P4 : Espérance, variance, moments…

I Espérance

I.1 Variables aléatoires réelles positives

Définition

Soit X une variable aléatoire discrète sur $(\Omega, \mathscr{A}, P)$, à valeurs dans $\mathbf{R}^+$. L'espérance de X est, par définition,

$$\mathrm{E}(X) = \sum_{x \in X(\Omega)} P(X = x)\ x$$

Explication $\mathrm{E}(X)$ est un élément de $\mathbf{R}^+ \cup \{+\infty\} = [0, +\infty]$: l'espérance est réelle (ou « finie ») si et seulement si la famille $(P(X = x)\ x)_{x \in X(\Omega)}$ est sommable. Dans le cas contraire, on a $\mathrm{E}(X) = +\infty$.

Remarque Cette définition diffère de celle donnée en mpsi pour les variables aléatoires sur un univers fini, qui est

$$\mathrm{E}(X) = \sum_{\omega \in \Omega} P(\{\omega\}) X(\omega)$$

(somme finie si Ω est fini). Elle lui est équivalente (pour le voir, il suffit dans la somme $\sum_{\omega \in \Omega} P(\{\omega\}) X(\omega)$ de regrouper les ω qui ont la même image par X). Mais elle a l'inconvénient de ne pas avoir de sens si Ω est infini non dénombrable. Passer de l'espérance comme somme indexée par Ω (quand c'est possible) à l'espérance comme somme indexée par $X(\Omega)$ est un « transfert », voir plus loin.

I.2 Espérance d'une loi géométrique

Proposition Si $X \sim \mathscr{G}(p)$, alors

$$\mathrm{E}(X) =$$

I.3 Une formule célèbre

Proposition Soit X une variable aléatoire à valeurs dans $\mathbf{N} \cup \{+\infty\}$. Montrer que

$$\mathrm{E}(X) = \sum_{n=1}^{+\infty} \mathrm{P}(X \geq n)$$

Il s'agit d'un bel exercice de sommabilité. Appliquer la formule à une variable géométrique, pour la tester. *On peut se contenter de savoir la montrer pour une variable aléatoire à valeurs dans* $\mathbf{N}$. *Il est néanmoins intéressant de savoir justifier que*

$$\mathrm{P}(X \geq n) \xrightarrow[n \to +\infty]{} \mathrm{P}(X = +\infty)$$

I.4 Variable aléatoire réelle ou complexe quelconque

Définition Soit X une variable aléatoire réelle ou complexe discrète sur $(\Omega, \mathscr{A}, P)$. Lorsque la famille $(P(X = x)\ x)_{x \in X(\Omega)}$ est sommable, on dit que X est d'espérance finie, et on définit

$$\mathrm{E}(X) = \sum_{x \in X(\Omega)} P(X = x)\ x$$

Dans le cas contraire, X n'a pas d'espérance (pas plus infinie que finie).

Vocabulaire Une expression équivalente à « X est d'espérance finie » est « X a un moment d'ordre 1 ». On a même rencontré dans certains énoncés : « X est intégrable ». Et « X est sommable » est bien compréhensible et peu choquant. Notons que le programme officialise la notation $X \in L^1$ pour signifier que X est d'espérance finie.

Rappel On dit donc que X est d'espérance finie lorsque la famille $(P(X = x)\ |x|)_{x \in X(\Omega)}$ est sommable. Ce qui est cohérent avec ce qui précède dans le cas où X est réelle positive.

Définition On dit qu'une variable aléatoire (réelle ou complexe) est centrée quand son espérance est nulle.

Ne pas confondre « centrée » et « symétrique » : si $P(X = 1) = P(X = 2) = P(X = 3) = P(X = -6) = 1/4$, *la variable aléatoire* X *est bien centrée. Pour autant,* X *et* $-X$ *n'ont pas même loi.*

I.5 Théorème de transfert

Théorème

Soit X un variable aléatoire discrète sur un espace probabilisé $(\Omega, \mathscr{A}, P)$, f une fonction définie sur $X(\Omega)$, à valeurs réelles ou complexes ; $f(X)$ est d'espérance finie si et seulement si la famille $\big(P(X=x)\ f(x)\big)_{x\in X(\Omega)}$ est sommable, et dans ce cas

$$\mathrm{E}\big(f(X)\big) = \sum_{x\in X(\Omega)} P(X=x)\ f(x)$$

C'est un résultat d'une utilité considérable ! On l'utilisera quelquefois dans des cas où X n'est pas à valeurs réelles. Par exemple, pour l'espérance d'une somme de deux variables aléatoires réelles X_1 et X_2, on peut considérer
$X\ :\ \omega \longmapsto (X_1(\omega), X_2(\omega))$ et $f\ :\ (x_1, x_2) \longmapsto x_1 + x_2$.
On peut alors traduire le résultat précédent dans ce contexte.

Démonstration En annexe. *Cette démonstration est un exercice de sommabilité, pas de probabilité.*

Corollaire Une variable aléatoire réelle ou complexe discrète X est d'espérance finie si et seulement si $|X|$ l'est, c'est-à-dire si et seulement si

$$\mathrm{E}\left(|X|\right) < +\infty$$

Et le cas échéant,

$$\mathrm{E}(|X|) = \sum_{x\in X(\Omega)} P(X=x)|x|$$

I.6 Quelques calculs

…Pour s'entraîner à utiliser le transfert :

Si $X \sim \mathscr{G}(p)$, calculer les espérances de $\dfrac{1}{X}$ et de X^2. Puis celle de a^X, où a est un réel strictement positif.

I.7 Cas d'un univers fini ou dénombrable

Proposition Soit X une variable aléatoire réelle ou complexe discrète sur un espace probabilisé $(\Omega, \mathscr{A}, P)$, Ω étant dénombrable (ou fini). X est d'espérance finie si et seulement si la famille $(P(\{\omega\})\ X(\omega))_{\omega\in\Omega}$ est sommable. Et si c'est le cas,

$$\mathrm{E}(X) = \sum_{\omega\in\Omega} P(\{\omega\})\ X(\omega)$$

Démonstration C'est une utilisation du transfert, mais un peu astucieuse : il suffit de remarquer que l'identité Id_Ω est une variable aléatoire discrète sur Ω. Et X une application définie sur Ω. On applique ce qui précède en remplaçant X par Id_Ω et f par X.

Remarque On retrouve en particulier la définition de l'espérance vue en mpsi pour les univers finis.

Remarque Soit X une variable discrète réelle ou complexe sur un espace probabilisé quelconque $(\Omega, \mathscr{A}, P)$. On munit $X(\Omega)$ (qui est fini ou dénombrable) de la probabilité P_X (la loi de X). Que peut-on dire de l'espérance de l'identité $(x \longmapsto x)$?

I.8 Propriétés de l'espérance

Toutes les variables aléatoires dont il est question dans ce paragraphe sont discrètes, réelles ou complexes, définies sur le même espace probabilisé $(\Omega, \mathscr{A}, P)$. Si on sait que l'espérance est une intégrale (au sens de la théorie de Lebesgue, celle de Riemann est ici inadaptée), les propriétés qui suivent doivent sembler assez naturelles.

a. Critère d'espérance finie

Proposition Si $|X| \leq Y$ et si Y est d'espérance finie, alors X est d'espérance finie, et

$$|\mathrm{E}(X)| \leq \mathrm{E}(Y)$$

Autrement dit, si $|X| \leq Y$ et si $Y \in L^1$, alors $X \in L^1$.

b. Linéarité de l'espérance

Proposition Si X et Y sont d'espérance finie, si $\lambda \in \mathbf{K}$, $\mathbf{K} = \mathbf{R}$ ou $\mathbf{C}$, $\lambda X + Y$ est d'espérance finie, et

$$\mathrm{E}(\lambda X + Y) = \lambda \mathrm{E}(X) + \mathrm{E}(Y)$$

Autrement dit, l'ensemble des variables aléatoires réelles ou complexes discrètes d'espérance finie sur $(\Omega, \mathscr{A}, P)$ est un espace vectoriel, et E est une forme linéaire sur cet espace.

c. Positivité, croissance

Proposition Si X est positive (i.e. à valeurs réelle positives),

$$\mathrm{E}(X) \geq 0$$

(si $X \geq 0$ presque sûrement, c'est encore vrai).

Si $X \leq Y$, si X et Y étant réelles et d'espérance finie,

$$\mathrm{E}(X) \leq \mathrm{E}(Y)$$

(si $X \leq Y$ presque sûrement, c'est encore vrai).

Si X est d'espérance finie,

$$|\mathrm{E}(X)| \leq \mathrm{E}(|X|)$$

d. Les variables positives et d'espérance nulle...

...sont les variables

e. On a déjà vu ce genre de propriété

Redisons-le, l'espérance est une intégrale...mais pas dans le cadre de la théorie de Riemann. La théorie de l'intégration de Lebesgue permet (ce n'est pas la seule) d'interpréter l'espérance comme une intégrale. Le savoir permet de ne pas être surpris par les propriétés ci-dessus.

I.9 Quelques propriétés supplémentaires

On ne considère dans cette section que des variables aléatoires réelles ou complexes discrètes sur un espace probabilisé.

a. Variables aléatoires p.s. constantes

Si $X = a$ presque sûrement, X est d'espérance finie, et

$$\mathrm{E}(X) =$$

b. Complément à la positivité

Si $X \geq 0$ (X réelle), et si X est d'espérance nulle, alors

c. Variables aléatoires bornées

Si $a \leq X \leq b$ (X réelle) est presque sûr, alors X est d'espérance finie.

I.10 Fonctions indicatrices

Si A est un événement, $\mathbf{1}_A$ est d'espérance finie, et

$$\mathrm{E}(\mathbf{1}_A) =$$

I.11 Espérance d'une loi binomiale

« La » bonne idée est de calculer cette espérance en disant que l'espérance d'une somme de v.a. est la somme des espérances. Or, si X suit une loi de Bernoulli $\mathscr{B}(p)$,

$$E(X) = p$$

Donc, si $X_1, \dots, X_n$ suivent toutes des lois $\mathscr{B}(p)$,

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = np$$

Finalement,

Proposition L'espérance d'une loi binomiale $\mathscr{B}(n, p)$ est np.

Remarque : faire les calculs directement est un intéressant exercice de sommation. On considère donc une variable aléatoire X qui suit une loi $\mathscr{B}(n, p)$:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= n \sum_{k=1}^{n} \frac{k}{n} \binom{n}{k} p^k (1-p)^{n-k} \\
&= n \sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\
&= np(p+1-p)^{n-1} \\
&= np
\end{aligned}$$

(avec une réindexation $j = k - 1$).

Insistons : ce n'est vraiment pas la bonne méthode. Mais la sommation effectuée est intéressante. La bonne méthode, c'est dire que

$$\mathrm{E}(X_1 + \cdots + X_n) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n)$$

où les X_i suivent une loi $\mathscr{B}(p)$ (ici, l'indépendance ne sert pas).

I.12 Espérance d'une loi uniforme sur un ensemble fini

Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathscr{A}, P)$. On suppose que X suit une loi uniforme sur $\{a_1, \dots, a_n\}$ où les a_i sont des réels distincts. Alors

$$E(X) =$$

II Moments

II.1 Définition

Définition Soit X une variable aléatoire réelle discrète sur $(\Omega, \mathscr{A}, P)$. Soit $p \geq 1$ (p réel). On dit que X admet un moment d'ordre p lorsque

$$\mathrm{E}\left(|X|^p\right) < +\infty$$

Le moment d'ordre p de X est alors $\mathrm{E}\left(X^p\right)$

Rappelons qu'admettre un moment d'ordre 1, c'est avoir une espérance finie.

II.2 Moments d'ordre 2

Une variable aléatoire discrète X admet un moment d'ordre 2 (et on note $X \in L^2$) lorsque

$$\mathrm{E}\left(|X|^2\right) < +\infty$$

(on peut bien sûr oublier la valeur absolue. Mais si le programme se limite aux variables réelles, un énoncé peut en revanche très bien s'intéresser à des variables aléatoires complexes, et alors il ne faudrait pas omettre le module).

II.3 Moments d'ordres 1 et 2

Proposition Si une variable aléatoire (réelle discrète) admet un moment d'ordre 2, elle est d'espérance finie.

On peut résumer en $L^2 \subset L^1$.
Autrement dit, il est plus facile d'avoir un moment d'ordre 1 qu'un moment d'ordre 2. D'une manière générale, il es plus facile d'avoir un moment d'ordre r qu'un moment d'ordre r' si $r < r'$.
Pour la proposition, deux démonstrations, à connaître toutes les deux!

Démonstration 1 : Cauchy-Schwarz. . .voir paragraphe suivant.

Démonstration 2 : Simple mais légèrement astucieuse et l'idée est intéressante. . .On commence par constater simplement que, pour tout $\omega \in \Omega$,

$$|X(\omega)| \leq X^2(\omega) \qquad \text{ou} \qquad |X(\omega)| \leq 1$$

ce qui permet la majoration

$$|X(\omega)| \leq$$

II.4 Inégalité de Cauchy-Schwarz

Proposition Si deux variables aléatoires réelles discrètes X et Y sur $(\Omega, \mathscr{A}, P)$ ont des moments d'ordre 2, leur produit est d'espérance finie, et

$$(\mathrm{E}(XY))^2 \leq \mathrm{E}(X^2)\ \mathrm{E}(Y^2)$$

ou encore

$$|\mathrm{E}(XY)| \leq \sqrt{\mathrm{E}(X^2)}\ \sqrt{\mathrm{E}(Y^2)}$$

…ce qui précise le résultat du paragraphe précédent : en prenant $Y = 1$, on obtient

$$(\mathrm{E}(X))^2 \leq \mathrm{E}(X^2)$$

Démonstration Classique, rappelons que l'espérance est une intégrale. Ce qu'on ne peut pas utiliser, mais ce qui rend les résultats plus évidents.

Notations Le début peut se résumer en :
Si X et Y sont dans L^2 alors $XY \in L^1$.

II.5 Deux espaces vectoriels

L^1 et L^2 sont deux espaces vectoriels (résultat très mal écrit, mais prenons-le comme un slogan). Le second est un sous-espace du premier.

II.6 Variance et écart-type

Ce sont des notions « euclidiennes », voir exercice; cet aspect n'est pas explicitement au programme, car l'application $(X,Y) \mapsto E(XY)$ n'est pas tout-à-fait un produit scalaire sur $\mathscr{L}^2(\Omega,\mathscr{A},P)$: elle n'est « que » bilinéaire symétrique positive. Pour obtenir un vrai produit scalaire, il faut considérer l'ensemble quotient de $\mathscr{L}^2(\Omega,\mathscr{A},P)$ par la relation d'égalité presque sûre, qui est une relation d'équivalence. L'ensemble quotient est en général noté $L^2(\Omega,\mathscr{A},P)$

a. Définition

Soit X une variable aléatoire réelle discrète admettant un moment d'ordre 2. On peut alors définir sa variance :

$$\mathrm{V}(X) = \mathrm{E}\left[(X-\mathrm{E}(X))^2\right]$$

et son écart type :

$$\sigma(X) = \sqrt{\mathrm{V}(X)}$$

b. Formules

Soit X une variable aléatoire réelle discrète admettant un moment d'ordre 2. *Plus bref : soit $X \in L^2$*. On a

$$\mathrm{V}(X) = \sum_{x\in X(\Omega)} P(X=x)(x-\mathrm{E}(X))^2$$

Et

$$\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$$

II.7 Homogénéité, invariance par translation

Proposition Soit X une variable aléatoire réelle discrète, admettant un moment d'ordre 2. Alors, pour tous réels a et b,

$$\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$$

$$\sigma(aX+b) = |a|\,\sigma(X)$$

Définition On dit qu'une variable aléatoire est réduite lorsque sa variance (ou son écart type) est égal à 1.

Proposition Si X est une variable aléatoire admettant un moment d'ordre 2, si $m = \mathrm{E}(X)$, si $\sigma = \sigma(X)$,

$$\frac{X-m}{\sigma}$$

est centrée réduite.

II.8 Un exercice d'oral

Exercice : Soit X une variable aléatoire à valeurs dans $[a,b]$, $a<b$. Montrer que

$$V(X) \leq \frac{(b-a)^2}{4}$$

Trouver un cas dans lequel cette inégalité est une égalité.

II.9 Remarque sur la notation de l'espérance

La majuscule du E de l'espérance n'est pas agréable dans les calculs : on a intérêt à réserver les majuscules pour les variables aléatoires, les minuscules pour les nombres réels. Noter $m = \mathrm{E}(X)$ si X est une variable aléatoire n'est donc pas une mauvaise idée. On constate que
$\mathrm{E}(X-m) = \mathrm{E}(X) - m = 0$
est plus facile à écrire (et à lire) que $\mathrm{E}\,(X - \mathrm{E}(X)) = \mathrm{E}(X) - \mathrm{E}(X) = 0$
De même, il est plus agréable, pour la formule de calcul de la variance, de calculer
$\mathrm{E}\left((X-m)^2\right)$
que $\mathrm{E}\left((X-\mathrm{E}(X))^2\right)$

II.10 Exemple : variance d'une loi géométrique

Proposition Si $X \sim \mathscr{G}(p)$, alors

$$\mathrm{V}(X) =$$

$$\sigma(X) =$$

et la variable aléatoire est centrée réduite.

III Covariance

III.1 Définition, formule

Définition Si X et Y sont deux variables aléatoires discrètes réelles admettant des moments d'ordre 2, on définit leur covariance :

$$\mathrm{Cov}(X,Y) = \mathrm{E}\left((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right)$$

Remarque

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$$

$$\mathrm{Cov}(X,X) =$$

Formule

$$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$$

III.2 Couple de variables aléatoires

Proposition (variance d'une somme) Soit X, Y deux variables aléatoires discrètes réelles admettant des moments d'ordre 2. Alors

$$\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y) + 2\mathrm{Cov}(X,Y)$$

Proposition (espérance d'un produit de variables indépendantes) Soit X, Y deux variables aléatoires indépendantes admettant une espérance finie. Alors XY a une espérance finie, et

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$$

Cette propriété, très importante, s'étend à un nombre fini quelconque de variables aléatoires indépendantes.
On voit donc que si X et Y sont dans L^1 et indépendantes, alors $XY \in L^1$.
En revanche, si X et Y ne sont pas indépendantes, l'appartenance de X et Y à L^1 ne suffit pas pour assurer que $XY \in L^1$. Il faut supposer X et Y dans L^2.

Démonstration Comme pour la démonstration de l'espérance d'une somme, on utilise le théorème de transfert, appliqué à la variable aléatoire discrète (mais pas réelle) $Z = (X, Y)$, et à l'application $f\ :\ (x, y) \mapsto xy$ définie sur $\mathbf{R} \times \mathbf{R}$ qui est bien, elle, à valeurs réelles : $XY = f(Z)$ est d'espérance finie si et seulement si la famille

$$\left(xyP\left((X,Y)=(x,y)\right)\right)_{(x,y)\in Z(\Omega)}$$

est sommable. Ce qui équivaut à dire que

$$\left(xyP\left((X,Y)=(x,y)\right)\right)_{(x,y)\in X(\Omega)\times Y(\Omega)}$$

l'est (si $(x, y) \notin Z(\Omega)$, $P\left((X,Y)=(x,y)\right) = 0$, on rajoute donc éventuellement des termes nuls, ce qui ne change ni la sommabilité, ni, le cas échéant, la somme). Mais, par indépendance de X et Y, la famille $\left(xyP\left((X,Y)=(x,y)\right)\right)_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est une famille produit :

$$\forall (x,y) \in X(\Omega)\times Y(\Omega) \qquad xyP\left((X,Y)=(x,y)\right) = (xP(X=x))\left(yP(Y=y)\right)$$

ce qui donne directement le résultat (dans le chapitre sur la sommabilité, ce résultat sur les familles produits est donné, dans le cadre du programme, pour des suites doubles. Mais une famille double $\left(\alpha_i\beta_j\right)_{(i,j)\in I\times J}$, où I et J sont dénombrables, peut être réindexée en suite double puisqu'il existe des bijections de $\mathbf{N}$ sur I et de $\mathbf{N}$ sur J).

Proposition (covariance de variables indépendantes) Soit X, Y deux variables aléatoires indépendantes admettant des moments d'ordre 2. Alors

$$\mathrm{Cov}(X,Y) = 0$$

$$\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$$

La réciproque est fausse On peut lire un contre-exemple sur le tableau suivant :

X \ Y	-1	0	1
-1	0	1/4	0
0	1/4	0	1/4
1	0	1/4	0

Ou, de manière constructive, considérer deux variables aléatoires U et V de même loi, et définir $X = U + V$ et $Y = U - V$. On a Y et $XY = U^2 - V^2$ centrées, donc $\mathrm{Cov}(X, Y) = 0$. Or rien n'oblige X et Y à être indépendantes. Par exemple, si U et V sont les résultats de deux lancers de dés indépendants…

III.3 Famille finie de variables aléatoires

Proposition Si $X_1,\dots,X_p$ admettent des moments d'ordre 2, alors

$$V(X_1+X_2+\cdots+X_p)=\sum_{i=1}^{p}V(X_i)\ +\ 2\sum_{1\leq i<j\leq n}\mathrm{Cov}(X_i,X_j)$$

Proposition Si $X_1,\dots,X_p$ admettent des moments d'ordre 2 et sont deux à deux indépendantes, alors

$$V\left(\sum_{i=1}^{p}X_i\right)=\sum_{i=1}^{p}V(X_i)$$

On observera d'ailleurs qu'il suffit que les X_i soient « décorrélées », l'indépendance n'est pas nécessaire.

III.4 Variance d'une loi binomiale

a. La bonne méthode

On commence par le calcul de la variance d'une loi de Bernoulli, puis on utilise le résultat précédent.

b. La mauvaise méthode

Si X suit une loi $\mathscr{B}(n,p)$, par théorème de transfert, on a :

$$\begin{aligned}
\mathrm{E}(X^2) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= n \sum_{k=1}^{n} k \frac{k}{n} \binom{n}{k} p^k (1-p)^{n-k} \\
&= n \sum_{k=1}^{n} k \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= n \sum_{k=1}^{n} (k-1+1) \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= n(n-1) \sum_{k=1}^{n} \frac{k-1}{n-1} \binom{n-1}{k-1} p^k (1-p)^{n-k} \ + \ np \\
&= n(n-1) \sum_{j=0}^{n-2} \binom{n-2}{j} p^{j+2} (1-p)^{n-2-j} \ + \ np \\
&= n(n-1)p^2 + np
\end{aligned}$$

Ce calcul est une sommation intéressante, mais ce n'est pas la bonne méthode pour obtenir le résultat.

IV Inégalités de Markov et de Bienaymé-Tchebychev

IV.1 Inégalité de Markov

Proposition (**Inégalité de Markov**)

Soit X une variable aléatoire réelle discrète définie sur un espace probabilisé $(\Omega, \mathscr{A}, P)$, et d'espérance finie. Alors, pour tout réel strictement positif a,

$$P(|X| \geq a) \leq \frac{\mathrm{E}(|X|)}{a}$$

Cette inégalité majore la probabilité pour une variable aléatoire de prendre de grandes valeurs (en valeur absolue).

Corollaire Soit X une variable aléatoire réelle discrète définie sur un espace probabilisé $(\Omega, \mathscr{A}, P)$. Soit $k \in [1, +\infty[$ tel que X ait un moment d'ordre k. Alors, pour tout réel strictement positif a,

$$P(|X| \geq a) \leq \frac{\mathrm{E}(|X|^k)}{a^k}$$

Démonstration de l'inégalité de Markov : première rédaction

Notons $Y = |X|$, et rappelons que

$$\mathrm{E}(Y) = \sum_{y \in Y(\Omega)} P(Y = y)\ y$$

Notons $A = \{y \in Y(\Omega)\ ;\ y \geq a\}$. Alors

$$\begin{aligned} \mathrm{E}(Y) &\geq \sum_{y \in A} P(Y = y)\ y \\ &\geq a \sum_{y \in A} P(Y = y) \end{aligned}$$

Mais, par σ-additivité,

$$\sum_{y \in A} P(Y = y) = P(Y \in A) = P(Y \geq a)$$

Démonstration de l'inégalité de Markov : deuxième rédaction
Par théorème de transfert,

$$\mathrm{E}(|X|) = \sum_{x \in X(\Omega)} P(X = x)|x|$$

Notons $A = \{x \in X(\Omega)\ ;\ |x| \geq a\}$. Alors

$$\begin{aligned}\mathrm{E}(|X|) &\geq \sum_{x \in A} P(X = x)|x| \\ &\geq a \sum_{x \in A} P(X = x)\end{aligned}$$

Mais, par σ-additivité,

$$\sum_{x \in A} P(X = x) = P(X \in A) = P(|X| \geq a)$$

Démonstration de l'inégalité de Markov : troisième rédaction
Par linéarité de l'espérance,

$$\begin{aligned}\mathrm{E}(|X|) &= \mathrm{E}\left(|X|\mathbf{1}_{|X| \geq a}\right) + \mathrm{E}\left(|X|\mathbf{1}_{|X| < a}\right) \\ &\geq \mathrm{E}\left(|X|\mathbf{1}_{|X| \geq a}\right) \\ &\geq \mathrm{E}\left(a\mathbf{1}_{|X| \geq a}\right) \\ &= a\mathrm{E}\left(\mathbf{1}_{|X| \geq a}\right) \\ &= a\mathrm{P}\left(|X| \geq a\right)\end{aligned}$$

Cette rédaction est à tous points de vue beaucoup plus intéressante que les deux premières : pas besoin de la discrétion des variables, ce type de découpage est parfois très efficace.

IV.2 Inégalité de Bienaymé-Tchebychev

Proposition (Inégalité de Bienaymé-Tchebychev)
Soit X une variable aléatoire réelle discrète définie sur un espace probabilisé $(\Omega, \mathscr{A}, P)$, admettant un moment d'ordre 2. Alors, pour tout réel strictement positif a,

$$P(|X - \mathrm{E}(X)| \geq a) \;\leq\; \frac{\mathrm{V}(X)}{a^2}$$

Démonstration On applique le corollaire de l'inégalité de Markov, avec $k = 2$, à la variable aléatoire $X - \mathrm{E}(X)$.

IV.3 Deux exercices

a. Un classique incontournable (Chernov)

Exercice : Soit X une variable aléatoire réelle, telle que, pour tout $\lambda > 0$, $E\left(e^{\lambda X}\right) < +\infty$. Soit $a \in \mathbf{R}$. Montrer

$$\forall t > 0 \qquad P(X \geq a) \leq e^{-ta} E\left(e^{tX}\right)$$

b. Une inégalité élégante mais...

...pas évidente (oral X).
Exercice : Soit X une variable aléatoire réelle admettant un moment d'ordre 2. Montrer, si $\lambda \in \mathbf{R}_*^+$, que $P(X \geq E(X) + \lambda) \leq \dfrac{V(X)}{V(X) + \lambda^2}$.
(*Inégalité de Tchebychev-Cantelli.*)

V Loi faible des grands nombres

Proposition Soit $(X_n)_{n \in \mathbf{N}}$ une suite de variables aléatoires discrètes réelles deux à deux indépendantes sur un même espace probabilisé. On suppose que les X_n sont toutes de même loi, et admettent un moment d'ordre 2. On note $m = \mathrm{E}(X_1)$ et, si $n \geq 1$,

$$S_n = X_1 + X_2 + \ldots + X_n$$

Alors, pour tout $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - m\right| \geq \epsilon\right) \xrightarrow[n\to+\infty]{} 0$$

Démonstration Appliquons l'inégalité de Tchebychev. Notons

$$Y = \frac{S_n}{n}$$

Alors

$$V(Y) =$$

et on conclut assez rapidement.

VI Annexe : démonstration des propriétés de l'espérance

Ces preuves sont sans importance pour les concours.

Démonstration d théorème de transfert Notons $Y = f(X) = f \circ X$. Et commençons par nous occuper de la première proposition :

« $f(X)$ est d'espérance finie si et seulement si la famille $\big(P(X = x)\ f(x)\big)_{x\in X(\Omega}$) est sommable »

Par définition, Y est d'espérance finie si et seulement si la famille $\big(P(Y = y)y\big)_{y\in Y(\Omega)}$ est sommable. Par définition encore, Y est d'espérance finie si et seulement si la famille $\big(P(Y = y)|y|\big)_{y\in Y(\Omega)}$ est sommable.

Rappelons que $X(\Omega)$ est fini ou dénombrable, et que par conséquent $Y(\Omega) = f(X(\Omega))$ l'est (vu en **I.3.**).

On a $Y(\Omega) = f(X(\Omega))$. Il est donc naturel de partitionner $X(\Omega)$ en classant ses éléments suivant leur image par f. On définit donc, pour tout $y \in Y(\Omega)$,

$$I_y = f^{-1}(\{y\}) = \{x \in X(\Omega)\ ;\ f(x) = y\}$$

L'évènement $(Y = y)$ est l'évènement $\big(f(X) = y\big)$, qui peut donc s'écrire $\big(X \in f^{-1}(\{y\})\big)$, ou encore $\big(X \in I_y\big)$. Par additivité ou σ-additivité (suivant si I_k est fini ou non),

$$P(X \in I_y) = \sum_{x\in I_y} P(X = x)$$

(la sommabilité de la famille au second membre est automatique, par σ-additivité de la probabilité P). C'est-à-dire

$$P(Y = y) = \sum_{x\in I_y} P(X = x)$$

Multiplions par $|y|$, en remarquant que $f(x) = y$ pour tout $x \in I_y$:

$$|y|\ P(Y = y) = \sum_{x\in I_y} |f(x)|P(X = x)$$

Comme $X(\Omega)$ est la réunion disjointe des I_y, c'est-à-dire

$$X(\Omega) = \bigcup_{y\in Y(\Omega)} I_y \qquad \text{et} \qquad y \neq y' \Rightarrow I_y \cap I_{y'} = \emptyset$$

le théorème de sommabilité par paquets dit que la famille $\big(|y|\ P(Y=y)\big)_{y\in Y(\Omega)}$ est sommable si et seulement si la famille $\big(|f(x)|P(X=x)\big)_{x\in X(\Omega)}$ l'est. On obtient donc la première proposition.

Pour la deuxième proposition, le travail est presque fait : la sommabilité de la famille $\big(P(X=x)\ f(x)\big)_{x\in X(\Omega)}$ permet de faire une sommation par paquets en utilisant la partition $\big(I_y\big)_{y\in Y(\Omega)}$
(on note toujours $Y=f(X)=f\circ X$) :

$$\begin{aligned}
\sum_{x\in X(\Omega)} P(X=x)\ f(x) &= \sum_{y\in Y(\Omega)}\left(\sum_{x\in I_y} P(X=x)f(x)\right)\\
&= \sum_{y\in Y(\Omega}\left(y\sum_{x\in I_y} P(X=x)\right)\\
&= \sum_{y\in Y(\Omega)} yP(Y=y)\\
&= E(Y)\\
&= E\big(f(X)\big)
\end{aligned}$$

Démonstration du critère d'espérance finie On peut noter $X(\Omega)=\{x_i\ ;\ i\in I\}$ (I fini ou dénombrable) et $Y(\Omega)=\{y_j\ ;\ j\in J\}$ (J fini ou dénombrable). La famille $\big(y_j\ P(Y=y_j)\big)_{j\in J}$ est sommable (et c'est une famille de réels positifs, car les y_j sont réels positifs). Le théorème des probabilités totales dit que, pour tout j,

$$y_j\ P(Y=y_j)=\sum_{i\in I} y_j\ P(X=x_i,Y=y_j)$$

et donc, par théorème de sommabilité par paquets, la famille

$$\big(y_j\ P(X=x_i,Y=y_j)\big)_{(i,j)\in I\times J}$$

est sommable. Or, pour tous i et j, deux cas se présentent :

— Si $P(X=x_i,Y=y_j)=0$, alors
$|x_i|P(X=x_i,Y=y_j)\le y_j\ P(X=x_i,Y=y_j)$ (en effet, $0\le 0\ldots$).

— Si $P(X=x_i,Y=y_j)>0$, alors $|x_i|\le y_j$ (car il existe $\omega\in\Omega$ tel que $X(\omega)=x_i$ et $Y(\omega)=y_j$), or $|X|\le Y$.

Bref, pour tous i et j, $|x_i|P(X = x_i, Y = y_j) \leq y_j\ P(X = x_i, Y = y_j)$. Et donc, par majoration, la famille

$$\big(|x_i|\ P(X = x_i, Y = y_j)\big)_{(i,j)\in I\times J}$$

est sommable. Or, par théorème des probabilités totales, pour tout $i \in I$,

$$\sum_{j\in J} |x_i|\ P(X = x_i, Y = y_j) = |x_i|\ P(X = x_i)$$

Donc la famille $(|x_i|\ P(X = x_i))_{i\in I}$ est sommable, ce qui signifie que X est d'espérance finie, et

$$\begin{aligned} E(|X|) &= \sum_{i\in I} |x_i|\ P(X = x_i) \qquad \text{(transfert)} \\ &= \sum_{i\in I}\left(|x_i| \sum_{j\in J} P(X = x_i, Y = y_j)\right) \\ &= \sum_{i\in I}\left(\sum_{j\in J} |x_i| P(X = x_i, Y = y_j)\right) \end{aligned}$$

Utilisons alors la sommation par paquets :

$$\begin{aligned} E(|X|) &= \sum_{(i,j)\in I\times J} |x_i|\ P(X = x_i, Y = y_j) \\ &\leq \sum_{(i,j)\in I\times J} y_j\ P(X = x_i, Y = y_j) \\ &= \sum_{j\in J}\left(y_j \sum_{i\in I} P(X = x_i, Y = y_j)\right) \\ &= E(Y) \end{aligned}$$

Démonstration de la linéarité On commence par remarquer que $E(\lambda X) = \lambda E(X)$ est très simple en distinguant les cas $\lambda = 0$ et $\lambda \neq 0$. Dans ce dernier cas, il suffit de partir de la définition et de remarquer que, pour tout $x \in X(\Omega)$, $P(X = x) = P(\lambda X = \lambda x)$.

Montrons maintenant que, si X et Y sont **à valeurs réelles positives** et d'espérances finies, $X+Y$ est d'espérance finie.
On applique la formule de transfert à la variable aléatoire $Z=(X,Y)$ (c'est bien une variable aléatoire discrète puisque $X(\Omega)$ et $Y(\Omega)$ étant finis ou dénombrables, $X(\Omega)\times Y(\Omega)$ l'est, or $Z(\Omega)\subset X(\Omega)\times Y(\Omega)$), et à l'application $f : (x,y)\mapsto x+y$ qui est bien à valeurs réelles : $X+Y=f(Z)$ est d'espérance finie si et seulement si la famille

$$\Big((x+y)P\big((X,Y)=(x,y)\big)\Big)_{(x,y)\in Z(\Omega)}$$

est sommable. Ce qui équivaut à dire que

$$\Big((x+y)P\big((X,Y)=(x,y)\big)\Big)_{(x,y)\in X(\Omega)\times Y(\Omega)}$$

l'est (si $(x,y)\not\in Z(\Omega)$, $P\big((X,Y)=(x,y)\big)=0$, rajouter des termes nuls ne change pas la sommabilité, ni d'ailleurs la somme). Mais la famille $(xP(X=x))_{x\in X(\Omega)}$ est sommable, de somme $E(X)$. Qui plus est, pour tout $x\in X(\Omega)$,

$$P(X=x)=\sum_{y\in Y(\Omega)} P(X=x,Y=y)$$

par formule des probabilités totales. On en déduit (par paquets, en écrivant la partition $X(\Omega)\times Y(\Omega)=\bigcup_{x\in X(\Omega)}\{x\}\times Y(\Omega))$ que la famille
$\Big(xP\big((X,Y)=(x,y)\big)\Big)_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est sommable de somme $E(X)$.
De même, la famille $\Big(yP\big((X,Y)=(x,y)\big)\Big)_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est sommable de somme $E(Y)$.
On en déduit bien ce qu'on voulait : la famille $\Big((x+y)P\big((X,Y)=(x,y)\big)\Big)_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est sommable, de somme $E(X)+E(Y)$. Ou encore, $X+Y$ est d'espérance finie $X+Y$.

Supposons maintenant X et Y deux variables aléatoires discrètes à valeurs réelles, non nécessairement positives. Si X et Y sont d'espérance finie, $|X|$ et $|Y|$ le sont, donc $|X|+|Y|$ d'après ce qui vient d'être vu. Et donc, comme $|X+Y|\leq|X|+|Y|$, par le critère ci-dessus, $X+Y$ est d'espérance finie. Et en sommant par paquets les familles sommables
$\Big(xP\big((X,Y)=(x,y)\big)\Big)_{(x,y)\in X(\Omega)\times Y(\Omega)}$ et $\Big(yP\big((X,Y)=(x,y)\big)\Big)_{(x,y)\in X(\Omega)\times Y(\Omega)}$ on obtient de nouveau, cette fois sans hypothèse de signe,

$$E(X+Y)=E(X)+E(Y)$$

Fin des démonstrations Le plus dur est fait. En effet, comme il est bien clair que l'espérance d'une variable aléatoire positive est positive, si $X \leq Y$ et X, Y d'espérance finie, $Y - X \geq 0$, donc $E(Y - X) \geq 0$, et donc en utilisant la linéarité, $E(Y) \geq E(X)$.

Mais $X \leq |X|$ et $-X \leq |X|$, donc si X est d'espérance finie, on a $E(X) \leq E(|X|)$ et $-E(X) = E(-X) \leq E(|X|)$. Ces deux inégalités donnent

$$|E(X)| \leq E(|X|)$$

Table des matières