

Chapitre 14

Variables aléatoires

Très souvent, on peut associer à chaque issue d'une expérience aléatoire un résultat, notamment numérique, qui correspond à l'observation d'un des aspects de l'expérience. Par exemple, si on lance deux dés, un rouge et un vert, on peut s'intéresser au résultat du dé rouge, à celui du dé vert, à la somme des deux, à la couleur de celui (ou ceux) qui donne(nt) le plus grand résultat. Si l'on observe le déplacement aléatoire d'une particule dans l'espace, on peut s'intéresser à la position, à chaque seconde, de la particule, mais aussi à sa vitesse, au temps nécessaire pour que la particule atteigne, éventuellement, une position fixée, etc...

Dans tout le chapitre, (Ω, \mathcal{A}, P) est un espace probabilisé.

I. Définitions, premières propriétés

Définition – Variable aléatoire

Une **variable aléatoire discrète** sur (Ω, \mathcal{A}) est une application définie sur Ω , et vérifiant les conditions suivantes :

- L'image $X(\Omega)$ de X est finie ou dénombrable,
- Pour tout $x \in X(\Omega)$, $X^{-1}(\{x\}) \in \mathcal{A}$.

Pour tout $x \in X(\Omega)$, l'événement $X^{-1}(\{x\})$ est noté $\{X = x\}$ ou $(X = x)$.

Lorsque X est à valeurs dans \mathbb{R} , on dit que X est une variable aléatoire **réelle**.

Remarques

- On parle aussi souvent de variable aléatoire sur (Ω, \mathcal{A}, P) , mais la définition d'une variable aléatoire n'utilise pas la probabilité P .
- Dans ce cours, toutes les variables aléatoires seront implicitement supposées discrètes.
- On rappelle que $X^{-1}(\{x\}) = \{\omega \in \Omega; X(\omega) = x\}$. Plus généralement, si U est un sous-ensemble de $X(\Omega)$, $X^{-1}(U) = \{\omega \in \Omega; X(\omega) \in U\}$. Le fait d'employer cette notation ne signifie absolument pas que X est bijective!
- Si X est une variable aléatoire sur (Ω, \mathcal{A}) , $X(\Omega)$ est fini ou dénombrable, donc on peut le décrire en extension sous la forme $X(\Omega) = \{x_n; n \in I\}$, où I est une partie de \mathbb{N} .

Alors la famille $((X = x_n))_{n \in I}$ est un système complet d'événements.

- Lorsque Ω est fini, si X est une application définie sur Ω , $X(\Omega)$ est également fini. Sachant de plus que $\mathcal{A} = \mathcal{P}(\Omega)$, la deuxième condition de la définition ci-dessus est aussi remplie. Une variable aléatoire est donc tout simplement, dans ce cadre, une application définie sur Ω . On parle de variable aléatoire sur Ω , au lieu de $(\Omega, \mathcal{P}(\Omega))$.

Propriété

Soit X une variable aléatoire sur (Ω, \mathcal{A}) et U un sous-ensemble de $X(\Omega) : U \subset X(\Omega)$. Alors $X^{-1}(U) \in \mathcal{A}$. L'événement $X^{-1}(U)$ est noté $\{X \in U\}$ ou $(X \in U)$.

Démonstration – L'ensemble U est fini ou dénombrable en tant que sous-ensemble de $X(\Omega)$, on peut le décrire en extension sous la forme $U = \{x_n; n \in I\}$, où I est une partie de \mathbb{N} . Alors

$$X^{-1}(U) = \bigcup_{n \in I} X^{-1}(\{x_n\});$$

c'est un élément de \mathcal{A} en tant que réunion finie ou dénombrable d'éléments de \mathcal{A} . \square

Notation – Soit X une variable aléatoire réelle sur (Ω, \mathcal{A}) et $x \in \mathbb{R}$. Lorsque $U =]-\infty, x] \cap X(\Omega)$, l'événement $(X \in U)$ est noté plus simplement $(X \leq x)$. On définit de façon analogue les événements $(X < x)$, $(X \geq x)$ et $(X > x)$.

Exemple – On modélise le lancer de deux dés, un rouge et un vert, par le choix de $\Omega = \llbracket 1, 6 \rrbracket^2$, muni de la probabilité uniforme. Pour tout $(i, j) \in \Omega$, i est le résultat du dé rouge, j celui du dé vert. La fonction X qui à (i, j) associe $i + j$ est une variable aléatoire sur Ω . Elle prend toutes les valeurs de $\llbracket 2, 12 \rrbracket$. Par exemple,

$$\begin{aligned} (X = 2) &= \{1, 1\} && \text{avec } P(X = 2) = \frac{1}{36}, \\ (X = 4) &= \{(1, 3), (2, 2), (3, 1)\} && \text{avec } P(X = 4) = \frac{3}{36} = \frac{1}{12}, \\ (X = 7) &= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} && \text{avec } P(X = 7) = \frac{6}{36} = \frac{1}{6}, \end{aligned}$$

Propriété/Définition

Soit X une variable aléatoire sur (Ω, \mathcal{A}) et f une fonction définie sur $X(\Omega)$. Alors $f \circ X$ est une variable aléatoire sur (Ω, \mathcal{A}) , plus souvent notée $f(X)$.

Démonstration – L'image de X est finie ou dénombrable, donc celle de $f(X)$ également. De plus, soit a un élément de $f(X(\Omega))$ (image de $f(X)$); alors

$$(f \circ X)^{-1}(\{a\}) = (X \in f^{-1}(\{a\})).$$

Or $f^{-1}(\{a\}) \subset X(\Omega)$, donc d'après la propriété précédente, $(f \circ X)^{-1}(\{a\}) \in \mathcal{A}$, ce qui prouve le résultat. \square

Exemple – Si X est une variable aléatoire réelle, X^2 est une variable aléatoire. Si X est à valeurs strictement positives, $\ln(X)$ est une variable aléatoire.

II. Loi d'une variable aléatoire

1. Généralités

Définition – Loi d'une variable aléatoire

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) .

On appelle **loi** de la variable aléatoire X la fonction définie sur $X(\Omega)$ par :

$$\forall x \in X(\Omega), \quad P_X(x) = P(X = x).$$

Remarque – La loi de X permet de définir une probabilité sur $(X(\Omega), \mathcal{P}(X(\Omega)))$.

Propriété

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) . On décrit $X(\Omega)$ en extension sous la forme $X(\Omega) = \{x_n; n \in I\}$ où I est une partie de \mathbb{N} .

Alors, pour tout $U \subset X(\Omega)$, on a

$$P(X \in U) = \sum_{x_n \in U} P(X = x_n).$$

Rappel – Lorsque $X(\Omega)$ est dénombrable et décrit en extension sous la forme $\{x_n; n \in \mathbb{N}\}$, U est fini ou dénombrable, et peut-être décrit en extension sous la forme $\{x_{\varphi(1)}, \dots, x_{\varphi(m)}\}$ (où $m = \text{card}(U)$) ou $\{x_{\varphi(k)}; k \in \mathbb{N}\}$ (où φ est une bijection de \mathbb{N} sur \mathbb{N}). Alors $\sum_{x_n \in U} P(X = x_n)$ s'exprime comme une somme finie, ou une somme de série convergente :

$$\sum_{x_n \in U} P(X = x_n) = \sum_{k=1}^m P(X = x_{\varphi(k)}) \quad \text{ou} \quad \sum_{x_n \in U} P(X = x_n) = \sum_{k=0}^{+\infty} P(X = x_{\varphi(k)}).$$

Par exemple, si $X(\Omega) = \mathbb{N}$ et $U = 2\mathbb{N} = \{2k; k \in \mathbb{N}\}$, alors $P(X \in U) = \sum_{k=0}^{+\infty} P(X = 2k)$.

Démonstration de la propriété – L'événement $(X \in U)$ est la réunion des événements deux à deux disjoints $(X = x_n)$ pour les x_n de U , d'où le résultat par définition d'une probabilité (et notamment, la somme précédente ne dépend pas de la façon de décrire U en extension). \square

Remarque – Dans le cas dénombrable, la série $\sum_{n \geq 0} P(X = x_n)$ converge et a pour somme 1. De plus, pour tout événement $A \in \mathcal{A}$, on a d'après la formule des probabilités totales,

$$P(A) = \sum_{n=0}^{+\infty} P(A | X = x_n) P(X = x_n).$$

Définition – Fonction de répartition

Soit X une variable aléatoire réelle sur (Ω, \mathcal{A}, P) .

On appelle **fonction de répartition** de X la fonction F_X définie sur \mathbb{R} par :

$$\forall x \in \mathbb{R}, \quad F_X(x) = P(X \leq x).$$

Propriété

Soit X une variable aléatoire réelle sur (Ω, \mathcal{A}, P) et F_X sa fonction de répartition. Alors :

- F_X est croissante sur \mathbb{R} .
- $F_X(x) \xrightarrow{x \rightarrow -\infty} 0$ et $F_X(x) \xrightarrow{x \rightarrow +\infty} 1$.

Démonstration

• Soit $(x, y) \in \mathbb{R}^2$ tel que $x \leq y$; alors $(X \leq x) \subset (X \leq y)$, et donc $P(X \leq x) \leq P(X \leq y)$, i.e., $F_X(x) \leq F_X(y)$: la fonction F_X est croissante.

• D'après le premier point, F_X a une limite ℓ en $+\infty$, et donc $F_X(n) \xrightarrow{n \rightarrow +\infty} \ell$. Or on remarque que $\bigcup_{n=0}^{+\infty} (X \leq n) = \Omega$, donc par propriété de continuité croissante,

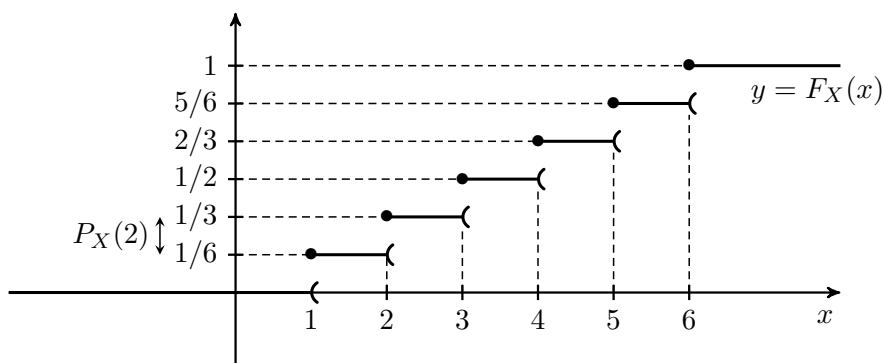
$$F_X(n) = P(X \leq n) \xrightarrow{n \rightarrow +\infty} P(\Omega) = 1.$$

On a donc $\ell = 1$.

On procède de même pour la limite en $-\infty$ en utilisant la propriété de continuité décroissante et le fait que $\bigcap_{n=0}^{+\infty} (X \leq -n) = \emptyset$ avec $P(\emptyset) = 0$. \square

Remarques

- La fonction de répartition d'une variable aléatoire réelle X est une fonction « en escalier » (pas tout à fait au sens mathématique), chaque « marche » correspondant au passage en abscisse d'une valeur prise par X . Ci-dessous, on donne la fonction de répartition correspondant au résultat du lancer d'un dé équilibré.



- Les fonctions F_X et P_X sont liées : si $X(\Omega) = \mathbb{N}$ par exemple, on a, pour tout $n \in \mathbb{N}$,

$$F_X(n) = \sum_{k=0}^n P(X = k) = \sum_{k=0}^n P_X(k)$$

et pour $n \geq 1$,

$$P_X(n) = P(X \leq n) - P(X \leq n-1) = F_X(n) - F_X(n-1).$$

Les valeurs de P_X correspondent aux hauteurs des « marches », sur le dessin précédent, $P_X(n)$ est la hauteur de la marche au point d'abscisse n .

Comme on l'a vu plus haut, si X est une variable aléatoire sur (Ω, \mathcal{A}) , la donnée d'une probabilité sur (Ω, \mathcal{A}) définit la loi de X , qui s'identifie à la donnée des $P(X = x)$ pour $x \in X(\Omega)$. Inversement, il est en fait possible de *choisir* des lois, ce qui peut être très utile lors de l'étape de modélisation :

Propriété (admise : démonstration hors programme)

Soit X une variable aléatoire sur (Ω, \mathcal{A}) . On décrit $X(\Omega)$ en extension sous la forme $X(\Omega) = \{x_n; n \in I\}$, où I est une partie de \mathbb{N} .

Soit $(p_n)_{n \in I}$ une famille ou une suite de réels positifs vérifiant

$$\sum_{n \in I} p_n = 1 \quad (\text{si } X(\Omega) \text{ est fini}) \quad \text{ou} \quad \begin{cases} \sum_{n \geq 0} p_n \text{ converge} \\ +\infty \\ \sum_{n=0} p_n = 1 \end{cases} \quad (\text{si } X(\Omega) \text{ est dénombrable})$$

Alors il existe une probabilité P sur (Ω, \mathcal{A}) telle que, pour tout $n \in I$, $P(X = x_n) = p_n$.

Remarque – En pratique, très souvent, une expérience aléatoire est en fait décrite par des données sur une ou plusieurs variables aléatoires. La modélisation par le choix de (Ω, \mathcal{A}) vient après, et elle n'est parfois pas nécessaire, ou admise. Par exemple :

- L'évolution d'un arbre généalogique peut être décrite par le nombre aléatoire de descendants directs de chaque individu, mais un choix de (Ω, \mathcal{A}) n'est pas du tout évident.
- Imaginons un système dont les états à différentes dates sont repérés par les entiers naturels ou relatifs (on pourra penser à la position d'une particule, à un stock de marchandises). L'évolution du système est décrite par les probabilités de transition de l'état i à l'état j . Supposons que les

transitions se font entre états voisins dans \mathbb{Z} (de k à $k + 1$ ou $k - 1$), et notons X_n l'état du système au rang n . La description du système se fait en donnant, pour tout $(n, k) \in \mathbb{N} \times \mathbb{Z}$, la probabilité

$$P(X_{n+1} = k + 1 | X_n = k).$$

On peut choisir

$$\Omega = \{(u_n)_{n \in \mathbb{N}} \in \mathbb{Z}^{\mathbb{N}}; \forall n \in \mathbb{N}, |u_{n+1} - u_n| = 1\},$$

mais ce n'est pas nécessairement utile de le préciser pour étudier le système.

2. Lois usuelles

La propriété précédente permet de définir des lois par la simple vérification qu'une série est à termes positifs, convergente et de somme 1 (ou qu'une famille finie de nombres positifs a pour somme 1). Ceci permet de définir les lois fondamentales suivantes ; pour chaque exemple, on donne un exemple de situation ainsi modélisée.

a. Loi uniforme

Définition

On dit qu'une variable aléatoire X sur (Ω, \mathcal{A}, P) suit la **loi uniforme** si $X(\Omega)$ est fini et si les événements $(X = x)$ pour $x \in X(\Omega)$ sont équiprobables.

Exemples

- La loi uniforme modélise par exemple le résultat d'un lancer de dé équilibré.
- Dans la modélisation du jeu de pile ou face infini faite dans le chapitre **Espaces probabilisés**, la variable aléatoire X qui donne le résultat des n premiers lancers suit la loi uniforme : pour tout $(u_1, \dots, u_n) \in \{0, 1\}^n$ (qui est de cardinal 2^n),

$$P(X = (u_1, \dots, u_n)) = \frac{1}{2^n}.$$

b. Loi de Bernoulli

Définition

Soit $p \in [0, 1]$. On dit qu'une variable aléatoire X sur (Ω, \mathcal{A}, P) suit la **loi de Bernoulli** de paramètre p si $X(\Omega) = \{0, 1\}$ et si

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

Ceci se note $X \hookrightarrow \mathcal{B}(p)$.

Remarque – On note très souvent $q = 1 - p$.

Exemples

- La loi de Bernoulli modélise un lancer de pièce, p représentant par exemple la probabilité d'obtenir « pile ».
- Plus généralement, la loi de Bernoulli modélise toutes les épreuves de Bernoulli, c'est-à-dire ayant deux résultats possibles ; celui de probabilité p est souvent interprété comme succès.

En Python, on peut simuler ainsi une expérience de Bernoulli de paramètre p (on supposera importé le module `random`) :

```

1 | def sim_bernoulli(p):
2 |     x = random.random()
3 |     if x < p:
4 |         return 1
5 |     else:
6 |         return 0

```

Propriété – Lien avec les fonctions indicatrices

- Soit A un événement de probabilité p , avec $A \neq \emptyset$ et $A \neq \Omega$. Alors $\mathbb{1}_A$ est une variable aléatoire sur (Ω, \mathcal{A}, P) qui suit la loi de Bernoulli de paramètre p .
- Inversement, soit X une variable aléatoire sur (Ω, \mathcal{A}, P) qui suit la loi de Bernoulli de paramètre p . Alors $X = \mathbb{1}_A$, avec $A = (X = 1)$ de probabilité p .

Démonstration

- La fonction $\mathbb{1}_A$ prend les valeurs 0 et 1, et $P(\mathbb{1}_A = 1) = P(A) = p$.
- Les deux fonctions X et $\mathbb{1}_{(X=1)}$ prennent la valeur 1 sur $(X = 1)$ et 0 sur $(X = 0)$, avec $(X = 0) \cup (X = 1) = \Omega$, donc ces fonctions sont égales. On a $P(X = 1) = p$ par définition. \square

c. Loi binomiale

Définition

Soient $n \in \mathbb{N}^*$ et $p \in [0,1]$. On dit qu'une variable aléatoire X sur (Ω, \mathcal{A}, P) suit la **loi binomiale** de paramètres n et p si $X(\Omega) = \llbracket 0, n \rrbracket$ et si

$$\forall k \in \llbracket 0, n \rrbracket, \quad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Ceci se note $X \hookrightarrow \mathcal{B}(n, p)$.

Remarque – On définit bien ainsi une loi, car d'après la formule du binôme de Newton,

$$\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1.$$

Interprétation – Le nombre S de succès lors d'une succession de n épreuves de Bernoulli de paramètre p mutuellement indépendantes suit la loi binomiale de paramètres n et p . En effet, la variable aléatoire S est à valeurs dans $\llbracket 0, n \rrbracket$ et, pour $k \in \llbracket 0, n \rrbracket$, l'événement $(S = k)$ est la réunion des événements consistants à fixer k succès et $n - k$ échecs. Ces événements sont deux à deux incompatibles, sont au nombre de $\binom{n}{k}$, et chacun est de probabilité $p^k (1-p)^{n-k}$ par indépendance mutuelle. On a donc

$$P(S = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Exemples

- Le nombre de « pile » obtenus lors de n lancers successifs mutuellement indépendants d'une pièce suit la loi binomiale de paramètres n et p , où p est la probabilité d'obtenir « pile » à un lancer donné.
- On effectue n tirages avec remise dans une urne contenant des boules indiscernables, rouges en proportion p et vertes en proportion $q = 1 - p$. La variable aléatoire donnant le nombre de boules rouges tirées suit la loi binomiale de paramètres n et p .

En Python, on peut simuler ainsi une suite de n épreuves de Bernoulli de paramètre p :

```
1 def sim_tirages(n,p):
2     L = []
3     for i in range(n):
4         x = random.random()
5         if x < p:
6             L.append(1)
7         else:
8             L.append(0)
9     return L
```

On peut simuler la variable aléatoire S de la façon suivante :

```

1 | def sim_nb_succes(n,p):
2 |     S = 0
3 |     for i in range(n):
4 |         x = random.random()
5 |         if x < p:
6 |             S += 1
7 |     return S

```

On peut alors simuler la loi $\mathcal{B}(n,p)$ de la façon suivante : on répète N fois la simulation ci-dessus, et on calcule, pour tout $k \in \llbracket 0,n \rrbracket$ la fréquence relative du résultat k lors de ces N expériences :

```

1 | def loi_binomiale(n,p,N):
2 |     L = []
3 |     for i in range(N):
4 |         S = sim_nb_succes(n,p)
5 |         L.append(S)
6 |     return [L.count(k)/float(N) for k in range(n+1)]

```

d. Loi géométrique

Définition

Soit $p \in]0,1[$. On dit qu'une variable aléatoire X sur (Ω, \mathcal{A}, P) suit la **loi géométrique** de paramètre p si $X(\Omega) \supset \mathbb{N}^*$ et si

$$\forall k \in \mathbb{N}^*, \quad P(X = k) = p(1-p)^{k-1}.$$

Ceci se note $X \hookrightarrow \mathcal{G}(p)$.

Remarques

- C'est le premier exemple que l'on rencontre de variable aléatoire prenant un nombre infini de valeurs.
- On définit bien une loi car la série géométrique de raison $(1-p) \in]0,1[$ est à termes positifs, elle converge, et

$$\sum_{k=1}^{+\infty} p(1-p)^{k-1} = p \sum_{k=0}^{+\infty} (1-p)^k = \frac{p}{1-(1-p)} = 1.$$

Exemples

- Considérons le jeu de pile ou face infini, avec p la probabilité d'obtenir « pile ». Pour $k \in \mathbb{N}^*$, l'événement « pile apparaît pour la première fois au rang k » a pour probabilité $p(1-p)^{k-1}$ ($k-1$ échecs suivis d'un succès).
- Plus généralement, la loi géométrique peut être interprétée comme loi du rang du premier succès dans une suite illimitée d'épreuves de Bernoulli mutuellement indépendantes et de même paramètre p .

Il est parfois utile d'autoriser que X prenne d'autres valeurs que celles de \mathbb{N}^* , avec probabilité nulle, notamment, en lien avec l'interprétation précédente, si aucun succès ne survient.

- La loi géométrique est aussi souvent utilisée pour modéliser des durées de fonctionnement de composants, machines, etc...

Remarque – On peut remplacer $X(\Omega) = \mathbb{N}^*$ par $X(\Omega) = \mathbb{N}$ avec :

$$\forall k \in \mathbb{N}, \quad P(X = k) = p(1-p)^k.$$

Dans ce cas, cette loi s'interprète comme loi du nombre d'échecs avant le premier succès.

e. Loi de Poisson

Définition

Soit $\lambda \in \mathbb{R}_+^*$. On dit qu'une variable aléatoire X sur (Ω, \mathcal{A}, P) suit la **loi de Poisson** de paramètre λ si $X(\Omega) = \mathbb{N}$ et si

$$\forall k \in \mathbb{N}, \quad P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Ceci se note $X \hookrightarrow \mathcal{P}(\lambda)$.

Remarque – On définit bien ainsi une loi, car on reconnaît la série exponentielle de λ , qui est à termes positifs, convergente, avec

$$\sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Le théorème suivant établit un lien asymptotique entre loi binomiale et loi de Poisson :

Théorème – Approximation de la loi binomiale par la loi de Poisson

Soient $(p_n)_{n \in \mathbb{N}}$ une suite d'éléments de $[0,1]$, $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires sur (Ω, \mathcal{A}, P) et $\lambda \in \mathbb{R}_+^*$. On fait les hypothèses suivantes :

- Pour tout $n \in \mathbb{N}$, X_n suit la loi binomiale de paramètres n et p_n ,
- $n p_n \xrightarrow[n \rightarrow +\infty]{} \lambda$.

Alors, pour tout $k \in \mathbb{N}$,

$$P(X_n = k) \xrightarrow[n \rightarrow +\infty]{} e^{-\lambda} \frac{\lambda^k}{k!}$$

Démonstration – Soit $k \in \mathbb{N}$. Alors, pour $n \geq k$ assez grand, $p_n \in]0,1[$ et on a

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{k!} p_n^k (1 - p_n)^{n-k} \\ &\underset{n \rightarrow +\infty}{\sim} \frac{n^k}{k!} p_n^k (1 - p_n)^{n-k}. \end{aligned}$$

Tout d'abord, $(n p_n)^k \xrightarrow[n \rightarrow +\infty]{} \lambda^k$. De plus, $n p_n \rightarrow \lambda$, donc $p_n \rightarrow 0^+$ et, lorsque $n \rightarrow +\infty$,

$$(1 - p_n)^{n-k} = \exp((n-k) \ln(1 - p_n)) = \exp((n-k)(-p_n + o(p_n))).$$

Or

$$(n-k)(-p_n + o(p_n)) = -n p_n + o(n p_n) \underset{n \rightarrow +\infty}{\sim} -n p_n \xrightarrow[n \rightarrow +\infty]{} -\lambda.$$

Par continuité de l'exponentielle et d'après ce qui précède, on a bien

$$P(X_n = k) \xrightarrow[n \rightarrow +\infty]{} e^{-\lambda} \frac{\lambda^k}{k!} \quad \square$$

Remarques

- Dans les calculs, on peut donc approcher $\binom{n}{k} p_n^k (1 - p_n)^{n-k}$ par $e^{-np} \frac{(np)^k}{k!}$.

Cela permet d'éviter des calculs de coefficients du binôme, qui font intervenir des quotients de grands nombres.

- On considère que l'approximation est intéressante lorsque $p \leq 0,1$, $n \geq 30$ et $np < 15$.

Exemple – On lance 100 fois un dé équilibré à 20 faces et on compte le nombre N de 20 obtenus. Ce nombre suit une loi binomiale $\mathcal{B}(100,1/20)$, on a donc, pour tout $k \in \llbracket 0,100 \rrbracket$,

$$P(N = k) = \binom{100}{k} \frac{1}{20^k} \left(\frac{19}{20}\right)^{100-k}$$

On est dans les conditions de l'approximation avec $np = 100/20 = 5$, on peut donc approcher $P(N = k)$ par $e^{-5} 5^k/k!$. Pour $k = 2$ par exemple, on a

$$\binom{100}{2} \frac{1}{20^2} \left(\frac{19}{20}\right)^{98} \approx 0,081 \quad \text{et} \quad e^{-5} \frac{5^2}{2!} \approx 0,084. \quad \square$$

Le programme suivant permet d'utiliser cette approximation :

```
1 | from math import exp, factorial
2 |
3 | def approx_poisson(n,p):
4 |     return [exp(-n*p)*(n*p)**k/factorial(k) for k in range(n+1)]
```

On peut alors tester par exemple l'approximation de $\mathcal{B}(30,0.1)$ par $\mathcal{P}(3)$ (listes B et A), ainsi qu'une simulation de cette approximation (liste L); dans ce qui suit, on n'affiche que les 10 premières valeurs, en arrondissant à 4 décimales pour B et A :

```
1 | from scipy.special import binom
2 |
3 | # Loi binomiale B(30,0.1)
4 | B = [ binom(30,k)*(0.1**k)*(0.9**(30-k)) for k in range(31) ]
5 | B = [ float("%.4f" % x) for x in B ]
6 |
7 | # Approximation par P(3)
8 | A = approx_poisson(30,0.1)
9 | A = [ float("%.4f" % x) for x in A ]
10 |
11 | # Simulation de B(30,0.1)
12 | L = loi_binomiale(30,0.1,10000)
13 |
14 | for k in range(10):
15 |     print "P( X =",k,") :",B[k],",",A[k],",",L[k]
```

Voici un résultat possible :

```
P( X = 0 ) : 0.0424 , 0.0498 , 0.0424
P( X = 1 ) : 0.1413 , 0.1494 , 0.139
P( X = 2 ) : 0.2277 , 0.224 , 0.2332
P( X = 3 ) : 0.2361 , 0.224 , 0.2358
P( X = 4 ) : 0.1771 , 0.168 , 0.1743
P( X = 5 ) : 0.1023 , 0.1008 , 0.1014
P( X = 6 ) : 0.0474 , 0.0504 , 0.047
P( X = 7 ) : 0.018 , 0.0216 , 0.0187
P( X = 8 ) : 0.0058 , 0.0081 , 0.006
P( X = 9 ) : 0.0016 , 0.0027 , 0.0019
```

Remarque – On s'intéresse à la loi du nombre d'occurrences d'un phénomène dans un intervalle de temps $[0,T]$. On fait les hypothèses suivantes :

- il existe $a \in \mathbb{R}$ tel que la probabilité que le phénomène se produise une fois dans un intervalle de temps de petite longueur h est ah ;
- la probabilité qu'il se produise plus d'une fois est négligeable (en fait, un $o(h)$);
- les nombres d'occurrences du phénomène dans des intervalles disjoints sont mutuellement indépendants.

On subdivise $[0, T]$ en intervalles de longueur T/n . D'après les hypothèses précédentes, on peut considérer que le nombre d'occurrences du phénomène dans l'intervalle $[0, T]$ suit la loi binomiale $\mathcal{B}(n, aT/n)$. D'après le résultat d'approximation précédent, pour n grand, on peut approcher cette loi par la loi de Poisson $\mathcal{P}(aT)$ (le paramètre λ s'identifie donc à aT).

Pour cette raison, la loi de Poisson est dite **loi des événements rares**; elle est souvent utilisée pour modéliser le nombre d'occurrences d'un phénomène dans un intervalle de temps fixé, ce phénomène étant « rare » dans un court intervalle de temps, mais observé sur un grand nombre de tels intervalles. Par exemple, on peut modéliser ainsi le nombre de véhicules passant devant un point d'observation, de clients entrant dans un magasin, de catastrophes naturelles, de désintégrations de noyaux radioactifs (lorsque la source est éloignée, les mesures faites par un compteur Geiger font effectivement apparaître une loi de Poisson).

III. Familles de variables aléatoires

1. Couple de variables aléatoires

Propriété/Définition

Soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}) .

L'application $\omega \mapsto (X(\omega), Y(\omega))$ est une variable aléatoire sur (Ω, \mathcal{A}) , appelée **couple** (X, Y) .

Démonstration – Les ensembles $X(\Omega)$ et $Y(\Omega)$ sont finis ou dénombrables, donc $X(\Omega) \times Y(\Omega)$ est fini ou dénombrable. L'image de (X, Y) est contenue dans $X(\Omega) \times Y(\Omega)$, elle est donc aussi finie ou dénombrable. Notons $Z = (X, Y)$. Pour tout (x, y) de $Z(\Omega)$,

$$Z^{-1}(\{(x, y)\}) = \{\omega \in \Omega; (X(\omega), Y(\omega)) = (x, y)\} = X^{-1}(x) \cap Y^{-1}(y);$$

c'est un événement en tant qu'intersection de deux événements. □

Notation

- L'événement $((X, Y) = (x, y)) = (X = x) \cap (Y = y)$ est plus souvent noté $(X = x, Y = y)$.
- Si $A \subset X(\Omega)$ et $B \subset Y(\Omega)$, l'événement $((X, Y) \in A \times B)$, c'est-à-dire $(X \in A) \cap (Y \in B)$, est plus souvent noté $(X \in A, Y \in B)$.

Corollaire

L'ensemble des variables aléatoires sur (Ω, \mathcal{A}) à valeurs dans \mathbb{K} ($\mathbb{K} = \mathbb{R}$ ou \mathbb{C}) est un \mathbb{K} -espace vectoriel (pour les lois d'addition et de multiplication par un scalaire).

Démonstration – C'est un sous-ensemble de l'espace vectoriel des applications de Ω dans \mathbb{K} , qui est non vide (la fonction nulle est une variable aléatoire) Enfin, soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}) à valeurs dans \mathbb{K} et soit $\lambda \in \mathbb{K}$. On définit la fonction $f : (x, y) \mapsto \lambda x + y$ sur \mathbb{K}^2 . Alors $\lambda X + Y = f(X, Y)$, qui est une variable aléatoire car le couple (X, Y) est une variable aléatoire. □

Définition

Soit (X, Y) un couple de variables aléatoires sur (Ω, \mathcal{A}, P) . On appelle :

- **loi conjointe** de X et Y la loi du couple (X, Y) .
- **lois marginales** du couple (X, Y) les lois de X et de Y .

Propriété

Soit (X, Y) un couple de variables aléatoires sur (Ω, \mathcal{A}, P) .

La loi du couple (X, Y) détermine entièrement ses lois marginales par les relations

$$\forall x \in X(\Omega), \quad P(X = x) = \sum_{y \in Y(\Omega)} P(X = x, Y = y),$$

$$\forall y \in Y(\Omega), \quad P(Y = y) = \sum_{x \in X(\Omega)} P(X = x, Y = y).$$

En revanche, les lois marginales du couple (X, Y) ne déterminent pas la loi conjointe de X et Y .

Démonstration – La première égalité est immédiate en remarquant que $((Y = y))_{y \in Y(\Omega)}$ est un système complet dénombrable d'événements ; de même pour la seconde, avec $((X = x))_{x \in X(\Omega)}$.

En revanche, considérons l'exemple suivant, où l'on définit les lois de deux couples (X_1, Y_1) et (X_2, Y_2) :

(x, y)	(0,0)	(0,1)	(1,0)	(1,1)
$P(X_1 = x, Y_1 = y)$	0,25	0,25	0,25	0,25
$P(X_2 = x, Y_2 = y)$	0,3	0,2	0,2	0,3

Dans les deux cas, les lois marginales sont les mêmes, car pour $i \in \{1, 2\}$,

$$P(X_i = 0) = P(X_i = 1) = P(Y_i = 0) = P(Y_i = 1) = 0,5$$

mais les lois conjointes ne sont pas les mêmes (car $P(X_1 = 0, Y_1 = 0) \neq P(X_2 = 0, Y_2 = 0)$ par exemple).

Les lois marginales du couple (X, Y) ne déterminent donc pas la loi conjointe de X et Y . \square

2. Conditionnement et indépendance

Définition – Loi conditionnelle

Soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) et $y \in Y(\Omega)$ tel que $P(Y = y) > 0$.

On appelle **loi conditionnelle** de X sachant $(Y = y)$ la fonction

$$\begin{cases} X(\Omega) & \rightarrow [0,1] \\ x & \mapsto P(X = x | Y = y) \end{cases}$$

C'est la loi de X en tant que variable aléatoire sur l'espace probabilisé $(\Omega, \mathcal{A}, P_{(Y=y)})$.

On rappelle que pour tout $x \in X(\Omega)$,

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

Exemple – Dans l'exemple de la propriété précédente, on a

$$P(Y_2 = 0) = P(X_2 = 0, Y_2 = 0) + P(X_2 = 1, Y_2 = 0) = 0,3 + 0,2 = 0,5 > 0.$$

La loi de X_2 sachant $(Y_2 = 0)$ est caractérisée par les deux nombres

$$P(X_2 = 0 | Y_2 = 0) = \frac{0,3}{0,5} = 0,6 \quad \text{et} \quad P(X_2 = 1 | Y_2 = 0) = \frac{0,2}{0,5} = 0,4.$$

Définition – Indépendance de variables aléatoires

- Soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) .

On dit que X et Y sont **indépendantes** si pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$, les événements $(X = x)$ et $(Y = y)$ sont indépendants, *i.e.*

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

- Soit I un ensemble d'indices. Pour tout $i \in I$, soit X_i une variable aléatoire sur (Ω, \mathcal{A}, P) .

On dit que les variables aléatoires X_i , pour $i \in I$, sont **mutuellement indépendantes** si, pour toute famille $(x_i)_{i \in I}$ telle que pour tout $i \in I$, $x_i \in X_i(\Omega)$, les événements $(X_i = x_i)$ pour $i \in I$ sont mutuellement indépendants, *i.e.* : pour toute partie finie $J \subset I$,

$$P\left(\bigcap_{j \in J} (X_j = x_j)\right) = \prod_{j \in J} P(X_j = x_j).$$

Propriété (admise : démonstration hors programme)

- Soient X et Y deux variables aléatoires indépendantes sur (Ω, \mathcal{A}, P) , A un sous-ensemble de $X(\Omega)$ et B un sous-ensemble de $Y(\Omega)$.

Alors les événements $(X \in A)$ et $(Y \in B)$ sont indépendants, *i.e.*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

- Soit $(X_i)_{i \in I}$ une famille de variables aléatoires mutuellement indépendantes sur (Ω, \mathcal{A}, P) .

Alors, pour toute famille $(A_i)_{i \in I}$ telle que pour tout $i \in I$, $A_i \subset X_i(\Omega)$, les événements $(X_i \in A_i)$ pour $i \in I$ sont mutuellement indépendants, *i.e.* : pour toute partie finie $J \subset I$,

$$P\left(\bigcap_{j \in J} (X_j \in A_j)\right) = \prod_{j \in J} P(X_j \in A_j).$$

Propriété

Soient X et Y deux variables aléatoires indépendantes sur (Ω, \mathcal{A}, P) .

Soient f et g des fonctions définies respectivement sur $X(\Omega)$ et $Y(\Omega)$.

Alors les variables aléatoires $f(X)$ et $g(Y)$ sont indépendantes.

Démonstration – Soit $a \in f(X(\Omega))$ et $b \in g(Y(\Omega))$. Alors

$$P(f(X) = a, g(Y) = b) = P(X \in f^{-1}(\{a\}), Y \in g^{-1}(\{b\})).$$

Par indépendance de X et Y , et d'après la propriété précédente,

$$P(f(X) = a, g(Y) = b) = P(X \in f^{-1}(\{a\}))P(Y \in g^{-1}(\{b\})) = P(f(X) = a)P(g(Y) = b),$$

d'où le résultat. \square

3. Quelques propriétés des lois usuelles

Propriété – Somme de variables de Bernoulli

Soient X_1, \dots, X_n des variables aléatoires mutuellement indépendantes sur (Ω, \mathcal{A}, P) , suivant chacune la loi de Bernoulli $\mathcal{B}(p)$.

Alors la variable aléatoire $X_1 + \dots + X_n$ suit la loi binomiale $\mathcal{B}(n, p)$.

Démonstration – La démonstration est identique à celle donnée plus haut en interprétation de la loi $\mathcal{B}(n, p)$. \square

Remarque – Des sommes de variables de Bernoulli, comme dans la propriété précédente, sont très utiles pour compter le nombre de succès dans une succession d'épreuves de Bernoulli. On rappelle de plus que de telles variables de Bernoulli peuvent être vues comme des fonctions indicatrices.

Propriété – Caractérisation des lois géométriques comme lois sans mémoire

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) telle que $X(\Omega) = \mathbb{N}^*$.

Les propriétés suivantes sont équivalentes :

1. Il existe $p \in]0, 1[$ tel que $X \hookrightarrow \mathcal{G}(p)$.
2. $P(X = 1) > 0$, $P(X > n) > 0$ pour tout $n \in \mathbb{N}$ et

$$\forall (n, k) \in \mathbb{N}^2, \quad P(X > n + k | X > n) = P(X > k).$$

La loi d'une variable aléatoire vérifiant **2** est dite **loi sans mémoire** (ou sans vieillissement).

Ainsi, les lois géométriques sont exactement les lois sans mémoire.

Démonstration

1 \Rightarrow **2** : supposons que $X \hookrightarrow \mathcal{G}(p)$ avec $p \in]0, 1[$. Alors $P(X = 1) = p > 0$ et, pour tout $n \in \mathbb{N}$,

$$P(X > n) = \sum_{j=n+1}^{+\infty} P(X = j) = \sum_{j=n+1}^{+\infty} p(1-p)^{j-1} = p \frac{(1-p)^n}{1 - (1-p)} = (1-p)^n.$$

En particulier, $P(X > n) > 0$ pour tout $n \in \mathbb{N}$. Soit $(n, k) \in \mathbb{N}^2$. Alors

$$\begin{aligned} P(X > n + k | X > n) &= \frac{P(X > n + k, X > n)}{P(X > n)} \\ &= \frac{P(X > n + k)}{P(X > n)} = \frac{(1-p)^{n+k}}{(1-p)^n} = (1-p)^k = P(X > k). \end{aligned}$$

2 \Rightarrow **1** : posons $p = P(X = 1) > 0$. On a aussi $p = 1 - P(X > 1) < 1$. Soit, pour tout $n \in \mathbb{N}$, $x_n = P(X > n)$. D'après la propriété d'absence de mémoire,

$$x_{n+1} = P(X > n + 1) = P(X > n + 1 | X > n) P(X > n) = P(X > 1) P(X > n) = (1-p) x_n.$$

La suite $(x_n)_{n \in \mathbb{N}}$ est donc géométrique de raison $1-p$ et de premier terme $x_0 = P(X > 0) = 1$, donc pour tout $n \in \mathbb{N}$, $x_n = (1-p)^n$. Alors, pour tout $n \in \mathbb{N}^*$,

$$\begin{aligned} P(X = n) &= P(X > n - 1) - P(X > n) = (1-p)^{n-1} - (1-p)^n \\ &= (1-p)^{n-1} (1 - (1-p)) \\ &= p (1-p)^{n-1}. \end{aligned}$$

Finalement, $p \in]0, 1[$ et $X \hookrightarrow \mathcal{G}(p)$. \square

Remarque – Comme on l’a dit plus haut, la loi $\mathcal{G}(p)$ modélise souvent une durée de fonctionnement, ou plus généralement un temps d’attente avant qu’un phénomène se produise. La propriété d’absence de mémoire signifie que ce temps d’attente est indépendant de l’étape à laquelle on commence à attendre.

4. Indépendance et modélisation

Comme nous l’avons déjà vu, la modélisation d’une expérience aléatoire par le choix de (Ω, \mathcal{A}, P) n’est pas toujours évidente. En fait, elle n’est parfois pas utile, le fait de préciser les conditions de l’expérience, ce qui est plus intuitif, étant souvent suffisant. C’est ce que permet de faire le résultat suivant :

Théorème (admis : démonstration hors programme)

Soit I un ensemble d’indices fini ou dénombrable. Pour tout $i \in I$, on se donne une loi discrète \mathcal{L}_i (ce qui revient à se donner une famille ou une suite de nombres positifs de somme 1).

Alors il existe un espace probabilisé (Ω, \mathcal{A}, P) et une famille $(X_i)_{i \in I}$ de variables aléatoires sur (Ω, \mathcal{A}, P) , mutuellement indépendantes, tels que pour tout $i \in I$, X_i suit la loi \mathcal{L}_i .

Il est ainsi possible de modéliser une succession, finie ou infinie, d’expériences aléatoires mutuellement indépendantes, par le choix des lois de variables aléatoires, sans avoir à préciser (Ω, \mathcal{A}, P) .

Exemples

- Un jeu de pile ou face, fini ou infini, avec indépendance mutuelle des différents lancers, pourra être modélisé par le choix d’une suite $(X_i)_{i \in I}$, finie ou infinie, de variables de Bernoulli mutuellement indépendantes de même paramètre p . Pour tout $i \in I$, X_i représente le résultat du i -ième lancer (1 pour « pile », de probabilité p , 0 pour « face », par exemple).
- On considère la situation suivante : une urne contient des jetons rouges en proportion p , et blancs en proportion $1 - p$; N personnes tirent successivement, avec remise, n jetons dans l’urne, le gain de chaque personne étant lié au nombre de jetons rouges tirés.

On pourra modéliser cette situation par une famille (X_1, \dots, X_N) de N variables aléatoires mutuellement indépendantes, suivant chacune la loi binomiale $\mathcal{B}(n, p)$. Pour tout $i \in \llbracket 1, N \rrbracket$, X_i représente le nombre de jetons rouges tirés par le i -ième participant.

IV. Espérance

Définition – Espérance

Soit X une variable aléatoire réelle sur (Ω, \mathcal{A}, P) , avec $X(\Omega)$ dénombrable ; on décrit $X(\Omega)$ en extension sous la forme $\{x_n; n \in \mathbb{N}\}$.

On dit que X est d’**espérance finie** si la série

$$\sum_{n \geq 0} x_n P(X = x_n)$$

est absolument convergente.

Dans ce cas, la somme de cette série est appelée **espérance** de X , et notée $E(X)$, c’est-à-dire,

$$E(X) = \sum_{n=0}^{+\infty} x_n P(X = x_n).$$

Remarques

- L'espérance de X est à interpréter comme moyenne pondérée des valeurs de X . Par exemple en physique, elle représente l'énergie moyenne de systèmes à spectre discret (comme un atome confiné dans une boîte).
- La notion d'espérance de X dépend de X uniquement à travers sa loi.
- La définition précédente semble dépendre du choix des x_n (c'est-à-dire de l'ordre d'énumération des éléments de $X(\Omega)$). On *admettra* que lorsque X est d'espérance finie, la somme définissant $E(X)$ ne dépend pas de l'ordre d'énumération.
- Si $X(\Omega)$ est fini avec $X(\Omega) = \{x_1, \dots, x_m\}$, alors X est d'espérance finie, et $E(X)$ est simplement définie par :

$$E(X) = \sum_{n=1}^m x_n P(X = x_n).$$

- S'il existe $a \in \mathbb{R}$ tel que $P(X = a) = 1$, alors X est d'espérance finie égale à a .
- Si Ω est fini, on a la relation $E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$.

Propriété – Espérance correspondant aux lois usuelles

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) .

- Si X suit la loi uniforme avec $X(\Omega) = \{x_1, \dots, x_m\}$, alors X est d'espérance finie avec

$$E(X) = \frac{1}{m} \sum_{n=1}^m x_n.$$

- Si $X \hookrightarrow \mathcal{B}(p)$, alors X est d'espérance finie et $E(X) = p$.
- Si $X \hookrightarrow \mathcal{B}(n, p)$, alors X est d'espérance finie et $E(X) = np$.
- Si $X \hookrightarrow \mathcal{G}(p)$, alors X est d'espérance finie et $E(X) = \frac{1}{p}$.
- Si $X \hookrightarrow \mathcal{P}(\lambda)$, alors X est d'espérance finie et $E(X) = \lambda$.

Démonstration

- Pour tout $n \in \llbracket 1, m \rrbracket$, $P(X = x_n) = 1/m$, d'où le résultat.
- Si $X \hookrightarrow \mathcal{B}(p)$, on a $E(X) = 0 \times (1 - p) + 1 \times p = p$.
- Si $X \hookrightarrow \mathcal{B}(n, p)$,

$$E(X) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k}.$$

Avec le changement d'indice $j = k - 1$, on obtient

$$\begin{aligned} E(X) &= n \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j+1} (1-p)^{(n-1)-j} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j} = np (p + (1-p))^{n-1} = np. \end{aligned}$$

- Supposons que $X \hookrightarrow \mathcal{G}(p)$. La série (à termes positifs) $\sum_{n \geq 1} np(1-p)^{n-1}$ est convergente : on reconnaît la dérivée de la série géométrique évaluée en $1-p$ avec $|1-p| < 1$. Donc X est d'espérance finie et

$$E(X) = p \frac{1}{(1 - (1-p))^2} = \frac{1}{p}.$$

- Supposons que $X \hookrightarrow \mathcal{P}(\lambda)$. Pour tout $n \in \mathbb{N}^*$,

$$n e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \frac{\lambda^n}{(n-1)!},$$

terme général (positif) d'une série convergente (série exponentielle). Donc X est d'espérance finie et avec un changement d'indice, on obtient

$$E(X) = \lambda e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} e^\lambda = \lambda.$$

□

Propriété

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) à valeurs dans \mathbb{N} .

La variable aléatoire X est d'espérance finie si et seulement si la série $\sum_{n \geq 1} P(X \geq n)$ converge, et dans ce cas on a

$$E(X) = \sum_{n=1}^{+\infty} P(X \geq n).$$

Démonstration – Pour tout $n \in \mathbb{N}$, on a

$$(X \geq n) = (X = n) \cup (X \geq n + 1),$$

ces deux événements étant incompatibles, et donc

$$P(X = n) = P(X \geq n) - P(X \geq n + 1).$$

Alors, pour tout $p \in \mathbb{N}^*$,

$$\begin{aligned} \sum_{n=0}^p n P(X = n) &= \sum_{n=0}^p n (P(X \geq n) - P(X \geq n + 1)) \\ &= \sum_{n=0}^p n P(X \geq n) - \sum_{n=1}^{p+1} (n-1) P(X \geq n) \end{aligned}$$

après séparation des sommes et changement d'indice dans la deuxième somme. Finalement,

$$\sum_{n=0}^p n P(X = n) = \left(\sum_{n=1}^p P(X \geq n) \right) - p P(X \geq p + 1). \quad (14.1)$$

Si X est d'espérance finie, alors on peut écrire

$$0 \leq p P(X \geq p + 1) = p \sum_{n=p+1}^{+\infty} P(X = n) \leq \sum_{n=p+1}^{+\infty} n P(X = n) \xrightarrow{p \rightarrow +\infty} 0$$

en tant que reste d'une série convergente. On en déduit que $\sum_{n \geq 1} P(X \geq n)$ converge ainsi que l'égalité souhaitée en faisant tendre p vers $+\infty$.

Par positivité des termes, et d'après (14.1), si $\sum_{n \geq 1} P(X \geq n)$ converge, alors

$$\sum_{n \geq 1} n P(X = n)$$

converge (la suite de ses sommes partielles est majorée) donc X est d'espérance finie. On conclut comme précédemment. □

Théorème de transfert (admis : démonstration hors-programme)

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) avec $X(\Omega)$ dénombrable ; on décrit $X(\Omega)$ en extension sous la forme $\{x_n; n \in \mathbb{N}\}$. Soit $f : X(\Omega) \rightarrow \mathbb{R}$ une fonction.

La variable aléatoire $f(X)$ est d'espérance finie si et seulement si la série $\sum_{n \geq 0} f(x_n) P(X = x_n)$ converge absolument, et dans ce cas, on a

$$E(f(X)) = \sum_{n=0}^{+\infty} f(x_n) P(X = x_n).$$

Remarque – Si l'on appliquait la définition de l'espérance pour $f(X)$, on devrait déterminer la loi de $f(X)$: on devrait décrire $f(X(\Omega))$ en extension sous la forme $\{y_n; n \in I\}$ (I fini ou $I = \mathbb{N}$) puis considérer la somme finie ou la série $\sum_{n \in I} y_n P(f(X) = y_n)$.

L'immense avantage du théorème de transfert est de montrer qu'il suffit en fait de considérer la loi de X . On a *transféré* le calcul de $E(f(X))$ sur la variable aléatoire X . Ceci est particulièrement intéressant lorsque f n'est pas injective.

Exemple – Soit X une variable aléatoire suivant la loi géométrique de paramètre p . D'après le théorème de transfert, si la série

$$\sum_{n \geq 1} (-1)^n p (1-p)^{n-1}$$

converge absolument, alors $(-1)^X$ est d'espérance finie et la somme de cette série est $E((-1)^X)$. On reconnaît (à un facteur $-p$ près) la série géométrique de raison $p-1$ avec $|p-1| < 1$, donc absolument convergente. On en déduit que $(-1)^X$ est d'espérance finie avec

$$E((-1)^X) = \sum_{n=1}^{+\infty} (-1)^n p (1-p)^{n-1} = -p \frac{1}{1-(p-1)} = \frac{p}{p-2}.$$

Théorème – Quelques propriétés de l'espérance

Soient X et Y deux variables aléatoires d'espérance finie sur (Ω, \mathcal{A}, P) et $\lambda \in \mathbb{R}$. Alors :

- **Linéarité** : $\lambda X + Y$ est d'espérance finie et $E(\lambda X + Y) = \lambda E(X) + E(Y)$.
- **Positivité** : si $P(X \geq 0) = 1$, alors $E(X) \geq 0$.
- **Croissance** : si $P(X \leq Y) = 1$, alors $E(X) \leq E(Y)$.

Démonstration

- La démonstration de la linéarité de l'espérance n'est pas exigible.

Considérons le couple (X, Y) et lorsque $X(\Omega) \times Y(\Omega)$ est dénombrable, décrivons-le en extension sous la forme $\{(x_n, y_n)\}; n \in \mathbb{N}\}$. Soit f une fonction définie sur $X(\Omega) \times Y(\Omega)$, à valeurs dans \mathbb{R} ; d'après le théorème de transfert, la série $\sum_{n \geq 0} f(x_n, y_n) P(X = x_n, Y = y_n)$ est absolument convergente si et seulement si $f(X, Y)$ est d'espérance finie, et dans ce cas

$$E(f(X, Y)) = \sum_{n=0}^{+\infty} f(x_n, y_n) P(X = x_n, Y = y_n).$$

Nous allons utiliser ce résultat avec $f : (x, y) \mapsto x$, $f : (x, y) \mapsto y$ et $f : (x, y) \mapsto \lambda x + y$. Les séries

$$\sum_{n \geq 0} x_n P(X = x_n, Y = y_n) \quad \text{et} \quad \sum_{n \geq 0} y_n P(X = x_n, Y = y_n)$$

sont absolument convergentes car X et Y sont d'espérance finie. Par combinaison linéaire, la série

$$\sum_{n \geq 0} (\lambda x_n + y_n) P(X = x_n, Y = y_n)$$

est absolument convergente, donc $\lambda X + Y$ est d'espérance finie ; on a alors

$$E(\lambda X + Y) = \lambda \sum_{n=0}^{+\infty} x_n P(X = x_n, Y = y_n) + \sum_{n=0}^{+\infty} y_n P(X = x_n, Y = y_n) = \lambda E(X) + E(Y).$$

On adapte la démonstration avec des sommes finies si $X(\Omega) \times Y(\Omega)$ est fini.

• On décrit $X(\Omega)$ en extension sous la forme $\{x_n; n \in I\}$. On a $P(X < 0) = 0$, donc pour tout n tel que $x_n < 0$, $x_n P(X = x_n) = 0$. Donc on peut écrire $E(X)$ comme somme d'une série (ou somme finie) à termes positifs, d'où $E(X) \geq 0$.

• Cela résulte des deux points précédents. □

Application – On retrouve facilement l'espérance d'une variable aléatoire suivant la loi $\mathcal{B}(n, p)$ en utilisant la linéarité de l'espérance : soient X_1, \dots, X_n des variables aléatoires mutuellement indépendantes suivant la même loi $\mathcal{B}(p)$ (on sait qu'il existe un espace probabilisé portant de telles lois). Alors on sait que $S = X_1 + \dots + X_n$ suit la loi $\mathcal{B}(n, p)$. Par linéarité de l'espérance, on a donc

$$E(S) = \sum_{k=1}^n E(X_k) = np$$

car $E(X_k) = p$ pour tout k . L'espérance ne dépendant que de la loi, on obtient ainsi l'espérance de toutes les variables aléatoires suivant la loi $\mathcal{B}(n, p)$.

Propriété

Soient X et Y deux variables aléatoires indépendantes sur (Ω, \mathcal{A}, P) , d'espérance finie. Alors XY est d'espérance finie et

$$E(XY) = E(X)E(Y).$$

La réciproque est fautive en général.

La démonstration est hors-programme dans le cas général. Dans le cas des univers finis, elle a été donnée en première année. □

Exemple – Marche aléatoire

Reprenons un exemple décrit plus haut : une particule peut occuper différentes positions repérées par les entiers relatifs. À intervalle régulier, la particule peut passer de la position i à la position $i + 1$ avec probabilité $p \in]0, 1[$, ou à la position $i - 1$ avec probabilité $q = 1 - p$. On suppose qu'un mouvement ne dépend que de la position à partir de laquelle il est fait. Pour $n \geq 1$, on note X_n la variable aléatoire représentant la position de la particule après n mouvements ; X_0 est la variable aléatoire nulle (la position initiale est 0). On admet l'existence d'un espace probabilisé (Ω, \mathcal{A}, P) modélisant cette expérience.

On cherche à étudier différents aspects de cette marche aléatoire.

• **Loi de X_1 et X_2** : X_1 prend les valeurs 1 et -1 , avec $P(X_1 = 1) = p$, $P(X_1 = -1) = q$. On en déduit que X_2 prend les valeurs $-2, 0$ et 2 . D'après la formule des probabilités totales,

$$\begin{aligned} P(X_2 = 2) &= P(X_2 = 2 | X_1 = 1)P(X_1 = 1) + P(X_2 = 2 | X_1 = -1)P(X_1 = -1) \\ &= pP(X_1 = 1) + 0 \times P(X_1 = -1) = p^2, \end{aligned}$$

$$P(X_2 = 0) = P(X_2 = 0 | X_1 = 1)P(X_1 = 1) + P(X_2 = 0 | X_1 = -1)P(X_1 = -1) = 2pq,$$

$$\begin{aligned} P(X_2 = -2) &= P(X_2 = -2 | X_1 = 1)P(X_1 = 1) + P(X_2 = -2 | X_1 = -1)P(X_1 = -1) \\ &= 0 \times P(X_1 = 1) + qP(X_1 = -1) = q^2. \end{aligned}$$

• La particule ne peut revenir en 0 qu'après un nombre pair de mouvements, ainsi, pour tout $n \in \mathbb{N}$, $P(X_{2n+1} = 0) = 0$. Pour $n \in \mathbb{N}$, la particule est à l'origine après $2n$ mouvements si et

seulement si elle a effectué n mouvements à droite et n mouvements à gauche. Le nombre de mouvements à droite parmi les $2n$ premiers suit la loi $\mathcal{B}(2n, p)$, donc

$$P(X_{2n} = 0) = \binom{2n}{n} p^n (1-p)^{2n-n} = \frac{(2n)!}{(n!)^2} (p(1-p))^n.$$

D'après la formule de Stirling,

$$\frac{(2n)!}{(n!)^2} \sim \frac{\left(\frac{2n}{e}\right)^{2n} \sqrt{4\pi n}}{\left(\frac{n}{e}\right)^{2n} 2\pi n} = \frac{4^n}{\sqrt{n\pi}}$$

et finalement,

$$P(X_{2n} = 0) \sim \frac{1}{\sqrt{n\pi}} (4p(1-p))^n.$$

• La variable aléatoire $\mathbb{1}_{(X_2=0)} + \dots + \mathbb{1}_{(X_{2n}=0)}$ représente le nombre de retours à l'origine au cours des $2n$ premiers mouvements. Par linéarité de l'espérance (pour tout $A \in \mathcal{A}$, la variable aléatoire $\mathbb{1}_A$ est d'espérance finie égale à $P(A)$),

$$E(\mathbb{1}_{(X_2=0)} + \dots + \mathbb{1}_{(X_{2n}=0)}) = \sum_{k=1}^n P(X_{2k} = 0).$$

Remarquons que l'on a calculé cette espérance sans déterminer la loi du nombre de retours.

– Si $p \neq 1/2$, $0 < 4p(1-p) < 1$, et par comparaison de séries à termes positifs, la série de terme général $P(X_{2n} = 0)$ converge. L'espérance du nombre de retours à l'origine est majorée indépendamment du nombre de mouvements.

– Si $p = 1/2$, $P(X_{2n} = 0) \sim \frac{1}{\sqrt{n\pi}}$ et la série de terme général $P(X_{2n} = 0)$ (à termes positifs) diverge par comparaison avec une série de Riemann d'exposant $1/2 < 1$. Un résultat sur les sommes partielles de séries à termes positifs divergentes, puis une comparaison série/intégrale (que nous ne détaillons pas ici), montrent alors que

$$\sum_{k=1}^n P(X_{2k} = 0) \sim \sum_{k=1}^n \frac{1}{\sqrt{k\pi}} \sim 2\sqrt{\frac{n}{\pi}}.$$

Cette espérance tend vers $+\infty$ lorsque $n \rightarrow +\infty$: en un temps illimité, il y a en moyenne une infinité de retours à l'origine !

V. Séries génératrices des variables aléatoires à valeurs dans \mathbb{N}

Propriété/Définition – Série génératrice

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{N} .

Alors, pour tout $t \in [-1, 1]$, la variable aléatoire t^X est d'espérance finie. On pose, pour tout $t \in [-1, 1]$,

$$G_X(t) = E(t^X), \quad \text{et on a} \quad G_X(t) = \sum_{n=0}^{+\infty} P(X = n) t^n.$$

La fonction G_X est la somme d'une série entière de rayon de convergence au moins égal à 1. Elle est appelée **série génératrice** (ou **fonction génératrice**) de X .

Démonstration – On peut considérer que $X(\Omega) = \mathbb{N}$. Soit $t \in [-1,1]$. D'après le théorème de transfert, t^X est d'espérance finie si et seulement si la série

$$\sum_{n \geq 0} P(X = n) t^n$$

converge absolument. Or, pour tout $n \in \mathbb{N}$, $|P(X = n) t^n| \leq P(X = n)$, et $\sum_{n \geq 0} P(X = n)$ converge (et sa somme vaut 1). Par comparaison, on en déduit l'existence de $E(t^X)$; la formule donnant $E(t^X)$ provient aussi du théorème de transfert.

Sachant que la série entière définissant G_X converge absolument en tout point de $[-1,1]$, son rayon de convergence est au moins égal à 1. \square

Remarques

- On a $G_X(1) = \sum_{n=0}^{+\infty} P(X = n) = 1$.
- Lorsque $X(\Omega)$ est fini, G_X est un polynôme (et $R = +\infty$).

Propriété

La loi d'une variable aléatoire à valeurs dans \mathbb{N} est caractérisée par sa série génératrice : soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{N} , telles que $X(\Omega) = Y(\Omega)$ et $G_X(t) = G_Y(t)$ pour tout $t \in]-r, r[$ (pour un certain $r \in]0,1[$). Alors X et Y ont la même loi.

Démonstration – Si $G_X(t) = G_Y(t)$ pour tout $t \in [-1,1]$, alors par unicité du développement en série entière, $P(X = n) = P(Y = n)$ pour tout $n \in \mathbb{N}$. \square

Remarque – La série génératrice de X contient donc toute l'information sur la loi de X . On a en fait, d'après l'expression des coefficients d'une série entière : pour tout $n \in \mathbb{N}$,

$$P(X = n) = \frac{G_X^{(n)}(0)}{n!}$$

Propriété – Lien avec l'espérance

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{N} . Alors, pour que X soit d'espérance finie, il faut et il suffit que G_X soit dérivable à gauche en 1. Dans ce cas, on a

$$E(X) = G'_X(1).$$

Démonstration (non exigible)

\Rightarrow Posons, pour tout $n \in \mathbb{N}$, $f_n : t \mapsto P(X = n) t^n$. La série de fonctions $\sum_{n \geq 0} f_n$ converge simplement sur $[-1,1]$; pour tout $n \in \mathbb{N}$, f_n est de classe \mathcal{C}^1 sur $[-1,1]$ avec pour tout $n \in \mathbb{N}^*$ et $t \in [-1,1]$,

$$|f'_n(t)| = |n P(X = n) t^{n-1}| \leq n P(X = n).$$

Le majorant est le terme général d'une série convergente car X est d'espérance finie. D'après le théorème de la classe \mathcal{C}^1 pour les séries de fonctions, G_X est de classe \mathcal{C}^1 sur $[-1,1]$, et en particulier dérivable à gauche en 1. On a de plus

$$G'_X(1) = \sum_{n=0}^{+\infty} f'_n(1) = \sum_{n=1}^{+\infty} n P(X = n) = E(X).$$

\Leftarrow Soit $p \in \mathbb{N}^*$. Pour tout $t \in [0,1[$,

$$\frac{G_X(t) - G_X(1)}{t - 1} \geq \sum_{n=0}^p P(X = n) \frac{t^n - 1}{t - 1} = \sum_{n=1}^p P(X = n) (1 + t + \dots + t^{n-1}),$$

l'inégalité étant valable par positivité des termes. Lorsque $t \rightarrow 1^-$, on en déduit que

$$\sum_{n=1}^p n P(X = n) \leq G'_X(1).$$

pour tout $p \in \mathbb{N}^*$. La série à termes positifs $\sum_{n \geq 0} n P(X = n)$ est donc à sommes partielles majorées indépendamment de p , donc convergente, ce qui entraîne (à nouveau par positivité des termes) que X est d'espérance finie. \square

Propriété – Séries génératrices correspondant aux lois usuelles

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) .

- Si $X \hookrightarrow \mathcal{B}(p)$, alors pour tout $t \in \mathbb{R}$, $G_X(t) = 1 - p + pt$.
- Si $X \hookrightarrow \mathcal{B}(n, p)$, alors pour tout $t \in \mathbb{R}$, $G_X(t) = (1 - p + pt)^n$.
- Si $X \hookrightarrow \mathcal{G}(p)$, alors pour tout t tel que $|(1 - p)t| < 1$, $G_X(t) = \frac{pt}{1 - (1 - p)t}$.
- Si $X \hookrightarrow \mathcal{P}(\lambda)$, alors pour tout $t \in \mathbb{R}$, $G_X(t) = e^{\lambda(t-1)}$.

Démonstration

- Si $X \hookrightarrow \mathcal{B}(p)$, on a pour tout $t \in \mathbb{R}$, $G_X(t) = P(X = 0) + P(X = 1)t = 1 - p + pt$.
- Si $X \hookrightarrow \mathcal{B}(n, p)$, on a pour tout $t \in \mathbb{R}$,

$$G_X(t) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} t^k = \sum_{k=0}^n \binom{n}{k} (pt)^k (1 - p)^{n-k} = (1 - p + pt)^n$$

d'après la formule du binôme de Newton.

- Supposons que $X \hookrightarrow \mathcal{G}(p)$. La série génératrice de X est la fonction somme de la série entière

$$\sum_{n \geq 1} p (1 - p)^{n-1} t^n.$$

On reconnaît une série géométrique de raison $(1 - p)t$. Elle converge si et seulement si $|(1 - p)t| < 1$, et dans ce cas

$$G_X(t) = pt \sum_{n=0}^{+\infty} ((1 - p)t)^n = \frac{pt}{1 - (1 - p)t}.$$

- Supposons que $X \hookrightarrow \mathcal{P}(\lambda)$. La série génératrice de X est la fonction somme de la série entière

$$\sum_{n \geq 0} e^{-\lambda} \frac{\lambda^n}{n!} t^n.$$

On reconnaît une série exponentielle; elle converge pour tout $t \in \mathbb{R}$, et

$$\forall t \in \mathbb{R}, \quad G_X(t) = e^{-\lambda} \sum_{n=0}^{+\infty} \frac{(\lambda t)^n}{n!} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}. \quad \square$$

Propriété – Somme de deux variables aléatoires indépendantes

Soient X et Y deux variables aléatoires indépendantes sur (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{N} .

Alors, pour tout $t \in [-1, 1]$,

$$G_{X+Y}(t) = G_X(t) G_Y(t).$$

Démonstration – La variable $X + Y$ est à valeurs dans \mathbb{N} de même que X et Y . Les variables X et Y sont indépendantes, donc pour tout $t \in [-1,1]$, t^X et t^Y sont indépendantes. On en déduit que

$$G_{X+Y}(t) = E(t^{X+Y}) = E(t^X t^Y) = E(t^X)E(t^Y) = G_X(t) G_Y(t).$$

□

Remarque – Soit $n \in \mathbb{N}$; on a

$$(X + Y = n) = \bigcup_{k=0}^n (X = k, Y = n - k),$$

ces événements étant deux à deux incompatibles, d'où, par indépendance,

$$P(X + Y = n) = \sum_{k=0}^n P(X = k, Y = n - k) = \sum_{k=0}^n P(X = k) P(Y = n - k).$$

On connaît donc la loi de $X + Y$. Par produit de Cauchy de deux séries entières absolument convergentes, on a pour tout $t \in [-1,1]$,

$$G_X(t) G_Y(t) = \sum_{n=0}^{+\infty} \left(\sum_{k=0}^n P(X = k) P(Y = n - k) \right) t^n = \sum_{n=0}^{+\infty} P(X + Y = n) t^n = G_{X+Y}(t),$$

ce qui donne une autre démonstration de la propriété précédente.

Corollaire – Somme de variables aléatoires suivant une loi de Poisson

Soient X et Y deux variables aléatoires indépendantes sur (Ω, \mathcal{A}, P) et λ, μ deux réels strictement positifs. On suppose que $X \hookrightarrow \mathcal{P}(\lambda)$ et $Y \hookrightarrow \mathcal{P}(\mu)$.

Alors $X + Y \hookrightarrow \mathcal{P}(\lambda + \mu)$.

Démonstration – Pour tout $n \in \mathbb{N}$, $P(X = n, Y = 0) = P(X = n) P(Y = 0)$ par indépendance, donc $P(X + Y = n) > 0$. On en déduit que $(X + Y)(\Omega) = \mathbb{N}$. De plus, pour tout $t \in [-1,1]$ (en fait pour tout $t \in \mathbb{R}$),

$$G_{X+Y}(t) = G_X(t) G_Y(t) = e^{\lambda(t-1)} e^{\mu(t-1)} = e^{(\lambda+\mu)(t-1)}.$$

La série génératrice caractérisant la loi, on en déduit que $X + Y \hookrightarrow \mathcal{P}(\lambda + \mu)$. □

VI. Variance

1. Généralités

L'espérance de X correspond à la moyenne pondérée des valeurs de X , mais ne décrit pas comment sont réparties les valeurs de X autour de cette moyenne. C'est l'intérêt des notions de variance et d'écart-type.

Propriété

Soit X une variable aléatoire réelle sur (Ω, \mathcal{A}, P) . On suppose que X^2 est d'espérance finie. Alors :

- X est d'espérance finie.
- $(X - E(X))^2$ est d'espérance finie.

Démonstration

• Le problème ne se pose que si $X(\Omega)$ est dénombrable. On écrit $X(\Omega) = \{x_n; n \in \mathbb{N}\}$. La variable aléatoire X^2 est d'espérance finie, donc d'après le théorème de transfert, $\sum_{n \geq 0} x_n^2 P(X = x_n)$ converge et sa somme est $E(X^2)$. Pour tout $p \in \mathbb{N}$, on a d'après l'inégalité de Cauchy-Schwarz,

$$\begin{aligned} \sum_{n=0}^p |x_n| P(X = x_n) &= \sum_{n=0}^p \left(|x_n| \sqrt{P(X = x_n)} \right) \sqrt{P(X = x_n)} \\ &\leq \left(\sum_{n=0}^p x_n^2 P(X = x_n) \sum_{n=0}^p P(X = x_n) \right)^{1/2} \\ &\leq \left(\sum_{n=0}^{+\infty} x_n^2 P(X = x_n) \sum_{n=0}^{+\infty} P(X = x_n) \right)^{1/2} \\ &= \sqrt{E(X^2)} \end{aligned}$$

car $\sum_{n=0}^{+\infty} P(X = x_n) = 1$.

Les sommes partielles de la série à termes positifs $\sum_{n \geq 0} |x_n| P(X = x_n)$ sont majorées indépendamment de p , donc cette série converge, ce qui prouve le résultat. En passant à la limite dans les inégalités précédentes, on obtient même : $E(|X|) \leq \sqrt{E(X^2)}$.

• On a $(X - E(X))^2 = X^2 - 2E(X)X + E(X)^2$. Si X^2 est d'espérance finie, X également, et donc par combinaison linéaire, $(X - E(X))^2$ est d'espérance finie. \square

Cette propriété permet de donner la définition suivante :

Propriété/Définition – Variance et écart-type

Soit X une variable aléatoire réelle sur (Ω, \mathcal{A}, P) . On dit que X admet une **variance** (ou admet un **moment d'ordre 2**) si X^2 est d'espérance finie. Dans ce cas :

• On appelle **variance** de X le réel positif

$$V(X) = E((X - E(X))^2).$$

On a aussi $V(X) = E(X^2) - E(X)^2$.

• On appelle **écart-type** de X le réel positif $\sigma(X) = \sqrt{V(X)}$.

Démonstration de la seconde expression de $V(X)$

D'après la propriété précédente, $(X - E(X))^2 = X^2 - 2E(X)X + E(X)^2$ est d'espérance finie ; par linéarité de l'espérance,

$$V(X) = E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2. \quad \square$$

Remarques

• Si X^2 est d'espérance finie, le moment d'ordre 2 de X est le réel positif $E(X^2)$.

• Si $X(\Omega) = \{x_n; n \in \mathbb{N}\}$, d'après le théorème de transfert, X a une variance si et seulement si la série à termes positifs $\sum_{n \geq 0} x_n^2 P(X = x_n)$ converge, et dans ce cas,

$$V(X) = \sum_{n=0}^{+\infty} (x_n - E(X))^2 P(X = x_n).$$

• Si X admet une variance et $m = E(X)$, on a $V(X) = 0$ si et seulement si $P(X = m) = 1$.

Exemple – Soit X une variable aléatoire prenant les valeurs 1 et -1 et suivant la loi uniforme, et soit Y la variable aléatoire nulle. Alors X et Y sont toutes les deux d'espérance nulle. Pourtant, elles se comportent très différemment ; la variance est un moyen de mesurer cette différence : on a

$$V(X) = E((X - 0)^2) = E(X^2) = 1 \quad \text{et} \quad V(Y) = 0.$$

Propriété

Soit X une variable aléatoire réelle sur (Ω, \mathcal{A}, P) , admettant une variance, et $(a, b) \in \mathbb{R}^2$.
Alors $aX + b$ admet une variance et on a : $V(aX + b) = a^2 V(X)$.

Démonstration – On a $(aX + b)^2 = a^2 X^2 + 2abX + b^2$ et X^2 est d'espérance finie donc X également. Par combinaison linéaire, $aX + b$ a une variance et par linéarité de l'espérance,

$$\begin{aligned} E((aX + b)^2) &= a^2 E(X^2) + 2abE(X) + b^2 \\ (E(aX + b))^2 &= (aE(X) + b)^2 = a^2 E(X)^2 + 2abE(X) + b^2. \end{aligned}$$

Par différence, on en déduit que

$$V(aX + b) = a^2(E(X^2) - E(X)^2) = a^2 V(X). \quad \square$$

Remarque – Cette propriété est cohérente avec l'interprétation de $V(X)$ et $\sigma(X)$ comme indicateurs de dispersion des valeurs de X autour de son espérance : ajouter une même valeur b à toutes les valeurs de X ne modifie pas la variance et l'écart-type, multiplier toutes les valeurs de X par un réel a multiplie l'écart-type par $|a|$.

Propriété – Lien entre série génératrice et variance

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{N} .
Pour que X admette une variance, il faut et il suffit que G_X soit deux fois dérivable à gauche en 1. Dans ce cas,

$$V(X) = G_X''(1) + G_X'(1) - G_X'(1)^2.$$

Ce résultat est admis (démonstration non exigible). Il s'agit d'adapter la démonstration faisant le lien entre l'existence de $E(X)$ et celle de $G_X'(1)$. Expliquons simplement comment retrouver la formule donnant $V(X)$: en cas d'existence, on montre que $G_X'(t)$ et $G_X''(t)$ se calculent, pour $t \in [-1, 1]$, par dérivation terme à terme avec

$$\begin{aligned} G_X'(t) &= \sum_{n=1}^{+\infty} n P(X = n) t^{n-1}, & G_X''(t) &= \sum_{n=2}^{+\infty} n(n-1) P(X = n) t^{n-2} \\ G_X'(1) &= \sum_{n=0}^{+\infty} n P(X = n) = E(X), & G_X''(1) &= \sum_{n=0}^{+\infty} n(n-1) P(X = n) = E(X(X-1)). \end{aligned}$$

D'après le théorème de transfert, et par linéarité de l'espérance,

$$V(X) = E(X^2) - E(X)^2 = E(X(X-1)) + E(X) - E(X)^2 = G_X''(1) + G_X'(1) - G_X'(1)^2. \quad \square$$

Propriété – Variance correspondant aux lois usuelles

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) .

- Si $X \hookrightarrow \mathcal{B}(p)$, alors X admet une variance et $V(X) = p(1-p)$.
- Si $X \hookrightarrow \mathcal{B}(n, p)$, alors X admet une variance et $V(X) = np(1-p)$.
- Si $X \hookrightarrow \mathcal{G}(p)$, alors X admet une variance et $V(X) = \frac{1-p}{p^2}$.
- Si $X \hookrightarrow \mathcal{P}(\lambda)$, alors X admet une variance et $V(X) = \lambda$.

Démonstration

- Si $X \hookrightarrow \mathcal{B}(p)$, on a $E(X^2) = 0^2 \times (1-p) + 1^2 \times p = p$. Alors

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p).$$

- Si $X \hookrightarrow \mathcal{B}(n,p)$, on sait que $G_X(t) = (1-p+pt)^n$ pour tout $t \in \mathbb{R}$. La fonction G_X est deux fois dérivable en 1, donc X admet une variance, et

$$V(X) = G_X''(1) + G_X'(1) - G_X'(1)^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

- Supposons que $X \hookrightarrow \mathcal{G}(p)$. On sait que $G_X(t) = \frac{pt}{1-(1-p)t}$ notamment pour tout $t \in [-1,1]$. La fonction G_X est deux fois dérivable sur $[-1,1]$, avec

$$\forall t \in [-1,1], \quad G_X'(t) = \frac{p}{(1-(1-p)t)^2}, \quad G_X''(t) = \frac{2p(1-p)}{(1-(1-p)t)^3}.$$

En particulier, X admet une variance, et

$$V(X) = G_X''(1) + G_X'(1) - G_X'(1)^2 = \frac{2p(1-p)}{p^3} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

- Supposons que $X \hookrightarrow \mathcal{P}(\lambda)$. On sait que $G_X(t) = e^{\lambda(t-1)}$ pour tout $t \in \mathbb{R}$. La fonction G_X est deux fois dérivable en 1, donc X admet une variance, et

$$V(X) = G_X''(1) + G_X'(1) - G_X'(1)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \quad \square$$

Remarque – On peut calculer toutes ces variances directement à partir du théorème de transfert.

2. Covariance et corrélation

Propriété – Inégalité de Cauchy-Schwarz

Soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) , admettant une variance.

Alors XY est d'espérance finie et

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}.$$

Démonstration – On a $|XY| \leq X^2 + Y^2$; en adaptant la démonstration de la linéarité de l'espérance, on en déduit que XY est d'espérance finie. Quant à l'inégalité de Cauchy-Schwarz, on procède comme pour un produit scalaire, en considérant la fonction polynomiale de degré au plus 2

$$\lambda \mapsto E((\lambda X + Y)^2) = \lambda^2 E(X^2) + 2\lambda E(XY) + E(Y^2),$$

à valeurs positives. □

Définition

Soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) , admettant une variance.

- On appelle **covariance** de X et Y le réel

$$\text{Cov}(X, Y) = E([X - E(X)][Y - E(Y)]) = E(XY) - E(X)E(Y).$$

- Si $\sigma(X)$ et $\sigma(Y)$ sont non nuls, on appelle **coefficient de corrélation** de X et Y le réel

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Démonstration de l'existence de $\text{Cov}(X, Y)$, et de la seconde formule

On a $[X - E(X)][Y - E(Y)] = XY - E(X)Y - E(Y)X + E(X)E(Y)$. Les variables aléatoires X et Y ont une variance, donc le produit XY est d'espérance finie et par combinaison linéaire, $[X - E(X)][Y - E(Y)]$ est d'espérance finie. Par linéarité de l'espérance, on a

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) = E(XY) - E(X)E(Y). \quad \square$$

Remarques

- Si X admet une variance, $\text{Cov}(X, X) = V(X)$.
- Si X et Y admettent une variance, $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

Propriété

Soient X et Y deux variables aléatoires indépendantes sur (Ω, \mathcal{A}, P) , admettant une variance.

Alors $\text{Cov}(X, Y) = 0$.

Démonstration – On a $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$ par indépendance. \square

Remarque – La réciproque de la propriété précédente est fautive comme le montre l'exemple suivant : soit X une variable aléatoire d'image $\{-1, 0, 1\}$, de loi uniforme, et soit $Y = X^2$. Alors $E(XY) = E(X) = 0$ (on a $XY = X^3 = X$) donc $\text{Cov}(X, Y) = 0$, mais X et Y ne sont pas indépendantes car

$$P(Y = 0 | X = 1) = 0 \neq \frac{1}{3} = P(Y = 0).$$

Exemple – Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes suivant la loi $\mathcal{B}(p)$ avec $p \in]0, 1[$. Posons, pour tout $n \in \mathbb{N}^*$, $Y_n = X_n X_{n+1}$. Pour tout n , X_n est la fonction indicatrice de l'événement $(X_n = 1)$, et Y_n est la fonction indicatrice de l'événement $(X_n = 1) \cap (X_{n+1} = 1)$, de probabilité $p^2 \in]0, 1[$ par indépendance. En particulier, $Y_n \hookrightarrow \mathcal{B}(p^2)$. La variable Y_n indique deux succès consécutifs aux rangs n et $n + 1$.

De la même façon, pour tout $n \in \mathbb{N}^*$, $Y_n Y_{n+1} = X_n X_{n+1} X_{n+2} \hookrightarrow \mathcal{B}(p^3)$, donc

$$\text{Cov}(Y_n, Y_{n+1}) = E(Y_n Y_{n+1}) - E(Y_n)E(Y_{n+1}) = p^3 - p^4 = p^3(1 - p).$$

Notamment, Y_n et Y_{n+1} ne sont pas indépendantes.

En revanche, si $j \geq i + 2$, on remarque que $Y_i Y_j$ est la fonction indicatrice de

$$(Y_i Y_j = 1) = (X_i = 1) \cap (X_{i+1} = 1) \cap (X_j = 1) \cap (X_{j+1} = 1),$$

de probabilité p^4 par indépendance, et donc $E(Y_i Y_j) = p^4$, puis

$$\text{Cov}(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i)E(Y_j) = p^4 - p^2 p^2 = 0.$$

Attention, on ne peut pas en déduire que Y_i et Y_j sont indépendantes (c'est vrai, mais il faudrait le prouver en revenant par exemple à la définition).

Propriété

Soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) , admettant une variance.

Alors

$$|\text{Cov}(X, Y)| \leq \sigma(X) \sigma(Y),$$

En particulier, si $\sigma(X) \neq 0$ et $\sigma(Y) \neq 0$,

$$\rho(X, Y) \in [-1, 1].$$

Démonstration – D'après l'inégalité de Cauchy-Schwarz,

$$|\text{Cov}(X, Y)| = |E([X - E(X)][Y - E(Y)])| \leq (E((X - E(X))^2)E((Y - E(Y))^2))^{1/2} = \sigma(X)\sigma(Y).$$

L'encadrement de $\rho(X, Y)$ s'ensuit directement. \square

Remarque – Le coefficient de corrélation mesure en quelque sorte la dépendance entre X et Y . Lorsque $|\rho(X, Y)|$ est proche de 1, une information sur X apporte une information sur Y . Lorsque X et Y sont indépendantes, $\rho(X, Y) = 0$, mais la réciproque est fautive.

Propriété – Variance d'une somme de variables aléatoires

Soient X_1, \dots, X_n des variables aléatoires sur (Ω, \mathcal{A}, P) , admettant une variance.

Alors :

- $\sum_{k=1}^n X_k$ admet une variance et

$$V\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n V(X_k) + 2 \sum_{i < j} \text{Cov}(X_i, X_j).$$

- Si de plus X_1, \dots, X_n sont deux à deux indépendantes, on a

$$V\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n V(X_k).$$

Démonstration

- On a

$$\left(\sum_{k=1}^n X_k\right)^2 = \sum_{k=1}^n X_k^2 + 2 \sum_{i < j} X_i X_j.$$

Les X_k ont toutes une variance, donc les $X_i X_j$ sont d'espérance finie, et par combinaison linéaire $(\sum_{k=1}^n X_k)^2$ est d'espérance finie (i.e., $\sum_{k=1}^n X_k$ admet une variance). De plus, par linéarité de l'espérance,

$$E\left(\left(\sum_{k=1}^n X_k\right)^2\right) = \sum_{k=1}^n E(X_k^2) + 2 \sum_{i < j} E(X_i X_j).$$

D'autre part,

$$\left(E\left(\sum_{k=1}^n X_k\right)\right)^2 = \left(\sum_{k=1}^n E(X_k)\right)^2 = \sum_{k=1}^n (E(X_k))^2 + 2 \sum_{i < j} E(X_i)E(X_j).$$

On en déduit le résultat par différence.

- Si les X_k sont deux à deux indépendantes, on a, pour tout $(i, j) \in \llbracket 1, n \rrbracket^2$ tel que $i < j$, $\text{Cov}(X_i, X_j) = 0$, d'où l'égalité souhaitée. \square

Application – Soient X_1, \dots, X_n des variables aléatoires mutuellement indépendantes suivant la même loi $\mathcal{B}(p)$ et soit $S = X_1 + \dots + X_n$. D'après la propriété précédente, S a une variance et

$$V(S) = \sum_{k=1}^n V(X_k) = np(1-p).$$

On sait aussi que S suit la loi $\mathcal{B}(n, p)$. La variance ne dépendant que de la loi, on en déduit que pour toute variable aléatoire X qui suit la loi $\mathcal{B}(n, p)$, on a $V(X) = np(1-p)$. On retrouve donc la valeur de $V(X)$ déterminée plus tôt par un calcul direct.

3. Estimations de la dispersion

La variance s'interprète comme indicateur de dispersion. Dans ce paragraphe, nous allons montrer plus précisément comment la variance (ou l'écart-type) permet de mesurer cette dispersion.

Théorème – Inégalité de Markov

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) , positive, d'espérance finie.

Alors, pour tout $\varepsilon > 0$,

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

Démonstration – Soit $\varepsilon > 0$ fixé. On décrit $X(\Omega)$ en extension sous la forme $\{x_n; n \in I\}$. Soit $U = [\varepsilon, +\infty[$. Par positivité de X ,

$$E(X) \geq \sum_{x_n \in U} x_n P(X = x_n) \geq \varepsilon \sum_{x_n \in U} P(X = x_n)$$

car $x_n \geq \varepsilon$ si $x_n \in U$. Alors

$$E(X) \geq \varepsilon P(X \in U) = \varepsilon P(X \geq \varepsilon),$$

d'où le résultat. □

Théorème – Inégalité de Bienaymé - Tchebychev

Soit X une variable aléatoire sur (Ω, \mathcal{A}, P) , admettant une variance.

Alors, pour tout $\varepsilon > 0$,

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\sigma(X)^2}{\varepsilon^2}.$$

Démonstration – Soit $\varepsilon > 0$ fixé. La variable aléatoire X admet une variance donc est d'espérance finie et, en posant $Y = (X - E(X))^2$, alors Y est une variable aléatoire positive d'espérance finie. De plus, on remarque que

$$(|X - E(X)| \geq \varepsilon) = (Y \geq \varepsilon^2).$$

Alors, d'après l'inégalité de Markov,

$$P(|X - E(X)| \geq \varepsilon) = P(Y \geq \varepsilon^2) \leq \frac{E(Y)}{\varepsilon^2} = \frac{\sigma(X)^2}{\varepsilon^2}. \quad \square$$

Remarque – L'inégalité de Bienaymé - Tchebychev permet de majorer la probabilité que X s'écarte d'au moins ε de son espérance, *i.e.*, de sa moyenne. On voit que cette majoration fait intervenir l'écart-type de X ; plus précisément, plus $\sigma(X)$ est petit, plus la probabilité précédente est faible, c'est-à-dire, plus grande est la probabilité que X soit proche de son espérance. Cela confirme l'interprétation de $\sigma(X)$ et $V(X)$ comme indicateurs de dispersion.

Exemple – Notons $m = E(X)$ et $\sigma = \sigma(X)$. Pour $\varepsilon = 2\sigma$, on obtient

$$P(|X - m| \geq 2\sigma) \leq \frac{1}{4},$$

ou de façon équivalente,

$$P(m - 2\sigma < X < m + 2\sigma) \geq \frac{3}{4}.$$

La probabilité que X soit au plus à 2 écarts-types de son espérance est donc au moins 3/4. En revanche, pour $\varepsilon = \sigma$, l'inégalité ne donnerait pas de résultat intéressant.

Théorème – Loi faible des grands nombres

Soit $(X_n)_{n \in \mathbb{N}^*}$ une famille de variables aléatoires sur (Ω, \mathcal{A}, P) . On suppose que les variables aléatoires X_n

- sont deux à deux indépendantes,
- ont la même loi et admettent une variance.

On note $m = E(X_1)$, $\sigma = \sigma(X_1)$ et pour tout $n \in \mathbb{N}^*$, $S_n = X_1 + \dots + X_n$.

Alors, pour tout $\varepsilon > 0$,

$$P\left(\left|\frac{1}{n}S_n - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2},$$

et en particulier,

$$P\left(\left|\frac{1}{n}S_n - m\right| \geq \varepsilon\right) \xrightarrow{n \rightarrow +\infty} 0.$$

Démonstration – Les variables aléatoires X_n admettent une variance donc également une espérance. Sachant qu'elles ont la même loi, elles ont la même espérance et la même variance (par exemple celles de X_1 , m et σ^2). De plus, par linéarité de l'espérance, on a pour tout $n \in \mathbb{N}^*$,

$$E\left(\frac{1}{n}S_n\right) = \frac{1}{n}nE(X_1) = m,$$

et d'après les propriétés de la variance,

$$V\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}V(S_n) = \frac{1}{n}V(X_1)$$

par indépendance deux à deux des X_k . Ainsi, $\sigma\left(\frac{S_n}{n}\right)^2 = \frac{\sigma^2}{n}$.

Soit $\varepsilon > 0$ fixé. D'après l'inégalité de Bienaymé-Tchebychev appliquée à S_n/n , on a

$$P\left(\left|\frac{1}{n}S_n - m\right| \geq \varepsilon\right) \leq \frac{\sigma(S_n/n)^2}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0. \quad \square$$

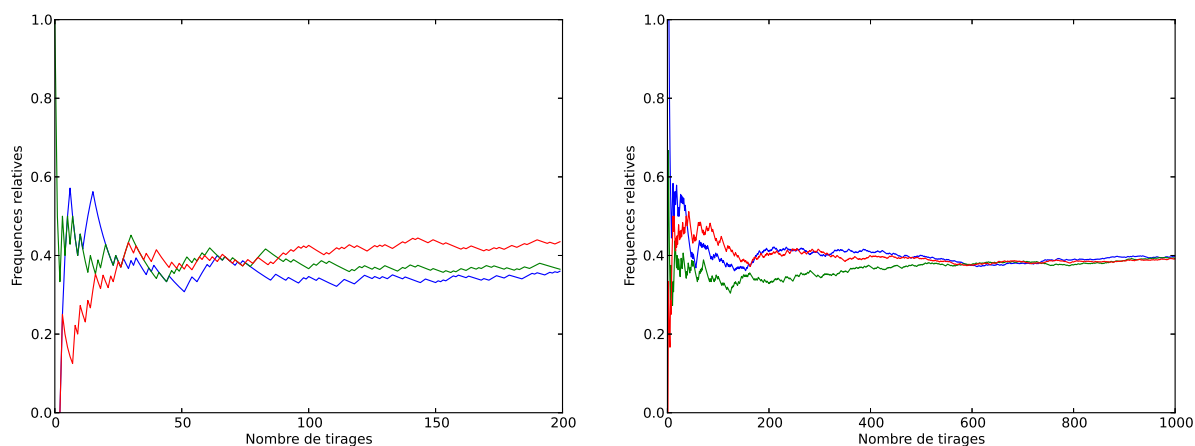
Remarques

• Imaginons que l'on répète indéfiniment une même expérience aléatoire en observant, à chaque étape, un certain résultat; cette situation est modélisée par une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires mutuellement indépendantes et de même loi, X_n représentant le résultat observé à la n -ième étape. Alors S_n/n représente la moyenne *empirique* des résultats au cours des n premières expériences.

Notons m l'espérance commune à toutes les variables X_n . La loi faible des grands nombres affirme que pour tout $\varepsilon > 0$, la probabilité que S_n/n s'écarte de m d'au moins ε tend vers 0 lorsque le nombre d'expériences tend vers $+\infty$. De façon équivalente, la probabilité que cette moyenne vérifie $m - \varepsilon < S_n/n < m + \varepsilon$ tend vers 1.

• Par exemple, considérons un jeu de pile ou face infini (ou toute autre expérience de Bernoulli reproduite indéfiniment) et notons X_n l'indicatrice de l'événement « le n -ième lancer donne pile ». Pour tout $n \in \mathbb{N}^*$, $X_n \hookrightarrow \mathcal{B}(p)$, $E(X_n) = p$ et $V(X_n) = p(1-p)$. Si les X_n sont deux à deux indépendantes, le théorème précédent affirme que la moyenne S_n/n du nombre de « pile » au cours des n premiers lancers sera « proche » de p (à ε près) avec une probabilité tendant vers 1 lorsque $n \rightarrow +\infty$. En un certain sens, la moyenne se stabilise vers p lorsque le nombre de lancers augmente.

Ci-dessous, on a représenté les fréquences relatives d'apparition de « pile » au cours des n premiers lancers, pour $n \in [1, 200]$ puis pour $n \in [1, 1000]$. Dans chaque cas, on a effectué trois simulations (courbes des différentes couleurs).



Il faut bien comprendre que ce théorème ne dicte pas à une expérience « concrète » comment elle va se dérouler pour « assurer » l'équilibre. Le théorème s'inscrit à *l'intérieur du modèle*, mais est cohérent avec l'approche intuitive des probabilités comme fréquence relative de réalisation lors d'un grand nombre de répétitions.

- Ce théorème peut jouer un rôle dans la validation du modèle : si on suppose une pièce équilibrée et que toutes les observations montrent une convergence vers $p \neq 1/2$, alors le modèle est sans doute à revoir. Il permet d'estimer certains paramètres (par observation d'un échantillon, comme par exemple lors d'un sondage), l'inégalité du théorème permettant de mesurer le risque d'erreur. Ces deux remarques relèvent de la théorie des Statistiques.

- Le théorème précédent n'affirme pas que $S_n(\omega)/n$ tend vers m pour toute issue ω (ce qui est faux en général) ; il ne faudrait donc pas s'étonner d'une issue ω pour laquelle $(S_n(\omega)/n)_{n \in \mathbb{N}^*}$ ne converge pas vers m , ou même, ne converge pas : dans le jeu de pile ou face infini avec une pièce équilibrée, il est possible d'obtenir pile à chaque tirage (même si l'événement associé est de probabilité nulle), et pour cette issue ω de l'expérience, $(S_n(\omega)/n)$ est constante égale à 1.

Exemple – On fait un test de qualité dans une production de N articles. Soit p la proportion d'articles défectueux. On vérifie n articles pris au hasard dans le stock, ce que l'on modélise par une famille (X_1, \dots, X_n) de variables aléatoires de Bernoulli mutuellement indépendantes de paramètre p (X_k prend la valeur 1 si le k -ième article testé est défectueux). Avec les notations précédentes, S_n/n est la proportion d'articles défectueux dans l'échantillon testé. On sait que pour tout $\varepsilon > 0$,

$$P\left(\left|\frac{1}{n}S_n - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2},$$

la dernière inégalité provenant de l'étude de la fonction trinôme $p \mapsto p(1-p)$. Choisissons par exemple $\varepsilon = 10^{-2}$; alors le majorant vaut $2500/n$. Ainsi, en testant n pièces, on peut affirmer avec un risque d'erreur d'au plus $2500/n$, que la proportion observée est une valeur approchée de p à 10^{-2} près. On voit que, avec la précision voulue, minimiser le risque d'erreur implique de tester un nombre assez grand d'articles : la convergence du majorant n'est pas très rapide.

Le tableau suivant récapitule certaines caractéristiques des lois usuelles :

Nom	Notation	Condition	Image	$P(X = k)$	$E(X)$	$V(X)$	$G_X(t)$
Bernoulli	$\mathcal{B}(p)$	$p \in [0,1]$	$\{0,1\}$	$P(X = 1) = p$	p	$p(1 - p)$	$1 - p + pt$
Binomiale	$\mathcal{B}(n,p)$	$n \in \mathbb{N}^*, p \in [0,1]$	$\llbracket 0, n \rrbracket$	$\binom{n}{k} p^k (1 - p)^{n-k}$	np	$np(1 - p)$	$(1 - p + pt)^n$
Géométrique	$\mathcal{G}(p)$	$p \in]0,1[$	\mathbb{N}^*	$p(1 - p)^{k-1}$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$	$\frac{pt}{1 - (1 - p)t}$
Poisson	$\mathcal{P}(\lambda)$	$\lambda > 0$	\mathbb{N}	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ	$e^{\lambda(t-1)}$