Cramér's theorem

Jan M. Swart (Czech Academy of Sciences)

Wednesday, April 28th, 2021

Cramér's theorem

Let $(X_k)_{k\geq 1}$ be i.i.d. real random variables with law μ.
Assume that

$$Z(\lambda) := \mathbb{E}[e^{\lambda X_1}] = \int \mu(\mathrm{d}x)\, e^{\lambda x} < \infty \qquad (\lambda \in \mathbb{R}).$$

and set $\rho := \langle \mu \rangle = \mathbb{E}[X_1]$.

Theorem (Harald Cramér, 1938)

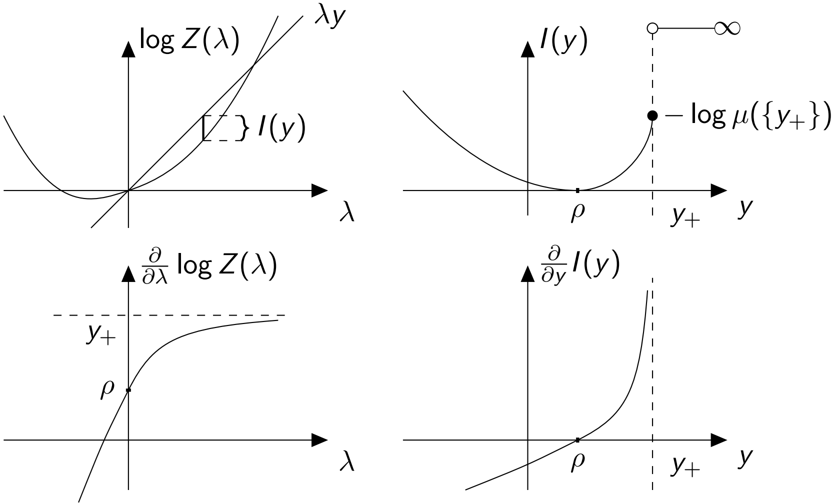
$$\text{(i)} \quad \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big[\frac{1}{n}\sum_{k=1}^{n} X_k \geq y\Big] = -I(y) \qquad (y > \rho),$$

$$\text{(ii)} \quad \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big[\frac{1}{n}\sum_{k=1}^{n} X_k \leq y\Big] = -I(y) \qquad (y < \rho),$$

where I is defined by

$$I(y) := \sup_{\lambda\in\mathbb{R}} \big[\lambda y - \log Z(\lambda)\big] \qquad (y \in \mathbb{R}).$$

Cramér's theorem



The upper bound

Proof We only prove (i). Since $1_{[0,\infty)}(z) \leq e^z$, we can estimate for $y \in \mathbb{R}$ and $\lambda > 0$,

$$\begin{aligned}
\mathbb{P}\Big[\frac{1}{n}\sum_{k=1}^n X_k \geq y\Big] &= \mathbb{P}\Big[\frac{1}{n}\sum_{k=1}^n (X_k - y) \geq 0\Big] = \mathbb{P}\Big[\lambda\sum_{k=1}^n (X_k - y) \geq 0\Big] \\
&\leq \mathbb{E}\big[e^{\lambda\sum_{k=1}^n (X_k - y)}\big] = \prod_{k=1}^n \mathbb{E}\big[e^{\lambda(X_k - y)}\big] = e^{-n\lambda y}\mathbb{E}\big[e^{\lambda X_1}\big]^n \\
&= e^{(\log Z(\lambda) - \lambda y)n}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\frac{1}{n}\log\mathbb{P}\Big[\frac{1}{n}\sum_{k=1}^n X_k \leq y\Big] &\leq \inf_{\lambda>0}\big[\log Z(\lambda) - \lambda y\big] \\
&= -\sup_{\lambda>0}\big[\lambda y - \log Z(\lambda)\big].
\end{aligned}$$

The upper bound

The function $\lambda \mapsto \lambda y - \log Z(\lambda)$ assumes its maximum in the point λ that is uniquely characterised by

$$\frac{\partial}{\partial\lambda} \log Z(\lambda) = y.$$

If $y > \rho$, then the maximum is achieved for $\lambda > 0$ and hence

$$\sup_{\lambda > 0} \big[\lambda y - \log Z(\lambda)\big] = \sup_{\lambda \in \mathbb{R}} \big[\lambda y - \log Z(\lambda)\big] =: I(y).$$

This completes the proof of the upper bound.

The lower bound

To prove the lower bound, we first treat two trivial cases. Recall that y_+ is the supremum of the support of the law of X_1.

If $y > y_+$, then $\mathbb{P}[\frac{1}{n}\sum_{k=1}^n X_k \geq y] = 0$ for all $n \geq 1$ while $I(y) = \infty$ so (i) holds trivially.

If $y = y_+$, then $\mathbb{P}[\frac{1}{n}\sum_{k=1}^n X_k \geq y] = \mathbb{P}[X_1 = y_+]^n$ while $I(y_+) = -\log\mathbb{P}[X_1 = y_+]$ by, hence again (i) holds.

It remains to treat the case $\rho < y < y_+$.

The lower bound

Now $I(y) = y\lambda_\circ - \log Z(\lambda_\circ)$, where is uniquely characterized by the requirement that

$$\langle \mu_{\lambda_\circ} \rangle = \partial \log Z(\lambda_\circ) = y.$$

The idea of the proof is to replace the law μ of the $(X_k)_{k\geq 1}$ by $\mu_{\lambda_\circ}$ at an exponential cost of size $I(y)$.

$$\begin{aligned}
&\mathbb{P}\Big[\frac{1}{n}\sum_{k=1}^{n} X_k \geq y\Big] = \mathbb{P}\Big[\sum_{k=1}^{n}(X_k - y) \geq 0\Big] \\
&= \int \mu(\mathrm{d}x_1)\cdots\int \mu(\mathrm{d}x_n) 1_{\{\sum_{k=1}^{n}(x_k - y) \geq 0\}} \\
&= Z(\lambda_\circ)^n e^{-n\lambda_\circ y} \int \mu_{\lambda_\circ}(\mathrm{d}x_1)\cdots\int \mu_{\lambda_\circ}(\mathrm{d}x_n) \\
&\qquad \times e^{-\lambda_\circ \sum_{k=1}^{n}(x_k - y)} 1_{\{\sum_{k=1}^{n}(x_k - y) \geq 0\}} \\
&= e^{-nI(y)} \mathbb{E}\Big[e^{-\lambda_\circ \sum_{k=1}^{n}(\hat{X}_k - y)} 1_{\{\sum_{k=1}^{n}(\hat{X}_k - y) \geq 0\}}\Big].
\end{aligned}$$

The lower bound

To complete the proof, we must show that

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\big[e^{-\lambda_\circ\sum_{k=1}^n(\hat X_k-y)}1_{\{\sum_{k=1}^n(\hat X_k-y)\geq 0\}}\big]\geq 0,$$

where $(\hat X_k)_{k\geq 1}$ are i.i.d. with law $\mu_{\lambda_\circ}$, so $\mathbb{E}[\hat X_k]=y$ and $\sigma^2:=\mathrm{Var}(\hat X_k)>0$.
We estimate

$$\begin{aligned}&\mathbb{E}\big[e^{-\lambda_\circ\sum_{k=1}^n(\hat X_k-y)}1_{\{\sum_{k=1}^n(\hat X_k-y)\geq 0\}}\big]\\ &\quad\geq\mathbb{P}\Big[0\leq\sum_{k=1}^n(\hat X_k-y)\leq\sigma\sqrt{n}\Big]e^{-\sigma\lambda_\circ\sqrt{n}},\end{aligned}$$

where by the central limit theorem,

$$\mathbb{P}\Big[0\leq\sum_{k=1}^n(\hat X_k-y)\leq\sigma\sqrt{n}\Big]\underset{n\to\infty}{\longrightarrow}\frac{1}{\sqrt{2\pi}}\int_0^1 e^{-z^2/2}\mathrm{d}z=:\theta>0.$$

Use of the CLT

Remark Our proof shows that for $\rho < y < y_+$,

$$e^{-nI(y) - O(\sqrt{n})} \leq \mathbb{P}[\frac{1}{n}\sum_{k=1}^{n} X_k \geq y] \leq e^{-nI(y)} \qquad \text{as } n \to \infty,$$

where the lower bound is based on the central limit theorem. Alternatively, one can use the weak law of large numbers to deduce that

$$\mathbb{P}\big[0 \leq \sum_{k=1}^{n} (\tilde{X}_k - y) \leq \varepsilon\big] \underset{n\to\infty}{\longrightarrow} 1 \qquad (\varepsilon > 0)$$

when $\tilde{X}_k$ have law μ_λ with $\lambda > \lambda_\circ$.
Since $\varepsilon > 0$ and $\lambda > \lambda_\circ$ are arbitrary, this is enough to derive the LDP.

Upper limit of the range

Remark We have proved a bit more than the LDP.
If $y_+ < \infty$ and $\mu(\{y_+\}) > 0$, then the LDP tells us that

$$\limsup_{n\to\infty} \mu_n([y_+, \infty)) \leq - \inf_{y\in[y_+,\infty)} I(y) = -I(y_+),$$

but as shown in Excercise 1.11, the complementary statement for the limit inferior does not follow from the LDP since $[y_+, \infty)$ is not an open set.

The multi-dimensional case

Remark The multi-dimensional Cramér's theorem can be proved in the same way, but we will give a different proof based on the Gärtner-Ellis theorem.

Remark Cramér's theorem remains true if the assumption $Z(\lambda) := \mathbb{E}[e^{\lambda X_1}] < \infty$ for all $\lambda \in \mathbb{R}$ is weakened to $Z(\lambda) < \infty$ for λ in an open neighbourhood of the origin. Our proof in the multi-dimensional case will partly cover this regime but not fully.

Conditioning on rare events

Remark For $\rho < y < y_+$, it can be shown that for fixed $m \geq 1$,

$$\mathbb{P}\big[X_1 \in \mathrm{d}x_1, \ldots, X_m \in \mathrm{d}x_m \,\big|\, \tfrac{1}{n}\sum_{k=1}^{n} X_k \geq y\big] \underset{n\to\infty}{\Longrightarrow} \mu_{\lambda_\circ}(\mathrm{d}x_1)\cdots\mu_{\lambda_\circ}(\mathrm{d}x_m).$$

This means that conditioned on the rare event $\frac{1}{n}\sum_{k=1}^{n} X_k \geq y$, in the limit $n \to \infty$, the random variables $X_1, \ldots, X_n$ are approximately distributed as if they are i.i.d. with common law $\mu_{\lambda_\circ}$.