A short proof of Cramér’s theorem in $\mathbb{R}$

Raphaël Cerf and Pierre Petit

Abstract

We give a short proof of Cramér’s large deviations theorem based on convex duality. This proof does not resort to the law of large numbers or any other limit theorem.

The most fundamental result in probability theory is the law of large numbers for a sequence $(X_n)_{n \geq 1}$ of independent and identically distributed real-valued random variables. Define the empirical mean of $(X_n)_{n \geq 1}$ by

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$ 

The law of large numbers asserts that the empirical mean $\bar{X}_n$ converges (almost surely) towards the theoretical mean $\mathbb{E}(X_1)$ provided that $\mathbb{E}(|X_1|)$ is finite. The next fundamental results are the central limit theorem and Cramér’s theorem. Both are refinements of the law of large numbers in two different directions. The central limit theorem describes the random fluctuations of $\bar{X}_n$ around $\mathbb{E}(X_1)$, in fact “small deviations” of order $1/\sqrt{n}$. Cramér’s theorem estimates the probability that $\bar{X}_n$ deviates significantly from $\mathbb{E}(X_1)$:

$$P(\bar{X}_n \geq \mathbb{E}(X_1) + \varepsilon)$$

for $\varepsilon > 0$, that is, the probability of “large deviation” events. It turns out that this probability decays exponentially fast with $n$ when $|X_1|$ admits exponential moments. The first estimate of this kind can be traced back to Cramér’s paper [6], which deals with variables possessing a density and exponential moments. In [5] Chernoff relaxed the first assumption. Bahadur [2] finally gave a proof without any assumption on the law of $X_1$. Coming from statistical mechanics, Lanford imported the subadditivity argument in the proof [10]. The result of Cramér was generalized in several directions. Bahadur and Zabell [3] extended Cramér’s theory to infinite-dimensional topological vector spaces. Hammersley [9] considered superadditive sequences in $\mathbb{R}$; taking advantage of the order structure of $\mathbb{R}$, he found a quite different efficient proof of the result (we thank Olivier Garet for drawing our attention to Hammersley’s paper [9]). The texts of Azencott [1], Deuschel and Stroock [8], Dembo and Zeitouni [7], and Cerf [4] take stock of the foregoing improvements. The proofs of Cramér’s theorem in $\mathbb{R}$ presented in these texts resort either to the law of large numbers (see, e.g., [7]), Mosco’s theorem (see, e.g., [4]), or another limit theorem. We give here a
direct proof of Cramér’s theorem in \( \mathbb{R} \) which combines the ideas of Hammersley, Lanford, Bahadur, and Zabell, with two simplifying features: we first prove the dual version of Cramér’s theorem (in the sense of convex functions) and we use conditioning by a compact convex set. Not only is the proof clearer, but it can easily adapt to infinite-dimensional spaces and even nonindependent variables.

**Cramér’s theorem.** Let \((X_n)_{n \geq 1}\) be a sequence of independent and identically distributed real-valued random variables and let \(\overline{X}_n\) be the empirical mean:

\[
\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.
\]

For all \(x \in \mathbb{R}\) the sequence

\[
\frac{1}{n} \log P(\overline{X}_n \geq x)
\]

converges in \([-\infty, 0]\) and

\[
\lim_{n \to \infty} \frac{1}{n} \log P(\overline{X}_n \geq x) = \inf_{\lambda \geq 0} \left( \log E(e^{\lambda X_1}) - \lambda x \right).
\]

Let us define the **entropy** of the sequence \((X_n)_{n \geq 1}\) by

\[
\forall x \in \mathbb{R} \quad s(x) = \sup_{n \geq 1} \frac{1}{n} \log P(\overline{X}_n \geq x)
\]

and the **pressure** of \(X_1\) (or the log-Laplace transform of the law of \(X_1\)) by

\[
\forall \lambda \in \mathbb{R} \quad p(\lambda) = \log E(e^{\lambda X_1}).
\]

The entropy and the pressure may take infinite values. We work in \([-\infty, +\infty]\) with the natural extensions of addition and the order relation. Our strategy is to show a dual version, in the sense of convex functions, of Cramér’s theorem.

**Dual equality.** For all \(\lambda \geq 0\),

\[
p(\lambda) = \sup_{u \in \mathbb{R}} \left( \lambda u + s(u) \right).
\]

**Proof.** The classical Chebyshev’s inequality will yield one part of the proof of the above equality. To prove the other part we will condition \(\overline{X}_n\) to be bounded by \(K\) and then let \(K\) grow towards \(+\infty\).

Since \(X_1, \ldots, X_n\) are independent and identically distributed, we have, for all \(\lambda \geq 0\) and \(n \geq 1\),

\[
E(e^{n\lambda \overline{X}_n}) = E(e^{\lambda X_1})^n.
\]
Thus, for \( u \in \mathbb{R} \) and \( n \geq 1 \), it follows from Chebyshev’s inequality that
\[
p(\lambda) = \log \mathbb{E}(e^{\lambda X_1}) = \frac{1}{n} \log \mathbb{E}(e^{n\lambda X_n}) \geq \frac{1}{n} \log (e^{n\lambda u} \mathbb{P}(e^{n\lambda X_n} \geq e^{n\lambda u})) \geq \lambda u + \frac{1}{n} \log \mathbb{P}(X_n \geq u).
\]
Hence, taking the supremum over \( n \geq 1 \) and then over \( u \in \mathbb{R} \), we get
\[
\forall \lambda \geq 0 \quad p(\lambda) \geq \sup_{u \in \mathbb{R}} (\lambda u + s(u)).
\]
Next, we prove equality for \( \lambda = 0 \). Let us state a simple and useful inequality.

**Useful inequality.** For all \( x \in \mathbb{R} \) and \( n \geq 1 \),
\[
\mathbb{P}(X_1 \geq x)^n \leq \mathbb{P}(X_n \geq x) \leq e^{n s(x)}.
\]
In particular, \( s(u) \geq \log \mathbb{P}(X_1 \geq u) \). Hence, letting \( u \) go to \( -\infty \), we see that
\[
\sup_{u \in \mathbb{R}} s(u) = 0 = p(0).
\]
Now we prove the converse inequality for \( \lambda > 0 \). Let \( \lambda > 0 \) and \( K > 0 \). If \( X \) is a random variable, \( 1_{X \leq x} \) will denote the random variable that is 1 if \( X \leq x \) and 0 otherwise. For all \( n \geq 1 \), using the fact that \( X_1, \ldots, X_n \) are independent and identically distributed, we have
\[
\log \mathbb{E}(e^{\lambda X_1 1_{|X_1| \leq K}}) = \frac{1}{n} \log \mathbb{E}(e^{\lambda(X_1 + \cdots + X_n) 1_{|X_1| \leq K} \cdots 1_{|X_n| \leq K}}) \leq \frac{1}{n} \log \mathbb{E}(e^{n\lambda X_n 1_{|X_n| \leq K}})
\]
since
\[
\{|X_1| \leq K\} \cap \cdots \cap \{|X_n| \leq K\} \subset \{|X_n| \leq K\}
\]
where \( \{X \leq x\} \) denotes the event that the random variable \( X \) is at most \( x \). Then, writing \( \exp(n\lambda X_n) \) as an integral, we get
\[
\log \mathbb{E}(e^{\lambda X_1 1_{|X_1| \leq K}}) \leq \frac{1}{n} \log \mathbb{E}\left(e^{-n\lambda K} + \int_{-K}^{\infty} n\lambda e^{n\lambda u} du 1_{|X_n| \leq K}\right) \leq \frac{1}{n} \log \left(e^{-n\lambda K} + \int_{-\infty}^{+\infty} \mathbb{E}(1_{-K \leq u \leq X_n} 1_{|X_n| \leq K} n\lambda e^{n\lambda u} du)\right),
\]
the last step being a consequence of Fubini’s theorem. Since
\[
\mathbb{E}(1_{-K \leq u \leq X_n} 1_{|X_n| \leq K}) \leq \mathbb{E}(1_{X_n \geq u} 1_{-K \leq u \leq K}) = \mathbb{P}(X_n \geq u) 1_{|u| \leq K} \leq e^{n s(u)} 1_{|u| \leq K},
\]
...
we get
\[
\log E \left( e^{\lambda X_1 1_{X_1 \leq K}} \right) \leq \frac{1}{n} \log \left( e^{-n\lambda K} + \int_{-K}^{K} n\lambda e^{n(\lambda u + s(u))} du \right) \\
\leq \frac{1}{n} \log \left( e^{-n\lambda K} + 2Kn\lambda \exp \left( n \sup_{u \in \mathbb{R}} (\lambda u + s(u)) \right) \right).
\]

Let \( K \) be large enough so that (recall that the supremum of \( s \) is 0, whence there is some \( u \in \mathbb{R} \) such that \( s(u) > -\infty \))
\[-\lambda K < \sup_{u \in \mathbb{R}} (\lambda u + s(u)).\]

Then
\[
\log E \left( e^{\lambda X_1 1_{X_1 \leq K}} \right) \leq \frac{1}{n} \log \left( (1 + 2Kn\lambda) \exp \left( n \sup_{u \in \mathbb{R}} (\lambda u + s(u)) \right) \right) \\
\leq \frac{1}{n} \log(1 + 2Kn\lambda) + \sup_{u \in \mathbb{R}} (\lambda u + s(u)).
\]

Sending \( n \) to \( \infty \) we obtain
\[
\log E \left( e^{\lambda X_1 1_{X_1 \leq K}} \right) \leq \sup_{u \in \mathbb{R}} (\lambda u + s(u)).
\]

Finally, sending \( K \) to \(+\infty\) we get
\[
p(\lambda) \leq \sup_{u \in \mathbb{R}} (\lambda u + s(u)). \quad \Box
\]

To deduce Cramér’s theorem from the dual equality, we need to prove that the function \( s \) is concave.

**Convergence.** For all \( x \in \mathbb{R} \),
\[
\frac{1}{n} \log P(X_n \geq x)
\]
converges in \( [-\infty, 0] \) towards \( s(x) \).

**Proof.** Let \( x \in \mathbb{R} \). The proof relies on the subadditivity of the sequence \(-\log P(X_n \geq x)\). Suppose that \( P(X_1 \geq x) > 0 \) and let \( m \geq 1 \). For \( n \geq m \), let \( n = mq_n + r_n \) be the Euclidean division of \( n \) by \( m \). Then
\[
\{X_n \geq x\} \supset \bigcap_{k=0}^{q_n-1} \left\{ \frac{1}{m} \sum_{i=mk+1}^{m(k+1)} X_i \geq x \right\} \cap \bigcap_{i=mq_n+1}^{n} \{X_i \geq x\}
\]
since $\overline{X}_n$ is a weighted mean of the random variables appearing in the events of the right-hand side. Since the $X_i$ are i.i.d., for all $k \in \{0, \ldots, q_n - 1\}$,

$$\mathbb{P}\left(\frac{1}{m} \sum_{i=mk+1}^{mk+1} X_i \geq x\right) = \mathbb{P}(\overline{X}_m \geq x)$$

and, for all $i \in \{mq_n + 1, \ldots, n\}$, $\mathbb{P}(X_i \geq x) = \mathbb{P}(X_1 \geq x)$. Therefore

$$\mathbb{P}(\overline{X}_n \geq x) \geq \mathbb{P}(\overline{X}_m \geq x)^{q_n} \mathbb{P}(X_1 \geq x)^r_n \geq \mathbb{P}(\overline{X}_m \geq x)^{q_n} \mathbb{P}(X_1 \geq x)^r_n.$$

Now, taking logarithms, dividing by $n$, sending $n$ to $\infty$, and remembering that $q_n/n \rightarrow 1/m$, we get

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \geq x) \geq \frac{1}{m} \log \mathbb{P}(\overline{X}_m \geq x).$$

Taking the supremum over $m \geq 1$, we conclude that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \geq x) = s(x).$$

If $x$ is such that $\mathbb{P}(X_1 \geq x) = 0$, then

$$\forall n \geq 1 \quad \frac{1}{n} \log \mathbb{P}(\overline{X}_n \geq x) = -\infty$$

whence the sequence converges towards $-\infty = s(x)$. \hfill \Box

**Concavity.** The function $s : \mathbb{R} \rightarrow [-\infty, 0]$ is concave.

**Proof.** For $x, y \in \mathbb{R}$ and $n \geq 1$,

$$\{\overline{X}_{2n} \geq \frac{1}{2}(x + y)\} \supseteq \left\{\frac{1}{n} \sum_{i=1}^{n} X_i \geq x\right\} \cap \left\{\frac{1}{2n} \sum_{i=n+1}^{2n} X_i \geq y\right\}$$

so that, taking the logarithm of the probability of each side, dividing by $n$, and sending $n$ to $\infty$, we get

$$s\left(\frac{1}{2}(x + y)\right) \geq \frac{1}{2} (s(x) + s(y)).$$

Iterating, we have for every dyadic rational $\alpha \in [0, 1]$

$$s(\alpha x + (1 - \alpha)y) \geq \alpha s(x) + (1 - \alpha)s(y).$$

Since $s$ is nonincreasing, we conclude that $s$ is concave. \hfill \Box

We now finish the proof of Cramér’s theorem. At this point we know that for all $x \in \mathbb{R}$

$$\inf_{\lambda \geq 0} (p(\lambda) - \lambda x) = \inf_{\lambda \geq 0} \sup_{u \in \mathbb{R}} (\lambda(u - x) + s(u)).$$
It remains to prove that the latter quantity equals $s(x)$. The result is standard in convex function theory and known as Fenchel-Legendre duality. Let us give an elementary proof in our setting. The right-hand side of the previous equation is clearly greater than or equal to $s(x)$: take $u = x$. To prove the converse inequality we set

$$c = \inf \{ x \in \mathbb{R} : \mathbb{P}(X_1 \geq x) = 0 \}$$

and we distinguish the two cases $x < c$ and $x \geq c$.

• Suppose $x < c$. So $s(x) > -\infty$ (remember the useful inequality). Since $s$ is concave and nonincreasing let

$$-\lambda = \lim_{u \to x^-} \frac{s(u) - s(x)}{u - x} \leq 0$$

be the left derivative of $s$ at the point $x$. Then, by concavity,

$$\forall u \in \mathbb{R} \quad s(u) \leq s(x) - \lambda(u - x)$$

from which the result follows.

• Suppose $x \geq c$. Note that

$$\mathbb{P}(X_1 > c) = \lim_{\varepsilon \to 0^+} \mathbb{P}(X_1 \geq c + \varepsilon) = 0.$$

Then, for all $\lambda \geq 0$ and $\varepsilon > 0$,

$$p(\lambda) - \lambda x = \log \mathbb{E} \left( e^{\lambda(X_1-x)} \left(1_{X_1 < x - \varepsilon} + 1_{x - \varepsilon \leq X_1 \leq c} \right) \right)$$

$$\leq \log \left( e^{-\lambda \varepsilon} + \mathbb{P}(X_1 \geq x - \varepsilon) \right).$$

Taking the infimum over $\lambda \geq 0$ and sending $\varepsilon$ to 0 we get

$$\inf_{\lambda \geq 0} (p(\lambda) - \lambda x) \leq \log \mathbb{P}(X_1 \geq x) \leq s(x). \qed$$

Acknowledgments. We would like to thank Cécile Dubois and Yann Fuchs for their careful reading, and the referees for their remarks.


Mathématique, bâtiment 425, Université Paris Sud, 91405 Orsay Cedex
rcerf@math.u-psud.fr
pierre.petit@normalesup.org