Démonstration. Afin de simplifier les écritures, on notera dans ce qui suit Λ et Λ^* pour Λ_μ et Λ^*_μ.
1) Montrons que $h(x) \geq \Lambda^*(x)$. Par l'inégalité de Chebychev , on a pour tout $\lambda \geq 0$,

$$\mathbb{P}(S_n \geq xn) \leq \mathbb{P}(e^{\lambda S_n} \geq e^{\lambda nx}) \leq e^{-\lambda nx}\mathbb{E}[e^{\lambda S_n}] = e^{n\Lambda(\lambda)-\lambda nx}$$

et donc

$$\frac{1}{n}\log \mathbb{P}(S_n \geq nx) \leq \Lambda(\lambda) - \lambda x. \tag{14.1}$$

Ainsi, $\forall \lambda \geq 0 \quad \forall x \in \mathbb{R} \quad \Lambda(\lambda) + h(x) \geq \lambda x$, ce qui nous donne

$$\forall \lambda \geq 0 \quad \Lambda(\lambda) \geq h^*(\lambda) = \sup_{x\in\mathbb{R}} \left(\lambda x - h(x)\right).$$

Si $x \geq \mathbb{E}X_1$, l'inégalité (14.1) demeure valide pour $\lambda \leq 0$, car on a alors $\Lambda(\lambda) - \lambda x \geq 0$ pour $x \geq \mathbb{E}X_1$ (grâce à l'inégalité de Jensen). Ainsi

$$\forall \lambda \in \mathbb{R} \quad \forall x \geq \mathbb{E}X_1 \quad \Lambda(\lambda) + h(x) \geq \lambda x,$$

d'où l'on tire

$$h(x) \geq \Lambda^*(x) \text{ pour } x \geq \mathbb{E}[X_1].$$

2) On va commencer par montrer que $\Lambda(\lambda) \leq h^*(\lambda)$ pour tout $\lambda \geq 0$. Traitons à part le cas $\lambda = 0$: on a

$$h^*(0) = \sup_{x\in\mathbb{R}} -h(x) \geq \sup_{x\in\mathbb{R}} -\log \mathbb{P}(S_1 \geq x) = 0 = \Lambda(0).$$

Supposons maintenant $\lambda > 0$. On a

$$\begin{aligned}
\Lambda(\lambda) &= \log \mathbb{E}[e^{\lambda X}] = \overline{\lim_{M\to+\infty}} \log \mathbb{E}[e^{\lambda X} \mathbb{1}_{\{|X|\leq M\}}] \\
&\leq \overline{\lim_{M\to+\infty}} \inf_{n\geq 1} \frac{1}{n} \log \mathbb{E}[e^{\lambda S_n} \mathbb{1}_{\{|S_n|\leq nM\}}].
\end{aligned}$$

On peut écrire

$$\begin{aligned}
&\mathbb{E}[e^{\lambda S_n}\mathbb{1}_{\{|S_n|\leq nM\}}] = \int_{[-nM,nM]} e^{\lambda u} d\mathbb{P}_{S_n}(u) \\
=\ & \int_{[-nM,nM]} \left(e^{-\lambda nM} + \int_{[-nM,nM]} \mathbb{1}_{\{x\leq u\}} \lambda e^{\lambda x}\ d\lambda(x)\right) d\mathbb{P}_{S_n}(u) \\
\leq\ & e^{-\lambda nM} + \int_{[-nM,nM]} \lambda e^{\lambda x} \mathbb{P}(S_n \geq x)\ d\lambda(x) \\
&= e^{-\lambda nM} + \int_{[-M,M]} n\lambda e^{\lambda nx} \mathbb{P}(S_n \geq nx)\ d\lambda(x) \\
\leq\ & e^{-\lambda nM} + \int_{[-M,M]} n\lambda e^{\lambda nx} e^{-nh(x)}\ d\lambda(x) \\
&\leq e^{-\lambda nM} + 2Mn\lambda e^{nh^*(\lambda)}
\end{aligned}$$

Comme $h^*(0) = \sup_{x\in\mathbb{R}} -h(x) \geq 0$, il existe $x \in \mathbb{R}$ avec $h(x) < +\infty$, ce qui entraîne que $h^*(\lambda) > -\infty$. On peut donc choisir M tel que $-M\lambda \leq h^*(\lambda)$. On a alors pour tout $n \geq 1$

$$\log \mathbb{E}[e^{\lambda S_n} \mathbb{1}_{\{|S_n|\leq nM\}}] \leq \log(2Mn\lambda + 1) + nh^*(\lambda),$$

d'où

$$\inf_{n\geq 1} \frac{1}{n} \log \mathbb{E}[e^{\lambda S_n} \mathbb{1}_{\{|S_n|\leq nM\}}] \leq h^*(\lambda),$$

et finalement $\Lambda(\lambda) \leq h^*(\lambda)$.

3) Pour conclure, il nous reste à montrer que $h(x) \leq \Lambda^*(x)$.
Notons $c = \inf\{x \in \mathbb{R} : \mathbb{P}(X_1 \geq x) = 0\}$.

— si $x < c$. Il nous suffit de montrer qu'il existe λ réel avec $h(x) \leq \lambda x - \Lambda(\lambda)$. Or, comme h est convexe croissante, finie sur $]-\infty, c[$, il existe λ réel positif tel que pour tout y, $h(y) \geq h(x) + \lambda(y - x)$, soit $\lambda x - h(x) \geq \lambda y - h(y)$. En prenant le supremum en y, on obtient $\lambda x - h(x) \geq h^*(\lambda)$, mais comme on a montré que $h^*(\lambda) = \Lambda(\lambda)$ pour $\lambda \geq 0$, cela achève la preuve dans ce cas.

— si $x > c$, comme $\Lambda(\lambda) \leq c\lambda$ pour tout $\lambda \geq 0$, on a

$\Lambda^*(x) \geq \sup_{\lambda\in\mathbb{R}_+} (x - c)\lambda = +\infty$, et l'inégalité est évidente.

— si $x = c$, il est facile de voir que $\mathbb{P}(S_n \geq nc) = \mathbb{P}(X_1 = c)^n$, d'où $h(c) = -\log \mathbb{P}(X_1 = c)$. Par ailleurs, pour $\varepsilon, \lambda > 0$, on a

$$\mathbb{E}(e^{\lambda X_1}) \leq e^{\lambda(c-\varepsilon)} + e^{\lambda c}\mathbb{P}(c - \varepsilon < X_1 \leq c),$$

d'où $\lambda c - \Lambda(\lambda) \geq -\log(e^{-\lambda\varepsilon} + \mathbb{P}(c - \varepsilon < X_1 \leq c))$ et

$$\begin{aligned}\Lambda^*(c) \geq \overline{\lim_{\lambda\to+\infty}}\ \lambda c - \Lambda(\lambda) &\geq \overline{\lim_{\lambda\to+\infty}} -\log(e^{-\lambda\varepsilon} + \mathbb{P}(c - \varepsilon < X_1 \leq c)) \\ &= -\log(\mathbb{P}(c - \varepsilon < X_1 \leq c)).\end{aligned}$$

En faisant tendre ε vers 0, on obtient donc $\Lambda^*(c) \geq h(c)$.

$\square$

Le résultat utilisé dans la pratique est le suivant.

Corollaire 14.3.4 (inégalité de Chernov)**.** *Soit $(X_n, n \geq 1)$ une suite de variables aléatoires indépendantes de même loi, intégrable* ~~*et $A = ess \sup X_1$*~~ *. Alors, pour tout $x > \mathbb{E}(X_1)$,*

$$\mathbb{P}(X_1 + \cdots + X_n \geq xn) \leq e^{-n\Lambda^*(x)}. \tag{14.2}$$

On a $\Lambda^*(x) > 0$ dès qu'il existe $\varepsilon > 0$ tel que $\Lambda(\varepsilon) < +\infty$.
~~*Cette décroissance exponentielle (14.2) s'étend à $x = A$ si et seulement si $\mathbb{P}(X_1 = A) < 1$.*~~

Démonstration. (14.2) vient de l'identité $h(x) = \Lambda^*(x)$ pour $x \geq \mathbb{E}(X_1)$ et de la définition de h. En appliquant le théorème de convergence monotone à la partie négative de X_1, on a $\lim_{N\to+\infty} \mathbb{E}(\max(-N, X_1)) = \mathbb{E}(X_1)$, donc on peut trouver N tel que $x > \mathbb{E}(\max(-N, X_1))$. Notons ν la loi de $\max(-N, X_1)$ et prenons a avec $x > a > \mathbb{E}\max(-N, X_1)$. Pour tout $\lambda > 0$, on a

$$\Lambda^*(x) \geq -\Lambda(\lambda) + \lambda x \geq -\Lambda_\nu(\lambda) + \lambda x,$$

donc pour avoir $\Lambda^*(x) > 0$, il suffit de trouver $\lambda > 0$ tel que $\Lambda_\nu(\lambda) < \lambda a$. Posons $\phi(\lambda) = \int e^{\lambda t}\ d\nu(t)$. Pour $\lambda \in]0, \varepsilon/2[$, on a

$$|\frac{e^{\lambda t}-1}{\lambda}| \leq |t| \max(1, e^{\lambda t}) \leq N + \frac{2}{\varepsilon} e^{\varepsilon t} \text{ pour } \nu \text{ presque tout } t,$$

d'où, avec le théorème de convergence dominée : $\phi(\lambda) = 1 + \lambda \int t\ d\nu(t) + o(\lambda)$, lorsque λ tend vers 0 par valeurs supérieures. Finalement

$$\Lambda_\nu(\lambda) = \lambda \mathbb{E}(\max(-N, X_1)) + o(\lambda),$$

ce qui donne le résultat voulu en prenant $\lambda > 0$ suffisamment petit.

□

Remarque. Les estimées de grandes déviations s'étendent aux intervalles de la forme $[x, +\infty[$, aux intervalles ouverts, fermés ou semi-ouverts. Il s'agit du *principe de grandes déviations*, prouvé par Cramér pour une suite de variables indépendantes de même loi sur $\mathbb{R}$. Nous ne l'énonçons pas ici, car il s'agit d'un résultat compliqué, qui a déjà fait l'objet de divers livres (pas toujours faciles d'accès), comme celui de Varadhan.

Exemple. Voyons maintenant sur un exemple comment utiliser ces résultats. Étudions le cas simple où les variables $(X_n)_{n\geq 1}$ sont indépendantes de loi de Bernoulli de paramètre p. Dans ce cas, S_n correspond au nombre de "pile" obtenu lors de n lancers indépendants d'une pièce de monnaie.

Le théorème précédent implique le résultat suivant : pour tout $m \in [p, 1[$, on a

$$\begin{aligned}
\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(S_n \geq nm) &= \sup_{s\geq 0} (m \log s - \log G_X(s)) \\
&= m \log \frac{m(1-p)}{p(1-m)} - \log\left(1 - \left(1 - \frac{m(1-p)}{p(1-m)}\right) p\right),
\end{aligned}$$

où $G_X(s) = 1 - (1-s)p$ est la fonction génératrice de la loi de Bernoulli de paramètre $0 < p < 1$ (car pour une loi de Bernoulli, $\Lambda(\lambda) = \log G_X(e^\lambda)$).