Lois de probabilités discrètes

Pour trouver un modèle décrivant un ensemble de données, il est nécessaire de connaître parfaitement les lois statistiques les plus utilisées. Le choix d'une loi est lié :

- à la nature du phénomène étudié afin de choisir entre loi discrète et loi continue,
- à la forme de la distribution (histogramme),
- à la connaissance et à l'interprétation des principales caractéristiques de l'ensemble de données : espérance, médiane, variance, écart-type, coefficients d'asymétrie et de dissymétrie, etc.,
- au nombre de paramètres des lois, une loi dépendant de plusieurs paramètres peut s'adapter plus facilement à une distribution.

5.1 Définition d'une variable discrète

Une *variable aléatoire discrète* prend ses valeurs sur un ensemble fini ou dénombrable de points. La loi de probabilité d'une telle variable est appelée *loi discrète*.

Une loi de probabilité discrète est caractérisée par l'énumération des valeurs x_i, appartenant à $\mathbb{R}$ ou à un intervalle de $\mathbb{R}$, prises par la variable aléatoire X et par les probabilités associées, c'est-à-dire les nombres réels positifs p_i tels que :

$$\Pr(X = x_i) = p_i \qquad 0 \leqslant p_i \leqslant 1 \quad \sum_i p_i = 1$$

La *fonction de répartition* est une fonction en escalier, constante sur tout intervalle $[x_i, x_{i+1}[$, admettant en chaque point x_i un saut égal à $p_{i+1} = \Pr(X = x_{i+1})$.

5.1.1 Moments

$$E(X) = \sum_i x_i p_i \qquad \mathrm{Var}(X) = \sum_i x_i^2 p_i - [E(X)]^2$$

5.1.2 Domaine d'utilisation

Les lois discrètes sont utilisées pour modéliser les résultats des jeux de hasard, les sondages d'opinion, les phénomènes biologiques, les processus aléatoires (files d'attente, évolution de l'état de matériels)... Les plus utilisées sont la loi uniforme, la loi binomiale et les lois dérivées, la loi hypergéométrique, la loi de Poisson.

Exemple 5.1

Soit X la variable aléatoire prenant trois valeurs 0, 1, 2 avec les probabilités :

$$\Pr(X = 0) = 1/2 \qquad \Pr(X = 1) = 1/3 \qquad \Pr(X = 2) = 1/6$$

$(1/2 + 1/3 + 1/6 = 1)$

Espérance mathématique :

$$E(X) = 0 \times 1/2 + 1 \times 1/3 + 2 \times 1/6 = 2/3$$

Variance :

$$\mathrm{Var}(X) = (\sigma_x)^2 = E(X^2) - [E(X)]^2 = 1 \times 1/3 + 4 \times 1/6 - (2/3)^2 = 5/9$$

Les figures 5.1 et 5.2 représentent l'histogramme et la fonction de répartition de la loi donnée dans l'exemple 5.1.

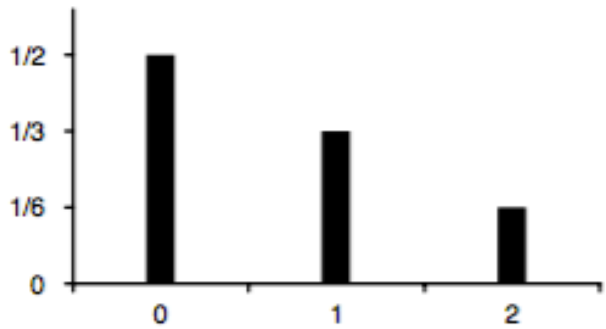

Figure 5.1 – Histogramme de la loi donnée dans l'exemple 5.1.

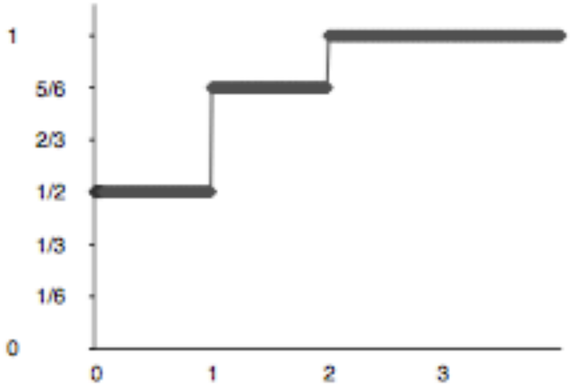

Figure 5.2 – Fonction de répartition de la loi donnée dans l'exemple 5.1.

5.2 Loi de Dirac

5.2.1 Définition

La loi de Dirac au point a de $\mathbb{R}$ est la loi de probabilité δ_a, définie sur $\mathbb{R}$ par :

$$\delta_a(x) = 1 \quad \text{si } x = a \qquad \delta_a(x) = 0 \quad \text{si } x \neq a$$

Cette loi est la loi la plus simple, associée à un phénomène déterministe X dont le résultat de toute expérience est égale à a.

5.2.2 Moments

$$E(X) = a \qquad \mathrm{Var}(X) = 0$$

5.2.3 Généralisation

Soit A un ensemble de n nombres réels distincts a_i et n nombres réels p_i tels que :

$$0 \leqslant p_i \leqslant 1 \qquad \sum_{i=1}^{n} p_i = 1$$

La combinaison linéaire $\sum_{i=1}^{n} p_i \, \delta_{a_i}$ définit une loi de probabilité associée à une variable aléatoire discrète X telle que pour tout indice i :

$$\Pr(X = a_i) = p_i$$

Si, pour toutes les valeurs de l'indice i, $p_i = 1/n$, la loi de la variable aléatoire X est la *loi de probabilité uniforme sur A*.

5.3 Loi uniforme

5.3.1 Définition

La *loi uniforme* sur $[1, n]$ est la loi de probabilité d'une variable aléatoire X prenant chaque valeur de l'ensemble $(1, 2 \dots n)$ avec la même probabilité :

$$\Pr(X = k) = 1/n \text{ pour tout entier } k \text{ compris entre 1 et } n$$

Plus généralement, soit Ω un ensemble fini de cardinal n. La *loi de probabilité équidistribuée* ou *uniforme* sur Ω est la loi définie sur Ω par la probabilité :

$$\Pr(\omega) = 1/n \text{ pour tout élément } \omega \text{ de } \Omega$$

Pour toute partie finie A de Ω, on a :

$$\Pr(A) = \frac{\operatorname{card}(A)}{\operatorname{card}(\Omega)}$$

5.3.2 Moments

$$E(X) = \frac{1}{n}\sum_{k=1}^{n} k = \frac{n+1}{2}$$

$$\begin{aligned}\operatorname{Var}(X) &= \frac{1}{n}\sum_{k=1}^{n} k^2 - \left(\frac{n+1}{2}\right)^2 \\ &= \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}\end{aligned}$$

5.3.3 Domaine d'utilisation

La loi de probabilité uniforme intervient dans de nombreux domaines comme les jeux de pile ou face ou les jeux de dés (avec une pièce ou un dé parfaitement équilibré(e)), les jeux de cartes, les loteries, les sondages...

5.4 Loi binomiale ou loi des tirages avec remise

5.4.1 Définition et propriétés d'une variable de Bernoulli

On considère une expérience dont le résultat ne peut prendre que deux valeurs appelées, par convention, *succès* ou *échec* : un candidat est reçu ou non à un examen, une pièce usinée est bonne ou défectueuse, une porte est ouverte ou fermée...

À une expérience de ce type, est associée une variable aléatoire X prenant la valeur 1 pour le succès et la valeur 0 pour l'échec, avec les probabilités respectives p et $(1-p) = q$. Cette variable est appelée *variable de Bernoulli*[1].

La *loi de probabilité* d'une variable de Bernoulli est définie par :

$$\Pr(X = 1) = p$$
$$\Pr(X = 0) = 1 - p = q$$

Ses *moments* sont :

$$E(X) = p$$
$$\mathrm{Var}(X) = p(1-p) = pq$$

Domaine d'utilisation : elle est utilisée pour modéliser des matériels qui seront soit survivants (valeur 1), soit défaillants (valeur 0) à un instant donné.

Elle s'applique aux jeux de hasard de type binaire comme pile ou face...

5.4.2 Définition d'une variable binomiale

On réalise n épreuves indépendantes de la même expérience telles que :
- chaque épreuve ne peut avoir que deux résultats, s'excluant mutuellement, soit le succès, soit l'échec,
- la probabilité p de succès est constante à chaque épreuve, la probabilité d'échec est également constante et égale à $1 - p = q$.

■ Probabilité d'obtenir *k* succès au cours de ces *n* épreuves

Soit X la variable aléatoire qui « compte » le nombre de succès au cours de n épreuves.

- Si au cours de n épreuves, on obtient k succès, on obtient également $(n-k)$ échecs. La probabilité de réalisation d'un tel événement est égale à $p^k(1-p)^{n-k}$ (les épreuves sont indépendantes).
- Il y a différentes façons d'obtenir k succès au cours de n épreuves, chaque épreuve pouvant être, indépendamment les unes des autres, un succès ou un échec. Le nombre de réalisations possibles de l'événement « obtenir k succès au cours de n épreuves » est le nombre de combinaisons sans répétitions de n objets pris k à k, soit :

$$C_n^k = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

D'où :

$$\Pr(X = k) = C_n^k\, p^k\,(1-p)^{n-k} = \frac{n!}{k!\,(n-k)!}\, p^k\,(1-p)^{n-k}$$

Cette expression étant un terme du développement du binôme $[p + (1-p)]^n$, la variable X est appelée ***variable binomiale***. On vérifie facilement que :

$$\sum_{k=0}^{n} \Pr(X = k) = 1$$

La loi de la variable X est appelée ***loi binomiale de paramètres*** (n, p), notée $B(n\,;p)$.

Une variable binomiale est égale à la somme de n variables aléatoires indépendantes de Bernoulli, la loi binomiale est donc la ***loi des épreuves répétées***. Elle est également la loi ***d'un tirage sans remise*** dans une urne contenant des boules de deux types différents.

5.4.3 Propriétés de la loi binomiale

■ Histogramme

La loi binomiale étant une loi discrète, son histogramme est en bâtons. La hauteur des bâtons, proportionnelle à la quantité $\Pr(X = k)$, croît de façon monotone, puis décroît également de façon monotone. En effet :

$$\frac{\Pr(X=k)}{\Pr(X=k-1)} = \frac{C_n^k\, p^k\, (1-p)^{n-k}}{C_n^{k-1}\, p^{k-1}\, (1-p)^{n-k+1}} = \frac{n-k+1}{k} \times \frac{p}{1-p}$$

$\Pr(X=k) \geqslant \Pr(X=k-1)$ si $(n-k+1)p \geqslant k(1-p)$ ou $k \leqslant p(n+1)$

La probabilité $\Pr(X = k)$ croît si k varie de 0 à une valeur k' égale à la partie entière de $p(n+1)$, puis décroît si k varie de k' à n.

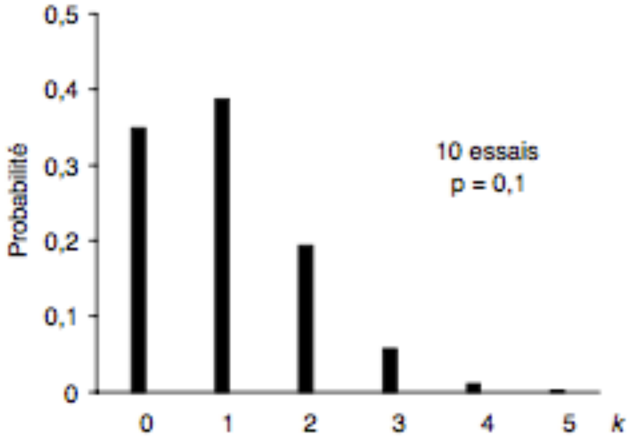

Figure 5.3 – Histogramme de la loi $B(10\,;0{,}10)$.

Tables de la loi binomiale

La table 1 donne les probabilités individuelles, $\Pr(X = k)$, et la table 2, les probabilités cumulées $\sum_{i=0}^{k} \Pr(X=i)$, pour toutes les valeurs de i et k et pour des valeurs du nombre n d'épreuves et de la probabilité p.

Moments

Une variable binomiale, suivant la loi $B(n\,;p)$, peut être considérée comme la somme de n variables indépendantes de Bernoulli. D'où les résultats :

$$E(X) = np \qquad \mathrm{Var}(X) = np(1-p) = npq$$

monotone, puis décroît également de façon monotone. En effet :

$$\frac{\Pr(X = k)}{\Pr(X = k-1)} = \frac{C_n^k \, p^k \, (1-p)^{n-k}}{C_n^{k-1} \, p^{k-1} \, (1-p)^{n-k+1}} = \frac{n-k+1}{k} \times \frac{p}{1-p}$$

$\Pr(X = k) \geqslant \Pr(X = k-1)$ si $(n-k+1)p \geqslant k(1-p)$ ou $k \leqslant p(n+1)$

La probabilité $\Pr(X = k)$ croît si k varie de 0 à une valeur k' égale à la partie entière de $p(n+1)$, puis décroît si k varie de k' à n.

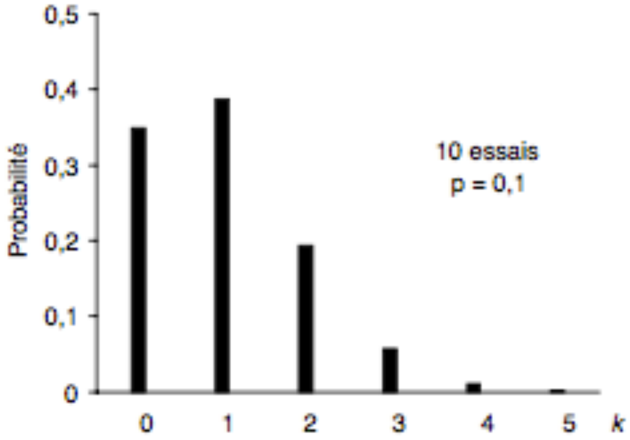

Figure 5.3 – Histogramme de la loi $B(10\,;0{,}10)$.

Tables de la loi binomiale

La table 1 donne les probabilités individuelles, $\Pr(X = k)$, et la table 2, les probabilités cumulées $\sum_{i=0}^{k} \Pr(X = i)$, pour toutes les valeurs de i et k et pour des valeurs du nombre n d'épreuves et de la probabilité p.

Moments

Une variable binomiale, suivant la loi $B(n\,;p)$, peut être considérée comme la somme de n variables indépendantes de Bernoulli. D'où les résultats :

$$E(X) = np \qquad \mathrm{Var}(X) = np(1-p) = npq$$

Coefficients d'asymétrie et d'aplatissement

$$\gamma_1 = \frac{1-2p}{\sqrt{npq}} \qquad \gamma_2 = 3 + \frac{1-6p}{npq}$$

Domaine d'utilisation

- La loi binomiale décrit des phénomènes ne pouvant prendre que deux états s'excluant mutuellement, succès ou échec dans un jeu, bonne pièce ou pièce défectueuse dans une fabrication, lot acceptable ou lot refusé, défaillance ou fonctionnement d'un matériel…
- Elle est utilisée dans le domaine technique pour déterminer la probabilité de défaillance à la sollicitation de matériels, en contrôle qualité, mais elle ne peut s'appliquer rigoureusement que si les expériences sont non exhaustives, c'est la loi du *tirage avec remise.*
- Les événements considérés doivent être indépendants et la probabilité de réalisation d'un événement doit être constante.

Somme de variables aléatoires, binomiales, indépendantes et de même paramètre

De la définition résulte la propriété suivante :

La somme de n variables aléatoires binomiales indépendantes et de même paramètre p est une variable aléatoire binomiale.

$$\left.\begin{array}{l}\text{Loi de la variable } X \ : \ B\left(n_1\,;p\right)\\ \text{Loi de la variable } Y \ : \ B\left(n_2\,;\,p\right)\\ X \text{ et } Y \text{ indépendantes}\end{array}\right\}\ \text{loi de } S = X + Y : B\left(n_1 + n_2\,;\,p\right)$$

Exemple 5.2

On veut réaliser une étude clinique sur des malades se présentant à une consultation hospitalière. Pour cette étude, seuls les malades répondant à un ensemble de critères C sont retenus. Des statistiques antérieures ont montré que 20 % des consultants présentent les critères C.

10 malades viennent consulter le premier jour.

Soit X la variable aléatoire « nombre de malades retenus » c'est-à-dire répondant à l'ensemble des critères C. La variable X suit *la loi binomiale* $B(10\,;0{,}20)$.

■ Loi de Pascal

□ Définition

La loi de Pascal est la loi de la variable Z « loi du nombre d'essais nécessaires » pour obtenir exactement k fois un événement de probabilité p, les hypothèses étant les mêmes que pour la loi binomiale (la probabilité p est constante au cours des essais).

$$\Pr(Z = z) = p\, C_{z-1}^{k-1}\, p^{k-1}\ (1-p)^{z-k} \qquad z \geqslant k \quad z \in \mathbb{N}^*$$

(On a obtenu un succès à l'essai n° z (de probabilité p) et $k-1$ succès au cours des $z-1$ essais précédents.)

□ Moments

$$E(Z) = \frac{k}{p} \qquad \operatorname{Var}(Z) = \frac{k\ (1-p)}{p^2}$$

□ Coefficients d'asymétrie et d'aplatissement

$$\gamma_1 = \frac{2-p}{\sqrt{k(1-p)}} \qquad \gamma_2 = 3 + \frac{p^2 + 6(1-p)}{k(1-p)}$$

■ Loi binomiale négative

□ Définition

À partir de la loi de Pascal, on définit la *loi binomiale négative*, ou loi de la variable aléatoire $T = Z - k$:

$$\Pr(T = t) = C_{k+t-1}^{k-1}\, p^k\, (1-p)^t = C_{k+t-1}^{t}\, p^k\, (1-p)^t$$

□ Moments, coefficients d'asymétrie et d'aplatissement

Si on pose $P = (1-p)/p$ et $Q = 1/p$, on obtient :

$$E(T) = kP \qquad \operatorname{Var}(T) = kPQ \qquad \gamma_1 = \frac{P+Q}{\sqrt{kPQ}} \qquad \gamma_2 = 3 + \frac{1+6PQ}{kPQ}$$

valeurs que l'on peut comparer à celles de la loi binomiale $B(k\,;p)$.

La probabilité qu'aucun malade ne soit recruté ce jour est égale à :

$$\Pr(X = 0) = C_{10}^{0}\,(0{,}20)^{0}\,(0{,}80)^{10} = 0{,}107$$

La probabilité pour qu'il y ait au moins un malade recruté est égale à :

$$\Pr(X \geqslant 1) = 1 - \Pr(X = 0) = 1 - 0{,}107 = 0{,}803$$

5.4.4 Lois voisines de la loi binomiale

■ Loi géométrique

□ Définition

La loi géométrique est la loi de la variable Y « loi du nombre d'essais nécessaires » pour qu'un événement de probabilité p apparaisse pour la première fois, les hypothèses étant les mêmes que pour la loi binomiale, en particulier, la ***probabilité*** ***p*** reste ***constante*** au cours des essais :

$$\Pr(Y = k) = p\left(1 - p\right)^{k-1} \qquad k \in \mathbb{N}^{*}$$

(Il y a eu $k - 1$ échecs avant d'obtenir le succès au $k^{\text{ème}}$ essai).

□ Moments

$$E(Y) = \frac{1}{p} \qquad \operatorname{Var}(Y) = \frac{1-p}{p^{2}}$$

□ Coefficients d'asymétrie et d'aplatissement

$$\gamma_1 = \frac{2-p}{\sqrt{1-p}} \qquad \gamma_2 = 3 + \frac{(p-3)^{2} - 3}{1-p}$$

Exemple 5.3

Un certain matériel a une probabilité $p = 0{,}02$ constante de défaillance à chaque mise en service. On procède à l'expérience suivante, l'appareil est mis en marche, arrêté, remis en marche, arrêté, jusqu'à ce qu'il tombe en panne. Le nombre d'essais nécessaires pour obtenir la panne est une variable aléatoire suivant la loi géométrique de paramètre p. La probabilité que ce matériel tombe en panne (pour la première fois) au dixième essai est égale à :

$$\Pr(Y = 10) = (0{,}02)(1 - 0{,}02)^{9} = 0{,}0167$$

■ Comparaison des lois binomiale et de Pascal

□ Loi binomiale

- Elle compte le nombre de succès au cours de n épreuves.
- Le nombre n d'épreuves est fixé.
- Le nombre de succès est une variable aléatoire pouvant prendre toutes les valeurs entières entre 0 et n.

□ Loi de Pascal

- Elle compte le nombre d'essais nécessaires pour obtenir k succès.
- Le nombre k de succès est fixé.
- Le nombre d'épreuves est une variable aléatoire pouvant prendre toutes les valeurs entières entre k et l'infini.
- Elle compte le nombre d'essais nécessaires pour obtenir k succès.

5.5 Loi multinomiale

5.5.1 Définition

La loi multinomiale est une généralisation de la loi binomiale.

Une population P est composée d'individus appartenant à k types différents, dans des proportions $p_1, p_2 \dots p_k$ telles que $\sum_{i=1}^{k} p_i = 1$.

On tire un échantillon de n individus, de façon équiprobable et indépendante et on s'intéresse à la composition de l'échantillon.

Soit X_i la variable aléatoire représentant le nombre d'individus de type i dans l'échantillon. Par définition, le vecteur $\underline{X} = (X_1, \dots, X_k)$ est un vecteur aléatoire suivant une *loi multinomiale de paramètres* $(n\,; p_1, \dots, p_k)$, notée $M(n\,; p_1, \dots, p_k)$.

$$\begin{aligned}\Pr[X = (x_1, \dots, x_k)] &= \mathrm{C}_n^{x_1}\, \mathrm{C}_{n-x_1}^{x_2} \dots \mathrm{C}_{n-(x_1+\dots+x_{k-1})}^{x_k}\, p_1^{x_1} \dots p_k^{x_k} \\ &= \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}\end{aligned}$$

$$\sum_{i=1}^{k} p_i = 1 \qquad \sum_{i=1}^{k} x_i = n$$

Le coefficient $\dfrac{n!}{x_1! \dots x_k!}$ est le nombre de partitions d'un échantillon de taille n en sous-populations d'effectifs x_i (voir annexe 1).

5.5.2 Propriétés

La loi marginale de X_i est une loi binomiale $B(n\,;p_i)$. En effet, un élément tiré est :

– soit du type i avec la probabilité p_i,

– soit de n'importe quel autre type avec la probabilité $(1 - p_i)$.

Le nombre d'individus du type i dans l'échantillon suit donc la loi binomiale $B(n\,;p_i)$. D'où :

$$E(X_i) = np_i \qquad \mathrm{Var}(X_i) = np_i(1 - p_i)$$

Les variables X_i et X_k ne sont pas indépendantes.

Le couple (X_i, X_k) suit une loi multinomiale de dimension 3. En effet, un élément tiré est :

– soit du type i (probabilité p_i),

– soit du type k (probabilité p_k),

– soit de n'importe quel autre type (probabilité $1 - p_i - p_k$).

En partant de ces propriétés, on démontre que :

$$E\,(X_iX_k) = n\,(n-1)\;p_i\,p_k$$

$$\mathrm{Cov}\,(X_i,\;X_k) = E\,(X_iX_k) - E\,(X_i)\;E\,(X_k) = -n\,p_i\,p_k$$

Les variables X_i et X_k ne peuvent donc pas être indépendantes.

5.5.3 Domaine d'utilisation

La loi multinomiale est utilisée en statistique.

Soit X une variable aléatoire continue de densité $f(x)$. On suppose que l'espace D_X des valeurs prises par cette variable est partagé en k classes distinctes C_i d'extrémités e_{i-1} et e_i, par exemple tranches d'âges, de revenus, d'impôts...

On considère un échantillon $(X_1, \dots, X_n)$ de n observations de cette variable et on cherche le nombre de points N_i de l'échantillon dans la classe C_i.

Le vecteur $(N_1, \ldots, N_k)$ suit la loi multinomiale de paramètres $(n, p_1, \ldots, p_k)$ avec :

$$p_i = \int_{e_{i-1}}^{e_i} f(x)\, \mathrm{d}x$$

Connaissant p_i pour chaque valeur de i, on en déduit la composition de l'échantillon. C'est la méthode qui peut être utilisée, par exemple, pour construire un histogramme.

Exemple 5.4

Un produit d'éclairage de l'entreprise M peut présenter des défectuosités regroupées en trois catégories : défectuosité critique, défectuosité majeure, défectuosité mineure.

Un contrôle final est effectué une semaine après la sortie du produit pour vérifier si certaines défectuosités se seraient développées au cours de cette période. Le résultat du contrôle est le suivant :

- 80 % du produit ne présente aucune défectuosité (ensemble E_1),
- 10 % du produit présente des défectuosités mineures (ensemble E_2),
- 6 % du produit présente des défectuosités majeures (ensemble E_3),
- 4 % du produit présente des défectuosités critiques (ensemble E_4).

Un échantillon de taille $n = 20$ est prélevé au hasard dans un grand lot et vérifié selon les critères précédents.

Soit X_i le nombre d'unités appartenant au sous-ensemble E_i dans l'échantillon contrôlé. L'ensemble (X_1, X_2, X_3, X_4) suit une loi multinomiale qui a pour paramètres les pourcentages donnés par le contrôle. D'où la probabilité :

$$\Pr(X_1 = x_1,\ X_2 = x_2,\ X_3 = x_3,\ X_4 = x_4)$$
$$= \mathrm{C}_{20}^{x_1}\, \mathrm{C}_{20-x_1}^{x_2}\, \mathrm{C}_{20-(x_1+x_2)}^{x_3}\, \mathrm{C}_{20-(x_1+x_2+x_3)}^{x_4}\, (0{,}80)^{x_1}\, (0{,}10)^{x_2}\, (0{,}06)^{x_3}\, (0{,}04)^{x_4}$$
$$= \frac{20!}{x_1!\,x_2!\,x_3!\,x_4!}\, (0{,}80)^{x_1}\, (0{,}10)^{x_2}\, (0{,}06)^{x_3}\, (0{,}04)^{x_4}$$

Espérance mathématique des variables X_i :

$$E(X_1) = 0{,}80 \times 20 = 16 \qquad E(X_2) = 0{,}10 \times 20 = 2$$
$$E(X_3) = 0{,}06 \times 20 = 1{,}2 \qquad E(X_2) = 0{,}04 \times 20 = 0{,}8$$

La probabilité cherchée est égale à :

$$\Pr(X = x) = \frac{C_{Np}^{x} C_{N-Np}^{n-x}}{C_{N}^{n}}$$

Les valeurs extrêmes de x sont :

$$\min x = \max\{0, n - N(1 - p)\} \quad \max x = \min\{n, Np\}$$

Le quotient n/N est appelé *taux de sondage*.

La variable aléatoire ainsi définie suit *une loi hypergéométrique* $H(N\,;n\,;p)$. Cette loi dépend de trois paramètres N, n et p.

5.6.2 Moments

Une variable aléatoire X suivant une loi hypergéométrique $H(N\,;n\,;p)$ peut être considérée comme une somme de n variables aléatoires de Bernoulli, $X_1 \ldots X_n$, non indépendantes, correspondant aux tirages successifs de n individus, tirage sans remise. On en déduit le calcul des moments d'ordre 1 et 2.

☐ Espérance mathématique

– La variable aléatoire X_1 correspond au tirage du premier individu :

$$\Pr(X_1 = 1) = p$$

$$\Pr(X_1 = 0) = 1 - p$$

d'où :

$$E(X_1) = p \quad \mathrm{Var}(X_1) = p(1 - p)$$

La variable aléatoire X_2 correspond au tirage du 2^{e} individu :

$$\Pr(X_2 = 1) = \Pr(X_2 = 1/X_1 = 1)\Pr(X_1 = 1) + \Pr(X_2 = 1/X_1 = 0)\Pr(X_1 = 0)$$

$$\Pr(X_2 = 1) = \frac{Np - 1}{N - 1}p + \frac{Np}{N - 1}(1 - p) = p$$

d'où :

$$\Pr(X_2 = 1) = p$$

$$\Pr(X_2 = 0) = 1 - p$$

On peut calculer différentes probabilités :

$\Pr(X_1 = 10,\ X_2 = 6,\ X_3 = 3,\ X_4 = 1) = 0{,}0001439$

$\Pr(X_1 = 20) = 0{,}0115292$

$\Pr(X_1 = 15,\ X_2 = 5) = 0{,}0054549$

5.6 Loi hypergéométrique ou loi du tirage exhaustif

5.6.1 Définition

On considère un tirage ***équiprobable sans remise*** ou tirage ***exhaustif*** dans une population d'effectif N, cette population étant composée de deux parties disjointes :

- une partie A à Np éléments (éléments possédant un certain caractère),
- une partie B à $(N - Np)$ éléments (éléments n'ayant pas ce caractère).

C'est le cas, par exemple, d'un lot de N pièces comprenant Np pièces défectueuses et donc $(N - Np)$ pièces fonctionnant bien.

Quelle est la probabilité qu'un sous-ensemble de n éléments contienne x éléments de l'ensemble A ?

Les éléments peuvent être tirés, soit un par un, soit d'un seul coup, mais ***sans remise.***

Soit X la variable aléatoire représentant le nombre d'individus ayant le caractère considéré dans l'échantillon. On veut calculer $\Pr(X = x)$.

Le nombre d'échantillons de taille n est égal à C_N^n

Un échantillon de taille n comprend :

- x individus, pris parmi les Np individus ayant le caractère considéré, donc C_{Np}^x choix possibles,
- $(n-x)$ individus pris parmi les $(N-Np)$ individus n'ayant pas ce caractère, donc C_{N-Np}^{n-x} choix possibles.

Le nombre d'échantillons de taille n, comprenant x individus pris parmi les Np individus, est donc égal à $C_{Np}^x\, C_{N-Np}^{n-x}$

La probabilité cherchée est égale à :

$$\Pr(X = x) = \frac{C_{Np}^{x}\, C_{N-Np}^{n-x}}{C_{N}^{n}}$$

Les valeurs extrêmes de x sont :

$$\min x = \max\{0, n - N(1 - p)\} \quad \max x = \min\{n, Np\}$$

Le quotient n/N est appelé *taux de sondage*.

La variable aléatoire ainsi définie suit *une loi hypergéométrique* $H(N\ ; n\ ; p)$. Cette loi dépend de trois paramètres N, n et p.

5.6.2 Moments

Une variable aléatoire X suivant une loi hypergéométrique $H(N\ ; n\ ; p)$ peut être considérée comme une somme de n variables aléatoires de Bernoulli, $X_1 \ldots X_n$, non indépendantes, correspondant aux tirages successifs de n individus, tirage sans remise. On en déduit le calcul des moments d'ordre 1 et 2.

□ Espérance mathématique

– La variable aléatoire X_1 correspond au tirage du premier individu :

$$\Pr(X_1 = 1) = p$$

$$\Pr(X_1 = 0) = 1 - p$$

d'où :

$$E(X_1) = p \quad \mathrm{Var}(X_1) = p(1 - p)$$

La variable aléatoire X_2 correspond au tirage du 2^e individu :

$$\Pr(X_2 = 1) = \Pr(X_2 = 1/X_1 = 1)\Pr(X_1 = 1) + \Pr(X_2 = 1/X_1 = 0)\Pr(X_1 = 0)$$

$$\Pr(X_2 = 1) = \frac{Np - 1}{N - 1}\,p + \frac{Np}{N - 1}\,(1 - p) = p$$

d'où :

$$\Pr(X_2 = 1) = p$$

$$\Pr(X_2 = 0) = 1 - p$$