MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.265/15.070J — Fall 2013

Lecture 3 — 9/11/2013

Large deviations Theory. Cramér's Theorem

Content.

1. Cramér's Theorem.
2. Rate function and properties.
3. Change of measure technique.

1 Cramér's Theorem

We have established in the previous lecture that under some assumptions on the Moment Generating Function (MGF) $M(\theta)$, an i.i.d. sequence of random variables $X_i, 1 \leq i \leq n$ with mean μ satisfies $\mathbb{P}(S_n \geq a) \leq \exp(-nI(a))$, where $S_n = n^{-1}\sum_{1\leq i\leq n} X_i$, and $I(a) \triangleq \sup_\theta(\theta a - \log M(\theta))$ is the Legendre transform. The function $I(a)$ is also commonly called the *rate* function in the theory of Large Deviations. The bound implies

$$\limsup_n \frac{\log \mathbb{P}(S_n \geq a)}{n} \leq -I(a),$$

and we have indicated that the bound is tight. Namely, ideally we would like to establish the limit

$$\limsup_n \frac{\log \mathbb{P}(S_n \geq a)}{n} = -I(a),$$

Furthermore, we might be interested in more complicated rare events, beyond the interval $[a, \infty)$. For example, the likelihood that $\mathbb{P}(S_n \in A)$ for some set $A \subset \mathbb{R}$ not containing the mean value μ. The Large Deviations theory says that roughly speaking

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{P}(S_n \in A) = -\inf_{x\in A} I(x), \quad (1)$$

but unfortunately this statement is not precisely correct. Consider the following example. Let X be an integer-valued random variable, and $A = \{\frac{m}{p} : m \in \mathcal{Z}, p \text{ is odd prime.}\}$. Then for prime n, we have $\mathbb{P}(S_n \in A) = 1$; but for $n = 2^k$, we have $P(S_n \in A) = 0$. As a result, the limit $\lim_{n\to\infty} \frac{\log P(S_n \in A)}{n}$ in this case does not exist.

The sense in which the identity (1) is given by the Cramér's Theorem below.

Theorem 1 (**Cramér's Theorem**). *Given a sequence of i.i.d. real valued random variables $X_i, i \geq 1$ with a common moment generating function $M(\theta) = E[\exp(\theta X_1)]$ the following holds:*

(a) For any closed set $F \subseteq \mathbb{R}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \in F) \leq -\inf_{x\in F} I(x),$$

(b) For any open set $U \subseteq \mathbb{R}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \in U) \geq -\inf_{x\in U} I(x).$$

We will prove the theorem only for the special case when $\mathcal{D}(M) = \mathbb{R}$ (namely, the MGF is finite everywhere) and when the support of X is entire $\mathbb{R}$. Namely for every $K > 0$, $\mathbb{P}(X > K) > 0$ and $\mathbb{P}(X < -K) > 0$. For example a Gaussian random variable satisfies this property.

To see the power of the theorem, let us apply it to the tail of S_n. In the following section we will establish that $I(x)$ is a non-decreasing function on the interval $[\mu, \infty)$. Furthermore, we will establish that if it is finite in some interval containing x it is also continuous at x. Thus fix a and suppose I is finite in an interval containing a. Taking F to be the closed set $[a, \infty)$ with $a > \mu$, we obtain from the

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \in [a,\infty)) &\leq -\min_{x\geq a} I(x) \\ &= -I(a).\end{aligned}$$

Applying the second part of Cramér's Theorem, we obtain

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \in [a,\infty)) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \in (a,\infty)) \\ \geq -\inf_{x>a} I(x) \\ = -I(a).\end{aligned}$$

Thus in this special case indeed the large deviations limit exists:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq a) = -I(a).$$

The limit is insensitive to whether the inequality is strict, in the sense that we also have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n > a) = -I(a).$$

2 Properties of the rate function I

Before we prove this theorem, we will need to establish several properties of $I(x)$ and $M(\theta)$.

Proposition 1. *The rate function I satisfies the following properties*

(a) I is a convex non-negative function satisfying $I(\mu) = 0$. Furthermore, it is an increasing function on $[\mu, \infty)$ and a decreasing function on $(-\infty, \mu]$. Finally $I(x) = \sup_{\theta \geq 0}(\theta x - \log M(\theta))$ for every $x \geq \mu$ and $I(x) = \sup_{\theta \leq 0}(\theta x - \log M(\theta))$ for every $x \leq \mu$.

(b) Suppose in addition that $\mathcal{D}(M) = \mathbb{R}$ and the support of X_1 is $\mathbb{R}$. Then, I is a finite continuous function on $\mathbb{R}$. Furthermore, for every $x \in \mathbb{R}$ we have $I(x) = \theta_0 x - \log M(\theta_0)$, for some $\theta_0 = \theta_0(x)$ satisfying

$$x = \frac{\dot{M}(\theta_0)}{M(\theta_0)}. \tag{2}$$

Proof of part (a). Convexity is due to the fact that $I(x)$ is point-wise supremum. Precisely, consider $\lambda \in (0, 1)$

$$\begin{aligned} I(\lambda x + (1-\lambda)y) =& \sup_\theta [\theta(\lambda x + (1-x)y) - \log M(\theta)] \\ =& \sup[\lambda(x - \log M(\theta)) + (1-\lambda)(y - \log M(\theta))] \\ \leq& \lambda \sup_\theta (x - \log M(\theta)) + (1-\lambda) \sup_\theta (y - \log M(\theta)) \\ =& \lambda I(x) + (1-\lambda) I(y). \end{aligned}$$

This establishes the convexity. Now since $M(0) = 1$ then $I(x) \geq 0 \cdot x - \log M(0) = 0$ and the non-negativity is established. By Jensen's inequality, we have that

$$M(\theta) = \mathbb{E}[\exp(\theta X_1)] \geq \exp(\theta \mathbb{E}[X_1]) = \exp(\theta\mu).$$

Therefore, $\log M(\theta) \geq \theta\mu$, namely, $\theta\mu - \log M(\theta) \leq 0$, implying $I(\mu) = 0 = \min_{x\in\mathbb{R}} I(x)$.

Furthermore, if $x > \mu$, then for $\theta < 0$ we have $\theta x - \log M(\theta) \leq \theta(x - \mu) < 0$. This means that $\sup_\theta(\theta x - \log M(\theta))$ must be equal to $\sup_{\theta\geq 0}(\theta x - \log M(\theta))$. Similarly we show that when $x < \mu$, we have $I(x) = \sup_{\theta\leq 0}(\theta x - \log M(\theta))$.

Next, the monotonicity follows from convexity. Specifically, the existence of real numbers $\mu \leq x < y$ such that $I(x) > I(y) \geq I(\mu) = 0$ violates convexity (check). This completes the proof of part (a).

□

Proof of part (b). For any $K > 0$ we have

$$
\begin{aligned}
\liminf_{\theta\to\infty} \frac{\log M(\theta)}{\theta} &= \liminf_{\theta\to\infty} \frac{\log\left(\int \exp(\theta x)\,\mathrm{d}P(x)\right)}{\theta} \\
&\geq \liminf_{\theta\to\infty} \frac{1}{\theta}\log\left(\int_K^\infty \exp(\theta x)\,\mathrm{d}P(x)\right) \\
&\geq \liminf_{\theta\to\infty} \frac{1}{\theta}\log\left(\exp(K\theta)\mathbb{P}([K,\infty])\right) \\
&= K + \liminf_{\theta\to\infty} \frac{1}{\theta}\log\mathbb{P}([K,\infty]) \\
&= K \text{ (since } \mathit{supp}(X_1) = \mathbb{R}\text{, we have } \mathbb{P}([K,\infty)) > 0.)
\end{aligned}
$$

Since K is arbitrary,

$$\liminf_{\theta\to\infty} \frac{1}{\theta}\log M(\theta) = \infty$$

Similarly,

$$\liminf_{\theta\to-\infty} -\frac{1}{\theta}\log M(\theta) = \infty$$

Therefore,

$$\lim_{\theta\to\infty} \theta x - \log M(\theta) = \lim_{\theta\to\infty} \theta(x - \frac{1}{\theta}\log M(\theta)) \to -\infty$$

Therefore, for each x as $|\theta| \to \infty$, we have that

$$\lim_{|\theta|\to\infty} \theta x - \log M(\theta) = -\infty$$

From the previous lecture we know that $M(\theta)$ is differentiable (hence continuous). Therefore the supremum of $\theta x - \log M(\theta)$ is achieved at some finite value $\theta_0 = \theta_0(x)$, namely,

$$I(x) = \theta_0 x - \log M(\theta_0) < \infty,$$

where θ_0 is found by setting the derivative of $\theta x - \log M(\theta)$ to zero. Namely, θ_0 must satisfy (2). Since I is a finite convex function on $\mathbb{R}$ it is also continuous (verify this). This completes the proof of part (b). $\square$

3 Proof of Cramér's Theorem

Now we are equipped to proving the Cramér's Theorem.

Proof of Cramér's Theorem. Part (a). Fix a closed set $F \subset \mathbb{R}$. Let $\alpha_+ = \min\{x \in [\mu, +\infty) \cap F\}$ and $\alpha_- = \max\{x \in (-\infty, \mu] \cap F\}$. Note that α_+ and α_- exist since F is closed. If $\alpha_+ = \mu$ then $I(\mu) = 0 = \min_{x \in \mathbb{R}} I(x)$. Note that $\log \mathbb{P}(S_n \in F) \leq 0$, and the statement (a) follows trivially. Similarly, if $\alpha_- = \mu$, we also have statement (a). Thus, assume $\alpha_- < \mu < \alpha_+$. Then

$$\mathbb{P}\left(S_n \in F\right) \leq \mathbb{P}\left(S_n \in [\alpha_+, \infty)\right) + \mathbb{P}\left(S_n \in (-\infty, \alpha_-]\right)$$

Define

$$x_n \triangleq \mathbb{P}\left(S_n \in [\alpha_+, \infty)\right), \ y_n \triangleq \mathbb{P}\left(S_n \in (-\infty, \alpha_-]\right).$$

We already showed that

$$\mathbb{P}\left(S_n \geq \alpha_+\right) \leq \exp(-n(\theta\alpha_+ - \log M(\theta))), \ \forall \theta \geq 0.$$

from which we have

$$\frac{1}{n} \log \mathbb{P}\left(S_n \geq \alpha_+\right) \leq -\left(\theta\alpha_+ - \log M(\theta)\right), \ \forall \theta \geq 0.$$
$$\Rightarrow \frac{1}{n} \log \mathbb{P}\left(S_n \geq \alpha_+\right) \leq -\sup_{\theta \geq 0}(\theta\alpha_+ - \log M(\theta)) = -I(\alpha_+)$$

The second equality in the last equation is due to the fact that the supremum in $I(x)$ is achieved at $\theta \geq 0$, which was established as a part of Proposition 1. Thus, we have

$$\limsup_n \frac{1}{n} \log \mathbb{P}\left(S_n \geq \alpha_+\right) \leq -I(\alpha_+) \tag{3}$$

Similarly, we have

$$\limsup_n \frac{1}{n} \log \mathbb{P}\left(S_n \leq \alpha_-\right) \leq -I(\alpha_-) \tag{4}$$

Applying Proposition 1 we have $I(\alpha_+) = \min_{x \geq \alpha_+} I(x)$ and $I(\alpha_-) = \min_{x \leq \alpha_-} I(x)$. Thus

$$\min\{I(\alpha_+), I(\alpha_-)\} = \inf_{x \in F} I(x) \tag{5}$$

From (3)-(5), we have that

$$\limsup_n \frac{1}{n} \log x_n \leq -\inf_{x \in F} I(x), \ \limsup_n \frac{1}{n} \log y_n \leq -\inf_{x \in F} I(x), \qquad (6)$$

which implies that

$$\limsup_n \frac{1}{n} \log(x_n + y_n) \leq -\inf_{x \in F} I(x).$$

(you are asked to establish the last implication as an exercise). We have established

$$\limsup_n \frac{1}{n} \log \mathbb{P}\left(S_n \in F\right) \leq -\inf_{x \in F} I(x) \qquad (7)$$

Proof of the upper bound in statement (a) is complete. □

Proof of Cramér's Theorem. Part (b). Fix an open set $U \subset \mathbb{R}$. Fix $\epsilon > 0$ and find y such that $I(y) \leq \inf_{x \in U}((x)$. It is sufficient to show that

$$\liminf_{n \to \infty} \frac{1}{n} \mathbb{P}\left(S_n \in U\right) \geq -I(y), \qquad (8)$$

since it will imply

$$\liminf_{n \to \infty} \frac{1}{n} \mathbb{P}\left(S_n \in U\right) \geq -\inf_{x \in U} I(x) + \epsilon,$$

and since $\epsilon > 0$ was arbitrary, it will imply the result.

Thus we now establish (8). Assume $y > \mu$. The case $y < \mu$ is treated similarly. Find $\theta_0 = \theta_0(y)$ such that

$$I(y) = \theta_0 y - \log M(\theta_0).$$

Such θ_0 exists by Proposition 1. Since $y > \mu$, then again by Proposition 1 we may assume $\theta_0 \geq 0$.

We will use the change-of-measure technique to obtain the cover bound. For this, consider a new random variable let X_{θ_0} be a random variable defined by

$$\mathbb{P}(X_{\theta_0} \leq z) = \frac{1}{M(\theta_0)} \int_{-\infty}^{z} \exp(\theta_0 x) \, \mathrm{d}P(x)$$

Now,

$$\begin{aligned} \mathbb{E}[X_{\theta_0}] &= \frac{1}{M(\theta_0)} \int_{-\infty}^{\infty} x \exp(\theta_0 x) \, \mathrm{d}P(x) \\ &= \frac{\dot{M}(\theta_0)}{M(\theta_0)} \\ &= y, \end{aligned}$$

where the second equality was established in the previous lecture, and the last equality follows by the choice of θ_0 and Proposition 1. Since U is open we can find $\delta > 0$ be small enough so that $(y-\delta, y+\delta) \subset U$. Thus, we have

$$\begin{aligned}
&\mathbb{P}(S_n \in U) \\
&\quad \geq \mathbb{P}(S_n \in (y-\delta, y+\delta)) \\
&\quad = \int_{|\frac{1}{n}\sum x_i - y|<\delta} \mathrm{d}P(x_1)\cdots \mathrm{d}P(x_n) \\
&\quad = \int_{|\frac{1}{n}\sum x_i - y|<\delta} \exp(-\theta_0 \sum_i x_i) M^n(\theta_0) \prod_{1\leq i\leq n} M^{-1}(\theta_0)\exp(\theta_0 x_i) dP(x_i).
\end{aligned} \tag{9}$$

Since θ_0 is non-negative, we obtain a bound

$$\begin{aligned}
&\mathbb{P}(S_n \in (y-\delta, y+\delta)) \\
&\quad \geq \exp(-\theta_0 y n - \theta_0 n\delta) M^n(\theta_0) \int_{|\frac{1}{n}\sum x_i - y|<\delta} \prod_{1\leq i\leq n} M^{-1}(\theta_0)\exp(\theta_0 x_i) dP(x_i)
\end{aligned}$$

However, we recognize the integral on the right-hand side of the inequality above as the that the average $n^{-1}\sum_{1\leq i\leq n} Y_i$ of n i.i.d. random variables $Y_i, 1 \leq i \leq n$ distributed according to the distribution of X_{θ_0} belongs to the interval $(y-\delta, y+\delta)$. Recall, however that $\mathbb{E}[Y_i] = \mathbb{E}[X_{\theta_0}] = y$ (this is how X_{θ_0} was designed). Thus by the Weak Law of Large Numbers, this probability converges to unity. As a result

$$\lim_{n\to\infty} n^{-1} \log \int_{|\frac{1}{n}\sum x_i - y|<\delta} \prod_{1\leq i\leq n} M^{-1}(\theta_0)\exp(\theta_0 x_i) dP(x_i) = 0.$$

We obtain

$$\begin{aligned}
\liminf_{n\to\infty} n^{-1} \log \mathbb{P}(S_n \in U) &\geq -\theta_0 y - \theta_0 \delta + \log M(\theta_0) \\
&= -I(y) - \theta_0\delta.
\end{aligned}$$

Recalling that θ_0 depends on y only and sending δ to zero, we obtain (8). This completes the proof of part (b). □

15.070J / 6.265J Advanced Stochastic Processes
Fall 2013