PROBABILITÉS ET VARIABLES ALÉATOIRES

Les probabilités finies

Définitions et premières propriétés

• Une expérience aléatoire est définie intuitivement comme une expérience dont le résultat ne dépend que du hasard, par exemple « lancer un dé » ou « tirer une pièce à pile ou face ».
L'ensemble des résultats possibles est appelé univers, souvent noté Ω.
Tout élément $\omega \in \Omega$ de l'univers Ω correspond à un résultat possible, ou une issue possible, de cette expérience aléatoire.
Un événement A est un sous-ensemble de l'univers $A \subset \Omega$.
Un événement élémentaire est un singleton $\{\omega\} \subset \Omega$.

Dans la suite de cette partie l'univers Ω est un ensemble fini, et les événements de Ω sont notés A, B, C, etc.

• Une probabilité sur un ensemble fini $\Omega \neq \emptyset$ est une application P des sous-ensembles de Ω à valeurs positives $P : \mathcal{P}(\Omega) \to [0,\, 1]$ vérifiant ces deux conditions :

$$P(\Omega) = 1$$

$$\text{ET} \; \forall\, (A,B) \in \mathcal{P}(\Omega)^2 \quad A \cap B = \emptyset \Longrightarrow P(A \cup B) = P(A) + P(B)$$

Cette égalité est appelée σ-additivité.

★ Le lancer d'un dé est une expérience aléatoire associée à l'univers $\Omega = [\![1, 6]\!]$ des tirages possibles de ce dé.
La probabilité d'un dé est, sauf exception, l'application P définie par $P(A) = \dfrac{\text{card } A}{6}$. Cette application P est bien une probabilité du fait que $P(\Omega) = 1$ et que l'application card de sous-ensembles finis et disjoints est σ-additive : $A \cap B = \emptyset \Longrightarrow \text{card}\,(A \cup B) = \text{card } A + \text{card } B$.
L'événement « tirer un nombre impair » est le sous-ensemble $I = \{1, 3, 5\}$ de $\Omega = [\![1, 6]\!]$. La probabilité de tirer un nombre impair est donc $P(I) = 3/6 = 1/2$.
Cette application probabilité P change si le dé est pipé : toutes les valeurs élémentaires n'ont alors pas la même probabilité.

• L'événement contraire $\overline{A}$ de $A \subset \Omega$, est le sous-ensemble complémentaire de A noté $\overline{A} = \Omega \setminus A = \big\{x \in \Omega \;\big/\; x \notin A\big\}$.
L'événement $\emptyset \subset \Omega$ est dit impossible ou irréalisable.
L'événement Ω est dit certain.
L'événement « A et B » correspond à $A \cap B$.
De même « A ou B » est l'événement $A \cup B$.
Deux événements A et B sont par définition incompatibles lorsque ces sous-ensembles sont disjoints : $A \cap B = \emptyset$.
« L'événement A entraîne l'événement B » signifie l'inclusion $A \subset B$.

$\ast$ La probabilité d'événements incompatibles est additive.
L'événement contraire de A vérifie $A \cap \overline{A} = \emptyset$ et $A \cup \overline{A} = \Omega$.

■ La probabilité de l'ensemble vide est $P(\emptyset) = 0$.
La probabilité de l'événement contraire vérifie $P(A) + P(\overline{A}) = 1$.
L'application probabilité est croissante par rapport à l'inclusion :

$$\forall\, (A,B) \in \mathcal{P}(\Omega)^2 \quad A \subset B \subset \Omega \Longrightarrow P(A) \leq P(B)$$

◊ La σ-additivité appliquée aux égalités $\emptyset \cap \emptyset = \emptyset$ et $\emptyset \cup \emptyset = \emptyset$ énonce donc $P(\emptyset) + P(\emptyset) = P(\emptyset) \in [0,\, 1]$. Ainsi $P(\emptyset) = 0$.

◊ L'événement contraire $\overline{A} = \Omega \setminus A$ vérifie $A \cap \overline{A} = \emptyset$ et $A \cup \overline{A} = \Omega$. Ainsi $1 = P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A})$.

◊ Supposons $A \subset B \subset \Omega$, et notons $C = B \setminus A$. Ainsi, $A \cup C = B$ et $A \cap C = \emptyset$. Donc $P(B) = P(A \cup C) = P(A) + P(C) \geq P(A)$ car $P(C) \geq 0$.

$\ast$ L'ensemble vide n'est pas un univers.
Par l'absurde, si $\Omega = \emptyset$ alors $1 = P(\Omega) = P(\emptyset) = 0$.

∘ Un événement A est dit presque impossible lorsque sa probabilité est nul, $P(A) = 0$, cela ne signifie pas nécessairement $A = \emptyset$.
Un événement B est dit presque certain lorsque sa probabilité est unitaire, $P(B) = 1$, cela ne signifie pas nécessairement $B = \Omega$. Dans ce cas l'événement contraire $\overline{B}$ est presque impossible, $P(\overline{B}) = 0$.
L'événement « tomber sur la tranche » T est presque impossible dans l'univers $\Omega = \{T, P, F\}$ d'un lancer de pièce à pile ou face :

$$\Omega \neq \{P, F\} \qquad P(\{T\}) = P(\emptyset) = 0 \qquad P(\{P, F\}) = P(\Omega) = 1$$

Réunion et intersection d'événements

■ Deux événements quelconques A et B de Ω vérifient cette égalité :

$$\forall (A,B) \in \mathcal{P}(\Omega)^2 \quad P(A) + P(B) = P(A \cup B) + P(A \cap B)$$

◊ En notant $C = A \cap B$ et $A' = A \setminus C$, ces sous-ensembles vérifient $A' \cap C = \emptyset$, $A' \cup C = A$, $A' \cap B = \emptyset$ et $A' \cup B = A \cup B$. La σ-additivité des probabilités énoncent ces égalités :

$$P(A) = P(A') + P(A \cap B) \qquad P(B) = P(B') + P(A \cap B)$$
$$\begin{aligned} P(A) + P(B) &= (P(A') + P(A \cap B)) + P(B) \\ &= (P(A') + P(B)) + P(A \cap B) \\ &= P(A' \cup B) + P(A \cap B) = P(A \cup B) + P(A \cap B) \end{aligned}$$

∗ Trois événements A, B et C sont disjoints deux à deux si et seulement si $A \cap B = A \cap C = B \cap C = \emptyset$.
Ils vérifient alors $P(A \cup B \cup C) = P(A) + P(B) + P(C)$.

∗ Si trois événements A, B et C disjoints deux à deux alors $A \cup B$ et C sont disjoints, etc. La démonstration repose sur la distributivité de l'intersection $\cap$ par rapport à la réunion $\cup$:

$$\begin{aligned} &A \cap B = A \cap C = B \cap C = \emptyset \\ \Longrightarrow\ &(A \cup B) \cap C = (A \cap C) \cup (B \cap C) = \emptyset \end{aligned}$$

• Plus généralement cette condition caractérise les familles d'événements $(A_k)_{k=1}^n$ disjoints deux à deux :

$$\forall (i,j) \in [\![1,n]\!] \quad i \neq j \Longrightarrow A_i \cap A_j = \emptyset$$

∗ Toutes les familles d'événements $(A_k)_{k=1}^n$ disjoints deux à deux vérifient l'égalité $P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$.
La démonstration s'effectue par récurrence à partir de la σ-additivité.

∗ Trois événements quelconques vérifient la formule suivante :

$$\begin{aligned} P(A \cup B \cup C) = \quad & P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & + P(A \cap B \cap C) \end{aligned}$$

◊ La démonstration exploite la distributivité de l'intersection $\cap$ sur la réunion $\cup$: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$. Elle applique plusieurs fois la formule précédente sur la probabilité d'une réunion de deux

ensembles quelconque :

$$\begin{aligned} P(X \cup Y) &= P(X) + P(Y) - P(X \cap Y) \\ P((A \cup B) \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A \cup B) + P(C) - P((A \cap C) \cup (B \cap C)) \\ &= (P(A) + P(B) - P(A \cap B)) + P(C) \\ &\quad - (P(A \cap C) + (B \cap C) - P(A \cap C \cap B \cap C)) \\ &= P(A) + P(B) - P(A \cap B) + P(C) \\ &\quad - P(A \cap C) - (B \cap C) + P(A \cap B \cap C) \end{aligned}$$

★ Tout événement $A \subset \Omega$ de l'univers fini Ω est réunion disjointe et finie d'événements élémentaires :

$$A = \bigcup_{\omega \in A} \{\omega\} \qquad P(A) = \sum_{\omega \in A} P(\{\omega\})$$

Il suffit de connaître la probabilité de chaque événement élémentaire ω de Ω pour connaître l'application probabilité sur $\mathcal{P}(\Omega)$.

• Un système complet d'événements $(A_k)_{k=1}^n$ de l'univers Ω est une famille d'événements disjoints deux à deux dont la réunion est l'univers complet :

$$\bigcup_{k=1}^{n} A_k = \Omega \qquad \text{ET} \qquad \forall (i,j) \in [\![1,n]\!]^2 \quad i \neq j \Longrightarrow A_i \cap A_j = \emptyset$$

∗ Par exemple $(A, \overline{A})$ forment un système complet d'événements.

► Déterminer une condition nécessaire et suffisante sur $(x,y) \in \mathbb{R}^2$ de façon à ce que l'univers $\Omega = \{a,b,c\}$ possède une probabilité P vérifiant $P(\{a,b\}) = x$ et $P(\{a,c\}) = y$.

▷ Si cette application probabilité existe alors elle vérifie ces égalités :

$$\begin{aligned} P(\{a,b\}) + P(\{a,c\}) &= P(\{a,b,c\}) + P(\{a\}) \\ &= P(\Omega) + P(\{a\}) = 1 + P(\{a\}) \\ P(\{a\}) &= x + y - 1 \geq 0 \\ P(\{b\}) &= P(\{a,b\}) - P(\{a\}) \\ &= x - (x + y - 1) = 1 - y \geq 0 \\ P(\{c\}) &= y - (x + y - 1) = 1 - x \geq 0 \end{aligned}$$

Réciproquement les trois conditions affines $x + y \geq 1$, $0 \leq x$ et $0 \leq y$ suffisent pour vérifier toutes les conditions imposées par une probabilité : la probabilité de chaque sous-ensemble est positive, additive sur les sous-ensembles disjoints, et la probabilité de l'univers est $P(\{a,b,c\}) = 1$.

Probabilité uniforme

• Une probabilité est uniforme lorsque chacun des événements élémentaires ont la même probabilité :

$$P(A) = \frac{\text{card } A}{\text{card } \Omega} \qquad P(\{\omega\}) = \frac{1}{\text{card } \Omega} \quad \text{pour tout } \omega \in \Omega$$

∗ La probabilité uniforme est bien une probabilité.
Un jeu de dé a habituellement une probabilité uniforme.
Les probabilités uniformes sur un univers fini se ramènent le plus souvent à des questions de dénombrement.

► Une urne $\mathcal{U}$ contient 2 boules rouges, 4 boules vertes et 5 boules bleues. Les probabilités de tirage de chaque boule sont identiques.
Déterminer la probabilité p_1 de tirer sans remise un sous-ensemble de trois boules d'une même couleur, celle p_2 de faire un tirage bicolore, et celle p_3 d'un tirage tricolore.
Déterminer de même les trois probabilités p'_1, p'_2 et p'_3 lorsque ce tirage est effectué avec remise.
Calculer enfin les probabilités q et q' de faire un tirage dans l'ordre, d'une boule rouge, puis verte puis bleue, lors du tirage successif des boules une à une, dans les deux cas sans remise et avec remise.

▷ Un tirage sans remise de trois boules consiste en un sous-ensemble de 3 éléments parmi l'ensemble $\mathcal{U}$ des 11 boules de l'urne. Il y a $\binom{11}{3} = 165$ possibiltés.
Le nombre de sous-ensembles de 3 boules vertes parmi les 4 possibles est $\binom{4}{3} = 4$, et celui de 3 boules bleues parmi 5 est $\binom{5}{3} = 10$. Le tirage de 3 boules rouges est impossible.
La probabilité d'un tirage unicolore est $p_1 = 14/165$.
Le tirage est bicolore consiste en 2 boules d'une couleur et 1 d'une autre couleur. Six associations de couleurs sont possibles et disjoints les uns des autres. Chaque cas est un produit de possibilités :

$$\begin{array}{llll}
(2R,1V) & \binom{2}{2} \times \binom{4}{1} = 4 & (1R,2V) & \binom{2}{1} \times \binom{4}{2} = 12 \\
(2R,1B) & \binom{2}{2} \times \binom{5}{1} = 5 & (1B,2R) & \binom{2}{1} \times \binom{5}{2} = 20 \\
(2V,1B) & \binom{4}{2} \times \binom{5}{1} = 30 & (1B,2V) & \binom{4}{1} \times \binom{5}{2} = 40
\end{array}$$

La probabilité p_2 est $p_2 = (4+12+5+20+30+40)/165 = 37/55$.
Les sous-ensembles de 3 boules de couleurs différentes sont obtenus en prenant une boule de chaque couleur : Le nombre de ces sous-ensembles donne la probabilité p_3 :

$$\binom{2}{1} \times \binom{4}{1} \times \binom{5}{1} = 40 \qquad p_3 = 40/165 = 8/33$$

$$p_1 + p_2 + p_3 = 114 + 111 + 40165 = 1$$

Cette dernière égalité effectue une vérification ou permet de déterminer l'une des trois probabilités à partir des deux autres.

▷ Un tirage avec remise correspond aux éléments du produit cartésien $\mathcal{U}^3$ de l'ensemble $\mathcal{U}$ de toutes les boules, qui permet 11^3 tirages possibles.
Le tirage est unicolore lorsque les boules sont toutes les trois rouges, avec 2^3 possibilités, toutes les trois vertes dans 4^3 cas ou toutes les trois bleues dans 5^3 cas.
La probabilité du tirage unicolore est $p'_1 = \dfrac{8+64+125}{1331} = \dfrac{197}{1331}$.
Le tirage bicolore de la couleur c répétée deux fois et de la couleur γ est de l'une des trois formes (γ,c,c), (c,γ,c) ou (c,c,γ) définie par la position de la couleur γ. Le nombre de possibilités pour ces trois sous-cas est le même.
Ce tableau récapitule le nombre de possibilités des six groupes de couleurs possibles, dans cet ordre :

$$\begin{array}{llll}
(R,R,V) & 2^2 \times 4 = 16 & (R,V,V) & 2 \times 4^2 = 32 \\
(R,R,B) & 2^2 \times 5 = 20 & (R,B,B) & 2 \times 5^2 = 50 \\
(V,V,B) & 4^2 \times 5 = 80 & (V,B,B) & 4 \times 5^2 = 100
\end{array}$$

Le nombre de triplets bicolore est à multiplier par trois pour la raison précédente d'ordre d'énumération, d'où la probabilité p'_2 :

$$p'_2 = \frac{3 \times (16+32+20+50+80+100)}{1331} = \frac{893}{1331}$$

Enfin le nombre de possibilités d'énumérer les 3 couleurs dans n'importe quel ordre est $3! = 6$, et le nombre de tirages dans l'ordre (R, V, B) est $2 \times 4 \times 5$, d'où la probabilité p'_3 :

$$p'_3 = 6 \times \frac{2 \times 4 \times 5}{1331} = \frac{240}{1331} \qquad p'_1 + p'_2 + p'_3 = 1$$

▷ Les probabilités q et q' sont obtenues par produit à partir de chaque boule :

$$q = \frac{2 \times 4 \times 5}{11 \times 10 \times 9} = \frac{400}{990} = \frac{4}{99} \qquad q' = \frac{2 \times 4 \times 5}{11 \times 11 \times 11} = \frac{40}{1331}$$

Remarquons que ces probabilités sont 6 fois plus faibles que p_3 et p'_3 à cause de l'ordre imposé des couleurs, parmi les 6 possibles.

Probabilité conditionnelle

Dans cette partie C est un événement de Ω vérifiant $P(C) > 0$.

• La « probabilité conditionnelle de A sachant C » est définie ainsi :

$$P(A \mid C) = P_C(A) = \frac{P(A \cap C)}{P(C)}$$

▪ L'application $P_C(\bullet)$ est une probabilité sur l'univers Ω.

◊ La probabilité de l'univers est $P_C(\Omega) = \dfrac{P(\Omega \cap C)}{P(C)} = \dfrac{P(C)}{P(C)} = 1$.

Si A et B sont des événements incompatibles de Ω, alors leurs probabilités conditionnelles s'additionnent par distributivité de l'intersection sur la réunion :

$$A \cap B = \emptyset \quad \text{donc} \quad (A \cap C) \cap (B \cap C) = A \cap B \cap C = \emptyset$$

$$\begin{aligned} P_C(A \cup B) &= \frac{P((A \cup B) \cap C)}{P(C)} = \frac{P((A \cap C) \cup (B \cap C))}{P(C)} \\ &= \frac{P(A \cap C) + P(B \cap B)}{P(C)} = \frac{P(A \cap C)}{P(C)} + \frac{P(B \cap C))}{P(C)} \\ &= P_C(A) + P_C(B) \end{aligned}$$

Formule de Bayes

★ Ces égalités sont des conséquences directes des probabilités conditionnelles :

$$P(A \cap B) = P(B)\, P_B(A) = P(A)\, P_A(B) \text{ si } P(A) > 0 \text{ et } P(B) > 0$$

$$P_A(B) = \frac{P_B(A)\, P(B)}{P(A)}$$

▶ Une machine A fabrique 90% de la production, et la machine B les 10% restants. Sur l'ensemble 3% des pièces sont défaillantes, et il y en a autant qui proviennent de la machine A que de la machine B. Quel est le taux de pièces défaillantes pour chacune de ces machines ?

▷ Notons $P(F)$ la probabilité d'une pièce défaillante, les événements A [respectivement B] énoncent que la pièce a été fabriquée par la machine A [respectivement B] :

$$P(F) = 0{,}03 \qquad P(A) = 0{,}9 \qquad P(B) = 0{,}1$$
$$P_F(A) = P_F(B) = 0{,}5$$

Ce calcul porte sur les probabilités conditionnelles $P_A(F)$ et $P_B(F)$:

$$P(A \cap F) = P_F(A)\, P(F) = P_A(F)\, P(A)$$

$$P_A(F) = \frac{P_F(A)\, P(F)}{P(A)} = \frac{0{,}5 \times 0{,}03}{0{,}9} \approx 1{,}67\%$$

$$P_B(F) = \frac{P_F(B)\, P(F)}{P(B)} = \frac{0{,}5 \times 0{,}03}{0{,}1} = 15\%$$

Formule des probabilités totales

★ Les probabilités conditionnelles sur un système complet d'événements permettent de retrouver la probabilité totale de A à partir des probabilités conditionnelles :

Par exemple lorsque le système de complet d'événements est $(C, \overline{C})$:

$$\begin{aligned} P(A) &= P(A \cap (C \cup \overline{C})) = P((A \cap C) \cup (A \cap \overline{C})) \\ &= P(A \cap C) + P(A \cap \overline{C}) = P(C)\, P_C(A) + P(\overline{C})\, P_{\overline{C}}(A) \end{aligned}$$

Plus généralement à partir d'un système complet $(C_k)_{k=1}^n$ d'événements, la probabilité de A est cette somme :

$$P(A) = \sum_{k=1}^{n} P(C_k)\, P_A(C_k)$$

◊ Cette démonstration est rédigée pour un système complet de 3 événements (C, C', C'') :

$$C \cup C' \cup C'' = \Omega \qquad C \cap C' = C \cap C'' = C' \cap C'' = \emptyset$$

$$
\begin{aligned}
P(A) &= P(A \cap (C \cup C' \cup C'')) \\
&= P((A \cap C) \cup (A \cap C') \cup (A \cap C'')) \\
&= P(A \cap C) + P(A \cap C') + P(A \cap CB'') \\
&= P(C)\, P_C(A) + P(C')\, P_{C'}(A) + P(C'')\, P_{C''}(A)
\end{aligned}
$$

▶ Un test sanguin détecte une pathologie présente chez une personne sur 10 000. Ce test est positif chez 99% des malades et faussement positif chez 0,1% des personnes non atteintes. Un individu passe ce test et obtient un résultat positif ; quelle est sa probabilité d'être malade ?

▷ L'univers Ω est l'ensemble de la population avec quatre sous-ensembles des personnes malades M ou non $\overline{M}$, détectée D ou non $\overline{D}$ associée à ces probabilités :

$$P(M) = 10^{-4} \qquad P_M(D) = 0{,}99 \qquad P_{\overline{M}}(D) = 10^{-3}$$

La probabilité recherchée est la probabilité conditionnelle $P_D(M)$ obtenue à partir de la probabilité de détection :

$$
\begin{aligned}
P(M \cap D) &= P(M)\, P_M(D) = 9{,}9\, 10^{-5} \\
P(D) &= P(M)\, P_M(D) + P(\overline{M})\, P_{\overline{M}}(D) \approx 1{,}099\, 10^{-3} \\
P_D(M) &= \frac{P(M \cap D)}{P(D)} = \frac{P(M)\, P_M(D)}{P(D)} \approx 9\%
\end{aligned}
$$

Formule des probabiltés composées

■ La formule des probabilités composées énonce ce produit :

$$
\begin{aligned}
&P(\bigcap_{i=1}^{n} A_i) \\
&= P(A_1) \times P_{A_1}(A_2) \times P_{A_1 \cap A_2}(A_3) \times \cdots \times P_{A_1 \cap A_2 \cap \cdots A_{n-1}}(A_n)
\end{aligned}
$$

▶ Un paquet contient 20 gourmandises emballées individuellement dans des petites boites en matière plastique. Le propriétaire du paquet neuf en offre une puis remet l'emballage vide dans le paquet. Il se sert ensuite une première fois et remet l'emballage vide, fait de même la deuxième et la troisième fois. Quelle est la probabilité qu'il mange en tout trois gourmandises ?

▷ Notons G_1, G_2 et G_3 les événements qui consistent à manger une gourmandise la première, deuxième ou troisième fois. Les probabilités

conditionnelles interviennent pour connaitre le nombre d'emballages vides dans le paquet :

$$P(M_1) = \frac{19}{20} \qquad P_{M_1}(M_2) = \frac{18}{20} \qquad P_{M_1 \cap M_2}(M_3) = \frac{17}{20}$$

$$
\begin{aligned}
P(M_1 \cap M_2) &= P(M_1) \times P_{M_1}(M_2) = \frac{19}{20}\,\frac{18}{20} = 85{,}5\% \\
P(M_1 \cap M_2 \cap M_3) &= P(M_1 \cap M_2) \times P_{M_1 \cap M_2}(M_3) \\
&= P(M_1) \times P_{M_1}(M_2) \times P_{M_1 \cap M_2}(M_3) \\
&= \frac{19}{20}\,\frac{18}{20}\,\frac{17}{20} \approx 72{,}7\%
\end{aligned}
$$

Événements indépendants

• Par définition deux événements A et B sont indépendants si et seulement si $P(A \cap B) = P(A)\, P(B)$.
Dit autrement $P_B(A) = P(A)$ dans le cas où $P(B) > 0$.

∗ Les événements contraires d'événements indépendants sont indépendants. Les trois couples $(A, \overline{B})$, $(\overline{A}, B)$ et $(\overline{A}, \overline{B})$ sont indépendants dès que les événements A et B le sont :

$$
\begin{aligned}
P(A) &= P(A \cap \overline{B}) + P(A \cap B) \\
P(A \cap \overline{B}) &= P(A) - P(A \cap B) = P(A) - P(A)\, P(B) \\
&= P(A)(1 - P(B)) = P(A)\, P(\overline{B})
\end{aligned}
$$

La formule est symétrique pour traiter le cas des événements $\overline{A}$ et B ; cette égalité appliquée successivement à $\overline{A}$ et B, puis $\overline{A}$ et $\overline{B}$ permet de conclure dans le dernier cas.

• Sur le même principe trois événements sont mutuellements indépendants si et seulement si ils vérifient toutes ces égalités :

$$
\begin{aligned}
P(A \cap B \cap C) &= P(A)\, P(B)\, P(C) \qquad & P(A \cap B) &= P(A)\, P(B) \\
P(B \cap C) &= P(B)\, P(C) \qquad & P(A \cap C) &= P(A)\, P(C)
\end{aligned}
$$

★ Trois événements deux à deux indépendants ne sont pas nécessairement indépendants dans leur ensemble.

▶ Prenons l'univers Ω d'un tirage à pile ou face de deux pièces ; les deux pièces sont régulières sans influence de l'une sur l'autre. La probabilité est uniforme, chacune de ces 4 événement élémentaire est de probabilité 1/4.

$$\Omega = \{(P,P),(P,F),(F,P),(F,F)\}$$
$$P(P,P) = P(P,F) = P(F,P) = P(F,F) = 1/4$$

Notons F' [respectivement F''] l'événement que la première [seconde pièce] pièce tombe sur face, et M l'événement que les deux pièces tombent sur la même face.
Montrer que ces trois événements F', F'' et M pris deux à deux sont indépendants, mais que ces trois événements ne sont pas mutuellement indépendants.

⊳ Les événements F' et F'' sont indépendants, de même F' et M, et F'' et M :

$$\begin{aligned}
&F' = \{(F,P),(F,F)\} && \text{de probabilité } 1/2\\
&F'' = \{(P,F),(F,F)\} && \text{de probabilité } 1/2\\
&M = \{(P,P),(F,F)\} && \text{de probabilité } 1/2\\
&F'\cap F'' = F'\cap M = F''\cap M = \{(F,F)\} && \\
& && \text{de probabilité } 1/4 = 1/2\times 1/2\\
&F'\cap F''\cap M = F'\cap F'' && \text{de probabilité } 1/4 \neq 1/8 = 1/2^3
\end{aligned}$$

∗ Si les événements A, B et C sont mutuellement indépendants alors $A\cap B$ et C sont indépendants, de même $A\cup B$ et C, etc.

$$\begin{aligned}
P((A\cap B)\cap C) &= P(A\cap B\cap C)\\
&= P(A)\,P(B)\,P(C) = P(A\cap B)\,P(C)\\
P((A\cup B)\cap C) &= P((A\cap B)\cap(A\cap C))\\
&= P(A\cap C) + P(B\cap C) - P(A\cap B\cap C)\\
&= P(A)\,P(C) + P(B)\,P(C) - P(A)\,P(B)\,P(C)\\
&= (P(A) + P(B) - P(A\cap B))\,P(C)\\
&= P(A\cup B)\,P(C)
\end{aligned}$$

• Plus généralement n événements $(A_k)_{k=1}^n$ sont mutuellement indépendants si et seulement si les probabilités pour toutes les intersections de ces événements sont les produits des probababilités de chaque événement.

$$\forall \mathcal{I} \subset [\![1,n]\!] \quad P(\bigcap_{i\in\mathcal{I}} A_i) = \prod_{i\in\mathcal{I}} P(A_i)$$

∗ Il ne suffit pas que cette égalité soit uniquement valable pour pour intersections de deux événements.

Par convention l'intersection sur un sous-ensemble vide d'indice $\mathcal{I} = \emptyset$ est l'événement Ω de probabilité 1 qui est aussi la valeur du produit vide.
La définition précédente énumère donc 2^n cas possibles pour $\mathcal{I}$, mais seulement $2^n - n - 1$ correspondent à plusieurs événements. n cas correspondent à l'événement A_k pour $\mathcal{I} = \{k\}$, et 1 à $\mathcal{I} = \emptyset$.

★ La probabilité d'une réunion d'événements mutuellements indépendants $(A_k)_{k=1}^n$ se calcule à partir de l'intersection des événements contraires :

$$\begin{aligned}
\overline{\bigcup_{k=1}^n A_k} &= \bigcap_{k=1}^n \overline{A_k}\\
P(\bigcap_{k=1}^n \overline{A_k}) &= \prod_{k=1}^n P(\overline{A_k}) = \prod_{k=1}^n (1 - P(A_k))\\
P(\bigcup_{k=1}^n A_k) &= 1 - P(\overline{\bigcap_{k=1}^n A_k} = 1 - \prod_{k=1}^n (1 - P(A_k))
\end{aligned}$$

Exemple d'une suite de probabilités

▶ Un étudiant arrive chaque jour à l'heure ou en retard en cours. L'événement « être en retard le jour n » est noté R_n. Il est à l'heure le premier jour, et la probabilité qu'il soit en retard un jour suivant dépend uniquement s'il a été à l'heure ou non la veille, et est indépendants des éventuels retards des autres jours précédents :

$$p_n = P(R_n) \qquad p_1 = P(R_1) = 0$$
$$P(R_{n+1} \mid R_n) = a/2 \qquad P(R_{n+1} \mid \overline{R_n}) = a \in [0,\,1]$$

Exprimer p_{n+1} en fonction de p_n, déterminer λ de façon à ce que la suite $q_n = p_n + \lambda$ soit géometrique, puis en déduire p_n en fonction de n uniquement, et préciser la limite de la probabilité quand n tend vers $+\infty$.
Encadrer la probabilité q_n pour que cet étudiant soit sanctionné à cause d'un retard deux jours successifs.
Quelle est la probabilité p' pour que cet étudiant arrive à l'heure les N jours de l'année scolaire, ou au contraire la probabilité p'' qu'il soit en retard au moins une fois.
Préciser les applications numériques pour un taux $a = 2\%$.

▷ Le suite $(p_n)_{n\in\mathbb{N}^*}$ est définie par récurrence à partir de probabilités conditionnelles pour $n\in\mathbb{N}^*$.

$$\begin{aligned} p_{n+1} &= P(R_{n+1}) = P(R_n)\,P_{R_n}(R_{n+1}) + P(\overline{R_n})\,P_{\overline{R_n}}(R_{n+1}) \\ &= \frac{a}{2}\,p_n + a(1-p_n) = a - \frac{a}{2}\,p_n \end{aligned}$$

La valeur λ recherchée simplifie les termes constants dans la relation de récurrence :

$$p_{n+1} = q_{n+1} - \lambda = a - \frac{aq_n}{2} + \frac{a\lambda}{2} \qquad -\lambda = a + \frac{a\lambda}{2} \qquad \lambda = \frac{-2a}{2+a}$$

Cette valeur de λ aboutit à la suite géométrique $q_{n+1} = -aq_n/2$, ainsi $q_n = q_1(-a/2)^{n-1}$ où $q_1 = \lambda$ enfin $p_n = q_n - \lambda$:

$$\begin{aligned} p_n &= q_n - \lambda = q_1(-a/2)^{n-1} - \lambda \\ &= \lambda((-a/2)^{n-1} - 1) = \frac{2a}{2+a}(1-(a/2)^{n-1}) \end{aligned}$$

$$\lim_{n\to\infty} p_n = \frac{2a}{2+a}$$

▷ La probabilité q_n d'un retard les deux jours consécutifs $n \geq 2$ et $n+1$ est obtenue par les probabilités conditionnelles :

$$\begin{aligned} q_n &= P(R_n \cap R_{n+1}) = P(R_n)\,P_{R_n}(R_{n+1}) = p_n\,a/2 \\ &= \frac{a^2}{2+a}(1-(a/2)^{n-1}) \in \left[\frac{a^2}{2(2+a)}, \frac{a^2}{2+a}\right] \qquad \text{si } n \geq 2, \text{ et } q_1 = 0 \end{aligned}$$

▷ L'événement étudié est $\overline{R_1}\cap\overline{R_2}\cap\overline{R_3}\cap\cdots\cap\overline{R_N}$ de probabilité p'. Chaque jour l'étudiant est à l'heure, et était à l'heure la veille. La probabilité d'être à l'heure un jour dépend uniquement du précédent, et est $1-a$ pour tous les jours à partir du deuxième. La probabilité sur l'année est donc $p' = (1-a)^{N-1}$ car $p_1 = 0$.
Le fait d'être en retard au moins un jour se décompose en événements disjoints, en retard dès le deuxième jour, ou bien à l'heure les deux premiers et en retard le troisième, etc. La probabilité de cette réunion d'événements est une somme qui aboutit à une série géométrique :

$$\begin{aligned} p'' &= a + (1-a)a + (1-a)^2 a + \cdots + (1-a)^{N-2} a \\ &= \sum_{k=0}^{N-2} a(1-a)^k = a\frac{1-(1-a)^{N-1}}{1-(1-a)} = 1-(1-a)^{N-1} \end{aligned}$$

$$p' + p'' = 1$$

Variables aléatoires réelles

Définition

• Une variable aléatoire X réelle est une application $X : \Omega \to \mathbb{R}$ de l'univers Ω dans $\mathbb{R}$.

• La loi de probabilité de la variable aléatoire X, appelée plus rapidement loi de X et est définie par $P(X=a)$ où $a\in\mathbb{R}$:

$$P(X=a) = P(\{\omega\in\Omega \;/\; X(\omega)=a\}) = P(X^{-1}(\{a\}))\in[0,\,1]$$

$*$ Ainsi $X = a$ est une abréviation de l'événement suivant :
$\{\omega\in\Omega \;/\; X(\omega) = a\} = X^{-1}(\{a\}) \qquad P(X=a) = P(X^{-1}(\{a\})) \subset \Omega$

Dans la suite les variables aléatoires sont finies, c'est-à-dire que l'ensemble $X(\Omega)$ est fini. Cela est le cas dès que l'univers Ω est fini, mais ce n'est pas nécessaire.

★ Ainsi $P(X=a) = 0$ dès que $a\in\mathbb{R}$ n'est pas l'une des valeurs de cet ensemble fini $X(\Omega)\subset\mathbb{R}$.

★ Plus généralement pour tout sous-ensemble $F\subset\mathbb{R}$ de $\mathbb{R}$ vérifie que $F\cap X(\Omega)$ est un sous-ensemble fini de $\mathbb{R}$:

$$\begin{aligned} P(X\in F) &= P(\{\omega\in\Omega \;/\; X(\omega)\in F\}) = P(X^{-1}(F)) \\ &= \sum_{a\in F\cap X(\Omega)} P(X=a)\in[0,\,1] \end{aligned}$$

★ Dans certains cas la loi de probabilité est confondue avec l'application probabilité. Prenons l'exemple d'un dé à six faces.
Notons d'une part l'univers $\Omega = \{F_1, F_2, F_3, F_4, F_5, F_6\}$ des six faces du dé, et la variable aléatoire $D : \Omega \to [\![1,6]\!]$ définie par la valeur $k\in[\![1,6]\!]$ de la face F_k, ainsi $D(F_k) = k$; dans ce cas la loi de probabilité de D est définie par $P_\Omega(D=k)\in[0,\,1]$ lorsque $k\in[\![1,6]\!]$.
Notons d'autre part l'univers $\mathcal{U} = \{1,2,3,4,5,6\} = [\![1,6]\!]$ dont l'application probabilité d'un dé est définie par la probabilité élémentaire $P_{\mathcal{U}}(\{k\})$.
Ces deux modélisations sont équivalentes : $P_\Omega(D=k) = P_{\mathcal{U}}(\{k\})$. L'une utilise une probabilité dans l'univers des faces F_k dont la valeur est donnée par la variable aléatoire D, et l'autre étudie une probabilité dont le paramètre est directement dans le sous-ensemble

des entiers $[\![1,6]\!]$.

$*$ La fonction de répartition $F_X : \mathbb{R} \to [0,\,1]$ de la variable aléatoire X est définie par $F_X(x) = P(X \leq x)$. Cette application est croissante et vérifie ces propriétés :

$$\lim_{x\to-\infty} F_X(x) = 0 \qquad \lim_{x\to+\infty} F_X(x) = 1$$

$$P(a < X \leq b) = F_X(b) - F_X(a) \quad \text{pour } (a,b)\in\mathbb{R}^2 \text{ et } a \leq b$$

$\bigstar$ Dans l'univers du lancer de deux dés indépendants aux faces $[\![1,6]\!]^2$ équiprobables la somme S des valeurs de ces deux dés est une variable aléatoire à valeurs dans $[\![2,12]\!]$.
Le tableau suivant récapitule la loi de cette variable aléatoire S en énumérant les 36 coups possibles :

$$\begin{array}{ll} P(S=2) = P(S=12) = 1/36 & P(S=3) = P(S=11) = 2/36 \\ P(S=4) = P(S=10) = 3/36 & P(S=5) = P(S=9) = 4/36 \\ P(S=6) = P(S=8) = 5/36 & P(S=7) = 6/36 \end{array}$$

$*$ La loi d'une variable aléatoire X constante de valeur α est définie ainsi :

$$X : \Omega \to \mathbb{R} \qquad X(\omega) = \alpha \quad \text{pour tout } \omega\in\Omega$$

$$P(X=\alpha) = 1 \qquad P(X=\lambda) = 0 \quad \text{pour tout } \lambda \neq \alpha$$

$*$ La loi uniforme d'une variable aléatoire ayant $n \geq 1$ valeurs $(v_k)_{k=1}^n$ différentes deux à deux est par définition celle-ci :

$$P(X=v_k) = \frac{1}{n} \quad \text{pour tout } k\in[\![1,n]\!]$$

$$P(X=u) = 0 \quad \text{si } u\notin \left\{v_k \;/\; k\in[\![1,n]\!]\right\}$$

Deux lois usuelles

Loi de Bernoulli

$\bullet$ Une variable aléatoire X de Bernoulli $\mathcal{B}(p)$ de paramètre $p\in[0,\,1]$ est définie par exactement deux valeurs 0 et 1 possibles :

$$P(X=1) = p \qquad P(X=0) = 1-p$$

$$P(X=a) = 0 \quad \text{pour tout } a \neq 0 \text{ et } a \neq 1$$

$*$ Cette loi modélise un univers à deux valeurs $\Omega = \{0,1\}$. La probabilité sur cet univers est bien $p + (1-p) = 1$.

$*$ Le tirage d'une pièce « à pile ou face » suit une loi de Bernoulli de valeur $p = 1/2$, en accordant par exemple de façon arbitraire la valeur 1 au côté face.

Loi binomiale

$\bullet$ Par définition une variable aléatoire X suit la loi binomiale $\mathcal{B}(p,n)$ de paramètre $p\in[0,\,1]$ et $n\in\mathbb{N}$ à cette condition :

$$P(X=k) = \binom{n}{k} p^k(1-p)^{n-k} \quad \text{pour tout } k\in[\![0,n]\!]$$

$$P(X=x) = 0 \quad \text{si } x\notin[\![0,n]\!]$$

$*$ Toutes les probabilités élémentaires $P(X=k)$ sont positives et de somme 1 par la formule du binome de Newton, donc cette loi binomiale est bien une loi de probabilité :

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

$*$ La loi binomiale pour $n=0$ est la loi constante de valeur 0, et la loi binomiale pou $n=1$ est la loi de Bernoulli de même paramètre p.

$\blacktriangleright\!\!\!>$ Quelle est en fonction de p et n la valeur maximale de la probabilité $k\in[\![0,n]\!]$ de la loi binomiale $\mathcal{B}(p,n)$?

$\vartriangleright\!\!\!>$ Notons $0 < p_k = P(X=k) = \binom{n}{k} p^k(1-p)^{n-k}$ pour $k\in[\![0,n]\!]$,
Le quotient p_{k+1}/p_k précise les variations de p_k :

$$\frac{p_{k+1}}{p_k} = \frac{\binom{n}{k+1}}{\binom{n}{k}} \frac{p}{1-p} = \frac{k!\,(n-k)!}{(k+1)!\,(n-k-1)!} \frac{p}{1-p} = \frac{(n-k)p}{(k+1)(1-p)}$$

$$\begin{aligned} 1 \leq \frac{p_{k+1}}{p_k} &\iff (k+1)(1-p) \leq (n-k)p \\ &\iff 1+k-kp-p \leq np-kp \\ &\iff 1+k \leq (n+1)p \iff 1+k \leq \lfloor (n+1)p \rfloor = k_0 \end{aligned}$$

Le quotient précédent démontre ces inégalités où $k_0 = \lfloor (n+1)p \rfloor$:

$$0 < p_0 \leq p_1 \leq \cdots \leq p_{k_0-1} \leq p_{k_0}$$
$$p_{k_0} \geq p_{k_0+1} \geq \cdots \geq p_{n-2} \geq p_{n-1} \geq p_n > 0$$

La probabilité maximale est donc obtenue pour k_0, en particulier $k_0 = \lfloor (n+1)/2 \rfloor$ lorsque $p = 1/2$, et correspond au plus grand coefficient du binome d'ordre n.

Variables aléatoires indépendantes

• Deux variables aléatoires X et Y sur un même univers Ω définissent un couple de variables aléatoires (X, Y).
La loi conjointe de X et de Y est la loi de ce couple de variable aléatoire $P((X,Y) = (a,b))$.
Réciproquement les lois X et Y du couple (X,Y) de variables aléatoires sont appelées lois marginales de la loi conjointe.

∗ La loi conjointe $P((X,Y) = (a,b))$ permet de retrouver les lois marginales :

$$P(X = a) = \sum_{b \in Y(\Omega)} P((X,Y) = (a,b))$$
$$P(Y = b) = \sum_{a \in X(\Omega)} P((X,Y) = (a,b))$$

⋆ Réciproquement connaitre les deux lois marginales ne permet pas de trouver la loi conjointe. Les trois expériences suivantes ont les mêmes lois marginales de répartition uniforme mais la loi conjointe est différente.
La première expérience lance deux dés indépendants. La deuxième consiste à lancer un dé prenant la variable aléatoire X, et la variable aléatoire Y est la même $Y = X$. La troisième consiste aussi à lancer un seul dé dont la valeur est la variable aléatoire X, et la variable aléatoire Y est $Y = 7 - X$.
Dans ces trois cas les lois marginales sont les lois uniformes sur $[\![1,6]\!]$, mais les trois lois conjointes sont différentes comme le résume ces tableaux :

Expérience 1 :	$P((X,Y) = (a,b)) = \frac{1}{36}$	pour $(a,b) \in [\![1,6]\!]^2$
Expérience 2 :	$P((X,Y) = (a,a)) = \frac{1}{6}$	pour $a \in [\![1,6]\!]$
	$P((X,Y) = (a,b)) = 0$	pour $a \neq b$
Expérience 3 :	$P((X,Y) = (a,7-a)) = \frac{1}{6}$	pour $a \in [\![1,6]\!]$
	$P((X,Y) = (a,b)) = 0$	pour $a + b \neq 7$

• Deux variables aléatoires X et Y sont indépendantes si et seulement si pour toutes les valeurs possibles de ces variables la loi conjointe et les deux lois marginales vérifient cette égalité :

$$P(X = a)\, P(X = b) = P((X,Y) = (a,b))$$

∗ Cette définition peut aussi bien être donnée pour tous les événements possibles $X(\omega) \in A$ et $Y(\omega) \in B$ que pour toutes les conditions élémentaires $X = a$ et $Y = b$:

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

◊ Le passage de l'un à l'autre s'effectue par somme sur des sous-ensembles fini des valeurs possibles $X(\Omega)$ et $Y(\Omega)$:

$$P(X \in A) = \sum_{a \in A} P(X = a)$$
$$P(Y \in B) = \sum_{b \in B} P(Y = b)$$
$$P((X,Y) \in A \times B) = \sum_{(a,b) \in A \times B} P((X,Y) = (a,b))$$

En toute rigueur les ensembles finis d'indices devraient être $A \cap X(\Omega)$ à la place de A, $B \cap Y(\Omega)$ à la place de B et $(A \times B) \cap (X(\Omega) \times Y(\Omega))$ à la place de $A \times B$.

⋆ Deux événements peuvent être indépendant sans que les lois soient indépendantes. Notons (X, Y) les lois modélisant le lancer de deux dés, et $S = X + Y$ la variable aléatoire somme.
Les lois X et S ne sont pas indépendantes, par exemple :

$$P((X,S) = (6,2)) = 0 \neq P(X = 6)\, P(S = 2) = \frac{1}{6 \times 36}$$

Les événements $X = k$ et $S = 7$ sont indépendants :

$$P(S = 7) = \frac{6}{36} \qquad \text{car } S = 7 \text{ correspond à 6 tirages possibles sur 36}$$

$$P((X,S) = (k,7)) = P((X,Y) = (k, 7-k)) = \frac{1}{36}$$
$$= P(X = k)\, P(S = 7)$$

• Trois variables aléatoires X, Y et Z sont mutuellement indépendantes si et seulement si ces lois vérifient toutes ces égalités :

$$P(X=a)\,P(Y=b)\,P(Z=c) = P((X,Y,Z)=(a,b,c))$$
$$P(X=a)\,P(Y=b) = P((X,Y)=(a,b))$$
$$P(X=a)\,P(Z=c) = P((X,Z)=(a,c))$$
$$P(Y=b)\,P(Z=c) = P((Y,Z)=(b,c))$$

Plus généralement les n variables aléatoires $(X_k)_{k=1}^n$ sont mutuellement indépendantes si et seulement si pour toutes les valeurs $(a_k)_{k=1}^n \in \mathbb{R}^n$ possibles les n événements $X_k = a_k$ sont mutuellement indépendants.
Les conditions $X_k = a_k$ peuvent être remplacées par $X_k \in A_k$ dans ces égalités de produits.

$*$ Si les variables aléatoires X et Y sont indépendantes alors les compositions par des applications réelles $f \circ X$ et $g \circ Y$ sont des variables aléatoires indépendantes.
Ces événements sont égaux :

$$f(X(\omega)) \in A \Longleftrightarrow X(\omega) \in f^{-1}(A)$$
$$g(Y(\omega)) \in B \Longleftrightarrow Y(\omega) \in g^{-1}(B)$$

Somme de variables de Bernoulli indépendantes

■ La loi d'une somme de n variables de Bernouilli $(X_i)_{i=1}^n$ de paramètre p est la loi binomiale $\mathcal{B}(p,n)$:

$$P(\sum_{i=1}^n X_i = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

♦ Les n variables aléatoires de Bernoulli X_i sont mutuellement indépendantes. La probabilité des intersections des n cas est le produit des probabilités p ou $1-p$:

$$\begin{aligned} &P((X_1,\cdots,X_k,X_{k+1},X_n) = (1,\cdots,1,0,\cdots,0)) \\ &= P(X_1=1)\cdots P(X_k=1)\,P(X_{k+1}=0)\cdots P(X_n=0) \\ &= p^k(1-p)^{n-k} \end{aligned}$$

Il existe en fait $\binom{n}{k}$ sous-ensembles de k indices parmi les n variables aléatoires qui déterminent les k indices i des variables aléatoires X_i de valeur 1 pour que la somme soit k :

$$S = \sum_{i=1}^n X_i \qquad P(S=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Espérance et moments des variables aléatoires

Espérance

• L'espérance $\mathrm{E}(X) \in \mathbb{R}$ d'une variable aléatoire réelle X est sa moyenne pondérée :

$$\mathrm{E}(X) = \sum_{\omega\in\Omega} X(\omega)\,P(\{\omega\}) = \sum_{x\in X(\Omega)} x\,P(X=x)$$

Une variable aléatoire X est centrée lorsque $\mathrm{E}(X) = 0$.

■ L'espérance est linéaire, positive et croissante :

$$\mathrm{E}(\lambda X) = \lambda\,\mathrm{E}(X) \qquad \mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$$
$$\forall\,\omega\in\Omega \quad X(\omega) \geq 0 \Longrightarrow \mathrm{E}(X) \geq 0$$
$$\forall\,\omega\in\Omega \quad X(\omega) \leq Y(\omega) \Longrightarrow \mathrm{E}(X) \leq \mathrm{E}(Y)$$

◊ Ces démonstrations reposent sur des manipulations de sommes :

$$\mathrm{E}(\lambda X) = \sum_{\omega\in\Omega} \lambda\,X(\omega)\,P(\{\omega\}) = \lambda \sum_{\omega\in\Omega} X(\omega)\,P(\{\omega\}) = \lambda\,\mathrm{E}(X)$$

$$\begin{aligned} \mathrm{E}(X+Y) &= \sum_{\omega\in\Omega} (X(\omega)+Y(\omega))\,P(\{\omega\}) \\ &= \sum_{\omega\in\Omega} X(\omega)\,P(\{\omega\}) + \sum_{\omega\in\Omega} Y(\omega)\,P(\{\omega\}) = \mathrm{E}(X) + E(Y) \end{aligned}$$

◊ La propriétés suivantes reposent sur les inégalités $P(\omega) \geq 0$ et $X(\omega) \geq 0$.

$$\mathrm{E}(X) = \sum_{\omega\in\Omega} \lambda\,X(\omega)\,P(\{\omega\}) \geq 0$$

★ L'espérance d'une variable aléatoire constante $X = a$ est a :

$$\sum_{X(\Omega)=\{a\}} x\,P(X=x) = a \times 1 = a$$

L'espérance d'une variable X de Bernoulli $\mathcal{B}(p)$ est $\mathrm{E}(X) = p$:

$$\sum_{X(\Omega)=\{0,1\}} x\,P(X=x) = p \times 1 + (1-p) \times 0 = p$$

L'espérance d'une variable X binomiale $\mathcal{B}(p,n)$ est $\mathrm{E}(X) = np$.

◊ Une démonstration possible porte sur de manipulations de la formule du binome :

$$\sum_{x\in[\![0,n]\!]} x\,P(X=x) = \sum_{k=0}^{n} k\binom{n}{k} p^k(1-p)^{(n-k)}$$
$$=\sum_{k=1}^{n} k\binom{n}{k} p^k(1-p)^{(n-k)} = \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k(1-p)^{(n-k)}$$
$$=n\sum_{k=0}^{n-1}\binom{n}{k-1} p\,p^{k-1}(1-p)^{(n-1-k)} = np(1+(1-p))^{n-1} = np$$

∗ La composition par la fonction réelle f aboutit à

$$\mathrm{E}(f\circ X) = \mathrm{E}(f(X)) = \sum_{x\in X(\Omega)} f(x)\,P(X=x)$$

∗ L'espérance de $Y = aX + b$ est $\mathrm{E}(Y) = a\,\mathrm{E}(X) + b$ par linéarité :

$$\mathrm{E}(Y) = \mathrm{E}(aX+b) = a\,\mathrm{E}(X) + \mathrm{E}(b) = a\,\mathrm{E}(X) + b$$

Espérance d'un produit de variables indépendantes

■ Espérance du produit de variables aléatoires indépendantes est le produit des espérances : $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$.

◊ L'espérance du produit des deux lois marginales indépendantes de la loi conjointe est la suivante :

$$\begin{aligned}\mathrm{E}(XY) &= \sum_{(x,y)\in(X(\Omega)\times Y(\Omega))} xy\,P((X,Y)=(x,y))\\ &= \sum_{(x,y)\in(X(\Omega)\times Y(\Omega))} xy\,P(X=x)\,P(Y=y)\\ &= \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} xy\,P(X=x)\,P(Y=y)\\ &= \sum_{x\in X(\Omega)} x\,P(X=x)\sum_{y\in Y(\Omega)} y\;P(Y=y) = \mathrm{E}(X)\;\mathrm{E}(Y)\end{aligned}$$

►► Un trousseau a huit clefs similaires, mais une seule ouvre une porte donnée. Combien faut-il en moyenne d'essais pour ouvrir la porte, d'une part en prenant ces clefs méthodiquement, sans les remélanger, et d'autre part en les confondant de nouveau entre deux essais.

▻▻ Notons A_k l'événement consistant à ouvrir la serrure à l'essai $k \in [\![1,8]\!]$. Des probabilités conditionnelles obtenues successivement pour $k=0$, $k=1$, $k=2$, etc. donnent la probabilité de chaque test, et notons X la variable aléatoire décrivant le nombre d'essais à effectuer :

$$P_(A_1) = \frac{1}{8} \qquad P(\overline{A_1}) = \frac{7}{8} \qquad P(X=1) = \frac{1}{8}$$
$$P_{\overline{A_1}}(A_2) = \frac{1}{7} \qquad P(X=2) = P(\overline{A_1})\,P_{\overline{A_1}}(A_2) = \frac{7}{8}\frac{1}{7} = \frac{1}{8}$$
$$P(\overline{A_1}\cap\overline{A_2}) = 1 - P(X=1) - P(X=2) = \frac{6}{8}$$
$$P_{\overline{A_1}\cap\overline{A_2}}(A_3) = \frac{1}{6}$$
$$P(X=3) = P(\overline{A_1}\cap\overline{A_2})\,P_{\overline{A_1}\cap\overline{A_2}}(A_3) = \frac{6}{8}\frac{1}{6} = \frac{1}{8}$$
$$P_{\overline{A_1}\cap\cdots\cap\overline{A_k}}(A_{k+1}) = \frac{1}{8-k}$$
$$P(\overline{A_1}\cap\cdots\cap\overline{A_k}) = 1 - P(X\le k) = \frac{k}{8} = \frac{8-k}{8}$$
$$P(X=k+1) = P(\overline{A_1}\cap\cdots\cap\overline{A_k})\,P_{\overline{A_1}\cap\cdots\cap\overline{A_k}}(A_{k+1}) = \frac{1}{8}$$
$$\mathrm{E}(X) = \sum_{k=1}^{8} k\,P(X=k) = \frac{1}{8}\sum_{k=1}^{8} k = \frac{8\times 9/2}{8} = \frac{9}{2}$$

Cette probabilité uniforme de réussite de 1/8 par essai était prévisible, on peut aussi étudier en quelle position est la bonne clé parmi les 8, en dénombrant toutes les énumérations possibles des clefs, avec 8! cas en tout, et 7! cas où la bonne clé est en position k.

▻▻ La méthode est la même si les clefs ont été mélangées de nouveau entre chaque tirage :

$$P_(A_1) = \frac{1}{8} \qquad P(X=1) = \frac{1}{8}$$
$$P_{\overline{A_1}}(A_2) = \frac{1}{8} \qquad P(X=2) = P_{\overline{A_1}}(A_2)\,P(A_2) = \frac{7}{8}\times\frac{1}{8}$$
$$P_{\overline{A_1}\cap\overline{A_2}}(A_3) = \frac{1}{8} \qquad P(\overline{A_1}\cap\overline{A_2}) = P(\overline{A_1})\,P(\overline{A_2}) = \left(\frac{7}{8}\right)^2$$
$$P(X=3) = \frac{1}{8}\left(\frac{7}{8}\right)^2$$

$$P(\overline{A_1}\cap\cdots\cap\overline{A_k}) = P(\overline{A_1})\,P(\overline{A_2}\,\cdots\,\overline{A_k}) = \Big(\frac{7}{8}\Big)^k$$
$$P(X=k) = \frac{1}{8}\Big(\frac{7}{8}\Big)^k$$

La somme $\displaystyle\sum_{k=1}^{N} kx^k$ correspond à la dérivée d'une série géométrique :

$$\begin{aligned}\sum_{k=1}^{N} k\,x^k = x\sum_{k=1}^{N} k\,x^{k-1} = x\sum_{k=1}^{N}(x^k)' = x\Big(\sum_{k=0}^{N} x^k\Big)' = x\Big(\frac{1-x^{N+1}}{1-x}\Big)'\\ = x\,\frac{1-x^{N+1}-(N+1)x^N(1-x)}{(1-x)^2}\\ = \frac{x+Nx^{N+2}-(N+1)x^{N+1}}{(1-x)^2}\end{aligned}$$

$$\sum_{k=1}^{+\infty} k\,x^k = \frac{x}{(1-x)^2}$$

Cette variable aléatoire X n'est pas finie, elle prend ses valeurs dans $\mathbb{N}$, et est appelée variable aléatoire discrète. L'espérance est la valeur limite de la série, et non une somme finie :

$$\mathrm{E}(X) = \sum_{k=1}^{+\infty}\frac{k}{8}\Big(\frac{7}{8}\Big)^k = \frac{1}{8}\sum_{k=1}^{+\infty}\Big(\frac{7}{8}\Big)^k = \frac{1}{8}\,\frac{7/8}{1/8^2} = 7 > \frac{9}{2}$$

Variance et écart-type d'une variable aléatoire

• La variance $\mathrm{V}(X)$ d'une variable aléatoire réelle X mesure le carré des écarts entre les valeurs de X et son espérance :

$$\mathrm{V}(X) = \mathrm{E}((X-\mathrm{E}(X))^2) = \mathrm{E}(X^2)-\mathrm{E}(X)^2$$

◊ La démonstration de la dernière égalité repose sur la linéarité de l'espérance ; notons $\lambda = \mathrm{E}(X)$, pour mettre en évidence cette propriété ; ainsi $\mathrm{E}(\lambda^2) = \lambda^2$ pour la loi constante λ^2 :

$$\begin{aligned}\mathrm{V}(X) &= \mathrm{E}((X-\mathrm{E}(X))^2) = \mathrm{E}((X-\lambda)^2) = \mathrm{E}(X^2-2\lambda X+\lambda^2)\\ &= \mathrm{E}(X^2) - 2\lambda\,\mathrm{E}(X) + \mathrm{E}(\lambda^2) = \mathrm{E}(X^2) - 2\,\mathrm{E}(X)^2 + \mathrm{E}(X)^2\\ &= \mathrm{E}(X^2)-\mathrm{E}(X)^2\end{aligned}$$

∗ La variance est une valeur positive car espérance de variable aléatoire $(X-\mathrm{E}(X))^2$ à valeurs positives :

$$\forall\,\omega\in\Omega\quad (X(\omega)-\mathrm{E}(X))^2 \geq 0 \Longrightarrow \mathrm{E}((X-\mathrm{E}(X))^2)$$

• L'écart-type d'une variable aléatoire X est $\sigma(X) = \sqrt{\mathrm{V}(X)}$.

∗ La variance d'une variable aléatoire constante $X = a$ est nulle.

$$\mathrm{V}(X) = \mathrm{E}((X-\mathrm{E}(X)^2) = \mathrm{E}((X-a)^2) = \mathrm{E}(0) = 0$$

∗ La variance d'une variable aléatoire X de Bernoulli $\mathcal{B}(p)$ est $\mathrm{V}(X) = p(1-p)$, où $X^2 = X$ car X est de valeurs 0 ou 1.

$$\begin{aligned}\mathrm{V}(X) &= \mathrm{E}((X-p)^2) = \mathrm{E}(X^2-2pX+p^2) = \mathrm{E}(X-2pX+p^2)\\ &= \mathrm{E}(X) - 2p\,\mathrm{E}(X) + p^2 = p - p^2 = p(1-p)\end{aligned}$$

∗ La variance d'une variable aléatoire X de loi binomiale $\mathcal{B}(p,n)$ $\mathrm{V}(X) = np(1-p)$.

◊ Une démonstration possible repose sur la manipulation des coefficients du binome :

$$k^2 = (k(k-1)+k) \qquad k\binom{n}{k} = \binom{n-1}{k-1} \qquad \text{pour } 1\leq k\leq n$$

$$k(k-1)\binom{n}{k} = n(k-1)\binom{n-1}{k-1} = n(n-1)\binom{n-2}{k-2} \qquad \text{pour } 2\leq k\leq n.$$

$$\begin{aligned}&\mathrm{E}((X-\mathrm{E}(X))^2) = \mathrm{E}(X^2)-\mathrm{E}(X)^2\\ &= \sum_{k=0}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k} - (np)^2\\ &= \sum_{k=1}^{n}(k(k-1)+k)\binom{n}{k}p^k(1-p)^{n-k} - (np)^2\\ &= \sum_{k=2}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k} + \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k} - (np)^2\\ &= \sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2}p^k(1-p)^{n-k} + \sum_{k=1}^{n} n\binom{n}{k}p^k(1-p)^{n-k}\\ &\qquad - (np)^2\\ &= \sum_{k=0}^{n-2} n(n-1)\binom{n-2}{k}p^{k+2}(1-p)^{n-2-k}\\ &\qquad + \sum_{k=0}^{n-1} n\binom{n-1}{k-1}p^{k+1}(1-p)^{n-1-k} - (np)^2\end{aligned}$$

$$= n(n-1)p^2(p+(1-p))^{n-2} + np(p+(1-p))^{n-1} - (np)^2$$
$$= n^2p^2 - np^2 + np - n^2p^2 = np(1-p)$$

∗ La variance de $Y = aX + b$ est $\mathrm{V}(Y) = a^2\,\mathrm{V}(X)$ par linéarité :

$$\begin{aligned}\mathrm{V}(aX+b) &= \mathrm{E}((Y-\mathrm{E}(Y))^2) = \mathrm{E}((aX+b-a\,\mathrm{E}(X)-b)^2)\\ &= \mathrm{E}((a^2(X-\mathrm{E}(X))^2) = a^2\,\mathrm{V}(X)\\ \sigma(aX+b) &= \sqrt{a^2\,\mathrm{V}(X)} = |a|\,\mathrm{V}(X)\end{aligned}$$

◦ Une variable aléatoire X est réduite si et seulement si $\mathrm{V}(X) = 1$.

Moment d'une variable aléatoire

• Plus généralement le moment d'ordre $k \in \mathbb{N}^*$ de la variable aléatoire X est $m_k = \mathrm{E}(X^k) \in \mathbb{R}$.

∗ Le moment centré $\mu_k = \mathrm{E}((X-\mathrm{E}(X))^k)$ s'obtient par développement en fonction du moment $m_k = \mathrm{E}(X^k)$:

$$\begin{aligned}\mu_k &= \mathrm{E}((X-\mathrm{E}(X))^k) = \mathrm{E}\Big(\sum_{i=0}^{k}\binom{k}{i}\mathrm{E}(X)^{k-i}X^i\Big)\\ &= \sum_{i=0}^{k}\binom{k}{i}\mathrm{E}(X)^{k-i}\,\mathrm{E}(X^i) = \sum_{i=0}^{k}\binom{k}{i}\mathrm{E}(X)^{k-i}m_i\end{aligned}$$

Inégalité de Markov et deTchebychev d'une variable aléatoire

■ L'inégalité de Markov d'une variable aléatoire à valeurs positives minore une probabilité par l'espérance :

$$P(X \geq \lambda\,\mathrm{E}(X)) \leq \frac{1}{\lambda} \qquad \text{où } \lambda > 0 \text{ et } X(\omega) \geq 0$$

◊ La démonstration repose sur la positivité des probabilités :

$$\begin{aligned}\mathrm{E}(X) &= \sum_{a\in X(\Omega)} a\,P(X=a)\\ &\geq \sum_{\substack{a\in X(\Omega)\\ a\geq\lambda\,\mathrm{E}(X)}} a\,P(X=a) \geq \lambda\,\mathrm{E}(X)\,P(X\geq\lambda\,\mathrm{E}(X))\end{aligned}$$

$$\mathrm{E}(X) \geq \lambda\,\mathrm{E}(X)\,P(X\geq\lambda\,\mathrm{E}(X)) \quad \text{donc} \quad P(X\geq\lambda\,\mathrm{E}(X)) \leq \frac{1}{\lambda}$$

□ L'inégalité de Bienaymé-Tchebychev se déduit de l'inégalité de Markov pour faire intervenir la variance :

$$P(|X-\mathrm{E}(X)| \geq \gamma) \leq \frac{\mathrm{V}(X)}{\gamma^2} \qquad \text{où } \gamma > 0$$

◊ La démonstration applique l'inégalité de Markov à la variable aléatoire positive $Y = (X-\mathrm{E}(X))^2$ et $\lambda = \gamma^2/\,\mathrm{E}(Y) = \gamma^2/\,\mathrm{V}(X)$:

$$P(Y \geq \lambda\,\mathrm{E}(Y)) = P(|X-\mathrm{E}(X)| \geq \gamma) \leq \frac{1}{\lambda} = \frac{\mathrm{V}(X)}{\gamma^2}$$

∗ Soit une famille $(X_k)_{k=1}^n$ de n variables indépendantes d'espérance $m = \mathrm{E}(X_k)$ et d'écart-type $\sigma = \sqrt{\mathrm{V}(X_k)}$. L'espérance et la variance de la loi moyenne M de ces n lois aboutit par l'inégalité de Bienaymé-Tchebychev :

$$S = \sum_{k=1}^{n} X_k \qquad M = \frac{S}{n} \qquad \mathrm{E}(M) = m$$
$$\mathrm{V}(S) = n\sigma^2 \qquad \mathrm{V}(M) = \frac{\sigma^2}{n}$$
$$P(|M-m| \geq \alpha) \leq \frac{\sigma^2}{n\alpha^2}$$

Covariance de deux variables aléatoires

• Par définition la convariance de deux variables aléatoires réelles X et Y est $\mathrm{cov}(X,Y)$:

$$\mathrm{cov}(X,Y) = \mathrm{E}((X-\mathrm{E}(X))(Y-\mathrm{E}(Y))) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$$

◊ La dernière égalité repose sur la distributivité du produit et cette première égalité pour toutes les valeurs possibles de y :

$$\mathrm{E}(X) = \sum_x x\,P(X=x) = \sum_{x,y} x\,P((X,Y)=(x,y))$$

$$\begin{aligned}
&\mathrm{E}((X-\mathrm{E}(X))(Y-\mathrm{E}(Y)))\\
&= \sum_{(x,y)} (x-\mathrm{E}(X))(y-\mathrm{E}(Y))\,P((X,Y)=(x,y))\\
&= \sum_{(x,y)} (xy - x\ \mathrm{E}(Y) - \mathrm{E}(X)\,y + \mathrm{E}(X)\ \mathrm{E}(Y))P((X,Y)=(x,y))\\
&= \sum_{(x,y)} xy\,P((X,Y)=(x,y)) - \mathrm{E}(Y)\sum_x x\,P(X=x)\\
&\qquad\qquad - \mathrm{E}(X)\sum_y y\,P(Y=y) + \mathrm{E}(X)\ \mathrm{E}(Y)\\
&= \mathrm{E}(XY) - 2\ \mathrm{E}(X)\ \mathrm{E}(Y) + \mathrm{E}(X)\ \mathrm{E}(Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)
\end{aligned}$$

Les sommes sur (x,y), x et y sont finies pour ces variables aléatoires finies et portent sur $(x,y)\in X(\Omega)\times Y(\Omega)$, $x\in X(\Omega)$ et $y\in Y(\Omega)$.

■ La covariance de deux variables aléatoires indépendantes est nulle.

◊ La démonstration repose sur l'espérance de deux variables X et Y indépendantes :

$$\mathrm{E}(XY) = \mathrm{E}(X)\ \mathrm{E}(Y) \qquad \mathrm{cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\ \mathrm{E}(Y)$$

∗ Deux variables peuvent avoir une covariance nulle sans être indépendante : Soit R la variable aléatoire d'un jeu de pile ou face équilibré de valeurs ± 1, et D un jeu de dé, les variables aléatoires centrées X et Y sont $X = 2D-7 \in \{\pm 5, \pm 3, \pm 1\}$, et $Y = RX$. Les variables R et X sont indépendantes :

$$\begin{aligned}
\mathrm{E}(D) &= \frac{7}{2} \qquad \mathrm{E}(X) = 2\,\mathrm{E}(D) - 7 = 0 \qquad \mathrm{E}(R) = 0\\
\mathrm{E}(Y) &= \mathrm{E}(RD) = \mathrm{E}(R)\ \mathrm{E}(D) = 0\\
P((X,Y)=(5,5)) &= P((D,R)=(6,1)) = \frac{1}{12}\\
&\neq P((X,Y)=(1,3)) = 0\\
\mathrm{cov}(X,Y) &= \mathrm{E}(XY) - \mathrm{E}(X)\ \mathrm{E}(Y)\\
&= \mathrm{E}(RX^2) - \mathrm{E}(R)\,\mathrm{E}(X)^2\\
&= \mathrm{E}(R)\,(\mathrm{E}(X^2) - \mathrm{E}(X)^2) = \mathrm{E}(R)\ \mathrm{V}(X) = 0
\end{aligned}$$

∗ La définition $\mathrm{cov}(X,Y) = \mathrm{E}((X-\mathrm{E}(X))(Y-\mathrm{E}(Y)))$ que la convariance est positive si les variations des deux variables X et Y ont

tendance à être dans le même sens : $X-\mathrm{E}(X)$ et $Y-\mathrm{E}(Y)$ souvent du même signe.

■ La variance d'une somme fait intervenir la covariance :

$$\mathrm{V}(X+Y) = \mathrm{V}(X) + V(Y) + 2\ \mathrm{cov}(X,Y)$$

◊ La démonstration passe par la linéarité de l'espérance :

$$\begin{aligned}
\mathrm{V}(X+Y) &= \mathrm{E}((X+Y) - \mathrm{E}(X) - \mathrm{E}(Y))^2\\
&= \mathrm{E}(((X-\mathrm{E}(X)) + (Y-\mathrm{E}(Y)))^2)\\
&= \mathrm{E}((X-\mathrm{E}(X))^2 + (Y-\mathrm{E}(Y))^2 + 2(X-\mathrm{E}(X))(Y-\mathrm{E}(Y)))\\
&= \mathrm{E}((X-\mathrm{E}(X))^2) + \mathrm{E}((Y-\mathrm{E}(Y))^2)\\
&\qquad + 2\ \mathrm{E}((X-\mathrm{E}(X))(Y-\mathrm{E}(Y)))\\
&= \mathrm{V}(X) + \mathrm{V}(Y) + 2\ \mathrm{cov}(X,Y)
\end{aligned}$$

∗ covariance symétrique et linéaire par rapport à chaque variable :

$$\begin{gathered}
\mathrm{cov}(X,X) = \mathrm{V}(X) \qquad \mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)\\
\mathrm{cov}(\lambda X, Y) = \mathrm{cov}(X, \lambda Y) = \lambda\ \mathrm{cov}(X,Y)\\
\mathrm{cov}(X+Y, Z) = \mathrm{cov}(X,Z) + \mathrm{cov}(Y,Z)
\end{gathered}$$