

Méthodes probabilistes pour le traitement automatique
de la langue

LI 031

Table des matières

1	Éléments de la théorie des probabilités	2
1.1	Ensemble fondamental et événements	2
1.1.1	Opérations sur les événements	2
1.1.2	Propriétés des opérations sur les événements	4
1.2	Probabilités	5
1.2.1	Approche axiomatique	5
1.2.2	Quelques théorèmes élémentaires	6
1.3	Ensembles fondamentaux à événements élémentaires équiprobables	7
1.4	Probabilités conditionnelles	8
1.4.1	Présentation intuitive	8
1.4.2	Généralisation	9
1.4.3	Événements indépendants	10
1.4.4	La règle de multiplication	11
1.4.5	Formule des probabilités totales	12
1.5	Théorème de Bayes	14
1.5.1	Signification du théorème de Bayes	15
1.6	Variables aléatoires	17
1.6.1	variables aléatoires discrètes	19
1.6.2	Variables aléatoires conjointes	21
1.6.3	Variables aléatoires indépendantes	22

Chapitre 1

Eléments de la théorie des probabilités

1.1 Ensemble fondamental et événements

On considère une expérience dont l'issue n'est pas prévisible mais dont l'ensemble des résultats possibles (on dit aussi **issues**) est connu et appelé **ensemble fondamental**, noté S .

Exemples :

- si le résultat de l'expérience équivaut à la détermination du sexe d'un nouveau né, alors :

$$S = \{g, f\}$$

- si le résultat est l'ordre d'arrivée d'une course entre 7 chevaux numérotés de 1 à 7, alors :

$$S = \{\text{toutes les permutations de } 1 \dots 7\}$$

- si l'expérience consiste à jeter 2 dés, alors :

$$S = \{PP, PF, FP, FF\}$$

- si l'expérience consiste à choisir un mot au hasard dans un livre français, alors :

$$S = \{\text{ensemble des mots du français}\}$$

Tout sous-ensemble E de S est appelé un **événement**. Un événement est donc un ensemble formé de divers résultats possibles de l'expérience.

Si le résultat de l'expérience est compris dans E , alors on dit que E est **réalisé**.

Exemples :

- Dans le premier exemple, si $E = \{g\}$ alors E est l'événement *l'enfant est un garçon*.
- Dans le second exemple, si $E = \{\text{tous les résultats de } S \text{ commençant par } 3\}$ alors E correspond à l'événement *le cheval 3 remporte la course*.

1.1.1 Opérations sur les événements

Un événement étant un ensemble, on peut combiner des événements grâce aux opérateurs ensemblistes :

union L'événement $E \cup F$ est composé des résultats appartenant à E ou à F . L'événement $E \cup F$ est réalisé si soit E soit F l'est, comme l'illustre le diagramme de Venn de la figure 1.8.

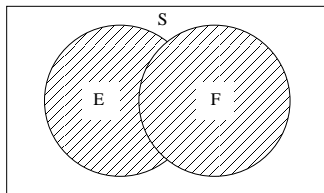


FIG. 1.1 – $E \cup F$

Exemple :

– Dans le cas du premier exemple, si $E = \{g\}$ et $F = \{f\}$, alors $E \cup F = \{f, g\}$, ce qui revient à dire que $E \cup F$ est l'ensemble S tout entier.

intersection L'événement $E \cap F$ (on note aussi EF) est composé des résultats appartenant à E et à F . Cela veut dire que EF est réalisé si E est réalisé et F l'est aussi, comme l'illustre le diagramme de Venn de la figure 1.2.

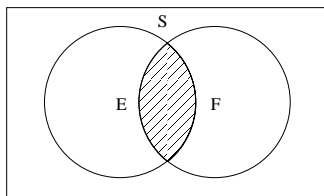


FIG. 1.2 – $E \cap F$

Exemple :

– Si $E = \{PP, PF, FP\}$ (au moins une pièce est pile) et $F = \{PF, FP, FF\}$ (au moins une pile est face), alors, l'événement $EF = \{PF, FP\}$ est l'événement *une pièce est pile et l'autre face*.

événement vide Si $E = \{g\}$ et $F = \{f\}$, alors l'événement EF ne pourra jamais être réalisé, on appellera un tel événement l'**événement vide**, que l'on notera \emptyset . Si $EF = \emptyset$, alors on dit que E et F sont **mutuellement exclusifs**

complémentation Finalement, pour chaque événement E , l'événement \bar{E} est composé des résultats qui sont dans S et qui ne sont pas dans E , comme l'illustre le diagramme de Venn de la figure 1.3. \bar{E} est appelé le complémentaire de E dans S . Si $E = \{g\}$ alors $\bar{E} = \{f\}$.

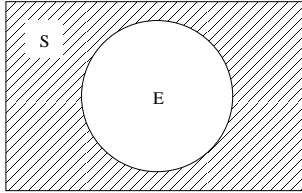


FIG. 1.3 – \bar{E}

1.1.2 Propriétés des opérations sur les événements

Propriétés algébriques Les opérations d'intersection, d'union et de complémentation d'événements obéissent à certaines règles d'algèbre :

- Commutativité $E \cup F = F \cup E$, $EF = FE$
- Associativité $(E \cup F) \cup G = E \cup (F \cup G)$, $(EF)G = E(FG)$
- Distributivité $(E \cup F)G = EG \cup FG$, $EF \cup G = (E \cup G)(F \cup G)$

Lois de De Morgan – E et F étant des événements, on a :

$$\overline{E \cup F} = \bar{E} \cap \bar{F}$$

généralisation, si E_i sont des événements :

$$\overline{\bigcup_{i=1}^n E_i} = \bigcap_{i=1}^n \bar{E}_i$$

démonstration

Pour démontrer cette loi, on commence par montrer que tout élément de $\overline{\bigcup_{i=1}^n E_i}$ est un élément de $\bigcap_{i=1}^n \bar{E}_i$, puis, inversement que tout élément du second est aussi un élément du premier.

Supposons que $x \in \overline{\bigcup_{i=1}^n E_i}$. Alors $x \notin \bigcup_{i=1}^n E_i$, ce qui signifie que x n'est contenu dans aucun des événements $E_1 \dots E_n$. Ceci implique que x est contenu dans tous les événements $\bar{E}_1 \dots \bar{E}_n$. Il est donc contenu dans $\bigcap_{i=1}^n \bar{E}_i$.

Pour montrer la réciproque, on suppose que $x \in \bigcap_{i=1}^n \bar{E}_i$, alors x appartient à chaque $\bar{E}_1 \dots \bar{E}_n$ et donc $x \notin \bigcup_{i=1}^n E_i$, ce qui entraîne finalement que $x \in \overline{\bigcup_{i=1}^n E_i}$ \square

– E et F étant des événements, on a :

$$\overline{E \cap F} = \bar{E} \cup \bar{F}$$

généralisation, si E_i sont des événements :

$$\overline{\bigcap_{i=1}^n E_i} = \bigcup_{i=1}^n \bar{E}_i$$

démonstration

On peut montrer cete loi en utilisant la précédente. On a en effet

$$\overline{\bigcup_{i=1}^n E_i} = \bigcap_{i=1}^n \overline{E_i}$$

ce qui, du fait que $\overline{\overline{E}} = E$ nous donne :

$$\bigcup_{i=1}^n \overline{E_i} = \overline{\bigcap_{i=1}^n E_i}$$

En prenant le complémentaire des deux membres de cette équation, on obtient :

$$\bigcap_{i=1}^n E_i = \overline{\bigcup_{i=1}^n \overline{E_i}}$$

□

1.2 Probabilités

Un moyen de définir la probabilité d'un événement est de le faire en termes de fréquence relative. Une telle définition est habituellement formulée ainsi : on suppose qu'une expérience d'ensemble fondamental S est exécutée plusieurs fois sous les mêmes conditions. Pour chaque événement E de S on définit $n(E)$ comme le nombre de fois où l'événement E survient lors des n premières répétitions de l'expérience. Alors $P(E)$, la probabilité de l'événement E est définie comme la limite du rapport du nombre de fois où E survient sur le nombre total des répétitions :

$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$$

Bien que la définition précédente soit intuitivement commode, et qu'elle doive toujours rester à l'esprit, elle possède un sérieux inconvénient. Nous ne savons pas, en fait, si $n(E)$ va converger vers une limite constante qui sera la même pour chaque séquence de répétitions de l'expérience. Dans le cas du jet d'une pièce par exemple, peut-on être sûr que la proportion de piles sur les n premiers jets va tendre vers une limite donnée lorsque n grandit à l'infini ? En plus, même si elle converge vers une certaine valeur, peut-on être sûr que nous obtiendrons de nouveau la même proportion limite de piles si l'expérience est entièrement répétée une deuxième fois ?

C'est la raison pour laquelle on préfère admettre pour les probabilités un ensemble d'axiomes simples et intuitivement acceptables. En particulier, nous admettrons que pour chaque événement E de l'ensemble fondamental S il existe une valeur $P(E)$ appelée probabilité de E . Nous admettrons alors que ces probabilités satisfont à un certain groupe d'axiomes qui seront introduits dans la section suivante. Ces axiomes sont, espérons que le lecteur en conviendra, en accord avec notre notion intuitive des probabilités.

1.2.1 Approche axiomatique

On considère que pour chaque événement E de l'ensemble fondamental S , il existe une valeur $P(E)$ appelée **probabilité de E** qui vérifie les trois axiomes suivants :

1.

$$0 \leq P(E) \leq 1$$

2.

$$P(S) = 1$$

3. Pour chaque séquence d'événements mutuellement exclusif $E_1, E_2 \dots E_n$ (c'est à dire d'événements pour lesquels $E_i E_j = \emptyset$ si $i \neq j$),

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$$

Pour chaque séquence d'événements mutuellement exclusifs, la probabilité qu'au moins l'un de ces événements survienne est simplement la somme de leur probabilités respectives.

Exemple : En supposant qu'à l'issue du lancer d'un dé les six faces ont les mêmes chances d'apparaître, on aura :

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

Du troisième axiome, il résulte que la probabilité de tirer un nombre pair est :

$$P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{3}{6}$$

1.2.2 Quelques théorèmes élémentaires

1.

$$P(\overline{E}) = 1 - P(E)$$

démonstration : $1 = P(S) = P(E \cup \overline{E}) = P(E) + P(\overline{E})$

2. Si $E \subset F$, alors $P(E) \leq P(F)$

\Rightarrow **mettre la démonstration**

3. relation entre la probabilité de la réunion de deux événements d'une part, les probabilités individuelles et la probabilité de l'intersection d'autre part :

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

Démonstration

Ce théorème peut être démontré de façon graphique en utilisant le diagramme de Venn de la figure 1.4. La section I représente tous les éléments de E qui ne sont pas dans F (c'est à dire $E\overline{F}$); la section II représente tous ceux qui sont dans E et dans F (c'est à dire EF) et la section III, tous ceux de F qui ne sont pas dans E (c'est à dire $\overline{E}F$)

On peut voir, d'après la figure 1.4 que :

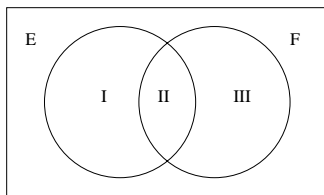


FIG. 1.4 – Illustration du théorème $P(E \cup F) = P(E) + P(F) - P(EF)$

$$E \cup F = I \cup II \cup III$$

$$E = I \cup II$$

$$F = II \cup III$$

I, II et III étant disjoints, on a d'après l'axiome 2 :

$$P(E \cup F) = P(I) + P(II) + P(III)$$

$$P(E) = P(I) + P(II)$$

$$P(F) = P(II) + P(III)$$

On a donc :

$$P(E \cup F) = P(E) + P(F) - P(II)$$

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

1.3 Ensembles fondamentaux à événements élémentaires équiprobables

Pour de nombreuses expériences, il est naturel d'admettre que chaque événement élémentaire a la même probabilité d'apparaître.

Si $S = \{1, 2 \dots N\}$, on a :

$$P(\{1\}) = P(\{2\}) = \dots P(\{N\})$$

Ce qui implique du fait des axiomes 2 et 3 :

$$\forall i 1 \leq i \leq N, P(\{i\}) = \frac{1}{N}$$

De ceci et de l'axiome 3, il résulte que pour tout événement E :

$$P(E) = \frac{\text{nombre d'éléments dans E}}{\text{nombre d'éléments dans S}}$$

⇒ effectuer le calcul

Exemple : Si deux dés sont jetés, quelle est la probabilité que la somme des faces soit 7?

Solution : On fait l'hypothèse que les 36 issues possibles sont équiprobables. Puisqu'il y a 6 issues, à savoir (1, 6), (2, 5), (3, 4), (4, 3), (5, 2) et (6, 1), qui donnent une somme de 7 pour les deux dés, la probabilité est $6/36 = 1/6$.

Exemple : Si deux boules sont tirées au hasard d'un bol en contenant 6 blanches et 5 noires, quelle est la probabilité qu'une des boules tirées soit blanche et l'autre noire?

Solution : Si nous considérons l'ordre dans lequel les boules sont choisies comme significatif, l'ensemble fondamental comprend $11 \times 10 = 110$ points. De plus, il y a $6 \times 5 = 30$ manières de tirer pour lesquelles la première boule est blanche et l'autre noire. On compte de même $5 \times 6 = 30$ manières de tirer pour lesquelles la première boule est noire et la seconde blanche. De ce fait, si tirer au hasard signifie que chacun des 110 points de l'ensemble fondamental a la même probabilité de survenir, nous voyons que la probabilité cherchée est :

$$\frac{30 + 30}{110} = \frac{6}{11}$$

1.4 Probabilités conditionnelles

Le concept de **probabilité conditionnelle** est l'un des concepts les plus importants de la théorie des probabilités. L'importance de ce concept est de deux ordres :

- On s'intéresse souvent à calculer des probabilités lorsqu'une partie de l'information concernant le résultat de l'expérience est disponible dans une telle situation, les probabilités cherchées sont justement des probabilités conditionnelles.
- Même lorsqu'aucune information partielle n'est disponible, il est souvent avantageux d'utiliser un détour par certaines probabilités conditionnelles pour réussir le calcul des probabilités cherchées.

1.4.1 Présentation intuitive

Supposons que nous jetions deux dés et que chacun des 36 événements élémentaires ait la même probabilité de survenir, soit $\frac{1}{36}$. Supposons encore que nous puissions observer le premier dé, qui donne un 3. Sur la base de cette information, quelle est dès lors la probabilité que la somme des deux dés donne 8

Pour calculer cette probabilité, on peut procéder comme suit : le dé initial étant un 3, il ne peut plus y avoir que 6 événements dans notre expérience, à savoir : (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6). Puisque chacun de ces événements a originellement la même probabilité d'apparaître, ils au-

ront encore des probabilités égales.

Autrement dit, étant donné que le premier dé est un 3, la probabilité (conditionnelle) de chacun des événements devient $\frac{1}{6}$, tandis que la probabilité (conditionnelle) des 30 autres événements de l'ensemble fondamental devient 0. La probabilité recherchée est donc $\frac{1}{6}$.

Si nous désignons respectivement par E et F les événements *la somme des dés est 8* et *le premier dé donne 3*, comme l'illustre la figure 1.5, une probabilité comme celle calculée dans l'exemple ci-dessus est appelée probabilité conditionnelle que E apparaisse sachant que F est réalisée, elle est notée $P(E|F)$ (*probabilité de E sachant F*).

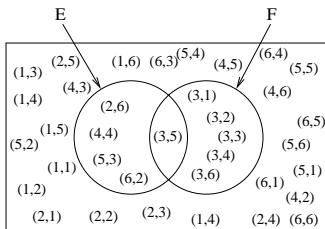


FIG. 1.5 – $P(d_1 + d_2 = 8 | d_1 = 3)$

1.4.2 Généralisation

On s'inspire de la même démarche pour dériver une formule générale donnant $P(E|F)$ pour tout événement E et F :

Si F est réalisé, alors E apparaîtra chaque fois que l'on aura affaire à un événement de E et de F à la fois, en d'autres termes, ce sera un élément de EF . Par ailleurs, comme nous savons que F est réalisé, cet ensemble devient notre nouvel ensemble fondamental, que l'on appellera *ensemble fondamental réduit*. Par conséquent, la probabilité conditionnelle de l'événement E sera donnée par comparaison de la probabilité non conditionnelle de EF avec la probabilité non conditionnelle de F . On débouche ainsi sur la définition suivante : Si $P(F) > 0$, la probabilité conditionnelle de E sera :

$$P(E|F) = \frac{P(EF)}{P(F)} \quad (1.1)$$

En multipliant par $P(F)$ les deux membres de l'équation, on obtient :

$$P(EF) = P(F)P(E|F) \quad (1.2)$$

Cette équation signifie que la probabilité que E et F apparaissent à la fois est égale à la probabilité que F apparaisse multipliée par la probabilité conditionnelle de E si l'on sait que F est survenu.

Exemple :

Une urne contient 10 billes blanches, 5 jaunes et 10 noires. Une bille est tirée au hasard de l'urne et l'on constate qu'elle n'est pas noire. Quelle est la probabilité qu'elle soit jaune ?

Solution : Soit J l'événement *la bille tirée est jaune* et soit \overline{N} *elle n'est pas noire*. De 1.1 on tire :

$$P(J|\overline{N}) = \frac{P(J\overline{N})}{P(\overline{N})}$$

Cependant, $J\overline{N} = J$ puisque la bille sera à la fois jaune et non noire si et seulement si elle est jaune. Nous obtenons ainsi, en supposant que chacune des 25 billes a la même chance d'être choisie :

$$P(J|\overline{N}) = \frac{\frac{5}{25}}{\frac{15}{25}} = \frac{1}{3}$$

Il faut noter que l'on aurait aussi pu déduire cette probabilité en travaillant directement avec l'ensemble fondamental réduit. Comme nous savons en effet que la bille choisie n'est pas noire, le problème se réduit à calculer la probabilité qu'une bille soit jaune lorsqu'elle est choisie au hasard dans une urne en contenant 10 blanches et 5 jaunes. Cette probabilité est évidemment $5/15 = 1/3$.

Lorsque l'on sait les événements équiprobables, il est souvent plus facile de calculer une probabilité conditionnelle en considérant l'ensemble fondamental réduit qu'en invoquant la formule 1.1.

Exemple : Une urne contient 8 boules rouges et 4 blanches. On tire sans remise 2 boules de l'urne et on admet qu'à chaque étape tous les tirages possibles sont équiprobables. Quelle est la probabilité que les 2 boules tirées soient rouges ?

Solution : Appelons R_1 et R_2 respectivement les événements *la première boule est rouge* et *la seconde est rouge*. Si la première boule sélectionnée est rouge, il reste dès lors 7 boules rouges et 4 boules blanches. Donc $P(R_2|R_1) = 7/11$; comme $P(R_1)$ vaut $8/12$, la probabilité recherchée est :

$$P(R_2R_1) = P(R_2|R_1)P(R_1) = 2/3 \times 7/11 = 14/33$$

1.4.3 Événements indépendants

La probabilité conditionnelle de E sachant que F est réalisé n'est en général pas égale à $P(E)$, la probabilité non conditionnelle de E . En d'autres termes, le fait de savoir que F est survenu influence la probabilité de E . Dans les cas où $P(E|F)$ est bien égal à $P(E)$, l'événement E est dit **indépendant** de F . Plus précisément, E est indépendant de F si le fait de savoir que F est survenu ne change pas la probabilité de E .

Du fait que $P(E|F) = \frac{P(EF)}{P(F)}$, on voit que l'indépendance de E et F équivaut à :

$$P(EF) = P(E)P(F)$$

Cette équation est symétrique en E et F , il en résulte que lorsque E est indépendant de F , F l'est aussi de E .

Exemple : On tire au hasard une carte d'un paquet de 52 cartes à jouer ordinaires. Désignons par E l'événement *la carte tirée est un as* et par F *elle est un pique*. Alors E et F sont indépendants. En effet, $P(EF) = 1/52$ d'une part, alors que d'autre part $P(E) = 4/52$ et $P(F) = 13/52$.

Exemple : On jette deux pièces et on suppose que les 4 résultats possibles sont équiprobables. On désigne par A *la première pièce est pile* et par B *la seconde est face*. A et B sont indépendants puisque $P(AB) = P(\{(P, F)\}) = 1/4$ d'une part, et $P(A) = P(\{(P, P), (P, F)\}) = 1/2$ $P(B) = P(\{(P, F), (F, F)\}) = 1/2$ d'autre part.

Il est important de ne pas confondre les deux notions d'événements mutuellement exclusifs et d'événements indépendants, deux notions qui entretiennent un certain rapport. En effet, si l'on sait que deux événements E et F sont mutuellement exclusifs alors on sait que si E est réalisé, F ne peut pas l'être, et vice versa. Par conséquent, la connaissance de E va modifier la connaissance de F , les deux événements ne sont donc pas indépendants. Ce raisonnement peut être illustré graphiquement à l'aide des diagrammes de Venn de la figure 1.6.

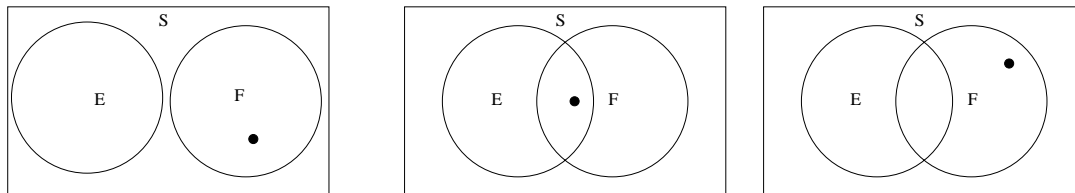


FIG. 1.6 – Événements indépendants et événements mutuellement exclusifs

Dans le diagramme de gauche, les événements E et F sont indépendants. Si F est réalisé, que l'on représente par le point noir représentant l'issue élémentaire correspondant au résultat de l'expérience, alors E ne peut pas l'être car le point noir ne peut appartenir à E et à F . Par contre, dans les deux diagrammes de droite, le fait de savoir que le point noir appartienne à F n'exclue pas qu'il appartienne à E , comme l'illustre le diagramme du milieu.

1.4.4 La règle de multiplication

On peut généraliser l'équation 1.2 à l'intersection d'un nombre arbitraire d'événements. Cette règle est appelée règle de multiplication.

$$P(E_1 E_2 \dots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \dots P(E_n|E_1 \dots E_{n-1})$$

On démontre cette loi en appliquant la définition des probabilités conditionnelles au membre de droite :

$$P(E_1) \frac{P(E_2 E_1)}{P(E_1)} \frac{P(E_3 E_1 E_2)}{P(E_2 E_1)} \cdots \frac{P(E_1 \dots E_n)}{P(E_1 \dots E_{n-1})} = P(E_1 E_2 \dots E_n)$$

1.4.5 Formule des probabilités totales

Soient E et F deux événements quelconques. Nous pouvons écrire E sous la forme $E = EF \cup E\bar{F}$, car tout élément de E doit se trouver soit dans E et dans F à la fois, soit dans E mais pas dans F , comme l'illustre la figure 1.7. Du fait que EF et $E\bar{F}$ s'excluent mutuellement, on peut écrire :

$$\begin{aligned} P(E) &= P(EF) + P(E\bar{F}) \\ &= P(E|F)P(F) + P(E|\bar{F})P(\bar{F}) \\ &= P(E|F)P(F) + P(E|\bar{F})[1 - P(F)] \end{aligned}$$

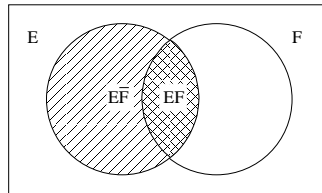


FIG. 1.7 - $E = EF \cup E\bar{F}$

L'équation 1.4.5, appelée **formule des probabilités totales**, peut être interprétée de la façon suivante : la probabilité de l'événement E est une moyenne pondérée de la probabilité conditionnelle de E lorsque F est apparu et de la probabilité du même E lorsque F n'est pas apparu, les poids étant les probabilités des événements conditionnants.

L'intérêt de cette formule est qu'elle permet de déterminer la probabilité d'un événement en commençant par le conditionner selon l'apparition ou non d'un autre événement.

Exemple : Une compagnie d'assurance estime que les gens peuvent être répartis en deux classes : ceux qui sont enclins aux accidents et ceux qui ne le sont pas. Les statistiques montrent qu'un individu enclin aux accidents a une probabilité de 0,4 d'en avoir un dans l'espace d'un an ; cette probabilité tombe à 0,2 pour les gens à risque modéré. On suppose que 30% de la population appartient à la classe à haut risque. Quelle est alors la probabilité qu'un nouvel assuré soit victime d'un accident durant l'année qui suit la signature de son contrat ?

Solution : Nous obtiendrons la probabilité de l'événement cité en le conditionnant selon que le signataire de la police est ou n'est pas enclin aux accidents. On note par A_1 l'événement *le signataire aura un accident dans l'année qui suit l'établissement du contrat* et par A *le signataire est enclin aux accidents*. La probabilité $P(A_1)$ voulue est alors donnée par :

$$\begin{aligned}
P(A_1) &= P(A_1|A)P(A) + P(A_1|\bar{A})P(\bar{A}) \\
&= .4 \times .3 + .2 \times .7 = .26
\end{aligned}$$

Exemple(suite) : Un nouveau signataire a un accident dans l'année qui suit l'établissement de son contrat. Quelle est la probabilité qu'il fasse partie de la classe à haut risque ?

Solution : Cette probabilité est $P(A|A_1)$, donnée par :

$$\begin{aligned}
P(A|A_1) &= \frac{P(AA_1)}{P(A_1)} \\
&= \frac{P(A)P(A_1|A)}{P(A_1)} \\
&= \frac{.3 \times .4}{.26} = \frac{6}{13}
\end{aligned}$$

Exemple : Un laboratoire d'analyse du sang assure avec une fiabilité de 95% la détection d'une certaine maladie lorsqu'elle est effectivement présente. Cependant, le test indique aussi un résultat faussement *positif* pour 1% des personnes réellement saines à qui on le fait subir (c'est à dire qu'une personne saine testée sera déclarée malade une fois sur cent). Si 0,5% de la population porte effectivement la maladie, quelle est la probabilité qu'une personne soit vraiment malade lorsqu'on la déclare telle sur la base du test ?

Solution : Soit D l'événement *la personne soumise au test est porteuse de la maladie* et E l'événement *le résultat du test est positif*. La probabilité $P(D|E)$ voulue se calcule de la façon suivante :

$$\begin{aligned}
P(D|E) &= \frac{P(DE)}{P(E)} \\
&= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|\bar{D})P(\bar{D})} \\
&= \frac{.95 \times .005}{.95 \times .005 + .01 \times .995} \\
&= \frac{95}{294} \simeq .323
\end{aligned}$$

Ainsi 32% seulement des personnes dont les résultats au test sont positifs ont vraiment la maladie.

Autre formulation : Puisque 0,5% de la population est réellement affectée par cette maladie, sur 200 personnes testées, 1 en moyenne l'aura. Le test détectera ce cas avec une probabilité de 0,95%. En moyenne donc, sur 200 personnes testées, on détectera correctement 0,95% cas.

D'autre part, parmi les 199 personnes saines, le test va à tort détecter $199 \times 0,01$ cas de maladie. Si l'on résume, à 0,95 cas de maladie correctement détectés, s'ajoutent en moyenne 1,99 cas faussement positifs (cas des personnes saines en réalité). Dès lors, la proportion de résultats corrects quand le test est positif n'est que de :

$$\frac{.95}{.95 + 199 \times .01} = \frac{95}{294} \simeq .323$$

Généralisation de la formule des probabilités totales :

Si $F_1 \dots F_n$ sont des événements s'excluant mutuellement, tels que :

$$\bigcup_{i=1}^n F_i = S$$

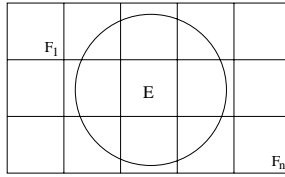


FIG. 1.8 - $\bigcup_{i=1}^n F_i = S$

On peut écrire :

$$E = \bigcup_{i=1}^n EF_i$$

En utilisant le fait que les événements EF_i s'excluent mutuellement, on peut écrire :

$$\begin{aligned} P(E) &= \sum_{i=1}^n P(EF_i) \\ &= \sum_{i=1}^n P(E|F_i)P(F_i) \end{aligned}$$

Cette équation montre qu'étant donné un jeu d'événements $F_1 \dots F_n$ desquels un et un seul surviendra, on peut calculer $P(E)$ en commençant par conditionner selon les F_i .

1.5 Théorème de Bayes

Supposons que E se soit réalisé et que nous cherchions à déterminer la probabilité que l'un des F_j se soit aussi réalisé. On déduit de l'équation ci-dessus le théorème suivant, appelé théorème de Bayes :

$$\begin{aligned}
P(F_j|E) &= \frac{P(EF_j)}{P(E)} \\
&= \frac{P(E|F_j)P(F_j)}{P(E)} \\
&= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)}
\end{aligned}$$

1.5.1 Signification du théorème de Bayes

Sous cette forme, le théorème de Bayes peut sembler bien abstrait, se présenter comme une conséquence triviale des diverses définitions que nous avons adoptées. Son intérêt est pourtant considérable car il permet de modifier notre estimation des probabilités en fonction d'une information nouvelle.

Les $F_1 \dots F_n$ sont les diverses modalités possibles que nous pouvons imaginer pour un certain événement; compte tenu de ce que nous savons des conditions dans lesquelles cet événement s'est produit (ou se produira), nous attribuons à ces éventualités les probabilités $P(F_1) \dots P(F_n)$.

Considérons un événement A dont la réalisation est liée d'une certaine façon aux F_i ; compte tenu de ce que nous savons de ces liens, nous pouvons attribuer à A une probabilité $P(A|F_1)$ de se réaliser dans l'éventualité F_1 , $P(A|F_2)$ de se réaliser dans l'éventualité $F_2 \dots$

Les nombres $P(F_i)$ et $P(A|F_i)$ résument ainsi notre connaissance des événements F_i et de leurs liens avec l'événement A .

Nous apprenons que l'événement A est réalisé, cette information nouvelle rend en partie caduc notre système de probabilités : les $P(A|F_i)$ restent valables puisqu'ils correspondent à un certain mécanisme interne des phénomènes, mais les $P(F_i)$ doivent être remplacés par des probabilités *sachant que A s'est réalisé* c'est à dire par des $P(F_i|A)$, qu'il nous suffit de calculer grâce à l'équation XX.

Nous pouvons passer ainsi de *probabilités a priori* à des *probabilités a posteriori*. Dans le cas où l'événement A est une conséquence des éventualités F_i , où par conséquent les F_i sont les causes possibles de A , le théorème de Bayes permet de calculer la probabilité de ces causes sachant que telle conséquence a eu lieu; la relation XX est souvent désignée par le terme *formule de la probabilité des causes*

Exemple : Problème de Ward Edwards

Nous disposons de deux urnes d'apparence identiques. Dans l'une, l'urne 1, nous mettons 700 boules rouges et 300 boules blanches, dans l'autre, l'urne 2, 300 boules rouges et 700 boules blanches.

Nous choisissons une de ces urnes au hasard sans regarder son contenu.

Je vous parie 10 F que l'urne choisie est l'urne 1, êtes-vous prêt à parier 10 F que c'est l'urne 2? (autrement dit, je vous donnerai 10 F si, vérification faite, c'est l'urne 2, vous m'en donnerez 10 dans le cas contraire).

Nous tirons successivement 12 boules dans l'urne choisie, en remettant à chaque tirage la boule obtenue. Nous obtenons la séquence : $RBRRRBRRRBRR$, soit 4 blanches et 8 rouges.

Etes-vous encore prêt à parier 10 F que l'urne choisie est l'urne 2?

Sinon, acceptez-vous ce pari si j'augmente ma mise jusqu'à 12 F (autrement dit, je vous donnerai 10 F si, vérification faite, c'est l'urne 2, vous m'en donnerez 12 dans le cas contraire) jusqu'à 14F, 20F, ...? Quel niveau vous semble équilibrer les espérances?

Même question si nous tirons seulement 4 boules et que nous obtenions 4 boules rouges.

La plupart des personnes interrogées répondent oui à la question 1. Elles estiment en effet, ce qui est exact, que les chances de gain sont égales; le jeu est équilibré.

⇒ **mettre l'équation** $P(U_1) \times C_a = P(U_2) \times C_b$

Mais après la seconde étape (le tirage des 8 boules rouges et des 4 blanches) chacun pense (ce qui est encore exact) que l'urne choisie est plus probablement l'urne 1 puisque l'on a tiré plus de boules rouges que de blanches. Le jeu devient donc déséquilibré. L'écart qui semble nécessaire entre les enjeux pour rétablir l'équilibre est une mesure de ce plus probablement. En général, la réponse montre que la confiance dans l'égalité des chances d'avoir choisi l'une ou l'autre urne est un peu ébranlée par le résultat du tirage; un enjeu de 15 ou 20F contre l'urne 2 semble suffisant à la plupart des personnes pour équilibrer les espérances (ce qui est grossièrement faux).

Quelle est la réponse rigoureuse?

Appelons U_1 l'événement *l'urne choisie est l'urne 1* et U_2 *l'urne choisie est l'urne 2*. Avant l'expérience, les probabilités attachées à ces deux événements sont :

$$P(U_1) = P(U_2) = \frac{1}{2}$$

Si l'on désigne respectivement par R et B les événements *la boule choisie est rouge* et *la boule choisie est blanche*, les probabilités de tirer des boules rouges ou blanches dans chacune des urnes sont :

$$\begin{aligned} P(R|U_1) &= 0,7 & P(B|U_1) &= 0,3 \\ P(R|U_2) &= 0,3 & P(B|U_2) &= 0,7 \end{aligned}$$

La probabilité d'obtenir la séquence S décrite dans l'énoncé, si l'urne choisie est l'urne 1 ou 2, est d'après le principe des probabilités conditionnelles :

$$P(S|U_1) = 0,7^8 \times 0,3^4 \quad P(S|U_2) = 0,7^4 \times 0,3^8$$

En appliquant le théorème de Bayes, nous obtenons :

$$P(U_1|S) = \frac{P(U_1)P(S|U_1)}{P(U_1)P(S|U_1) + P(U_2)P(S|U_2)} = 0,967$$

Autrement dit, une fois connue la séquence S , la probabilité pour que l'urne choisie soit l'urne 1 est égale à 96,7%.

L'enjeu qui équilibre le pari a une valeur X telle que :

$$\begin{aligned} 3,3 \times X &= 96,7 \times 10F \\ X &= 293F \end{aligned}$$

La réponse est exactement la même dans le cas où seulement 4 boules, toutes rouges, ont été tirées. Une attitude fondée sur la seule intuition a décidément, dans ce cas, bien peu de chance d'être correcte.

1.6 Variables aléatoires

Une variable aléatoire est une fonction¹ qui associe un nombre réel à tous les éléments de S , comme l'illustre la figure 1.9 ($X : S \rightarrow R$). Les variables aléatoires permettent ainsi de passer d'événements, qui peuvent être de natures très diverse, à des nombres, et d'effectuer des calculs sur ces derniers.

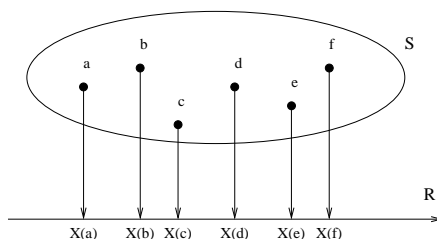


FIG. 1.9 – Représentation graphique d'une variable aléatoire X

Une variable aléatoire permet aussi de n'associer qu'une valeur à un ensemble d'issues élémentaires, dans le cas où à l'issue d'une expérience on s'intéresse plus à une fonction du résultat qu'au résultat lui-même.

¹Il peut sembler étrange d'appeler variable une fonction, c'est là une convention qui a la vie dure.

Lorsqu'on joue aux dés, par exemple, certains jeux accordent de l'importance à la somme obtenue avec deux dés, 7 par exemple, plutôt qu'à la question de savoir si c'est la paire (1, 6) qui est apparue, ou (2, 5) On peut alors définir une variable aléatoire X qui représente la somme des deux dés, comme l'illustre la figure 1.10.

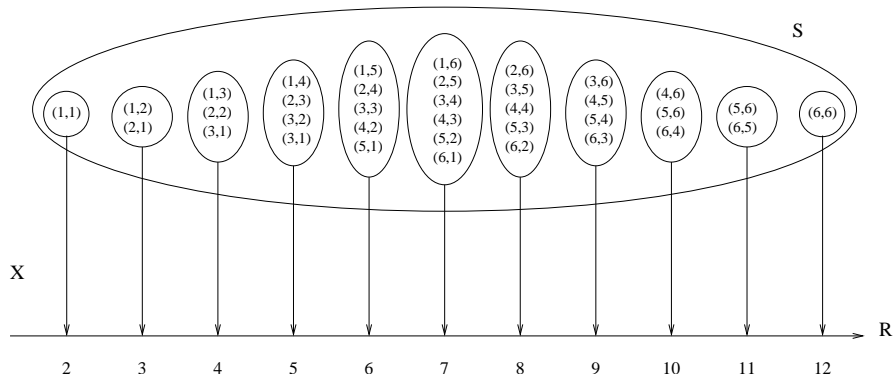


FIG. 1.10 – Variable aléatoire correspondant à la somme de deux dés

Dans le cas du jet d'une pièce, il peut être plus intéressant de connaître le nombre de fois où pile est apparue plutôt que la séquence détaillée des piles et des faces. De telles grandeurs peuvent être représentées par des variables aléatoires.

Du fait que la valeur d'une variable aléatoire est déterminée par le résultat de l'expérience, il est possible d'attribuer une probabilité aux différentes valeurs que la variable aléatoire peut prendre. On note la probabilité que la variable aléatoire X prenne la valeur i par : $P(X = i)$.

Exemple :

Notre expérience consiste à jeter 3 pièces équilibrées. Si l'on désigne le nombre de faces par Y , Y est une variable aléatoire et peut prendre les valeurs 0, 1, 2, 3 avec pour probabilité respectivement :

$$\begin{aligned}
 P(Y = 0) &= P((P, P, P)) = \frac{1}{8} \\
 P(Y = 1) &= P((P, P, F), (P, F, P), (F, P, P)) = \frac{3}{8} \\
 P(Y = 2) &= P((P, F, F), (F, P, F), (F, F, P)) = \frac{3}{8} \\
 P(Y = 3) &= P((F, F, F)) = \frac{1}{8}
 \end{aligned}$$

Du fait que Y doit nécessairement prendre l'une des valeurs 0, 1, 2, 3, on aura :

$$1 = P\left(\bigcup_{i=0}^3 (Y = i)\right) = \sum_{i=0}^3 P(Y = i)$$

1.6.1 variables aléatoires discrètes

Une variable aléatoire ne pouvant prendre qu'une quantité dénombrable de valeurs est dite **discrète**. Pour une telle variable aléatoire X , on définit sa **loi de probabilité** p_X par²

$$p(a) = P(X = a) = P(E) \text{ avec } E = \{i \in S | X(i) = a\}$$

Cette loi de probabilité ne peut être positive que pour un ensemble dénombrable d'arguments. En d'autres termes, si X peut prendre les valeurs x_1, x_2, \dots , alors :

$$p(x_i) \geq 0 \quad i = 1, 2, \dots \quad p(x) = 0 \text{ pour toutes les autres valeurs de } x$$

Du fait que X doit bien prendre l'une de ces valeurs x_i , on aura :

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Si une variable aléatoire X suit une loi de distribution p on notera $X \sim p$

Il est souvent instructif de représenter la loi de probabilité sur un graphique en reportant $p(x_i)$ sur l'axe des y et x_i sur l'axe des x . La figure 1.11 représente le graphe de la loi de probabilité d'une variable aléatoire comptant la somme des nombres obtenus lors du jet de deux dés équilibrés.

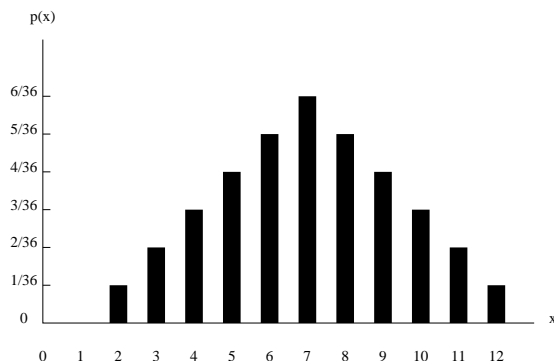


FIG. 1.11 – Représentation graphique de la loi d'une variable aléatoire discrète

La loi d'une variable aléatoire est constituée par l'ensemble des probabilités des valeurs que peut prendre cette variable. Dans certains cas, le nombre de valeurs possibles de la variable est extrêmement élevé ou même infini. Il est intéressant, dans ces cas, de dégager certaines caractéristiques qui permettent d'esquisser l'allure générale de ces lois. Nous définirons ici deux caractéristiques des lois de probabilités ; l'une, permettant de situer la variable, ce sera l'*espérance*, l'autre, permettant d'estimer la dispersion de la variable, ce sera la *variance*.

²Lorsque le contexte permet de savoir sans ambiguïté de quelle variable aléatoire on parle, la loi de probabilité de la variable est notée p , sans indice.

Espérance

Pour une variable aléatoire discrète X de loi $p(\cdot)$, on définit l'espérance de X , notée $E[X]$, par l'expression :

$$E[X] = \sum_{x|p(x)>0} xp(x)$$

En termes concrets, l'espérance de X est la moyenne pondérée des valeurs que X peut prendre, les poids étant les probabilités que ces valeurs soient prises.

Exemple : L'espérance de la v.a. X , résultat du lancer d'un dé équilibré est :

$$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{7}{2}$$

On peut donner une interprétation plus intuitive de la notion d'espérance en recourant à l'interprétation des probabilités comme mesures de fréquences relatives. Rappelons que dans cette perspective, la probabilité d'un événement E est la proportion du nombre de réalisation de E dans une séquence infiniment longue d'expériences identiques. Supposons que l'on se livre à jeu d'argent pour lequel on connaît la probabilité de chaque gain. On représente par X la variable aléatoire représentant le gain. Cette v.a. peut prendre les valeurs x_1, \dots, x_n avec les probabilités $p(x_1), \dots, p(x_n)$. Selon l'interprétation des probabilités données ci-dessus, en répétant un grand nombre de fois le jeu, la proportion de jeux pour laquelle la somme x_i aura été gagnée va se rapprocher de $p(x_i)$. L'espérance de la v.a. X est alors le gain moyen par jeu : $\sum_{i=1}^n x_i p(x_i)$

Variance

Une variable aléatoire X et sa loi de probabilité $p(\cdot)$ étant données, $E[X]$ nous donne une moyenne pondérée des valeurs possibles de X . Mais elle ne nous dit rien des variations de X autour de l'espérance. On peut s'en rendre compte grâce aux exemples suivants. Soit les variables aléatoires suivantes :

X avec $p_X(0) = 1$ Y avec $p_Y(-1) = \frac{1}{2}, p_Y(1) = \frac{1}{2}$ Z avec $p_Z(-100) = \frac{1}{2}, p_Z(100) = \frac{1}{2}$
Si toutes ont la même espérance, à savoir 0, il y a de bien plus grands écarts entre les différentes valeurs de Y qu'entre celles de W (qui est constante) et de plus grands écarts entre celles de Z qu'entre celles de Y .

Comme on s'attend dans un certain nombre de cas à voir une variable X prendre ses valeurs autour de son espérance $E[X]$, il paraît raisonnable de mesurer les variations de X en considérant l'écart moyen entre X et son espérance. Cela reviendrait à s'intéresser à la grandeur $E[|X - \mu|]$, où $\mu = E[X]$. On préfère plutôt utiliser l'espérance du carré de l'écart entre X et son espérance, que l'on appelle **variance** de X , notée $Var(X)$:

$$Var(X) = E[(X - E[X])^2]$$

Qui peut aussi être calculée de la façon suivante :

$$Var(X) = E[X^2] - (E[X])^2$$

Si X est le résultat du lancer d'un dé équilibré, alors $E[X] = \frac{7}{2}$ comme nous l'avons vu ci-dessus. De plus :

$$E[X^2] = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} = 91 \times \frac{1}{6}$$

et donc :

$$Var(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

La racine carrée de $Var(X)$ est appelée l'**écart-type** de X , qui se note σ . On a donc :

$$\sigma = \sqrt{Var(X)}$$

1.6.2 Variables aléatoires conjointes

Nous n'avons traité jusqu'ici que des distributions de variables isolées. Or, il est souvent nécessaire de considérer des événements relatifs à deux variables simultanément. On définit, pour traiter de tels problèmes, la **loi de probabilité conjointe** de deux variables X et Y :

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

On peut déduire de la loi de probabilité conjointe de deux variables X et Y la **loi de probabilité marginale** de X de la façon suivante :

$$p_X(x) = P(X = x) = \sum_{y|p(x,y)>0} p(x, y)$$

De façon similaire :

$$p_Y(y) = P(Y = y) = \sum_{x|p(x,y)>0} p(x, y)$$

Exemple :

On tire au hasard 3 boules d'une urne contenant 3 rouges, 4 blanches et 5 bleues. X et Y désignent respectivement le nombre de boules rouges et celui de boules blanches tirées. La loi de probabilité conjointe $p(i, j) = P(X = i, Y = j)$ de X et Y est représentée dans un tableau à deux entrées :

		j				
		0	1	2	3	$P(X = i)$
i	0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
	1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
	2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
	3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
$P(Y = j)$		$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

On remarquera que la loi marginale de X est calculée en faisant les totaux par ligne, tandis que celle de Y l'est en faisant les totaux par colonne. C'est le fait que les lois de X et Y individuellement puissent être lues dans les marges du tableau qui leur vaut leur nom de lois marginales.

1.6.3 Variables aléatoires indépendantes

Deux variables aléatoires discrètes X et Y de loi conjointe p sont dites **indépendantes** si, pour tout x et tout y :

$$p(x, y) = p_X(x)p_Y(y)$$