OPERATIONS SUR LES VARIABLES ALEATOIRES DISCRETES

I - Couples de variables aléatoires discrètes

Il s'agit de mettre en place des outils pour comparer deux variables aléatoires (ce que l'on pourra généraliser à plusieurs), ou pour déterminer la loi d'une variable aléatoire lorsque l'on ne la connaît que par rapport à d'autres.

1) Loi conjointe

<u>Définition</u> : Etant données deux variables aléatoires discrètes X et Y définies sur le même univers Ω telles que $X(\Omega)=\{x_i \,/\, i\in I\}$ et $Y(\Omega)=\{y_j \,/\, j\in J\}$, on appelle loi conjointe du couple (X,Y) l'ensemble formé par les couples (x_i,y_j) et par les probabilités $p_{i,j}=P[(X=x_i)\cap(Y=y_j)]$ pour $i\in I$ et $j\in J$. Les lois de X et de Y sont appelées lois marginales du couple (X,Y).
Ces nombres $p_{i,j}$ appartiennent à $[0,1]$ et $\displaystyle\sum_{i,j} p_{i,j}=1$.

Ils sont donnés dans un tableau lorsque X et Y sont finies.
<u>Exemple 1</u> : Une urne contient 4 boules rouges, 3 blanches et 2 vertes. On tire simultanément 2 boules et on note X le nombre de boules rouges tirées et Y le nombre de boules blanches tirées. Donc $X(\Omega)=Y(\Omega)=\{0,1,2\}$.

Il y a équiprobabilité et $\mathrm{Card}(\Omega)=\dbinom{9}{2}=36$.

Si $i+j\geq 3$, $P[(X=i)\cap(Y=j)]=0$ car on ne tire que 2 boules.

Sinon $P[(X=i)\cap(Y=j)]=\dfrac{\dbinom{4}{i}\dbinom{3}{j}\dbinom{2}{2-i-j}}{36}$ car on tire i rouges parmi 4, j blanches parmi 3 et le reste ($2-i-j$) parmi les 2 vertes.
On résume ces résultats dans un tableau à double entrée :

X \ Y	$j=0$	$j=1$	$j=2$	$P(X=i)$
$i=0$	$\frac{1}{36}$	$\frac{6}{36}=\frac{1}{6}$	$\frac{3}{36}=\frac{1}{12}$	$\frac{10}{36}=\frac{5}{18}$
$i=1$	$\frac{8}{36}=\frac{2}{9}$	$\frac{12}{36}=\frac{1}{3}$	0	$\frac{20}{36}=\frac{5}{9}$
$i=2$	$\frac{6}{36}=\frac{1}{3}$	0	0	$\frac{6}{36}=\frac{1}{6}$
$P(Y=j)$	$\frac{15}{36}=\frac{5}{12}$	$\frac{18}{36}=\frac{1}{2}$	$\frac{3}{36}=\frac{1}{12}$	1

Dans le cas de variables infinies, on donne la formule générale.
<u>Exemple 2</u> : On reprend la même urne (4 boules rouges, 3 blanches et 2 vertes). Mais on effectue des tirages successifs d'une boule avec remise. X est le rang de la première boule verte et Y le rang de la deuxième. Donc $X(\Omega)=[\![1,+\infty[\![$ et $Y(\Omega)=[\![2,+\infty[\![$.
On remarque d'abord que $X<Y$. Donc $P[(X=i)\cap(Y=j)]=0$ si $i\geq j$.
Ensuite, en notant V_k l'événement « la $k^{\text{ème}}$ boule tirée est verte », si $1\leq i<j$:

$(X=i)\cap(Y=j)=\overline{V_1}\cap\ldots\cap\overline{V_{i-1}}\cap V_i\cap\overline{V_{i+1}}\cap\ldots\cap\overline{V_{j-1}}\cap V_j$

Puisqu'il y a remise, les tirages sont indépendants, et donc :

$P[(X=i)\cap(Y=j)]=P(\overline{V_1})\ldots P(\overline{V_{i-1}})P(V_i)P(\overline{V_{i+1}})\ldots P(\overline{V_{j-1}})P(V_j)$.

Donc si $1\le i<j$: $P[(X=i)\cap(Y=j)]=\left(\frac{2}{9}\right)^2\left(\frac{7}{9}\right)^{j-2}$.

2) Lois marginales

Définition : Etant données deux variables aléatoires discrètes X et Y définies sur le même univers Ω, les lois de X et de Y sont appelées lois marginales du couple (X,Y).

On complète le tableau en remarquant, par exemple, que $X=0$ si et seulement si $X=0$ et $Y=0$, ou $X=0$ et $Y=1$, ou $X=0$ et $Y=2$. Avec la loi conjointe, on peut donc calculer $P(X=0)$: il suffit d'additionner tous les éléments de la première ligne. Le même raisonnement est utilisé pour toutes les lignes et toutes les colonnes.

Plus généralement, les événements $(X=x_i)$ forment un système complet d'événements, de même d'ailleurs que les événements $(Y=y_j)$. Donc :

La loi conjointe du couple (X,Y) permet de déterminer les deux lois marginales :

$$\forall i\in I\quad P(X=x_i)=\sum_{j\in J}P[(X=x_i)\cap(Y=y_j)]$$

$$\forall j\in J\quad P(Y=y_j)=\sum_{i\in I}P[(X=x_i)\cap(Y=y_j)]$$

On peut alors en déduire les autres caractéristiques de X et Y.

Dans le cas de variables discrètes infinies, il s'agit de sommes de séries.

Exemple 2 : $\forall i\ge 1\quad P(X=i)=\sum_{j=2}^{+\infty}P[(X=i)\cap(Y=j)]=\sum_{j=i+1}^{+\infty}\left(\frac{2}{9}\right)^2\left(\frac{7}{9}\right)^{j-2}$.

$$P(X=i)=\left(\frac{2}{9}\right)^2\sum_{j=i+1}^{+\infty}\left(\frac{7}{9}\right)^{j-2}=\left(\frac{2}{7}\right)^2\sum_{j=i+1}^{+\infty}\left(\frac{7}{9}\right)^{j}=\left(\frac{2}{7}\right)^2\left[\frac{1}{1-\frac{7}{9}}-\sum_{j=0}^{i}\left(\frac{7}{9}\right)^{j}\right].$$

$$P(X=i)=\left(\frac{2}{7}\right)^2\left[\frac{9}{2}-\frac{1-\left(\frac{7}{9}\right)^{i+1}}{1-\frac{7}{9}}\right]=\frac{9}{2}\left(\frac{2}{7}\right)^2\left(\frac{7}{9}\right)^{i+1}=\frac{2}{9}\left(\frac{7}{9}\right)^{i-1}.$$

On retrouve le fait que X suive la loi géométrique de paramètre $\frac{2}{9}$.

$$\forall j\ge 2\quad P(Y=j)=\sum_{i=1}^{+\infty}P[(X=i)\cap(Y=j)]=\sum_{i=1}^{j-1}\left(\frac{2}{9}\right)^2\left(\frac{7}{9}\right)^{j-2}.$$

Donc : $\forall j\ge 2\quad P(Y=j)=(j-1)\left(\frac{2}{9}\right)^2\left(\frac{7}{9}\right)^{j-2}$.

3) Lois conditionnelles

Inversement, les lois marginales ne permettent pas en général de déterminer la loi conjointe, puisque dans la plupart des cas, il n'y a pas de lien entre $P(A\cap B)$, $P(A)$ et $P(B)$. L'un des rares cas où l'on peut exprimer $P(A\cap B)$ en fonction de $P(A)$ et $P(B)$

est l'indépendance : $P(A \cap B) = P(A)P(B)$. Dans les autres cas, il faut au moins connaître $P_B(A)$ ou $P_A(B)$.

<u>Définition</u> :
Si $P(Y = y_j) \neq 0$, la loi conditionnelle de X sachant $(Y = y_j)$ est déterminée par les nombres $P_{(Y=y_j)}(X = x_i) = \dfrac{P[(X = x_i) \cap (Y = y_j)]}{P(Y = y_j)}$ pour tout $i \in I$.
Si $P(X = x_i) \neq 0$, la loi conditionnelle de Y sachant $(X = x_i)$ est déterminée par les nombres $P_{(X=x_i)}(Y = y_j) = \dfrac{P[(X = x_i) \cap (Y = y_j)]}{P(X = x_i)}$ pour tout $j \in J$.

On peut remarquer que si $P(Y = y_j) \neq 0$ et si $P(X = x_i) \neq 0$:

$$P[(X = x_i) \cap (Y = y_j)] = P_{(Y=y_j)}(X = x_i)P(Y = y_j) = P_{(X=x_i)}(Y = y_j)P(X = x_i)$$

Les lois conditionnelles de X sachant $(Y = y_j)$ ou de Y sachant $(X = x_i)$ peuvent donc permettre de déterminer (en utilisant la formule des probabilités totales) la loi conjointe du couple (X,Y) ou une loi marginale si l'on connaît l'autre.

Les lois conditionnelles permettent de déterminer la loi conjointe et les lois marginales du couple (X,Y) quand on connaît une des lois marginales :

$$P[(X = x_i) \cap (Y = y_j)] = P_{(Y=y_j)}(X = x_i)P(Y = y_j) = P_{(X=x_i)}(Y = y_j)P(X = x_i)$$

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

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

<u>*Exemple 3*</u> : On admet que le nombre X de voyageurs arrivant dans la station de RER à Vincennes entre 17h45 et 18h suit une loi de Poisson de paramètre 30, et que 3 voyageurs sur 5 se dirigent vers Paris, leurs comportements étant indépendants. Il s'agit de déterminer la loi du nombre Y de voyageurs qui prennent à Vincennes le RER pour Paris entre 17h45 et 18h. Tout d'abord : $Y(\Omega) = \mathbb{N}$.

On sait que $X(\Omega) = \mathbb{N}$ et que : $P(X = i) = \dfrac{30^i}{i!}e^{-30}$ pour tout i.

La loi conditionnelle de Y sachant $(X = i)$ est la loi binomiale $\mathcal{B}\left(i, \dfrac{3}{5}\right)$ car si le nombre de voyageurs est i, on a une épreuve de Bernoulli (le voyageur choisit une direction) de succès « il va à Paris », répétée i fois de manière indépendante et dans les mêmes conditions. Donc :

- Si $0 \leq j \leq i$: $P_{(X=i)}(Y = j) = \dbinom{i}{j}\left(\dfrac{3}{5}\right)^j\left(\dfrac{2}{5}\right)^{i-j}$.
- Si $j > i$: $P_{(X=i)}(Y = j) = 0$.

Donc : $P(Y = j) = \displaystyle\sum_{i=0}^{+\infty} P_{(X=i)}(Y = j)P(X = i) = \sum_{i=j}^{+\infty} \binom{i}{j}\left(\frac{3}{5}\right)^j\left(\frac{2}{5}\right)^{i-j}\frac{30^i}{i!}e^{-30}$.

$$P(Y = j) = \left(\frac{3}{5}\right)^j\left(\frac{2}{5}\right)^{-j}e^{-30}\sum_{i=j}^{+\infty}\frac{i!}{j!(i-j)!}\left(\frac{2}{5}\right)^i\frac{30^i}{i!} = \left(\frac{3}{2}\right)^j\frac{e^{-30}}{j!}\sum_{i=j}^{+\infty}\frac{1}{(i-j)!}(12)^i$$

$$P(Y=j)=\left(\frac{3}{2}\right)^j\frac{e^{-30}}{j!}\sum_{k=0}^{+\infty}\frac{1}{k!}(12)^{k+j}=(18)^j\frac{e^{-30}}{j!}\sum_{k=0}^{+\infty}\frac{(12)^k}{k!}=(18)^j\frac{e^{-30}}{j!}e^{12}.$$

Donc : $P(Y=j)=\dfrac{18^j}{j!}e^{-18}$ pour tout j.

La loi de Y est la loi de Poisson de paramètre 18.

4) Indépendance

Définition : Les deux variables aléatoires X et Y sont indépendantes si pour tout i de I et tout j de J : $P[(X=x_i)\cap(Y=y_j)]=P(X=x_i)P(Y=y_j)$.

Dans ce cas, les deux lois marginales permettent de trouver la loi conjointe : la loi conjointe est le produit des deux lois marginales.
Dans le cas du premier exemple, X et Y ne sont pas indépendantes car par exemple : $P[(X=1)\cap(Y=2)]\neq P(X=1)P(Y=2)$ (Un contre-exemple suffit).

Exemple : X et Y sont les résultats de deux lancers de dés honnêtes.

Pour tous i et j appartenant à $\{1,...,6\}$: $P(X=i)=P(Y=j)=\dfrac{1}{6}$.

La loi conjointe est donnée par : $P[(X=i)\cap(Y=j)]=\dfrac{1}{36}=P(X=i)P(Y=j)$

Pour démontrer l'indépendance, il faut vérifier la propriété pour tous les i et j.
La notion d'indépendance peut être étendue à plusieurs variables $X_1,...,X_n$. Comme pour les événements, on définit deux types d'indépendance.

Les variables aléatoires $X_1,...,X_n$ sont deux à deux indépendantes si quels que soient i et j distincts pris dans $[\![1,n]\!]$, les variables X_i et X_j sont indépendantes.
Les variables aléatoires $X_1,...,X_n$ sont mutuellement indépendantes si quelles que soient les valeurs $x_1\in X_1(\Omega)$, …, $x_n\in X_n(\Omega)$, on a :

$$P[(X_1=x_1)\cap...\cap(X_n=x_n)]=P(X_1=x_1)\times...\times P(X_n=x_n).$$

Des variables mutuellement indépendantes sont deux à deux indépendantes, mais la réciproque est fausse.

II – Opérations sur les variables aléatoires discrètes

1) Généralités

Soit $Z=f(X,Y)$ une variable aléatoire obtenue à partir du couple (X,Y) : par exemple, $Z=X+Y$ ou $Z=XY$ ou $Z=\mathrm{Max}(X,Y)$ ou $Z=\mathrm{Min}(X,Y)$ ou …
La loi conjointe du couple (X,Y) permet de déterminer la loi de la variable Z et son espérance. On commence par déterminer l'ensemble $Z(\Omega)$ des valeurs prises par Z.

Pour chaque valeur $z\in Z(\Omega)$, il peut exister plusieurs couples (i,j) tels que $i\in I$, $j\in J$ et $f(x_i,y_j)=z$: on notera $K(z)$ l'ensemble de ces couples.
La loi de probabilité de $Z=f(X,Y)$ est définie par :

$$\forall z\in Z(\Omega)\quad P(Z=z)=\sum_{(i,j)\in K(z)}P[(X=x_i)\cap(Y=y_j)]$$

L'espérance de Z est alors : $E(Z)=\displaystyle\sum_{(i,j)\in I\times J}f(x_i,y_j)P[(X=x_i)\cap(Y=y_j)]$.

En effet, pour tout $z\in Z(\Omega)$:

$$zP(Z=z)=\sum_{(i,j)\in K(z)}zP[(X=x_i)\cap(Y=y_j)]=\sum_{(i,j)\in K(z)}f(x_i,y_j)P[(X=x_i)\cap(Y=y_j)]$$

Et l'ensemble des $K(z)$ pour $z \in Z(\Omega)$ représente une partition de $I \times J$.

<u>Exemple 1</u> : Déduire de la loi du tableau de l'exemple précédent la loi de Z nombre de boules bleues tirées. On a la relation : $X + Y + Z = 2$, donc $Z = 2 - X - Y$.

L'ensemble des valeurs prises par Z est : $Z(\Omega) = \{0,1,2\}$.

$Z = 0 \Leftrightarrow X + Y = 2$. Donc :

$$P(Z = 0) = P[(X = 0) \cap (Y = 2)] + P[(X = 1) \cap (Y = 1)] + P[(X = 2) \cap (Y = 0)] = \frac{7}{12}.$$

$$P(Z = 1) = P(X + Y = 1) = P[(X = 0) \cap (Y = 1)] + P[(X = 1) \cap (Y = 0)] = \frac{7}{18}.$$

$$P(Z = 2) = P(X + Y = 0) = P[(X = 0) \cap (Y = 0)] = \frac{1}{36}.$$

Il y a deux méthodes pour calculer $E(Z)$:

- Directement, à partir de la loi de Z : $E(Z) = 0 \times \frac{7}{12} + 1 \times \frac{7}{18} + 2 \times \frac{1}{36} = \frac{4}{9}$.
- A partir de la loi conjointe du couple (X,Y) :

$$E(Z) = \sum_{i=0}^{2} \sum_{j=0}^{2} (2 - i - j) P[(X = i) \cap (Y = j)]$$

$$\begin{aligned} E(Z) = {} & 2P[(X = 0) \cap (Y = 0)] + 1P[(X = 0) \cap (Y = 1)] + 0P[(X = 0) \cap (Y = 2)] \\ & + 1P[(X = 1) \cap (Y = 0)] + 0P[(X = 1) \cap (Y = 1)] - 1P[(X = 1) \cap (Y = 2)] \\ & + 0P[(X = 2) \cap (Y = 0)] - 1P[(X = 2) \cap (Y = 1)] - 2P[(X = 2) \cap (Y = 2)] \end{aligned}$$

$$E(Z) = 2 \times \frac{1}{36} + 1 \times \frac{6}{36} + 1 \times \frac{8}{36} = \frac{16}{36} = \frac{4}{9}.$$

<u>Exemple 2</u> : On peut s'intéresser par exemple à la variable $Z = Y - X$ qui représente le temps entre la première et la deuxième boule verte. D'abord : $Z(\Omega) = \mathbb{N}^*$.

Pour tout entier $k \geq 1$: $K(k) = \{(i, j) \in (\mathbb{N}^*)^2 \,/\, i \geq 1,\ j \geq 2 \text{ et } k = j - i\}$.

Donc : $P(Z = k) = \displaystyle\sum_{i=1}^{+\infty} P[(X = i) \cap (Y = i + k)] = \sum_{i=1}^{+\infty} \left(\frac{2}{9}\right)^2 \left(\frac{7}{9}\right)^{i+k-2} = \left(\frac{2}{7}\right)^2 \left(\frac{7}{9}\right)^k \sum_{i=1}^{+\infty} \left(\frac{7}{9}\right)^i$

$$P(Z = k) = \left(\frac{2}{7}\right)^2 \left(\frac{7}{9}\right)^k \left(\frac{1}{1 - \frac{7}{9}} - 1\right) = \frac{2}{7} \left(\frac{7}{9}\right)^k.$$

Pour calculer $E(Z)$, deux méthodes :

- Directement : $E(Z) = \displaystyle\sum_{k=1}^{+\infty} k P(Z = k) = \frac{2}{9} \sum_{k=1}^{+\infty} k \left(\frac{7}{9}\right)^{k-1} = \frac{2}{9} \times \frac{1}{\left(1 - \frac{7}{9}\right)^2} = \frac{9}{2}$.
- A partir de la loi conjointe du couple (X,Y) :

$$E(Z) = \sum_{i=1}^{+\infty} \sum_{j=2}^{+\infty} (j - i) P[(X = i) \cap (Y = j)] = \sum_{i=1}^{+\infty} \sum_{j=i+1}^{+\infty} (j - i) \left(\frac{2}{9}\right)^2 \left(\frac{7}{9}\right)^{j-2}$$

$$E(Z) = \left(\frac{2}{9}\right)^2 \sum_{i=1}^{+\infty} \sum_{k=1}^{+\infty} k \left(\frac{7}{9}\right)^{k+i-2} = \left(\frac{2}{9}\right)^2 \sum_{i=1}^{+\infty} \left(\frac{7}{9}\right)^{i-1} \left[\sum_{k=1}^{+\infty} k \left(\frac{7}{9}\right)^{k-1}\right] \qquad (k = j - i)$$

$$E(Z) = \left(\frac{2}{9}\right)^2 \sum_{i=1}^{+\infty} \left(\frac{7}{9}\right)^{i-1} \times \frac{1}{\left(1 - \frac{7}{9}\right)^2} = \sum_{i=1}^{+\infty} \left(\frac{7}{9}\right)^{i-1} = \sum_{k=0}^{+\infty} \left(\frac{7}{9}\right)^k = \frac{1}{1 - \frac{7}{9}} = \frac{9}{2}.$$

2) Somme de variables aléatoires

C'est un cas particulier : $Z = X + Y$. Donc $Z(\Omega) = \left\{x_i + y_j / (i, j) \in I \times J\right\}$.

La loi de probabilité de $Z = X + Y$ est définie par :
$\forall z \in Z(\Omega) \quad P(Z = z) = \sum_{i \in I} P[(X = x_i) \cap (Y = z - x_i)] = \sum_{j \in J} P[(Y = y_j) \cap (X = z - y_j)]$

Pour la somme de deux variables aléatoires, on a vu que $E(X + Y) = E(X) + E(Y)$.
Pour la variance, en posant $Z = X + Y$:
$Z^2 = X^2 + Y^2 + 2XY$ et donc : $E(Z^2) = E(X^2) + E(Y^2) + 2E(XY)$.
$[E(Z)]^2 = [E(X) + E(Y)]^2 = [E(X)]^2 + [E(Y)]^2 + 2E(X)E(Y)$.
Donc : $V(Z) = V(X) + V(Y) + 2[E(XY) - E(X)E(Y)]$.

<u>Définition</u> : La covariance du couple (X,Y) est : $\text{cov}(X,Y) = E(XY) - E(X)E(Y)$.

Le calcul de l'espérance $E(XY)$ se fait suivant la règle précédente : $E(XY) = \sum_{(i,j) \in I \times J} x_i y_j P[(X = x_i) \cap (Y = y_j)]$.

<u>Exemple 1</u> : $E(XY) = \sum_{i=0}^{2} \sum_{j=0}^{2} ijP[(X = i) \cap (Y = j)] = \frac{1}{3}$ (un seul terme non nul).

$E(X) = 0 \times \frac{10}{36} + 1 \times \frac{20}{36} + 2 \times \frac{6}{36} = \frac{32}{36} = \frac{8}{9}$ et $E(Y) = 0 \times \frac{15}{36} + 1 \times \frac{18}{36} + 2 \times \frac{3}{36} = \frac{24}{36} = \frac{2}{3}$

Et : $\text{cov}(X,Y) = \frac{1}{3} - \frac{8}{9} \times \frac{2}{3} = -\frac{7}{27}$.

<u>Théorème</u> : Pour toutes variables aléatoires X et Y :
$E(X + Y) = E(X) + E(Y) \qquad V(X + Y) = V(X) + V(Y) + 2\,\text{cov}(X,Y)$

Il faut donc faire attention car en général : $V(X + Y) \neq V(X) + V(Y)$.

<u>Définition</u> : Si $\text{cov}(X,Y) = 0$, on dit que les variables aléatoires sont non corrélées.

<u>Cas de deux variables aléatoires indépendantes</u> :
Pour tous i et j : $P[(X = x_i) \cap (Y = y_j)] = P(X = x_i)P(Y = y_j)$. Donc :
$E(XY) = \sum_{(i,j) \in I \times J} x_i y_j P(X = x_i)P(Y = y_j)$. Donc :

$$E(XY) = \sum_{i \in I} x_i P(X = x_i) \left[\sum_{j \in J} y_j P(Y = y_j)\right] = \sum_{i \in I} x_i P(X = x_i) E(Y).$$

Donc : $E(XY) = \left[\sum_{i \in I} x_i P(X = x_i)\right] E(Y) = E(X)E(Y)$.

<u>Théorème</u> : Lorsque deux variables aléatoires X et Y sont indépendantes, en plus des autres propriétés usuelles :
$E(XY) = E(X)E(Y) \qquad \text{cov}(X,Y) = 0 \qquad V(X + Y) = V(X) + V(Y)$

En particulier deux variables aléatoires indépendantes sont non corrélées car $E(XY) = E(X)E(Y)$. Mais elles peuvent être non corrélées sans être indépendantes.

<u>Cas de n variables aléatoires</u> : $E(X_1 + ... + X_n) = E(X_1) + ... + E(X_n)$.
$V(X_1 + ... + X_n) = V(X_1) + ... + V(X_n) + 2 \sum_{1 \le i < j \le n} \text{cov}(X_i, X_j)$.
$V(X_1 + ... + X_n) = V(X_1) + ... + V(X_n)$ si $X_1, ..., X_n$ sont deux à deux indépendantes.

Ceci permet une autre démonstration de l'espérance et de la variance de la loi binomiale. En effet, dans chacune des n épreuves d'un schéma de Bernoulli, on peut définir la variable de Bernoulli associée : elle prend la valeur 1 en cas de succès et la valeur 0 sinon. On obtient ainsi n variables $X_1,...,X_n$ deux à deux indépendantes puisque les épreuves sont indépendantes. Le nombre total de succès est $X = X_1 + ... + X_n$ et les variables $X_1,...,X_n$ sont toutes identiques. Donc :

$E(X) = E(X_1) + ... + E(X_n) = nE(X_1) = np$.

$V(X) = V(X_1) + ... + V(X_n) = nV(X_1) = npq$.

Dans le cas de la loi hypergéométrique, on peut raisonner exactement de la même manière. On a toujours $X = X_1 + ... + X_n$, mais les variables $X_1,...,X_n$ ne sont pas indépendantes (tirages sans remise).

On a toujours $E(X) = E(X_1) + ... + E(X_n) = nE(X_1) = np$.

Mais la variance n'est pas la somme des variances.

3) Stabilité des lois

Stabilité de la loi de Poisson : Si X et Y sont deux variables aléatoires indépendantes qui suivent respectivement les lois de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$, alors la somme $X+Y$ suit la loi de Poisson $\mathcal{P}(\lambda+\mu)$.

Démonstration : $X(\Omega) = Y(\Omega) = \mathbb{N}$, donc $(X+Y)(\Omega) = \mathbb{N}$.

$$P(X+Y=k) = \sum_{i=0}^{k} P[(X=i)\cap(Y=k-i)] = \sum_{i=0}^{k} P(X=i)P(Y=k-i) \text{ (indépendance)}$$

$$P(X+Y=k) = \sum_{i=0}^{k} \frac{\lambda^i}{i!}e^{-\lambda}\,\frac{\mu^{k-i}}{(k-i)!}e^{-\mu} = \frac{1}{k!}e^{-(\lambda+\mu)}\sum_{i=0}^{k}\binom{k}{i}\lambda^i\mu^{k-i} = \frac{(\lambda+\mu)^k}{k!}e^{-(\lambda+\mu)}.$$

Donc la somme $X+Y$ suit la loi de Poisson $\mathcal{P}(\lambda+\mu)$.

Stabilité de la loi binomiale : Si X et Y sont deux variables aléatoires indépendantes qui suivent respectivement les lois binomiales $\mathcal{B}(n,p)$ et $\mathcal{B}(m,p)$, alors la somme $X+Y$ suit la loi binomiale $\mathcal{B}(n+m,p)$.

Démonstration : $X = X_1 + ... + X_n$ où $X_1,...,X_n$ sont n variables de Bernoulli indépendantes de même paramètre p.

De même, $Y = Y_1 + ... + Y_m$ où $Y_1,...,Y_m$ sont m variables de Bernoulli indépendantes de même paramètre p.

Donc $X+Y = X_1 + ... + X_n + Y_1 + ... + Y_m$ est la somme de $n+m$ variables de Bernoulli de même paramètre p. Elles sont toutes indépendantes car X et Y le sont.

Donc $X+Y$ suit la loi binomiale $\mathcal{B}(n+m,p)$.

Remarque : Ces deux résultats s'étendent par récurrence à n variables aléatoires mutuellement indépendantes.

III – Convergence

1) Convergence en probabilité

Définition : Une suite (X_n) de variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$ converge en probabilité vers une variable aléatoire X définie sur $(\Omega, \mathcal{A}, P)$ si : $\forall \varepsilon > 0 \quad \lim_{n\to+\infty} P(|X_n - X| \geq \varepsilon) = 0$.

2) Loi faible des grands nombres

Rappel de l'inégalité de Bienaymé-Tchebychev : Si X est une variable aléatoire d'espérance m et d'écart-type σ, alors $\forall \varepsilon > 0 \quad P(|X-m| \geq \varepsilon) \leq \dfrac{\sigma^2}{\varepsilon^2}$.

On va appliquer à une suite de variables aléatoires.

Théorème : Soit (X_n) une suite de variables aléatoires deux à deux indépendantes et de même loi, donc de même espérance m et de même écart-type σ.
Alors la suite de variables aléatoires $Y_n = \dfrac{X_1 + \ldots + X_n}{n}$ converge « en probabilité » vers la variable certaine égale à m.

Cela signifie donc que : $\forall \varepsilon > 0 \quad \lim\limits_{n \to +\infty} P(|Y_n - m| \geq \varepsilon) = 0$.

Démonstration :

$E(Y_n) = \dfrac{1}{n} E(X_1 + \ldots + X_n) = \dfrac{1}{n}[E(X_1) + \ldots + E(X_n)] = \dfrac{1}{n} \times nm$. Donc $E(Y_n) = m$.

$V(Y_n) = \dfrac{1}{n^2} V(X_1 + \ldots + X_n)$. Or les variables sont deux à deux indépendantes. Donc :

$V(Y_n) = \dfrac{1}{n^2}[V(X_1) + \ldots + V(X_n)] = \dfrac{1}{n^2} \times n\sigma^2 = \dfrac{\sigma^2}{n}$. Donc : $\sigma(Y_n) = \dfrac{\sigma}{\sqrt{n}}$.

On applique l'inégalité de Bienaymé-Tchebychev à Y_n. Pour tout $\varepsilon > 0$:

$0 \leq P(|Y_n - m| \geq \varepsilon) \leq \dfrac{\sigma^2}{n\varepsilon^2}$. Or $\lim\limits_{n \to +\infty} \dfrac{\sigma^2}{n\varepsilon^2} = 0$. Donc $\lim\limits_{n \to +\infty} P(|Y_n - m| \geq \varepsilon) = 0$.

Interprétation : Plus n est grand, moins les valeurs de Y_n s'écartent de m. Y_n a même espérance que les X_n mais est de moins en moins dispersée.

C'est ce qui justifie l'utilisation des statistiques pour évaluer certaines quantités.

Exemple : On dispose d'une pièce et on veut évaluer la probabilité p d'obtenir Pile.

On joue indéfiniment à Pile ou Face avec cette pièce. Soit X_n la variable aléatoire qui vaut 0 si l'on obtient Face au n-ième lancer et 1 si l'on obtient Pile. On a ainsi une suite de variables aléatoires indépendantes deux à deux (car les lancers sont indépendants) et de même loi (car on lance toujours la même pièce) : elles suivent toutes la loi de Bernoulli $\mathcal{B}(p)$. Donc : $E(X_n) = p$ et $\sigma(X_n) = \sqrt{p(1-p)}$.

Y_n est la fréquence d'apparition de Pile. Donc d'après la loi des grands nombres, la fréquence d'apparition de Pile tend vers la probabilité p d'obtenir Pile et :

$$\forall \varepsilon > 0 \quad P(|Y_n - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2} \geq 1 - \frac{1}{4n\varepsilon^2}$$

car : $p(1-p) = \dfrac{1}{4} - \left(\dfrac{1}{2} - p\right)^2$, donc $p(1-p) \leq \dfrac{1}{4}$, donc $\dfrac{p(1-p)}{n\varepsilon^2} \leq \dfrac{1}{4n\varepsilon^2}$.

On peut donc dire par exemple que si $n = 500$ et $\varepsilon = 0{,}1$: $P(|Y_{500} - p| < 0{,}1) \geq 0{,}95$.

Au bout de 500 lancers, la différence entre la fréquence d'apparition de Pile et la probabilité d'obtenir Pile est inférieure à 0,1 avec une probabilité supérieure à 0,95.

Si l'on effectue 500 lancers et si l'on calcule la fréquence d'apparition de Pile au cours de ces 500 lancers, on obtient une valeur approchée de p à 10^{-1} près avec un risque d'erreur inférieur à 5%.

Si l'on veut une valeur approchée à 10^{-2} près avec le même risque d'erreur, il faudrait effectuer au moins 50 000 lancers car $1-\dfrac{1}{4n(0{,}01)^2} \geq 0{,}95 \Leftrightarrow n \geq 50000$.

Lorsque l'on veut obtenir une valeur approchée de la valeur moyenne m d'un caractère quantitatif sur une population suffisamment importante, il suffit d'effectuer la moyenne des valeurs du caractère sur un échantillon de n individus. Plus n est grand, meilleure est l'approximation.