

Appendice 1: Grandes déviations

Gérard Letac¹

¹Professeur émérite, Institut de Mathématiques de Toulouse., Toulouse

3 juillet 2022



Théorème des grandes déviations

1 Appendice 1 : Grandes déviations

1.1 Le théorème des grandes déviations

Si X_1, \dots, X_n sont des variables aléatoires indépendantes et de même loi, de moyenne m et telles qu'il existe $\alpha > 0$ avec $\mathbb{E}(e^{\alpha|X_n|}) < \infty$, et si on note $S_n = X_1 + \dots + X_n$, le théorème suivant calcule pour $a > m$ le nombre $\lim_{n \rightarrow \infty} (\Pr(S_n \geq na))^{1/n}$.

THÉORÈME 0.1 Inégalité des grandes déviations

Soit μ une mesure positive sur \mathbb{R} non concentrée en un point et telle que l'intervalle des θ réels satisfaisant $L(\theta) = \int_{-\infty}^{\infty} e^{\theta x} \mu(dx) < \infty$ ait un intérieur Θ non vide. On considère la fonction strictement convexe sur Θ égale à $k = \log L$ et l'intervalle ouvert $M = k'(\Theta)$, et on note par $\psi : M \rightarrow \Theta$ la fonction réciproque de k' .

Soit $m = k'(\theta)$ fixé dans M . Soient X_1, \dots, X_n des variables aléatoires indépendantes et de même loi $e^{\theta x - k(\theta)} \mu(dx)$. Soit enfin $a \in M$ avec $m < a$ et les nombres

$$u_n = \Pr\left(\frac{1}{n}(X_1 + \dots + X_n) \geq a\right) h(m, a) = - \int_m^a (a-x) \psi'(x) dx = a(\psi(m) - \psi(a)) + k(\psi(a)) - k(\psi(m)).$$

Dans ces conditions on a

1. (Inégalité des grandes déviations) $u_n^{1/n} \leq e^{h(m,a)}$.
2. (Théorème des grandes déviations) $\lim_{n \rightarrow \infty} u_n^{1/n} = e^{h(m,a)}$.

Commentaires : 1) Une insupportable confusion règne dans la littérature d'enseignement concernant ce résultat, dû à Cramer (1938), principalement à cause de ses généralisations à des hypothèses plus faibles (et peu intéressantes) dans \mathbb{R} ainsi qu'à \mathbb{R}^d , où les résultats n'ont pas l'harmonie du résultat ci dessus.

2) Dans sa présentation, le théorème fait jouer un rôle symétrique à toute la famille de lois de probabilités $e^{\theta x - k(\theta)} \mu(dx)$ quand θ varie dans Θ . Cette famille est appelée une famille exponentielle naturelle engendrée par μ . Attention, μ n'est pas unique : μ' engendre la même famille exponentielle, c'est à dire le même *ensemble* de probabilités, indépendamment du paramétrage, si et seulement si il existe a et b réels tels que $\mu'(dx) = e^{ax+b} \mu(dx)$. Il est clair que la loi d'une variable aléatoire réelle X telle qu'il existe $\alpha > 0$ avec $\mathbb{E}(e^{\alpha|X|}) < \infty$ appartient à une famille exponentielle naturelle : il suffit de prendre pour μ la loi de X . Toutefois, pour la loi de X donnée, souvent avec un paramètre, il n'est pas toujours apparent de relier cette loi avec la famille exponentielle à laquelle elle appartient. Par exemple la loi de Bernoulli $(1-p)\delta_0 + p\delta_1$ appartient à la famille exponentielle engendrée par $\mu = \delta_0 + \delta_1$: prendre $p = \frac{e^\theta}{1+e^\theta}$.

3) Implicitement, l'énoncé utilise des résultats simples comme le fait que Θ soit un intervalle et comme la convexité de k , qui se démontrent comme le 1) et le 6) du théorème du cours de Deug. De plus, il est facile de voir que avec les notations du théorème, l'espérance des X_i est $m = k'(\theta)$ et leur variance est $k''(\theta) = 1/\psi'(m)$.

4) La partie 2) du théorème est plus difficile. La partie 1) est comme on va le voir amusante et élémentaire. Elle fournit une démonstration de poche de la loi forte des grands nombres qui affirme que si X_1, \dots, X_n, \dots sont des variables aléatoires indépendantes et de même loi de moyenne m , alors $\Pr(\lim_{n \rightarrow \infty} S_n/n = m) = 1$. Si on fait l'hypothèse supplémentaire de l'existence de moments exponentiels, c'est à dire qu'il existe $\alpha > 0$ avec $\mathbb{E}(e^{\alpha|X_n|}) < \infty$, alors l'inégalité des grandes déviations et le critère de Cauchy, du fait que $h(m, a) < 0$, entraîne que la série $\sum u_n$ converge, et on procède alors comme au Théorème du cours pour conclure avec le lemme de Borel.

5) Travaux pratiques : Voici quelques mesures μ classiques, et les lois et les fonctions $h(m, a)$ qui vont avec.

Loi de Bernoulli : $\mu = \delta_0 + \delta_1$, $L(\theta) = 1 + e^\theta$, $\Theta = \mathbb{R}$, $k(\theta) = \log(1 + e^\theta)$, $k'(\theta) = \frac{e^\theta}{1+e^\theta}$, $M =]0, 1[$, $\psi(m) = \log \frac{m}{1-m}$, $k(\psi(m)) = -\log(1 - m)$ et

$$h(m, a) = a \log \frac{a}{m} + (1 - a) \log \frac{1 - a}{1 - m}.$$

Loi de Poisson : $\mu = \sum_{n=0}^{\infty} \frac{1}{n!} \delta_n$, $L(\theta) = \exp e^\theta$, $\Theta = \mathbb{R}$, $k(\theta) = e^\theta$, $k'(\theta) = e^\theta$, $M =]0, \infty[$, $\psi(m) = \log m$, $k(\psi(m)) = m$ et

$$h(m, a) = a \log \frac{m}{a} + a - m.$$

Loi gamma : Soit $\alpha > 0$ fixé. $\mu = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \mathbf{1}_{]0, \infty[}(x) dx$, $L(\theta) = \frac{1}{(-\theta)^\alpha}$ si $\theta \in \Theta =]-\infty, 0[$, $k(\theta) = \alpha \log(-\theta)$, $k'(\theta) = \frac{\alpha}{-\theta}$, $M =]0, \infty[$, $\psi(m) = \frac{\alpha}{-m}$, $k(\psi(m)) = \alpha \log \frac{m}{\alpha}$ et

$$h(m, a) = \alpha - \alpha \frac{a}{m} + \alpha \log \frac{a}{m}.$$

Loi normale : Soit $\sigma > 0$ fixé. $\mu = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{x^2}{2\sigma^2}) dx$, $L(\theta) = \exp(\frac{\sigma^2\theta^2}{2})$, $\Theta = \mathbb{R}$, $k(\theta) = \frac{\sigma^2\theta^2}{2}$, $k'(\theta) = \sigma^2\theta$, $M = \mathbb{R}$, $\psi(m) = \frac{m}{\sigma^2}$, $k(\psi(m)) = \frac{m^2}{2\sigma^2}$ et

$$h(m, a) = -\frac{(a - m)^2}{2\sigma^2}.$$

Démonstration de l'inégalité des grandes déviations : Notons $S_n = X_1 + \dots + X_n$. Pour tout $t > 0$ tel que $\theta + t \in \Theta$ l'astuce est d'observer que les deux évènements $\{S_n/n \geq a\}$ et $\{e^{tS_n} \geq e^{nta}\}$ sont les mêmes (comme à la Prop.). On écrit, à l'aide de l'inégalité de Markov (voir cours de Deug, Prop.) appliquée à $Y = e^{tS_n}$ et à $y = e^{nta}$:

$$u_n = \Pr(e^{tS_n} \geq e^{nta}) \leq e^{-nta} \mathbb{E}(e^{tS_n}) = [e^{-ta} \mathbb{E}(e^{tX_1})]^n = [e^{-ta} \frac{L(\theta+t)}{L(\theta)}]^n.$$

Donc $u_n^{1/n} \leq e^{-ta+k(\theta+t)-k(\theta)}$. Observons ensuite que $t \mapsto -ta+k(\theta+t)-k(\theta)$ est convexe sur l'intervalle ouvert $(-\theta + \Theta) \cap]0, \infty[$ et que sa dérivée s'y annule au point $t = \psi(a) - \psi(\theta)$, c'est à dire tel que $k'(\theta+t) = a$. La valeur de $-ta+k(\theta+t)-k(\theta)$ en ce point est exactement $h(m, a)$ et le résultat est montré.

Démonstration du théorème des grandes déviations : On pose désormais $\tau = \psi(a) > \theta$. Avec cette notation, on remarque que

$$h(m, a) = -a(\tau - \theta) + k(\tau) - k(\theta).$$

L'astuce de Harald Cramer ici est d'introduire les variables aléatoires Y_1, \dots, Y_n indépendantes et de même loi $e^{\tau x - k(\tau)} \mu(dx)$. Si on a lu le commentaire 2, on remarque que cette loi appartient à la même famille exponentielle naturelle que la loi des X_i . L'espérance de Y_i est $a = k'(\tau)$. On pose ensuite $U_n = Y_1 + \dots + Y_n$, $V_n = U_n - na$ et $v_n = \mathbb{E}[e^{-(\tau-\theta)V_n} \mathbf{1}_{V_n \geq 0}]$. L'espérance de V_n est zéro. On montre alors l'identité remarquable

$$u_n = e^{nh(m,a)} v_n.$$

Pour le voir, on introduit la mesure positive μ_n sur \mathbb{R} égale à la n ième puissance de convolution μ^{*n} , c'est à dire de transformée de Laplace $L(\theta)^n$. La loi de S_n est donc $e^{\theta s - nk(\theta)} \mu_n(ds)$, comme on le vérifie en calculant la transformée de Laplace de cette loi et en voyant qu'elle est égale à celle de S_n , soit

$$\mathbb{E}(e^{tS_n}) = \left(\frac{L(\theta+t)}{L(\theta)} \right)^n = e^{nk(\theta+t) - nk(\theta)}$$

pour tout $t \in -\theta + \Theta$. De même, la loi de U_n est $e^{\tau u - nk(\tau)} \mu_n(du)$. Par conséquent

$$\begin{aligned} e^{nh(m,a)} \mathbb{E}[e^{-(\tau-\theta)V_n} \mathbf{1}_{V_n \geq 0}] &= e^{n[h(m,a) + (\tau-\theta)a]} \mathbb{E}[e^{-(\tau-\theta)U_n} \mathbf{1}_{U_n \geq na}] \\ &= e^{n[k(\tau) - k(\theta)]} \int_{na}^{\infty} e^{-(\tau-\theta)u} e^{\tau u - nk(\tau)} \mu_n(du) = \int_{na}^{\infty} e^{\theta u - nk(\theta)} \mu_n(du) = \Pr(S_n \geq na) = u_n, \end{aligned}$$

et l'identité annoncée $u_n = e^{nh(m,a)} v_n$ est montrée. On peut remarquer qu'elle nous donne au passage une seconde démonstration, moins élémentaire, de la partie 1, puisque trivialement $v_n < 1$. Cette partie algébrique étant faite, pour voir que la limite de $u_n^{1/n}$ est $e^{h(m,a)}$, il suffit de montrer que $\lim_{n \rightarrow \infty} v_n^{1/n} = 1$. C'est la partie plus difficile.

Commençons par un lemme classique :

LEMME 0.2

Si f est une variable aléatoire positive alors l'ensemble des $s \in \mathbb{R}$ tels que $\mathbb{E}(f^s) < \infty$ est un intervalle I et $s \mapsto [\mathbb{E}(f^s)]^{1/s}$ est croissante sur $]0, \infty[\cap I$.

Démonstration du lemme : On pourrait utiliser une inégalité classique de Hölder. Utilisons plutôt ici l'outil familier de la convexité du logarithme de la transformée de Laplace. Soit $(1-p)\delta_0 + p\nu(df)$ la loi de f avec $\nu(df)$ probabilité sur $]0, \infty[$ et $0 < p \leq 1$ (tout est trivial si $p = 0$). Soit $\mu(dx)$ l'image de

$\nu(df)$ par $f \mapsto x = \log f$. Soit T la transformée de Laplace de μ , soit I son domaine de finitude et soit $t(s) = \log T(s)$. Sur I on a $\mathbb{E}(f^s) = pe^{t(s)}$. Enfin si $0 < s < s_1$ sont dans I , comme t est convexe on a

$$t(s) = t\left(\frac{s}{s_1}s_1 + \left(1 - \frac{s}{s_1}\right)0\right) \leq \frac{s}{s_1}t(s_1) + \left(1 - \frac{s}{s_1}\right)t(0).$$

Comme $t(0) = 0$ (car μ est une probabilité) on obtient que $\frac{1}{s}t(s) \leq \frac{1}{s_1}t(s_1)$. Comme

$$p^{\frac{1}{s}} = \exp\left(\frac{1}{s} \log p\right) \leq \exp\left(\frac{1}{s_1} \log p\right) = p^{\frac{1}{s_1}},$$

le lemme est montré.

Démonstration Achéons alors la démonstration du théorème. On pose

$$A_n = \{V_n \geq 0\}, \quad B_n = \{e^{-(\tau-\theta)V_n/n^{3/4}} \geq \frac{1}{2}\}.$$

On a alors

$$v_n^{n^{-3/4}} \stackrel{(1)}{\geq} \mathbb{E}[e^{-(\tau-\theta)V_n/n^{3/4}} \mathbf{1}_{V_n \geq 0}] \stackrel{(2)}{\geq} \frac{1}{2} \Pr(A_n \cap B_n) \stackrel{(3)}{\geq} \frac{1}{2} (\Pr(A_n) - \Pr(B_n)).$$

Dans cette chaîne d'inégalités, (1) vient du lemme appliqué à $f = e^{-(\tau-\theta)V_n} \mathbf{1}_{V_n \geq 0}$ et au couple $s_1 = 1$ et $s = 1/n^{3/4}$, (2) est l'inégalité de Markov appliquée à $Y = f^{n^{-3/4}}$ et $y = 1/2$, et (3) vient du fait que si A et B sont deux évènements alors $A \subset (A \cap B) \cup B$ et donc $\Pr(A \cap B) \geq \Pr(A) - \Pr(B)$. Faisons alors tendre n vers l'infini. D'après le théorème central limite, la loi de V_n/\sqrt{n} tend vers une loi normale centrée. On en déduit que $\Pr(A_n)$ tend vers $1/2$ et, puisque B_n est aussi $B_n = \left\{\frac{V_n}{\sqrt{n}} \geq \frac{\log 2}{(\tau-\theta)} n^{1/4}\right\}$, on en déduit que $\Pr(B_n)$ tend vers 0. Par conséquent, la limite inférieure de $v_n^{n^{-3/4}}$ est $\geq 1/4$. Mais $\liminf \frac{1}{n^{3/4}} \log v_n \geq -\log 4$ entraîne naturellement que $\liminf \frac{1}{n} \log v_n \geq 0$. Comme $\log v_n \leq 0$ la limite de $\frac{1}{n} \log v_n$ est bien 0 et le théorème des grandes déviations est démontré.

Références