Somme de variables aléatoires, concentration et loi des grands nombres

maths-mde.fr

1 Somme de variables aléatoires

1.1 Linéarité de l'espérance

Définition. Rappels sur l'espérance mathématique

Soit $X = \{x_1; x_2; \dots; x_n\}$ une variable aléatoire discrète telle que $P(X = x_i) = p_i$ pour $1 \leqslant i \leqslant n$, $n \in \mathbb{N}*$.
L'espérance mathématique de X est le nombre $E(X)$ défini par :

$$E(X) = \sum_{k=1}^{n} x_i \times p_i.$$

Exemple. X est une variable aléatoire de loi de probabilité :

$X = x_i$	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_i)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Alors,

$$E(X) = 2 \times \frac{1}{36} + 3 \times \frac{1}{18} + 4 \times \frac{1}{12} + \dots + 11 \times \frac{1}{18} + 12 \times \frac{1}{36}$$

$$\boxed{E(X) = 7.}$$

Propriété. Linéarité de l'espérance

Soient X et Y deux variables aléatoires. L'espérance de la variable aléatoire $X + Y$ est :

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

De plus, pour tout réel a, l'espérance de la variable aléatoire aX est :

$$E(aX) = aE(X).$$

Exemple. Xavier et Yvone vont au restaurant. Xavier hésite de façon équiprobable entre les menus à 18 €, 24 € et 36 €. Yvonne hésite, quant à elle, de façon équiprobable entre les menus à 24 € et 36 €.
On note X et Y les variables aléatoires représentant le prix du menu choisi respectivement par Xavier et Yvonne. Alors, $X = \{18; 24; 36\}$ et $Y = \{24; 36\}$. De plus,

$$\begin{aligned} E(X) &= \frac{1}{3} \times 18 + \frac{1}{3} \times 24 + \frac{1}{3} \times 36 \\ &= 6 + 8 + 12 \\ &= 26 \end{aligned}$$

et

$$\begin{aligned} E(Y) &= \frac{1}{2} \times 24 + \frac{1}{2} \times 36 \\ &= 12 + 18 \\ &= 30. \end{aligned}$$

Ainsi,

$$E(X + Y) = E(X) + E(Y) = 26 + 30 = 56.$$

On peut ainsi dire que le prix moyen de la facture de ce repas est 56 €.

Si Xavier va seul au restaurant 5 fois, et s'il choisit à chaque fois au hasard un menu parmi les 3 cités précédemment, la facture moyenne totale correspondra à $E(5X)$, soit à $5E(X)$, et donc à $5 \times 26 = 130$ €.

1.2 Variance de deux variables aléatoires indépendantes

Définition. Variables aléatoires indépendantes

Soient X et Y deux variables aléatoires.
X et Y sont indépendantes si :

$$P((X = x) \cap (Y = y)) = P(X = x) \times P(Y = y).$$

Propriété

Soit X une variable aléatoire. Pour tout réel a, la variance de la variable aléatoire aX est :

$$V(aX) = a^2 V(X).$$

Son écart-type est donc :

$$\sigma(aX) = |a|\sigma(X).$$

Exemple. Reprenons le cas de Xavier de l'exemple précédent.

$$\begin{aligned} V(X) &= \frac{1}{3}[(18-26)^2 + (24-26)^2 + (36-30)^2] \\ &= 56. \end{aligned}$$

Ainsi,

$$V(5X) = 5^2 V(X) = 25 \times 56 = 1\,400.$$

De plus,

$$\sigma(5X) = 5\sigma(X)) = 5\sqrt{V(X)} = 10\sqrt{14} \approx 37,42.$$

Remarque. Ce dernier résultat peut-être interprété comme une marge d'erreur par rapport à l'espérance : on peut alors dire que pour 5 repas, Xavier paiera 130 € avec une marge d'erreur de 37,42 €.
Cependant, gardons à l'esprit que ce ne sont que des probabilités... et que l'intervalle $[130 - 37,42\ ;\ 130 + 37,42]$ est très approximatif !

Propriété

Soient X et Y deux variables aléatoires *indépendantes*.
La variance de la variable aléatoire $X + Y$ est :

$$V(X + Y) = V(X) + V(Y).$$

1.3 Somme et moyenne de variables indépendantes

1.3.1 Généralités

Propriété. Somme de variables aléatoires indépendantes

Soient X_1, ..., X_n n variables aléatoires indépendantes. On pose alors :

$$S_n = \sum_{k=1}^{n} X_k = X_1 + \cdots + X_n \quad \text{et} \quad M_n = \frac{S_n}{n}.$$

Alors,

— $E(S_n) = \displaystyle\sum_{k=1}^{n} E(X_k)$ et $V(S_n) = \displaystyle\sum_{k=1}^{n} V(X_k)$.

— $E(M_n) = \displaystyle\frac{1}{n}\sum_{k=1}^{n} E(X_k)$ et $V(M_n) = \displaystyle\frac{1}{n^2}\sum_{k=1}^{n} V(X_k)$.

2 Inégalité de Bienaymé-Tchebychev

2.1 Introduction

Considérons deux variables aléatoires X et Y telles que $X \leqslant Y$.
Nous pouvons dans un premier temps écrire que si $X \geqslant 0$ alors $E(X) \geqslant 0$.
De plus,

$$E(Y) - E(X) = E(Y - X)$$

et comme $Y - X \geqslant 0$, on en déduit :

$$E(Y-X) \geqslant 0 \qquad \text{soit :} \qquad E(X) \leqslant E(Y).$$

De ces résultats, on peut démontrer le théorème suivant :

Propriété. Inégalité de Markov

Soit X une variable aléatoire discrète d'espérance finie, et à valeurs positives. Alors :

$$\forall a > 0, \quad P(X \geqslant a) \leqslant \frac{E(X)}{a}.$$

2.2 Inégalité de Bienaymé-Tchebychev

À l'aide de l'inégalité de Markov, on déduit le théorème suivant :

Propriété. Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire d'espérance μ et de variance $V(X)$.
Quel que soit le réel $\delta > 0$,

$$P(|X - \mu| \geqslant \delta) \leqslant \frac{V(X)}{\delta^2}.$$

De ce théorème (fondamental), on peut notamment conclure que pour une variable aléatoire quelconque X d'écart-type σ et d'espérance μ :

$$P(|X - \mu| \geqslant 2\sigma) \leqslant \frac{\sigma^2}{(2\sigma)^2}$$

soit :

$$P(|X - \mu| \geqslant 2\sigma) \leqslant \frac{1}{4}.$$

Cela signifie alors que la probabilité que les valeurs prises par X diffèrent de 2σ de sa moyenne est inférieure à 0,25.

Remarque. Par simulation (manuelle ou informatique), on s'aperçoit en fait que cette probabilité est très souvent majorée par 0,05.

3 Loi faible des grands nombres

3.1 Inégalité de concentration

Propriété

Soient X_1, X_2, $\ldots$, X_n n variables aléatoires discrètes réelles indépendantes ayant toutes la même loi d'espérance μ et de variance σ^2. Alors,

$$\forall \delta > 0, \quad P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geqslant \delta\right) \leqslant \frac{\sigma^2}{n\delta^2}.$$

3.2 Loi des grands nombres

Propriété. Théorème de Khintchine

Soient X_1, ..., X_n n variables aléatoires indépendantes de même loi, d'espérance μ. Alors, pour tout réel $\delta > 0$,

$$\lim_{n \to +\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geqslant \delta\right) = 0.$$

Remarque. Ce dernier théorème, souvent appelé « loi faible des grands nombres », bien qu'au premier abord simpliste puisque résultant de l'inégalité de concentration prise lorsque n tend vers $+\infty$, est très important. Il permet notamment de justifier le principe des sondages. En effet, il permet d'interpréter la probabilité (valeur théorique) comme une fréquence (valeur observée) et présente l'espérance comme une moyenne. Ce théorème signifie que la *moyenne empirique*, c'est-à-dire moyenne calculée sur les valeurs observée sur un échantillon de population, converge vers l'espérance quand la taille de cet échantillon devient très grand.