Compare this with Girsanov's theorem (see Remark 2.9 and equation (2.17)). In fact, due to the continuity of Brownian paths, using approximation arguments, note that these finite-dimensional calculations already imply that

$$\mathbb{E}\left[\mathbb{I}_{\left\{\max_{t\in[0,T]}(B_t+\mu t)<x\right\}}\right] = \mathbb{E}\left[\mathbb{I}_{\left\{\max\limits_{t\in[0,T]} B_t<x\right\}} \exp\left(\mu B_T - \frac{1}{2}\mu^2 T\right)\right].$$

2.4.2. *Hitting time of Brownian motion with drift.* Let $\Omega \doteq \mathcal{C}[0,\infty)$, $\mathcal{F}_t^0 \doteq \sigma(\omega_s, s\le t)$, let $\mathbb{P}$ be Wiener measure and let $\tilde{\mathbb{P}}$ be the distribution induced by Brownian motion with drift. For $a>0$, let

$$\tau_a \doteq \inf\{t>0 : w_t = a\}.$$

Under $\tilde{\mathbb{P}}$, τ_a is the hitting time of Brownian motion with drift μ. By Girsanov's theorem, we have for $t\in[0,\infty)$,

$$\tilde{\mathbb{P}}\left(\tau_a\le t\right) = \tilde{\mathbb{E}}\left[\mathbb{I}_{\{\tau_a\le t\}}\right] = \mathbb{E}\left[\mathbb{I}_{\{\tau_a\le t\}}Z_t\right].$$

Conditioning on the σ-field $\mathcal{F}^0_{t\wedge\tau_a}$, using the fact that $\{\tau_a\le t\}$ is $\mathcal{F}^0_{t\wedge\tau_a}$-measurable and Z is an $\mathcal{F}^0_t$-martingale, we find that

$$\begin{aligned}\mathbb{E}\left[\mathbb{I}_{\{\tau_a\le t\}}Z_t\right] = \mathbb{E}\left[\mathbb{I}_{\{\tau_a\le t\}}Z_{t\wedge\tau_a}\right] &= \mathbb{E}\left[\mathbb{I}_{\{\tau_a\le t\}}\exp\left(\mu a - \frac{1}{2}\mu^2\tau_a\right)\right]\\ &= \int_0^t \exp\left(\mu a - \frac{1}{2}\mu^2 s\right)\frac{a}{s^{3/2}}\varphi\left(\frac{a}{\sqrt{s}}\right)ds,\end{aligned}$$

where φ is the standard normal density and we have used the known explicit expression for the density of the first hitting time for Brownian motion (see, for example, relation (8.5) on page 96 of [1]). Differentiating the last integral and completing the square in the exponent, we see that the density f_{τ_a} of τ_a is given by

$$f_{\tau_a}(t) = \frac{a}{t^{3/2}}\varphi\left(\frac{a-\mu t}{\sqrt{t}}\right).$$

Another very important application of Girsanov's theorem is in establishing existence and uniqueness of so-called weak solutions to stochastic differential equations. This will be covered in the next section.

2.4.3. *Large Deviations.* Large deviations is an asymptotic theory that characterizes the probabilities of rare events. Roughly speaking, the theory of large deviations captures the exponential rate of decay of the probabilities of a parametrized sequence of events that converge exponentially fast to zero (say, as the parameter tends to infinity). In large deviations theory, the rigorous statement that the probabilities of this parametrized sequence decays to zero exponentially is referred to as the *large deviations principle* (or LDP, for short), and the rate of decay is usually expressed as an optimization or variational problem involving the so-called *rate function.* In turn, the solution of the optimization problem provides insight into the most likely way in which rare events occur. We shall demonstrate this general idea via a simple concrete example of the sum of i.i.d. random variables in $\mathbb{R}$. In what follows, $M_1(\mathbb{R}^d)$ refers to the space of probability measures on $\mathbb{R}^d$.

Consider a sequence of i.i.d. d-dimensional random vectors $X_1, X_2, \ldots, X_n, \ldots$ with $X_1 \sim \mu \in M_1(\mathbb{R}^d)$. Consider the empirical mean, defined by

$$\overline{S}_n \doteq \frac{1}{n} \sum_{i=1}^{d} X_i$$

and let μ_n denote the distribution on $\mathbb{R}^d$ induced by $\overline{S}_n : \overline{S}_n \sim \mu_n$. If $\overline{x} = \mathbb{E}[X_1] < \infty$ and $\mathbb{E}\left[|X_1 - \overline{x}|^2\right] < \infty$ then a simple application of Chebyshev's inequality and the observation that

$$\mathbb{E}\left[|\overline{S}_n - \overline{x}|^2\right] = \frac{1}{n^2} \sum_{j=1}^{n} \mathbb{E}\left[|X_j - \overline{x}|^2\right] = \frac{1}{n} \mathbb{E}\left[|X_1 - \overline{x}|^2\right] \to 0$$

as $n \to \infty$ shows that the following weak law of large numbers holds:

$$\overline{S}_n \xrightarrow{p} \overline{x}.$$

(Recall that $\xrightarrow{p}$ implies convergence in probability.) Therefore, given any closed set F such that $\overline{x} \notin F$, it follows that

$$\mu_n(F) \to 0$$

as $n \to \infty$. A natural question to ask in this context is what is the rate of convergence of this sequence of probabilities to zero ? We will in fact establish a so-called large deviation principle for the sequence of measures $\{\mu_n\}$. Let $\mathcal{B} = \mathcal{B}(\mathbb{R}^d)$ denote the Borel σ-field on $\mathbb{R}^d$.

Definition 2.17. A sequence of measures $\{\mu_n\}$ on $(\mathbb{R}^d, \mathcal{B})$ is said to solve a *large deviation principle* with *rate function* I if

(1) $I : \mathbb{R}^d \mapsto [0, \infty)$ is a lower semicontinuous function, i.e., the level set $\Psi_I(\alpha) \doteq \{x : I(x) \leq \alpha\}$ is closed for every $\alpha \in [0, \infty)$;

(2) For all closed sets $F \subset \mathcal{B}$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mu_n(F) \leq - \inf_{x \in F} I(x); \tag{2.24}$$

(3) For all open sets $G \subset \mathcal{B}$,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mu_n(G) \geq - \inf_{x \in G} I(x). \tag{2.25}$$

Moreover, if for every $\alpha \in [0, \infty)$, $\Psi_I(\alpha)$ is a compact set, then I is said to be a *good rate function*.

Remark 2.18. On metric spaces, and hence in particular on $\mathbb{R}^d$, a function f is lower semicontinuous if and only if for every sequence $\{x_n\}$ such that $x_n \to x$ as $n \to \infty$,

$$\liminf_{n \to \infty} f(x_n) \geq f(x).$$

Before we state the main theorem, we need to introduce the following definitions. Let Λ be the *log moment generating function* associated with μ:

$$\Lambda(\lambda) \doteq \log M(\lambda) \doteq \log \mathbb{E}\left[e^{\langle \lambda, X_1 \rangle}\right], \qquad \lambda \in \mathbb{R}^d.$$

Note that $\Lambda(0) = 0$ and $\Lambda(\lambda) > -\infty$ for all λ.

Definition 2.19. *The Fenchel-Legendre transform of Λ is defined to be*

$$\Lambda^*(x) \doteq \sup_{\lambda \in \mathbb{R}^d} \left[\langle \lambda, x \rangle - \Lambda(\lambda)\right] \qquad x \in \mathbb{R}^d.$$

Define

$$\mathcal{D}_\Lambda \doteq \{\lambda \in \mathbb{R}^d : \Lambda(\lambda) < \infty\}.$$

and, likewise,

$$D_{\Lambda^*} \doteq \left\{x \in \mathbb{R}^d : \Lambda^*(x) < \infty\right\}.$$

For simplicity, we will make the following assumption on μ.

Assumption 1. $0 \in \mathcal{D}_\Lambda^\circ$, *i.e., zero lies in the interior of $\mathcal{D}_\Lambda$.*

The proof of the theorem makes use of several properties of Λ and Λ^* that are summarized in Proposition A.2 of the Appendix. For simplicity, we will restrict our focus to the case when $d = 1$ and the following assumption is satisfied by Λ.

Theorem 2.20. *Suppose $d = 1$. If Assumption 1 holds, then $\{\mu_n\}$ satisfies the LDP in $(\mathbb{R}, \mathcal{B})$ with good rate function $\Lambda^*(\cdot)$.*

Proof. We first note that the fact that $\Lambda^*(\cdot)$ is a good rate function follows from Proposition A.2(b). Therefore, in order to establish the LDP, it suffices to establish the large deviations upper and lower bounds stated in (2.24) and (2.25), respectively, with I replaced by Λ^*. Note that by Proposition A.2(f), Assumption 1 implies that $\overline{x} \in (-\infty, \infty)$.

Proof of the Large Deviation Upper Bound.

Step 1: Proof for closed half-intervals $F = [x, \infty)$, $F = (-\infty, x]$:

For all x and every $\lambda \geq 0$, we have

$$\begin{aligned}\mu_n([x,\infty)) = \mathbb{E}\left[1_{\{\overline{S}_n - x \geq 0\}}\right] \leq \mathbb{E}\left[e^{n\lambda(\overline{S}_n - x)}\right] &= e^{-n\lambda x} \prod_{i=1}^n \mathbb{E}\left[e^{\lambda x_i}\right] \\ &= e^{-n(\lambda x - \Lambda(\lambda))}.\end{aligned}$$

Since $\overline{x} < \infty$, taking the infimum over $\lambda > 0$ in the last display and using (A.62), we conclude that for every $x > \overline{x}$,

$$\mu_n([x,\infty)) \leq e^{-n\Lambda^*(x)}. \tag{2.26}$$

In a similar fashion, since $\overline{x} > -\infty$ for $x < \overline{x}$, we can take the infimum over all $\lambda < 0$ and use (A.63) to obtain

$$\mu_n((-\infty, x]) \leq e^{-n\Lambda^*(x)}. \tag{2.27}$$

Step 2: Proof for general closed sets F.

Let F be a non-empty closed set. Note that (2.24) holds trivially when $I_F \doteq \inf_{x\in F} \Lambda^*(x) = 1$. Therefore we can assume without loss of generality that $I_F > 0$. Now, since $\overline{x} < \infty$, by Proposition A.2 $\Lambda^*(\overline{x}) = 0$ and since $I_F > 0$, $\overline{x} \in F^c$. Let (x_-, x_+) be the union of all the open intervals $(a, b) \in F^c$ that contain $\overline{x}$. Note that $x_- < x_+$ and either $x_- < \infty$ or $x_+ < \infty$ since F is non-empty. If x_- is finite, then $x_- \in F$, and consequently $\Lambda^*(x_-) \geq I_F$. Likewise, $\Lambda^*(x_+) \geq I_F$ whenever x_+ is finite. We thus have (2.26) for $x = x_+$ and (2.27) for $x = x_-$. Therefore, the union bound ensures that

$$\mu_n(F) \leq \mu_n((-\infty, x_-]) + \mu_n([x_+, \infty)) \leq 2e^{-nI_F}$$

and the upper bound follows by taking logarithms and sending $n \to \infty$ on both sides.

Proof of the Large Deviation Lower Bound.

Step 1: Proof for intervals $\mathcal{O} = (x - \delta, x + \delta)$

We start with the claim that to complete Step 1 it suffices to show that for every $\delta > 0$ and every marginal law $\mu \in M_1(\mathbb{R})$,

$$\liminf_{n\to\infty} \frac{1}{n}\mu_n((-\delta, \delta)) \geq \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = -\Lambda^*(0). \tag{2.28}$$

Then the transformation $Y = X - x$ yields

$$\Lambda_Y(\lambda) = \Lambda(\lambda) - \lambda x,$$

and hence also yields

$$\Lambda_Y^*(\cdot) = \Lambda^*(\cdot + x).$$

Therefore (2.28) implies that for every x and every $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n((x - \delta, x + \delta)) \geq -\Lambda^*(x). \tag{2.29}$$

Step 1a: Proof of (2.28) assuming $\mu((-\infty, 0)) > 0$ and $\mu((0, \infty)) > 0$ and that μ is supported on a bounded subset of $\mathbb{R}$.

By the former assumption, one can conclude that $\Lambda(\lambda) \to \infty$ as $|\lambda| \to \infty$, and it is easy to see that by the latter assumption, $\Lambda(\cdot)$ is finite everywhere. Accordingly, $\Lambda(\cdot)$ is a continuous differentiable function by (e) of Proposition A.2. Hence, there exists a finite η such that

$$\Lambda(\eta) = \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) \qquad \text{and} \qquad \frac{d\Lambda}{d\lambda}(\eta) = 0.$$

Define a new probability measure $\tilde{\mu}$ in terms of μ via

$$\frac{d\tilde{\mu}}{d\mu}(x) = e^{\eta x - \Lambda(\eta)} = \frac{e^{\eta x}}{M(\eta)}$$

and observe that $\tilde{\mu}$ is a probability measure because

$$\int_{\mathbb{R}} d\tilde{\mu} = \frac{1}{M(\eta)} \int_{\mathbb{R}} e^{\eta x} dx = 1.$$

Let $\mathbb{E}_{\tilde{\mu}}$ denote expectation with respect to each X_i having marginal $\tilde{\mu}$. By (A.65) and the choice of η, it follows that

$$\mathbb{E}_{\tilde{\mu}}[X_1] = \frac{1}{M(\eta)} \int_{\mathbb{R}} x e^{\eta x} dx = \frac{d\Lambda}{d\lambda}(\eta) = 0.$$

Hence, by the law of large numbers, we have

$$\lim_{n\to\infty} \tilde{\mu}_n((-\varepsilon, \varepsilon)) = 1. \tag{2.30}$$

In large deviations theory, $\tilde{\mu}$ is sometimes referred to as the *tilted distribution*, which is a new distribution absolutely continuous with respect to the original distribution that has a different mean. The tilted distribution is chosen so that the "large deviation" event becomes a "typical" event. Then, as made precise below, the probability of the rare event is governed by the likelihood ratio (or Radon-Nikodym derivative) between the tilted distribution and the original distribution. More concretely, let $\tilde{\mu}_n$ be the distribution of $\overline{S}_n$ on $\mathbb{R}$, when X_i are i.i.d. random variables, each with law $\tilde{\mu}$. Then, for every $\varepsilon > 0$ and $n \in \mathbb{N}$,

$$\begin{aligned}
\mu_n((-\varepsilon, \varepsilon)) &= \int_{\{|\sum_{i=1}^n x_i| < n\varepsilon\}} \mu(dx_n) \\
&\geq e^{-n\varepsilon|\eta|} \int_{\{|\sum_{i=1}^n x_i| < n\varepsilon\}} \exp\left(\eta \sum_{i=1}^n x_i\right) \mu(dx_1)\cdots\mu(dx_n) \\
&= e^{-n\varepsilon|\eta|} e^{n\Lambda(\eta)} \tilde{\mu}_n((-\varepsilon, \varepsilon)).
\end{aligned}$$

Combining the last assertion with (2.30) it follows that for every $0 < \varepsilon < \delta$,

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \mu_n((-\delta, \delta)) &\geq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n((-\varepsilon, \varepsilon)) \\
&\geq \Lambda(\eta) - \varepsilon|\eta|,
\end{aligned}$$

and (2.28) follows by considering the limit as $\varepsilon \to 0$.

Step 1b. Proof of (2.28) when μ is of unbounded support, while both $\mu((-\infty, 0)) > 0$ and $\mu((0, \infty)) > 0$.

Fix M large enough such that $\mu([-M, 0)) > 0$ as well as $\mu((0, M]) > 0$, and let

$$\Lambda_M(\lambda) \doteq \log \int_{-M}^{M} e^{\lambda x} \, d\mu.$$

Let ν denote the law of X_1 conditioned on $\{|X_1| \leq M\}$, and let ν_n be the law of $\overline{S}_n$ conditioned on $\{|X_i| \leq M, i = 1, \ldots, n\}$. Then, for all n and every $\delta > 0$,

$$\mu_n((-\delta, \delta)) \geq \nu_n((-\delta, \delta)) \mu([-M, M])^n.$$

By the preceding proof, (2.28) holds with μ_n and Λ, respectively, replaced by ν_n and the cumulant generating function of ν, which equals $\Lambda_M(\cdot) -$

$\log \mu([-M, M])$. Therefore, we conclude that

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \mu_n((-\delta, \delta)) \\ &\geq \log \mu([-M, M]) + \liminf_{n\to\infty} \frac{1}{n} \log \nu_n((-\delta, \delta)) \\ &\geq \inf_{\lambda\in\mathbb{R}} \Lambda_M(\lambda).\end{aligned}$$

Set $I_M \doteq -\inf_{\lambda\in\mathbb{R}} \Lambda_M(\lambda)$ and $I^* \doteq \limsup_{M\to\infty} I_M$. Then it follows that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n((-\delta, \delta)) \geq -I^*. \tag{2.31}$$

Since $\Lambda_M(\cdot)$ is non-decreasing in M, so is $-I_M$. Moreover, $-I_M \leq \Lambda_M(0) \leq \Lambda(0) = 0$, and hence $-I^* \leq 0$. Now, since $-I_M$ is finite for all M large enough, $-I^* > -\infty$. Therefore, the level sets $\{\lambda : \Lambda_M(\lambda) \leq -I^*\}$ are non-empty, compact sets that are nested with respect to M, and hence there exists at least one point, denoted by λ_0, in their intersection. By the monotone convergence theorem, $\Lambda(\lambda_0) = \lim_{M\to\infty} \Lambda_M(\lambda_0) \leq -I^*$, and consequently the bound (2.31) yields (2.28).

Step 1c: Proof of (2.28) for arbitrary μ satisfying Assumption 1.

Observe that if either $\mu((-\infty, 0)) = 0$ or $\mu((0, \infty)) = 0$, then $\Lambda(\cdot)$ is a monotone function with $\inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = \log \mu(\{0\})$. Hence, in this case, (2.28) follows from

$$\mu_n((-\delta, \delta)) \geq \mu_n(\{0\}) = \mu(\{0\})^n.$$

Step 2: The large deviation lower bound for general sets. For any open set G, any $x \in G$ and all $\delta > 0$ small enough, $(x - \delta, x + \delta) \subset G$. So the large deviation lower bound follows from (2.29). □

Remark 2.21. Theorem 2.20 possesses a multivariate counterpart. In particular, it can be shown that when $\mathcal{D}_\Lambda = \mathbb{R}^d$, then $\{\mu_n\}$ satisfies the LDP with good convex rate function $\Lambda^*(\cdot)$. The proof is similar in spirit to the one-dimensional case. The difference is that the monotonicity properties stated in Proposition A.2(c) no longer hold. Thus the method (in Step 2 of the large deviation upper bound) of enclosing any closed set F in two half spaces is no longer valid. Instead, upper bounds on the probabilities that $\{\mu_n\}$ assign to balls are deduced by a Chebyshev-type inequality. Since compact sets can be covered by a finite number of balls, a union bound immediately yields the result for compact sets. The result is then extended to all closed sets using an additional approximation argument. On the other hand, the lower bound for balls centered around $y = \nabla\Lambda(\eta)$ follows essentially the same argument as in the one-dimensional case. An additional regularization argument is required to extend this to $y \in \mathcal{D}_{\Lambda^*}$ such that $y \notin \{\nabla\Lambda(\lambda) : \lambda \in \mathbb{R}^d\}$. We omit the details of this proof.