Estimations des grands écarts

Arnaud Girand

17 juin 2012

Référence :
- [Les01], p. 16–20

Prérequis :
- formule de Stirling.

Proposition 1
Soit $p \in (0,1)$.
On considère une suite $(X_n)_{n\geq 1}$ de variables aléatoires i.i.d de loi $\mathcal{B}(p)$ et on pose :

$$\forall n \geq 1,\, S_n := \sum_{i=1}^{n} X_i \sim \mathcal{B}(n,p)$$

Pour $\varepsilon \in (0, 1-p)$ on définit également :

$$h_+(\varepsilon) := (p+\varepsilon)\ln\left(\frac{p+\varepsilon}{p}\right) + (1-p-\varepsilon)\ln\left(\frac{1-p-\varepsilon}{1-p}\right)$$

Alors :
(i) $h_+(\varepsilon) > 0$ *;*
(ii)

$$\forall n \geq 1, \quad \mathbb{P}\left(\frac{S_n}{n} \geq p+\varepsilon\right) \leq e^{-nh_+(\varepsilon)}$$

(iii) cette inégalité est optimale, i.e :

$$\frac{1}{n}\ln\left(\mathbb{P}\left(\frac{S_n}{n} \geq p+\varepsilon\right)\right) \xrightarrow[n\to\infty]{} -h_+(\varepsilon)$$

Démonstration :

(i) et (ii) Soit $t > 0$. Alors :

$$\begin{aligned}
\mathbb{P}\left(\frac{S_n}{n} \geq p+\varepsilon\right) &= \mathbb{P}\left(e^{t(S_n - np - n\varepsilon)} \geq 1\right) \\
&\leq \mathbb{E}(e^{t(S_n-np-n\varepsilon)}) \text{ par inégalité de Markov} \\
&= e^{-nt(p+\varepsilon)}\mathbb{E}(e^{tS_n}) \\
&= e^{-nt(p+\varepsilon)} \sum_{k=0}^{\infty} e^{tk}\mathbb{P}(S_n = k) \text{ par lemme de transfert} \\
&= e^{-nt(p+\varepsilon)} \sum_{k=0}^{n} e^{tk} C_n^k p^k (1-p)^{n-k} \\
&= e^{-nt(p+\varepsilon)} \sum_{k=0}^{n} e^{tk} C_n^k p^k (1-p)^{n-k} \\
&= e^{-nt(p+\varepsilon)}(1-p+pe^t)^n \\
&= e^{-n(t(p+\varepsilon) - \ln(1-p+pe^t))}
\end{aligned}$$

Posons :

$$h := \sup_{t>0}(t(p+\varepsilon) - \ln(1-p+pe^t))$$

On a alors :

$$\mathbb{P}\left(\frac{S_n}{n} \geq p+\varepsilon\right) \leq e^{-nh}$$

On considère à présent la fonction $g : (t \in \mathbb{R}_+^*) \mapsto t(p+\varepsilon) - \ln(1-p+pe^t)$. Comme $g(0) = 0$, $h = \sup(g) \leq 0$. De plus $\forall t > 0$, $g'(t) = p + \varepsilon - \dfrac{pe^t}{1-p+pe^t}$ et donc $g'(0) = \varepsilon > 0$, ce qui implique que g est strictement croissante au voisinage de 0 et donc que $h > 0$. Enfin si on pose $t_0 := \ln\left(\dfrac{(p+\varepsilon)(1-p)}{p(1-p-\varepsilon)}\right)$ alors une rapide étude de fonction montre que $h = g(t_0) = h_+\varepsilon$, d'où les points (i) et (ii).

(iii) Par (ii) on a :

$$\frac{1}{n}\ln\left(\mathbb{P}\left(\frac{S_n}{n} \geq p+\varepsilon\right)\right) \leq -h_+(\varepsilon) \tag{1}$$

Nous allons donc tenter de minorer cette quantité. Pour $n \geq 1$ on pose $k_n := \lceil n(p+\varepsilon)\rceil$ et on remarque que $\mathbb{P}\left(\frac{S_n}{n} \geq p+\varepsilon\right) = \mathbb{P}(S_n \geq n(p+\varepsilon)) \geq \mathbb{P}(S_n = k_n) = C_n^{k_n} p^{k_n}(1-p)^{n-k_n}$. En appliquant la formule de Stirling à chacune[1] des factorielles présentes dans $C_n^{k_n}$ on trouve alors :

$$\mathbb{P}(S_n = k_n) \sim \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{k_n(n-k_n)}}\left(\frac{np}{k_n}\right)^{k_n}\left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n} \tag{2}$$

Or $k_n = n(p+\varepsilon) + o(n)$ et $n - k_n = n(1-p-\varepsilon) + o(n)$, donc $\mathbb{P}(S_n = k_n) \xrightarrow[n\to\infty]{} 0$, on peut passer au logarithme dans l'équivalent (2), d'où :

$$\frac{1}{n}\ln(\mathbb{P}(S_n = k_n)) \sim \frac{-1}{2n}\ln(2\pi) + \frac{1}{2n}\ln\left(\frac{n}{k_n(n-k_n)}\right) + \frac{1}{n}\ln\left(\left(\frac{np}{k_n}\right)^{k_n}\right) + \frac{1}{n}\ln\left(\left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n}\right)$$

On sait que $\dfrac{n}{k_n(n-k_n)} \xrightarrow[n\to\infty]{} 0$. Remarquons ensuite que :

$$\begin{aligned}\forall n \geq 1,\ \ln\left(\left(\frac{np}{k_n}\right)^{k_n}\right) &= k_n \ln\left(\frac{np}{k_n}\right) = n(p+\varepsilon)\ln\left(\frac{np}{k_n}\frac{p+\varepsilon}{p+\varepsilon}\right) + (k_n - n(p+\varepsilon))\ln\left(\frac{np}{k_n}\right)\\ &= n(p+\varepsilon)\ln\left(\frac{p}{p+\varepsilon}\right) + n(p+\varepsilon)\ln\left(\frac{n(p+\varepsilon)}{k_n}\right) + (k_n - n(p+\varepsilon))\ln\left(\frac{np}{k_n}\right)\end{aligned}$$

Ainsi, comme $(k_n - n(p+\varepsilon))_n$ est bornée (par 1) on obtient :

$$\frac{1}{n}\ln\left(\left(\frac{np}{k_n}\right)^{k_n}\right) \xrightarrow[n\to\infty]{} (p+\varepsilon)\ln\left(\frac{p}{p+\varepsilon}\right)$$

De même :

$$\frac{1}{n}\ln\left(\left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n}\right) \xrightarrow[n\to\infty]{} (1-p-\varepsilon)\ln\left(\frac{1-p}{1-p-\varepsilon}\right)$$

Et donc in fine :

$$\frac{1}{n}\mathbb{P}(S_n = k_n) \xrightarrow[n\to\infty]{} -h_+(\varepsilon)$$

Ce qui implique , comme $\mathbb{P}\left(\frac{S_n}{n} \geq p+\varepsilon\right) \geq \mathbb{P}(S_n = k_n)$ que :

$$\liminf_{n\to\infty}\frac{1}{n}\ln\left(\mathbb{P}\left(\frac{S_n}{n} \geq p+\varepsilon\right)\right) \geq -h_+(\varepsilon) \tag{3}$$

Or par (1) :

$$\limsup_{n\to\infty}\frac{1}{n}\ln\left(\mathbb{P}\left(\frac{S_n}{n} \geq p+\varepsilon\right)\right) \leq -h_+(\varepsilon) \tag{4}$$

Et donc, en combinant (3) et (4) :

$$\frac{1}{n}\ln\left(\mathbb{P}\left(\frac{S_n}{n} \geq p+\varepsilon\right)\right) \xrightarrow[n\to\infty]{} -h_+(\varepsilon)$$

1. J'aimerais que ce soit une mauvaise blague. J'aimerais tellement.

Détails supplémentaires :

– Rappelons la très divertissante *formule de Stirling* :

$$n! \sim_{n\to\infty} \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$

– Soient u et v deux suites de $\mathbb{R}_+^*$ convergeant vers une limite $\ell \neq 1$ telles que $u_n \sim v_n$. Montrons que $\ln(u_n) \sim \ln(v_n)$. Remarquons que comme $\lim_n v_n \neq 1$, il existe $N \geq 0$ tel que $\forall n \geq N$, $v_n \neq 1$. On peut alors écrire :

$$\forall n \geq N,\ \frac{\ln(u_n)}{\ln(v_n)} - 1 = \frac{\ln(u_n) - \ln(v_n)}{\ln(v_n)} = \frac{\ln\left(\frac{u_n}{v_n}\right)}{\ln(v_n)} \xrightarrow[n\to\infty]{} 0$$

D'où le résultat.

Références

[Les01] Emmanuel Lesigne. *Pile ou face.* Ellipses, 2001.