

COMPTEUR LOGLOG

Leçons : 260, 264

References:

Théorème

Soit $\varepsilon, \delta > 0$.

Il existe une suite de variables aléatoires $(X_i)_{i \in \mathbb{N}}$

telles que $\mathbb{P}(|X_n - n| \geq \varepsilon n) \leq \delta$ pour tout $n \in \mathbb{N}$.

et presque sûrement, X_n se code en $O(\log_2(\log_2(n)))$ bits

Résumé

I - On a, pour $n \in \mathbb{N}$, $2^{\log_2 n} = n$. On va estimer $\lfloor \log_2 n \rfloor$ avec une variable aléatoire dont l'espérance est n et la variance $\frac{n(n-1)}{2}$.

En posant $(\xi_{i,n})_{\substack{i \geq 0 \\ n \geq 0}}$ une suite de variables aléatoires indépendantes, suivant une loi de Bernoulli $\text{Ber}(2^{-i})$, et S_n une somme de ces v.a.

II - Calcul d'un estimateur non biaisé de n et de la précision en proba.

III - Calcul de l'ordre de grandeur de cet estimateur en proba.

I] • Prenons $\varepsilon, \delta > 0$.

Posons $(\xi_{i,k})_{\substack{i \geq 0 \\ k \geq 0}}$ des v.a. i. de loi $\text{Ber}(2^{-i})$.

$$S_0 = 0 \text{ et } S_{n+1} = \sum_{i=0}^{\infty} \xi_{i,n+1} \quad S_n + \sum_{i=0}^{\infty} \xi_{i,n+1}$$

On a alors, d'une part,

$$\mathbb{E}(2^{S_0}) = \mathbb{E}(2^0) = 1.$$

on veut approcher $\lfloor \log_2 n \rfloor$, donc à chaque fois qu'on augmente de 1, on divise par 2 la proba d'augmenter à nouveau... c'est ça le log!

On calcule l'espérance de 2^{S_n} .

$$= E\left(\sum_{k=0}^{+\infty} 2^{S_{n+1}} \mathbb{1}_{S_n=k}\right) \text{ puis permutation par Fubini.}$$

Et, on a

$$\begin{aligned} E(2^{S_{n+1}}) &= \sum_{k=0}^{+\infty} E\left[2^{S_{n+1}} \mathbb{1}_{S_n=k}\right] \quad \text{par le théorème de convergence monotone,} \\ &= \sum_{k=0}^{+\infty} E\left[2^{k+\tilde{S}_{k,n+1}} \mathbb{1}_{S_n=k}\right] \quad \text{et car } E(2^{S_{n+1}}) \mathbb{1}_{S_n=k} \text{ est la limite} \\ &= \sum_{k=0}^{+\infty} 2^k E(2^{\tilde{S}_{k,n+1}}) E(\mathbb{1}_{S_n=k}) \quad \text{d'une somme, et donc d'une suite croissante.} \\ &= \sum_{k=0}^{+\infty} 2^k E(2^{\tilde{S}_{k,n+1}}) E(\mathbb{1}_{S_n=k}) \quad \text{par indépendance.} \\ &= \sum_{k=0}^{+\infty} 2^k \cdot (2^1 \cdot 2^{-k} + 2^0 \cdot (1 - 2^{-k})) E(\mathbb{1}_{S_n=k}) \\ &= \sum_{k=0}^{+\infty} 2^k E(\mathbb{1}_{S_n=k}) + \sum_{k=0}^{+\infty} E(\mathbb{1}_{S_n=k}) \\ &= E(2^{S_n}) + 1. \end{aligned}$$

D'où, par récurrence, $E(2^{S_n}) = n+1$.

• Calculons ensuite la variance de 2^{S_n} , pour avoir un résultat de précision.

$$\ast \text{Var}(2^{S_0}) = \text{Var}(1) = 0.$$

$$\begin{aligned} \ast \text{Var}(2^{S_{n+1}}) &= E((2^{S_{n+1}})^2) - E(2^{S_{n+1}})^2 \\ &= E(4^{S_{n+1}} - (n+2)^2) \\ &= \sum_{k=0}^{+\infty} E\left[(4^{S_{n+1}} - (n+2)^2) \mathbb{1}_{S_n=k}\right] \quad (\text{convergence dominée}) \\ &= \sum_{k=0}^{+\infty} (4^k E(4^{\tilde{S}_{k,n+1}}) - (n+2)^2) E(\mathbb{1}_{S_n=k}) \quad (\text{indépendance}) \\ &= \sum_{k=0}^{+\infty} (4^k (4 \cdot 2^{-k} + 1(1 - 2^{-k})) - (n+1)^2 - 2(n+1) - 1) E(\mathbb{1}_{S_n=k}) \\ &= \sum_{k=0}^{+\infty} (4 \cdot 2^k + 4^k - 2^k - (n+1)^2 - 2(n+1) - 1) E(\mathbb{1}_{S_n=k}) \\ &= E(4^{S_n} - (n+1)^2) + 3 E(2^{S_n}) - 2(n+1) - 1 \\ &= \text{Var}(2^{S_n}) + 3(n+1) - 2(n+1) - 1 \\ &= \text{Var}(2^{S_n}) + n. \end{aligned}$$

d'où $\text{Var}(2^{S_n}) = \sum_{k=1}^n k - 1 = \frac{n(n-1)}{2}$.

II • Construisons notre estimateur, en posant :

$$X_n = 2^{S_n} - 1.$$

Et posons $X_n^{(1)}, \dots, X_n^{(k)}$ k copies indépendantes de X_n .

On a alors naturellement envie de poser $\bar{X}_n = \frac{1}{k} \sum_{i=1}^k X_n^{(i)}$.

Cela donne :

$$\bullet \mathbb{E}(\bar{X}_n) = \frac{1}{k} \sum_{i=1}^k \mathbb{E}(X_n^{(i)}) = \mathbb{E}(X_n).$$

$$\bullet \text{Var}(\bar{X}_n) = \frac{1}{k^2} \cdot k \text{Var}(X_n) \quad \text{par indépendance} \triangle$$

$$= \frac{1}{k} \cdot \frac{n(n-1)}{2}.$$

• Dès lors, on obtient :

$$\mathbb{P}(|\bar{X}_n - n| \geq \varepsilon n) \leq \frac{\text{Var} \bar{X}_n}{(\varepsilon n)^2} \leq \frac{1}{2k\varepsilon^2} \leq \delta \text{ pour } k \text{ assez grand.}$$

III • Enfin, regardons l'ordre de grandeur des S_n en posant :

$$N_n = \lfloor \ln_2 S_n \rfloor + 1 + \dots + \lfloor \ln_2 S_n \rfloor + 1 = k(\lfloor \ln_2 S_n \rfloor + 1)$$

On va montrer $N_n \leq 3k \ln_2(\ln_2(n))$ presque sûrement :

$$\begin{aligned} \mathbb{P}(N_n \geq 3k \ln_2(\ln_2(n))) &\leq k \mathbb{P}(\lfloor \ln_2 S_n \rfloor + 1 \geq 3 \cdot \ln_2(\ln_2(n))) \\ &\leq k \mathbb{P}(\ln_2 S_n \geq 2 \cdot \ln_2(\ln_2(n))) \quad \text{pour } n \geq 4. \\ &\leq k \mathbb{P}(S_n \geq \ln_2(n)^2) \\ &\leq k \mathbb{P}(2^{S_n} - 1 \geq n^{\ln_2(n)} - 1) \\ &\leq k \cdot \frac{\mathbb{E}(2^{S_n} - 1)}{n^{\ln_2(n)} - 1} \\ &\leq \frac{k}{n^2} \quad \text{pour } n \text{ assez grand.} \end{aligned}$$

Des lors on a la convergence de la somme

$$\sum_{n=0}^{+\infty} P(N_n > 3k \ln(\ln(n))) < +\infty.$$

Donc, d'après Borel-Cantelli,

$$P(\exists N \forall n \geq N \quad N_n < 3k \ln(\ln(n))) = 1.$$