1. Introduction

The goal of this lecture note is to give a very brief introduction to the theory of large deviations. In probability theory, a recurrent theme is the convergence of random variables, and the law of large numbers (LLN) is among the fundamental results concerning their "typical" behaviors. But beyond the scope, mathematicians yearn for more a refined picture of the "atypical configurations" or "rare events" that deviates from the expected outcome either by a small or a large amount. In this pursuit, the former is addressed in the central limit theorem (CLT) while the latter give rise to the large deviation principle (LDP). To demonstrate these, we will like to begin with the following example.

Example 1.1 (Bernoulli trial)**.** A Bernoulli trial is a series of identical and independent experiments with exactly two possible outcomes, say 0 and 1, which may be modeled as a sequence of i.i.d. random variables $(X_i)_{i=1}^{\infty}$ with the Bernoulli distribution:

$$\mathbb{P}(X_i = 1) = p \quad \text{and} \quad \mathbb{P}(X_i = 0) = 1 - p \quad (p \in (0,1)).$$

The core questions regarding the trials revolve around the number of $1'$s in an n -trial, which is mathematically phrased as a random variable $S_n = \sum_{i=1}^{n} X_i$. It is not hard to determine the distribution of S_n:

$$\mathbb{P}(S_n = \lfloor nx \rfloor) = \binom{n}{\lfloor nx \rfloor} p^{\lfloor ns \rfloor}(1-p)^{n-\lfloor nx \rfloor} \quad \text{if } s \in [0,1],$$

for which, here and throughout study, we intend to examine the logarithm of the probability the associated events. Specifically, by Stirling's approximation ($\log n! = n \log n - n + O(\log n)$),

$$\begin{aligned} &n^{-1} \log \mathbb{P}(S_n = \lfloor nx \rfloor) \\ = &\frac{\lfloor nx \rfloor}{n} \log\left(\frac{p}{\frac{\lfloor nx \rfloor}{n}}\right) + \left(1 - \frac{\lfloor nx \rfloor}{n}\right) \log\left(\frac{1-p}{1-\frac{\lfloor nx \rfloor}{n}}\right) + O\left(\frac{\log n}{n}\right). \end{aligned}$$

Asymptotically, given that $x \mapsto -x \log x$ is uniformly continuous on $[0,1]$,

$$\lim_{n \to \infty} n^{-1} \log \mathbb{P}(S_n = \lfloor ns \rfloor) = -I(s),$$

where

$$I(s) = x \log\left(\frac{x}{p}\right) + (1-x) \log\left(\frac{1-x}{1-p}\right)$$

is plotted in Figure 1. In the language of large deviation theory, the sequence $n^{-1} S_n$ is said to satisfy the large deviation principle (which will be defined rigorously later) with rate function I.

Some comments are in order.

- $I(x)$ is a convex function with

$$I'(x) = \log\left(\frac{x}{1-x}\right) - \log\left(\frac{p}{1-p}\right) \quad \text{and} \quad I''(x) = \frac{1}{x} + \frac{1}{1-x}.$$

  In particular, its minimum is attained by a unique point $x = p = \mathbb{E}(X_1)$. This implies the law of large numbers by the Borel-Cantelli lemma.

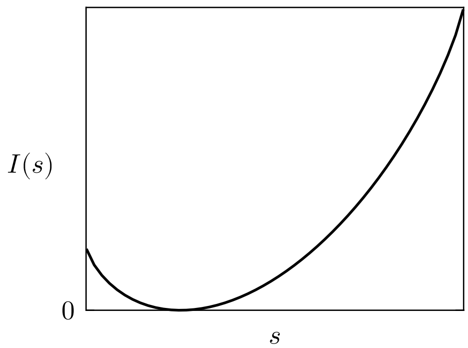

FIGURE 1. rate function $I(s)$

- Observe that the second-order Taylor expansion

$$I(x) = \frac{(x-p)^2}{2p(1-p)} + O(|x-p|^2).$$

When considering a small deviation $x = p + y/\sqrt{n}$, this is "roughly" consistent with the central limit theorem:

$$\mathbb{P}\left(\frac{\sqrt{n}(n^{-1}S_n - p)}{\sqrt{p(1-p)}} = \frac{y}{\sqrt{p(1-p)}} \right) \approx \frac{1}{\sqrt{2\pi p(1-p)}} e^{\frac{y^2}{2p(1-p)}}.$$

1.1. **The large deviation principle.**

Throughout the note, we denote by $\{\mu_\varepsilon\}$ be the collection of probability measures on a common measurable space $(\mathcal{X}, \mathcal{B})$ whose limiting behavior as $\varepsilon \to 0$ is of interest. Within the scope, μ_ε is always complete and $\mathcal{X}$ is always a Hausdorff topological space associated with the completed Borel σ-algebra $\mathcal{B}_{\mathcal{X}}$.

Definition 1.2 (rate function)**.** Let $I : \mathcal{X} \to [0, \infty]$ be a function with *sublevel sets* $\Psi_I(\alpha) := \{x : I(x) \leq \alpha\}$ and *effective domain* $\mathcal{D}_I := \{x : I(x) < \infty\}$.

- I is called a *rate function* if it is lower semicontinuous.
- A rate function I is said to be *good* if $\Psi_I(\alpha)$ is compact for all $\alpha < \infty$.

Definition 1.3 (large deviation principle)**.** $\{\mu_\varepsilon\}_\varepsilon$ is said to satisfy the *large deviation principle with rate* $I : \mathcal{X} \to [0, \infty]$ if for every $\Gamma \in \mathcal{B}$,

$$-\inf_{x \in \mathring{\Gamma}} I(x) \leq \liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(\Gamma) \leq \limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(\Gamma) \leq -\inf_{x \in \overline{\Gamma}} I(x). \tag{1.1}$$

Remark 1.4. *Suppose $\mathcal{B}_{\mathcal{X}} \subseteq \mathcal{B}$. Then, $\{\mu_\varepsilon\}_\varepsilon$ satisfies the large deviation principle with rate I if and only if the following hold.*

- *(upper bound) For any closed set $F \subset \mathcal{X}$,*
$$\limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(F) \leq -\inf_{x \in F} I(x). \tag{1.2}$$
- *(lower bound) For any open set $G \subset \mathcal{X}$,*
$$\liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(G) \geq -\inf_{x \in G} I(x). \tag{1.3}$$

Definition 1.5. Suppose $\mathcal{B}$ contains all compact subsets of $\mathcal{X}$. Then, $\{\mu_\varepsilon\}$ is said to be *exponentially tight* if for every $\alpha < \infty$, there is a compact set K such that

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K^c) < -\alpha.$$

Remark 1.6. *Suppose $\mathcal{B}_\mathcal{X} \subset \mathcal{B}$. If $\{\mu_\varepsilon\}$ is exponentially tight, then the following hold.*

- *(upper bound) If (1.2) holds for every compact sets, so does it for every closed set*
- *(lower bound) If (1.3) holds for every open set, then I is good.*

Notably, any $\{\mu_\varepsilon\}$ admitting a rate function satisfying the relaxed conditions is said to satisfy the weak LDP.

The *large deviation principle* (LDP) characterizes the limiting behavior of a sequence of measures.