LARGE DEVIATIONS OF STOCHASTIC PROCESSES

Contents

1. Introduction

In this talk we give a self-contained introduction to large deviations theory, culminating in a few results about large deviations theorems for a few stochastic processes of interest.

For some motivation, we note that many probabilistic problems fall into the following general framework: Let $\{\mu_n\}$ be a sequence of measures on a space $\mathcal{X}$. (Often these are the laws of some appropriately-scaled sequence of X-valued random variables defined on a common probability space.) Then, for some set $A \subseteq \mathcal{X}$, there exists a constant $c(A)$ such that we have $\mu_n(A) \approx \exp(-nc(A))$. Large deviations theory is essentially the study of determining when such an approximation holds, and, if it does hold, what is the value of $c(A)$. In other words, it is often said that large deviations theory is that it is the study of "the exponential rate of decay of rare events".

For the sake of concreteness, let's see some specific basic examples. For one, suppose that $X_1, X_2, \dots$ are iid standard normal random variables defined on the same space, and write $S_n = \frac{1}{n}(X_1 + \cdots + X_n)$. Note that S_n is distributed as $N(0, \frac{1}{n})$, hence for any $\delta > 0$ we have $\mathbb{P}(|S_n| \geq \delta) = \mathbb{P}(Z \geq \delta\sqrt{n})$, where Z represents a standard normal random variable. Now recall that we have

$$\exp\left(-\frac{x^2}{2}\right) \leq \mathbb{P}(Z \geq x) \leq \frac{1}{x}\exp\left(-\frac{x^2}{2}\right) \tag{1}$$

for all $x > 0$. Therefore, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(|S_n| \geq \delta) = \frac{\delta^2}{2}. \tag{2}$$

This can be interpreted, of course, as the heuristic $\mathbb{P}(|S_n| \geq \delta) \approx \exp(-n\frac{\delta^2}{2})$. In this talk we will see a much more general result about empirical averages of iid random variables.

For a simple example outside the setting of averages, suppose that $X_1, X_2, \dots$ are iid from some distribution μ on $\mathbb{R}$, and set $M_n = \max\{X_1, \dots X_n\}$. Note that for any $x \in \mathbb{R}$ we have

$$\mathbb{P}(M_n \leq x) = (\mu((-\infty, x]))^n \tag{3}$$

hence

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(M_n \leq x) = \log \mu((-\infty, x]). \tag{4}$$

In this case we get $\mathbb{P}(M_n \leq x) = \exp(-n \log \mu((-\infty, x]))$ with exact equality.

As a final example, suppose that Σ is a finite set and that μ is some probability measure on Σ, which we write as a vector in $\mathbb{R}^{|\Sigma|}$. Now let $X_1, X_2, \ldots$ be iid samples from μ and write $\overline{\mu}_n^X$ for their empirical measure $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. Take some probability measure ν on Σ, and note that we have:

$$\mathbb{P}(\overline{\mu}_n^X = \nu) \approx \frac{n!}{(n\nu)_1! \cdots (n\nu)_{|\Sigma|}!} \mu_1^{(n\nu)_1} \cdots \mu_{|\Sigma|}^{(n\nu)_{|\Sigma|}} = n! \prod_{i=1}^{|\Sigma|} \frac{\mu_i^{(n\nu)_i}}{(n\nu)_i!} \tag{5}$$

(Note that, if the entries of ν are all rational, then the above holds with equality whenever n is a multiple of the least common denominator of all the entries of ν.) Now use Stirling's approximation to get:

$$\begin{aligned}
\frac{1}{n} \log \mathbb{P}(\overline{\mu}_n^X = \overline{\mu}_m^Y) &= \frac{1}{n} \log \left(n! \prod_{i=1}^{|\Sigma|} \frac{\mu_i^{(n\nu)_i}}{(n\nu)_i!} \right) \\
&= \frac{1}{n} \log(n!) + \frac{1}{n} \sum_{i=1}^{|\Sigma|} ((n\nu)_i \log(\mu_i) - \log((n\nu)_i!)) \\
&\sim \log(n) + \frac{1}{n} \sum_{i=1}^{|\Sigma|} ((n\nu)_i \log(\mu_i) - (n\nu)_i \log((n\nu)_i)) \\
&= \log(n) + \sum_{i=1}^{|\Sigma|} (\nu_i \log(\mu_i) - \nu_i \log(\nu_i)) - \log(n) \\
&= \sum_{i=1}^{|\Sigma|} (\nu_i \log(\mu_i) - \nu_i \log(\nu_i)) \\
&= \sum_{i=1}^{|\Sigma|} \nu_i \log\left(\frac{\mu_i}{\nu_i}\right)
\end{aligned}$$

Notice that this last term is just the relative entropy of ν to μ (also called the Kullback-Liebler divergence of ν from μ), denoted $H(\mu|\nu)$. Summarizing the above we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\overline{\mu}_n^X = \overline{\mu}_m^Y) \approx H(\nu|\mu) \tag{6}$$

which is again an approximation of the sort $\mathbb{P}(\overline{\mu}_n^X = \overline{\mu}_m^Y) \approx \exp(-nH(\nu|\mu))$ that we described generally.

2. Basic Theory

In this section we outline some basic definitions and properties of large deviations theory. In the subsequent sections we'll prove some powerful results of the forms described here and above. Throughout, let $\mathcal{X}$ denote a Hausdorff topological space.

Definition 2.1. We say that $\{\mu_\varepsilon\}_{\varepsilon>0}$ *satisfy a large deviation principle (LDP) in* $\mathcal{X}$ *with rate function* $I : \mathcal{X} \to [0,\infty]$ if I is lower semi-continuous and if for all sets $A \in \mathcal{B}(\mathcal{X})$ we have

$$\begin{aligned} -\inf\{I(x) : x \in A^\circ\} &\leq \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) \\ &\leq \limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) \leq -\inf\{I(x) : x \in \overline{A}\}. \end{aligned}$$

These are called the *LDP lower bound* and the *LDP upper bound* respectively.

The definition of the LDP is a bit dense but it can be made rather intuitive by interpreting I as describing the "rarity" of each element of $\mathcal{X}$. Then applying the LDP to some set A for which the upper bound and the lower bound agree is just the statement that the exponential rate of decay of the probabilites A is determined by the rarest outcome in A. When the bounds do not agree this intuition is still rather useful.

Also note that it is possible to consider the limit with respect to a continuous parameter $\varepsilon \to 0$ or with a discrete parameter $a_n \to 0$ as $n \to \infty$. For the sake of simplicity, we will not focus on the differences between these different types of LDPs.

Next we remark that the topology of $\mathcal{X}$ should be regarded as a central part of the LDP. In particular, the interior and closure operations in the definitions are necessary; there exist plenty of examples which show that the limit need not exist in general. It is straightforward to prove that $\{\mu_\varepsilon\}_{\varepsilon>0}$ satisfies the LDP with rate function I iff it satisfies the LDP lower bound for all open sets and the LDP upper bound for all closed sets.

Now we make some remarks about the differing natures of the lower and upper bounds. First of all, we note that the LDP lower bound is a "local property" in the following sense: Suppose that $\{\mu_\varepsilon\}_{\varepsilon>0}$ and I are such that, for any point $x \in \mathcal{X}$ and any open neighborhood U containing x, we have

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(U) \geq -I(x). \tag{7}$$

Then, it follows that $\{\mu_\varepsilon\}_{\varepsilon>0}$ satisfies the LDP with rate function I. In fact, it suffices to check this property for some basis (or even a subbasis) for the topology of $\mathcal{X}$, for example the open balls when $\mathcal{X}$ is a metric space.

There is also plenty to be said about the nature of the upper bound, but we will not need to understand such properties for the purposes of the talk today. Instead, we'll state and prove an analytical lemma that will be used often when proving the large deviations upper bound:

Lemma 2.2 (winner-take-all lemma)**.** For any nonnegative families of reals $\{a_\varepsilon^1\}_{\varepsilon>0}, \ldots$ $\{a_\varepsilon^N\}_{\varepsilon>0}$, we have the identity

$$\limsup_{\varepsilon\to 0} \varepsilon \log\left(\sum_{i=1}^{N} a_\varepsilon^i\right) = \max_{1\le i\le N}\left\{\limsup_{\varepsilon\to 0} \varepsilon\log a_\varepsilon^i\right\}. \tag{8}$$

Proof. Note that for any $\varepsilon > 0$ we have

$$\begin{aligned} 0 &\le \log\left(\sum_{i=1}^{N} a_\varepsilon^i\right) - \max_{1\le i\le N}\left\{\log a_\varepsilon^i\right\} \\ &= \log\left(\sum_{i=1}^{N} \frac{a_\varepsilon^i}{\max_{1\le i\le N}\left\{\log a_\varepsilon^i\right\}.}\right) \le \log N. \end{aligned}$$

So multiplying by ε taking lim sup as $\varepsilon \to 0$ gives

$$\limsup_{\varepsilon\to 0} \varepsilon \log\left(\sum_{i=1}^{N} a_\varepsilon^i\right) = \limsup_{\varepsilon\to 0} \max_{1\le i\le N}\left\{\varepsilon\log a_\varepsilon^i\right\}. \tag{9}$$

Finally, we claim that we have

$$\limsup_{\varepsilon\to 0} \max_{1\le i\le N}\left\{\varepsilon\log a_\varepsilon^i\right\} = \max_{1\le i\le N}\left\{\limsup_{\varepsilon\to 0} \varepsilon\log a_\varepsilon^i\right\}. \tag{10}$$

If $\max_{1\le i\le N}\left\{\limsup_{\varepsilon\to 0}\varepsilon\log a_\varepsilon^i\right\} = A$, then there is some $j \in \{1,\dots N\}$ and some sequence $\{\varepsilon_k\}_{k=1}^\infty$ such that $\lim_{k\to\infty}\varepsilon_k \log a_{\varepsilon_k}^j = A$. But $\max_{1\le i\le N}\left\{\varepsilon_k\log a_{\varepsilon_k}^i\right\} \ge \varepsilon_k \log a_{\varepsilon_k}^j$ holds for all k, so we have $\limsup_{\varepsilon\to 0}\max_{1\le i\le N}\left\{\varepsilon\log a_\varepsilon^i\right\} \ge A$. Conversely, suppose that $\limsup_{\varepsilon\to 0}\max_{1\le i\le N}\left\{\varepsilon\log a_\varepsilon^i\right\} = B$ so that there exists some $\{\varepsilon_k\}_{k=1}^\infty$ with $\lim_{k\to\infty}\max_{1\le i\le N}\left\{\varepsilon\log a_\varepsilon^i\right\} = B$. Then there must exist some $j \in \{1,\dots N\}$ such that $\max_{1\le i\le N}\left\{\varepsilon\log a_\varepsilon^i\right\} = \varepsilon\log a_\varepsilon^j$ holds for infinitely many k. In particular, there is a subsequence $\{k_n\}_{n=1}^\infty$ satisfying

$$B = \lim_{n\to\infty}\left\{\varepsilon_{k_n}\log a_{\varepsilon_{k_n}}^j\right\} \le \limsup_{\varepsilon\to 0}\left\{\varepsilon\log a_\varepsilon^j\right\} \le \max_{1\le i\le N}\left\{\limsup_{\varepsilon\to 0}\varepsilon\log a_\varepsilon^i\right\}. \tag{11}$$

This proves (10) and finishes the claim. □

Next, we make some remarks about rate functions. A rate function is called *good* if its sublevel sets are compact, which in particular implies that inf in the LDP upper bound is achieved. Under some mild conditions, it can be shown that there is at most one rate function for which a family of measures $\{\mu_\varepsilon\}_{\varepsilon>0}$ can satisfy an LDP.

Finally, we remark that there is a wide array of tools that can be used tp study how to "move" a LDP from one space to another. The following is the most basic and most important example of this:

Lemma 2.3 (Contraction principle)**.** Suppose that $\mathcal{X}$ is a Hausdorff topological space and that $\{\mu_\varepsilon\}_{\varepsilon>0}$ on $\mathcal{X}$ satisfies the LDP with the good rate function I. If $\mathcal{Y}$ is a Hausdorff topological space and $f : \mathcal{X} \to \mathcal{Y}$ is continuous, then $\{f_*\mu_\varepsilon\}_{\varepsilon>0}$ satisfies the LDP with good rate function defined via $J(y) = \inf\{I(x) : f(x) = y\}$.

Proof. For the upper bound, we note that $f^{-1}(A)$ is closed whenever $A \subseteq \mathcal{Y}$ is closed, and also that we have

$$\begin{aligned}\limsup_{\varepsilon\to 0} \varepsilon \log(f_*\mu)(A) &= \limsup_{\varepsilon\to 0} \varepsilon \log \mu(f^{-1}(A)) \\ &\leq \inf\{I(x) : x \in f^{-1}(A)\} \\ &= \inf\{\inf\{I(x) : f(x) = y\} : y \in A\} \\ &= \inf\{J(y) : y \in A\}.\end{aligned}$$

Since $f^{-1}(A)$ is also open whenever A is open, we similarly have:

$$\begin{aligned}\liminf_{\varepsilon\to 0} \varepsilon \log(f_*\mu)(A) &= \liminf_{\varepsilon\to 0} \varepsilon \log \mu(f^{-1}(A)) \\ &\geq \inf\{I(x) : x \in f^{-1}(A)\} \\ &= \inf\{J(y) : y \in A\}.\end{aligned}$$

It only remains to show that J is a good rate function. Since f is continuous and I is good, it suffices to show that for any $\alpha \geq 0$ we have:

$$\{y \in \mathcal{Y} : J(y) \leq \alpha\} = f(\{x \in \mathcal{X} : I(x) \leq \alpha\}) \tag{12}$$

If $y \in f(\{x \in \mathcal{X} : I(x) \leq \alpha\})$, then of course $J(y) \leq \alpha$, so we only need to prove the converse. If $J(y) \leq \alpha$, then there exists some sequence $\{x_n\}_{n=1}^\infty$ in $f^{-1}(\{y\})$ with $I(x_n) \downarrow \alpha' \leq \alpha$. Then $\{x_n\}_{n=1}^\infty$ eventually lies in $f^{-1}(\{y\}) \cap \{x \in \mathcal{X} : I(x) \leq \alpha + 1\}$ which is compact, hence there is a subsequence $\{n_k\}_{k=1}^\infty$ with $x_{n_k} \to z \in f^{-1}(\{y\}) \cap \{x \in \mathcal{X} : I(x) \leq \alpha + 1\}$. In other words, we have $z \in \mathcal{X}$ with $f(z) = y$ and $I(z) \leq \liminf_{k\to\infty} I(x_{n_k}) = \alpha' \leq \alpha$ since I is lower semi-continuous. This gives the opposite inclusion and hence proves the claim. □

Be warned that, if I is not good, then J may fail to be lower semi-continuous.

3. LDPs for Averages of Random Variables

In this section we state and prove some of the "classical" results of large deviations theory, which are primarily concerned with averages of random variables.

The simplest setting, which we consider first is the case of sums of iid random variables in $\mathbb{R}$. We prove this case in full detail, as it provides the main structure that we will use for more complicated results later.

Theorem 3.1 (Crámer)**.** Let $X_1, X_2, \ldots$ be iid real-valued andom variables on a common probability space, and assume that we have $\mathbb{E}[\exp(\lambda X_1)] < \infty$ for all $\lambda \in \mathbb{R}$. Then the empirical averages $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ satisfy a large deviation principle in $\mathbb{R}$ with rate function

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - \Lambda(\lambda)), \tag{13}$$

where

$$\Lambda(\lambda) = \log \mathbb{E}[\exp(\lambda X_1)] \tag{14}$$

is the logarithmic moment generating function of X_1.

Proof. We separate the proof into three main steps.

Rate Function. Note that $\Lambda(0) = 0$, so we have $\Lambda^*(x) \geq 0$ for all $x \in \mathbb{R}$. Moreover, Λ^* is the pointwise supremum of continuous (in fact, affine) functions, so it is convex and lower semi-continuous. Next note that, by Jensen's inequality, we have $\Lambda(\lambda) = \log \mathbb{E}[\exp(\lambda X_1)] \geq \lambda \mathbb{E}[X_1]$ hence $\lambda \mathbb{E}[X_1] - \Lambda(\lambda) \leq 0$ for any $\lambda \in \mathbb{R}$. Taking the supremum and using nonnegativity gives $\Lambda^*(\mathbb{E}[X_1]) = 0$. Finally, note that if we have $x \geq \mathbb{E}[X_1]$ and λ, then

$$\lambda x - \Lambda(\lambda) \leq \lambda \mathbb{E}[X_1] - \Lambda(\lambda) = 0 \tag{15}$$

hence the supremum can be taken over $\lambda \geq 0$. Since each function $\lambda x - \Lambda(\lambda)$ is non-decreasing for $\lambda \geq 0$, this implies that Λ^* is non-decreasing on $(\mathbb{E}[X_1], \infty)$. Similarly, we can show that Λ^* is non-increasing on $(-\infty, \mathbb{E}[X_1])$. Now we proceed to the main proof.

Upper Bound. First we make a Chernoff-type bound. For any $x \geq \mathbb{E}[X_1]$ and any $\lambda \geq 0$, we have

$$\begin{aligned}
\mathbb{P}(S_n \geq x) &= \mathbb{P}\left(\sum_{i=1}^n X_i \geq nx\right) \\
&= \mathbb{P}\left(\exp\left(\lambda \sum_{i=1}^n X_i\right) \leq \exp(n\lambda x)\right) \\
&\leq \exp(-n\lambda x)\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^n X_i\right)\right] \\
&= \exp(-n\lambda x)\left(\mathbb{E}\left[\exp\left(\lambda X_1\right)\right]\right)^n \\
&= \exp(-n\lambda x + n\Lambda(\lambda))
\end{aligned}$$

Now optimize this over all $\lambda \geq 0$ and we get $\mathbb{P}(S_n \geq x) \leq \exp(-n\Lambda^*(x))$. A similar argument shows that for $x \leq \mathbb{E}[X_1]$ we have $\mathbb{P}(S_n \leq x) \leq \exp(-n\Lambda^*(x))$.

Now let $F \subseteq \mathbb{R}$ be any closed set. If $F = \emptyset$, then there is nothing to prove, so suppose F is non-empty. Write $I_F = \inf\{I(x) : x \in F\}$. If $I_F = 0$, then there is nothing to prove, so suppose $I_F > 0$. Then $\mathbb{E}[X_1] \notin F$, so we can let (x_+, x_-) be the union of all open intervals (a, b) satisfying $\mathbb{E}[X_1] \in (a, b) \subseteq \mathbb{R} \setminus F$. Since F is nonempty, at least one of x_+ or x_- must be finite. If x_+ is finite then we have $x_+ \in F$, hence $\Lambda^*(x_+) \geq I_F$. Then use the Chernoff bound above to get $\mathbb{P}(S_n \geq x_+) \leq \exp(-n\Lambda^*(x_+)) \leq \exp(-nI_F)$. On the other hand, if x_- is finite then we have $x_- \in F$ and this implies $\mathbb{P}(S_n \leq x_-) \leq \exp(-n\Lambda^*(x_-)) \leq \exp(-nI_F)$. Combining these, we have

$$\mathbb{P}(S_n \in F) \leq \mathbb{P}(S_n \leq x_-) + \mathbb{P}(S_n \geq x_+) \leq 2\exp(-nI_F), \tag{16}$$

and this is the desired upper bound.

Lower Bound. Take any $x \in \mathbb{R}$. Note that if we define the random variables $Y_1, Y_2, \ldots$ via $Y_i = X_i - x$, then we have $\Lambda_Y(\lambda) = \log \mathbb{E}[\exp(\lambda Y_1)] = \log \mathbb{E}[\exp(\lambda X_1)] - \lambda x = \Lambda(\lambda) - \lambda x$ hence $\Lambda'(\lambda) = \Lambda_Y'(\lambda) + x$. Since Λ_Y is smooth and convex hence

attains a minimum, there must exist some $\eta \in \mathbb{R}$ such that $\Lambda'_Y(\eta) = 0$ and hence that $\Lambda'(\eta) = x$. Now let $\tilde{\mu}$ denote the probability measure on $\mathbb{R}$ which has Radon-Nikodym derivative

$$\frac{d\tilde{\mu}}{d\mu}(t) = \exp(\eta t - \Lambda(\eta)) \tag{17}$$

Note that a random variable Z_1 distributed according to $\tilde{\mu}$ has $\mathbb{E}[Z_1] = \Lambda'(\eta) = x$.

Now keep x as above and take any $\delta > 0$. Note that we have:

$$\begin{aligned}
\mathbb{P}(S_n \in (x-\delta, x+\delta)) &= \int_{\{|S_n - x| < \delta\}} d\mu^{\otimes n}(t_1, \dots t_n) \\
&\geq \exp(-n\eta(x+\delta)) \int_{\{|S_n - x| < \delta\}} \exp\left(\eta \sum_{i=1}^n t_i\right) d\mu^{\otimes n}(t_1, \dots t_n) \\
&= \exp(-n\eta(x+\delta) + n\Lambda(\eta)) \int_{\{|S_n - x| < \delta\}} \exp\left(\eta \sum_{i=1}^n t_i - n\Lambda(\eta)\right) d\mu^{\otimes n}(t_1, \dots t_n) \\
&= \exp(-n\eta(x+\delta) + n\Lambda(\eta)) \int_{\{|S_n - x| < \delta\}} d\tilde{\mu}^{\otimes n}(t_1, \dots t_n)
\end{aligned}$$

Note that the integral term is just the probability that the empirical mean of iid samples from $\tilde{\mu}$ lies within distance δ of x. By the weak law of large numbers, we know that this probability goes to one as n goes to infinity. Now we can take normalized logarithmic limits to get:

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \in (x-\delta, x+\delta)) \\
&\quad \geq -\eta(x+\delta) + \Lambda(\eta) + \liminf_{n\to\infty} \frac{1}{n} \log \left(\int_{\{|S_n - x| < \delta\}} d\tilde{\mu}^{\otimes n}(t_1, \dots t_n) \right) \\
&\quad = -\eta(x+\delta) + \Lambda(\eta) \\
&\quad = -\left(\eta x - \Lambda(\eta)\right) - \eta\delta \\
&\quad \geq -\Lambda^*(x) - \eta\delta
\end{aligned}$$

Now suppose that $0 < \delta' < \delta$. The argument above shows

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \in (x-\delta, x+\delta)) \\
&\quad \geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \in (x-\delta', x+\delta')) \\
&\quad \geq -\Lambda^*(x) - \eta\delta'
\end{aligned}$$

so taking $\delta' \to 0$ gives

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \in (x-\delta, x+\delta)) \geq -\Lambda^*(x) \tag{18}$$

As we have seen, this inequality is sufficient for establishing the lower bound, so the theorem is proved. □

The exponential moment constraint can be significantly relaxed, but, for the sake of simplicity, we do not study the general case here. We also remark that the same theorem holds in $\mathbb{R}^d$ with some easy modifications to the statement and some significant modifications to the proof. (Primarily, the proof of the upper bound is more complicated since it does not make sense for Λ^* to be "monotonic on either side of $\mathbb{E}[X_1]$ in higher dimensions.)

Theorem 3.2 (Crámer)**.** Let $X_1, X_2, \dots$ be iid $\mathbb{R}^d$-valued random variables on a common probability space, and assume that we have $\mathbb{E}[\exp(\langle \lambda, X_1 \rangle)] < \infty$ for all $\lambda \in \mathbb{R}^d$. Then the empirical averages $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ satisfy a large deviation principle in $\mathbb{R}^d$ with rate function

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} (\langle \lambda, x \rangle - \Lambda(\lambda)), \tag{19}$$

where

$$\Lambda(\lambda) = \log \mathbb{E}[\exp(\langle \lambda, X_1 \rangle)] \tag{20}$$

is the logarithmic moment generating function of X_1.

We also note that the proofs above only rely on the independence of the variables insofar as the existence of the limit $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[\exp(\lambda S_n)]$ is concerned. For this reason, it is natural to guess that it is possible to extend the result to a large class of "weakly dependent" random variables. The Gärtner-Ellis theorem makes this precise, but, for the sake of brevity we do not describe this result in detail.

Next we show another classical result, which proves the existence of a large deviations principle for the empirical measure of iid random variables defined on a finite state space.

Theorem 3.3 (Sanov)**.** Let Σ be a set with $|\Sigma| = d$ and let $\mathcal{M}_1(\Sigma)$ denote the collection of all probability measures on Σ. Fix $\mu \in \mathcal{M}_1(\Sigma)$ and and let $\overline{\mu}_n$ denote the empirical measure of n iid samples from μ. Then, the laws of $\{\overline{\mu}_n\}_{n=1}^\infty$ satisfy a large deviations principle in $\mathcal{M}_1(\Sigma)$ with the good rate function $H(\nu|\mu) = \sum_{i=1}^d \nu_i \log(\mu_i/\nu_i)$.

Proof. Without loss of generality we may write the elements of Σ as $\{1, \dots d\}$, and we can view $\mathcal{M}_1(\Sigma)$ as a convex subset of $\mathbb{R}^d$, usually called the d-simplex. Note that, under this perspective, the empirical measure $\overline{\mu}_n$ is just the average of iid $\mathbb{R}^d$-valued random variables $X_1, X_2, \dots$ whose common distribution $\tilde{\mu}$ is such that $\tilde{\mu}(\{e_i\}) = \mu(\{i\})$ holds for all i. For simplicity, write $\mu(\{i\}) = \mu_i$ and similarly for other probability measures on Σ.

By Crámer's theorem in $\mathbb{R}^d$, we know that the laws of $\{\overline{\mu}_n\}_{n=1}^\infty$ satisfy a LDP in $\mathbb{R}^d$ with rate function Λ^*. Since $\overline{\mu}_n \in \mathcal{M}_1(\Sigma)$ holds almost surely, it follows that the laws of $\{\overline{\mu}_n\}_{n=1}^\infty$ satisfy a LDP in $\mathcal{M}_1(\Sigma)$ with rate function Λ^*. Next we show that we have $\Lambda^*(\nu) = H(\nu|\mu)$ for all $\nu \in \mathcal{M}_1(\Sigma)$.

First, note that by Jensen's inequality we have

$$\Lambda(\lambda) = \log\left(\sum_{i=1}^{d} e^{\lambda_i}\mu_i\right) = \log\left(\sum_{i=1}^{d} e^{\lambda_i}\frac{\mu_i}{\nu_i}\nu_i\right)$$
$$\geq \sum_{i=1}^{d}\left(\lambda_i + \log\left(\frac{\mu_i}{\nu_i}\right)\right)\nu_i = \langle\lambda,\nu\rangle - H(\nu|\mu),$$

hence

$$\Lambda^*(\nu) = \sup_{\lambda\in\mathbb{R}}(\langle\lambda,\nu\rangle - \Lambda(\lambda)) \leq H(\nu|\mu). \tag{21}$$

For the opposite inequality, first consider the case that ν is not absolutely continuous with respect to μ and hence that $H(\nu|\mu) = \infty$. That is, there exists some $j \in \Sigma$ with $\mu_j = 0$ and $\nu_j > 0$. Now for $c > 0$ define $\lambda \in \mathbb{R}^d$ via $\lambda_j = c$ and $\lambda_i = 0$ for all $i \neq j$. It follows that we have

$$\langle\lambda,\nu\rangle - \Lambda(\lambda) = c\nu_j - \log\left(\sum_{i\neq j}\mu_i\right) = c\nu_j. \tag{22}$$

Taking $c \to \infty$ shows that we have $\Lambda^*(\nu) = \infty$, so $\Lambda^*(\nu) \geq H(\nu|\mu)$ holds. Otherwise the fraction ν_i/μ_i is well-defined for all i, and we can define $\lambda_i = \log(\nu_i/\mu_i)$. Plugging this into the above gives

$$\Lambda^*(\nu) \geq \langle\lambda,\nu\rangle - \Lambda(\lambda) = H(\nu|\mu), \tag{23}$$

so $\Lambda^*(\nu) \geq H(\nu|\mu)$ holds. We have hence proven $\Lambda^*(\nu) = H(\nu|\mu)$, with ∞ being a possible value.

It only remains to show that $H(\nu|\mu)$ is good. Since $\mathcal{M}_1(\Sigma)$ is compact, we only need to show that $\{\nu \in \mathcal{M}_1(\Sigma) : H(\nu|\mu) \leq \alpha\}$ is closed for any $\alpha \geq 0$. Indeed, if $\{\nu^{(n)}\}_{n=1}^{\infty}$ lie in this set and satisfy $\nu^{(n)} \to \nu \in \mathcal{M}_1(\Sigma)$, then we have

$$H(\nu|\mu) = \sum_{i=1}^{d}\nu_i\log\left(\frac{\mu_i}{\nu_i}\right) = \lim_{n\to\infty}\sum_{i=1}^{d}\nu_i^n\log\left(\frac{\mu_i}{\nu_i^n}\right) \leq \alpha, \tag{24}$$

and the result follows. □

As suggested by the notation, Sanov's theorem holds much mroe generally. In fact, whenever Σ is a Polish space and $\mathcal{M}_1(\Sigma)$ is given the topology of weak convergence, then the same statement holds. However, the proof is significantly more involved, with many more topological and measure-theoretic considerations.

Also, note that we derived Sanov's theorem as a consequence of Crámer's theorem in $\mathbb{R}^d$. However, our heuristic calculations at the beginning of the talk showed that we could roughly justify the same conclusion with more combinatorial methods. It is not hard to show that one can prove the Sanov's theorem rigorously via combinatorial methods, without first proving Crámer's theorem.

4. LDPs for Stochastic Processes

In this section we develop a few large deviations result about continuous-time stochastic processes of interest. The theorems of the previous section serve as a good template for the work here, and we will follow essentially the arguments even in this more complicated case.

Throughout this section, let $\{B_t\}_{t\geq 0}$ denote a Brownian motion on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $T > 0$ denote some fixed (deterministic) time. Let $C([0,T];\mathbb{R})$ denote the space of continuous functions from $[0,T]$ to $\mathbb{R}$, and let $C_0([0,T];\mathbb{R}) \subseteq C([0,T];\mathbb{R})$ denote the subspace of functions with value zero at time zero.

Also recall that a function $f : [0,T] \to \mathbb{R}$ is called *absolutely continuous* if for any $\varepsilon > 0$ there exists some $\delta > 0$ such that any disjoint open intervals $\{(a_k, b_k)\}_{k=1}^{\infty}$ with $\sum_{k=1}^{\infty}(b_k - a_k) < \delta$ satisfy $\sum_{k=1}^{\infty} |f(b_k) - f(a_k))| < \varepsilon$. An absolutely continuous function f has a well-defined derivative Lebesgue-almost-everywhere, which we denote f'. We write $H_1([0,T];\mathbb{R}) \subseteq C_0([0,T];\mathbb{R})$ for the set of functions $f \in C_0([0,T];\mathbb{R})$ which are absolutely continuous and which satisfy $\int_0^T |f'(t)|^2 dt < \infty$.

Theorem 4.1 (Schilder)**.** The laws of $\{\sqrt{\varepsilon}B_t\}_{\varepsilon>0}$ satisfy a large deviation principle in $C_0([0,T];\mathbb{R})$ with good rate function

$$I(\phi) = \begin{cases} \frac{1}{2}\int_0^T |\phi'(t)|^2 dt & \text{if } \phi \in H_1 \\ \infty & \text{else} \end{cases}. \tag{25}$$

Proof. As before, we provide the proof in three main steps.

Rate Function. First let us prove that I is a good rate function, i.e. that for any $\alpha \geq 0$ the set $K_\alpha = \{\phi \in C_0([0,T];\mathbb{R}) : I(\phi) \leq \alpha\}$ is compact in $C_0([0,T];\mathbb{R})$. This will rely heavily on some functional analysis: Note first that for any $\phi \in K_\alpha$, we have, by the fundamental theorem of calculus and Cauchy-Schwarz:

$$\begin{aligned} \sup_{t\in[0,T]} |\phi(t)| &\leq \sup_{t\in[0,T]} \int_0^t |\phi'(s)| ds \\ &\leq \sup_{t\in[0,T]} \left(\int_0^t 1 ds\right)^{\frac{1}{2}} \left(\int_0^t |\phi'(s)|^2 ds\right)^{\frac{1}{2}} \leq \sup_{t\in[0,T]} \sqrt{\alpha t} = \sqrt{2\alpha T}, \end{aligned}$$

hence K_α is uniformly bounded. A similar argument shows that, for any $t_1, t_2 \in [0,T]$ with $t_1 < t_2$ and any $\phi \in K_\alpha$, we have:

$$\begin{aligned} |\phi(t_1) - \phi(t_2)| &\leq \int_{t_1}^{t_2} |\phi'(s)| ds \\ &\leq \left(\int_{t_1}^{t_2} 1 ds\right)^{\frac{1}{2}} \left(\int_{t_1}^{t_2} |\phi'(s)|^2 ds\right)^{\frac{1}{2}} \leq \sqrt{2\alpha(t_2 - t_1)}. \end{aligned}$$

This proves that we have $|\phi(t_1) - \phi(t_2)| \leq \sqrt{2\alpha|t_1 - t_2|}$ for all $\phi \in K$ and all $t_1, t_2 \in [0,T]$, i.e. that K_α is uniformly Hölder continuous with modulus $1/2$. In particular this shows that K_α is uniformly equicontinuous. So, by the Arzela-Ascoli theorem, it follows that K_α is pre-compact in $C_0([0,T];\mathbb{R})$.

Note also that H_1 can be made into a Hilbert space with the inner product $\langle \phi_1, \phi_2 \rangle_{H_1} = \int_0^T \phi_1'(t)\phi_2'(t)dt$, and then it follows that $I(\phi) = ||\phi||_{H_1}^2$. We can write $K_\alpha = \{\phi \in H_1 : ||\phi||_{H_1} \leq \sqrt{\alpha}\}$, and, by the Banach-Alaoglu theorem, this set is weakly-compact.

Therefore, any sequence $\{\phi_n\}_{n=1}^\infty$ in K_α admits a single subsequence $\{\phi_{n_k}\}_{k=1}^\infty$ which converges to $\phi \in C_0([0,T];\mathbb{R})$ in $C_0([0,T];\mathbb{R})$ and which converges to $\tilde{\phi} \in K_\alpha$ weakly in H_1. For each $t \in [0,T]$ define $\psi \in H_1$ via $\psi(s) = s$ for $s \in [0,t]$ and $\psi(s) = t$ for $s \in [t,T]$. Then we have that $\langle \phi, \psi \rangle_{H_1} = \int_0^t \phi'(s)ds = \phi(t)$, so the point evaluations are continuous with respect to the weak topology on H_1. Of course, the point evaluations are also continuous with respect to $C_0([0,T];\mathbb{R})$. This means $\phi_{n_k}(t)$ converges to both $\phi(t)$ and $\tilde{\phi}(t)$ for each $t \in [0,T]$, hence $\phi = \tilde{\phi}$. We have shown that any sequence in K_α contains a subsequence which converges in $C_0([0,T];\mathbb{R})$ to an element of K_α, hence that K_α is compact in the topology of $C_0([0,T];\mathbb{R})$. This shows that I is a good rate function.

Upper Bound. Write $\rho : C([0,T];\mathbb{R})^2 \to [0,\infty)$ for $\rho(f,g) = \sup_{t\in[0,T]} |f(t)-g(t)|$ and for a compact set $K \subseteq C([0,T];\mathbb{R})$ write $\rho(f,K) = \inf_{g\in K} \rho(f,g)$. Now for arbitrary N, write $\hat{B}^N$ for the piecewise-linear process which is equal to B for $t = jT/N$ for integers j. Note that we have

$$\left\{\rho(\sqrt{\varepsilon}B, K_\alpha) \geq \delta\right\} \subseteq \left\{I(\sqrt{\varepsilon}\hat{B}^N) > \alpha\right\} \cup \left\{\rho(\sqrt{\varepsilon}B, \sqrt{\varepsilon}\hat{B}^N) \geq \delta\right\}, \tag{26}$$

so the union bound and the winner-take-all lemma imply that we have

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(\rho(\sqrt{\varepsilon}B, K_\alpha) \geq \delta) \\
&\quad \leq \max\left\{ \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(I(\sqrt{\varepsilon}\hat{B}^N) > \alpha), \right. \\
&\qquad\qquad \left. \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(\rho(\sqrt{\varepsilon}B, \sqrt{\varepsilon}\hat{B}^N) \geq \delta) \right\}.
\end{aligned}$$

For the first term in the maximum, note that the derivative of $\hat{B}^N$ is defined at all points in $[0,T]$ except those of the form jT/N. Moreover, the derivative between these values is just the slope of the secant line which is equal to the size of the increment time N/T. As the increments are Gaussian with variance T/N, we have

$$I(\hat{B}^N) \stackrel{d}{=} \frac{\varepsilon}{2} \sum_{i=1}^N W_i^2 \tag{27}$$

where $W_1, W_2, \ldots$ are iid standard Gaussians. By Chernoff's bound and the moment generating function for chi-squared random variable, we have, for any $\theta \in (0, \frac{1}{2})$:

$$\mathbb{P}(I(\sqrt{\varepsilon}\hat{B}^N) > \alpha) = \mathbb{P}\left(\frac{\varepsilon}{2}\sum_{i=1}^{N} W_i^2 > \alpha\right)$$
$$\leq \exp\left(-\frac{2\alpha\theta}{\varepsilon}\right)\exp\left(-\frac{N}{2}\log(1-2\theta)\right).$$

This shows that we have, $\limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{P}(I(\sqrt{\varepsilon}\hat{B}^N) > \alpha) \leq -2\alpha\theta$, so, taking $\theta \uparrow \frac{1}{2}$ gives $\limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{P}(I(\sqrt{\varepsilon}\hat{B}^N) > \alpha) \leq -\alpha$.

For the second term in the maximum, make the bound:

$$\begin{aligned}
\mathbb{P}(\rho(\sqrt{\varepsilon}B, \sqrt{\varepsilon}\hat{B}^N) \geq \delta) &= \mathbb{P}\left(\sup_{t\in[0,T]} |B_t - \hat{B}_t^N| \geq \frac{\delta}{\sqrt{\varepsilon}}\right) \\
&= N\mathbb{P}\left(\sup_{t\in[0,\frac{T}{N})} |B_t - \hat{B}_t^N| \geq \frac{\delta}{\sqrt{\varepsilon}}\right) \\
&= N\mathbb{P}\left(\sup_{t\in[0,\frac{T}{N})} \left|B_t - \frac{t}{T/N}B_{\frac{T}{N}}\right| \geq \frac{\delta}{\sqrt{\varepsilon}}\right) \\
&= N\mathbb{P}\left(\sup_{t\in[0,\frac{T}{N})} |B_t| \geq \frac{\delta}{2\sqrt{\varepsilon}}\right) \\
&= 4N\mathbb{P}\left(B_{\frac{T}{N}} \geq \frac{\delta}{2\sqrt{\varepsilon}}\right) \leq \frac{4N^{\frac{3}{2}}}{\sqrt{2\pi T}}\exp\left(-\frac{N\delta^2}{4\varepsilon T}\right).
\end{aligned}$$

Now we see that the normalized logarithmic limit is

$$\limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{P}(\rho(\sqrt{\varepsilon}B, \sqrt{\varepsilon}\hat{B}^N) \geq \delta)$$
$$\leq \limsup_{\varepsilon\to 0}\varepsilon\log\left(\frac{4N^{\frac{3}{2}}}{\sqrt{2\pi T}}\right) - \frac{N\delta^2}{4T} \leq -\frac{N\delta^2}{4T}.$$

Combining these bounds, we have shown:

$$\limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{P}(\rho(\sqrt{\varepsilon}B, K_\alpha) \geq \delta) \leq \max\left\{-\alpha, -\frac{N\delta^2}{4T}\right\} \to -\alpha \tag{28}$$

where the limit sends $N \to \infty$.

Now we use this to prove the large deviations upper bound. For any closed set $F \subseteq C_0([0,T];\mathbb{R})$, write $I_F = \inf\{I(\phi) : \phi \in F\}$. If $I_F = 0$, then there is nothing to prove, so assume $I_F > 0$. For $\gamma \in (0, I_F)$, the set $K_{I_F-\gamma}$ is compact. Moreover, $K_{I_F-\gamma}$ is disjoint from F. Therefore, there exists some $\delta > 0$ such that $\rho(\phi, K_{I_F-\gamma}) \geq \delta$ holds for all $\phi \in F$. Putting this all together, we have

$$\limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{P}(\sqrt{\varepsilon}B \in F)$$
$$\leq \limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{P}(\rho(\sqrt{\varepsilon}B, K_{I_F-\gamma}) \geq \delta) \leq -I_F + \gamma$$

Now take $\gamma \downarrow 0$ to get

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(\sqrt{\varepsilon}B \in F) \leq -I_F \tag{29}$$

which is the desired upper bound.

Lower Bound. Take an arbitrary point $\phi \in C_0([0,T];\mathbb{R})$ and any $\delta > 0$. Note that by Girsanov's theorem we have:

$$\begin{aligned}
&\mathbb{P}(\sqrt{\varepsilon}B \in B_\delta(\phi)) \\
&= \mathbb{P}\left(B - \frac{1}{\sqrt{\varepsilon}}\phi \in B_{\frac{\delta}{\sqrt{\varepsilon}}}(0)\right) \\
&= \mathbb{E}\left[\exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^T \phi'(t)dB_t - \frac{1}{2\varepsilon}\int_0^T |\phi'(t)|^2 dt\right); B \in B_{\frac{\delta}{\sqrt{\varepsilon}}}(0)\right] \\
&= \exp\left(-\frac{1}{\varepsilon}I(\phi)\right)\mathbb{E}\left[\exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^T \phi'(t)dB_t\right); B \in B_{\frac{\delta}{\sqrt{\varepsilon}}}(0)\right].
\end{aligned}$$

Now note that we have $\mathbb{P}(B \in B_{\delta\varepsilon^{-1/2}}(0)) \to 1$ as $\varepsilon \to 0$, hence there exists some sufficiently small $\varepsilon > 0$ which gives $\mathbb{P}(B \in B_{\delta\varepsilon^{-1/2}}(0)) \geq \frac{3}{4}$. Likewise, we have by Chebeshev's inequality:

$$\begin{aligned}
&\mathbb{P}\left(\exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^T \phi'(t)dB_t\right) \geq \exp\left(-2\sqrt{\frac{2I(\phi)}{\varepsilon}}\right)\right) \\
&\qquad = 1 - \mathbb{P}\left(\int_0^T \phi'(t)dB_t \geq 2\sqrt{2I(\phi)}\right) \\
&\qquad \geq 1 - \frac{\mathbb{E}\left[\left(\int_0^T \phi'(t)dB_t\right)^2\right]}{8I(\phi)} = 1 - \frac{1}{4} = \frac{3}{4}
\end{aligned}$$

Therefore, the intersection of these two sets is nonempty and has probability at least $\frac{1}{2}$. Combining this with the above, this implies

$$\mathbb{P}(\sqrt{\varepsilon}B \in B_\delta(\phi)) \geq \frac{1}{2}\exp\left(-\frac{1}{\varepsilon}I(\phi)\right)\exp\left(-2\sqrt{\frac{2I(\phi)}{\varepsilon}}\right). \tag{30}$$

Finally, this gives

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(\sqrt{\varepsilon}B \in B_\delta(\phi)) \geq I(\phi), \tag{31}$$

which we have seen is sufficient for establishing the lower bound. □

The general methods we used to prove Crámer's theorem and Schilder's theorem are standard in the theory of large deviations: The upper bound is proved via concentration of measure, and the lower bound is proved by applying a suitable change of measure that makes a certain uncommon event into a common event. However, the actual details needed to carry out these steps are quite different.

Next we remark Schilder's theorem can be easily extended to a small class of diffusions. In the following result, the limit is taken as the scale of the noise goes to zero, so these results are extremely useful in practical engineering situations where one is interested in fine properties of dynamical systems in the small noise regime.

Theorem 4.2 (Freidlin-Wentzell)**.** Let $\mu : \mathbb{R} \to \mathbb{R}$ be a Lipschitz function, and, for $\varepsilon > 0$, write $\{X_t^\varepsilon\}_{t\in[0,T]}$ for the solution to the SDE $dX_t = \mu(X_t)dt + \sqrt{\varepsilon}dB_t$ with $X_0 = 0$. Then, the laws of $\{X_t^\varepsilon\}_{t\in[0,T]}$ satisfy a large deviations principle in $C_0([0,T];\mathbb{R})$ with good rate function

$$I_\mu(\phi) = \begin{cases} \frac{1}{2}\int_0^T |\phi'(t) - \mu(\phi(t))|^2 dt & \text{if } \phi \in H_1 \\ \infty & \text{else} \end{cases}. \tag{32}$$

Proof. Let $F : C_0([0,T];\mathbb{R}) \to C_0([0,T];\mathbb{R})$ be the function such that $F(\phi) = x$ whenever x is a solution to the integral equation

$$x(t) = \int_0^t \mu(x(s))ds + \phi(t) \tag{33}$$

for all $t \in [0,T]$.

To see that F is continuous, note that for any $\phi_1, \phi_2 \in C_0([0,T];\mathbb{R})$ we have that $x_1 = F(\phi_1)$ and $x_2 = F(\phi_2)$ satisfy

$$\begin{aligned} |x_1(t) - x_2(t)| &= \left| \int_0^t (\mu(x_1(s)) - \mu(x_2(s)))\, ds + (\phi_1(t) - \phi_2(t)) \right| \\ &\leq \int_0^t |\mu(x_1(s)) - \mu(x_2(s))|\, ds + |\phi_1(t) - \phi_2(t)| \\ &\leq K \int_0^t |x_1(s) - x_2(s)|\, ds + |\phi_1(t) - \phi_2(t)| \end{aligned}$$

for all $t \in [0,T]$, and where K is the Lipschitz constant of μ. Taking suprema gives

$$\sup_{s\in[0,t]} |x_1(s) - x_2(s)| \leq K \int_0^t \sup_{u\in[0,s]} |x_1(u) - x_2(u)|\, ds + \sup_{s\in[0,t]} |\phi_1(s) - \phi_2(s)|\,. \tag{34}$$

Now apply Gronwall's inequality to get

$$\sup_{t\in[0,T]} |x_1(t) - x_2(t)| \leq \sup_{t\in[0,T]} |\phi_1(t) - \phi_2(t)|\, e^{KT}, \tag{35}$$

which implies that F is continuous (in fact, Lipschitz continuous with Lipschitz constant 1).

Now we note that the law of $\{X_t^\varepsilon\}_{t\in[0,T]}$ is just the pushforward of the law of $\{\sqrt{\varepsilon}B_t\}_{t\in[0,T]}$ by F. Hence, the contraction principle implies that the laws of $\{X_t^\varepsilon\}_{t\in[0,T]}$ satisfy a large deviations principle with good rate function given by $J(x) = \inf\{I(\phi) : x = F(\phi)\}$. To finish the proof, we only need to show $J = I_\mu$. To do this, first note that F is a bijection of $C_0([0,T];\mathbb{R})$: It is surjective since for any $x \in C_0([0,T];\mathbb{R})$ we can set $\phi \in C_0([0,T];\mathbb{R})$ via $\phi(t) = x(t) - \int_0^T \mu(x(s))ds$

and we have $F(\phi) = x$, and it is injective since $F(\phi_1) = F(\phi_2) = x$ imply $\phi_1(t) = x(t) - \int_0^T \mu(x(s))ds = \phi_2(t)$. Moreover, is can be shown that, when $F(\phi) = x$, we have $x \in H_1$ iff $\phi \in H_1$. Putting this all together gives that $x = F(\phi)$ gives

$$J(x) = \frac{1}{2}\int_0^T |\phi'(t)|^2 ds = \frac{1}{2}\int_0^T |x'(t) - \mu(x(t))|^2 ds \tag{36}$$

when $x \in H_1$ and $J(x) = \infty$ otherwise. □

The result about can also be generalized to the laws of solutions of an SDE whose drift term depends on X_t. However, the same method of proof does not apply since there is no continuous mapping analogous to F. (Essentially, this is because the Ito integral is not defined pathwise.) We omit the proof of this result, but, for the sake of completeness, we state it now:

Theorem 4.3 (Freidlin-Wentzell)**.** Let $\mu, \sigma : \mathbb{R} \to \mathbb{R}$ be Lipschitz functions with $\sigma > 0$ everywhere, and, for $\varepsilon > 0$, write $\{X_t^\varepsilon\}_{t\in[0,T]}$ for the solution to the SDE $dX_t = \mu(X_t)dt + \sqrt{\varepsilon}\sigma(X_t)dB_t$ with $X_0 = 0$. Then, the laws of $\{X_t^\varepsilon\}_{t\in[0,T]}$ satisfy a large deviations principle in $C_0([0,T];\mathbb{R})$ with good rate function

$$I_{\mu,\sigma}(\phi) = \frac{1}{2}\int_0^T \frac{|\phi'(t) - \mu(\phi(t))|^2}{\sigma^2(\phi(t))} dt. \tag{37}$$

The theorems of this section all have counterparts in $\mathbb{R}^d$ with some modifications. In general, the statements of such results are easy to intuit, but the proofs are sometimes much more difficult to establish.