Sommes de variables aléatoires

Dans ce chapitre, les expériences aléatoires seront toujours associées à des univers ***finis****.*

1 Histoire des mathématiques

La parution de l'Ars Conjectandi de Jacques Bernoulli (1713), reprenant notamment d'anciens travaux de Huygens, marque une rupture dans l'histoire des probabilités. On y trouve la première étude de la distribution binomiale, introduite dans le cadre d'un tirage sans remise pour un modèle d'urne. Un résultat majeur de cet ouvrage est son « théorème d'or », la loi des grands nombres, qui relie fréquences et probabilité, valide le principe de l'échantillonnage et est le premier exemple de « théorème limite » en théorie des probabilités. Le mathématicien français Bienaymé (en 1853, publication en 1867) et le mathématicien russe Tchebychev (en 1867) démontrent l'inégalité qui porte leur nom, en parlant de fréquences d'échantillons plutôt que de variables aléatoires. Ils fournissent ainsi la possibilité d'une démonstration plus simple de la loi des grands nombres.

Au début du XIXe siècle, la modélisation des erreurs de mesure va devenir centrale pour faire de la statistique une science à part entière. Lagrange et Laplace développent une approche probabiliste de la théorie des erreurs. Gauss (1809, 1821), après Legendre (1805), imagine une méthode des moindres carrés qu'il applique avec succès à la prédiction de la position d'un astéroïde. Il y propose de comprendre l'écart-type comme une « erreur moyenne à craindre ».

L'introduction de méthodes statistiques en sociologie est l'œuvre du mathématicien et astronome belge Quételet dans les années 1830. Il réfléchit à la distribution de données autour de la moyenne, ce qui sera approfondi notamment par l'Anglais Galton.

2 Rappels

Définition 1
Une **variable aléatoire finie** est une variable qui ne prend qu'un nombre fini de valeurs x_1, x_2, ... , x_n, chacune avec une certaine probabilité.

On la définit en associant à tout événement élémentaire de l'univers Ω un nombre réel.

La **loi de probabilité** d'une v.a. X est l'application qui à toute valeur x_i prise par X associe la probabilité $P(X = x_i)$. On la représente habituellement à l'aide d'un tableau.

Par exemple, on lance deux dés et on considère la v.a. X correspondant à la somme des deux résultats.

Remarque évidente : La somme des probabilités associées à toutes les valeurs prises par la v.a. vaut 1.

Définition 2
Soit X une variable aléatoire prenant les valeurs x_1, x_2, ... , x_n.
Posons : $p_1 = P(X = x_1)$, $p_2 = P(X = x_2)$, etc.
L'espérance de X, notée $E(X)$ est le nombre :

$$E(X) = p_1 \times x_1 + p_2 \times x_2 + \cdots + p_n \times x_n$$

La variance de X, notée $V(X)$ est le nombre :

$$V(X) = p_1 \times (x_1 - E(X))^2 + p_2 \times (x_2 - E(X))^2 + \cdots + p_n \times (x_n - E(X))^2$$

L'écart-type de X, noté $\sigma(X)$ est la racine carrée de la variance :

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

Notation : Le signe $\sum$ (grand sigma) signifie « somme ». Il est associé à un ou plusieurs indices.

Exemple : $\sum_{k=1}^{4} x_k$ signifie : somme des x_k pour k variant de 1 à 4, c'est à dire : $x_1 + x_2 + x_3 + x_4$.

Avec cette notation, on peut écrire :

$$E(X) = \sum_{k=1}^{n} p_k \times x_k \quad \text{et} \quad V(X) = \sum_{k=1}^{n} p_k \times (x_k - E(X))^2.$$

3 Somme de deux variables aléatoires

On considère deux v.a. X et Y définies sur le **même univers** Ω (fini). Soit a un réel quelconque.

Pour tout événement élémentaire ω, on note $X(\omega)$ la valeur correspondante de X et $Y(\omega)$, la valeur correspondante de Y. On définit une nouvelle variable aléatoire Z en posant : $Z(\omega) = X(\omega) + Y(\omega)$.

La v.a. Z est par définition la somme des v.a. X et Y ; on la note $X + Y$.

Définition 3
La variable aléatoire somme de X et Y est définie pour tout $\omega \in \Omega$ par : $(X+Y)(\omega) = X(\omega) + Y(\omega)$.
De même, la variable aléatoire aX est définie pour tout $\omega \in \Omega$ par : $(aX)(\omega) = aX(\omega)$.

Exemple : on lance deux dés. X représente le résultat du premier dé, Y le résultat du second. La variable aléatoire $X + Y$ représente la somme des deux dés ; elle prend les valeurs 2 ; 3 ; 4 ; ... ; 12. Il y a une façon d'avoir une somme de 2, deux façon d'avoir 3, etc. On trouve donc la loi de probabilité :

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

Théorème 1 (Linéarité de l'espérance)
Soient X et Y deux variables aléatoires et a un réel. Alors :
$E(X + Y) = E(X) + E(Y)$ et $E(aX) = aE(X)$

Démontrons d'abord une propriété :

Propriété 1
$E(X) = \sum_{i=1}^{n} P(X = x_i) \times x_i = \sum_{\omega \in \Omega} P(\omega) \times X(\omega)$

Démonstration :
Soient $x_1 ; ... ; x_n$ les valeurs prises par X. Soit A_1 : l'ensemble des issues ω pour lesquelles $X(\omega) = x_1$; A_2 : l'ensemble des issues ω pour lesquelles $X(\omega) = x_2$; ... ; A_n : l'ensemble des issues ω pour lesquelles $X(\omega) = x_n$.

Les ensembles A_1, ..., A_n forment une partition de Ω et pour tout i, $P(X = x_i) = \sum_{\omega \in A_i} P(\omega)$ (on regroupe les issues qui correspondent à la même valeur de X. Alors :

$\sum_{\omega \in \Omega} P(\omega) \times X(\omega) = \sum_{\omega \in A_1} P(\omega) \times X(\omega) + \sum_{\omega \in A_2} P(\omega) \times X(\omega) + ... + \sum_{\omega \in A_n} P(\omega) \times X(\omega)$

$= \left[\sum_{\omega \in A_1} P(\omega)\right] \times x_1 + \left[\sum_{\omega \in A_2} P(\omega)\right] \times x_2 + ... + \left[\sum_{\omega \in A_n} P(\omega)\right] \times x_n$

$= P(X = x_1) \times x_1 + P(X = x_2) \times x_2 + ... + P(X = x_n) \times x_n = E(X)$ ■

Démonstration du théorème :

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

$$E(aX) = \sum_{\omega\in\Omega} P(\omega)\times aX(\omega) = a\sum_{\omega\in\Omega} P(\omega)X(\omega) = aE(X)$$

Exemple : Reprenons l'exemple précédent.

On a : $E(X) = E(Y) = \dfrac{1}{6}(1+2+3+4+5+6) = \dfrac{21}{6} = \dfrac{7}{2}$

D'autre part :

$$E(X+Y) = \frac{1}{36}\times(1\times 2+2\times 3+3\times 4+4\times 5+5\times 6+6\times 7+5\times 8+4\times 9+3\times 10+2\times 11+1\times 12)$$

$$E(X+Y) = \frac{252}{36} = 7$$

On vérifie que : $E(X)+E(Y) = \dfrac{7}{2}+\dfrac{7}{2} = 7 = E(X+Y)$.

On dit que deux v.a. X et Y sont indépendantes si le fait qu'une prenne une certaine valeur n'a pas d'influence sur l'autre. Formellement, cela signifie que les deux événements $\{X = x_k\}$ et $\{Y = y_i\}$ sont indépendants, autrement dit : $P(\{X = x_k\}\cap\{Y = y_i\}) = P(X = x_k)\times P(Y = y_i)$ (d'après le cours de première).

En particulier, on obtient une succession d'épreuves indépendantes si le résultat de l'une n'a aucune influence sur les suivantes. Exemple typique : des lancers successifs de pièces ou de dés.

Théorème 2 (Propriétés de la variance)
Soient X et Y deux variables aléatoires et a un réel. Alors : $V(aX) = a^2V(X)$
Si de plus X et Y sont **indépendantes**, alors : $V(X+Y) = V(X)+V(Y)$

Exemple : Reprenons l'exemple précédent (somme de deux dés).

$$\begin{aligned}V(X) = V(Y) &= \frac{1}{6}\left[(1-\tfrac{7}{2})^2+(2-\tfrac{7}{2})^2+(3-\tfrac{7}{2})^2+(4-\tfrac{7}{2})^2+(5-\tfrac{7}{2})^2+(6-\tfrac{7}{2})^2\right]\\ &= \frac{1}{6}\left[\frac{25}{4}+\frac{9}{4}+\frac{1}{4}+\frac{1}{4}+\frac{9}{4}+\frac{25}{4}\right] = \frac{1}{6}\times\frac{70}{4} = \frac{35}{12}.\end{aligned}$$

$$\begin{aligned}V(X+Y) &= \frac{1}{36}\,[1\times(2-7)^2+2\times(3-7)^2+3\times(4-7)^2+4\times(5-7)^2+5\times(6-7)^2+6\times(7-7)^2\\ &\qquad +5\times(8-7)^2+4\times(9-7)^2+3\times(10-7)^2+2\times(11-7)^2+1\times(12-7)^2]\\ &= \frac{1}{36}(25+32+27+16+5+0+5+16+27+32+25) = \frac{1}{36}\times 210 = \frac{35}{6}\end{aligned}$$

On voit que $E(X)+E(Y) = \dfrac{35}{12}+\dfrac{35}{12} = \dfrac{35}{6} = E(X+Y)$.

Attention : Il faut impérativement que les v.a. soient indépendantes pour être sûr que les variances s'ajoutent.

Contre-exemple : On lance une pièce et on note 0 pour « pile » et 1 pour « face ». X représente le résultat de la face supérieure et Y le résultat de la face inférieure.

On trouve facilement : $E(X) = E(Y) = \dfrac{1}{2}(0+1) = \dfrac{1}{2}$

et $V(X) = V(Y) = \dfrac{1}{2}\left[\left(0-\tfrac{1}{2}\right)^2+\left(1-\tfrac{1}{2}\right)^2\right] = \dfrac{1}{2}\left(\dfrac{1}{4}+\dfrac{1}{4}\right) = \dfrac{1}{4}$

D'autre part, la variable aléatoire $X+Y$ prend toujours la valeur 1 (donc avec la probabilité 1).

Donc : $E(X+Y) = 1$ et $V(X+Y) = 1\times(1-1)^2 = 0$.

On voit que $V(X+Y)\neq V(X)+V(Y)$. Bien sûr, les v.a. X et Y ne sont pas du tout indépendantes.

4 Généralisation à plusieurs variables

Les théorèmes précédents se généralisent au cas de n variables à l'aide d'un raisonnement par récurrence.

Théorème 3
Soient X_1, X_2, ... X_n, n variables aléatoires sur un univers fini. Soit $S = X_1 + X_2 + ... + X_n$ la variable aléatoire somme. Alors : $E(S) = E(X_1) + E(X_2) + ... + E(X_n)$
Si de plus X_1, X_2, ... X_n sont mutuellement indépendantes, alors : $V(S) = V(X_1)+V(X_2)+...+V(X_n)$

Dire que X_1, X_2, ... X_n sont mutuellement indépendantes signifie que pour tout n-uplet $x_1 ; ... ; x_n$ de valeurs prises par $X_1 ; ... ; X_n$, on a toujours :

$P(X_1 = x_1 \cap X_2 = x_2 \cap ... \cap X_n = x_n) = P(X_1 = x_1) \times P(X_2 = x_2) \times ... \times P(X_n = x_n)$

Un cas particulier très important est celui où les variables X_1, X_2, ... X_n suivent toutes la même loi :

Définition 4
Un **échantillon** de taille n d'une loi de probabilité est une liste $(X_1 ; X_2 ; ... ; X_n)$ de variables aléatoires *indépendantes* suivant chacune cette loi.
On définit la v.a. somme : $S_n = X_1 + X_2 + ... + X_n$ et la v.a. moyenne : $M_n = \dfrac{S_n}{n}$ de cet échantillon.

Les variables X_i suivant toutes la même loi, posons pour tout i : $\mu = E(X_i)$ et $\mathcal{V} = V(X_i)$.

On applique le théorème 3 et on obtient :

$E(S_n) = E(X_1)+E(X_2)+...+E(X_n) = \mu+\mu+...+\mu = n\mu$, et de même : $V(S_n) = V(X_1)+...+V(x_n) = n\mathcal{V}$

On applique à présent les théorèmes 1 et 2 avec $a = \dfrac{1}{n}$ et on obtient :

$E(M_n) = \dfrac{1}{n}E(S_n) = \dfrac{n\mu}{n} = \mu$ et $V(M_n) = \left(\dfrac{1}{n}\right)^2 V(S_n) = \dfrac{n\mathcal{V}}{n^2} = \dfrac{\mathcal{V}}{n}$

Propriété 2
Soit un échantillon de n variables aléatoires *indépendantes* et de même loi, d'espérance μ et de variance $\mathcal{V}$. Soit S_n la v.a. somme et M_n la v.a. moyenne de cet échantillon. Alors :

$$E(S_n) = n\mu \ ; \quad V(S_n) = n\mathcal{V} \ ; \quad E(M_n) = \mu \ ; \quad V(M_n) = \frac{\mathcal{V}}{n}$$

5 Application à la loi binomiale

Rappel : la variable aléatoire T associée à une épreuve de Bernoulli suit la loi de Bernoulli de paramètre p où p est la probabilité d'un succès.

Donc T prend les valeurs 1 et 0 avec les probabilités p et $1 - p$, respectivement.

Par conséquent : $E(T) = (1 - p) \times 0 + p \times 1 = p$

et $V(T) = (1 - p) \times (0 - p)^2 + p \times (1 - p)^2 = p^2(1 - p) + p(1 - p)^2 = p(1 - p)(p + (1 - p)) = p(1 - p)$.

Considérons à présent une variable X qui compte le nombre de succès au cours d'un schéma de Bernoulli ; on sait que X suit la loi binomiale de paramètres n et p, où n est le nombre de répétitions d'une épreuve de Bernoulli et p la probabilité d'un succès.

On constate que X est la somme de n variables aléatoires *indépendantes* suivant chacune une loi de Bernoulli de paramètre p : $X = T_1 + T_2 + ... + T_n$ et pour tout i : $E(T_i) = p$ et $V(T_i) = p(1 - p)$.

D'après la propriété 2, on peut écrire :

$E(X) = nE(T_i) = np$

et $V(X) = nV(T_i) = np(1 - p)$

Propriété 3
Soit X une variable aléatoire suivant la loi binomiale de paramètres n et p. Alors :

$$E(X) = np \quad \text{et} \quad V(X) = np(1 - p)$$