Les Lois de Probabilité Discrètes

1. Introduction

Il est toujours possible d'associer à une variable aléatoire une probabilité et définir ainsi une **loi de probabilité**. Lorsque le nombre d'épreuves augmente indéfiniment, les **fréquences observées** pour le phénomène étudié **tendent vers les probabilités** et les distributions observées vers les distributions de probabilité ou loi de probabilité.

Identifier la loi de probabilité suivie par une variable aléatoire donnée est essentiel car cela conditionne le choix des méthodes employées pour répondre à une question biologique donnée

Par définition, les variables aléatoires **discrètes** prennent des valeurs entières discontinues sur un intervalle donné. Ce sont généralement le résultat de dénombrement.

Une distribution de probabilité suit une **loi uniforme** lorsque toutes les valeurs prises par la variable aléatoire sont **équiprobables**. Si n est le nombre de valeurs différentes prises par la variable aléatoire,

$$\forall\, i, \quad P(X = x_i) = \frac{1}{n}$$

Exemple :

La distribution des chiffres obtenus au lancer de dé (si ce dernier est non pipé) suit une loi uniforme dont la loi de probabilité est la suivante :

X	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

avec pour espérance : $E(X) = \frac{1}{6}\sum_{i=1}^{6} i = 3,5$ et pour variance $V(X) = \frac{1}{6}\sum_{i=1}^{6} i^2 - E(X)^2 = 2,92$

où les valeurs x_i correspondent au rang i de la variable X dans la série.

Dans le cas particulier d'une **loi discrète uniforme** où les valeurs de la variable aléatoire X correspondent au rang $x_i = i \ (\forall\ i \in [1, n])$

$$E(X) = \frac{n+1}{2} \qquad \text{et} \qquad V(X) = \frac{n^2 - 1}{12}$$

Soit un univers Ω constitué de **deux éventualités**, S pour succès et E pour échec

$$\Omega = \{E, S\}$$

sur lequel on construit une variable aléatoire discrète, « *nombre de succès* » telle que au cours d'**une** épreuve,

si S est réalisé, $X = 1$

si E est réalisé, $X = 0$

On appelle **variable de Bernoulli** ou variable *indicatrice*,

la variable aléatoire X telle que : $X : \Omega \rightarrow \mathbb{R}$

$$X(\Omega) = \{0,1\}$$

La **loi de probabilité** associée à la variable de Bernoulli X telle que,

$P(X = 0) = q$

$P(X = 1) = p$ avec $p + q = 1$

est appelée **loi de Bernoulli notée $\mathcal{B}(1, p)$**

$$\boxed{E(X) = p}$$

L'**espérance** de la variable de Bernoulli est

car par **définition** $E(X) = \sum_{i=1}^{2} x_i p_i = (0 \text{ x } q) + (1 \text{ x } p) = p$

$$\boxed{V(X) = pq}$$

La **variance** de la variable de Bernoulli est

car par **définition** $V(X) = \sum_{i=1}^{2} x_i^2 p_i - E(X)^2 = [(0 \text{ x } q) + (1 \text{ x } p)] - p^2$

d'où $V(X) = p - p^2 = p\,(1 - p) = pq$

Décrite pour la première fois par Isaac **Newton** en 1676 et démontrée pour la première fois par le mathématicien suisse Jacob **Bernoulli** en 1713, la **loi binomiale** est l'une des distributions de probabilité les plus fréquemment rencontrées en statistique appliquée.

Soit l'application $S_n : \Omega^n \to \mathbb{R}^n$

avec $S_n = X_1 + X_2 + \ldots + X_i + \ldots + X_n$ où X_i est une variable de **Bernoulli**

La **variable binomiale**, S_n, représente le **nombre de succès** obtenus lors de la répétition de n épreuves **identiques et indépendantes**, chaque épreuve ne pouvant donner que deux résultats possibles.

Ainsi la loi de probabilité suivie par **la somme de n variables de Bernoulli** où la probabilité associée au succès est p, est la **loi binomiale** de paramètres n et p.

$$S_n : \Omega^n \to \mathbb{R}^n$$

$$S_n = \sum_{i=1}^{n} X_i \quad \to \quad \mathcal{B}(n,p)$$

La probabilité que $S_n = k$, c'est à dire l'obtention de k succès au cours de n épreuves indépendantes est :

$$\boxed{P(S_n = k) = C_n^k p^k q^{n-k}}$$

Il est facile de démontrer que l'on a bien une loi de probabilité car :

$$\sum_{k=0}^{n} P(S_n = k) = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1 \text{ car } p+q = 1$$

Exemple :

Dans une expérience sur **le comportement du rat**, *rattus norvegicus*, on fait pénétrer successivement n rats dans un labyrinthe en forme de H. On étudie alors la probabilité que k rats empruntent la branche supérieure droite du H.

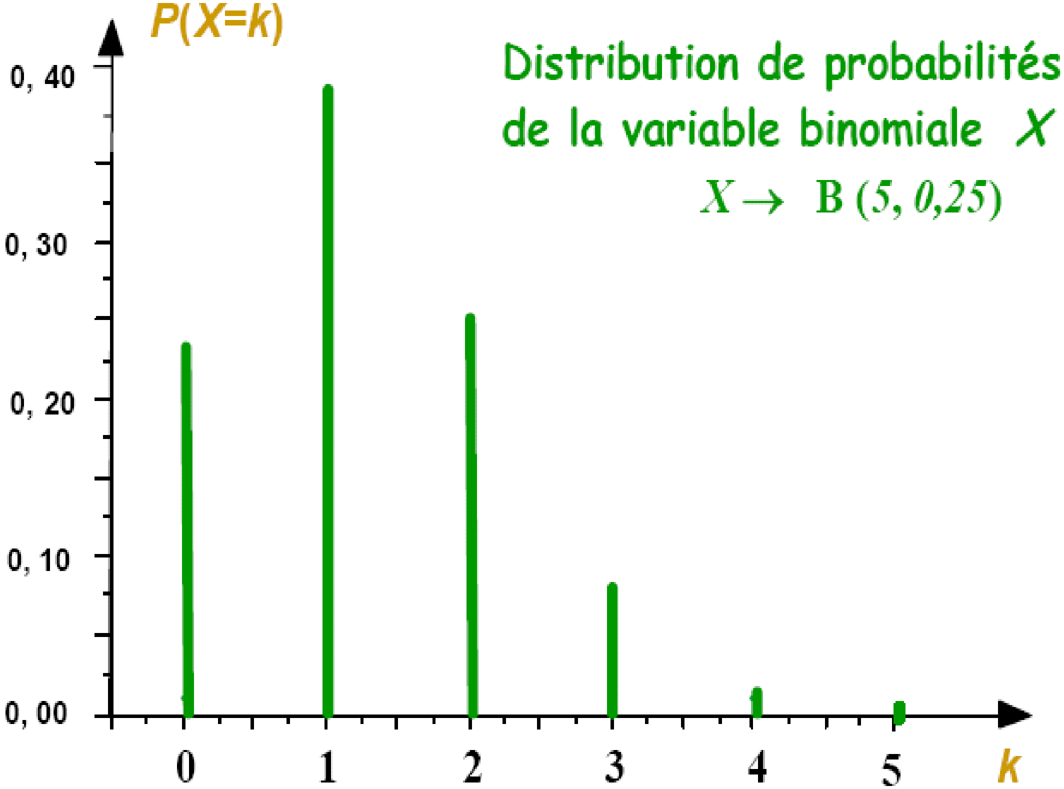
A chaque épreuve, deux évènements peuvent se produire : soit le rat suit l'itinéraire voulu (succès) soit il ne l'emprunte pas (échec). Sachant qu'il y a 4 itinéraires possibles (branches), la probabilité du succès $p = 1/4$.

Hypothèse :

- si les rats n'ont pas été conditionnés,
- si la branche supérieure droite ne comporte aucun élément attractif ou répulsif,
- si le choix de l'itinéraire d'un rat n'affecte pas le choix du suivant (odeurs)

alors : la variable aléatoire X « itinéraire emprunté pour x rats » suit une loi binomiale

$$\boxed{X \to \beta(n, \frac{1}{4})}$$



Nombre de rats ayant emprunt˘ la branche sup˘rieure droite du labyrinthe.

k	$P(X=k)$
0	$C_5^0\left(\frac{3}{4}\right)^5 = 0{,}237$
1	$C_5^1\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right) = 0{,}395$
2	$C_5^2\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2 = 0{,}264$
3	$C_5^3\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3 = 0{,}088$
4	$C_5^4\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^4 = 0{,}015$
5	$C_5^5\left(\frac{1}{4}\right)^5 = 0{,}001$

L'**espérance** d'une variable binomiale S_n est égale à

$$\boxed{E(S_n) = np}$$

en effet $E(S_n) = E(X_1 + X_2 + \ldots + X_i + \ldots + X_n)$

or $E(X_1 + X_2 + \ldots + X_i + \ldots + X_n) = \sum_{i=1}^{n} E(X_i)$ **propriété de l'espérance**

et $E(S_n) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p$ avec $E(X_i) = p$ variable de **Bernoulli**

d'où $E(S_n) = np$

La **variance** d'une variable binomiale S_n est égale à

$$\boxed{V(S_n) = npq}$$

en effet $V(S_n) = V(X_1 + X_2 + \ldots + X_i + \ldots + X_n)$

or $V(X_1 + X_2 + \ldots + X_i + \ldots + X_n) = \sum_{i=1}^{n} V(X_i)$ **propriété de la variance**

et $V(S_n) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} pq$ avec $V(X_i) = pq$ car variable de **Bernoulli**

d'où $V(S_n) = npq$

La **loi de Poisson** découverte au début du XIX[e] siècle par le magistrat français **Siméon-Denis Poisson** s'applique souvent aux phénomènes accidentels où la probabilité p est très faible ($p < 0{,}05$). Elle peut également dans certaines conditions être définie comme **limite d'une loi binomiale**.

Lorsque n devient grand, le calcul des probabilités d'une loi binomiale

$$P(S_n = k) = C_n^k p^k q^{n-k}$$

devient très fastidieux. On va donc, sous certaines conditions, trouver une approximation de p_k plus maniable.

Comportement asymptotique :

si $n \to \infty$ et $p \to 0$,

alors $X : \mathcal{B}(n,p) \to \mathcal{P}(\lambda)$ avec $np \to \lambda$

Remarque : Cette approximation est correcte si $n \geq 50$ et $np \leq 5$.

Exemple :

Soit une loi binomiale de paramètres (100 ; 0,01), les valeurs des probabilités pour k de 0 à 5 ainsi que leur approximation à 10^{-3} avec une loi de Poisson de paramètre ($\lambda = np = 1$) sont données dans le tableau ci-dessous :

k	0	1	2	3	4	5
$P(X=k)$	0,366	0,370	0,185	0,061	0,015	0,000
Approximation	0,368	0,368	0,184	0,061	0,015	0,003

Dans le cas de cet exemple où $n = 100$ et $np = 1$, l'approximation de la loi binomiale par une loi de poisson donne des valeurs de probabilités identiques à 10^{-3} près.

On appelle **processus poissonnien** (ou processus de Poisson), le modèle probabiliste des situations qui voient un flux d'évènements se produire les uns à la suite des autres de façon aléatoire (dans le temps et dans l'espace), obéissant aux conditions suivantes :

- la probabilité de réalisation de l'évènement au cours d'une petite période ou sur une petite portion d'espace Δt est proportionnelle à Δt soit $p\Delta t$.
- elle est indépendante de ce qui s'est produit antérieurement ou à côté,
- la probabilité de deux apparitions sur le même Δt est négligeable.

Ainsi, des évènements qui se réalisent de façon aléatoire comme des pannes de machines, des accidents d'avions, des fautes dans un texte, …peuvent être considérés comme relevant d'un processus poissonnien.

Une variable aléatoire X à valeurs dans $\mathbb{R}$ suit une **loi de Poisson de paramètre λ** (λ > 0) si les réels p_k sont donnés par

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

on note : $X \rightarrow \mathcal{P}(\lambda)$

Remarque : Une loi de Poisson est donnée par sa loi de probabilité :

(1) $\forall k, P(X = k) > 0$

(2) $\sum_{k \geq 0} P(X = k) = \sum_{k \geq 0} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k \geq 0} \frac{\lambda^k}{k!}$ or $\sum_{k \geq 0} \frac{\lambda^k}{k!} = e^{\lambda}$

d'où $\sum_{k \geq 0} P(X = k) = e^{-\lambda} e^{\lambda} = 1$

Exemple :

Une suspension bactérienne contient 5000 bactéries/litre. On ensemence à partir de cette suspension, 50 **boites de Pétri**, à raison d'1 cm^3 par boite. Si X représente le nombre de colonies par boite, alors la loi de probabilité de X est :

$$X \to \mathcal{P}\,(\lambda = 5)$$

La probabilité qu'il n'y ait **aucune** colonie sur la boite de Pétri est :

$P(X=0) = \dfrac{5^0\, e^{-5}}{0!} =$ **0,0067** soit approximativement **0,67 % de chance**.

La probabilité qu'il n'y ait **au moins une** colonie sur la boite de Pétri est :

$P(X > 0) = 1 - P(X = 0) = 1 - 0{,}0067 =$ **0,9933** soit **99,3 % de chance** d'avoir au moins une colonie bactérienne qui se développe dans la boite de Pétri. (voir **événement contraire**)

Comme pour la loi binomiale, il est possible d'utiliser une formule de **récurrence** pour calculer les valeurs des probabilités successives :

$$\boxed{P(X=k) = \frac{\lambda}{k} P(X = k-1)}$$

L'**espérance** d'une variable aléatoire de Poisson est

$$\boxed{E(X) = \lambda}$$

Par **définition** $E(X) = \sum_{k \geq 0} k p_k = \sum_{k \geq 0} k \frac{\lambda^k e^{-\lambda}}{k!}$ avec $k \in \mathbb{N}$ valeurs prises par la v.a. X

avec $\sum_{k \geq 0} \frac{\lambda^k}{k!} = \left[\lambda + \frac{\lambda^2}{1!} + \frac{\lambda^3}{2!} + \ldots + \frac{\lambda^{k+1}}{k!} \right] = \lambda \left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^k}{k!} \right]$

d'où $E(X) = \lambda e^{-\lambda} \sum_{k > 0} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$

La **variance** d'une variable de Poisson est

$$\boxed{V(X) = \lambda}$$

Par **définition**
$$V(X) = \sum_{k\geq 0} k^2 p_k - E(X)^2 = \sum_{k\geq 0} k^2 \frac{\lambda^k e^{-\lambda}}{k!} - \lambda^2$$

en posant $k^2 = k + k(k-1)$, alors

$$\sum_{k\geq 0} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k\geq 0} \frac{k}{k!} \lambda^k e^{-\lambda} + \sum_{k\geq 0} \frac{k(k-1)}{k!} \lambda^k e^{-\lambda}$$

d'où $\displaystyle\sum_{k\geq 0} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\left[\lambda + \frac{\lambda^2}{1!} + \frac{\lambda^3}{2!} + \ldots + \frac{\lambda^{k+1}}{k!}\right] + e^{-\lambda}\left[\lambda^2 + \frac{\lambda^3}{1!} + \frac{\lambda^4}{2!} + \ldots + \frac{\lambda^{k+2}}{k!}\right]$

d'où $\displaystyle\sum_{k\geq 0} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = \lambda^2 e^{-\lambda} \sum_{k\geq 0} \frac{\lambda^k}{k!} + \lambda e^{-\lambda} \sum_{k\geq 0} \frac{\lambda^k}{k!}$

$$\boxed{\text{d'où } V(X) = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} - \lambda^2 = \lambda}$$

Exemples :

Dans le cadre de la culture bactérienne, le nombre moyen de colonies attendu sur la boite de Pétri est : $E(X) = \lambda \ = $ **5 colonies.**

Ainsi si l'on effectue plusieurs cultures bactériennes (plusieurs boites de Pétri) à partir de la même solution initiale, on attend en moyenne **cinq colonies** pour l'ensemble des boites.

En ce qui concerne la variance et l'écart-type , on aura :

$V(X) = \lambda = 5$ et $\sigma(X) = \sqrt{V(X)} = $ **2,24 colonies.**

Exemples :

Dans le cadre de la culture bactérienne, le nombre moyen de colonies attendu sur la boite de Pétri est : $E(X) = \lambda = $ **5 colonies.**

Ainsi si l'on effectue plusieurs cultures bactériennes (plusieurs boites de Pétri) à partir de la même solution initiale, on attend en moyenne **cinq colonies** pour l'ensemble des boites.

En ce qui concerne la variance et l'écart-type , on aura :

$V(X) = \lambda = 5$ et $\sigma(X) = \sqrt{V(X)} = $ **2,24 colonies.**

Si X et Y sont deux variables aléatoires **indépendantes** suivant des **lois de Poisson** respectivement

$$X \rightarrow \mathcal{P}(\lambda) \text{ et } Y \rightarrow \mathcal{P}(\mu) \text{ alors } X+Y \rightarrow \mathcal{P}(\lambda+\mu)$$