Variable aléatoire, espérance, écart-type, le cas discret

Définitions

1) Soit une expérience aléatoire $\mathcal{E}$ et l'ensemble de ses issues (fini ou dénombrable) dénoté par U, son univers associé. Une *variable aléatoire* réelle X est une application de U dans $\mathbb{R}$ qui associe un nombre réel x_i à une issue donnée. La notation $X = x_i$ correspond à l'événement de $\wp(U)$ admettant pour image x_i. Ainsi, si une probabilité est définie sur $\wp(U)$ alors on notera la probabilité de l'événement admettant x_i pour image par $P(X = x_i)$ ou plus simplement (s'il n'y a pas de confusion possible) par p_i.

On appelle *loi de probabilité* (ou *distribution*) l'ensemble des couples $(x_i ; p_i)$.

Exemple. $\mathcal{E}$ est « lancer deux dés non pipés », $U = \{(i\,;j)$ où $1 \le i \le 6$ et $1 \le j \le 6\}$.

$X : (i\,;j) \to (-1)^{i+j}(i+j)$. On gagne cette somme si l'image est > 0 sinon on la perd !

$P(X=12) = 1/36 = P(X=2)$; $P(X=-11) = 1/18$ etc.

Question. Ce jeu est-il équitable ? Pour y répondre déterminons le « gain moyen ».

2) *L'espérance mathématique* d'une variable aléatoire X est la somme des images x_i pondérées par leur probabilité $P(X = x_i)$. En somme c'est une sorte de *valeur moyenne* obtenue après n itérations de l'expérience aléatoire $\mathcal{E}$.

Elle se note et se calcule par : $E(X) = \sum\limits_{1\le i\le n} x_i \cdot P(X = x_i) = \sum\limits_{1\le i\le n} x_i \cdot p_i$ ou encore μ_X

Exercice. Montrer que pour le jeu précédent $E(X) = 0$ et donc qu'il est équitable.

3) Pour mesurer la dispersion autour de la moyenne on introduit la notion de *variance* et *d'écart type* des concepts à la base issus des sciences *statistiques*.

Par définition la variance se note et s'obtient ainsi : $V(X) = E\left[(X-\mu_X)^2\right] = \sum\limits_i (x_i - \mu_X)^2 \cdot p_i$

Au verso figure une preuve (élémentaire) de la formule de König : $V(X) = E(X^2) - (E(X))^2$

Preuve de la formule de König.

$$V(X) = E\left[(X-\mu_X)^2\right] = \sum_i (x_i-\mu_X)^2 \cdot p_i = \sum_i x_i^2 \cdot p_i - 2\mu_X \sum_i x_i \cdot p_i + \mu_X^2 \sum_i p_i =$$

$$= E(X^2) - 2\mu_X^2 + \mu_X^2 = E(X^2) - (E(X))^2 = \mu_{X^2} - \mu_X^2$$

On appelle *écart type* le nombre $\sigma = \sqrt{V(X)}$. Ce nombre correspond encore mieux à une mesure de dispersion autour de μ puisqu'il admet les mêmes unités que μ.

4) Si X et Y sont deux variables aléatoires discrètes alors on définit $p_{ij} = P(X = x_i, Y = y_j)$, la *probabilité conjointe*. On dit que X et Y sont *indépendantes* si $\forall i$ et j on a $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$ ou plus simplement $p_{ij} = p_i \cdot p_j$.

Par définition on pose : $E(X+Y) = \sum\limits_{i,j} p_{ij}(x_i + y_j)$ et $E(X \cdot Y) = \sum\limits_{i,j} p_{ij}(x_i y_j)$

Théorème. Soit X et Y des variables aléatoires discrètes relatives à une épreuve $\mathcal{E}$. Alors

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

2) Si X et Y sont indépendantes alors $E(X\cdot Y)=E(X)\cdot E(Y)$

3) Si X et Y sont indépendantes alors $V(X+Y)=V(X)+V(Y)$

Preuves.

1) $E(X+Y)=\sum_{i,j}(x_i+y_j)p_{ij}=\sum_{i,j}(x_ip_{ij}+y_jp_{ij})=\sum_i x_i\sum_j p_{ij}+\sum_j y_j\sum_i p_{ij}=\sum_i x_ip_i+\sum_j y_jp_j=E(X)+E(Y)$

Remarque. Ci-dessus nous avons utilisé le fait que : $E(X)=\sum_i x_ip_i=\sum_{i,j}x_ip_{ij}$ où i et j parcourent toutes les valeurs possibles des variables aléatoires X et Y.

2)

$$\begin{aligned}E(XY)&=\sum_{i,j}x_iy_jp_{ij}=\sum_{i,j}x_iy_jP(x=x_i,y=y_j)\underset{indep.}{\overset{\text{var.}}{=}}\sum_{i,j}x_iy_jP(x=x_i)P(y=y_j)\\&=\sum_i x_iP(x=x_i)\cdot\sum_j y_jP(y=y_j)=E(X)\cdot E(Y)\end{aligned}$$

3)

$$\begin{aligned}V(X+Y)&\underset{König}{\overset{par}{=}}\sum_{i,j}(x_i+y_j)^2p_{ij}-\mu_{X+Y}^2=\sum_{i,j}(x_i^2+2x_iy_j+y_j^2)p_{ij}-(E(X+Y))^2\\&\underset{esper.}{\overset{linéar.}{=}}\sum_{i,j}x_i^2p_{ij}+2\sum_{i,j}x_iy_jp_{ij}+\sum_{i,j}y_j^2p_{ij}-(E(X)+E(Y))^2\\&=\sum_{i,j}x_i^2p_{ij}+2\sum_{i,j}x_iy_jp_{ij}+\sum_{i,j}y_j^2p_{ij}-\mu_X^2-2\mu_X\mu_Y-\mu_Y^2\\&\underset{Remar}{\overset{par}{=}}\sum_i x_i^2p_i+2\sum_{i,j}x_iy_jp_{ij}+\sum_j y_j^2p_j-\mu_X^2-2\mu_X\mu_Y-\mu_Y^2\\&\underset{indep}{\overset{par}{=}}\sum_i x_i^2p_i+2\sum_{i,j}x_iy_jp_ip_j+\sum_j y_j^2p_j-\mu_X^2-2\mu_X\mu_Y-\mu_Y^2\\&=\sum_i x_i^2p_i+2\sum_i x_ip_i\sum_j y_jp_j+\sum_j y_j^2p_j-\mu_X^2-2\mu_X\mu_Y-\mu_Y^2\\&=\mu_{X^2}+2\mu_X\mu_Y+\mu_{Y^2}-\mu_X^2-2\mu_X\mu_Y-\mu_Y^2\\&=\mu_{X^2}-\mu_X^2+\mu_{Y^2}-\mu_Y^2\\&\underset{König}{\overset{par}{=}}V(X)+V(Y)\end{aligned}$$

Soit une épreuve « genre binaire » (réussie ou ratée) répétée n-fois (épreuves identiques).
On note X_i la variable aléatoire prenant la valeur 1 si 'réussi', 0 sinon, au i-ième essai.
Une telle variable aléatoire s'appelle *de Bernoulli* (1685).
Supposons $P(X_i=1)=p$ (un nombre compris entre 0 et 1) et donc $P(X_i=0)=1-p$.
Posons $X=X_1+X_2+X_3+\cdots+X_n$ la variable aléatoire qui somme le nombre de réussites.
Alors $E(X)=E(X_1+X_2+X_3+\cdots+X_n)=E(X_1)+\cdots+E(X_n)=np$
et $V(X)=V(X_1+\cdots+X_n)=V(X_1)+\cdots+V(X_n)=\left(\mu_{X_1^2}-\mu_{X_1}^2\right)n=(p-p^2)n=(1-p)pn$

La variable aléatoire qui compte le nombre k de réussites sur n épreuves s'appelle une *loi binomiale (de paramètre n et p)*. On la note et elle s'obtient par :

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

Exercice. Montrer que **1)** $\sum_{k=0}^{n}P(X=k)=1$ **2)** $\mu=np$ **3)** $\sigma^2=V(X)=np(1-p)$