Cramér's Theorem

Large Deviations and Queues—Damon Wischik

Theorem 1 *Let $(X_n,\ n \in \mathbb{N})$ be a sequence of independent random variables each distributed like X, and let $S_n = X_1 + \cdots + X_n$. Let $\Lambda(\theta) = \log \mathbb{E}e^{\theta X}$, and let $\Lambda^*(x) = \sup_{\theta\in\mathbb{R}} \theta x - \Lambda(\theta)$. Suppose that Λ is finite in a neighbourhood of zero. Then for any measurable set $B \subseteq \mathbb{R}$*

$$-\inf_{x\in B^\circ} \Lambda^*(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in B\Big) \tag{1}$$

$$\leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in B\Big) \leq -\inf_{x\in \bar{B}} \Lambda^*(x). \tag{2}$$

Proof. We first establish the upper bound (1) for closed half-spaces, i.e. sets of the form $[x,\infty)$ and $(-\infty,x]$. We then extend it to all closed sets. We then establish the lower bound (2).

Upper bound for closed half-spaces. Write out the probability we wish to estimate:

$$\begin{aligned}\mathbb{P}\Big(\frac{S_n}{n} \in [x,\infty)\Big) &= \mathbb{P}\big(S_n \geq nx\big) = \mathbb{E}1_{S_n - nx \geq 0}\\ &\leq \mathbb{E}e^{\theta(S_n - nx)} = e^{-n\theta x}\big(\mathbb{E}e^{\theta X}\big)^n \quad \text{for all } \theta \geq 0.\end{aligned}$$

The inequality is known as the Chernoff bound. Assume for the moment that $x \geq \mathbb{E}X$. Taking logarithms,

$$\begin{aligned}\frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in [x,\infty)\Big) &\leq \inf_{\theta\geq 0}\big\{-\theta x + \Lambda(\theta)\big\}\\ &= -\sup_{\theta\geq 0}\big\{\theta x - \Lambda(\theta)\big\}\\ &= -\sup_{\theta\in\mathbb{R}}\big\{\theta x - \Lambda(\theta)\big\} \qquad (3)\\ &= -\Lambda^*(x).\end{aligned}$$

To see that the supremum can be taken over $\theta \in \mathbb{R}$ in (3), note that

$$\Lambda(\theta) = \log \mathbb{E}e^{\theta X} \geq \log e^{\theta \mathbb{E}X} = \theta\mathbb{E}X \quad \text{by Jensen's inequality}$$

and hence that $\theta x - \Lambda(\theta) \leq \theta(x - \mathbb{E}X)$; thus $\theta x - \Lambda(\theta) \leq 0$ whenever $\theta \leq 0$, and so the supremum in (3) is attained for $\theta \geq 0$. Finally, note that $\Lambda^*(x)$ is increasing in $x > \mathbb{E}X$, since for any $y \geq x$

$$\Lambda^*(x) = \sup_{\theta\geq 0} \theta x - \Lambda(\theta) \leq \sup_{\theta\geq 0} \theta y - \Lambda(\theta) = \Lambda^*(y).$$

Thus we have proved that for $x \geq \mathbb{E}X$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in [x,\infty)\Big) \leq -\inf_{y\in[x,\infty)} \Lambda^*(y). \tag{4}$$

It remains to deal with the case $x < \mathbb{E}X$. In this case, trivially,

$$\frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in [x,\infty)\Big) \leq 0$$

since $\mathbb{P}(\cdot) \leq 1$. Also note that $\Lambda^*(\mathbb{E}X) \leq 0$, using Jensen's inequality as before, and that $\Lambda^*(\cdot)$ is clearly non-negative; hence that $\Lambda^*(\mathbb{E}X) = 0$. This implies that

$$\frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in [x, \infty)\Big) \leq 0 = \inf_{y \in [x,\infty)} \Lambda^*(y).$$

So we have proved that (4) holds also for $x < \mathbb{E}X$. The proof of the upper bound for sets of the form $(-\infty, x]$ follows by considering the random variable $-X$.

LD upper bound for general closed sets. Let F be an arbitrary closed set. If F contains $\mathbb{E}X$, then the LD upper bound holds trivially since

$$\inf_{x \in F} \Lambda^*(x) = \Lambda^*(\mathbb{E}X) = 0.$$

Otherwise, F can be written as the union $F = F_1 \cup F_2$ where F_1 and F_2 are closed and

$$F_1 \subseteq [\mathbb{E}X, \infty) \text{ and } F_2 \subseteq (-\infty, \mathbb{E}X).$$

Suppose F_1 is non-empty, and let x be the infimum of F_1. By closure, $x \in F_1$. Now,

$$\begin{aligned}
\frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in F_1\Big) &\leq \frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in [x, \infty)\Big) \\
&\leq -\Lambda^*(x) \quad \text{by the upper bound for closed half-spaces} \\
&= -\inf_{y \in F_1} \Lambda^*(y)
\end{aligned}$$

where the last equality is by monotonicity of Λ^* on $[\mathbb{E}X, \infty)$, in which F_1 is contained. Similarly, by considering the LD upper bound for $(-\infty, x]$, where x is the supremum of F_2, we obtain

$$\frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in F_2\Big) \leq -\inf_{y \in F_2} \Lambda^*(y).$$

In other words, the LD upper bound holds for both of F_1 and F_2. Hence, by the principle of the largest term, it holds for $F = F_1 \cup F_2$.

LD lower bound. Let G be any open set, and let $x \in G$. We will show that

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in G\Big) \geq -\Lambda^*(x). \tag{5}$$

Taking the supremum over $x \in G$ will then yield the large deviations lower bound. We will proceed by calculating the value of $\Lambda^*(x)$. We will do this in two cases: first the case when $\mathbb{P}(X < x) = 0$ or $\mathbb{P}(X > x) = 0$, second the case when neither holds.

Suppose that $\mathbb{P}(X < x) = 0$. We can calculate Λ^* explicitly as follows:

$$\begin{aligned}
\Lambda^*(x) = \sup_{\theta \in \mathbb{R}} \Big\{\theta x - \Lambda(\theta)\Big\} &= -\inf_{\theta \in \mathbb{R}} \Big\{\Lambda(\theta) - \theta x\Big\} = -\inf_{\theta \in \mathbb{R}} \log \mathbb{E} e^{\theta(X - x)} \\
&= -\lim_{\theta \to -\infty} \log \mathbb{E} e^{\theta(X-x)} \quad \text{since } X \geq x \text{ almost surely} \\
&= -\log \lim_{\theta \to -\infty} \mathbb{E} e^{\theta(X-x)} \\
&= -\log \mathbb{E} 1_{X=x} \quad \text{by monotone convergence} \\
&= -\log \mathbb{P}(X = x).
\end{aligned}$$

If $\mathbb{P}(X = x) = 0$, then the lower bound in (5) is trivial. If $\mathbb{P}(X = x) = p > 0$ then

$$\begin{aligned}
\frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in (x - \delta, x + \delta)\Big) &\geq \frac{1}{n} \log \mathbb{P}\big(X_1 = \cdots = X_n = x\big) \\
&= \frac{1}{n} \log p^n = \log p = -\Lambda^*(x)
\end{aligned}$$

holds.

Assume now that $\mathbb{P}(X > x) > 0$ and $\mathbb{P}(X < x) > 0$. Again, we investigate the value of the lower bound:

$$\begin{aligned}\Lambda^*(x) &= \sup_{\theta\in\mathbb{R}} \theta x - \Lambda(\theta) \\ &= -\inf_{\theta\in\mathbb{R}} \Lambda(\theta) - \theta x = -\inf_{\theta\in\mathbb{R}} \log \mathbb{E}e^{\theta(X-x)}.\end{aligned}$$

Now, the function $g(\theta) = \Lambda(\theta) - \theta x$ satisfies $g(\theta) \to \infty$ as $|\theta| \to \infty$, by the assumption that there is probability mass both above and below x; and it inherits lower-semicontinuity from Λ (see Lemma 2). Any set of the form $\{g(\theta) \leq \alpha\}$ is thus bounded as well as closed, hence compact, and so g attains its infimum (see the note at the end of this proof), say

$$\Lambda^*(x) = \hat{\theta}x - \Lambda(\hat{\theta}).$$

We will use $\hat{\theta}$ to estimate the probability in question.

We will do this using a *tilted distribution*[1]. Let μ be the measure of X, and define a tilted measure $\tilde{\mu}$ by

$$\frac{d\tilde{\mu}}{d\mu}(x) = e^{\hat{\theta}x - \Lambda(\hat{\theta})}.$$

Let $\tilde{X}$ be a random variable drawn from $\tilde{\mu}$.

Observe that

$$\mathbb{E}\tilde{X} = \mathbb{E}Xe^{\hat{\theta}X-\Lambda(\hat{\theta})} = \Lambda'(\hat{\theta})$$

where the last equality comes from Lemma 2, making the assumption that Λ is differentiable at $\hat{\theta}$. (We will leave the case where it is not differentiable to Dembo & Zeitouni.) Note also that since the optimum in $\Lambda^*(x) = \sup_\theta \theta x - \Lambda(\theta)$ is attained at $\theta = \hat{\theta}$, it must be that $\Lambda'(\hat{\theta}) = x$. Thus $\mathbb{E}\tilde{X} = x$. (This tilted random variables captures the idea of being close in distribution to X, conditional on having a value close to x.)

We can now estimate the probability of interest, using the fact that (since G is open), the set $(x - \delta, x + \delta)$ is contained in G for sufficiently small δ. Let $\tilde{S}_n$ be the sum of n i.i.d. copies of $\tilde{X}$. Then

$$\begin{aligned}&\mathbb{P}\Big(\Big|\frac{S_n}{n} - x\Big| < \delta\Big) \\ &= \int\cdots\int_{|x_1+\cdots+x_n-nx|<n\delta} \mu(dx_1)\cdots\mu(dx_n) \\ &= \int\cdots\int_{|x_1+\cdots+x_n-nx|<n\delta} e^{-\hat{\theta}(x_1+\cdots+x_n)+n\Lambda(\hat{\theta})}\tilde{\mu}(dx_1)\cdots\tilde{\mu}(dx_n) \\ &= \mathbb{E}\Big(e^{-\hat{\theta}\tilde{S}_n+n\Lambda(\hat{\theta})}1_{|\tilde{S}_n/n-x|<\delta}\Big) \\ &\geq \mathbb{E}\Big(e^{-n(\hat{\theta}x-\Lambda(\hat{\theta})+|\hat{\theta}|\delta)}1_{|\tilde{S}_n/n-x|<\delta}\Big) \\ &= e^{-n(\hat{\theta}x-\Lambda(\hat{\theta})+|\hat{\theta}|\delta)}\mathbb{P}\Big(\Big|\frac{\tilde{S}_n}{n} - x\Big| < \delta\Big).\end{aligned}$$

By the weak law of large numbers, and the fact that our tilted distribution has mean x, the term $\mathbb{P}(\cdot)$ tends to 1 as $n \to \infty$. Taking logarithms and then lim inf,

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in G\Big) &\geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\Big|\frac{S_n}{n} - x\Big| < \delta\Big) \\ &\geq -(\hat{\theta}x - \Lambda(\hat{\theta}) + |\hat{\theta}|\delta).\end{aligned}$$

[1] What this means in practice is that $\mathbb{E}f(\tilde{X}) = \mathbb{E}\Big(f(X)\frac{d\tilde{\mu}}{d\mu}(X)\Big)$ or, in integral notation, $\int f(x)\tilde{\mu}(dx) = \int f(x)\frac{d\tilde{\mu}}{d\mu}(x)\mu(dx)$.

But δ was arbitrarily small, so

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{S_n}{n} \in G\Big) \geq -\Lambda^*(x).$$

This completes the proof. □

Here are some basic properties of Λ and Λ^*. Recall that a function $f : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is said to be *lower-semicontinuous* if $x_n \to x$ implies that $\liminf f(x_n) \geq f(x)$, or equivalently that any set $\{x : f(x) \leq \alpha\}$ for $\alpha \in \mathbb{R}$ is closed. It is not hard to prove that if K is a compact set and $\inf_{x\in K} f(x) < \infty$ then the infimum is attained at some $\hat{x} \in K$.

Lemma 2 (Properties of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda^*}$) *Assume that $\Lambda(\theta)$ is finite in a neighbourhood of $\theta = 0$. Then*

i. $\mathbb{E}X$ *is finite and equal to* $\Lambda'(0)$
ii. $\Lambda(0) = 0$
iii. Λ *is convex and lower-semicontinuous*
iv. Λ *is infinitely differentiable in the interior of* $\{\theta : \Lambda(\theta) < \infty\}$, *and* $\Lambda'(\theta) = \mathbb{E}(Xe^{\theta X})/\mathbb{E}e^{\theta X}$
v. $\Lambda^*(\mathbb{E}X) = 0$
vi. Λ^* *is non-negative, convex, and lower-semicontinuous*
vii. $(\Lambda^*)^* = \Lambda$