MODÉLISATION PROBABILISTE SUR UN UNIVERS FINI

1 EXPÉRIENCE ALÉATOIRE, ÉVÉNEMENTS ET VARIABLES ALÉATOIRES

- Le concept d'*expérience aléatoire* n'est pas mathématique à proprement parler. On appelle ainsi toute expérience — expérience matérielle ou expérience de pensée — susceptible a priori de *résultats* différents quand on la répète. L'ensemble des résultats observables d'une telle expérience est appelé son *univers* et généralement noté Ω. Plutôt que de résultats, on parle aussi souvent d'*issues* et de *réalisations*.
  - L'expérience d'un lancer de dé à 6 faces peut conduire à 6 résultats selon la face obtenue. Pour modéliser un tel lancer, on peut choisir pour univers Ω l'ensemble $[\![1,6]\!]$.
  - L'expérience du tirage de 2 boules successivement avec remise dans une urne contenant des boules noires et des boules blanches peut conduire à 4 résultats : NN, NB, BN et BB si on choisit d'associer « N » à la couleur noire et « B » à la couleur blanche. Pour modéliser un tel tirage, on peut choisir pour univers Ω l'ensemble $\{\mathrm{N},\mathrm{B}\}^2$ des 2-listes de $\{\mathrm{N},\mathrm{B}\}$, i.e. l'ensemble des mots de 2 lettres sur l'alphabet $\{\mathrm{N},\mathrm{B}\}$.

En MPSI, nos univers seront toujours des ensembles **FINIS**. Vous étudierez certains univers infinis en deuxième année.

Cette restriction du programme aux univers finis est uniquement technique. La théorie des probabilités sur un univers quelconque est techniquement délicate. L'ennui bien sûr, c'est qu'en limitant la taille des univers, on limite drastiquement le nombre des expériences aléatoires autorisées. Il nous sera par exemple impossible en MPSI :

- de jouer à pile ou face **INDÉFINIMENT** même si nous tricherons un peu parfois...
- de jouer aux fléchettes contre un disque de rayon 2 et de calculer la probabilité p d'atterrir dans le disque central de rayon 1 — intuitivement : $p=\dfrac{\pi\times 1^2}{\pi\times 2^2}=\dfrac{1}{4}$, mais comment justifier proprement ce rapport de surfaces ?

- Pour une expérience aléatoire donnée, on peut bien sûr appréhender chaque résultat isolément en tant qu'**ÉLÉMENT** de Ω, mais ce qui nous intéresse généralement, ce sont des regroupements de résultats définis génériquement par une propriété commune et appelés *événements*.
  - Dans le lancer de dé précédent, la propriété « La face obtenue est paire » est satisfaite par 3 résultats, en l'occurrence 2, 4 et 6, et sera identifiée à l'ensemble $\{2,4,6\}$.
  - Dans le tirage dans l'urne précédent, la propriété « La première boule est blanche » est satisfaite par 2 résultats, en l'occurrence BN et BB, et sera identifiée à l'ensemble $\{\mathrm{BN},\mathrm{BB}\}$.

  Plus généralement, toute partie de Ω est appelée un *événement*. L'ensemble des événements de l'expérience aléatoire étudiée est ainsi l'ensemble total $\mathscr{P}(\Omega)$ des parties de Ω.

Définition (Vocabulaire usuel sur les événements) Soit Ω un univers fini.

- Les singletons de Ω sont appelés les *événements élémentaires* de Ω. L'événement Ω est appelé l'*événement certain* et l'événement $\varnothing$ l'*événement impossible*.
- Pour tous événements $A,B\in\mathscr{P}(\Omega)$, l'événement $A\cup B$ est appelé l'*événement « A ou B »* et l'événement $A\cap B$ l'*événement « A et B »*. L'événement $\overline{A}$ est quant à lui appelé l'*événement contraire de A*. On dit enfin que les événements A et B sont *incompatibles* s'ils sont disjoints, i.e. si : $A\cap B=\varnothing$.
- On appelle *système complet d'événements* de Ω tout ensemble $\{A_1,\ldots,A_n\}$ d'événements deux à deux incompatibles dont la réunion est l'événement certain, i.e. pour lequel : $\Omega=\bigsqcup_{i=1}^{n}A_i$ — réunion **DISJOINTE**, attention !

Exemple Pour l'expérience aléatoire du lancer d'un dé à 6 faces, les événements « La face obtenue est paire » et « La face obtenue est impaire » forment un système complet d'événements car on obtient forcément une face paire ou une face impaire et jamais les deux en même temps.

Plus généralement, soient Ω l'univers fini d'une certaine expérience aléatoire et $A, B \in \mathscr{P}(\Omega)$ des événements. La paire d'événements $\left\{A, \overline{A}\right\}$ est alors un système complet d'événements car : $A \cup \overline{A} = \Omega$ et $A \cap \overline{A} = \varnothing$, et il n'est pas dur de vérifier que la collection $\left\{A \cap B, A \cap \overline{B}, \overline{A} \cap B, \overline{A} \cap \overline{B}\right\}$ en est un également, de même que $\left\{A, \overline{A} \cap B, \overline{A} \cap \overline{B}\right\}$. Dessinez quelques patates pour vous en convaincre !

Exemple Pour tout univers fini $\Omega = \left\{\omega_1, \ldots, \omega_n\right\}$: $\Omega = \bigsqcup_{i=1}^{n} \left\{\omega_i\right\}$, donc les événements élémentaires $\left\{\omega_1\right\}, \ldots, \left\{\omega_n\right\}$ forment un système complet d'événements.

La théorie des probabilités nous mettra généralement en présence d'événements plus compliqués à décrire que de simples « A ou B » ou « A et B ». Nous aurons par exemple affaire à des réunions d'intersection qui vous paniqueront au départ, mais que vous trouverez vite naturelles.

Exemple On lance une pièce n fois successivement et on note Ω un univers susceptible de modéliser cette expérience — la plupart du temps, vous verrez, nous n'aurons pas besoin de définir Ω. Dans ce cadre, pour tout $k \in [\![1, n]\!]$, on note F_k l'événement « On obtient face au $k^{\text{ème}}$ lancer ». Les événements $F_1, \ldots, F_n$ sont en quelque sorte les événements les plus basiques auxquels cette expérience aléatoire nous confronte. Il paraît raisonnable que tout événement puisse être écrit en fonction de ces événements basiques.

Par exemple :

— l'événement A « On obtient pile au moins une fois au cours des n lancers » peut être écrit : $A = \bigcup_{k=1}^{n} \overline{F_k}$,

— l'événement B « On n'obtient jamais pile » peut être écrit : $B = \bigcap_{k=1}^{n} F_k$,

— et l'événement C « On obtient deux faces consécutifs au cours des n lancers » peut être écrit : $C = \bigcup_{k=1}^{n-1} \left(F_k \cap F_{k+1}\right)$.

En réalité, la plupart du temps, les événements qu'on manipule sont construits à partir de grandeurs, numériques ou non, appelées *variables aléatoires*. Par exemple, quand on lance une pièce 10 fois successivement, l'événement « On a obtenu exactement 3 piles » peut être écrit « $N = 3$ » si on note N le nombre de piles obtenus. Ce nombre N n'est pas a priori une constante. Il DÉPEND des lancers effectués et doit donc être vu comme une FONCTION.

Définition **(Variable aléatoire)** Soient Ω un univers fini et E un ensemble quelconque.

- On appelle *variable aléatoire sur Ω à valeurs dans E* toute application X de Ω dans E.
  Dans la plupart des situations : $E = \mathbb{R}$ et on appelle X une *variable aléatoire réelle*.
- Pour toute partie A de E, l'événement $X^{-1}(A)$ est noté $\left\{X \in A\right\}$ ou $(X \in A)$:
$$\left\{X \in A\right\} = X^{-1}(A) = \left\{\omega \in \Omega / \quad X(\omega) \in A\right\}.$$
  Dans le cas où A est un singleton $\left\{x\right\}$, on emploie plutôt les notations $\left\{X = x\right\}$ ou $(X = x)$:
$$\left\{X = x\right\} = X^{-1}\Big(\left\{x\right\}\Big) = \left\{\omega \in \Omega / \quad X(\omega) = x\right\}.$$
  Enfin, dans le cas où : $E = \mathbb{R}$ et $A =]-\infty, x]$ pour un certain $x \in \mathbb{R}$, on emploie plutôt les notations $\left\{X \leqslant x\right\}$ et $(X \leqslant x)$ — même chose avec toutes les autres formes d'intervalles :
$$\left\{X \leqslant x\right\} = X^{-1}\big(]-\infty, x]\big) = \left\{\omega \in \Omega / \quad X(\omega) \leqslant x\right\}.$$

En dépit de son nom, une variable aléatoire sur Ω n'est pas une variable, c'est une fonction. De plus, en tant que fonction, une variable aléatoire associe fixement une valeur à tout élément de Ω et ceci n'a rien d'aléatoire. L'aléatoire fera irruption quand nous attribuerons à tout événement de Ω une probabilité, i.e. une certaine mesure de sa vraisemblance.

Exemple

- Pour modéliser le lancer d'un dé à 6 faces 2 fois, on peut choisir pour univers Ω l'ensemble $[\![1, 6]\!]^2$ des 2-listes de $[\![1, 6]\!]$. La fonction X_1 (resp. X_2) « valeur obtenue au premier (resp. deuxième) lancer » est une variable aléatoire réelle sur Ω d'image $[\![1, 6]\!]$. La fonction $S = X_1 + X_2$ en est une autre, d'image $[\![2, 12]\!]$.

  Par exemple, pour $\omega = (2, 5) \in \Omega$: $X_1(\omega) = 2$, $X_2(\omega) = 5$ et $S(\omega) = 7$.
- Pour modéliser le tirage simultané de 4 entiers entre 1 et 10, on peut choisir pour univers Ω l'ensemble $\mathscr{P}_4\big([\![1, 10]\!]\big)$ des 4-combinaisons de $[\![1, 10]\!]$. La fonction X « plus petit entier tiré » est une variable aléatoire réelle sur Ω d'image $[\![1, 7]\!]$ — comme on tire 4 entiers, il est impossible que le plus petit entier tiré vaille 8, 9 ou 10.

  Par exemple, pour $\omega = \left\{2, 5, 6, 8\right\} \in \Omega$: $X(\omega) = 2$.

Définition-théorème (Système complet d'événements associé à une variable aléatoire) Soient Ω un univers fini et X une variable aléatoire sur Ω. Les événements $\{X = x\}$, x décrivant $X(\Omega)$, forment un système complet d'événements de Ω appelé le *système complet d'événements associé à X*.

Démonstration Tout élément ω de Ω appartient à un et un seul événement de la forme $X^{-1}\big(\{x\}\big) = \{X = x\}$, en l'occurrence pour : $x = X(\omega)$. En d'autres termes : $\Omega = \bigsqcup_{x \in X(\Omega)} \{X = x\}$. ■

2 PROBABILITÉS SUR UN UNIVERS FINI

Le concept d'expérience aléatoire décrit des situations susceptibles de conduire à plusieurs résultats, mais rien ne nous permet pour le moment de mesurer la **VRAISEMBLANCE** de ces résultats les uns par rapport aux autres. Nous allons associer dans ce paragraphe à tout événement d'une expérience aléatoire une *probabilité*, i.e. un réel compris entre 0 et 1 dont la valeur 0 représente le plus bas niveau de vraisemblance et la valeur 1 le niveau le plus élevé.

Définition (Probabilité sur un univers fini) Soit Ω un univers fini. On appelle *probabilité sur* Ω toute application $P : \mathscr{P}(\Omega) \longrightarrow [0,1]$ telle que :

$$P(\Omega) = 1 \qquad \text{et :} \qquad \forall A, B \in \mathscr{P}(\Omega), \quad A \cap B = \varnothing \implies P(A \sqcup B) = P(A) + P(B) \quad (\textit{additivité}).$$

Le couple (Ω, P) est alors appelé un *espace probabilisé* (*fini*).

✖ **ATTENTION !** ✖ Le mot « probabilité » est utilisé de deux manières différentes qu'il convient de bien distinguer :

— La « probabilité de A » est le **RÉEL** $P(A)$ de $[0,1]$ qu'on associe à l'événement A pour mesurer sa vraisemblance.

— La « probabilité P » est l'**APPLICATION** de Ω dans $[0,1]$ qu'on choisit pour décrire la vraisemblance de **TOUS** les événements de Ω.

Les *probabilités uniformes* définies ci-dessous constituent l'exemple le plus naturel de probabilité sur un ensemble fini, mais ce ne sont pas les seules probabilités que nous aurons l'occasion d'utiliser — loin de là.

Définition-théorème (Probabilité uniforme) Soit Ω un univers fini. L'application $A \overset{P}{\longmapsto} \dfrac{|A|}{|\Omega|}$ est une probabilité sur Ω appelé sa *probabilité uniforme*.

Vous connaissez la relation : $P(A) = \dfrac{|A|}{|\Omega|}$ sous la forme suivante : $\dfrac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}$.

Exemple On lance une fois un dé équilibre à 6 faces. Quel espace probabilisé pour cette expérience aléatoire ? On peut choisir pour univers Ω l'ensemble $[\![1,6]\!]$ des résultats potentiels et pour probabilité P sur Ω la probabilité uniforme. Avec un dé pipé, on n'aurait bien sûr pas choisi pour P la probabilité uniforme.
Si on note A l'événement « On obtient un nombre premier » et B l'événement « On obtient un nombre pair », alors :

$$P(A) = P\Big(\{2,3,5\}\Big) = \frac{\big|\{2,3,5\}\big|}{|\Omega|} = \frac{3}{6} = \frac{1}{2} \qquad \text{et} \qquad P(B) = P\Big(\{2,4,6\}\Big) = \frac{\big|\{2,4,6\}\big|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

Dans ce cas particulier où : $P(A) = P(B)$, on dit que les événements A et B ont équiprobables. Plus généralement :

Définition (Événements équiprobables) Soient (Ω, P) un espace probabilisé fini et $A, B \in \mathscr{P}(\Omega)$. On dit que les événements A et B sont *équiprobables* si : $P(A) = P(B)$.

Exemple Soit $\Omega=\{\omega_1,\ldots,\omega_n\}$ un univers fini. On note P la probabilité uniforme sur Ω. Les événements élémentaires $\{\omega_1\},\ldots,\{\omega_n\}$ sont équiprobables de probabilité commune $\dfrac{1}{|\Omega|}$.

Théorème (Propriétés des probabilités) Soient (Ω,P) un espace probabilisé fini et $A,B,A_1,\ldots,A_n\in\mathscr{P}(\Omega)$.

(i) **Ensemble vide :** $P(\varnothing)=0$.

(ii) **Complémentaire et différence :** $P\big(\overline{A}\big)=1-P(A)$ et $P\big(A\setminus B\big)=P(A)-P(A\cap B)$.

(iii) **Croissance :** Si : $A\subset B$, alors : $P(A)\leqslant P(B)$.

(iv) **Réunion :** $P(A\cup B)=P(A)+P(B)-P(A\cap B)$. En général : $P\left(\bigcup_{k=1}^n A_k\right)\leqslant\sum_{k=1}^n P(A_k)$ (*sous-additivité*), mais cette inégalité est une égalité si $A_1,\ldots,A_n$ sont DEUX À DEUX INCOMPATIBLES : $P\left(\bigsqcup_{k=1}^n A_k\right)=\sum_{k=1}^n P(A_k)$.

Formule du crible : $P\left(\bigcup_{i=1}^n A_i\right)=\sum_{k=1}^n(-1)^{k+1}\sum_{1\leqslant i_1<\ldots<i_k\leqslant n}P\big(A_{i_1}\cap\ldots\cap A_{i_k}\big)$ (hors programme).

Démonstration

(i) et (ii) Comme : $A=\big(A\setminus B\big)\sqcup(A\cap B)$, alors : $P(A)=P\big(A\setminus B\big)+P(A\cap B)$.
Ainsi : $P\big(\overline{A}\big)=P\big(\Omega\setminus A\big)=P(\Omega)-P(\Omega\cap A)=1-P(A)$, mais aussi : $P(\varnothing)=1-P(\Omega)=1-1=0$.

(iii) Comme : $A\subset B$, alors : $B=A\sqcup\big(B\setminus A\big)$, donc : $P(B)=P(A)+P\big(B\setminus A\big)\geqslant P(A)+0=P(A)$.

(iv) Comme : $A=\big(A\setminus B\big)\sqcup(A\cap B)$, alors : $P(A)=P\big(A\setminus B\big)+P(A\cap B)$. Également : $A\cup B=\big(A\setminus B\big)\sqcup B$, donc : $P(A\cup B)=P\big(A\setminus B\big)+P(B)$. Ensuite, il suffit de mélanger.

Pour la sous-additivité, récurrence ! **Initialisation :** Pour tout $A_1\in\mathscr{P}(\Omega)$: $P(A_1)\leqslant P(A_1)$.

Hérédité : Soit $n\in\mathbb{N}^*$. On suppose que pour tous $A_1,\ldots,A_n\in\mathscr{P}(\Omega)$: $P\left(\bigcup_{k=1}^n A_k\right)\leqslant\sum_{k=1}^n P(A_k)$.

Alors pour tous $A_1,\ldots,A_{n+1}\in\mathscr{P}(\Omega)$: $P\left(\bigcup_{k=1}^{n+1} A_k\right)=P\left(\left(\bigcup_{k=1}^n A_k\right)\cup A_{n+1}\right)$

$$=P\left(\bigcup_{k=1}^n A_k\right)+P(A_{n+1})-P\left(\left(\bigcup_{k=1}^n A_k\right)\cap A_{n+1}\right)\leqslant P\left(\bigcup_{k=1}^n A_k\right)+P(A_{n+1})\overset{\text{HDR}}{\leqslant}\sum_{k=1}^n P(A_k)+P(A_{n+1})=\sum_{k=1}^{n+1}P(A_k).$$

La formule du crible sera prouvée au chapitre « Position et dispersion d'une variable aléatoire réelle ». ■

Nous avons présenté plus haut l'exemple naturel des probabilités uniformes. Notre prochain théorème décrit quant à lui l'ensemble de TOUTES les probabilités qu'on peut installer sur un univers fini. Se donner une probabilité P sur un univers Ω revient toujours à faire un choix dans ce vaste ensemble de toutes les probabilités possibles. Fixer P, c'est décider une fois pour toutes que P mesure fidèlement la vraisemblance des événements de Ω et qu'aucune autre probabilité ne le fait aussi bien. On ne travaille pas dans le cas d'un lancer de dé pipé avec la même probabilité que dans le cas d'un dé équilibré.

Théorème (Détermination d'une probabilité sur les événements élémentaires) Soient $\Omega=\{\omega_1,\ldots,\omega_r\}$ un univers fini et $p_1,\ldots,p_r\in[0,1]$ des réels pour lesquels : $\sum_{i=1}^r p_i=1$. Il existe une et une seule probabilité P sur Ω pour laquelle pour tout $i\in[\![1,r]\!]$: $P\big(\{\omega_i\}\big)=p_i$. En l'occurrence, pour tout $A\in\mathscr{P}(\Omega)$: $P(A)=\sum_{\substack{1\leqslant i\leqslant r\\ \omega_i\in A}}p_i$.

Démonstration

- **Analyse :** Soit P une probabilité sur Ω telle que pour tout $i\in[\![1,r]\!]$: $P\big(\{\omega_i\}\big)=p_i$. Dans ces conditions, pour tout $A\in\mathscr{P}(\Omega)$: $P(A)=P\left(\bigsqcup_{\substack{1\leqslant i\leqslant r\\ \omega_i\in A}}\{\omega_i\}\right)=\sum_{\substack{1\leqslant i\leqslant r\\ \omega_i\in A}}P\big(\{\omega_i\}\big)=\sum_{\substack{1\leqslant i\leqslant r\\ \omega_i\in A}}p_i=\sum_{i=1}^r\mathbb{1}_A(\omega_i)\,p_i$.
- **Synthèse :** Pour tout $A\in\mathscr{P}(\Omega)$, posons : $P(A)=\sum_{i=1}^r\mathbb{1}_A(\omega_i)\,p_i$.

— Pour tout $A\in\mathscr{P}(\Omega)$: $P(A)\in[0,1]$ car : $0\leqslant P(A)\leqslant\sum_{i=1}^r p_i=1$.

— En outre : $P(\Omega) = \sum_{i=1}^{r} \mathbb{1}_\Omega(\omega_i)\, p_i = \sum_{i=1}^{r} p_i = 1.$

— Également, pour tout $j \in [\![1,r]\!]$: $P\big(\{\omega_j\}\big) = \sum_{i=1}^{r} \mathbb{1}_{\{\omega_j\}}(\omega_i)\, p_i = \sum_{i=1}^{r} \delta_{ij}\, p_i = p_j.$

— Enfin, pour tous $A, B \in \mathscr{P}(\Omega)$ incompatibles : $\mathbb{1}_{A\sqcup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B} = \mathbb{1}_A + \mathbb{1}_B$, donc :

$$P(A \sqcup B) = \sum_{i=1}^{r} \mathbb{1}_{A\sqcup B}(\omega_i)\, p_i = \sum_{i=1}^{r} \mathbb{1}_A(\omega_i)\, p_i + \sum_{i=1}^{r} \mathbb{1}_B(\omega_i)\, p_i = P(A) + P(B).$$ ■

Exemple On lance une fois un dé pipé à 6 faces qui donne la face « 1 » avec probabilité $\frac{1}{4}$ et les autres faces avec probabilité $\frac{3}{20}$. Pour modéliser ce lancer, il paraît naturel qu'on choisisse pour univers Ω l'ensemble $[\![1,6]\!]$ des résultats possibles de l'expérience et pour probabilité P la probabilité définie par : $P\big(\{1\}\big) = \frac{1}{4}$ et pour tout $k \in [\![2,6]\!]$: $P\big(\{k\}\big) = \frac{3}{20}$. Il s'agit bien là d'une probabilité car : $\sum_{k=1}^{6} P\big(\{k\}\big) = \frac{1}{4} + 5 \times \frac{3}{20} = 1.$

L'événement A « On obtient une face impaire » a pour probabilité : $P(A) = P\big(\{1\}\big) + P\big(\{3\}\big) + P\big(\{5\}\big) = \frac{1}{4} + 2 \times \frac{3}{20} = \frac{11}{20}.$

3 Loi d'une variable aléatoire

Sur un espace probabilisé fini (Ω, P) fixé, P mesure la vraisemblance de tous les événements concevables, donc en particulier de tous les événements qu'on peut construire à partir d'une variable aléatoire X sur Ω. Les valeurs de X possèdent ainsi chacune leur propre vraisemblance, ce que la définition suivante formalise proprement.

Définition (Loi d'une variable aléatoire) Soient (Ω, P) un espace probabilisé fini, E un ensemble et $X : \Omega \longrightarrow E$ une variable aléatoire. On appelle *loi de X* l'application $x \overset{P_X}{\longmapsto} P(X = x)$ de $X(\Omega)$ dans $[0,1]$.

Remarque importante : $\sum_{x \in X(\Omega)} P(X = x) = 1.$

La loi de X est définie sur $X(\Omega)$, mais on peut calculer $P(X = x)$ pour tout $x \in E \setminus X(\Omega)$: $P(X = x) = P(\varnothing) = 0.$

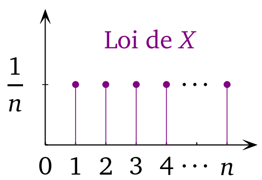
On donne à présent deux premiers exemples de lois uselles.

Définition (Loi uniforme) Soient X une variable aléatoire et E un ensemble fini non vide. On dit que X *suit la loi uniforme sur E*, ce qu'on note : $X \hookrightarrow \mathscr{U}(E)$, si pour tout $x \in E$: $P(X = x) = \frac{1}{|E|}$.

Dans ce cas, pour toute partie A de E : $P(X \in A) = \frac{|A|}{|E|}$.

La loi uniforme est la loi pour laquelle les événements $\{X = x\}$, x décrivant E, sont équiprobables.

La formule : $P(X \in A) = \frac{|A|}{|E|}$ est absolument essentielle en pratique et ramène tout problème de calcul sur la loi uniforme à un « simple » calcul de dénombrement.



Exemple

- Lorsqu'on choisit au hasard un entier X entre 1 et n, X suit la loi uniforme sur $[\![1,n]\!]$.
- Lorsqu'on lance simultanément un dé bleu et un dé rouge à 6 faces, le couple (B, R) des valeurs obtenues — B pour le dé bleu et R pour le rouge — suit la loi uniforme sur $[\![1,6]\!]^2$.
- La loi uniforme sur $\{-1, 1\}$ est souvent appelée la *loi de Rademacher*. On dit donc d'une variable aléatoire Y pour laquelle : $P(Y = -1) = P(Y = 1) = \frac{1}{2}$ qu'elle *suit la loi de Rademacher*.

Exemple Une urne contient 3 boules blanches et 5 noires. On en tire simultanément 4 boules. Avec quelles probabilité n'a-t-on tiré que des boules noires ? On peut bien sûr supposer sans perte de généralité que les 8 boules de l'urne sont numérotées de 1 à 8 histoire de les distinguer — de 1 à 3 pour les blanches et de 4 à 8 pour les noires. Notons alors X la 4-combinaison de $[\![1,8]\!]$ des boules tirées. Cette variable aléatoire suit clairement la loi uniforme sur l'ensemble $\mathscr{P}_4\big([\![1,8]\!]\big)$ des 4-combinaisons de $[\![1,8]\!]$ et la probabilité cherchée vaut : $P\Big(X \in \mathscr{P}_4\big([\![4,8]\!]\big)\Big) = \dfrac{\big|\mathscr{P}_4\big([\![4,8]\!]\big)\big|}{\big|\mathscr{P}_4\big([\![1,8]\!]\big)\big|} = \dfrac{\binom{5}{4}}{\binom{8}{4}} = \dfrac{1}{14}$.

Définition **(Loi de Bernoulli)**

- Soient X une variable aléatoire et $p \in [0,1]$. On dit que X *suit la loi de Bernoulli de paramètre* p, ce qu'on note $X \hookrightarrow \mathscr{B}(p)$, si : $P(X=1)=p$ et $P(X=0)=1-p$.
- **Exemple fondamental :**
  Pour tout espace probabilisé fini (Ω, P) et pour tout événement $A \in \mathscr{P}(\Omega)$ de probabilité p : $\mathbb{1}_A \hookrightarrow \mathscr{B}(p)$.

Les indicatrices jouent un rôle naturel en probabilités que nous exploiterons davantage au chapitre « Position et dispersion d'une variable aléatoire réelle », mais qui mérite d'être compris dès maintenant. Compter les bruns dans une assemblée, c'est donner la valeur « 1 » aux bruns et la valeur « 0 » aux autres et additionner toutes ces valeurs. Plus généralement :

Toute variable aléatoire qui représente un cardinal peut être exprimée comme une somme d'indicatrices, i.e. comme une somme de variables aléatoires de loi de Bernoulli.

Exemple On lance n fois un dé à 6 faces. Si on note, pour tout $k \in [\![1,n]\!]$, A_k l'événement « On obtient 6 au $k^{\text{ème}}$ lancer » et N le nombre de 6 obtenus, alors : $N = \mathbb{1}_{A_1} + \ldots + \mathbb{1}_{A_n}$.

Théorème **(Image d'une variable aléatoire par une fonction)** Soient (Ω, P) un espace probabilisé fini, F un ensemble, X une variable aléatoire sur Ω et $f : X(\Omega) \longrightarrow F$ une fonction. La variable aléatoire $f \circ X : \Omega \longrightarrow F$ est notée plus simplement $f(X)$.

La loi $P_{f(X)}$ de $f(X)$ est entièrement déterminée par f et la loi de X. Précisément, pour tout $y \in f\big(X(\Omega)\big)$:

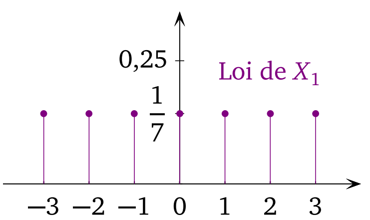
$$P\big(f(X)=y\big) = \sum_{\substack{x \in X(\Omega) \\ y=f(x)}} P(X=x).$$

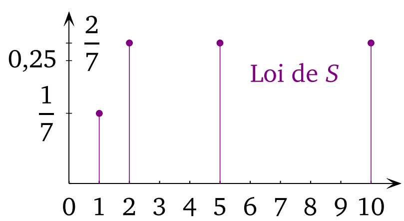
Démonstration Pour tout $y \in f\big(X(\Omega)\big)$, les éléments de l'événement $\big\{f(X)=y\big\}$ peuvent être regroupés en fonction de la valeur que X leur donne : $\big\{f(X)=y\big\} = \bigsqcup_{\substack{x \in X(\Omega) \\ y=f(x)}} \big\{X=x\big\}$. ■

Exemple On choisit un entier X au hasard entre -3 et 3. Quelle est la loi de la variable aléatoire X^2+1 ?

Démonstration Par définition : $X \hookrightarrow \mathscr{U}\big([\![-3,3]\!]\big)$, donc pour tout $k \in [\![-3,3]\!]$: $P(X=k) = \dfrac{1}{7}$.

La variable X^2+1 est quant à elle à valeurs dans $\big\{1,2,5,10\big\}$ et : $P\big(X^2+1=1\big) = P\big(X^2=0\big) = P(X=0) = \dfrac{1}{7}$ et : $P\big(X^2+1=2\big) = P\big(X^2=1\big) = P(X=1 \text{ ou } X=-1) = P(X=1)+P(X=-1) = \dfrac{1}{7}+\dfrac{1}{7} = \dfrac{2}{7}$. Par un calcul analogue : $P\big(X^2+1=5\big) = P\big(X^2+1=10\big) = \dfrac{2}{7}$.





Exemple On choisit un entier X au hasard entre 1 et $2n$. Quelle est la loi de $(-1)^X$?

Démonstration Par définition : $X \hookrightarrow \mathscr{U}\big([\![1,2n]\!]\big)$, donc pour tout $k \in [\![1,2n]\!]$: $P(X=k)=\dfrac{1}{2n}$. La variable $(-1)^X$ est quant à elle à valeurs dans $\big\{-1,1\big\}$, et le calcul qui suit montre que $(-1)^X$ suit la loi de Rademacher, i.e. la loi uniforme sur $\big\{-1,1\big\}$:

$$P\Big((-1)^X=1\Big)=\sum_{i=1}^{n}P(X=2i)=\sum_{i=1}^{n}\frac{1}{2n}=\frac{1}{2} \qquad \text{et} \qquad P\Big((-1)^X=-1\Big)=\sum_{i=0}^{n-1}P(X=2i+1)=\sum_{i=0}^{n-1}\frac{1}{2n}=\frac{1}{2}.$$

Un certain manque de rigueur ne s'est-il pas glissé hélas dans nos deux derniers exemples ? Pour modéliser une expérience aléatoire, nous commençions jusqu'ici par définir proprement un univers Ω sous-jacent, puis nous choisissions une probabilité P sur Ω. Dans les deux exemples précédents, au contraire, nous avons directement évoqué la variable aléatoire X étudiée et donné sa loi sans jamais définir le moindre espace probabilisé. Est-ce là une démarche rigoureuse ? La réponse est oui, mais le petit théorème suivant est hors programme.

Théorème (Définition implicite d'un espace probabilisé par la donnée d'une loi de variable aléatoire) Soient $E=\big\{x_1,\dots,x_r\big\}$ un ensemble et $p_1,\dots,p_r \in [0,1]$ des réels pour lesquels : $\displaystyle\sum_{i=1}^{r}p_i=1$. Il existe alors un espace probabilisé (Ω,P) et une variable aléatoire X sur Ω d'image E pour lesquels pour tout $i \in [\![1,r]\!]$: $P(X=x_i)=p_i$.

Démonstration On cherche un exemple, rien qu'un, d'espace probabilisé (Ω,P) et de variable aléatoire X qui répondent au problème posé. Eh bien, posons simplement : $\Omega=E$ et $X=\mathrm{Id}_E$, et notons P l'unique probabilité sur Ω définie pour tout $i \in [\![1,r]\!]$ par : $P\Big(\big\{x_i\big\}\Big)=p_i$. Comme voulu, pour tout $i \in [\![1,r]\!]$: $P(X=x_i)=P\Big(\big\{\omega \in E/\quad X(\omega)=x_i\big\}\Big)=P\Big(\big\{x_i\big\}\Big)=p_i$. ■

La loi d'une variable aléatoire n'est jamais que la donnée des valeurs x_i qu'elle prend, pondérées de certaines probabilités p_i. Le théorème qui précède affirme qu'il est suffisant de connaître ces données abstraitement pour qu'un contexte mathématique rigoureusement justifié leur corresponde. Ceci veut dire que quand on s'intéresse à une variable aléatoire, les couples $(x_1,p_1),\dots,(x_r,p_r)$ sont la seule chose qui compte, l'espace probabilisé (Ω,P) qui leur sert de support n'a aucune espèce d'importance. Il est voué à rester caché et rien ne garantit d'ailleurs qu'il est unique, de nombreux choix sont possibles.

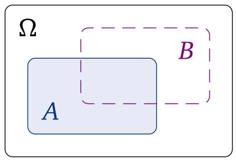
Si je lance un dé à 6 faces une fois, j'ai envie de choisir l'ensemble $[\![1,6]\!]$ pour univers Ω_1, mais si je relance le dé, l'univers $\Omega_2=[\![1,6]\!]^2$ paraît plus approprié. Toute probabilité d'un événement qui fait intervenir les deux lancers de dés doit être calculée dans le cadre de l'univers Ω_2, mais Ω_1 suffit si on s'intéresse seulement au premier lancer. Or va-t-on ainsi changer d'univers à chaque fois qu'on change de question ? Ce serait donner l'impression que nous ne sommes pas confrontés toujours à la même réalité avec une conception unique de ce qui est vraisemblable ou non.

La réponse des probabilistes à ce problème des univers multiples consiste à les négliger une bonne fois pour toutes et à ne jamais accorder le moindre regard à l'univers Ω. Sur quel Ω travaillerons-nous désormais ? Peu importe. Nous venons de voir que toute donnée intuitive d'une variable aléatoire peut être formalisée proprement, c'est suffisant. **NOUS NE DÉFINIRONS PLUS JAMAIS L'UNIVERS Ω DE NOS TRIBULATIONS PROBABILISTES.** Par exemple, si un exercice vous met dans la situation suivante : « On lance un dé à 6 faces une fois et on note X la face obtenue », vous pouvez affirmer que X est une variable aléatoire de loi uniforme sur $[\![1,6]\!]$ sans évoquer aucun espace probabilisé fini (Ω,P).

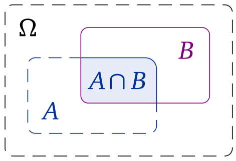
Exemple Une urne contient $3n$ boules numérotées de 1 à $2n$ avec, pour tout $k \in [\![1,n]\!]$, exactement 2 boules indiscernables de numéro $2k$ et exactement une boule de numéro $2k-1$. On tire de cette urne une boule au hasard et on note X son numéro. Il est alors clair que pour tout $k \in [\![1,n]\!]$: $P(X=2k)=\dfrac{2}{3n}$ et $P(X=2k-1)=\dfrac{1}{3n}$, et l'on a bien écrit ainsi la loi d'une variable aléatoire car : $\displaystyle\sum_{k=1}^{2n}P(X=k)=\sum_{k=1}^{n}P(X=2k)+\sum_{k=1}^{n}P(X=2k-1)=\sum_{k=1}^{n}\frac{2}{3n}+\sum_{k=1}^{n}\frac{1}{3n}=1$.

L'événement « On tire une boule de numéro pair » a pour probabilité : $\displaystyle P(X \text{ est pair})=\sum_{k=1}^{n}P(X=2k)=\sum_{k=1}^{n}\frac{2}{3n}=\frac{2}{3}$.

4 Probabilités conditionnelles



Soit (Ω, P) un espace probabilisé fini. Faisons l'hypothèse qu'un certain événement B est réalisé. Cette hypothèse modifie a priori la vraisemblance de tous les événements du contexte étudié. Par exemple, sous l'hypothèse que B est réalisé, on a bien envie de dire que B est réalisé « avec probabilité 1 » et que $\overline{B}$ l'est « avec probabilité nulle », pourtant il est faux a priori que : $P(B) = 1$ et $P(\overline{B}) = 0$. Comment tenir compte de notre hypothèse sur B ? La probabilité P mesurer la vraisemblance de tout événement, mais ceci avant toute hypothèse selon laquelle B est réalisé. Sous cette hypothèse, une autre mesure de vraisemblance doit être introduite, une autre probabilité sur Ω que nous noterons P_B et qu'on appelle la *probabilité conditionnelle sur Ω sachant B*. C'est pour cette nouvelle probabilité qu'on peut écrire : $P_B(B) = 1$ et $P_B(\overline{B}) = 0$.



À présent, comment calcule-t-on $P_B(A)$ pour un événement A quelconque ? La probabilité $P_B(A)$ mesure la vraisemblance de l'intersection $A \cap B$ au sens de la probabilité initiale P, mais si on veut que l'univers Ω lui-même ait une nouvelle vraisemblance de 1, il convient de poser : $P_B(A) = \dfrac{P(A \cap B)}{P(B)}$ sous l'hypothèse technique que : $P(B) > 0$. Cette définition fait de P_B une sorte de « probabilité d'oubli », une émanation de P aveugle à tout ce qui contredit B. La division par $P(B)$ n'est qu'une manière de normaliser P_B. Elle garantit l'égalité : $P_B(B) = 1$ selon laquelle « B est tout » — un nouvel univers apparent. Également : $P_B(\overline{B}) = 0$.

Définition (Probabilité conditionnelle) Soient (Ω, P) un espace probabilisé fini et $B \in \mathscr{P}(\Omega)$ pour lequel : $P(B) > 0$.
Pour tout $A \in \mathscr{P}(\Omega)$, le réel : $P(A|B) = P_B(A) = \dfrac{P(A \cap B)}{P(B)}$ est appelé la *probabilité conditionnelle de A sachant B*.
L'application P_B est alors une probabilité sur Ω appelée sa *probabilité conditionnelle sachant B (issue de P)*.

Démonstration Montrons que P_B est une probabilité sur Ω.

- Pour tout $A \in \mathscr{P}(\Omega)$: $P_B(A) \in [0,1]$ car par croissance de P : $0 \leqslant P(A \cap B) \leqslant P(B)$.
- Ensuite : $P_B(\Omega) = \dfrac{P(\Omega \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$.
- Enfin, pour tous $A, A' \in \mathscr{P}(\Omega)$ incompatibles, $A \cap B$ et $A' \cap B$ sont également incompatibles, donc :

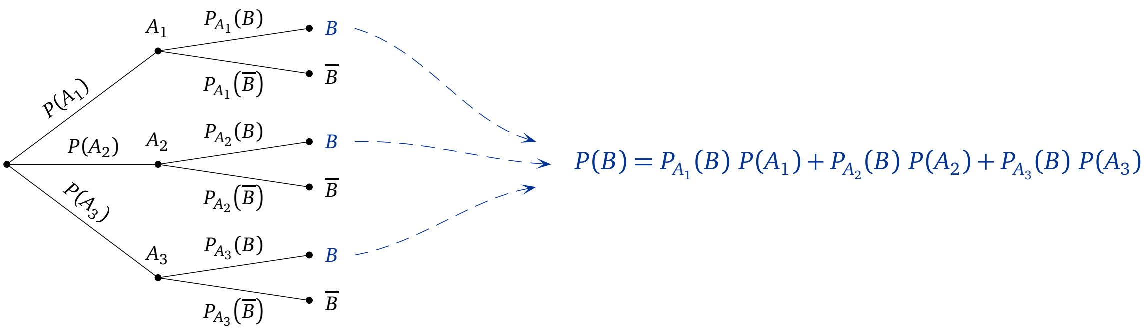
$$P_B(A \sqcup A') = \frac{P\big((A \sqcup A') \cap B\big)}{P(B)} = \frac{P\big((A \cap B) \sqcup (A' \cap B)\big)}{P(B)} = \frac{P(A \cap B) + P(A' \cap B)}{P(B)} = P_B(A) + P_B(A'). \quad \blacksquare$$

Théorème (Formule des probabilités totales) Soient (Ω, P) un espace probabilisé fini et $\{A_1, \ldots, A_n\}$ un système complet d'événements de Ω de probabilités strictement positives. Pour tout $B \in \mathscr{P}(\Omega)$: $P(B) = \sum_{i=1}^{n} P_{A_i}(B)\, P(A_i)$.

Démonstration

$$B = B \cap \Omega = B \cap \bigsqcup_{i=1}^{n} A_i = \bigsqcup_{i=1}^{n} (B \cap A_i) \quad \text{donc :} \quad P(B) = P\left(\bigsqcup_{i=1}^{n} (B \cap A_i)\right) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P_{A_i}(B)\, P(A_i). \quad \blacksquare$$

Vous avez déjà rencontré la formule des probabilités totales auparavant, mais sans doute à travers des *arbres de probabilité*, par exemple sous la forme suivante pour $n = 3$:



À partir d'aujourd'hui, vous gardez le droit de penser en termes d'arbres de probabilité si cela vous aide, mais **un arbre de probabilité ne sera plus considéré comme une preuve bien formalisée**. Pourquoi ? Parce qu'on ne peut pas résoudre des problèmes un peu sophistiqués sans maîtriser le formalisme de la formule des probabilités totales.

Exemple Dans une classe de 40 étudiants — 25 filles et 15 garçons — le professeur principal se propose de désigner brutalement deux délégués provisoires. Il prend une liste de la classe, ferme les yeux et pointe au hasard un premier nom avec la pointe de son stylo, puis de même un deuxième. Avec quelle probabilité le deuxième nom tiré est-il celui d'un garçon ?

Démonstration Notons G_1 (resp. G_2) l'événement « Le premier (resp. deuxième) nom tiré est celui d'un garçon ». Nous cherchons $P(G_2)$. De quelles informations disposons-nous ?

$$P(G_1) = \frac{15}{40} = \frac{3}{8}, \qquad P\left(\overline{G_1}\right) = \frac{25}{40} = \frac{5}{8}, \qquad P_{G_1}(G_2) = \frac{14}{39} \qquad \text{et} \qquad P_{\overline{G_1}}(G_2) = \frac{15}{39}.$$

Or $\left\{G_1, \overline{G_1}\right\}$ est un système complet d'événements de probabilités strictement positives, donc d'après la formule des probabilités totales : $P(G_2) = P_{G_1}(G_2)\, P(G_1) + P_{\overline{G_1}}(G_2)\, P\left(\overline{G_1}\right) = \dfrac{14}{39} \times \dfrac{3}{8} + \dfrac{15}{39} \times \dfrac{5}{8} = \dfrac{3}{8}$.

Cet exemple diffère sensiblement de ceux qui l'ont précédé. Alors que nous avons tué plus haut les univers Ω au profit des seules variables aléatoires, nous n'avons pas cette fois introduit de variables aléatoires, mais seulement des événements G_1 et G_2 et certaines probabilités, conditionnelles ou non, qui leur sont associées. De nouveau, il doit bien y avoir un espace probabilisé (Ω, P) quelque part, mais il reste dans l'ombre.

Théorème (Formules de Bayes) Soient (Ω, P) un espace probabilisé fini et $B \in \mathscr{P}(\Omega)$ pour lequel : $P(B) > 0$.

(i) Pour tout $A \in \mathscr{P}(\Omega)$, si : $P(A) > 0$, alors : $P_B(A) = \dfrac{P_A(B)\, P(A)}{P(B)}$.

(ii) Soit $\left\{A_1, \ldots, A_n\right\}$ un système complet d'événements de Ω de probabilités strictement positives.

Pour tout $j \in [\![1, n]\!]$: $P_B(A_j) = \dfrac{P_{A_j}(B)\, P(A_j)}{\displaystyle\sum_{i=1}^{n} P_{A_i}(B)\, P(A_i)}$.

Démonstration Pour l'assertion (i) : $P_B(A) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P_A(B)\, P(A)}{P(B)}$. L'assertion (ii) n'est quant à elle qu'un mélange peu subtil d'assertion (i) et de formule des probabilités totales. ■

Dans l'assertion (ii), pour tout $j \in [\![1, n]\!]$, $P(A_j)$ est souvent qualifiée de *probabilité a priori* et $P_B(A_j)$ de *probabilité a posteriori*. Avec cette terminologie, la formule de Bayes permet le passage de probabilités a priori à des probabilités a posteriori. Ce qu'il faut ici comprendre, c'est qu'une probabilité n'est pas tant une mesure du hasard dans une situation donnée qu'une mesure de l'information dont on dispose sur cette situation. Si on connaît les probabilités a priori $P(A_1), \ldots, P(A_n)$ et si tout à coup on apprend que l'événement B est réalisé, la formule de Bayes nous permet d'actualiser notre perception des événements $A_1, \ldots, A_n$ en nous donnant accès aux probabilités conditionnelles $P_B(A_1), \ldots, P_B(A_n)$ — connaissance plus fine puisque le conditionnement restreint le champ des possibles. La formule de Bayes est une formule de *révision des croyances*.

Exemple Juge au tribunal, je dois juger de la culpabilité d'une compagnie de taxis bleus. Un soir de brouillard, un taxi a percuté un piéton qui traversait la rue dans son bon droit, puis a pris la fuite. Un témoin affirme que le taxi était bleu et c'est sur la base de ce témoignage que le procès a été instruit. Or dans la ville, deux compagnies de taxis se partagent le marché, la compagnie des taxis bleus et la compagnie des taxis verts, mais tout de même les taxis verts dominent le marché au sens où 90% des taxis dans la ville sont verts.

On demande au témoin d'effectuer des tests de reconnaissance des couleurs pour mesurer la fiabilité de son témoignage. Il s'avère qu'il est fiable dans 90% des cas pour la couleur bleue et 80% des cas pour la couleur verte.

Dois-je condamner ou non la compagnie des taxis bleus ?

Démonstration Notons B l'événement « Le taxi coupable est bleu » et T_B l'événement « Le témoin croit savoir que le taxi coupable est bleu ». Je déciderai du sort de la compagnie des taxis bleus après avoir calculé $P_{T_B}(B)$ (probabilité a posteriori).

- De quelles informations disposons-nous ?

$$P\left(\overline{B}\right) = \frac{90}{100} = \frac{9}{10} \quad \text{et} \quad P(B) = \frac{1}{10} \quad \text{(probabilités a priori)} \quad \text{et} \quad P_B(T_B) = \frac{90}{100} = \frac{9}{10} \quad \text{et} \quad P_{\overline{B}}\left(\overline{T_B}\right) = \frac{80}{100} = \frac{8}{10}.$$

L'égalité : $P_{\overline{B}}\left(\overline{T_B}\right) = \dfrac{80}{100}$ repose sur l'hypothèse que le témoin ne pouvait hésiter qu'entre le vert et le bleu puisque tous les taxis sont verts ou bleus. Il en découle que : $P_{\overline{B}}(T_B) = 1 - P_{\overline{B}}\left(\overline{T_B}\right) = \dfrac{2}{10}$.

Finalement, d'après la formule de Bayes utilisée avec le système complet d'événements $\left\{B, \overline{B}\right\}$:

$$P_{T_B}(B) = \frac{P_B(T_B)\, P(B)}{P_B(T_B)\, P(B) + P_{\overline{B}}(T_B)\, P\left(\overline{B}\right)} = \frac{\frac{9}{10} \times \frac{1}{10}}{\frac{9}{10} \times \frac{1}{10} + \frac{2}{10} \times \frac{9}{10}} = \frac{1}{3}.$$

Conclusion : le témoignage du témoin est invalidé, je peux blanchir la compagnie des taxis bleus.

- Une remarque pour finir. Si le témoin avait été fiable à 95% pour la couleur verte au lieu de 80%, on aurait obtenu : $P_{T_B}(B) = \frac{2}{3}$ et là, j'aurais été bien embêté pour rendre mon jugement !

Théorème (Formule des probabilités composées) Soient (Ω, P) un espace probabilisé fini et $A_1, \ldots, A_n \in \mathscr{P}(\Omega)$ pour lesquels : $P(A_1 \cap \ldots \cap A_{n-1}) > 0$. Alors : $P(A_1 \cap \ldots \cap A_n) = P(A_1)\, P_{A_1}(A_2)\, P_{A_1 \cap A_2}(A_3) \, \ldots \, P_{A_1 \cap \ldots \cap A_{n-1}}(A_n)$.

Démonstration Comme : $(A_1 \cap \ldots \cap A_{n-1}) \subset (A_1 \cap \ldots \cap A_i)$ pour tout $i \in [\![1, n-1]\!]$ et : $P(A_1 \cap \ldots \cap A_{n-1}) > 0$, alors par croissance de P : $P(A_1 \cap \ldots \cap A_i) > 0$. Finalement, grâce à une simplification télescopique :

$$P(A_1)\, P_{A_1}(A_2) \, \ldots \, P_{A_1 \cap \ldots \cap A_{n-1}}(A_n) = P(A_1) \prod_{k=2}^{n} P_{A_1 \cap \ldots \cap A_{k-1}}(A_k) = P(A_1) \prod_{k=2}^{n} \frac{P(A_1 \cap \ldots \cap A_k)}{P(A_1 \cap \ldots \cap A_{k-1})} = P(A_1 \cap \ldots \cap A_n). \quad \blacksquare$$

Exemple J'ai 4 pièces de 1€ et 3 de 2€ en poche et je veux me faire un petit plaisir aujourd'hui. Pour ne pas tout dépenser, je tire 3 pièces au hasard de ce pactole, successivement. Avec quelle probabilité sortent-elles dans l'ordre : 2€ / 1€ / 2€ ?

Démonstration Pour tout $i \in [\![1, 3]\!]$, notons D_i l'événement « La $i^{\text{ème}}$ pièce tirée est une pièce de 2€ ». Nous cherchons $P\left(D_1 \cap \overline{D_2} \cap D_3\right)$. D'après la formule des probabilités composées :

$$P\left(D_1 \cap \overline{D_2} \cap D_3\right) = P(D_1)\, P_{D_1}\left(\overline{D_2}\right) P_{D_1 \cap \overline{D_2}}(D_3) = \frac{3}{7} \times \frac{4}{6} \times \frac{2}{5} = \frac{4}{35} \approx \frac{5}{35} = \frac{1}{7}.$$

Exemple Un commerçant met en vente 50 tickets d'un certain jeu dont exactement 3 sont gagnants. Je lui achète 6 tickets. Avec quelle probabilité en ai-je acheté au moins un gagnant ?

Démonstration Nous noterons G l'événement « L'un au moins des tickets achetés est gagnant », d'événement contraire $\overline{G}$ « Tous les tickets achetés sont perdants ». Nous allons calculer $P(G)$ grâce à $P\left(\overline{G}\right)$.

- **Première preuve à l'aide d'une variable aléatoire :** On peut numéroter sans perte de généralité les tickets gagnants 1, 2 et 3 et les tickets perdants de 4 à 50. Notons alors X la 6-combinaison de $[\![1, 50]\!]$ des tickets achetés. Clairement : $X \hookrightarrow \mathscr{U}\left(\mathscr{P}_6\left([\![1, 50]\!]\right)\right)$ et $\overline{G} = \left\{X \in \mathscr{P}_6\left([\![4, 50]\!]\right)\right\}$, donc :

$$P\left(\overline{G}\right) = \frac{\left|\mathscr{P}_6\left([\![4, 50]\!]\right)\right|}{\left|\mathscr{P}_6\left([\![1, 50]\!]\right)\right|} = \frac{\binom{47}{6}}{\binom{50}{6}} = \frac{47 \times 46 \times 45 \times 44 \times 43 \times 42}{50 \times 49 \times 48 \times 47 \times 46 \times 45} = \frac{44 \times 43 \times 42}{50 \times 49 \times 48} = \frac{473}{700}, \qquad \text{donc : } \quad P(G) = \frac{227}{700}.$$

- **Deuxième preuve sans variable aléatoire :** On peut considérer sans perte de généralité que les 6 tickets achetés l'ont été dans un certain ordre et pour tout $i \in [\![1, 6]\!]$, noter T_i l'événement « Le $i^{\text{ème}}$ ticket acheté est perdant ». Dans ces conditions : $P_{T_1 \cap \ldots \cap T_{i-1}}(T_i) = \frac{48 - i}{51 - i}$ pour tout $i \in [\![1, 6]\!]$, car sous l'hypothèse que les $i - 1$ premiers tickets achetés ont été perdants, il reste au commerçant $51 - i$ tickets à vendre dont $48 - i$ perdants. Finalement, d'après la formule des probabilités composées :

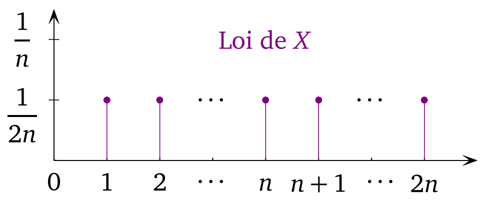
$$P\left(\overline{G}\right) = P(T_1 \cap \ldots \cap T_n) = P(T_1)\, P_{T_1}(T_2) \, \ldots \, P_{T_1 \cap \ldots \cap T_{n-1}}(T_n) = \prod_{i=1}^{6} \frac{48 - i}{51 - i}$$
$$= \frac{47 \times 46 \times 45 \times 44 \times 43 \times 42}{50 \times 49 \times 48 \times 47 \times 46 \times 45} = \frac{473}{700}, \qquad \text{donc : } \quad P(G) = \frac{227}{700}.$$

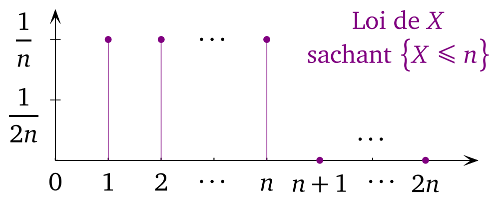
Nous achèverons ce paragraphe par un cas particulier de conditionnement appliqué au concept de variable aléatoire.

Définition (Lois conditionnelles d'une variable aléatoire) Soient (Ω, P) un espace probabilisé fini, X une variable aléatoire sur Ω et $A \in \mathscr{P}(\Omega)$ pour lequel : $P(A) > 0$. On appelle *loi conditionnelle de X sachant A* l'application $x \longmapsto P_A(X = x)$ de $X(\Omega)$ dans $[0, 1]$.

Exemple On choisit un entier X au hasard entre 1 et $2n$. Quelle est la loi de X sachant $\{X \leqslant n\}$?

Démonstration Par définition : $X \hookrightarrow \mathscr{U}([\![1,2n]\!])$, donc pour tout $k \in [\![1,2n]\!]$: $P(X=k) = \dfrac{1}{2n}$. Or : $P(X \leqslant n) = \displaystyle\sum_{k=1}^{n} P(X=k) = \sum_{k=1}^{n} \frac{1}{2n} = \frac{1}{2}$, donc pour tout $k \in [\![1,n]\!]$: $P_{\{X \leqslant n\}}(X=k) = \dfrac{P(X=k)}{P(X \leqslant n)} = \dfrac{1}{n}$ et pour tout $k \in [\![n+1,2n]\!]$: $P_{\{X \leqslant n\}}(X=k) = 0$. La loi conditionnelle de X sachant $\{X \leqslant n\}$ est ainsi la loi uniforme sur $[\![1,n]\!]$.





5 Indépendance

Intuitivement, étant donnés deux événements A et B pour lesquels : $P(B) > 0$, on a envie de dire que A et B sont *indépendants* si la probabilité $P(A)$ ne dépend pas de la réalisation de B, i.e. si : $P_B(A) = P(A)$, ce qui s'écrit aussi : $P(A \cap B) = P(A)\,P(B)$. Cette remarque conduit à la définition suivante.

Définition (Événements indépendants) Soit (Ω, P) un espace probabilisé fini.

- Soient $A, B \in \mathscr{P}(\Omega)$ des événements. On dit que A et B sont *indépendants* si : $P(A \cap B) = P(A)\,P(B)$.
- Soient $A_1, \ldots, A_n \in \mathscr{P}(\Omega)$ des événements. On dit que $A_1, \ldots, A_n$ sont (*mutuellement*) *indépendants* si pour toute partie I de $[\![1,n]\!]$: $P\left(\displaystyle\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$.

✖ Attention ! ✖ $\boxed{\text{(Mutuellement) indépendants} \implies \text{Deux à deux indépendants}}$ mais **la réciproque est fausse !** Posons par exemple : $\Omega = [\![1,4]\!]$ et notons P la probabilité uniforme sur Ω. Posons en outre : $A = \{1,2\}$, $B = \{1,3\}$ et $C = \{2,3\}$. Alors : $P(A) = P(B) = P(C) = \dfrac{1}{2}$ et $P(A \cap B) = P(A \cap C) = P(B \cap C) = \dfrac{1}{4}$, donc A, B et C sont deux à deux indépendants. Pourtant : $P(A \cap B \cap C) = P(\varnothing) = 0 \neq \dfrac{1}{8} = P(A)\,P(B)\,P(C)$, donc A, B et C ne sont pas (mutuellement) indépendants.

Théorème (Indépendance et complémentaires) Soient (Ω, P) un espace probabilisé fini et $A_1, \ldots, A_n \in \mathscr{P}(\Omega)$. Si $A_1, \ldots, A_n$ sont indépendants, les événements $A_1^\circ, \ldots, A_n^\circ$ le sont aussi pour tous $A_1^\circ \in \{A_1, \overline{A_1}\}, \ldots, A_n^\circ \in \{A_n, \overline{A_n}\}$.

Démonstration Nous nous contenterons du cas de deux événements A et B.

- Si A et B sont indépendants, $\overline{A}$ et B le sont aussi car :

$$P\big(\overline{A} \cap B\big) = P(B \setminus A) = P(B) - P(A \cap B) \overset{\text{Indép.}}{=} P(B) - P(A)\,P(B) = \big(1 - P(A)\big)\,P(B) = P\big(\overline{A}\big)\,P(B).$$

- Quitte à permuter A et B, le premier point montre que si A et B sont indépendants, A et $\overline{B}$ le sont aussi, et donc a fortiori $\overline{A}$ et $\overline{B}$ aussi, toujours d'après le premier point. ■

Le concept d'indépendance va maintenant nous permet de définir une loi usuelle très importante en pratique — la *loi binomiale*. Intéressons-nous pour cela à la répétition n fois indépendamment d'une expérience aléatoire à deux issues, disons « favorable » et « défavorable », de probabilité p pour l'issue favorable. Quelle est loi du nombre X d'issues favorables ? Pour tout $i \in [\![1,n]\!]$, notons F_i l'événement « La $i^{\text{ème}}$ expérience a une issue favorable ». Les événements $F_1, \ldots, F_n$ sont indépendants par hypothèse. La variable aléatoire X est à valeurs dans $[\![0,n]\!]$ et pour tout $k \in [\![0,n]\!]$, l'événement $\{X=k\}$ est réalisé si et seulement si on a obtenu k issues favorables et $n-k$ issues défavorables. Ainsi, si on note $\mathscr{P}_k\big([\![1,n]\!]\big)$ l'ensemble des parties de $[\![1,n]\!]$ de cardinal k : $\quad \{X=k\} = \bigsqcup_{I\in\mathscr{P}_k([\![1,n]\!])} \left(\bigcap_{i\in I} F_i \cap \bigcap_{j\in[\![1,n]\!]\setminus I} \overline{F_j}\right)$. Aussitôt :

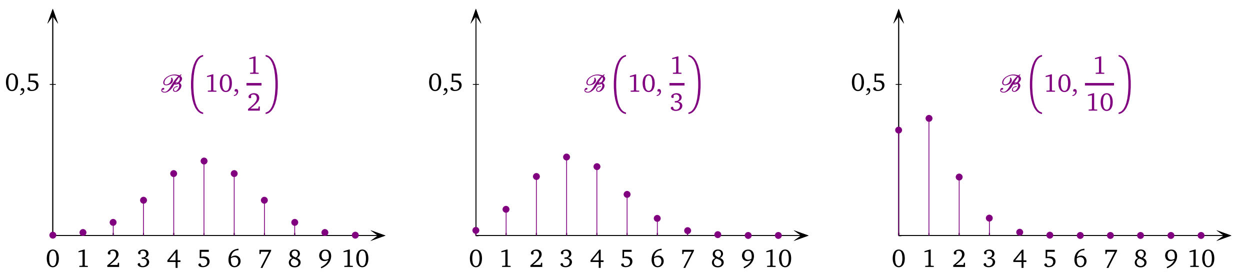
$$\begin{aligned} P(X=k) &= \sum_{I\in\mathscr{P}_k([\![1,n]\!])} P\left(\bigcap_{i\in I} F_i \cap \bigcap_{j\in[\![1,n]\!]\setminus I} \overline{F_j}\right) \overset{\text{Indép.}}{=} \sum_{I\in\mathscr{P}_k([\![1,n]\!])} \prod_{i\in I} P(F_i) \prod_{j\in[\![1,n]\!]\setminus I} P\big(\overline{F_j}\big) = \sum_{I\in\mathscr{P}_k([\![1,n]\!])} \prod_{i\in I} p \prod_{j\in[\![1,n]\!]\setminus I} (1-p) \\ &= \sum_{I\in\mathscr{P}_k([\![1,n]\!])} p^k(1-p)^{n-k} = \binom{n}{k} p^k(1-p)^{n-k}. \end{aligned}$$

Prenez le temps d'observer à quel point l'indépendance des événements $F_1,\dots,F_n$ a été cruciale dans ce calcul. Sans indépendance, on n'aurait pas su avancer.

A posteriori, la définition qui suit est légitimée par la formule du binôme car : $\displaystyle\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k} = \big(p+(1-p)\big)^n = 1$.

Définition **(Loi binomiale ou loi des tirages AVEC remise)** Soient X une variable aléatoire et $p \in [0,1]$. On dit que X *suit la loi binomiale de paramètre* (n,p), ce qu'on note : $\quad X \hookrightarrow \mathscr{B}(n,p)$, $\quad$ si pour tout $k \in [\![0,n]\!]$:

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



Lorsqu'on répète n fois indépendamment une expérience aléatoire à deux issues de probabilité p pour l'issue favorable, le **NOMBRE D'ISSUES FAVORABLES** suit la loi binomiale $\mathscr{B}(n,p)$.

L'expression « loi des tirages **AVEC** remise » par laquelle on décrit parfois la loi binomiale trouve son explication dans l'exemple suivant. Lorsqu'on tire **AVEC REMISE** — donc indépendamment — n boules dans une urne contenant des boules blanches en proportion p et des boules noires en proportion $1-p$, la variable aléatoire « nombre de boules blanches tirées » suit la loi binomiale $\mathscr{B}(n,p)$.

Enfin, clairement, la loi binomiale $\mathscr{B}(1,p)$ n'est autre que la loi de Bernoulli $\mathscr{B}(p)$!

Exemple On lance 5 fois un dé équilibré à 6 faces dont 2 blanches et 4 noires. Avec quelle probabilité obtient-on exactement 3 fois une face noire ?

Démonstration La face noire a une probabilité $\dfrac{4}{6} = \dfrac{2}{3}$ d'apparaître quand on lance le dé une fois. Par indépendance des lancers, le nombre N de faces noires obtenues à l'issue des 5 lancers suit donc la loi binomiale $\mathscr{B}\left(5,\dfrac{2}{3}\right)$, et donc enfin : $\quad P(N=3) = \dbinom{5}{3}\left(\dfrac{2}{3}\right)^3\left(1-\dfrac{2}{3}\right)^{5-3} = \dfrac{80}{243} \approx 0,33.$

On adapte à présent aux variables aléatoires le concept d'indépendance.

Définition **(Couple de variables aléatoires indépendantes)** Soient (Ω,P) un espace probabilisé fini et X et Y deux variables aléatoires sur Ω. On dit que X et Y sont *indépendantes* si pour tous $x \in X(\Omega)$ et $y \in Y(\Omega)$, les événements $\{X=x\}$ et $\{Y=y\}$ sont indépendants, i.e. :

$$P(X=x \text{ et } Y=y) \;=\; P(X=x)\,P(Y=y).$$

Théorème (Une propriété des couples de variables aléatoires indépendantes) Soient (Ω, P) un espace probabilisé fini et X et Y deux variables aléatoires sur Ω. Si X et Y sont indépendantes, alors pour toutes parties A de $X(\Omega)$ et B de $Y(\Omega)$, les événements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants, i.e. :

$$P(X \in A \text{ et } Y \in B) \;=\; P(X \in A)\, P(Y \in B).$$

Démonstration Pour tous $A \in \mathscr{P}\big(X(\Omega)\big)$ et $B \in \mathscr{P}\big(Y(\Omega)\big)$:

$$P(X \in A)\, P(Y \in B) = P\left(\bigsqcup_{x \in A} \{X = x\}\right) P\left(\bigsqcup_{y \in B} \{Y = y\}\right) = \left(\sum_{x \in A} P(X = x)\right)\left(\sum_{y \in B} P(Y = y)\right) = \sum_{(x,y) \in A \times B} P(X = x)\, P(Y = y)$$

$$\overset{\text{Indép.}}{=} \sum_{(x,y) \in A \times B} P(X = x \text{ et } Y = y) = P\left(\bigsqcup_{(x,y) \in A \times B} \{X = x \text{ et } Y = y\}\right) = P(X \in A \text{ et } Y \in B). \qquad \blacksquare$$

Exemple On lance un dé à 6 faces 2 fois et on note X_1 (resp. X_2) la face obtenue au premier (resp. second) lancer. Avec quelle probabilité obtient-on les deux fois une face impaire ? Les variables aléatoires X_1 et X_2 sont indépendantes par définition et suivent toutes deux la loi uniforme sur $[\![1,6]\!]$, donc :

$$P\big(X_1 \text{ et } X_2 \text{ sont impairs}\big) \overset{\text{Indép.}}{=} P(X_1 \text{ est impair})\, P(X_2 \text{ est impair}) = \frac{\big|\{1,3,5\}\big|}{6} \times \frac{\big|\{1,3,5\}\big|}{6} = \frac{1}{4}.$$

Définition-théorème (Ensemble fini de variables aléatoires mutuellement indépendantes) Soient (Ω, P) un espace probabilisé fini et $X_1, \ldots, X_n$ des variables aléatoires sur Ω.

- On dit que $X_1, \ldots, X_n$ sont (*mutuellement*) *indépendantes* si pour tous $x_1 \in X_1(\Omega), \ldots, x_n \in X_n(\Omega)$:

  $$P(X_1 = x_1 \text{ et } \ldots \text{ et } X_n = x_n) \;=\; P(X_1 = x_1) \;\ldots\; P(X_n = x_n).$$

  On peut montrer qu'il est équivalent d'exiger l'indépendance des événements $\{X = x_1\}, \ldots, \{X_n = x_n\}$ pour tous $x_1 \in X_1(\Omega), \ldots, x_n \in X_n(\Omega)$.
- On peut montrer également que si $X_1, \ldots, X_n$ sont (mutuellement) indépendantes, alors pour toutes parties A_1 de $X_1(\Omega)$… et A_n de $X_n(\Omega)$, les événements $\{X_1 \in A_1\}, \ldots, \{X_n \in A_n\}$ sont (mutuellement) indépendants. En particulier :

  $$P(X_1 \in A_1 \text{ et } \ldots \text{ et } X_n \in A_n) \;=\; P(X_1 \in A_1) \;\ldots\; P(X_n \in A_n).$$

✖ **Attention !** ✖ $\boxed{\text{(Mutuellement) indépendantes} \implies \text{DEUX À DEUX indépendantes}}$ mais **LA RÉCIPROQUE EST FAUSSE !**

Exemple Soient $p_1, \ldots, p_n \in [0,1]$ et $X_1, \ldots, X_n$ des variables aléatoires indépendantes de lois respectives $\mathscr{B}(p_1), \ldots, \mathscr{B}(p_n)$.
Alors : $X_1 \ldots X_n \hookrightarrow \mathscr{B}\big(p_1 \ldots p_n\big)$.

Démonstration La variable aléatoire $X_1 \ldots X_n$ est à valeurs dans $\{0,1\}$ et :

$$P(X_1 \ldots X_n = 1) = P(X_1 = 1 \text{ et } \ldots \text{ et } X_n = 1) \overset{\text{Indép.}}{=} P(X_1 = 1) \;\ldots\; P(X_n = 1) = p_1 \ldots p_n$$

et : $P(X_1 \ldots X_n = 0) = 1 - P(X_1 \ldots X_n = 1) = 1 - p_1 \ldots p_n$.

Exemple Un jeu de 32 cartes a été malicieusement truqué, on y a remplacé une carte autre que l'as de pique par un deuxième as de pique. On répète n fois avec remise l'expérience consistant à tirer simultanément 4 cartes. À partir de quelle valeur de n la probabilité de déceler la supercherie est-elle supérieure ou égale à 0,9 ?

Démonstration On peut numéroter sans perte de généralité les 2 as de pique 1 et 2 et numéroter les autres cartes une bonne fois pour toutes de 3 à 32. Pour tout $k \in [\![1,n]\!]$, notons X_k la 4-combinaison des cartes tirées au $k^{\text{ème}}$ tirage. Ces variables suivent toutes la loi uniforme sur $\mathscr{P}_4\big([\![1,32]\!]\big)$ et sont de plus indépendantes car les tirages sont sans remise.

- Calculons d'abord pour tout $k \in [\![1,n]\!]$ la probabilité de déceler la supercherie au $k^{\text{ème}}$ tirage, i.e. de l'événement $A_k = \{1 \in X_k \text{ et } 2 \in X_k\}$. Or il y a autant de tirages qui réalisent A_k que de 2-combinaisons de l'ensemble $[\![3,32]\!]$ des cartes qui ne sont pas les as de pique, donc : $\quad P(A_k) = \dfrac{\binom{30}{2}}{\binom{32}{4}} = \dfrac{3}{248}$.
- Notons à présent B l'événement « La supercherie finit par être décelée » d'événement contraire « La supercherie n'est jamais décelée » : $\quad \overline{B} = \overline{A_1} \cap \ldots \cap \overline{A_n}$. $\quad$ Parce que les variables aléatoires $X_1, \ldots, X_n$ le sont, les événements $A_1, \ldots, A_n$ sont indépendants, donc les événements $\overline{A_1}, \ldots, \overline{A_n}$ aussi, et du coup :
$$P(B) = 1 - P(\overline{B}) = 1 - P\big(\overline{A_1} \cap \ldots \cap \overline{A_n}\big) = 1 - P\big(\overline{A_1}\big) \ \ldots \ P\big(\overline{A_n}\big) = 1 - \left(1 - \frac{3}{248}\right)^n = 1 - \left(\frac{245}{248}\right)^n.$$
Finalement : $\quad P(B) \geqslant 0,9 \quad \Longleftrightarrow \quad \left(\dfrac{245}{248}\right)^n \leqslant \dfrac{1}{10} \quad \Longleftrightarrow \quad n \geqslant \dfrac{\ln 10}{\ln \dfrac{248}{245}} \quad \Longleftrightarrow \quad n \geqslant 190.$

Théorème (Indépendance des images de variables aléatoires indépendantes par des fonctions) Soient (Ω, P) un espace probabilisé fini, E et F deux ensembles, $X_1, \ldots, X_n$ des variables aléatoires sur Ω et $f : (X_1, \ldots, X_m)(\Omega) \longrightarrow E$ et $g : (X_{m+1}, \ldots, X_n)(\Omega) \longrightarrow F$ deux fonctions. Si $X_1, \ldots, X_n$ sont indépendantes, alors $f(X_1, \ldots, X_m)$ et $g(X_{m+1}, \ldots, X_n)$ le sont aussi.

L'énoncé paraît compliqué, mais il dit simplement, par exemple, que si trois variables aléatoires X, Y et Z définies sur un même espace probabilisé fini sont indépendantes, les variables aléatoires $X + Y$ et Z le sont aussi.

Démonstration Nous nous contenterons d'établir un cas particulier de ce résultat. Soient X et Y deux variables aléatoires indépendantes sur Ω et $f : X(\Omega) \longrightarrow F$ et $g : Y(\Omega) \longrightarrow G$ deux fonctions. Montrons que $f(X)$ et $g(Y)$ sont également indépendantes. Or pour tous $a \in f(X)(\Omega)$ et $b \in g(Y)(\Omega)$:
$$\begin{aligned} P\big(f(X) = a \text{ et } g(Y) = b\big) &= P\Big(X \in f^{-1}\big(\{a\}\big) \text{ et } Y \in g^{-1}\big(\{b\}\big)\Big) \overset{\text{Indép.}}{=} P\Big(X \in f^{-1}\big(\{a\}\big)\Big)\, P\Big(Y \in g^{-1}\big(\{b\}\big)\Big) \\ &= P\big(f(X) = a\big)\, P\big(g(Y) = b\big). \end{aligned}$$
■

6 Loi d'un couple de variables aléatoires

Nous parlerons surtout que de **couples** de variables aléatoires dans ce paragraphe, mais les définitions et résultats présentés se généralisent sans difficulté aux familles d'un nombre fini quelconque de variables aléatoires.

Remarquons pour commencer que les couples de variables aléatoires ne sont pas une notion nouvelle. Nous avons autorisé nos variables aléatoires à prendre leurs valeurs dans un ensemble quelconque, un couple (X, Y) de variables aléatoires n'est donc jamais qu'une variable aléatoire à valeurs dans un ensemble produit.

Définition-théorème (Couple de variables aléatoires, loi conjointe et lois marginales) Soient (Ω, P) un espace probabilisé fini et X et Y deux variables aléatoires sur Ω.

- La variable aléatoire $\left\{\begin{array}{ccc} \Omega & \longrightarrow & X(\Omega) \times Y(\Omega) \\ \omega & \longmapsto & \big(X(\omega), Y(\omega)\big) \end{array}\right.$ est notée simplement (X, Y) et sa loi $P_{(X,Y)}$ est souvent appelée la *loi conjointe de X et Y*.
- La loi P_X de X est appelée la *première loi marginale de* (X, Y) et la loi P_Y de Y sa *deuxième loi marginale*.
- Les événements $\{X = x \text{ et } Y = y\}$, (x, y) décrivant $(X, Y)(\Omega)$ ou $X(\Omega) \times Y(\Omega)$, forment un système complet d'événements appelé le *système complet d'événements associé à* (X, Y).

Le réel $P(X = x \text{ et } Y = y)$ ne dépend pas que des nombres $P(X = x)$ et $P(Y = y)$, i.e. des variables X et Y isolément, il dépend aussi du lien éventuellement étroit qui unit ces variables. En d'autres termes, on ne connaît pas la loi conjointe du couple (X, Y) quand on connaît séparément ses lois marginales.

L'indépendance de X et Y est la condition simple et exceptionnelle grâce à laquelle, en pratique, on arrive à relier les réels $P(X = x \text{ et } Y = y)$, $P(X = x)$ et $P(Y = y)$. En cas d'indépendance, la loi conjointe se déduit des lois marginales.

Exemple Une urne contient 3 boules blanches et 1 noire et on en tire successivement 2 boules sans remise. Le couple de couleurs ainsi tirées est noté (C_1, C_2) — avec « B » pour les blanches et « N » pour l'unique noire. Quelle est la loi de (C_1, C_2) ?

Démonstration $P(C_1 = \mathrm{B} \text{ et } C_2 = \mathrm{B}) = P(C_1 = \mathrm{B})\ P_{\{C_1=\mathrm{B}\}}(C_2 = \mathrm{B}) = \dfrac{3}{4} \times \dfrac{2}{3} = \dfrac{1}{2}$

et de même : $P(C_1 = \mathrm{B} \text{ et } C_2 = \mathrm{N}) = \dfrac{3}{4} \times \dfrac{1}{3} = \dfrac{1}{4}$,

$P(C_1 = \mathrm{N} \text{ et } C_2 = \mathrm{B}) = \dfrac{1}{4} \times \dfrac{3}{3} = \dfrac{1}{4}$ et $P(C_1 = \mathrm{N} \text{ et } C_2 = \mathrm{N}) = \dfrac{1}{4} \times \dfrac{0}{3} = 0.$

$C_1 \backslash C_2$	B	N
B	$\frac{1}{2}$	$\frac{1}{4}$
N	$\frac{1}{4}$	0

À défaut de pouvoir calculer en général la loi conjointe à partir des lois marginales, on peut au moins toujours faire le calcul inverse.

Théorème **(Calcul des lois marginales à partir de la loi conjointe)** Soient X et Y deux variables aléatoires sur Ω. La loi (conjointe) $P_{(X,Y)}$ du couple (X,Y) détermine entièrement ses lois marginales P_X et P_Y. Précisément, pour tout $x \in X(\Omega)$: $\displaystyle P(X = x) = \sum_{y \in Y(\Omega)} P(X = x \text{ et } Y = y)$ et pour tout $y \in Y(\Omega)$: $\displaystyle P(Y = y) = \sum_{x \in X(\Omega)} P(X = x \text{ et } Y = y)$.

Démonstration Pour tout $x \in X(\Omega)$:

$$P(X = x) = P\left(\{X = x\} \cap \bigsqcup_{y \in Y(\Omega)} \{Y = y\}\right) = P\left(\bigsqcup_{y \in Y(\Omega)} \{X = x \text{ et } Y = y\}\right) = \sum_{y \in Y(\Omega)} P(X = x \text{ et } Y = y). \qquad \blacksquare$$

On représente parfois la loi de (X,Y) comme un tableau à deux entrées donnant $P(X = x \text{ et } Y = y)$ en fonction des valeurs possibles de $x \in X(\Omega)$ en lignes et $y \in Y(\Omega)$ en colonnes. La figure ci-contre illustre la manière dont on peut calculer P_X et P_Y à partir de $P_{(X,Y)}$. On a noté : $X(\Omega) = \{x_1, \ldots, x_m\}$, $Y(\Omega) = \{y_1, \ldots, y_n\}$ et pour tout $(i,j) \in [\![1,m]\!] \times [\![1,n]\!]$:

$$p_{ij} = P(X = x_i \text{ et } Y = y_j).$$

$X \backslash Y$	y_1	y_2	$\cdots$	y_n	
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	$\longrightarrow \displaystyle\sum_{j=1}^n p_{1j} = P(X = x_1)$
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$\longrightarrow \displaystyle\sum_{j=1}^n p_{2j} = P(X = x_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	$\cdots$	p_{mn}	$\longrightarrow \displaystyle\sum_{j=1}^n p_{mj} = P(X = x_m)$

Première loi marginale P_X de (X,Y)

$$P(Y = y_1) = \sum_{i=1}^m p_{i1} \qquad \cdots \qquad \sum_{i=1}^m p_{in} = P(Y = y_n)$$

Deuxième loi marginale P_Y de (X,Y)

Le théorème qui suit est hors programme mais très naturel, et à mon avis éclairant. De quoi s'agit-il ? Pour modéliser le lancer d'un dé à 6 faces 2 fois, on peut aussi bien se donner deux variables aléatoires X_1 et X_2 indépendantes de loi uniforme sur $[\![1,6]\!]$ que se donner un couple (X_1, X_2) de loi uniforme sur $[\![1,6]\!]^2$, c'est pareil !

Théorème **(Loi uniforme et produit cartésien)** Soient E et F deux ensembles finis et X et Y deux variables aléatoires définies sur un même espace probabilisé fini. Les assertions suivantes sont équivalentes :

(i) $(X,Y) \hookrightarrow \mathcal{U}(E \times F)$.

(ii) X et Y sont indépendantes et : $X \hookrightarrow \mathcal{U}(E)$ et $Y \hookrightarrow \mathcal{U}(F)$.

Cet énoncé se généralise sans difficulté au cas d'un nombre fini de variables aléatoires.

Démonstration

(i) $\Longrightarrow$ (ii) Pour tout $x \in E$: $\displaystyle P(X = x) = \sum_{y \in F} P(X = x \text{ et } Y = y) = \sum_{y \in F} \frac{1}{|E \times F|} = \frac{1}{|E|} = P(X = x)$, donc : $X \hookrightarrow \mathcal{U}(E)$, et de même : $Y \hookrightarrow \mathcal{U}(F)$. A fortiori, X et Y sont indépendantes car pour tous $x \in E$ et $y \in F$: $\displaystyle P(X = x \text{ et } Y = y) = \frac{1}{|E \times F|} = \frac{1}{|E|} \times \frac{1}{|F|} = P(X = x)\ P(Y = y)$.

(ii) $\Longrightarrow$ (i) Le couple (X,Y) est à valeurs dans $E \times F$ et pour tout $(x,y) \in E \times F$:

$$P(X = x \text{ et } Y = y) \overset{\text{Indép.}}{=} P(X = x)\ P(Y = y) = \frac{1}{|E|} \times \frac{1}{|F|} = \frac{1}{|E \times F|}, \quad \text{donc :} \quad (X,Y) \hookrightarrow \mathcal{U}(E \times F). \qquad \blacksquare$$

Nous avons déjà appris à calculer la loi d'une variable aléatoire de la forme $f(Z)$. Dans le cas d'un couple $Z=(X,Y)$, on obtient l'énoncé suivant.

Théorème (Image d'un couple de variables aléatoires par une fonction et formule de transfert associée) Soient (Ω,P) un espace probabilisé fini, F un ensemble, X et Y deux variables aléatoires sur Ω et $f : (X,Y)(\Omega) \longrightarrow F$ une fonction.

La loi $P_{f(X,Y)}$ de $f(X,Y)$ est entièrement déterminée par f et la loi de (X,Y). Précisément, pour tout $z \in f(X,Y)(\Omega)$:

$$P\big(f(X,Y)=z\big) = \sum_{\substack{(x,y)\in(X,Y)(\Omega)\\ z=f(x,y)}} P(X=x \text{ et } Y=y).$$

✎ En pratique ✎ Cas particulier le plus important :

$$P(X+Y=z) = \sum_{\substack{(x,y)\in(X,Y)(\Omega)\\ z=x+y}} P(X=x \text{ et } Y=y).$$

Exemple On lance 2 fois un dé équilibré à 6 faces. La valeur obtenue au premier (resp. deuxième) lancer est notée X_1 (resp. X_2). Les variables aléatoires X_1 et X_2 sont deux variables aléatoires indépendantes de loi uniforme sur $[\![1,6]\!]$, mais on pourrait dire de façon équivalente, comme nous l'avons vu plus haut, que le couple (X_1,X_2) suis la loi uniforme sur $[\![1,6]\!]^2$.

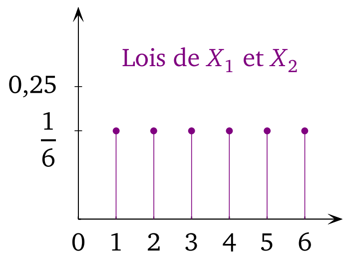
Nous allons tâcher de déterminer la loi de la somme $S=X_1+X_2$, la loi conditionnelle de X_1 sachant $\{S=4\}$ et la loi de l'écart $E=|X_1-X_2|$.

Démonstration

- **Loi de la somme $S=X_1+X_2$:** La variable S est à valeurs dans $[\![2,12]\!]$ et pour tout $s\in[\![2,12]\!]$:

$$P(S=s)=P(X_1+X_2=s) = \sum_{\substack{1\leqslant i,j\leqslant 6\\ i+j=s}} P(X_1=i \text{ et } X_2=j) \overset{\text{Indép.}}{=} \sum_{\substack{1\leqslant i,j\leqslant 6\\ i+j=s}} P(X_1=i)\,P(X_2=j) = \sum_{\substack{1\leqslant i,j\leqslant 6\\ i+j=s}} \frac{1}{36}.$$

  Or combien existe-t-il de couples $(i,j)\in[\![1,6]\!]^2$ pour lesquels : $i+j=s$? Réponse : autant que de couples $(i,s-i)$ pour lesquels : $1\leqslant i\leqslant 6$ et $1\leqslant s-i\leqslant 6$, i.e. : $1\leqslant i\leqslant 6$ et $s-6\leqslant i\leqslant s-1$. Il découle de ces encadrements que pour tout $s\in[\![2,7]\!]$: $P(S=s)=\sum_{i=1}^{s-1}\frac{1}{36}=\frac{s-1}{36}$, et pour tout $s\in[\![8,12]\!]$: $P(S=s)=\sum_{i=s-6}^{6}\frac{1}{36}=\frac{13-s}{36}$.

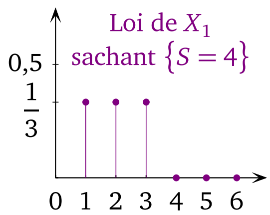


- **Loi conditionnelle de X_1 sachant $\{S=4\}$:** Pour tout $k\in[\![1,6]\!]$:

$$P(X_1=k \text{ et } S=4)=P(X_1=k \text{ et } X_2=4-k) \overset{\text{Indép.}}{=} P(X_1=k)\,\underbrace{P(X_2=4-k)}_{=0 \text{ si : } k\geqslant 4} = \begin{cases} \frac{1}{6}\times\frac{1}{6}=\frac{1}{36} & \text{si : } k\in[\![1,3]\!] \\ 0 & \text{sinon,} \end{cases}$$

  et comme : $P(S=4)=\frac{4-1}{36}=\frac{1}{12}$ d'après le point précédent :

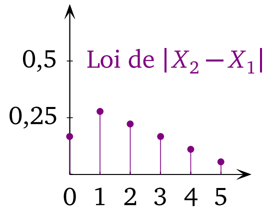
$$P_{\{S=4\}}(X_1=k)=\frac{P(X_1=k \text{ et } S=4)}{P(S=4)} = \begin{cases} \frac{1}{3} & \text{si : } k\in[\![1,3]\!] \\ 0 & \text{sinon.} \end{cases}$$



- **Loi de l'écart** $E = |X_1 - X_2|$ **:** La variable aléatoire $|X_2 - X_1|$ est à valeurs dans $[\![0,5]\!]$ et pour tout $d \in [\![0,5]\!]$:

$$P\big(|X_2 - X_1| = d\big) = P\big(X_2 - X_1 = d \text{ ou } X_1 - X_2 = d\big) = \begin{cases} P(X_1 = X_2) & \text{si : } \quad d = 0 \\ P(X_2 - X_1 = d) + P(X_1 - X_2 = d) & \text{sinon.} \end{cases}$$

Or : $\quad P(X_1 = X_2) = \displaystyle\sum_{k=1}^{6} P(X_1 = k \text{ et } X_2 = k) = \sum_{k=1}^{6} \frac{1}{36} = \frac{1}{6}$ $\quad$ et pour tout $d \in [\![1,5]\!]$:



$$\begin{aligned} P\big(|X_2 - X_1| = d\big) &= P(X_2 - X_1 = d) + P(X_1 - X_2 = d) \\ &= \sum_{i=1}^{6-d} P(X_1 = i \text{ et } X_2 = d+i) + \sum_{j=1}^{6-d} P(X_1 = d+j \text{ et } X_2 = j) \\ &= \sum_{i=1}^{6-d} \frac{1}{36} + \sum_{j=1}^{6-d} \frac{1}{36} = 2 \times \frac{6-d}{36} = \frac{6-d}{18}. \end{aligned}$$

Exemple Dans un centre d'appels, un employé effectue n appels téléphoniques vers n correspondants distincts dont chacun décroche avec probabilité p.

- On note N_1 le nombre de correspondants qui ont décroché. Quelle est donc la loi de N_1 ? Réponse **SANS CALCUL** : $N_1 \hookrightarrow \mathscr{B}(n,p)$, car les appels sont indépendants et la probabilité d'obtenir un correspondant ne dépend pas du correspondant choisi.
- L'employé rappelle un peu plus tard les $n - N_1$ correspondants qui n'ont pas décroché lors de sa première série d'appels. On note N_2 le nombre de ces correspondants qui décrochent cette fois et N le nombre total des correspondants qui ont décroché. Quelle est la loi de N ?

Démonstration On cherche la loi de $N = N_1 + N_2$.

- Pour tout $k \in [\![0,n]\!]$, sous l'hypothèse que $N_1 = k$, les $n-k$ appels de la deuxième série d'appels satisfont les mêmes hypothèses que ceux de la première série, donc la loi conditionnelle de N_2 sachant $\{N_1 = k\}$ est la loi binomiale $\mathscr{B}(n-k,p)$.
- La variable aléatoire N est à valeurs dans $[\![0,n]\!]$ et pour tout $s \in [\![0,n]\!]$:

$$\begin{aligned} P(N = s) = P(N_1 + N_2 = s) &= \sum_{\substack{0 \leqslant i,j \leqslant n \\ i+j=s}} P(N_1 = i \text{ et } N_2 = j) = \sum_{i=0}^{s} P(N_1 = i \text{ et } N_2 = s-i) \\ &= \sum_{i=0}^{s} P(N_1 = i)\, P_{\{N_1 = i\}}(N_2 = s-i) = \sum_{i=0}^{s} \binom{n}{i} p^i (1-p)^{n-i} \binom{n-i}{s-i} p^{s-i} (1-p)^{(n-i)-(s-i)}. \end{aligned}$$

Or avec des notations évidentes :

$$\binom{n}{i}\binom{n-i}{s-i} = \frac{n!}{i!(n-i)!} \times \frac{(n-i)!}{(s-i)!(n-s)!} = \frac{n!}{(n-s)!(s-i)!i!} = \frac{n!}{s!(n-s)!} \times \frac{s!}{i!(s-i)!} = \binom{n}{s}\binom{s}{i},$$

$$\begin{aligned} \text{donc :} \quad P(N = s) &= \sum_{i=0}^{s} \binom{n}{s}\binom{s}{i} p^s (1-p)^{2n-s-i} = \binom{n}{s} p^s (1-p)^{2n-s} \sum_{i=0}^{s} \binom{s}{i} (1-p)^{-i} \\ &\overset{\text{Binôme}}{=} \binom{n}{s} p^s (1-p)^{2n-s} \left(1 + \frac{1}{1-p}\right)^s = \binom{n}{s} p^s (2-p)^s (1-p)^{2n-2s} = \binom{n}{s} p^s (2-p)^s \big((1-p)^2\big)^{n-s} \\ &= \binom{n}{s} (2p-p^2)^s \big(1 - (2p-p^2)\big)^{n-s}. \end{aligned}$$

Finalement : $\quad N \hookrightarrow \mathscr{B}\big(n, 2p-p^2\big)$.

Théorème **(Sommes de variables aléatoires indépendantes de lois binomiales)** Soient $p \in [0,1]$, $m,n \in \mathbb{N}^*$ et $X, Y, X_1, \ldots, X_n$ des variables aléatoires définies sur un même espace probabilisé fini.

(i) Si : $\quad X \hookrightarrow \mathscr{B}(m,p) \quad$ et $\quad Y \hookrightarrow \mathscr{B}(n,p) \quad$ et si X et Y sont **INDÉPENDANTES**, alors : $\quad X + Y \hookrightarrow \mathscr{B}(m+n,p)$.

(ii) Si : $\quad X_i \hookrightarrow \mathscr{B}(p) \quad$ pour tout $i \in [\![1,n]\!]$ et si $X_1, \ldots, X_n$ sont **INDÉPENDANTES**, alors : $\quad \displaystyle\sum_{i=1}^{n} X_i \hookrightarrow \mathscr{B}(n,p)$.

Bref, quand on répète une expérience aléatoire à deux issues plusieurs fois indépendamment, si X compte le nombre de succès obtenus après m essais et Y le nombre de succès obtenus après n nouveaux essais, $X + Y$ compte le nombre de succès obtenus après $m + n$ essais !

Démonstration Il nous suffit bien sûr d'établir l'assertion (i). La variable aléatoire $X+Y$ est à valeurs dans $[\![0,m+n]\!]$ et pour tout $k\in[\![0,m+n]\!]$:

$$P(X+Y=k)=\sum_{i=0}^{k}P(X=i\text{ et }Y=k-i)\overset{\text{Indép.}}{=}\sum_{i=0}^{k}P(X=i)\,P(Y=k-i)$$
$$=\sum_{i=0}^{k}\binom{m}{i}p^i(1-p)^{m-i}\times\binom{n}{k-i}p^{k-i}(1-p)^{n-k+i}=p^k(1-p)^{m+n-k}\sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i}.$$

Or dans l'anneau des polynômes $\mathbb{R}[T]$: $(T+1)^m(T+1)^n=(T+1)^{m+n}$, donc après identification des coefficients de degré k : $\displaystyle\sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i}=\binom{m+n}{k}$ — ce qui généralise la formule de Vandermonde.

Conclusion : $P(X+Y=k)=\displaystyle\binom{m+n}{k}p^k(1-p)^{m+n-k}$, donc en effet : $X\hookrightarrow\mathscr{B}(m+n,p)$. ■

Il est enfin temps de conclure ce chapitre et nous le ferons en prouvant un petit théorème hors programme qui justifie la consistance de nombreux exercices. Nous avons justifié dans un précédent paragraphe qu'on peut toujours associer un espace probabilisé à la donnée d'une loi de variable aléatoire. Or de fait, on a souvent besoin de se donner non pas **UNE** variable aléatoire de loi donnée, mais une **FAMILLE** de variables aléatoires indépendantes de lois données. Considérons par exemple le début d'exercice suivant : « Soient X et Y deux variables aléatoires définies sur un même espace probabilisé. On suppose que X suit la loi uniforme sur $[\![0,n]\!]$ et Y la loi binomiale $\mathscr{B}\left(n,\dfrac{1}{2}\right)$. » Est-on bien sûr qu'un tel couple de variables aléatoires existe ?

Théorème **(Existence d'une famille finie de variables aléatoires de lois prescrites)** Soient $E=\{x_1,\dots,x_r\}$ et $F=\{y_1,\dots,y_s\}$ deux ensembles et $p_1,\dots,p_r,q_1,\dots,q_s\in[0,1]$ des réels pour lesquels : $\displaystyle\sum_{i=1}^{r}p_i=1$ et $\displaystyle\sum_{j=1}^{s}q_j=1.$

Il existe alors un espace probabilisé (Ω,P) et des variables aléatoires indépendantes X et Y sur Ω pour lesquels pour tout $i\in[\![1,r]\!]$: $P(X=x_i)=p_i$ et pour tout $j\in[\![1,s]\!]$: $P(Y=y_j)=q_j$.

Cet énoncé se généralise sans difficulté au cas d'un nombre fini de variables aléatoires.

Démonstration L'égalité : $\displaystyle\sum_{i,j}p_iq_j=\left(\sum_{i=1}^{r}p_i\right)\left(\sum_{j=1}^{s}q_j\right)=1$ justifie qu'il existe un espace probabilisé fini (Ω,P) et une variable aléatoire Z sur Ω pour laquelle pour tout $(i,j)\in[\![1,r]\!]\times[\![1,s]\!]$: $P\big(Z=(x_i,y_j)\big)=p_iq_j$. Notons alors X la première composante de Z et Y la seconde. Pour tout $i\in[\![1,r]\!]$:

$$P(X=x_i)=\sum_{j=1}^{s}P\big(Z=(x_i,y_j)\big)=\sum_{j=1}^{s}p_iq_j=p_i\sum_{j=1}^{s}q_j=p_i,$$

et de même, pour tout $j\in[\![1,s]\!]$: $P(Y=y_j)=q_j$. A fortiori, X et Y sont indépendantes car pour tous $i\in llb1,r]\!]$ et $j\in[\![1,s]\!]$: $P(X=x_i\text{ et }Y=y_j)=P\big(Z=(x_i,y_j)\big)=p_iq_j=P(X=x_i)\,P(Y=q_j)$. ■