Chapter 5: Large Deviations

Before formally introducing the large deviation principle, we first explain what the large deviation principle is about and look at a simple example carefully.

Suppose that $X_1, X_2, \ldots$ is a sequence of independent and identically distributed random variables with a common mean $EX_1 = \mu$ and that $S_n = X_1 + \cdots + X_n$. The weak law of large numbers tells us that, for any $\epsilon > 0$, as $n \to \infty$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \to 0.$$ 

But the weak law gives no information as the rate that the sequence above goes to zero. The large deviation principle tells us that in general the sequence above decays to zero exponentially fast and provides with rate of exponential decay. Let’s look at the following example.

**Example 1.** Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with $P(X_1 = 1) = P(X_1 = -1) = 1/2$. Then $EX_1 = 0$ and

$$Ee^\lambda X_1 = \frac{e^\lambda + e^{-\lambda}}{2}, \quad \lambda \in R.$$ 

Put $\Lambda(\lambda) = \ln Ee^\lambda X_1$ and define

$$\Lambda^*(x) = \sup_{\lambda \in R} \{\lambda x - \Lambda(\lambda)\}, \quad x \in R.$$ 

Elementary calculus tells us that

$$\Lambda^*(x) = \begin{cases} \frac{(1 + x) \ln(1 + x) + (1 - x) \ln(1 - x)}{2}, & x \in (-1, 1) \\ +\infty, & |x| \geq 1. \end{cases}$$

and that

$$\Lambda^*(x) = \sup_{\lambda > 0} \{\lambda x - \Lambda(\lambda)\}, \quad x > 0$$

and that

$$\Lambda^*(x) = \sup_{\lambda < 0} \{\lambda x - \Lambda(\lambda)\}, \quad x < 0.$$ 

For any $\epsilon > 0$ and $\lambda > 0$,

$$P\left(\frac{S_n}{n} \geq \epsilon\right) = E(1_{\left\{\frac{S_n}{n} - \epsilon \geq 0\right\}}) \leq E(e^{n\lambda(\frac{S_n}{n} - \epsilon)})$$

$$= e^{-n\lambda \epsilon} \prod_{j=1}^{n} E(e^{\lambda X_j}) = e^{-n(\lambda \epsilon - \Lambda(\lambda))}.$$
Thus
\[ P\left(\frac{S_n}{n} \geq \varepsilon\right) \leq e^{-n\Lambda^*(\varepsilon)}. \]
Similarly we can get that
\[ P\left(\frac{S_n}{n} \leq -\varepsilon\right) \leq e^{-n\Lambda^*(-\varepsilon)} = e^{-n\Lambda^*(\varepsilon)}. \]
Therefore we have
\[ P\left(|\frac{S_n}{n}| \geq \varepsilon\right) \leq 2e^{-n\Lambda^*(\varepsilon)}. \]
Therefore we have shown that as \( n \to \infty \),
\[ P\left(|\frac{S_n}{n}| \geq \varepsilon\right) \]
tends to zero exponentially fast and the exponential decay rate is at least \( \Lambda^*(\varepsilon) \). In fact one can show that the exponential decay rate is indeed \( \Lambda^*(\varepsilon) \).

Now we are going to look at the general case. We first introduce some definitions first.

Recall that a function \( f : R^d \to [-\infty, \infty] \) is a lower semicontinuous function if
\[ f(x) \leq \liminf_{y \to x} f(y), \quad \forall x \in R^d. \]
\( f \) is lower semicontinuous if and only if for any real number \( c \), \( \{x : f(x) > c\} \) is an open subset of \( R^d \). From this one can easily show that if \( \{f_a(x) : a \in A\} \) is a family of continuous functions on \( R^d \), then the function
\[ f(x) = \sup_{a \in A} f_a(x), \quad x \in R^d \]
is a lower semicontinuous function on \( R^d \).

**Definition 1** A rate function \( I \) is a lower semicontinuous function \( I : R^d \to [0, \infty] \). A good rate function is a rate function \( I \) for which all level sets \( \Psi_I(\alpha) = \{x : I(x) \leq \alpha\}, \alpha \geq 0 \), are compact subsets of \( R^d \).

**Definition 2** Suppose that \( \{\nu_n\} \) is a sequence of probability measures on \( R^d \). We say that \( \{\nu_n\} \) satisfies the large deviation principle with a rate function \( I \) if

(a) for any closed subset \( F \) of \( S \),
\[ \limsup_{n \to \infty} \frac{1}{n} \log \nu_n(F) \leq -\inf_{x \in F} I(x); \]

(b) for any open subset \( G \) of \( S \),
\[ \liminf_{n \to \infty} \frac{1}{n} \log \nu_n(G) \geq -\inf_{x \in G} I(x). \]
Suppose that \( \{\nu_n\} \) satisfies the large deviation principle with a rate function \( I \). If \( \Gamma \in \mathcal{R} \) is a continuity set of \( I \), i.e.,

\[
\inf_{x \in \Gamma} I(x) = \inf_{x \in \Gamma} I(x) := I_{\Gamma},
\]

then

\[
\lim_{n \to \infty} \frac{1}{n} \log \nu_n(\Gamma) = -I_{\Gamma}
\]

which can be written heuristically as

\[
\nu_n(\Gamma) \asymp \exp(-n I_{\Gamma}).
\]

Now suppose that \( X_1, X_2, \ldots \) are iid random variables. Let \( \mu = E X_1 \) and let \( \nu \) be the distribution of \( X_1 \). Let \( \nu_n \) be the distribution of \( \frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n} \). The logarithmic moment generating function associated with \( \nu \) is

\[
\Lambda(\lambda) := \log M(\lambda) := \log E(e^{\lambda X_1}), \quad \lambda \in \mathbb{R}.
\]

Obviously, \( \Lambda(0) = 0 \), \( \Lambda(\lambda) > -\infty \) for any \( \lambda \). It is possible to have \( \Lambda(\lambda) = \infty \).

**Definition 3** The Legendre-Fenchel transform of \( \Lambda(\lambda) \) is

\[
\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\}.
\]

From the definition one can easily see that \( \Lambda^* \) is a nonnegative lower semicontinuous function on \( \mathbb{R} \), so it is a rate function.

**Theorem 1** (Cramer’s Theorem) The sequence of probability measures \( \{\nu_n\} \) satisfies the large deviation principle with the convex rate function \( I(\cdot) = \Lambda^*(\cdot) \), namely

(a) for any closed subset \( F \) of \( \mathbb{R} \),

\[
\limsup_{n \to \infty} \frac{1}{n} \log \nu_n(F) \leq -\inf_{x \in F} I(x);
\] (1)

(b) for any open subset \( G \) of \( \mathbb{R} \),

\[
\liminf_{n \to \infty} \frac{1}{n} \log \nu_n(G) \geq -\inf_{x \in G} I(x).
\] (2)

First we state the properties of \( \Lambda^*(\cdot) \) and \( \Lambda \) that are needed for proving Cramer’s Theorem.

**Lemma 2** (a) \( \Lambda \) and \( \Lambda^* \) are both convex functions on \( \mathbb{R} \).

(b) If \( \mathcal{D}_\Lambda := \{\lambda : \Lambda(\lambda) < \infty\} = \{0\} \), then \( \Lambda^* \) is identically zero. If \( \Lambda(\lambda) < \infty \) for some \( \lambda > 0 \), then \( \mu < \infty \) and for all \( x \geq \mu \),

\[
\Lambda^*(x) = \sup_{\lambda \geq 0} \{\lambda x - \Lambda(\lambda)\},
\] (3)
is a nondecreasing function on $(\mu, \infty)$. Similarly, if $\Lambda(\lambda) < \infty$ for some $\lambda < 0$, then $\mu > -\infty$ and for all $x \leq \mu$,
\[
\Lambda^*(x) = \sup_{\lambda \leq 0} \{\lambda x - \Lambda(\lambda)\},
\]
(4)
is a nonincreasing function on $(-\infty, \mu)$. When $\mu$ is finite, $\Lambda^*(\mu) = 0$, and we always have
\[
\inf_{x \in \mathbb{R}} \Lambda^*(x) = 0.
\]
(5)

(c) $\Lambda(\cdot)$ is differentiable in $\mathcal{D}_\alpha^c$ with
\[
\Lambda'(\eta) = \frac{1}{M(\eta)} E(X_1 e^{\eta X_1})
\]
and
\[
\Lambda'(\eta) = y \implies \Lambda^*(y) = \eta y - \Lambda(\eta).
\]
(6)

Proof. (a) The convexity of $\Lambda$ follows from Hölder’s inequality
\[
\Lambda(\theta \lambda_1 + (1 - \theta) \lambda_2) = \log E[(e^{\lambda_1 X_1})^\theta (e^{\lambda_2 X_1})^{1-\theta}]
\leq \log \{E(e^{\lambda_1 X_1})^\theta E(e^{\lambda_2 X_1})^{1-\theta}\}
= \theta \Lambda(\lambda_1) + (1 - \theta) \Lambda(\lambda_2).
\]
The convexity of $\Lambda^*$ follows from its definition
\[
\theta \Lambda^*(x_1) + (1 - \theta) \Lambda^*(x_2) = \sup_{\lambda \in \mathbb{R}} \{\theta \lambda x_1 - \theta \Lambda(\lambda)\} + \sup_{\lambda \in \mathbb{R}} \{(1 - \theta) \lambda x_2 - (1 - \theta) \Lambda(\lambda)\}
\geq \sup_{\lambda \in \mathbb{R}} \{(\theta x_1 + (1 - \theta) x_2) \lambda - \Lambda(\lambda)\}
= \Lambda^*(\theta x_1 + (1 - \theta) x_2).
\]
(b) If $\mathcal{D}_\Lambda = \{0\}$, then $\Lambda^*(x) = \Lambda(0) = 0$ for all $x \in \mathbb{R}$. If $\Lambda(\lambda) = \log M(\lambda) < \infty$ for some $\lambda > 0$, then $\int_0^\infty x \, dv \leq M(\lambda)/\lambda < \infty$, implying $\mu = \int x \, dv < \infty$. Now, for all $\lambda \in \mathbb{R}$, by Jensen’s inequality
\[
\Lambda(\lambda) = \log E(e^{\lambda X_1}) \geq E(\log e^{\lambda X_1}) = \lambda \mu.
\]
If $\mu = -\infty$, then $\Lambda(\lambda) = \infty$ for $\lambda$ negative and (3) holds trivially. When $\mu$ is finite, it follows from the proceeding inequality that $\Lambda^*(\mu) = 0$. In this case, for any $x \geq \mu$ and every $\lambda < 0$,
\[
\lambda x - \Lambda(\lambda) \leq \lambda \mu - \Lambda(\lambda) \leq \Lambda^*(\mu) = 0,
\]
and (3) follows. Observe that (3) implies the monotonicity of $\Lambda^*$ on $(\mu, \infty)$, since for any $\lambda \geq 0$, $\lambda x - \Lambda(\lambda)$ is a nondecreasing function of $x$.

When $\Lambda(\lambda) < \infty$ for some $\lambda < 0$, both (4) and the monotonicity of $\Lambda^*$ on $(-\infty, \mu)$ follows by considering the logarithmic moment generating function of $-X_1$, for which the proceeding proof applies.
It remains to prove that \( \inf_{x \in \mathbb{R}} \Lambda^* (x) = 0 \). This is already established for \( \mathcal{D}_\Lambda = \{0\} \), in which case \( \Lambda^* \equiv 0 \), and when \( \mu \) is finite, in which case, \( \Lambda^* (\mu) = 0 \). Now consider the case \( \mu = -\infty \) while \( \Lambda(\lambda) < \infty \) for some \( \lambda > 0 \). Then by (3) and Chebyshev

\[
\log \nu([x, \infty)) \leq \inf_{\lambda \geq 0} \log E(e^{\lambda(X_1 - x)})
= -\sup_{\lambda \geq 0} \{\lambda x - \Lambda(\lambda)\} = -\Lambda^*(x).
\]

Hence

\[
\lim_{x \to -\infty} \Lambda^*(x) \leq \lim_{x \to -\infty} \{-\log \nu([x, \infty))\} = 0,
\]
and (5) follows.

The last remaining case, that of \( \mu = \infty \) while \( \Lambda(\lambda) < \infty \) for some \( \lambda < 0 \), is settled by considering the logarithmic moment generating function of \(-X_1\).

(c) The identity (6) follows by interchanging the order of differentiation and integration. This is justified by the dominated convergence theorem since \( f_{\epsilon}(x) := (e^{\epsilon x} - e^{\epsilon x})/\epsilon \) converges pointwise to \( xe^{\epsilon x} \) as \( \epsilon \to 0 \), and \( |f_{\epsilon}(x)| \leq e^{\epsilon x}(e^{\epsilon x} - 1)/\epsilon : = h(x) \) for every \( \epsilon \in (-\delta, \delta) \), while \( E|h(X_1)| < \infty \) for \( \delta \) small enough.

Let \( \Lambda'(\eta) = y \) and consider the function \( g(\lambda) := \lambda y - \Lambda(\lambda) \). Since \( g(\cdot) \) is a concave function and \( g'(\eta) = 0 \), it follows that \( g(\eta) = \sup_{\lambda \in \mathbb{R}} g(\lambda) \) and (7) is established.

**Proof of Cramer’s Theorem** (a) Let \( F \) be a nonempty closed subset of \( \mathbb{R} \). (1) holds trivially when \( I_F := \inf_{x \in F} \Lambda^*(x) = 0 \). Assume \( I_F > 0 \). It follows from part (b) of the lemma that \( \mu \) exists. For all \( x \) and every \( \lambda > 0 \), Chebyshev yields

\[
\nu_n([x, \infty)) = E(1_{\left\{ \frac{x}{\lambda} - x \geq 0 \right\}}) \leq E(e^{\lambda \frac{x}{\lambda} - x}) = e^{-n \lambda x} \prod_{i=1}^{n} E(e^{\lambda X_i}) = e^{-n(\lambda x - \Lambda(\lambda))}
\]

Therefore, if \( \mu < \infty \), then by (3) for every \( x > \mu \),

\[
\nu_n([x, \infty)) \leq e^{-n \Lambda^*(x)}.
\]

By a similar argument, if \( \mu > -\infty \) and \( x < \mu \), then

\[
\nu_n((\infty, x)) \leq e^{-n \Lambda^*(x)}.
\]

First, consider the case of \( \mu \) finite. Then \( \Lambda^*(\mu) = 0 \), and because by assumption \( I_F > 0 \), \( \mu \) must be in \( F^c \). Let \((x_-, x_+)\) be the component of \( F^c \) containing \( \mu \). Note that \( x_- < x_+ \) and either \( x_- \) or \( x_+ \) must be finite since \( F \) is nonempty. 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. Applying (9) for \( x = x_+ \) and (10) for \( x = x_- \) we get

\[
\nu_n(F) \leq \nu_n((\infty, x_-)) + \nu_n((x_+, \infty)) \leq 2e^{-I_F}
\]
and the upper bounds follows.  

Suppose now that $\mu = -\infty$. Then, since $\Lambda^*$ is nondecreasing, it follows from (5) that $\lim_{x \to -\infty} \Lambda^*(x) = 0$, and hence $x_+ = \inf\{x : x \in F\}$ is finite for otherwise $I_F = 0$. Since $F$ is closed, $x_+ \in F$ and consequently $\Lambda^*(x_+) \geq I_F$. Moreover, $F \subset [x_+, \infty)$ and, therefore, the large deviation upper bounds follows by applying (9) for $x = x_+$.

The case of $\mu = \infty$ is handled analogously.

(b) We prove next that for every $\delta > 0$

$$\liminf \frac{1}{n} \log \nu_n((-\delta, \delta)) \geq \inf_{\lambda \in \mathbb{R}} \Lambda(\lambda) = -\Lambda^*(0). \tag{11}$$

Since the transform $Y = X - x$ results in $\Lambda_Y(\lambda) = \Lambda(\lambda) - \lambda x$ and hence $\Lambda_Y^*(\cdot) = \Lambda^*(\cdot + x)$, it follows from the proceeding inequality that for every $x$ and every $\delta > 0$,

$$\liminf \frac{1}{n} \log \nu_n((x - \delta, x + \delta)) \geq -\Lambda^*(x). \tag{12}$$

For any open subset $G$ of $\mathbb{R}$ and any $x \in G$, and any $\delta > 0$ small enough, $(x - \delta, x + \delta) \subset G$. Thus the large deviation lower bounds follows from (12).

Turning to the proof of the key inequality (11), first suppose that $\nu((-\infty, 0)) > 0$, $\nu((0, \infty)) > 0$ and $\nu$ is supported by a bounded subset of $\mathbb{R}$. By the former assumption, $\Lambda(\lambda) \to \infty$ as $|\lambda| \to \infty$, and by the later assumption, $\Lambda(\cdot)$ is finite everywhere. Accordingly, $\Lambda(\cdot)$ is a continuous, differentiable function, and hence there exists a finite $\eta$ such that $\Lambda(\eta) = \inf_{\lambda \in \mathbb{R}} \Lambda(\lambda)$ and $\Lambda'(\eta) = 0$. Define a new probability measure $\tilde{\nu}$ in terms of $\nu$ by

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

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

$$\int_{\mathbb{R}} d\tilde{\nu} = \frac{1}{M(\eta)} \int e^{n\eta} d\nu = 1.$$  

Let $\tilde{\nu}_n$ be the distribution of $\frac{S_n}{n}$ when $X_1, X_2, \cdots$ are iid with distribution $\tilde{\nu}$. Note that for every $\epsilon > 0$,

$$\nu_n((-\epsilon, \epsilon)) = \int_{|\sum_{i=1}^{n} x_i| < n\epsilon} \nu(dx_1) \cdots \nu(dx_n)$$

$$\geq e^{-n\epsilon\eta} \int_{|\sum_{i=1}^{n} x_i| < n\epsilon} \exp(\eta \sum_{i=1}^{n} x_i) \nu(dx_1) \cdots \nu(dx_n)$$

$$= e^{-n\epsilon\eta} e^{n\Lambda(\eta)} \tilde{\nu}_n((-\epsilon, \epsilon)) \tag{13}$$

By (6) and the choice of $\eta$

$$E_{\tilde{\nu}}(X_1) = \frac{1}{M(\eta)} \int xe^{n\eta} d\nu = \Lambda'(\eta) = 0.$$  

Hence, by the weak law of large numbers,

$$\lim_{n \to \infty} \tilde{\nu}_n((-\epsilon, \epsilon)) = 1. \tag{14}$$
It now follows from (13) that for every $0 < \epsilon < \delta$,
\[
\liminf \frac{1}{n} \log \nu_n((-\delta, \delta)) \geq \liminf \frac{1}{n} \log \nu_n((-\epsilon, \epsilon)) \geq \Lambda(\eta) - \epsilon \|\eta\|
\]
and (11) follows by letting $\epsilon \to 0$.

Suppose that $\nu$ is of unbounded support, while both $\nu((-\infty, 0)) > 0$ and $\nu((0, \infty)) > 0$. Fix $M$ large enough so that $\nu([-M, 0)) > 0$ and $\nu((0, M]) > 0$, and let
\[
\Lambda_M(\lambda) = \log \int_{-M}^{M} e^{\lambda x} d\nu.
\]
Let $\mathcal{P}$ be the probability measure defined by
\[
\mathcal{P}(A) = \frac{\nu(A \cap [-M, M])}{\nu([-M, M])},
\]
and let $\mathcal{P}_n$ be the distribution of $\frac{S_n}{n}$ when $X_1, X_2, \cdots$ are iid with distribution $\mathcal{P}$. The for all $n$ and every $\delta > 0$,
\[
\nu_n((-\delta, \delta)) \geq \mathcal{P}_n((-\delta, \delta)) \nu([-M, M])^n.
\]
Observe that by the proceeding paragraph, (11) holds for $\mathcal{P}_n$. Therefore, with the logarithmic moment generating function associated with $\mathcal{P}$ being $\Lambda_M(\lambda) - \log \nu([-M, M])$,
\[
\liminf \frac{1}{n} \log \nu_n((-\delta, \delta)) \geq \log \nu([-M, M]) + \liminf \frac{1}{n} \log \mathcal{P}_n((-\delta, \delta)) \geq \inf_{\lambda \in \mathbb{R}} \Lambda_M(\lambda).
\]
With $I_M = -\inf_{\lambda \in \mathbb{R}} \Lambda_M(\lambda)$ and $I^* = \limsup_{M \to \infty} I_M$, it follows that
\[
\liminf \frac{1}{n} \log \nu_n((-\delta, \delta)) \geq -I^*.
\]
Note that $\Lambda_M(\cdot)$ is nondecreasing in $M$, and so is $-I_M$. Moreover, $-I_M \leq \Lambda_M(0) = \Lambda(0) = 0$, 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 nonempty, compact sets that are decreasing with respect to $M$, and hence there exists at least one point, denoted $\lambda_0$, in their intersection. By the monotone convergence theorem, $\Lambda(\lambda_0) = \lim \Lambda_M(\lambda_0) \leq -I^*$, and consequently the bounds (15) yields (11), now for $\nu$ of unbounded support.

The proof of (11) for an arbitrary $\mu$ is completed by observing that if either $\mu((-\infty, 0)) = 0$ or $\mu((0, \infty)) = 0$, then $\Lambda$ is a monotone function with $\inf_{\lambda \in \mathbb{R}} \Lambda(\lambda) = \log \nu(\{0\})$. Hence, in this case, (11) follows from
\[
\nu_n((-\delta, \delta)) \geq \nu_n(\{0\}) = \nu(\{0\})^n.
\]

**Lemma 3** If $0 \in \mathcal{D}^\nu_\Lambda$ then $\Lambda^*$ is a good rate function. Moreover, if $\mathcal{D}_\Lambda = \mathbb{R}$, then
\[
\lim_{|x| \to \infty} \frac{\Lambda^*(x)}{|x|} = \infty.
\]

7
Proof. As $0 \in \mathcal{D}_\Lambda^0$, there exist $\lambda_- < 0$ and $\lambda_+ > 0$ that are both in $\mathcal{D}_\Lambda$. Since for any $\lambda \in \mathbb{R}$,

$$\frac{\Lambda^*(x)}{|x|} \geq \lambda \text{sign}(x) - \frac{\Lambda(\lambda)}{|x|},$$

it follows that

$$\liminf_{|x| \to \infty} \frac{\Lambda^*(x)}{|x|} \geq \min\{\lambda_+, -\lambda_-\} > 0.$$

In particular, $\Lambda^*(x) \to \infty$ as $|x| \to \infty$, and its level sets are closed and bounded, hence compact. Thus $\Lambda^*$ is a good rate function. Note that (16) follows for $\mathcal{D}_\Lambda = \mathbb{R}$ by considering $-\lambda_- = \lambda_+ \to \infty$.

Example 2

(a) If the iid sequence $X_1, X_2, \cdots$ has the normal distribution with mean $\mu$ and variance $\sigma^2$ as their common distribution, then $M(\lambda) = \exp(\mu \lambda + \frac{\sigma^2 \lambda^2}{2})$, $\Lambda(\lambda) = \mu \lambda + \frac{\sigma^2 \lambda^2}{2}$, and $\Lambda^*(x) = \frac{(x-\mu)^2}{2\sigma^2}$.

(b) If the iid sequence $X_1, X_2, \cdots$ has the exponential distribution with parameter $\theta > 0$ as their common distribution, then $M(\lambda) = \frac{\theta}{\lambda}$ for $\lambda < \theta$, and $\infty$ otherwise; $\Lambda(\lambda) = \log \frac{\theta}{\lambda}$ for $\lambda < \theta$, and $\infty$ otherwise; $\Lambda^*(x) = \theta x - 1 - \log(\theta x)$ for $x > 0$, and $\infty$ otherwise.

(c) If the iid sequence $X_1, X_2, \cdots$ has the Bernoulli distribution with parameter $p \in (0, 1)$ as their common distribution, then $M(\lambda) = pe^\lambda + 1 - p$, $\Lambda^*(x) = x \log(\frac{x}{p}) + (1 - p) \log(\frac{1 - x}{1 - p})$ for $x \in [0, 1]$ and $\infty$ otherwise.

(d) If the iid sequence $X_1, X_2, \cdots$ has the Poisson distribution with parameter $\theta > 0$ as their common distribution, then $M(\lambda) = \exp(\theta (e^\lambda - 1))$ and $\Lambda^*(x) = \theta - x + x \log(\frac{\theta}{x})$ for $x \geq 0$ and $\infty$ otherwise.