RMS 134 [1543]

Soit $(X_n)_{n\geqslant 1}$, une suite de variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$, indépendantes, de même loi, à valeurs dans $\mathbb{N}$. Pour tout entier $n \geqslant 1$, on pose

$$C_n = \#\{X_1, \dots, X_n\}.$$

1❧ *Pour tout entier $n \geqslant 1$, l'application $C_n : \Omega \to \mathbb{N}$ est une variable aléatoire d'espérance finie.*

2❧ *Démontrer que, pour tout $n \in \mathbb{N}^*$ et tout $k \in \mathbb{N}$,*

$$\mathbf{E}(C_n) \leqslant k + n^2\,\mathbf{P}(X_1 \geqslant k).$$

3❧ *On suppose dorénavant que les variables aléatoires X_k sont d'espérance finie.*
Démontrer que

$$\mathbf{P}(X_1 \geqslant k) \underset{k\to+\infty}{=} o(1/k).$$

En déduire que

$$\mathbf{E}(C_n) \underset{n\to+\infty}{=} o(n).$$

1❧ Il est clair que les valeurs de C_n sont comprises entre 1 (cas où toutes les variables aléatoires X_k prennent la même valeur) et n (cas où toutes les variables aléatoires X_k prennent des valeurs distinctes).

L'ensemble $\mathbb{N}^n$ est dénombrable (produit cartésien d'un nombre fini d'ensembles dénombrables) et

$$\mathbb{N}^n = \bigsqcup_{1\leqslant k\leqslant n} E_k$$

où E_k est l'ensemble des n-uplets formés au moyen de k entiers naturels distincts.

Les ensembles E_k sont donc des ensembles dénombrables (ils sont contenus dans un ensemble dénombrable et manifestement ils sont infinis) et, pour tout entier $1 \leqslant k \leqslant n$,

$$C_n(\omega) = k \iff \exists\,(i_1, \dots, i_n) \in E_k, \quad \big(X_1(\omega), \dots, X_n(\omega)\big) = (i_1, \dots, i_n).$$

Autrement dit,

$$[C_n = k] = \bigsqcup_{(i_1,\dots,i_n)\in E_k} [X_1 = i_1] \cap \dots \cap [X_n = i_n]$$

ce qui prouve que

$$\forall\, 1 \leqslant k \leqslant n, \qquad [C_n = k] \in \mathcal{A}$$

et donc que C_n est bien une variable aléatoire discrète sur $(\Omega, \mathcal{A})$.

❧ Comme C_n est une variable aléatoire bornée, elle est d'espérance finie.

✍ *Plus précisément, $1 \leqslant \mathbf{E}(C_n) \leqslant n$, ce qui donne de l'intérêt aux estimations qui vont suivre.*

2❧ Soit $k \in \mathbb{N}$. De deux choses l'une :

— ou bien toutes les variables aléatoires X_1, ..., X_n sont strictement inférieures à k et dans ce cas, elles ne peuvent pas prendre plus de k valeurs distinctes ;
— ou bien l'une d'elles au moins est supérieure à k.

Autrement dit :

$$\Omega = [X_1 < k, \dots, X_n < k] \sqcup \big([X_1 \geqslant k] \cup \dots \cup [X_n \geqslant k]\big)$$

et

$$[X_1 < k, \dots, X_n < k] \subset [C_n \leqslant k].$$

Comme $C_n(\omega) \leqslant n$ pour tout $\omega \in \Omega$, on en déduit que

$$C_n \leqslant k \mathbb{1}_{[X_1<k,\dots,X_n<k]} + n \mathbb{1}_{[X_1 \geqslant k] \cup \dots \cup [X_n \geqslant k]}.$$

Comme l'espérance conserve les inégalités, on en déduit que

$$\mathbf{E}(C_n) \leqslant k \, \mathbf{P}([X_1 < k, \dots, X_n < k]) + n \, \mathbf{P}([X_1 \geqslant k] \cup \dots \cup [X_n \geqslant k]).$$

Les variables aléatoires X_k sont de même loi, donc

$$\mathbf{P}([X_1 \geqslant k] \cup \dots \cup [X_n \geqslant k]) \leqslant \sum_{i=1}^{n} \mathbf{P}(X_i \geqslant k) = n \, \mathbf{P}(X_1 \geqslant k)$$

et finalement

$$\mathbf{E}(C_n) \leqslant k \times 1 + n \times n \, \mathbf{P}(X_1 \geqslant k) = k + n^2 \, \mathbf{P}(X_1 \geqslant k).$$

✍ *Il n'y a plus de calcul de probabilité dans ce qui suit, c'est un résultat d'analyse. Nous considérerons donc deux suites* positives $(u_n)_{n \in \mathbb{N}}$ *et* $(\varepsilon_n)_{n \in \mathbb{N}}$ *telles que*

$$\forall\, 0 \leqslant k \leqslant n, \quad 0 \leqslant u_n \leqslant k + n^2 \frac{\varepsilon_k}{k} \qquad \textit{avec} \qquad \lim_{k \to +\infty} \varepsilon_k = 0.$$

et nous cherchons à en déduire que

$$\lim_{n \to +\infty} \frac{u_n}{n} = 0.$$

3❧ Fixons $0 < \eta < 1$ et posons $\varepsilon = \sqrt{\eta} \in \,]0, 1[$.
D'après les hypothèses,

$$\forall\, n \in \mathbb{N}, \ \forall\, 0 \leqslant k \leqslant n, \quad 0 \leqslant \frac{u_n}{n} \leqslant \frac{k}{n} + n \frac{\varepsilon_k}{k}$$

et, comme $\eta > 0$, il existe un entier $K_0 \in \mathbb{N}$ tel que

$$\forall\, k \geqslant K_0, \quad 0 \leqslant \varepsilon_k \leqslant \varepsilon.$$

Par conséquent,

$$\forall\, n > K_0, \ \forall\, K_0 \leqslant k \leqslant n, \quad 0 \leqslant \frac{u_n}{n} \leqslant \frac{k}{n} + n \frac{\varepsilon}{k}. \tag{$\star$}$$

✍ *Le principe du raisonnement qui va suivre est classique : c'est un passage à la borne inférieure. Si*

$$\forall\, x \in [a, b], \quad g(n) \leqslant f(x)$$

alors

$$g(n) \leqslant \inf_{x \in [a,b]} f(x).$$

La mise en œuvre de ce raisonnement est (un peu) compliquée par le fait que, ici, le majorant est une fonction de la variable entière k *et non pas une fonction de la variable* réelle x.

❧ Une étude rapide montre que la fonction f définie par

$$\forall\, x > 0, \quad f(x) = \frac{x}{n} + \frac{n\varepsilon}{x}$$

atteint son minimum en $x_0 = n\sqrt{\varepsilon}$. Comme $0 < \varepsilon < 1$, on a $n\sqrt{\varepsilon} < n$ pour tout n et la suite de terme général $y_n = n\sqrt{\varepsilon}$ tend vers $+\infty$, donc il existe donc un rang $N_1 \in \mathbb{N}$ tel que

$$\forall\, n \geqslant N_1, \quad K_0 \leqslant k_0 \overset{\text{déf.}}{=} \lfloor n\sqrt{\varepsilon} \rfloor \leqslant n.$$

✍ *À défaut de pouvoir majorer par le minimum de la fonction* f*, on va majorer par une valeur particulière de* f *qui est assez proche du minimum de* f*.*

On déduit donc de $(\star)$ que

$$\forall\, n > \max\{K_0, N_1\}, \quad 0 \leqslant \frac{u_n}{n} \leqslant f(k_0) = \frac{\lfloor n\sqrt{\varepsilon}\rfloor}{n} + n\frac{\varepsilon}{\lfloor n\sqrt{\varepsilon}\rfloor}$$
$$\leqslant \frac{n\sqrt{\varepsilon}+1}{n} + n\frac{\varepsilon}{n\sqrt{\varepsilon}} \leqslant \frac{1}{n} + 2\sqrt{\varepsilon}.$$

Comme la suite de terme général $1/n$ tend vers 0, il existe un rang $N_1 \in \mathbb{N}$ tel que

$$\forall\, n \geqslant N_2, \qquad \frac{1}{n} \leqslant \eta.$$

Finalement, en posant $N_0 = \min\{K_0, N_1, N_2\} \in \mathbb{N}$,

$$\forall\, n > N_0, \qquad 0 \leqslant \frac{u_n}{n} \leqslant 3\eta$$

et nous avons bien démontré que

$$\lim_{n\to+\infty} \frac{u_n}{n} = 0.$$