownian motion), or other stable laws. Another direction, which we shall pursue in this lecture, is to quantify deviations of the averages away from their limit. For example, we may wish to estimate

$$\mu \left\{ x \in X : \left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) - \int f \, d\mu \right| > \epsilon \right\},$$

for $\epsilon > 0$, as $n \to \infty$. The study of such problems is called Large Deviations and in the classical setting one obtains exponential decay with description of the rate in terms of the so-called thermodynamic formalism. We shall make this more precise shortly.

*These notes are an extended version of lectures delivered during the workshop "Statistical Properties of Nonequilibrium Dynamical Systems" at SUSTech, Shenzhen, China in July 2016. The author is grateful to the organisers of the school for the opportunity to speak and to the Department of Mathematics at SUSTech for its hospitality.

1.2 A little history

As is usual in the study of statistical properties in dynamical systems, large deviations were first considered for averages of i.i.d. random variables. The initial results were obtained in this setting by Harald Cramér in 1938. However, the probabilistic theory and its applications, particularly in statistical mechanics, exploded in the 1970s after the work of Donsker and Varadhan in the US and Wentzell in the Soviet Union. See [14] (available on the author's webpage) for a nice historical survey and references and the books [11] and [13] for more comprehensive accounts.

Early work on the application of large deviations to dynamical systems appears independently in the work of Denker [12], Lopes [18], Orey and Pelikan [22], Takahashi [29, 30], Vaienti [31], and Young [40]. A more abstract and general approach was given by Kifer [17] and it is essentially this approach that we shall follow.

More recent work has both extended that range of the classical theory where exponential bounds hold [4], [7], [8], [9], [10], [15], [24], [27], and has investigated that weaker bounds that hold for nonuniformly hyperbolic systems [19], [20], [26].

1.3 Level 1 and Level 2

When studying large deviations for dynamical systems we usually look at two levels of results. We begin by discussing so called Level 1 large deviations, which corresponds to the problem at the start. As before, suppose that $T : X \to X$ is an ergodic measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$, where X is a compact metric space and T is continuous, and let $f \in C(X, \mathbb{R})$. We shall use the notation

$$f^n(x) := \sum_{j=0}^{n-1} f(T^j x).$$

Level 1 large deviations involves estimating

$$\mu \left\{ x \in X : \frac{f^n(x)}{n} \in A \right\} \tag{2}$$

for subsets $A \subset \mathbb{R}$, as $n \to \infty$.

However, one can take a more abstract point of view. Let $\mathcal{M}(X)$ denote the set of Borel probability measures on X, equipped with the weak* topology. For $x \in X$, let

$$\nu_{x,n} := \frac{1}{n} \left(\delta_x + \delta_{Tx} + \cdots + \delta_{T^{n-1}x} \right),$$

where δ_x denotes the Dirac measure at x. We call the $\nu_{x,n}$ (normalised) empirical measures. Level 2 large deviations involves estimating

$$\mu \left\{ x \in X : \nu_{x,n} \in \mathcal{A} \right\}, \tag{3}$$

for subsets $\mathcal{A} \subset \mathcal{M}(X)$, as $n \to \infty$.

In both cases, the classical theory, which holds in the presence of uniform hyperbolicity, gives that when A does not contain $\int f \, d\mu$ or when $\mathcal{A}$ does not contain μ the quantities in

(2) and (3) decay exponentially fast with an explicit rate function depending on A or $\mathcal{A}$. This rate function is related to the thermodynamic formalism associated to the dynamical system.

Remark 1.1. It is worthwhile emphasising that this exponential decay is *not* related to a spectral gap. It is a fundamentally different type of phenomenon to exponential decay of correlations.

2 Background

2.1 Thermodynamic formalism

Let $T : X \to X$ be a continuous mapping of a compact metric space. (What follows will also hold *mutatis mutandis* for flows.) Let $\mathcal{M}(X)$ denote the set of Borel probability measures on X, equipped with the weak* topology, and let $\mathcal{M}_T(X)$ denote the subset of T-invariant probabilities. We assume the reader is familiar with the definitions of the topological entropy $h(T)$ in terms of spanning or separating sets and, for $m \in \mathcal{M}_T(X)$, the measure-theoretic entropy $h_m(T)$. We call the function $m \mapsto h_m(T)$ the *entropy map*. We also assume familiarity with the pressure function $P : C(X, \mathbb{R}) \to \mathbb{R}$, again defined in terms of spanning or separating sets. See [33] for all these definitions.

An important result is the so-called *Variational Principle* for pressure (and, by setting $f = 0$, for entropy). This is due to Walters [32] and also appears as Theorem 9.10 of [33].

Theorem 2.1 (Walters)**.** *For all $f \in C(X, \mathbb{R})$, we have*

$$P(f) = \sup \left\{ h_m(T) + \int f \, dm \, : \, m \in \mathcal{M}_T(X) \right\}.$$

Given $f \in C(X, \mathbb{R})$, a measure for which the supremum in the Variational Principle is attained is called an *equilibrium state* for f. Equilibrium states do not always exist but their existence for all $f \in C(X, \mathbb{R})$ is guaranteed if the entropy map is upper semi-continuous. Even if equilibrium states exist they are not necessarily unique but they are unique in certain nice settings (for example if T is a hyperbolic diffeomorphism or expanding map and f is Hölder continuous). If an equilibrium state is unique it is automatically ergodic.

2.2 Convex analysis and duality

We need to know a small amount about abstract convex analysis. Let $\mathcal{X}$ be a locally convex topological vector space over $\mathbb{R}$ and let $\mathcal{X}^*$ be its dual space (i.e. the space of continuous linear functional from $\mathcal{X}$ to $\mathbb{R}$). Let $\langle \cdot, \cdot \rangle : \mathcal{X}^* \times \mathcal{X} \to \mathbb{R}$ denote the duality pairing.

The *Fenchel-Legendre transform* of a function $\Phi : \mathcal{X}^* \to \mathbb{R} \cup \{\pm\infty\}$ is the function $\Phi^* : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ defined by

$$\Phi^*(x) = \sup_{\omega \in \mathcal{X}^*} \left(\langle \omega, x \rangle - \Phi(\omega) \right).$$

Theorem 2.2 (Duality Theorem [11], Lemma 4.5.8)**.** *Let* $\Psi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ *be a convex and lower semi-continuous function and define*

$$\Phi(\omega) = \sup_{x \in \mathcal{X}} \left(\langle \omega, x \rangle - \Psi(x) \right).$$

Then Ψ *is the Fenchel-Legendre transform of* Φ*.*

We will apply this with $\mathcal{X} = C(X, \mathbb{R})$. Functions on the dual will only be defined on $\mathcal{M}(X)$ but can be extended to the whole of $C(X, \mathbb{R})^*$ by setting them to be equal to $+\infty$ elsewhere.

3 An abstract large deviation theorem

3.1 Rate functions and large deviations

Let X be a compact metric space and let $T : X \to X$ be a continuous map such that $h(T) < \infty$. Suppose that the entropy map $m \mapsto h_m(T)$ is upper semi-continuous. (A suitably amended version of what follows also works for flows.) Let $g \in C(X, \mathbb{R})$ be a fixed potential such that g has a unique equilibrium state μ.

Rather than restricting to results about μ, we shall allow the flexibility of considering a sequence of probability measures $\mu_n \in \mathcal{M}(X)$ (which we do not assume to be T-invariant).

Assumption A. For each $f \in C(X, \mathbb{R})$, we have

$$\limsup_{n \to \infty} \frac{1}{n} \log \int e^{f^n(x)} \, d\mu_n \leq P(f + g) - P(g).$$

Assumption B. There exists a $\| \cdot \|_\infty$-dense subspace V of $C(X, \mathbb{R})$, such that, for every $f \in V$, $f + g$ has a unique equilibrium state and we have

$$\lim_{n \to \infty} \frac{1}{n} \log \int e^{f^n(x)} \, d\mu_n = P(f + g) - P(g).$$

The decay rate in the large deviations results will be obtained from the following function.

Definition 3.1. Define a rate function $I : \mathcal{M}(X) \to \mathbb{R} \cup \{+\infty\}$ by

$$I(m) = \sup \left\{ \int f \, dm - P(f + g) + P(g) : \, f \in C(X, \mathbb{R}) \right\}.$$

In fact, there is a more explicit formula for $I(m)$, given in the next lemma.

Lemma 3.2.

$$I(m) = \begin{cases} P(g) - \int g \, dm - h_m(T) & \text{if } m \in \mathcal{M}_T(X) \\ +\infty & \text{if } m \in \mathcal{M}(X) \setminus \mathcal{M}_T(X). \end{cases}$$

Proof. Write

$$\Theta_g(m) = \begin{cases} -h_m(T) - \int g\,dm & \text{if } m \in \mathcal{M}_T(X) \\ +\infty & \text{if } m \in \mathcal{M}(X) \setminus \mathcal{M}_T(X). \end{cases}$$

Since the entropy map is affine and upper semi-continuous, $\Theta_g : \mathcal{M}(X) \to \mathbb{R} \cup \{+\infty\}$ is convex and lower semi-continuous. We have

$$P(f+g) = \sup_{m \in \mathcal{M}_T(X)} \left(h_m(T) + \int (f+g)\,dm \right) = \sup_{m \in \mathcal{M}(X)} \left(\int f\,dm - \Theta_g(m) \right).$$

By the Duality Theorem, we have that

$$\Theta_g(m) = \sup_{f \in C(X,\mathbb{R})} \left(\int f\,dm - P(f+g) \right) = -P(g) + \sup_{f \in C(X,\mathbb{R})} \left(\int f\,dm - P(f+g) + P(g) \right).$$

Rearranging given

$$I(m) = \Theta_g(m) + P(g),$$

which gives the required formula. ☐

Corollary 3.3. *I is lower semi-continuous.*

It is now easy to get a large deviations upper bound with rate function I.

Theorem 3.4 (Kifer [17])**.** *Suppose Assumption A holds. Then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n \{x \in X : \nu_{x,n} \in \mathcal{K}\} \le -\inf\{I(m) : m \in \mathcal{K}\}, \tag{4}$$

for every closed $\mathcal{K} \subset \mathcal{M}(X)$.

Proof. For $\mathcal{A} \subset \mathcal{M}(X)$, write $\rho(\mathcal{A}) := \inf\{I(m) : m \in \mathcal{A}\}$. Let $\mathcal{K} \subset \mathcal{M}(X)$ be closed and hence compact. The Left Hand Side of inequality (4) is at most zero, so if $\rho(\mathcal{K}) \le 0$ there is nothing to prove. So suppose $\rho(\mathcal{K}) > 0$. For $\epsilon > 0$ and $f \in C(X,\mathbb{R})$, define

$$\mathcal{U}(f,\epsilon) := \left\{ m \in \mathcal{M}(X) : \int f\,dm - P(f+g) + P(g) > \rho(\mathcal{K}) - \epsilon \right\};$$

these are open sets. From the definition of I, it is clear that

$$\mathcal{K} \subset \{m \in \mathcal{M}(X) : I(m) > \rho(\mathcal{K}) - \epsilon\} = \bigcup_{f \in C(X,\mathbb{R})} \mathcal{U}(f,\epsilon),$$

so $\{\mathcal{U}(f,\epsilon)\}_{f \in C(X,\mathbb{R})}$ is an open cover of $\mathcal{K}$. Since $\mathcal{K}$ is compact, we can find $f_1, \ldots, f_k \in C(X,\mathbb{R})$ such that

$$\mathcal{K} \subset \bigcup_{i=1}^{k} \mathcal{U}(f_i,\epsilon).$$

Writing

$$A_{i,n} = \{x \in X : f_i^n(x) - n(P(f_i + g) - P(g) + \rho(\mathcal{K}) - \epsilon) > 0\},$$

we then have

$$\begin{aligned}
\mu_n \{x \in X : \nu_{x,n} \in \mathcal{K}\} &\leq \sum_{i=1}^{k} \mu_n\{x \in X : \nu_{x,n} \in \mathcal{U}(f_i, \epsilon)\} \\
&= \sum_{i=1}^{k} \mu_n \left\{x \in X : \int f_i \, d\nu_{x,n} > P(f_i + g) - P(g) + \rho(\mathcal{K}) - \epsilon\right\} \\
&= \sum_{i=1}^{k} \int \chi_{A_{i,n}} \, d\mu_n \\
&\leq \sum_{i=1}^{k} e^{-n(P(f_i+g)-P(g)+\rho(\mathcal{K})-\epsilon)} \int e^{f_i^n(x)} \chi_{A_{i,n}} \, d\mu_n \\
&\leq \sum_{i=1}^{k} e^{-n(P(f_i+g)-P(g)+\rho(\mathcal{K})-\epsilon)} \int e^{f_i^n(x)} \, d\mu_n,
\end{aligned}$$

since $e^{f_i^n(x)-n(P(f_i+g)-P(g)+\rho(\mathcal{K})-\epsilon)} > 1$ on $A_{i,x}$ and is positive everywhere. Taking logs, dividing by n and taking the lim sup, we get

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n\{x \in X : \nu_{x,n} \in \mathcal{K}\} \leq -\rho(\mathcal{K}) + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, the inequality (4) follows.

Finally, suppose that $\rho(\mathcal{K})$ is infinite. For $N > 0$ and $f \in C(X, \mathbb{R})$, write

$$\mathcal{V}(f, N) := \left\{m \in \mathcal{M}(X) : \int f \, dm - P(f + g) + P(g) > N\right\}.$$

Arguing as above, with $\mathcal{V}(f, N)$ replacing $\mathcal{U}(f, \epsilon)$, we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n\{x \in X : \nu_{x,n} \in \mathcal{K}\} \leq -N.$$

Since N can be taken arbitrarily large, (4) follows in this case also. □

Theorem 3.5. *Suppose Assumption A and Assumption B hold. Then*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n \{x \in X : \nu_{x,n} \in \mathcal{U}\} \geq -\inf\{I(m) : m \in \mathcal{U}\},$$

for every open $\mathcal{U} \subset \mathcal{M}(X)$.

Proof. Omitted. The theorem is essentially Theorem C from [9]. This improves on the statement in [17] (where it is the second part of Theorem 2.1) in that there is no requirement that V be spanned by a countable collection of elements. See the remarks after Theorem C and Appendix B of [9] for an illuminating discussion. □

A Level 1 large deviations result immediately follows from Theorem 3.4 and 3.5.

Corollary 3.6. *Let $f \in C(X,\mathbb{R})$. For $\alpha \in \mathbb{R}$, set*

$$J(\alpha) = \inf\left\{ I(m) : m \in \mathcal{M}_T(X) \text{ such that } \int f\, dm = \alpha \right\}.$$

Under the hypothesis of Theorem 3.4, we have

$$\limsup_{n\to\infty} \log \mu_n \left\{ x \in X : \frac{f^n(x)}{n} \in K \right\} \leq -\inf\{J(\alpha) : \alpha \in K\},$$

for every compact $K \subset \mathbb{R}$. Under the additional hypothesis of Theorem 3.5, we have

$$\liminf_{n\to\infty} \log \mu_n \left\{ x \in X : \frac{f^n(x)}{n} \in U \right\} \geq -\inf\{J(\alpha) : \alpha \in U\},$$

for every open $U \subset \mathbb{R}$.

Proof. Let

$$\mathcal{K} = \left\{ m \in \mathcal{M}(X) : \int f\, dm \in K \right\}.$$

Then $\mathcal{K}$ is closed and $\nu_{x,n} \in \mathcal{K}$ if and only if $f^n(x)/n \in K$ and so the first result follows from Theorem 3.4. The second result follows from Theorem 3.5 by a similar argument once we set $\mathcal{U} = \left\{ m \in \mathcal{M}(X) : \int f\, dm \in U \right\}$. □

As mentioned in the introduction, we note that for sets in $\mathcal{M}(X)$ which do not include μ we have exponential decay:

Lemma 3.7. *Suppose that μ is the unique equilibrium state for g. If $\mathcal{K} \subset \mathcal{M}(X)$ is compact and $\mu \notin \mathcal{K}$ then $\inf\{I(m) : m \in \mathcal{K}\} > 0$.*

Proof. It follows from Lemma 3.2 and the Variational Principle that $I(m) > 0$ whenever $m \neq \mu$. Since I is lower semi-continuous, the result follows. □

3.2 Applications

We formulated the results is the preceding section in terms of a sequence of (not necessarily invariant) probability measures μ_n. In applications, the three main examples of sequences μ_n we have in mind are the following:

1. $\mu_n = \mu$ for all n;

2. μ_n is a weighted average of periodic point measures,

$$\mu_n = \mu_n^{(\text{per})} = \frac{\sum_{T^n x = x} e^{g^n(x)} \nu_{x,n}}{\sum_{T^n x = x} e^{g^n(x)}};$$

3. (if T is non-invertible) μ_n is a weighted average of measures supported on pre-images,

$$\mu_n = \mu_n^{(\mathrm{pre})} = \frac{\sum_{T^n x = x_0} e^{g^n(x)} \delta_x}{\sum_{T^n x = x_0} e^{g^n(x)}},$$

for some $x_0 \in X$.

We then have the following results.

Theorem 3.8. *Let $T : X \to X$ be a uniformly hyperbolic diffeomorphism and let $g : X \to \mathbb{R}$ be Hölder continuous. Let μ be the unique equilibrium state for g. Then the conclusions of Theorems 3.4 and 3.5 hold for the sequences $\mu_n = \mu$ and $\mu_n = \mu_n^{(\mathrm{per})}$.*

Theorem 3.9. *Let $T : X \to X$ be a uniformly expanding map and let $g : X \to \mathbb{R}$ be Hölder continuous. Let μ be the unique equilibrium state for g. Then the conclusions of Theorems 3.4 and 3.5 hold for the sequences $\mu_n = \mu$, $\mu_n = \mu_n^{(\mathrm{per})}$ and $\mu_n = \mu_n^{(\mathrm{pre})}$.*

Assumptions A and B are standard for μ, $\mu_n^{(\mathrm{per})}$ and, for expanding maps, $\mu_n^{(\mathrm{pre})}$ in these settings, where V can be taken to be the space of Hölder continuous functions.

Theorem 3.9 covers the case where $T : X \to X$ is a hyperbolic rational map of the Riemann sphere restricted to its Julia set (see also [8]). A version of Theorem 3.4 for pre-images was proved for general rational maps in [25], subject to Urbanski's condition that $P(g) > \sup_{x \in X} g(x)$. Both upper and lower bounds were obtained for all Hölder g for rational maps satisfying the Topological Collet-Eckmann condition (TCE). An easy way of stating this is to say that T satisfies TCE if there exists $\lambda > 1$ and $r > 0$ such that for every $x \in X$, every $n \geq 1$, and every connected component W of $T^{-n}(B(x,r)$, we have

$$\mathrm{diam}(W) \leq \lambda^{-n}.$$

Rational maps satisfying TCE have a good thermodynamic formalism and Assumptions A and B hold with again V equal to the Hölder continuous functions.

Theorem 3.10 (Comman and Rivera-Letelier [9])**.** *Let $T : X \to X$ be a rational map, restricted to its Julia set, satisfying TCE and let $g : X \to \mathbb{R}$ be Hölder continuous. Let μ be the unique equilibrium state for g. Then the conclusions of Theorems 3.4 and 3.5 hold for $\mu_n = \mu$, $\mu_n = \mu_n^{(\mathrm{per})}$ and $\mu_n = \mu_n^{(\mathrm{pre})}$.*

3.3 Weak* convergence of μ_n

Suppose that we are in the setting of section 1.6 and that the sequence μ_n satisfies Assumptions A and B. It is not necessarily the case that μ_n converges to μ in the weak* topology. However, if the measures μ_n are (eventually) T-invariant, then we do have weak* convergence. We prove this in the next lemma and then give a counterexample to the general statement.

Lemma 3.11. *Suppose that the measures μ_n are T-invariant. Suppose Assumption (A) holds. Then μ_n converges to μ weak*, as $n \to \infty$.*

Proof. Let $f \in C(X, \mathbb{R})$ and let $\epsilon > 0$. Write

$$E(\epsilon, n) = \left\{ x \in X : \left| \frac{f^n(x)}{n} - \int f \, d\mu \right| \geq \epsilon \right\}.$$

Then, using the T-invariance of μ_n for the first equality,

$$\int f \, d\mu_n = \frac{1}{n} \int f^n \, d\mu_n = \frac{1}{n} \int_{E(\epsilon,n)} f^n \, d\mu_n + \frac{1}{n} \int_{X \setminus E(\epsilon,n)} f^n \, d\mu_n.$$

Now

$$\left| \frac{1}{n} \int_{E(\epsilon,n)} f^n \, d\mu_n \right| \leq \|f\|_\infty \mu_n(E(\epsilon, n)),$$

which converges to zero by Lemma 3.7. On the other hand,

$$\mu(X \setminus E(\epsilon, n)) \left(\int f \, d\mu - \epsilon \right) \leq \frac{1}{n} \int_{X \setminus E(\epsilon,n)} f^n \, d\mu_n \leq \mu(X \setminus E(\epsilon, n)) \left(\int f \, d\mu + \epsilon \right).$$

Since $\epsilon > 0$ is arbitrary, we conclude that

$$\lim_{n \to \infty} \int f \, d\mu_n = \int f \, d\mu.$$

□

Now let $T : X \to X$ be a mixing one-sided subshift of finite type and let $g : X \to \mathbb{R}$ be Hölder continuous. Consider the sequence $\mu_n^{(\mathrm{pre})}$ defined in section 1.6. Associated to g is the transfer operator $L_g : C(X, \mathbb{C}) \to C(X, \mathbb{C})$ defined by

$$(L_g w)(x_0) = \sum_{Tx = x_0} e^{g(x)} w(x).$$

By the well-known Ruelle–Perron–Frobenius Theorem [23], L_g has $e^{P(g)}$ as an eigenvalue and the dual operator has an associated eigenprobability m, i.e.

$$\int L_g w \, dm = e^{P(g)} \int w \, dm,$$

for all $w \in C(X, \mathbb{C})$. The measure m is equal to μ if and only if the constant functions are eigenfunctions for the eigenvalue $e^{P(g)}$. In general, this is not the case and then m is not T-invariant. However, in all cases, the measures $\mu_n^{(\mathrm{pre})}$ converge weak* to m.

Remark 3.12. A more abstract view is taken in [9]. Define a measure a sequence of measures Υ_n on $M(X)$ by

$$\Upsilon_n(\mathcal{A}) = \frac{\sum_{\substack{T^n x = x_0 \\ \nu_{x,n} \in \mathcal{A}}} e^{g^n(x)}}{\sum_{T^n x = x_0} e^{g^n(x)}}.$$

Then Υ_n converges weak* to the Dirac measure at μ.

4 Non-uniformly hyperbolic systems and local results

There are many classes of dynamical system which exhibit hyperbolic behaviour but for which the full Large Deviation Principle discussed about fails to hold. Nevertheless, one can obtain Level 1 results valid for deviations in a neighbourhood of the mean. These results hold for certain non-uniformly hyperbolic systems, specifically those modelled by Young towers with exponential return times.

The idea underpinning the theory of Young towers was introduced by Young in her study of quadratic interval maps [41]. She subsequently extended this to a more general theory, covering a wealth of examples, in her very influential paper [42]. (The objects we call Young towers are called Markov towers in [42].) To explain the idea, suppose that $T : X \to X$ is a $C^{1+\alpha}$ map and that $Y \subset X$ is a subset on which T is uniformly hyperbolic. We also suppose that T has an invariant measure μ that is absolutely continuous with respect to the ambient volume measure. Then there is a return time function $R : Y \to \mathbb{N} \cup \{\infty\}$ defined by

$$R(x) = \inf\{n \in \mathbb{N} : T^n x \in Y\}$$

and, by the Poincaré Recurrence Theorem, R is finite μ-almost everywhere. The Young tower is the set

$$\Delta = \{(x, k) : x \in Y,\ 0 \le k \le R(x) - 1\},$$

along with the dynamics $F : \Delta \to \Delta$ given by

$$F(x,k) = \begin{cases} (x, k+1) & \text{if } 0 \le k < R(x) - 1 \\ T^{R(x)}x & \text{if } k = R(x) - 1. \end{cases}$$

The dynamics of T and F are related by the map $\pi : \Delta \to X$ defined by $\pi(x, k) = T^k x$; clearly, $T \circ \pi = \pi \circ F$. To simplify matters considerably, one hopes that the map F is more amenable to analysis than T but that results for F can be pushed down to obtain results for T. Conditions that permit this are given as (P1)–(P5) in [42]. A key characteristic of the tower is the sequence

$$\tau_\Delta(n) = \mu\{x \in Y : R(x) > n\}.$$

If $\tau_\Delta(n) = O(e^{-\gamma n})$, for some $\gamma > 0$, then we say that F has exponential return times. For the rest of the section, we assume that this us the case. (We consider the situation of return times exhibiting polynomial decay in the next section.)

The following are examples of systems modelled by Young towers with exponential return times.

1. Planar periodic Lorentz gas. Let Ω be a disjoint union of strictly convex regions in $\mathbb{R}^2/\mathbb{Z}^2$ with C^3 boundaries. There is a natural billiard flow on $(\mathbb{R}^2/\mathbb{Z}^2 \setminus \Omega) \times S^1$, which induces a billiard map $T : X \to X$, where $X = \partial\Omega \times [-\pi/2, \pi/2]$. This is modelled by a Young tower with exponential return times in both the cases of finite horizon (Young [42]) and infinite horizon (Chernov [5]).

2. Piecewise hyperbolic attractors For example, Lotenz, Lozi and Belykh attractors, or higher dimensional systems obtained by coupling. (Chernov [6], Young [42].)

3. Hénon maps and other rank-one attractors. (Wang and Young [34], [35], [36], [38], Young [42].)

4. Non-uniformly expanding maps in one dimension. (Wang and Young [37], Young [42].)

We now state a large deviations result in this setting. Let $f : X \to \mathbb{R}$ be a Hölder continuous function. We assume that $\int f \, d\mu = 0$ and then the variance $\sigma^2(f)$ is defined by

$$\sigma^2(f) = \lim_{n\to\infty} \frac{1}{n} \int (f^n(x))^2 \, d\mu(x).$$

Theorem 4.1 ([20], [27]). *Let $T : X \to X$ be a system with an absolutely continuous invariant probability measure μ, modelled by a Young tower with exponential return times. Let $f : X \to \mathbb{R}$ be a Hölder continuous function with $\int f \, d\mu = 0$ and $\sigma^2(f) > 0$. Then there exist constants $a(f) < 0$ and $b(f) > 0$ such that, for any compact interval $K \subset [a(f), b(f)]$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \mu \left\{ x \in X \, : \, \frac{f^n(x)}{n} \in K \right\} = - \inf\{ J_f(\alpha) \, : \, \alpha \in K\},$$

where $J_f : [a(f), b(f)] \to \mathbb{R}$ is a strictly convex rate function, vanishing at $\int f \, d\mu$. Moreover, J_f is the Fenchel-Legendre transform of

$$e(t) = \lim_{n\to\infty} \frac{1}{n} \log \int e^{tf^n(x)} \, d\mu(x).$$

We will not give the complete proof of this theorem but we will identify the rate function J_f and explain why it is only defined on a neighbourhood of $\int f \, d\mu$. We begin by considering the transfer operator $\mathcal{P} : L^1(\Delta, \mu_\Delta)$ (where μ_Δ is the measure on Δ induced by μ) defined by

$$\int v \, (w \circ F) \, d\tilde{\mu} = \int (\mathcal{P} v) \, w \, d\mu_\Delta,$$

for $v \in L^1(\Delta, \mu_\Delta)$ and $w \in L^\infty(\Delta, \mu_\Delta)$. (Here we are glossing over some arguments that reduce to the case of a mixing Young tower map.) This operator does not have good spectral properties but Young showed that it is better behaved when restricted to a a Banach space $\mathcal{B}$ of weighted Lipschitz functions [42]. In particular, $\mathcal{P} : \mathcal{B} \to \mathcal{B}$ is quasi-compact.

Following [20], one can introduce a weighted version on $\mathcal{P}$. For $z \in \mathbb{C}$, define $\mathcal{P}_z : \mathcal{B} \to \mathcal{B}$ by $\mathcal{P}_z(v) = \mathcal{P}(e^{zf} v)$. Then each $\mathcal{P}_z$ is bounded and $z \mapsto \mathcal{P}_z$ is analytic on $\mathbb{C}$. Theorem 4.1 then follows from a result of Hennion and Hervé which appears as Theorem E* on page 84 of [16]. This abstract result has a number of hypothesis and these were checked in [20]. A more self-contained account appears in [27].

A notable feature of this result is that the rate function is only defined locally, in a neighbourhood of zero. This is related to the domain of analyticity of the function complex function

$$e(z) = \lim_{n\to\infty} \frac{1}{n} \log \int e^{zf^n(x)} \, d\mu(x).$$

Indeed, we have the following result of Ray-Bellet and Young [27] (which can also be extracted from [16]).

Lemma 4.2. *There exist $\theta > 0$ and $\omega > 0$ such that $e(z)$ is well-defined and analytic on*

$$\{z \in \mathbb{C} : |\mathrm{Re}(z)| < \theta, \, |\mathrm{Im}(z)| < \omega\}.$$

Proof. Lemma 4.2 is obtained by relating $e(z)$ to $\mathcal{P}_z$ via the formula

$$e(z) = \lim_{n \to \infty} \frac{1}{n} \log \int \mathcal{P}_z h \, d\nu, \tag{5}$$

where h and ν are, respectively, an eigenfunction and an eigenmeasure for $\mathcal{P}$ (normalised so that $\int h \, d\nu = 1$), and then using the spectral properties of the operator.

If z is real then, for $|z|$ sufficiently small, $\mathcal{P}_z$ has a spectral gap and a simple maximal eigenvalue $\lambda(z)$. (This can fail for large $|z|$ due to the contribution of the higher levels of the tower. More formally, one requires that the so-called *tail pressure* of zf is strictly small than $P(zf)$, the pressure of zf. See Theorem 3.1 of [27].) Hence there is a decomposition

$$\mathcal{P}_z^n v = \lambda(z)^n h_z \int v \, d\nu_z + \mathcal{R}_z^n v, \tag{6}$$

where h_z and ν_z are eigenfunctions/eigenmeasures for $\mathcal{P}_z$ and $\mathcal{R}_z$ has spectral radius strictly smaller than $|\lambda(z)|$.

By standard perturbation theory, the simple maximal eigenvalue $\lambda(z)$ and the above decomposition persists for $z \in \mathbb{C}$ with $\mathrm{Im}(z)$ sufficiently small. (Unlike the bound on the real part of z, the bound on the imaginary part is not particularly related to the tower construction.) Furthermore, since $z \mapsto \mathcal{P} - z$ is analytic, we also have that, for these values of z, $z \mapsto \lambda(z)$ is analytic. The proof is then completed by applying (5) to (6). □

We can then take

$$[a(f), b(f)] = [e'(-\theta), e'(\theta)]$$

in Theorem 4.1, with J_f defined by

$$J_f(\alpha) = \sup_{t \in [-\theta, \theta]} \left(t\alpha - e(t) \right).$$

5 Polynomial Estimates

The exponential estimates in the preceding section fail when the tower has polynomial tails but polynomial bounds still hold in this setting. The optimal result (which subsumes estimates in [20], [26]) is in [19]. In what follows, C will denote an arbitrary positive constant depending on the quantities indicated (although the dependence on the whole systems $T : X \to X$ has been suppressed).

Theorem 5.1 (Melbourne [19])**.** *Let $\beta > 0$. Suppose that $f \in L^\infty(X, \mu)$ and that*

$$\left| \int f \, (h \circ T^n) \, d\mu - \int f \, d\mu \int h \, d\mu \right| \le C(f) \|h\|_\infty n^{-\beta},$$

for all $h \in L^\infty(X,\mu)$ and all $n \geq 1$. Then

$$\mu\left\{x \in X : \left|\frac{f^n(x)}{n} - \int f\, d\mu\right| > \epsilon\right\} \leq C(f,\epsilon)n^{-\beta},$$

for all $n \geq 1$.

The proof relies on the following technical lemma and a general version of Markov's inequality: for $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ increasing,

$$\mu\{x \in X : |f(x)| \geq \epsilon\} \leq \frac{1}{\phi(\epsilon)} \int \phi(|f|)\, d\mu.$$

Lemma 5.2 (Melbourne [19], Lemma 2.1)**.** *Let $\beta > 0$ and $q \geq 1$. Let $f \in L^\infty(X,\mu)$ with $\int f\, d\mu = 0$. Suppose that*

$$\left|\int f\,(h \circ T^n)\, d\mu\right| \leq C(f)\|h\|_\infty n^{-\beta},$$

for all $h \in L^\infty(X,\mu)$ and $n \geq 1$. Then, for all sufficiently large n, we have

$$\int |f^n|^{2q}\, d\mu \leq C(\beta,q)C(f)^q\|f\|_\infty^{2q-1} \begin{cases} n^{2q-\beta} & \text{if } q > \beta, \\ n^q(\log n)^q & \text{if } q = \beta. \end{cases}$$

Proof. See [19]. The idea is to formulate the problem in terms of martingale differences and apply estimates from probability theory [21], [28]. □

Proof of Theorem 5.1. Without loss of generality, suppose that $\int f\, d\mu = 0$. By Markov's inequality with $\phi(t) = t^{2q}$,

$$\mu\left\{x \in X : \frac{|f^n(x)|}{n} > \epsilon\right\} \leq \epsilon^{-2q}n^{-2q}\int |f^n|^{2q}\, d\mu.$$

Now apply Lemma 5.2 with $q > \max\{1,\beta\}$ to get the result. □

Corollary 5.3. *Let $T : X \to X$ with ergodic invariant probability measure μ be a nonuniformly expanding or nonuniformly hyperbolic system modelled by a Young tower with polynomial tail. Then, for some $\beta > 0$,*

$$\mu\left\{x \in X : \left|\frac{f^n(x)}{n} - \int f\, d\mu\right| > \epsilon\right\} = O(n^{-\beta}).$$

The above theorem also yields a Level 2 result.

Theorem 5.4 (Melbourne [19], using an argument from [26])**.** *Suppose that Theorem 5.1 holds for all Hölder continuous functions f with exponent some fixed exponent $\alpha > 0$. Then if $\mathcal{K} \subset \mathcal{M}(X)$ is compact and $\mu \notin \mathcal{K}$ then we have*

$$\mu\{x \in X : \nu_{x,n} \in \mathcal{K}\} \leq C(\mathcal{K})n^{-\beta},$$

for all $n \geq 1$.

Proof. Let $C^\alpha(X,\mathbb{R})$ denote the space of α-Hölder continuous functions in X. Suppose $m \in \mathcal{K}$. Since $\mu \notin \mathcal{K}$ and $C^\alpha(X,\mathbb{R})$ is $\|\cdot\|_\infty$-dense in $C(X,\mathbb{R})$, there exists $\epsilon > 0$ and $f \in C^\alpha(X,\mathbb{R})$ such that $|\int f\,dm - \int f\,d\mu| > \epsilon$. Hence

$$\mathcal{K} \subset \bigcup_{f\in C^\alpha(X,\mathbb{R})} \bigcup_{\epsilon>0} \left\{ m \in \mathcal{M} : \left| \int f\,dm - \int f\,d\mu \right| > \epsilon \right\}.$$

Since $\mathcal{K}$ is compact, we can cover it by finitely many of these sets:

$$\mathcal{K} \subset \bigcup_{i=1}^{k} \left\{ m \in \mathcal{M} : \left| \int f_i\,dm - \int f_i\,d\mu \right| > \epsilon_i \right\},$$

for some $\epsilon_i > 0$ and $f_i \in C^\alpha(X,\mathbb{R})$, $i = 1,\ldots,k$. Therefore, we have

$$\mu\left\{x \in X : \nu_{x,n} \in \mathcal{K}\right\} \leq \sum_{i=1}^{k} \mu \left\{ x \in X : \left| \frac{f_i^n(x)}{n} - \int f_i\,d\mu \right| > \epsilon_i \right\} = O(n^{-\beta}).$$

□

The polynomial decay in the above results is genuine and is not simply a result of limitations of the techniques. Indeed, there exist examples of systems, modelled by Young towers with polynomial tails, with polynomial lower bounds for large deviations [19], [20], [26].

Example 5.5. As an example of this type of setting, we consider Manneville-Pomeau maps. Define $T : [0,1] \to [0,1]$ by

$$Tx = \begin{cases} x(1 + 2^\alpha x^\alpha) & 0 \leq x < 1/2, \\ 2x - 1 & 1/2 \leq x \leq 1, \end{cases}$$

for $0 < \alpha < 1$. This has a unique ergodic invariant probability measure μ equivalent to Lebesgue measure. For this system, the conclusions of Theorem 5.1 hold with $\beta = \alpha^{-1} - 1$. Furthermore, there is an open set of Hölder observables f for which

$$\lim_{n\to\infty} \frac{1}{\log n} \mu \left\{ x \in [0,1] : \left| \frac{f^n(x)}{n} - \int f\,d\mu \right| > \epsilon \right\} = -\beta$$

and if the limit is relaxed to a $\limsup$, it holds for an open dense set of f [19] (see also [20], [26]).

References

[1] V. Araújo, *Large deviations bound for semiflows over a non-uniformly expanding base*, Bulletin of the Brazilian Mathematical Society **38**, 335–376, 2007.

[2] V. Araújo and M. J. Pacifico, *Large deviations for non-uniformly expanding maps*, Journal of Statistical Physics **125**, 415–457, 2006.

[3] M. Benedicks and L.-S. Young, *Markov extensions and decay of correlations for certain Hénon maps*, Géométrie complexe et systèmes dynamiques (Orsay, 1995), Astérisque **261**, 13–56, 2000.

[4] T. Bomfim and P. Varandas, *The gluing orbit property, uniform hyperbolicity and large deviation principles for semiflows*, Journal of Differential Equations **267**, 228–266, 2019.

[5] N. Chernov, *Decay of correlations and dispersing billiards*, Journal of Statistical Physics **94**, 513–556, 1999.

[6] N. Chernov, *Statistical properties of piecewise smooth hyperbolic systems in high dimensions*, Discrete and Continuous Dynamical Systems **5**, 425–448, 1999.

[7] V. Climenhaga, D. Thompson and K. Yamamoto, *Large deviations for systems with non-uniform structure*, Transactions of the American Mathematical Society **369**, 4167–4192, 2017.

[8] H. Comman, *Strengthened large deviations for rational maps and full shifts, with unified proof*, Nonlinearity **22**, 1413–1429, 2009.

[9] H. Comman and J. Rivera-Letelier, *Large deviation principles for non-uniformly hyperbolic rational maps*, Ergodic Theory and Dynamical Systems **31**, 321–349, 2011.

[10] D. Constantine, J.-F. Lafont and D. Thompson, *The weak specification property for geodesic flows on* CAT(-1) *spaces*, arXiv:1606.06253, 2016, to appear in Groups, Geometry and Dynamics.

[11] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Applications of Mathematics **38**, Springer-Verlag, New York, 1998.

[12] M. Denker, *Large deviations and the pressure function*, Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Czech Academy of Sciences, 21–33, 1992.

[13] R. S. Ellis, Entropy, large deviations, and statistical mechanics, Springer-Verlag, New York, 1985.

[14] R. S. Ellis, *An overview of the theory of large deviations and applications to statistical mechanics*, Scandinavian Actuarial Journal **1**, 97–142, 1995.

[15] K. Gelfert and B. Schapira, *Pressures for geodesic flows of rank one manifolds*, Nonlinearity **27**, 1575–1594, 2014.

[16] H. Hennion and L. Hervé, *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness*, Lecture Notes in Mathematics **1766**, Springer–Verlag, Berlin Heidelberg, 2001.

[17] Y. Kifer, *Large deviations in dynamical systems and stochastic processes*, Transactions of the Americal Mathematical Society **321**, 505–524, 1990.

[18] A. Lopes, *Entropy and large deviation*, Nonlinearity **3**, 527–546, 1990.

[19] I. Melbourne, *Large and moderate deviations for slowly mixing dynamical systems*, Proceedings of the American Mathematical Society **137**, 1735–1741, 2009.

[20] I. Melbourne and M. Nicol, *Large deviations in nonuniformly hyperbolic systems*, Transactions of the American Mathematical Society **360**, 6661–6676, 2008.

[21] F. Merlevède, M. Peligrad and S. Utev, *Recent advances in invariance principles for stationary sequences*, Probability Surveys **3**, 1–36, 2006.

[22] S. Orey and S. Pelikan, *Deviations of trajectory averages and the defect in Pesin's formula for Anosov diffeomorphisms*, Transactions of the American Mathematical Society **315**, 741–753, 1989.

[23] W. Parry and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque 187-188, 1-268, 1990.

[24] C.-E. Pfister and W. Sullivan, *Large deviations estimates for dynamical systems without the specification property. Application to the β-shifts*, Nonlinearity **18**, 237–261, 2005.

[25] M. Pollicott and R. Sharp, *Large deviations and the distribution of pre-images of rational maps*, Communications in Mathematical Physics **181**, 733–739, 1996.

[26] M. Pollicott and R. Sharp, *Large deviations for intermittent maps*, Nonlinearity **22**, 2079–2092, 2009.

[27] L. Ray-Bellet and L.-S. Young, *Large deviations in non-uniformly hyperbolic dynamical systems*, Ergodic Theory and Dynamical Systems **28**, 587–612, 2008.

[28] E. Rio, *Théorie asymptotique des processus aléatoires faiblement dépendents*, Mathématiquess et Applications **31**, Springer-Verlag, Berlin, 2000.

[29] Y. Takahashi, *Entropy functional (free energy) for dynamical systems and theire random perturbations*, in Stochastic Analysis: Proceedings of the Taniguchi International Symposium on Stochastic Analysis, Katata and Kyoto, 1982, ed. K. Itô, North-Holland Mathematical Library **32**, 437–467, 1984.

[30] Y. Takahashi, *Asymptotic behaviour of measures of small tubes: Entropy, Lyapunov's exponent and large deviation*, World Scientific Advanced Series in Dynamical Systems, ed. N. Aoki, **5**, 1–21, 1987.

[31] S. Vaienti, *Computing the pressure for Axiom-A attractors by time series and large deviations for the Lyapunov exponent*, Journal of Statistical Physics **56**, 403–413, 1989.

[32] P. Walters, *A variational principle for the pressure of continuous transformations*, Americal Journal of Mathematics **17**, 937–971, 1976.

[33] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics **79**, Springer-Verlag, New York, 1982.

[34] Q. Wang and L.-S. Young, *Strange attractors with one direction ofinstability*, Communications in Mathematical Physics **218**, 1–97, 2001.

[35] Q. Wang and L.-S. Young, *From invariant curves to strange attractors*, Communications in Mathematical Physics **225**, 275–304, 2002.

[36] Q. Wang and L.-S. Young, *Strange attractors in periodically-kicked limit cycles and Hopf bifurcations*, Communications in Mathematical Physics **240**, 509–529, 2003.

[37] Q. Wang and L.-S. Young, *Nonuniformly expanding 1D maps*, Communications in Mathematical Physics **264**, 255–282, 2006.

[38] Q. Wang and L.-S. Young, *Toward a theory of rank one attractors*, Annals of Mathematics **167**, 349–380, 2008.

[39] Z. Yin and E. Chen, *Large-deviation properties in some non-uniformly hyperbolic systems via Pesin theory*, Dynamical Systems: An International Journal **31**, 483–495, 2016.

[40] L.-S. Young, *Some large deviation results for dynamical systems*, Transactions of the American Mathematical Society **318**, 525–543, 1990.

[41] L. S. Young, *Decay of correlations for certain quadratic maps*, Communications in Mathematical Physics **146**, 123–138, 1992.

[42] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Annals of Mathematics **147**, 585–650, 1998.