# CHAPITRE IV : LES LOIS DE PROBABILITE DISCRETES

## I Introduction

La plupart des phénomènes statistiques peuvent être décrits par un petit nombre de **modèles probabilistes ou lois de probabilité**. Lorsque cette représentation est possible, elle fournit une description beaucoup plus riche du phénomène que le simple calcul des caractéristiques de tendance centrale et de dispersion. Elle permet notamment de calculer la probabilité de certains événements, et par conséquent, de préciser dans une certaine mesure la représentation que l'on peut se faire de l'avenir. Il convient donc de connaître les modèles probabilistes les plus courants de façon à pouvoir rechercher dans ce catalogue celui qui est susceptible de convenir à la description du phénomène aléatoire déterminé.

Dans tous les cas, le processus est le suivant :

- l'observation du phénomène fournit une distribution expérimentale ou empirique ;
- l'analyse de cette distribution empirique – examen de la représentation graphique et calcul des caractéristiques de tendance centrale et de dispersion – donne une première idée de la nature du phénomène observé.

Au vu de ces premières conclusions, on choisit parmi les différents types de lois de distribution théorique celui qui paraît convenir. Il faut alors, au moyen de la série empirique, estimer les paramètres de cette loi. La substitution de la loi théorique à la distribution empirique n'est valable que si les valeurs observées et les valeurs théoriques résultant du modèle sont assez proches les unes des autres : il faut tester si les écarts observés entre les fréquences empiriques et les fréquences théoriques peuvent être raisonnablement attribués au hasard.

## II La loi binomiale

**La loi binomiale** intervient chaque fois que l'on considère **deux alternatives** dont **les probabilités restent constantes** au cours de cette suite d'épreuves.

### 1) Définition

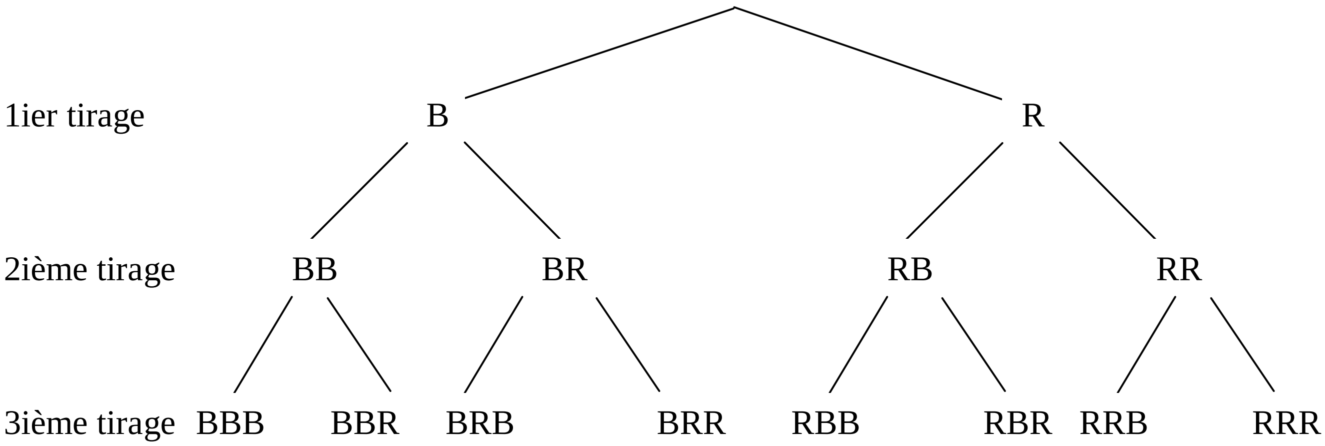
Soit un sac contenant N jetons de **deux** catégories :

- des jetons blancs (B) en proportion p ;
- des jetons rouges (R) en proportion q = 1 − p.

On effectue **n tirages successifs** d'un jeton, **en remettant** à chaque fois celui-ci dans le sac.

La variable aléatoire **binomiale X** est définie comme étant **le nombre de jetons blancs obtenus en ces n tirages**. Elle peut prendre les valeurs : 0, 1, ..., n. C'est une variable **discrète**.

Les différents événements possibles s'obtiennent suivant un schéma en arbre :




Au nième tirage, la variable aléatoire X prend la valeur k pour chacun des événements élémentaires correspondant à l'apparition de k jetons blancs. La probabilité de chacun de ces événements est égale à : $p^k q^{n-k}$ (formule des probabilités composées). Il y en a $C_n^k$ : autant de façons de choisir les rangs de tirage des k jetons blancs parmi les n rangs possibles.

Par suite, la probabilité pour que la variable aléatoire X prenne la valeur k correspondant à la réalisation de l'un ou l'autre de ces $C_n^k$ événements élémentaires est égale à : $C_n^k\ p^k q^{n-k}$ (formule des probabilités totales).

$$\boxed{\mathbf{p(X = k) = C_n^k\ p^k q^{n-k} \qquad avec\ q = 1 - p}}$$

**Remarques :** 1) $\sum_{k=0}^{n} p(X=k) = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1$ (formule du binôme).

2) La loi binomiale dépend de **deux paramètres**.

- n : le nombre de tirages successifs ou d'épreuves indépendant(e)s ;
- p : la probabilité de réalisation de l'événement étudié lors de chacun des tirages ou épreuves indépendant(e)s.

**Définition :** La probabilité que la variable binomiale X prenne la valeur k est

$$\boxed{\mathbf{dbinom(k, n, p)} = \mathbf{p(X = k)} = \mathbf{C_n^k\ p^k q^{n-k}} \qquad \mathbf{avec\ q = 1 - p}}$$

La notation dbinom(k, n, p) est celle du logiciel MathCad.

$C_n^k$ est noté combin(n, k) ou C(n, k) par le logiciel MathCad.

On notera X par $\mathbf{B_{n,\,p}}$ pour indiquer que la variable aléatoire X suit une loi binomiale de paramètres n et p.

## 2) Caractéristiques de la loi binomiale

### a) L'espérance mathématique

**Définition : L'espérance mathématique (ou moyenne)** de la loi binomiale est

$$\boxed{\mathbf{E(B_{n,\,p}) = np}}$$

### b) La variance et l'écart-type

**Définition : La variance** de la loi binomiale est

$$\boxed{\mathbf{V(B_{n,\,p}) = npq = np(1 - p)}}$$

**L'écart-type** de la loi binomiale est

$$\boxed{\sigma\,\mathbf{(B_{n,\,p}) = \sqrt{npq} = \sqrt{np(1 - p)}}}$$

## 3) La loi de Bernoulli

### a) Définition

**Définition : La loi de Bernoulli** est égale à la loi binomiale de paramètres n = 1 et p. C'est $\mathbf{B_{1,\,p}}$.

**Remarque :** Une loi binomiale $B_{n,\,p}$ est égale à la somme de n lois de Bernoulli $B_{1,\,p}$.

**b) <u>Caractéristiques</u>**
**<u>Espérance mathématique</u> :**

$$\boxed{\mathbf{E(B_{1,\,p}) = p}}$$

**<u>Variance</u> :**

$$\boxed{\mathbf{V(B_{1,\,p}) = pq = p(1 - p)}}$$

**<u>L'écart-type</u> :**

$$\boxed{\sigma\,\mathbf{(B_{1,\,p}) = \sqrt{pq} = \sqrt{p(1 - p)}}}$$

## 4) Calcul pratique des probabilités : table de la loi binomiale

Le calcul de la valeur de la probabilité attachée à chaque valeur de $B_{n,\,p}$ :

$p(X = k) = C_n^k\ p^k q^{n-k}$ devient fastidieux dès que n devient grand.

Des tables de la loi binomiale ont été établies. Elles dépendent des deux paramètres n et p. Les tables du National Bureau of Standards (« Tables of the binomial probability distribution n = 1, 2, ..., 49 ») donnent les probabilités et la fonction de répartition de la loi binomiale pour n inférieur à 50 et p variant de centième en centième : p = 0,01 ; 0,02 ; .... Les tables de Romig fournissent les mêmes données pour n compris entre 50 et 100. Ces tables ne sont utilisées que par les spécialistes : en effet, dès que le nombre n devient assez grand, la loi binomiale peut être approchée convenablement soit par la loi de Poisson (propriété qui sera vue dans le chapitre suivant), soit par la loi de Laplace-Gauss dont il existe des tables d'usage plus commode.

## 5) Approximation de la loi binomiale par la loi normale

Soit X une variable aléatoire binomiale $B_{n,\,p}$ dont le paramètre n croît indéfiniment, p n'étant pas trop voisin de 0 et de 1. Dans ces conditions, la loi binomiale tend vers la loi normale de paramètres $\mu$ **= np** et $\sigma$ **=** $\sqrt{\mathbf{npq}}$ **.**

$$\boxed{\mathbf{B_{n,\,p} \longrightarrow N_{np,\,\sqrt{npq}}}}$$

**<u>Considérations d'ordre pratique</u> :** 1) Habituellement, on considère que l'approximation de la loi binomiale par la loi normale est acceptable dès lors que
**n $\geq$ 30, np $\geq$ 15 et nq > 5.**

2) En pratique, cela signifie que les résultats trouvés en utilisant $B_{n,\,p}$ ou $N_{np,\,\sqrt{npq}}$ seront très proches.

## III La loi hypergéométrique

**La loi binomiale** correspond au tirage d'un échantillon **avec remise** dans une population comportant deux catégories d'individus.

**La loi hypergéométrique** correspond, au contraire, au tirage d'un échantillon **sans remise**. Elle a des propriétés moins simples et elle est d'utilisation moins commode que la loi binomiale. Toutefois, dès que l'effectif N de la population devient grand par rapport à celui de l'échantillon n, la loi hypergéométrique devient très proche de la loi binomiale et peut, en pratique, lui être assimilée.

### 1) Définition

On reprend le schéma du sac contenant N jetons de deux catégories :

- des jetons blancs (B) en proportion p ;
- des jetons rouges (R) en proportion $q = 1 - p$.

On effectue **n tirages successifs** d'un jeton, **sans remettre** celui-ci dans le sac avant le tirage suivant.

La variable aléatoire **hypergéométrique X** est définie comme étant **le nombre de jetons blancs obtenus en ces n tirages**.

Dans le cas de la loi binomiale, du fait de la remise du jeton dans le sac, les tirages successifs étaient indépendants. Pour la variable hypergéométrique, il en va autrement : la probabilité de choisir un jeton blanc au kième tirage dépend du résultat des tirages précédents. En effet, la composition du sac varie au fur et à mesure des épreuves. L'effectif du sac s'épuise peu à peu, d'où le nom de **tirage exhaustif** donné à ce mode de sélection d'un échantillon. Si l'échantillon est tiré au hasard, chacune des $C_N^n$ combinaisons que l'on peut faire en choisissant n jetons parmi les N contenus dans le sac sont équiprobables : ce sont les événements possibles.

Parmi celles-ci dénombrons celles qui correspondent à la présence de k jetons blancs et $(n - k)$ jetons rouges. Il y a $C_{Np}^k$ façons de choisir les k jetons blancs parmi les Np jetons blancs contenus dans le sac. A chacune de ces combinaisons correspondent $C_{Nq}^{n-k}$ façons de choisir les $(n - k)$ jetons rouges complémentaires parmi les Nq jetons rouges contenus dans le sac.

Il y a donc au total $C_{Np}^k \times C_{Nq}^{n-k}$ éventualités favorables à l'obtention de k jetons blancs. Le nombre k de jetons blancs dans l'échantillon ne peut pas être supérieur à l'effectif n de l'échantillon ni au nombre Np de jetons blancs présents dans le sac.

$$\boxed{\mathbf{k \leq \min(n, Np)}}$$

Le même raisonnement avec le nombre $(n - k)$ de jetons rouges conduit à :

$$\boxed{\mathbf{k \geq \max(0, n - Nq)}}$$

donc

$$\boxed{\max(0, n - Nq) \leq k \leq \min(n, Np)}$$

**Remarque :** La variable aléatoire hypergéométrique X est une **variable discrète** qui dépend de **trois** paramètres :

- N : l'effectif total de la population ;
- p : la proportion de jetons blancs dans celle-ci ;
- n : le nombre de tirages successifs.

**Définition :** La probabilité que la variable hypergéométrique X prenne la valeur k est

$$\boxed{p(X = k) = \frac{C_{Np}^{k} \times C_{Nq}^{n-k}}{C_{N}^{n}}}$$

Les valeurs possibles de cette variable sont les k tels que :

$$\boxed{\max(0, n - Nq) \leq k \leq \min(n, Np)}$$

On notera X par $H_{N,n,p}$ pour indiquer que la variable aléatoire X suit une loi hypergéométrique de paramètres N, n et p.

## 2) Caractéristiques de la loi hypergéométrique

### a) L'espérance mathématique

**Définition : L'espérance mathématique (ou moyenne)** de la distribution hypergéométrique est

$$\boxed{E(H_{N,n,p}) = np}$$

**Remarque :** Les lois binomiale et hypergéométrique ont donc même espérance mathématique.

### b) La variance et l'écart-type

**Définition : La variance** de la distribution hypergéométrique est

$$\boxed{V(H_{N,n,p}) = \frac{N - n}{N - 1}\, npq}$$

Le facteur $\mathbf{\dfrac{N - n}{N - 1}}$ est appelé **facteur d'exhaustivité**.

**Définition : L'écart-type** de la distribution hypergéométrique est

$$\boxed{\sigma(\mathbf{H_{N,n,p}}) = \sqrt{\mathbf{\frac{N - n}{N - 1}\, npq}} = \sqrt{\mathbf{\frac{N - n}{N - 1}\, np(1 - p)}}}$$

**Remarque :** Dès que l'on effectue plus d'un tirage, le coefficient $\frac{N-n}{N-1}$ est inférieur à 1. Cette variance est donc inférieure à celle de la loi binomiale qui est égale à npq.
Au fur et à mesure que l'effectif de l'échantillon se rapproche de celui de la population, la variance de la variable hypergéométrique diminue. A la limite, on prélève toute la population et la variance est nulle : on connaît exactement le nombre de jetons blancs contenus dans le sac. Cependant, lorsque l'effectif N de la population est grand par rapport à la taille n de l'échantillon, le coefficient est peu différent de 1.
$\frac{N-n}{N-1} \to 1$ lorsque $N \to +\infty$. Les deux modes de tirage de l'échantillon, exhaustif (loi hypergéométrique) et avec remise (loi binomiale) sont alors équivalents. En effet, la précision d'un sondage, dans le cas le plus fréquent où la taille de la population est grande et celle de l'échantillon est relativement petite, ne dépend en pratique que de l'effectif de l'échantillon et non de celui de la population. Autrement dit, la précision obtenue dépend davantage de l'effectif n de l'échantillon que **du taux de sondage** $\frac{n}{N}$.

### 3) Tendance de la loi hypergéométrique vers la loi binomiale

Quand l'effectif N de la population devient très grand, n et p demeurant fixes, la loi hypergéométrique tend vers la loi binomiale.

**Proposition : L'approximation de la loi hypergéométrique par la loi binomiale** est considérée comme valable dès que **le taux de sondage** $\frac{n}{N}$ **est inférieur à 10 % = 0,1**.

## IV La loi de Pascal

### 1) Définition

Soit un sac contenant N jetons de deux catégories :
- des jetons blancs (B) en proportion p ;
- des jetons rouges (R) en proportion q = 1− p.

On effectue des tirages d'un jeton **avec remise**.

La variable aléatoire de **<u>Pascal X</u>** est définie comme étant **<u>le nombre de tirages nécessaires pour obtenir r jetons blancs</u>**. Elle peut prendre les valeurs : r, r + 1, r + 2 , .....

$$\boxed{\mathbf{p(X = k) = C_{k-1}^{r-1}\, p^{r}\, q^{k-r} \quad avec\ q = 1 - p}}$$

**<u>Définition</u> :** La probabilité que la variable aléatoire de Pascal X prenne la valeur k est :

$$\boxed{\mathbf{p(X = k) = C_{k-1}^{r-1}\, p^{r}\, q^{k-r} \quad avec\ q = 1 - p}}$$

On notera X par **Pascal(r, p)** pour indiquer que la variable aléatoire X suit une loi de Pascal de paramètres r et p.

## 2) <u>Caractéristiques de la loi de Pascal</u>

### a) <u>L'espérance mathématique</u>

**<u>Définition</u> : <u>L'espérance mathématique (ou moyenne)</u>** de la loi de Pascal est

$$\boxed{\mathbf{E(Pascal(r, p)) = \frac{r}{p}}}$$

### b) <u>La variance et l'écart-type</u>

**<u>Définition</u> : <u>La variance</u>** de la loi de Pascal est

$$\boxed{\mathbf{V(Pascal(r, p)) = \frac{rq}{p^2} = \frac{r(1 - p)}{p^2}}}$$

.

**<u>Définition</u> : <u>L'écart-type</u>** de la loi de Pascal est

$$\boxed{\mathbf{\sigma(Pascal(r, p)) = \sqrt{\frac{rq}{p^2}} = \frac{\sqrt{r(1 - p)}}{p}}}$$

**<u>Remarques</u> :** 1) Si r = 1, la loi de Pascal est appelée **<u>loi géométrique</u>**.
2) La loi de Pascal est aussi appelée **<u>loi binomiale négative</u>**.

**Fin**