Somme de variables aléatoires

Prélude - Historique et contexte

Additionner des variables aléatoires, c'est par exemple cumuler les gains de plusieurs parties, les durées de plusieurs tâches ou les erreurs de plusieurs mesures. Au XVIII^{e} siècle, Abraham de Moivre étudie déjà la somme de nombreux lancers de pièces. Ces outils permettent de comprendre pourquoi la moyenne de nombreuses mesures est plus fiable qu'une mesure isolée : c'est le point de départ de la statistique moderne.

I Transformation affine

Rappel

Si X prend les valeurs $x_1, \dots, x_n$ avec les probabilités $p_1, \dots, p_n$:

$$E(X) = \sum_{i=1}^{n} p_i x_i \qquad V(X) = \sum_{i=1}^{n} p_i \big(x_i - E(X)\big)^2 = E\left(X^2\right) - E(X)^2 \qquad \sigma(X) = \sqrt{V(X)}.$$

Propriété

Pour tous réels a et b :

$$E(aX+b) = aE(X) + b \qquad V(aX+b) = a^2 V(X) \qquad \sigma(aX+b) = |a|\,\sigma(X).$$

Exemple

Un jeu rapporte X euros avec $E(X) = 2$ et $V(X) = 9$. Si tous les gains sont multipliés par 3 et qu'on paie 5 € pour jouer, le gain net est $Y = 3X - 5$: $E(Y) = 1$, $V(Y) = 81$, $\sigma(Y) = 9$.

II Somme de deux variables aléatoires

Définition

Soit X et Y deux variables aléatoires définies sur un même univers. $X + Y$ est la variable aléatoire qui à chaque issue associe la somme des valeurs prises par X et Y.

Propriété - Linéarité de l'espérance

$$E(X+Y) = E(X) + E(Y) \qquad E(aX+bY) = aE(X) + bE(Y).$$

Propriété - Variance d'une somme (admise)

Si X et Y sont **indépendantes** : $V(X+Y) = V(X) + V(Y)$.
Attention : ce n'est pas vrai en général, et $\sigma(X+Y) \neq \sigma(X) + \sigma(Y)$.

Exemple

On lance deux dés équilibrés à six faces. Pour un dé, $E(D) = 3{,}5$ et $V(D) = \frac{35}{12}$. La somme $S = D_1 + D_2$ vérifie $E(S) = 7$ et, les dés étant indépendants, $V(S) = \frac{35}{6}$.

III Échantillon et applications

Définition - Échantillon

Un **échantillon de taille** n d'une loi de probabilité est une liste $(X_1, \dots, X_n)$ de variables aléatoires indépendantes suivant toutes cette loi. On note $S_n = X_1 + \cdots + X_n$ et $M_n = \dfrac{S_n}{n}$ la moyenne.

Propriété

Si $E(X_i) = \mu$ et $V(X_i) = \sigma^2$:

$$E(S_n) = n\mu \qquad V(S_n) = n\sigma^2 \qquad \sigma(S_n) = \sqrt{n}\sigma$$

$$E(M_n) = \mu \qquad V(M_n) = \frac{\sigma^2}{n} \qquad \sigma(M_n) = \frac{\sigma}{\sqrt{n}}.$$

La moyenne a la même espérance que chaque X_i, mais son écart type est divisé par $\sqrt{n}$: elle est moins dispersée.

Propriété - Espérance et variance d'une loi binomiale

Si X suit $\mathscr{B}(n\ ;\ p)$, X est la somme de n variables de Bernoulli indépendantes de paramètre p, d'espérance p et de variance $p(1-p)$. Donc :

$$\boxed{E(X) = np \qquad V(X) = np(1-p) \qquad \sigma(X) = \sqrt{np(1-p)}}$$

Exemple

Un QCM de 20 questions à 4 choix est rempli au hasard. Le nombre X de bonnes réponses suit $\mathscr{B}(20\ ;\ 0{,}25)$: $E(X) = 5$, $V(X) = 3{,}75$. Avec $+1$ point par bonne réponse et $-0{,}5$ par mauvaise, la note est $N = X - 0{,}5(20 - X) = 1{,}5X - 10$, d'espérance $-2{,}5$.