Schéma de Bernoulli et loi binomiale

1. Une *épreuve de Bernoulli* est une expérience aléatoire ayant deux issues possibles appelées "succès" et "échec". On note p la probabilité de succès (la probabilité d'échec est donc $1 - p$).

2. Soit X la variable aléatoire réelle (v.a.r.) qui associe 1 au succès et 0 à l'échec.

   Proposition : la loi de probabilité de X est :
   $P_X\left(\{0\}\right) = P\left(\left[X = 0\right]\right) = 1 - p$ et $P_X\left(\{1\}\right) = P\left(\left[X = 1\right]\right) = p$

   Définition : cette loi est appelée *loi de Bernoulli* de paramètre p et notée $\mathcal{B}(p)$.

   On vérifie : $E(X) = p$ et $\operatorname{var}(X) = p(1-p)$.

3. Un *schéma de Bernoulli* (de taille n et de paramètre p) est une expérience aléatoire consistant à répéter n fois une épreuve de Bernoulli de paramètre p dans les mêmes conditions (on considérera que l'issue d'une épreuve ne dépend pas des issues précédentes).

   Pour $i \in [\![1, n]\!]$, soit X_i la v.a.r. qui associe 1 au succès et 0 à l'échec de la $i^{\text{ème}}$ épreuve de Bernoulli. Les v.a.r. $\left(X_i\right)_{i \in [\![1,n]\!]}$ ont même loi de probabilité, loi de Bernoulli de paramètre p, et sont indépendantes en probabilité par hypothèse.

4. Soit $Z = \sum_{i=1}^{n} X_i$ la v.a.r. indiquant le nombre de succès obtenus sur les n épreuves.

   Proposition : la loi de probabilité de Z est :
   $\forall k \in [\![0, n]\!], P_Z\left(\{k\}\right) = P\left(\left[Z = k\right]\right) = \binom{n}{k} p^k \left(1 - p\right)^{n-k}$ (que l'on note aussi $P\left(Z = k\right)$).

   Définition : cette loi est appelée *loi binomiale de paramètres n et p* et notée $\mathcal{B}(n, p)$.

Preuve de la proposition.
Soit k appartenant à $[\![0, n]\!]$, la v.a.r. Z est égale à k si et seulement si le cardinal de l'ensemble $\left\{i \in [\![1, n]\!], X_i = 1\right\}$ est égal à k. Soit A un sous-ensemble de $[\![0, n]\!]$ de cardinal k ; la probabilité que X_i soit égale à 1 si i appartient à A et 0 sinon est égale à $p^k\left(1 - p\right)^{n-k}$ parce que les v.a.r. sont indépendantes et que l'ordre dans lequel on obtient les k succès et les (n-k) échecs n'intervient pas. Il y a $\binom{n}{k}$ façons de choisir A ; on en déduit le résultat annoncé par additivité de la probabilité.

Commentaires

1) *Le schéma de Bernoulli et l'espace produit*
Pour écrire correctement le modèle associé au schéma de Bernoulli, il faut connaître la construction de l'espace produit (associé à une suite d'expériences aléatoires "indépendantes"), qui n'est pas explicitement au programme du Capes.

2) L'espérance et la variance d'une loi binomiale de paramètres n et p peuvent être calculées en utilisant les propriétés des coefficients binomiaux. On les obtient sans calcul en utilisant la

caractérisation de la loi $\mathcal{B}(n,p)$ comme la loi de la somme de n v.a.r. indépendantes et de même loi $\mathcal{B}(p)$. On a en effet : $E(Z)=\sum_{i=1}^{n}E(X_i)=np$ car l'espérance est linéaire et $\operatorname{var}(Z)=\sum_{i=1}^{n}\operatorname{var}(X_i)=np(1-p)$ car les variables $(X_i)_{i\in[\![1,n]\!]}$ sont indépendantes en probabilité.

3) *Loi de la somme de deux lois binomiales indépendantes de même paramètre*

Montrons sans calcul (c'est-à-dire en raisonnant sur les lois de probabilité) que si X_1 et X_2 dont deux variables indépendantes de lois $\mathcal{B}(n_1,p)$ et $\mathcal{B}(n_2,p)$ respectivement, alors la loi de X_1+X_2 est $\mathcal{B}(n_1+n_2,p)$.

La loi binomiale $\mathcal{B}(n,p)$ ($p\in\left]0,1\right[, n\in\mathbb{N}^*$) est définie par : $\forall k\in[\![0,n]\!]\ P(\{k\})=\binom{n}{k}p^k(1-p)^{n-k}$

et c'est la loi de la somme de n variables i.i.d. $\mathcal{B}(p)$ (*)

Soit $(U_i)_{i=1,\ldots,n_1+n_2}$ n_1+n_2 v.a.r. i.i.d. $\mathcal{B}(p)$, $Y_1=\sum_{1}^{n_1}U_i$ et $Y_2=\sum_{i=n_1+1}^{n_1+n_2}U_i$. Les v.a.r. Y_1 et Y_2 sont indépendantes car fonctions de deux familles de v.a.r. indépendantes $(U_i)_{i=1,\ldots,n_1}$ et $(U_i)_{i=n_1+1,\ldots,n_1+n_2}$.

En utilisant (*) on a : la loi de Y_1 est $\mathcal{B}(n_1,p)$, la loi de Y_2 est $\mathcal{B}(n_2,p)$ et la loi de Y_1+Y_2 est $\mathcal{B}(n_1+n_2,p)$.

On a donc montré que la somme de deux v.a.r. indépendantes de loi $\mathcal{B}(n_1,p)$ et $\mathcal{B}(n_2,p)$ respectivement est une loi $\mathcal{B}(n_1+n_2,p)$.

On ne retient des v.a.r. que leur loi de probabilité (ensemble des valeurs prises par les variables et probabilités correspondantes).

4) *Loi de la fréquence de succès sur n épreuves de Bernoulli indépendantes et approche fréquentiste de la probabilité*

Soit $F_n=\dfrac{Z}{n}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$, la fréquence de succès sur n épreuves de Bernoulli indépendantes.

La loi de probabilité de F_n se déduit de la loi de Z binomiale de taille n et de paramètre p :

$\forall k\in[\![0,n]\!], P([F_n=k/n])=\binom{n}{k}p^k(1-p)^{n-k}$; on vérifie $E(F_n)=p$ et $\operatorname{var}(F_n)=\dfrac{p(1-p)}{n}$.

On remarque que l'espérance de la fréquence ne dépend pas du nombre n d'épreuves et que la variance converge vers 0 lorsque le nombre d'épreuves tend vers l'infini.

C'est une condition suffisante de *convergence en probabilité* de F_n vers p qui s'énonce :

$\forall\varepsilon>0, P(|F_n-p|>\varepsilon)\xrightarrow{n\to\infty}0$.

Il s'agit de la loi faible des grands nombres, qui permet dans le cadre de la modélisation probabiliste d'estimer une probabilité par la fréquence de succès lors d'expériences précédentes supposées indépendantes (approche fréquentiste de la probabilité).