Probabilités sur un univers fini

1 Probabilité sur un ensemble fini

1.1 Espace probabilisable fini

Dans ce texte, on modélise des résultats d'expériences aléatoires ayant un nombre fini d'issues (tirage à pile ou face, loterie, tirage au sort, etc.). Cette modélisation mathématique commence par le choix d'un ensemble fini Ω appelé *univers des possibles*. Ses éléments, généralement désignés par la lettre grecque ω (qui se lit ≪ *omega* ≫), sont les résultats de l'expérience étudiée.

Les sous-ensembles de Ω sont appelés *événements*. L'ensemble des parties de Ω, que l'on note couramment $\mathcal{P}(\Omega)$ ou $\wp(\Omega)$, est donc l'ensemble des événements. Les singletons de Ω, qui sont les événements limités à un seul résultat, s'appellent *événements élémentaires*. L'ensemble vide $\varnothing$ est appelé *événement impossible*. L'ensemble ≪ plein ≫ Ω est appelé *événement certain*.

Exemple 1. On lance un dé classique à six faces. On choisit alors de modéliser l'issue de l'expérience au moyen de l'univers des possibles

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Les événements attachés à cette expérience sont généralement décrits par des phrases, que l'on modélise par des sous-ensembles de l'univers des possibles. En voici quelques exemples.

- L'événement [Le dé a fait un 3.] est l'événement élémentaire $\{3\}$.
- L'événement [Le dé a donné un résultat pair.] est l'événement $\{2, 4, 6\}$.
- L'événement [Le dé a fait un 7.] est l'événement impossible $\varnothing$.

Exemple 2. On effectue quatre lancers d'une pièce et on s'intéresse à la suite de résultats (pile/face). On code ≪ pile ≫ par le chiffre 1 et ≪ face ≫ par le chiffre 0. On choisit donc comme univers des possibles l'ensemble mathématique

$$\Omega = \{0; 1\}^4.$$

Voici alors quelques exemples d'événements décrits par des phrases, avec leur traduction ensembliste.

- L'événement [Les quatre lancers ont donné le même résultat.] s'écrit $\{(0, 0, 0, 0), (1, 1, 1, 1)\}$.
- L'événement [Il y a eu autant de ≪ pile ≫ que de ≪ face ≫.] s'écrit

$$\{(0, 0, 1, 1), (0, 1, 0, 1), (0, 1, 1, 0), (1, 0, 0, 1), (1, 0, 1, 0), (1, 1, 0, 0)\}$$

Exemple 3. On choisit six numéros distincts entre 1 et 49. On ne s'intéresse pas à l'ordre dans lequel les numéros ont été tirés, si bien que les tirages peuvent être vus comme des parties à 6 éléments de l'ensemble $[\![1, 49]\!]$. On choisit donc comme univers des possibles l'ensemble

$$\Omega = \mathcal{P}_6([\![1, 49]\!]).$$

Voici deux exemples d'événements décrits par des phrases, avec leur traduction ensembliste.

- L'événement [Les six numéros sont consécutifs.] s'écrit $\big\{\{k, k+1, k+2, k+3, k+4, k+5\}\ ;\ k \in [\![1, 44]\!]\big\}$.
- L'événement [Les six numéros sont impairs.] est plus difficile à formaliser de manière concise mais voici une solution

$$\{\{2k+1\ ;\ k \in \mathrm{T}\}\ ;\ \mathrm{T} \in \mathcal{P}_6([\![0, 24]\!])\}.$$

Voici une autre écriture, légèrement moins concise mais certainement plus compréhensible

$$\{\{2a+1, 2b+1, 2c+1, 2d+1, 2e+1, 2f+1\}\ ;\ \{a, b, c, d, e, f\} \in \mathcal{P}_6([\![0, 24]\!])\}.$$

Un *espace probabilisable fini* est la donnée d'un couple $(\Omega, \mathcal{B})$, où Ω est un ensemble fini et $\mathcal{B}$ est un sous-ensemble de $\mathcal{P}(\Omega)$ soumis à certaines lois qui ne seront pas détaillées dans ce texte. On se limitera en effet au cas où $\mathcal{B}$ est égal à $\mathcal{P}(\Omega)$, auquel cas les lois en question sont automatiquement vérifiées.

Les éléments de $\mathcal{B}$ sont appelés *événements*. Dans le cadre de ce texte, tout sous-ensemble de $\mathcal{P}(\Omega)$ est un événement.

1.2 Opérations sur les événements

Les événements sont modélisés par des sous-ensembles d'un ensemble Ω. Ils sont donc soumis aux opérations usuelles sur les ensembles : réunion, intersection, passage au complémentaire. Ils peuvent également être liés par une relation d'inclusion ou être disjoints. Ces opérations abstraites transcrivent alors la plupart du temps une opération logique intuitivement attachée au phénomène concret que l'on modélise. Les correspondances de vocabulaire sont détaillées dans ce paragraphe.

L'appartenance. Soit $\omega \in \Omega$. L'élément ω est un résultat de l'expérience [1]. Soit A un événement.

La formule $\omega \in \mathrm{A}$ signifie que ω est un élément de A. Dans le langage courant, on dira que ≪ *le résultat ω réalise l'événement* A ≫.

Dans l'exemple 1 du premier paragraphe, le résultat 3 réalise le premier événement mais pas les deux autres.

L'inclusion. Soient A et B deux événements. Dire que A est inclus dans B signifie que tout résultat ω qui réalise l'événement A réalise aussi l'événement B. On dit alors que ≪ *l'événement* A *entraîne l'événement* B ≫.

Il s'agit de comprendre qu'il s'agit d'une implication au même sens qu'en logique mathématique et qu'il est tout à fait possible qu'une implication soit vraie sans qu'elle traduise un lien de cause à effet dans la réalité [2].

Prenons l'exemple 1 du premier paragraphe (jet d'un dé à six faces). Dans cet exemple, l'événement [Le dé a fait 2 ou 4.] entraîne l'événement [Le dé a donné un résultat pair.]. Ceci est traduit mathématiquement par l'inclusion

$$\{2;4\} \subset \{2;4;6\}$$

mais l'implication se déduit directement de la formulation de ces événements — il y a un vrai lien de cause à effet logique dans ce cas précis.

Événement contraire/complémentaire. Le complémentaire d'un événement A de $\mathcal{P}(\Omega)$ est appelé *événement contraire* de A. On utilise généralement la notation

$$\overline{\mathrm{A}} = \Omega \setminus \mathrm{A}.$$

Toujours dans l'exemple d'un dé à six faces, notons A l'événement [Le dé a donné un résultat pair.]. Son événement contraire se formule [Le dé n'a pas donné un résultat pair.] et se reformule naturellement en [Le dé a donné un résultat impair.]. La modélisation mathématique s'écrit

$$\mathrm{A} = \{2,4,6\} \qquad \text{et} \qquad \overline{\mathrm{A}} = \{1,3,5\}.$$

Intersection. L'intersection de deux sous-ensembles A et B est l'ensemble des éléments ω de Ω qui appartiennent à A *et* à B en même temps. En d'autres termes, les résultats qui réalisent l'événement $\mathrm{A} \cap \mathrm{B}$ sont ceux qui réalisent simultanément les événements A et B.

Plus généralement, si $\mathrm{A}_1, \dots, \mathrm{A}_n$ sont des événements de $\mathcal{P}(\Omega)$, alors les résultats qui réalisent $\mathrm{A}_1 \cap \dots \cap \mathrm{A}_n$ sont les résultats qui réalisent simultanément *tous* les événements de la liste $(\mathrm{A}_i)_{1 \leqslant i \leqslant n}$. Une intersection multiple traduit un quantificateur universel.

Événements disjoints/incompatibles. D'un point de vue ensembliste, deux événements A et B sont disjoints si, et seulement si, leur intersection est vide. Dans le jargon probabiliste, on dit que l'intersection $\mathrm{A} \cap \mathrm{B}$ est *impossible* ; on dit aussi que les événements A et B sont *incompatibles.*

Dans chacun des exemples du premier paragraphe, les événements proposés sont deux à deux incompatibles.

Réunion. La réunion de deux événements A et B est l'ensemble des résultats ω qui réalisent au moins l'un de ces deux événements. En d'autres termes, les résultats qui réalisent l'événement $\mathrm{A} \cap \mathrm{B}$ sont ceux qui réalisent A *ou* B — il s'agit bien sûr d'un *ou inclusif* : on accepte aussi les résultats qui réalisent les deux événements à la fois ; en particulier, l'intersection $\mathrm{A} \cap \mathrm{B}$ entraîne toujours la réunion $\mathrm{A} \cup \mathrm{B}$.

Plus généralement, si $\mathrm{A}_1, \dots, \mathrm{A}_n$ sont des événements de $\mathcal{P}(\Omega)$, alors les résultats qui réalisent $\mathrm{A}_1 \cup \dots \cup \mathrm{A}_n$ sont les résultats qui réalisent au moins l'un des événements de la liste $(\mathrm{A}_i)_{1 \leqslant i \leqslant n}$. Une réunion multiple traduit un quantificateur existentiel.

Partition/système complet d'événements. Une *partition* de l'ensemble Ω est un ensemble de parties non vides de Ω, deux à deux disjointes, qui recouvrent l'ensemble Ω. Il s'agit donc d'un ensemble $\{\mathrm{A}_1, \dots, \mathrm{A}_n\}$ de parties de Ω vérifiant les conditions

1. Ce n'est pas forcément un résultat qui se produit réellement, mais un résultat qui est envisagé par le modèle mathématique.
2. Ainsi, l'événement vide $\varnothing$ entraîne n'importe quel autre événement mais il n'en est jamais la *cause*…

- $\forall i \in [\![1,n]\!], \quad \mathrm{A}_i \neq \varnothing$;
- $\forall (i,j) \in [\![1,n]\!]^2, \quad (i \neq j \Rightarrow \mathrm{A}_i \cap \mathrm{A}_j = \varnothing)$;
- $\mathrm{A}_1 \cup \ldots \cup \mathrm{A}_n = \Omega$.

Les deux dernières conditions signifient que tout élément de Ω appartient à l'un des A_i et à un seul.

Dans le langage des probabilités, une partition de l'univers Ω s'appelle *un système complet d'événéments*. Par exemple, l'ensemble des événements élémentaires est un système complet d'événements — attention, on parle bien de l'ensemble des singletons

$$\{\{\omega\} \; ; \; \omega \in \Omega\},$$

ce qui est très différent de l'ensemble des résultats, à savoir Ω.

En dénombrement, les partitions servent à décomposer des ensembles en ensembles plus simples afin d'obtenir des relations entre cardinaux. En probabilités, les systèmes complets d'événements jouent un rôle similaire, comme on le verra avec les *formules des probabilités totales*.

1.3 Probabilité

Définition. On se donne un espace probabilisable fini $(\Omega, \mathcal{P}(\Omega))$. Une *probabilité* sur cet espace probabilisé est une fonction

$$\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$$

vérifiant les conditions

- $\mathbb{P}(\Omega) = 1$;
- $\forall (\mathrm{A}, \mathrm{B}) \in \mathcal{P}(\Omega)^2$, $((\mathrm{A} \cap \mathrm{B}) = \varnothing \Rightarrow \mathbb{P}(\mathrm{A} \cup \mathrm{B}) = \mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{B}))$.

Le triplet $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est alors un *espace probabilisé fini*.

Exemple fondamental : la probabilité uniforme. Considérons un ensemble Ω fini et non vide. Pour toute partie A de Ω, on note alors

$$\mathbb{P}(\mathrm{A}) = \frac{\mathrm{Card}(\mathrm{A})}{\mathrm{Card}(\Omega)}.$$

La fonction $\mathbb{P}$ vérifie alors bien les conditions de la définition ci-dessus : c'est une probabilité sur l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$, que l'on appelle *probabilité uniforme* sur A. Pour ce choix de $\mathbb{P}$, tous les événements élémentaires ont la même probabilité, à savoir $1/\mathrm{Card}(\Omega)$.

Illustration de l'exemple fondamental. On dispose de deux dés à six faces. L'un est mauve, l'autre est bordeaux. On décide de modéliser l'issue d'un lancer de dés par un couple (m, b), où les éléments m et b, valeurs respectives des dés mauve et bordeaux, sont des entiers compris entre 1 et 6. On prend donc pour univers des possibles l'ensemble

$$\Omega = [\![1,6]\!]^2$$

et on munit l'ensemble des événements $\mathcal{P}(\Omega)$ de la probabilité uniforme $\mathbb{P}$, définie par

$$\forall \mathrm{A} \in \mathcal{P}(\Omega), \qquad \mathbb{P}(\mathrm{A}) = \frac{\mathrm{Card}(\mathrm{A})}{36}.$$

Dans de nombreux jeux utilisant deux dés, c'est la somme des deux dés qui importe et non leurs valeurs individuelles. Cette somme est comprise entre 2 et 12. Pour tout entier k compris entre 2 et 12, notons S_k l'événement [La somme des deux dés vaut k.] La modélisation choisie donne alors par exemple

$$\mathrm{S}_5 = \{(1,4), (2,3), (3,2), (4,1)\}$$

donc le cardinal de l'événement A_5 vaut 4 et sa probabilité vaut

$$\mathbb{P}(\mathrm{S}_5) = \frac{\mathrm{Card}(\mathrm{A}_5)}{\mathrm{Card}(\Omega)} = \frac{4}{36} = \frac{1}{9}.$$

Plus généralement, la probabilité des événements S_k est résumée dans le tableau suivant (les fractions n'ont pas été simplifiées pour mieux faire apparaître les variations de $\mathbb{P}(\mathrm{S}_k)$ en fonction de k).

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(\mathrm{S}_k)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Construction générale d'une probabilité. Étant donné un univers fini

$$\Omega = \{\omega_1, \ldots, \omega_n\},$$

se donner une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{P}(\Omega))$ revient à choisir des nombres $p_1, \ldots, p_n$ dans $[0, 1]$ dont la somme vaut 1, puis à poser

$$\forall \mathrm{A} \in \mathcal{P}(\Omega), \qquad \mathbb{P}(\mathrm{A}) = \sum_{i \in \mathrm{I}(\mathrm{A})} p_i \qquad \text{avec} \qquad \mathrm{I}(\mathrm{A}) = \{i \in [\![1, n]\!] \ ; \ \omega_i \in \mathrm{A}\}.$$

Une autre manière d'écrire cette formule est de faire intervenir la fonction caractéristique de A, notée $\mathbb{1}_\mathrm{A}$, définie par

$$\mathbb{1}_\mathrm{A}(\omega) = \begin{cases} 1 & \text{si } \omega \in \mathrm{A} \\ 0 & \text{sinon.} \end{cases}$$

Avec cette notation, la probabilité $\mathbb{P}$ s'écrit

$$\forall \mathrm{A} \in \mathcal{P}(\Omega), \qquad \mathbb{P}(\mathrm{A}) = \sum_{i=1}^{n} p_i \mathbb{1}_\mathrm{A}(\omega_i).$$

Exemple : une autre modélisation pour la somme de deux dés. On reprend l'expérience du lancer de deux dés mais, comme seule la somme nous intéresse, on décide de simplifier la modélisation en prenant

$$\Omega = [\![2, 12]\!]$$

et en définissant les probabilités des événements élémentaires selon le tableau de l'exemple précédent

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(\{k\})$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Considérons maintenant l'événement A décrit par la phrase [La somme des deux dés est paire.]. Dans ce modèle, cet événement s'écrit

$$\mathrm{A} = \{2, 4, 6, 8, 10, 12\}.$$

Sa probabilité vaut donc

$$\mathbb{P}(\mathrm{A}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) + \mathbb{P}(\{8\}) + \mathbb{P}(\{10\}) + \mathbb{P}(\{12\}) = \frac{1+3+5+5+3+1}{36} = \frac{18}{36} = \frac{1}{2}.$$

Les amateurs du jeu *Catane* se plaignent parfois que les scores réputés rares sortent plus souvent qu'ils ne devraient. Considérons donc l'événement

$$\mathrm{B} = \{2, 3, 11, 12\}.$$

On trouve alors

$$\mathbb{P}(\mathrm{B}) = \frac{1+2+2+1}{36} = \frac{6}{36} = \frac{1}{6}.$$

On s'attend donc à ce qu'un des scores les plus rares sorte environ une fois sur six. C'est donc loin d'être négligeable. C'est d'ailleurs la même probabilité que le résultat individuel 7.

Un peu de vocabulaire. Un événement de probabilité nulle est dit *quasi impossible.* Il arrive qu'un événement ait une probabilité nulle sans être l'ensemble vide. On prendra soin de ne pas confondre « impossible » et « quasi impossible ».

Un événement de probabilité 1 est dit *quasi certain.* Là encore, un tel événement n'est pas forcément égal à Ω.

Deux événements A et B tels que $\mathrm{A} \cap \mathrm{B}$ ait une probabilité nulle sont dits *quasi incompatibles.*

Règle de calcul. Pour tout couple (A, B) d'événements, le nombre $\mathbb{P}(\mathrm{A} \cup \mathrm{B})$ est donné par

$$\mathbb{P}(\mathrm{A} \cup \mathrm{B}) = \mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{B}) - \mathbb{P}(\mathrm{A} \cap \mathrm{B}).$$

On remarque en particulier la majoration $\mathbb{P}(\mathrm{A} \cup \mathrm{B}) \leqslant \mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{B})$, avec égalité si et seulement si les deux événements sont quasi incompatibles.

Une récurrence rapide permet d'obtenir la majoration

$$\mathbb{P}(\mathrm{A}_1 \cup \ldots \cup \mathrm{A}_n) \leqslant \sum_{k=1}^{n} \mathbb{P}(\mathrm{A}_k)$$

pour tout n-uplet d'événements $(\mathrm{A}_1, \ldots, \mathrm{A}_n)$. De plus, si les n événements sont quasi incompatibles deux à deux, alors la probabilité de leur réunion est donnée par la *formule d'additivité finie*

$$\mathbb{P}(\mathrm{A}_1 \cup \ldots \cup \mathrm{A}_n) = \sum_{k=1}^{n} \mathbb{P}(\mathrm{A}_k) \qquad \text{si les } \mathrm{A}_k \text{ sont quasi incompatibles deux à deux.}$$

En particulier, on peut écrire

$$\sum_{k=1}^{n} \mathbb{P}(\mathrm{A}_k) = 1 \qquad \text{si } \{\mathrm{A}_1, \ldots, \mathrm{A}_n\} \text{ est un système complet d'événements.}$$

À titre culturel, je donne la formule générale suivante, mais elle n'est pas à notre programme (formule du crible)

$$\mathbb{P}(\mathrm{A}_1 \cup \ldots \cup \mathrm{A}_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \mathbb{P}(\mathrm{A}_{i_1} \cap \ldots \cap \mathrm{A}_{i_k}).$$

Une autre règle de calcul. Soient A et B deux événements.

$$\mathbb{P}(\mathrm{A} \setminus \mathrm{B}) = \mathbb{P}(\mathrm{A}) - \mathbb{P}(\mathrm{A} \cap \mathrm{B}).$$

1.4 Probabilité conditionnelle

Considérons un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et un événement A tel que $\mathbb{P}(\mathrm{A})$ soit non nul. On définit la fonction

$$\begin{array}{cccc} \mathbb{P}_{\mathrm{A}} & : & \mathcal{P}(\Omega) & \to & [0,1] \\ & & \mathrm{B} & \mapsto & \frac{\mathbb{P}(\mathrm{B} \cap \mathrm{A})}{\mathbb{P}(\mathrm{A})}. \end{array}$$

Cette fonction est *la probabilité sachant* A. C'est une probabilité sur l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$. Le nombre $\mathbb{P}_{\mathrm{A}}(\mathrm{B})$ est également noté $\mathbb{P}(\mathrm{B}|\mathrm{A})$.

Exemple. Reprenons la dernière modélisation pour la somme de deux dés et prenons l'événement A décrit par la phrase [La somme des deux dés est paire.]. On a vu que $\mathbb{P}(\mathrm{A})$ vaut $1/2$. La probabilité $\mathbb{P}_{\mathrm{A}}$ est alors décrite par le tableau suivant

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}_{\mathrm{A}}(\{k\})$	1/18	0	3/18	0	5/18	0	5/18	0	3/18	0	1/18

Le principe est simple : les résultats qui n'appartiennent pas à A sont considérés comme quasi impossibles. Les autres résultats restent envisagés et leurs probabilités restent dans les mêmes rapports de proportionnalité.

Sachant A, tout événement contenant A est quasi certain.

Sachant A, tout événement inclus dans $\overline{\mathrm{A}}$ est quasi impossible.

Formule de Bayes. Prenons deux événements A et B de probabilité non nulle. Les probabilités conditionnelles $\mathbb{P}_{\mathrm{A}}(\mathrm{B})$ et $\mathbb{P}_{\mathrm{B}}(\mathrm{A})$ sont alors reliées par la formule

$$\mathbb{P}(\mathrm{A}|\mathrm{B}) = \frac{\mathbb{P}(\mathrm{B}|\mathrm{A})\mathbb{P}(\mathrm{A})}{\mathbb{P}(\mathrm{B})}.$$

Illustration sur un exercice. Sur l'île de Gaard est apparue la *maladie de Groubou*. Les Gaardiens sont atteints par la maladie de Groubou avec une probabilité de $0,3$. Pour déceler la maladie, on soumet les Gaardiens à un test partiellement fiable. La probabilité qu'un Gaardien non infecté réagisse négativement au test vaut $0,8$. La probabilité qu'un Gaardien infecté réagisse positivement au test vaut $0,9$.

Quelle est la probabilité pour un Gaardien pris au hasard ayant une réaction positive au test d'être atteint par la maladie de Groubou ?

Solution de cet exercice. Piochons un Gaardien au hasard et considérons les événements suivants.

- M =[Le Gaardien choisi a la maladie de Groubou.]
- R =[Le Gaardien choisi a une réaction positive au test.]

Avec ces notations, les données numériques de l'énoncé s'écrivent comme suit

$$\mathbb{P}(\mathrm{M}) = 0,3 \qquad \mathbb{P}(\overline{\mathrm{R}}|\overline{\mathrm{M}}) = 0,8 \qquad \mathbb{P}(\mathrm{R}|\mathrm{M}) = 0,9.$$

L'énoncé nous demande de calculer $\mathbb{P}(\mathrm{M}|\mathrm{R})$. Pour cela, il nous suffit d'appliquer la formule de Bayes mais il nous manque encore la valeur de $\mathbb{P}(\mathrm{R})$. Celle-ci se calcule en partitionnant l'événement R comme suit

$$\mathrm{R} = (\mathrm{R} \cap \mathrm{M}) \cup (\mathrm{R} \cap \overline{\mathrm{M}}).$$

Les événements $\mathrm{R} \cap \mathrm{M}$ et $\mathrm{R} \cap \overline{\mathrm{M}}$ sont incompatibles donc

$$\mathbb{P}(\mathrm{R}) = \mathbb{P}(\mathrm{R} \cap \mathrm{M}) + \mathbb{P}(\mathrm{R} \cap \overline{\mathrm{M}}).$$

On fait ensuite intervenir les probabilités conditionnelles

$$\mathbb{P}(\mathrm{R}) = \mathbb{P}(\mathrm{R}|\mathrm{M})\mathbb{P}(\mathrm{M}) + \mathbb{P}(\mathrm{R}|\overline{\mathrm{M}})\mathbb{P}(\overline{\mathrm{M}}).$$

On connaît $\mathbb{P}(\overline{\mathrm{M}}) = 1 - \mathbb{P}(\mathrm{M}) = 0,7$ et $\mathbb{P}_{\overline{\mathrm{M}}}(\mathrm{R}) = 1 - \mathbb{P}_{\overline{\mathrm{M}}}(\overline{\mathrm{R}}) = 0,2$. On obtient donc

$$\mathbb{P}(\mathrm{R}) = 0,9 \times 0,3 + 0,2 \times 0,7 = 0,27 + 0,14 = 0,41$$

puis, en reportant dans la formule de Bayes,

$$\mathbb{P}(\mathrm{M}|\mathrm{R}) = \frac{\mathbb{P}(\mathrm{R}|\mathrm{M})\mathbb{P}(\mathrm{M})}{\mathbb{P}(\mathrm{R})} = \frac{0,9 \times 0,3}{0,41} = \frac{0,27}{0,41} \simeq 0,66.$$

1.5 Formule des probabilités composées

Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé fini. On considère des événements $\mathrm{A}_1, \ldots, \mathrm{A}_n$ de cet espace probabilisé. La probabilité de leur intersection est alors donnée par

$$\mathbb{P}(\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_n) = \mathbb{P}(\mathrm{A}_1) \times \mathbb{P}(\mathrm{A}_2|\mathrm{A}_1) \times \mathbb{P}(\mathrm{A}_3|\mathrm{A}_1 \cap \mathrm{A}_2) \times \cdots \times \mathbb{P}(\mathrm{A}_n|\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_{n-1}).$$

Cette formule, appelée *formule des probabilités composées*, nécessite que les événements par lesquels on conditionne aient tous une probabilité non nulle. On peut remarquer que cette hypothèse se simplifie en

$$\mathbb{P}(\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_{n-1}) > 0$$

car toutes les autres intersections contiennent celle-ci. On peut remarquer aussi que dans le cas contraire, la probabilité de l'intersection $\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_n$ est nulle.

Cette formule est particulièrement bien adaptée aux expériences où s'effectuent plusieurs choix successifs avec modification de l'environnement, comme par exemple les tirages sans remise.

Exemple-exercice. Un sac contient n boules rouges ainsi que n boules vertes et n boules bleues. On tire n boules dans ce sac, successivement et sans remise. On demande d'exprimer la probabilité d'avoir pioché les n boules rouges.

Solution. Pour tout k dans $[\![1, n]\!]$, on note R_k l'événement [La k-ième boule piochée est rouge.]. On cherche donc à calculer la probabilité de l'événement $\mathrm{R}_1 \cap \ldots \cap \mathrm{R}_n$ et on va le faire au moyen de la formule des probabilités composées. Déjà, la probabilité de l'événement R_1 vaut $n/(3n)$, que l'on ne simplifie pas en $1/3$ en prévision des calculs ultérieurs.

Soit k dans $[\![1, n-1]\!]$. Si l'on suppose que l'événement $\mathrm{R}_1 \cap \ldots \cap \mathrm{R}_k$ s'est produit, alors le sac contient $n - k$ boules rouges ainsi que n boules vertes et n boules bleues ; la probabilité de piocher une boule rouge dans ces conditions vaut

$$\mathbb{P}(\mathrm{R}_{k+1}|\mathrm{R}_1 \cap \ldots \cap \mathrm{R}_k) = \frac{n-k}{3n-k}.$$

La formule des probabilités composées donne maintenant

$$\mathbb{P}(\mathrm{R}_1 \cap \ldots \cap \mathrm{R}_n) = \mathbb{P}(\mathrm{R}_1) \times \prod_{k=1}^{n-1} \mathbb{P}(\mathrm{R}_{k+1}|\mathrm{R}_1 \cap \ldots \cap \mathrm{R}_k) = \frac{n}{3n} \times \prod_{k=1}^{n-1} \frac{n-k}{3n-k} = \prod_{k=0}^{n-1} \frac{n-k}{3n-k} = \frac{n(n-1)\cdots 1}{3n(3n-1)\cdots(2n+1)}$$
$$= \frac{n!(2n)!}{(3n)!} = \binom{3n}{n}^{-1}.$$

Remarque. Cette valeur s'obtient par une autre modélisation : piocher n boules successivement sans remise revient à les piocher toutes en même temps car on ne s'intéresse pas à l'ordre dans lequel on les a piochées. On peut ainsi considérer un univers à $\binom{3n}{n}$ éléments constitué de tous les tirages possibles, que l'on considère comme équiprobables. L'événement mentionné dans cet exercice est alors un événement élémentaire.

Un autre exemple de ce type de raisonnement est l'étude du temps d'attente du premier « pile » lors d'une suite de tirages à pile ou face. Cependant, cette expérience nécessite une modélisation infinie donc elle ne sera abordée que dans le cours de deuxième année.

1.6 Formules des probabilités totales

On considère un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et un système complet d'événements $\{A_1, \ldots, A_n\}$ de cet espace probabilisé.

Soit B un événement. On remarque que l'événement B se partitionne en

$$B = \bigcup_{k=1}^{n} (A_k \cap B)$$

et que les événements $A_k \cap B$ sont deux à deux incompatibles. La formule d'additivité finie donne donc

$$\mathbb{P}(B) = \sum_{k=1}^{n} \mathbb{P}(A_k \cap B).$$

C'est la première *formule des probabilités totales.*

Dans le cas où les événements A_k ont tous une probabilité strictement positive, on peut faire intervenir des probabilités conditionnelles, pour obtenir

$$\mathbb{P}(B) = \sum_{k=1}^{n} \mathbb{P}(B|A_k) \times \mathbb{P}(A_k).$$

C'est la deuxième *formule des probabilités totales.* Dans la pratique, on utilise principalement celle-ci. Il est d'ailleurs stipulé dans le programme officiel qu'il est permis d'écrire cette formule y compris dans le cas où certaines des probabilités $\mathbb{P}(A_k)$ est nulle ; dans ce cas, le produit $\mathbb{P}(B|A_k)\mathbb{P}(A_k)$ doit être considéré comme nul par convention même si le facteur $\mathbb{P}(B|A_k)$ n'est théoriquement pas défini.

Exemple-exercice. On fixe une constante p dans $]0, 1[$. Pendant N années, un militaire est affecté chaque année au hasard dans l'une des quatre villes A, B, C, D. Quand il est dans une ville donnée une certaine année, la probabilité qu'il y soit encore l'année suivante vaut p, les trois autres destinations étant équiprobables.

On suppose qu'il est initialement affecté dans la ville A.

Pour tout entier n compris entre 0 et $N - 1$, on note A_n l'événement [Le militaire est affecté à la ville A pour l'année n.] et on pose $a_n = \mathbb{P}(A_n)$. On définit de même des événements B_n, C_n, D_n et leur probabilité b_n, c_n, d_n.

a. Que valent a_0, b_0, c_0, d_0 ?

b. Que vaut $a_n + b_n + c_n + d_n$?

c. Pour tout entier n, exprimer a_{n+1} en fonction de a_n, b_n, c_n, d_n, puis en fonction de a_n seulement.

d. En déduire une expression de a_n en fonction de n. Faire de même avec b_n, c_n, d_n.

Solution. a. L'événement A_0 est certain. Les événements B_0, C_0, D_0 sont impossibles.

$$a_0 = 1 \qquad \text{et} \qquad b_0 = c_0 = d_0 = 0.$$

b. Pour l'année n, le militaire est dans exactement l'une des quatre villes mentionnées donc $\{A_n, B_n, C_n, D_n\}$ est un système complet d'événements. On en déduit la relation

$$a_n + b_n + c_n + d_n = 1.$$

c. On applique la formule des probabilités totales avec le système complet d'événements $\{A_n, B_n, C_n, D_n\}$.

$$\mathbb{P}(A_{n+1}) = \mathbb{P}_{A_n}(A_{n+1})\mathbb{P}(A_n) + \mathbb{P}_{B_n}(A_{n+1})\mathbb{P}(B_n) + \mathbb{P}_{C_n}(A_{n+1})\mathbb{P}(C_n) + \mathbb{P}_{D_n}(A_{n+1})\mathbb{P}(D_n).$$

D'après les hypothèses de l'énoncé, la probabilité de transition $\mathbb{P}_{A_n}(A_{n+1})$ vaut p et les trois autres probabilités de transitions sont égales entre elles, si bien qu'elles valent $\frac{1-p}{3}$. On obtient donc

$$a_{n+1} = pa_n + \frac{1-p}{3}(b_n + c_n + d_n).$$

On connaît la relation $b_n + c_n + d_n = 1 - a_n$. On obtient donc

$$a_{n+1} = pa_n + \frac{1-p}{3}(1 - a_n) = \frac{4p-1}{3}a_n + \frac{1-p}{3}.$$

d. La suite finie $(a_n)_{0\leqslant n\leqslant N-1}$ est arithmético-géométrique. L'équation

$$x = \frac{4p-1}{3}x + \frac{1-p}{3}$$

a pour solution $x = 1/4$. On en déduit la relation

$$a_{n+1} - \frac{1}{4} = \frac{4p-1}{3}\left(a_n - \frac{1}{4}\right).$$

La suite de terme général $a_n - \frac{1}{4}$ est donc géométrique, de raison $\frac{4p-1}{3}$. Son premier terme vaut $\frac{3}{4}$. Il vient

$$\forall n \in [\![0, N-1]\!], \qquad a_n = \frac{1}{4} + \frac{3}{4} \times \left(\frac{4p-1}{3}\right)^n.$$

Pour les autres probabilités, on trouve la même relation de récurrence par le même procédé. Seul le premier terme change. On trouve cette fois

$$\forall n \in [\![0, N-1]\!], \qquad b_n = c_n = d_n = \frac{1}{4} - \frac{1}{4} \times \left(\frac{4p-1}{3}\right)^n.$$

1.7 Événements indépendants

On considère un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Soient A et B deux événements. Dire que ces événements sont *indépendants* signifie qu'ils vérifient la relation

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).$$

Si $\mathbb{P}(A)$ n'est pas nul, ça équivaut à l'égalité $\mathbb{P}_A(B) = \mathbb{P}(B)$ (en quelque sorte, le fait que A soit vérifié ou non n'influe pas sur les chances de réalisation de l'événement B).

Propriété. Si A et B sont indépendants, alors A et $\overline{B}$ sont indépendants. Idem pour $\overline{A}$ et B. Idem pour $\overline{A}$ et $\overline{B}$.

Exemple. Considérons un lancer de deux dés, modélisé par l'univers $\Omega = [\![1, 6]\!]^2$ muni de la probabilité uniforme. Prenons les événements A =[Le premier dé a fait 6.] et B =[La somme des deux dés est paire.]

Nous avons déjà vu que la probabilité de B vaut 1/2. L'événement A s'écrit

$$A = \{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}.$$

Sa probabilité vaut donc $\mathbb{P}(A) = 6/36 = 1/6$. On trouve donc $\mathbb{P}(A) \times \mathbb{P}(B) = 1/12$.

L'événement $A \cap B$ s'écrit [Le premier dé a fait 6 et la somme des deux dés est paire.]. On peut remarquer qu'il se réécrit [Le premier dé a fait 6 et le deuxième dé a fait un résultat pair.]. Cet événement est donc

$$A \cap B = \{(6, 2), (6, 4), (6, 6)\}.$$

Sa probabilité vaut donc $\mathbb{P}(A \cap B) = 3/36 = 1/12$.

L'égalité $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$ est vérifiée : les événements A et B sont indépendants.

Généralisation. Prenons des événements $A_1, \ldots, A_n$. Dire que ces n événements sont *mutuellement indépendants* signifie que pour tout entier k compris entre 2 et n, pour tout choix $(i_1, \ldots, i_k)$ d'indices distincts entre 1 et n, la formule

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j})$$

est vérifiée.

Par exemple, dire que trois événements A, B, C sont indépendants signifie

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \qquad \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C), \qquad \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C), \qquad \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

Exercice/contre-exemple. On se place dans le cadre de l'exemple précédent et on reprend les événements A et B. On note C l'événement [Le deuxième dé a fait 3.].

Vérifier que les événements A, B, C sont deux à deux indépendants mais qu'ils ne sont pas mutuellement indépendants (en d'autres termes, vérifier que les trois premières relations ci-dessus sont valides mais pas la quatrième).

Propriété. Soient $A_1, \ldots, A_n$ des événements mutuellement indépendants. On considère des événements $B_1, \ldots, B_n$ où, pour chaque indice k, l'événement B_k est A_k ou son événement contraire $\overline{A_k}$.

Alors les événements $B_1, \ldots, B_n$ sont mutuellement indépendants.

2 Variables aléatoires

2.1 Définition

Une *variable aléatoire* sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est une fonction $\mathrm{X} : \Omega \to \mathbb{R}$.

Dans le cas où l'espace probabilisé modélise une expérience aléatoire, une variable aléatoire modélise une grandeur attachée à cette expérience.

Étant donné une variable aléatoire X, l'ensemble $\mathrm{X}(\Omega)$ des valeurs prises par cette fonction est appelé *univers image*. Pour toute valeur k réelle, l'événement

$$\mathrm{X}^{-1}(k) = \{\omega \in \Omega \ ; \ \mathrm{X}(\omega) = k\}$$

est noté indifféremment[3] $[\mathrm{X} = k]$ ou $(\mathrm{X} = k)$. Cet abus de notation est analogue à ce que l'on rencontre dans les calculs en physique-chimie, où l'on manipule des grandeurs plus que des fonctions.

Exemple 1. On considère un lancer de deux dés à six faces, modélisé par l'univers $\Omega = [\![1, 6]\!]^2$, que l'on munit de la probabilité uniforme.

Pour tout résultat ω de l'expérience, on note $\mathrm{X}_1(\omega)$ le résultat du premier dé et $\mathrm{X}_2(\omega)$ le résultat du deuxième dé. Cette définition se résume par

$$\forall \omega \in [\![1, 6]\!]^2, \qquad \omega = (\mathrm{X}_1(\omega), \mathrm{X}_2(\omega)).$$

Les univers images sont

$$\mathrm{X}_1(\Omega) = [\![1, 6]\!] \qquad \text{et} \qquad \mathrm{X}_2(\Omega) = [\![1, 6]\!].$$

L'événement $[\mathrm{X}_1 = 3]$ s'écrit

$$[\mathrm{X}_1 = 3] = \{(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)\}.$$

L'événement $[\mathrm{X}_2 = 2]$ s'écrit

$$[\mathrm{X}_2 = 2] = \{(1,2), (2,2), (3,2), (4,2), (5,2), (6,2)\}.$$

Pour tout résultat ω de l'expérience, on note $\mathrm{S}(\omega)$ la somme des deux résultats, c'est-à-dire

$$\mathrm{S}(\omega) = \mathrm{X}_1(\omega) + \mathrm{X}_2(\omega).$$

La variable aléatoire S possède alors l'univers image

$$\mathrm{S}(\Omega) = [\![2, 12]\!].$$

L'événement $[\mathrm{S} = 8]$ s'écrit

$$[\mathrm{S} = 8] = [\mathrm{X}_1 + \mathrm{X}_2 = 8] = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}.$$

Exemple 2. On effectue quatre lancers d'une pièce et on code l'ensemble des suites de résultats (pile/face) par l'univers $\{0; 1\}^4$. Pour tout résultat ω de l'univers et tout indice i dans $[\![1, 4]\!]$, on note $\mathrm{X}_i(\omega)$ l'issue du i-ième lancer et on note $\mathrm{S}(\omega)$ le nombre de lancers qui ont donné « pile ». On remarque alors les identités

$$\omega = (\mathrm{X}_1(\omega), \mathrm{X}_2(\omega), \mathrm{X}_3(\omega), \mathrm{X}_4(\omega)) \qquad \text{et} \qquad \mathrm{S}(\omega) = \sum_{i=1}^{4} \mathrm{X}_i(\omega).$$

L'événement $[\mathrm{S} = 3]$ s'écrit

$$[\mathrm{S} = 3] = \{(0,1,1,1), (1,0,1,1), (1,1,0,1), (1,1,1,0)\}$$

ou encore

$$\begin{aligned}
[\mathrm{S} = 3] = & \ ([\mathrm{X}_1 = 0] \cap [\mathrm{X}_2 = 1] \cap [\mathrm{X}_3 = 1] \cap [\mathrm{X}_3 = 1]) \ \cup \\
& \ ([\mathrm{X}_1 = 1] \cap [\mathrm{X}_2 = 0] \cap [\mathrm{X}_3 = 1] \cap [\mathrm{X}_3 = 1]) \ \cup \\
& \ ([\mathrm{X}_1 = 1] \cap [\mathrm{X}_2 = 1] \cap [\mathrm{X}_3 = 0] \cap [\mathrm{X}_3 = 1]) \ \cup \\
& \ ([\mathrm{X}_1 = 1] \cap [\mathrm{X}_2 = 1] \cap [\mathrm{X}_3 = 1] \cap [\mathrm{X}_3 = 0]).
\end{aligned}$$

3. On trouve aussi parfois la notation $\{\mathrm{X} = k\}$.

2.2 Loi d'une variable aléatoire

Propriété et définition. Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. On peut alors définir une probabilité $\mathbb{P}^{\mathrm{X}}$ sur l'univers fini $\mathrm{X}(\Omega)$ par la formule

$$\forall \mathrm{A} \in \mathcal{P}(\mathrm{X}(\Omega)), \qquad \mathbb{P}^{\mathrm{X}}(\mathrm{A}) = \mathbb{P}([\mathrm{X} \in \mathrm{A}]).$$

La probabilité $\mathbb{P}^{\mathrm{X}}$ est alors appelée *loi* de la variable aléatoire X.

Comme on l'a vu au paragraphe 1.3, pour définir une probabilité sur un ensemble fini, il suffit de se donner les probabilités des événements élémentaires. C'est ce que l'on fait en général pour les lois de variables aléatoires.

Description classique d'une loi de probabilité. Pour décrire une variable aléatoire finie X sur l'espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, on donne
- l'univers image $\mathrm{X}(\Omega)$;
- pour chaque élément x de l'univers image $\mathrm{X}(\Omega)$, la valeur $\mathbb{P}([\mathrm{X} = x])$.

Illustrons ce procédé sur les exemples du paragraphe précédent.

Exemple 1. La loi de la variable aléatoire X_1 est donnée par $\mathrm{X}_1(\Omega) = [\![1, 6]\!]$ et le tableau suivant

k	1	2	3	4	5	6
$\mathbb{P}([\mathrm{X}_1 = k])$	1/6	1/6	1/6	1/6	1/6	1/6

Les valeurs prises par X_1 ont toutes la même probabilité de réalisation. On dit que X_1 suit une loi *uniforme*. La variable aléatoire X_2 suit la même loi que X_1 — attention, cela ne signifie en aucun cas que X_2 prend la même valeur que X_1.

La loi de la variable aléatoire $\mathrm{S} = \mathrm{X}_1 + \mathrm{X}_2$ est donnée par $\mathrm{S}(\Omega) = [\![2, 12]\!]$ et le tableau

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}([\mathrm{S} = k])$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Exemple 2. On munit l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$ de la probabilité uniforme $\mathbb{P}$. La loi de chacun des X_i est alors uniforme, donnée par

$$\mathrm{X}_i(\Omega) = \{0; 1\} \qquad \text{et} \qquad \mathbb{P}([\mathrm{X}_i = 0]) = \mathbb{P}([\mathrm{X}_i = 1]) = \frac{1}{2}.$$

La loi de S est donnée par $\mathrm{S}(\Omega) = [\![0, 4]\!]$ et par le tableau

k	0	1	2	3	4
$\mathbb{P}([\mathrm{S} = k])$	1/16	4/16	6/16	4/16	1/16

Construction générale. Soit n dans $\mathbb{N}^*$. Soient $x_1, \dots, x_n$ des nombres réels deux à deux distincts. Soient $p_1, \dots, p_n$ des éléments de $[0, 1]$ dont la somme vaut 1. On a vu au paragraphe 1.3 qu'il existe une probabilité $\mathbb{P}$ sur l'ensemble $\{x_1, \dots, x_n\}$ vérifiant la propriété

$$\forall i \in [\![1, n]\!], \qquad \mathbb{P}(\{x_i\}) = p_i.$$

Cette probabilité peut alors être vue comme la loi d'une certaine variable aléatoire sur un certain espace probabilisé. Il suffit pour cela de prendre $\Omega = \{x_1, \dots, x_n\}$ muni de la probabilité $\mathbb{P}$ évoquée précédemment et de définir

$$\mathrm{X} : \omega \mapsto \omega$$

de Ω vers $\mathbb{R}$.

2.3 Structure algébrique

Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé fini. L'ensemble des variables aléatoires sur cet espace probabilisé est $\mathbb{R}^{\Omega}$. À ce titre, c'est un espace vectoriel réel de dimension $\mathrm{Card}(\Omega)$.

Notons $\Omega = \{\omega_1, \dots, \omega_n\}$. Pour tout i dans $[\![1, n]\!]$, on considère la variable aléatoire X_i définie par

$$\mathrm{X}_i(\omega_j) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon.} \end{cases}$$

La famille $(\mathrm{X}_1, \dots, \mathrm{X}_n)$ est alors la base canonique de $\mathbb{R}^{\Omega}$. La décomposition d'une variable aléatoire X quelconque dans cette base est

$$\mathrm{X} = \sum_{i=1}^{n} \mathrm{X}(\omega_i) \mathrm{X}_i.$$

Remarque. La variable aléatoire X_i est la fonction indicatrice $\mathbb{1}_{\{\omega_i\}}$.

2.4 Lois de référence

Loi dégénérée. Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Dire que la loi de X est *dégénérée* signifie qu'il existe x dans $\mathrm{X}(\Omega)$ tel que l'événement $[\mathrm{X} = x]$ soit presque certain

$$\mathbb{P}([\mathrm{X} = x]) = 1.$$

Les variables aléatoires constantes suivent une loi dégénérée mais ce ne sont pas les seules. C'est le cas aussi par exemple de variables aléatoires prenant plusieurs valeurs lorsqu'on remplace la probabilité $\mathbb{P}$ par une probabilité conditionnelle bien choisie. Ce point sera évoqué dans le paragraphe 2.5

Loi uniforme. Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Dire que la loi de X est *uniforme* signifie que les valeurs prises par X sont équiprobables

$$\forall x \in \mathrm{X}(\Omega), \qquad \mathbb{P}([\mathrm{X} = x]) = \frac{1}{\mathrm{Card}(\mathrm{X}(\Omega))}.$$

Cette loi est fréquemment employée pour modéliser le résultat d'un lancer de dé ou le tirage d'une boule de loterie.

Dans le cas où l'univers image est un intervalle d'entiers $[\![a, b]\!]$, on note cette loi $\mathscr{U}([\![a, b]\!])$.

Loi de Bernoulli [4]. Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Soit p dans $]0, 1[$. Dire que X suit la *loi de Bernoulli* $\mathscr{B}(p)$ signifie que sa loi est décrite par

$$\mathrm{X}(\Omega) = \{0; 1\}, \qquad \mathrm{P}([\mathrm{X} = 1]) = p, \quad \mathrm{P}([\mathrm{X} = 0]) = 1 - p.$$

Une telle variable aléatoire est généralement utilisée pour modéliser l'issue (succès ou échec) d'une expérience.

Remarque. Soit A un sous-ensemble de X. Notons $p = \mathbb{P}(\mathrm{A})$. La fonction indicatrice $\mathbb{1}_{\mathrm{A}}$ est alors une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ dont la loi est $\mathscr{B}(p)$.

Réciproquement, si X suit la loi $\mathscr{B}(p)$, alors cette variable aléatoire coïncide avec la fonction $\mathbb{1}_{\mathrm{A}}$ en choisissant l'événement $\mathrm{A} = [\mathrm{X} = 1]$.

Loi binomiale. Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Soit p dans $]0, 1[$. Soit n dans $\mathbb{N}^*$. Dire que X suit la loi binomiale $\mathscr{B}(n, p)$ signifie que sa loi est décrite par

$$\mathrm{X}(\Omega) = [\![0, n]\!] \qquad \text{et} \qquad \forall k \in [\![0, n]\!], \quad \mathbb{P}([\mathrm{X} = k]) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Une telle loi est généralement utilisée pour modéliser le nombre de succès lors d'une succession de n expériences identiques et indépendantes, chacune ayant une probabilité de succès égale à p. La pertinence de cette modélisation sera justifiée dans le paragraphe 4

2.5 Loi conditionnelle

Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Soit A un événement de probabilité non nulle.

La variable aléatoire X est alors encore une variable aléatoire sur l'espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P}_{\mathrm{A}})$. Relativement à la probabilité $\mathbb{P}_{\mathrm{A}}$, elle possède une nouvelle loi, décrite par la donnée des nombres $\mathbb{P}_{\mathrm{A}}([\mathrm{X} = x])$ lorsque x décrit l'ensemble $\mathrm{X}(\Omega)$. Cette loi est en général différente de la loi relative à la probabilité $\mathbb{P}$. On l'appelle *loi conditionnelle* de X sachant A.

Exemple 1. Prenons une variable aléatoire X quelconque et un nombre x pour lequel $\mathbb{P}([\mathrm{X} = x])$ est non nul. Notons A l'événement $[\mathrm{X} = x]$.

La loi conditionnelle de X sachant A est alors dégénérée car $\mathbb{P}_{\mathrm{A}}([\mathrm{X} = x])$ vaut 1.

Exemple 2. Plaçons-nous dans le cadre de l'exemple 2 des deux paragraphes précédents et notons $\mathrm{A} = [\mathrm{S} = 3]$.

Je laisse aux lecteurs le soin de vérifier que la loi conditionnelle de X_1 sachant A est la loi de Bernoulli $\mathscr{B}(3/4)$, alors que la loi (tout court) de X_1 est la loi de Bernoulli $\mathscr{B}(1/2)$.

Les lois conditionnelles seront surtout à l'honneur dans le paragraphe 3.

4. N'oublions pas que Bernoulli n'est pas une nouille !

2.6 Fonction de répartition

Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. La *fonction de répartition* de X est la fonction

$$\mathrm{F_X} \; : \; \begin{array}{rcl} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \mathbb{P}(\mathrm{X} \leqslant x). \end{array}$$

C'est une fonction en escalier croissante. Si l'on note m la plus petite valeur prise par X et M la plus grande, on remarque alors que $\mathrm{F_X}$ vaut 0 sur $]-\infty, m[$ et 1 sur $[\mathrm{M}, +\infty[$.

Les fonctions de répartitions seront surtout employées dans le paragraphe 3.

2.7 Espérance

Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. *L'espérance* de X est le nombre

$$\mathbb{E}(\mathrm{X}) = \sum_{x \in \mathrm{X}(\Omega)} x \times \mathbb{P}([\mathrm{X} = x]).$$

C'est la moyenne des valeurs prises par X pondérées par leurs probabilités de réalisation. Une autre formule est

$$\mathbb{E}(\mathrm{X}) = \sum_{\omega \in \Omega} \mathrm{X}(\omega)\mathbb{P}(\{\omega\}).$$

Calculons [5] l'espérance des variables aléatoires qui suivent les lois de référence.

Loi dégénérée. On suppose que X suit la loi dégénérée en x, ce qui s'écrit $\mathbb{P}([\mathrm{X} = x]) = 1$. L'espérance de X vaut alors x et c'est logique : on calcule la moyenne avec une seule valeur.

Loi uniforme. On suppose que X suit la loi uniforme sur l'ensemble $\mathrm{X}(\Omega) = \{x_1, \dots, x_n\}$. Son espérance vaut alors

$$\mathbb{E}(\mathrm{X}) = \frac{x_1 + \ldots + x_n}{n}.$$

C'est la *moyenne arithmétique* des valeurs prises par X.
Dans le cas particulier de la loi uniforme sur l'intervalle entier $[\![a, b]\!]$, on trouve

$$\mathbb{E}(\mathrm{X}) = \frac{a + b}{2}.$$

Loi de Bernoulli. On suppose que X suit la loi de Bernoulli $\mathscr{B}(p)$. L'espérance de X vaut alors p.

$$\mathbb{E}(\mathrm{X}) = 1 \times \mathbb{P}([\mathrm{X} = 1]) + 0 \times \mathbb{P}([\mathrm{X} = 0]) = p.$$

Loi binomiale. On suppose que X suit la loi binomiale $\mathscr{B}(n, p)$. L'espérance de X vaut alors np. Démontrons-le.

$$\mathbb{E}(\mathrm{X}) = \sum_{k=0}^{n} k \times \mathbb{P}([\mathrm{X} = k]) = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k}.$$

Prenons k dans $[\![1, n]\!]$. On remarque la simplification

$$k\binom{n}{k} = k \times \frac{n!}{k!(n-k)!} = \frac{n!}{(k-1)!(n-k)!} = n \times \frac{(n-1)!}{(k-1)!(n-k)!} = n\binom{n-1}{k-1}.$$

On obtient donc

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

On effectue le décalage $\ell = k - 1$.

$$\mathbb{E}(\mathrm{X}) = n \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell+1} (1-p)^{n-1-\ell}.$$

On met p en facteur et on reconnaît la formule du binôme.

$$\mathbb{E}(\mathrm{X}) = np \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell} (1-p)^{n-1-\ell} = np(p + 1 - p)^{n-1} = np.$$

Dans le paragraphe 4, nous verrons une manière plus rapide d'obtenir ce résultat.

5. Lorsque le calcul n'est pas détaillé, il est de la responsabilité des lecteurs de l'effectuer au brouillon.

2.8 Linéarité de l'espérance

Propriété. L'espérance est une forme linéaire sur l'espace vectoriel $\mathbb{R}^\Omega$ des variables aléatoires sur l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$.

Démonstration. Prenons deux variables aléatoires X et Y ainsi qu'un scalaire λ. Pas besoin de connaître la loi de la variable aléatoire $\mathrm{X}+\lambda\mathrm{Y}$ pour avoir son espérance

$$\mathbb{E}(\mathrm{X}+\lambda\mathrm{Y}) = \sum_{\omega\in\Omega}(\mathrm{X}+\lambda\mathrm{Y})(\omega)\mathbb{P}(\{\omega\}) = \sum_{\omega\in\Omega}((\mathrm{X}(\omega)+\lambda\mathrm{Y}(\omega))\mathbb{P}(\{\omega\}) = \sum_{\omega\in\Omega}\mathrm{X}(\omega)\mathbb{P}(\{\omega\})+\lambda\sum_{\omega\in\Omega}\mathrm{Y}(\Omega)\mathbb{P}(\{\omega\}) = \mathbb{E}(\mathrm{X})+\lambda\mathbb{E}(\mathrm{Y}).$$

Cette propriété sert dans énormément d'exercices.

Variables aléatoires centrées. Comme toutes les applications linéaires, l'espérance possède un noyau. Ses éléments sont les variables aléatoires d'espérance nulle, que l'on appelle *variables aléatoires centrées.*

L'image de l'espérance est $\mathbb{R}$ donc, par la formule du rang, on obtient

$$\dim(\mathrm{Ker}(\mathbb{E})) = \mathrm{Card}(\Omega) - 1.$$

2.9 Formule du transfert

Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. On considère une fonction f définie sur une partie de $\mathbb{R}$ qui contient l'univers image $\mathrm{X}(\Omega)$.

La fonction $\mathrm{Y} = f\circ\mathrm{X}$ est alors une variable aléatoire sur l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Par un abus de notation, on la note couramment $f(\mathrm{X})$.

La loi de la variable aléatoire $f(\mathrm{X})$ est facile à déterminer si la fonction f n'est pas trop tordue mais il se trouve que la connaissance de cette loi n'est pas utile si le seul but est de calculer l'espérance de la variable aléatoire $f(\mathrm{X})$.

Formule du transfert. L'espérance de $f(\mathrm{X})$ est donnée par la formule

$$\mathbb{E}(f(\mathrm{X})) = \sum_{x\in\mathrm{X}(\Omega)} f(x)\mathbb{P}([\mathrm{X}=x]).$$

Cette formule dit que l'espérance de $f(\mathrm{X})$ relativement à la probabilité $\mathbb{P}$ est égale à l'espérance de la variable aléatoire f sur l'espace probabilisé fini $(\mathrm{X}(\Omega), \mathcal{P}(\mathrm{X}(\Omega)), \mathbb{P}^{\mathrm{X}})$. On « transfère » un calcul relatif à la probabilité $\mathbb{P}$ vers un calcul relatif à la loi de X, qui est en général ce qu'on connaît le mieux.

Exemple 1. On suppose que X suit la loi de Bernoulli $\mathscr{B}(p)$. Pour n'importe quelle fonction f définie au moins en 0 et en 1, on obtient alors

$$\mathbb{E}(f(\mathrm{X})) = f(0)\mathbb{P}([\mathrm{X}=0]) + f(1)\mathbb{P}([\mathrm{X}=1]) = (1-p)f(0) + pf(1).$$

En particulier, pour tout entier m strictement positif, on obtient

$$\mathbb{E}(\mathrm{X}^m) = p.$$

Exemple 2. On suppose que X suit la loi uniforme $\mathscr{U}(\llbracket 1, n\rrbracket)$. Prenons f quelconque pour commencer

$$\mathbb{E}(f(\mathrm{X})) = \frac{f(1)+\cdots+f(n)}{n}.$$

Prenons maintenant la fonction $f : x \mapsto x^2$.

$$\mathbb{E}(\mathrm{X}^2) = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{(n+1)(2n+1)}{6}.$$

Prenons maintenant la fonction $f : x \mapsto \mathrm{e}^x$.

$$\mathbb{E}(\mathrm{e}^{\mathrm{X}}) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{e}^k = \frac{\mathrm{e}}{n}\times\frac{1-\mathrm{e}^n}{1-\mathrm{e}}.$$

Exemple 3. On suppose que X suit la loi uniforme $\mathscr{B}(n,p)$. Fixons t dans $\mathbb{R}$ et définissons $f : k \mapsto t^k$ sur $\mathrm{X}(\Omega)$. On obtient alors

$$\mathbb{E}(f(\mathrm{X})) = \mathbb{E}(t^{\mathrm{X}}) = \sum_{k=0}^{n} t^k\mathbb{P}([\mathrm{X}=k]) = \sum_{k=0}^{n}\binom{n}{k}(tp)^k(1-p)^{n-k} = (tp+1-p)^n.$$

2.10 Moments

Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Pour tout entier s, le *moment* d'ordre s de la variable aléatoire X est le nombre
$$m_s(\mathrm{X}) = \mathbb{E}(\mathrm{X}^s).$$

On définit également le *moment centré* d'ordre s de la variable aléatoire X par
$$\mu_s(\mathrm{X}) = \mathbb{E}((\mathrm{X} - \mathbb{E}(\mathrm{X}))^s).$$

Le moment centré d'ordre s de X est le moment d'ordre s de la variable centrée $\mathrm{X} - \mathbb{E}(\mathrm{X})$.

Variance. La *variance* de la variable aléatoire X est son moment centré d'ordre 2
$$\mathbb{V}(\mathrm{X}) = \mu_2(\mathrm{X}) = \mathbb{E}((\mathrm{X} - \mathbb{E}(\mathrm{X}))^2).$$

Formule de Huygens. La variance se réécrit
$$\mathbb{V}(\mathrm{X}) = \mathbb{E}(\mathrm{X}^2) - (\mathbb{E}(\mathrm{X}))^2.$$

La démonstration de cette formule est laissée en exercice. On va maintenant utiliser cette formule pour calculer la variance dans le cas des lois classiques.

Loi dégénérée. Dans le cas d'une loi dégénérée, la variable aléatoire $\mathrm{X} - \mathbb{E}(\mathrm{X})$ est nulle avec une probabilité de 1 donc
$$\mathbb{V}(\mathrm{X}) = 0$$
dans ce cas.

Loi uniforme. On suppose que X suit la loi uniforme sur l'intervalle entier $[\![a, b]\!]$. Sa variance vaut alors
$$\mathbb{V}(\mathrm{X}) = \frac{(b - a + 1)^2 - 1}{12}.$$

Le calcul est laissé en exercice. J'ajoute que cette formule n'est pas à connaître, contrairement aux deux exemples qui suivent.

Loi de Bernoulli. On suppose que X suit la loi de Bernoulli $\mathscr{B}(p)$. Sa variance vaut alors $p(1-p)$, comme le prouve le calcul qui suit.
$$\mathbb{V}(\mathrm{X}) = \mathbb{E}(\mathrm{X}^2) - \mathbb{E}(\mathrm{X})^2 = p - p^2 = p(1-p).$$

Loi binomiale. On suppose que X suit la loi binomiale $\mathscr{B}(n, p)$. Sa variance vaut alors $np(1-p)$. Voici un calcul qui aboutit à cette formule. On va faire apparaître une simplification analogue à celle du calcul de l'espérance. Pour cela, on calcule l'espérance de $\mathrm{X}(\mathrm{X}-1)$ en passant par la formule de transfert.
$$\mathbb{E}(\mathrm{X}(\mathrm{X}-1)) = \sum_{k=0}^{n} k(k-1)\mathbb{P}(\mathrm{X} = k) = \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k}.$$

On observe cette fois l'identité $k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}$, ce qui mène à
$$\mathbb{E}(\mathrm{X}(\mathrm{X}-1)) = n(n-1)\sum_{k=2}^{n} \binom{n-2}{k-2} p^k (1-p)^{n-k}.$$

On effectue le décalage d'indice $\ell = k - 2$.
$$\mathbb{E}(\mathrm{X}(\mathrm{X}-1)) = n(n-1)\sum_{\ell=0}^{n-2} \binom{n-2}{\ell} p^{\ell+2} (1-p)^{n-2-\ell} = n(n-1)p^2(p+1-p)^{n-2} = n(n-1)p^2.$$

On en déduit ensuite
$$\mathbb{E}(\mathrm{X}^2) = \mathbb{E}(\mathrm{X}(\mathrm{X}-1)) + \mathbb{E}(\mathrm{X}) = n(n-1)p^2 + np = np((n-1)p+1)$$

puis
$$\mathbb{V}(\mathrm{X}) = \mathbb{E}(\mathrm{X}^2) - \mathbb{E}(\mathrm{X})^2 = np((n-1)p + 1 - np) = np(1-p).$$

Un calcul plus direct sera présenté dans le paragraphe 4.

Règle de calcul pour la variance. Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Soient a et b deux constantes réelles. La variance de la variable aléatoire $a\mathrm{X} + b$ est alors donnée par

$$\mathbb{V}(a\mathrm{X} + b) = a^2 \mathbb{V}(\mathrm{X}).$$

Cette formule est laissée en exercice. Le fait que b ne joue aucun rôle ne doit pas surprendre : ce terme disparaît car on commence par centrer la variable.

2.11 Propriétés de positivité

Positivité de l'espérance. Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. On suppose que l'univers image de X est inclus dans $[0, +\infty[$. Alors l'espérance de X est positive.

De plus, si l'espérance de X est nulle, alors la loi de X est dégénérée en 0.

Démonstration. L'espérance de X est donnée par

$$\mathbb{E}(\mathrm{X}) = \sum_{x \in \mathrm{X}(\Omega)} x \times \mathbb{P}([\mathrm{X} = x]).$$

Dans cette somme, tous les termes sont positifs donc cette somme est positive.

Supposons maintenant que cette somme est nulle. On en déduit que tous ses termes sont nuls. Ainsi, pour toute valeur x strictement positive dans $\mathrm{X}(\Omega)$, le nombre $\mathbb{P}([\mathrm{X} = x])$ est nul. La seule valeur x pour laquelle $\mathbb{P}([\mathrm{X} = x])$ peut être non nul est donc $x = 0$.

On connaît aussi la formule
$$\sum_{x \in \mathrm{X}(\Omega)} \mathbb{P}([\mathrm{X} = x]) = 1.$$

On en déduit que le nombre $\mathbb{P}([\mathrm{X} = 0])$ vaut 1.

Positivité de la variance. Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. La variance de X est positive.

De plus, si la variance de X est nulle, alors la loi de X est dégénérée en $\mathbb{E}(\mathrm{X})$.

Ceci est un corollaire direct de la propriété précédente, où l'on remplace X par $(\mathrm{X} - \mathbb{E}(\mathrm{X}))^2$.

Écart-type. L'écart-type de la variable aléatoire X est le nombre

$$\sigma(\mathrm{X}) = \sqrt{\mathbb{V}(\mathrm{X})}.$$

L'écart-type est *positivement homogène* de degré 1, en ce sens qu'il vérifie l'identité

$$\sigma(a\mathrm{X} + b) = |a| \sigma(\mathrm{X})$$

lorsque a et b sont des constantes.

2.12 Inégalités de Markov et de Bienaymé-Tchebychev

Inégalité de Markov. Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. On suppose que l'univers image $\mathrm{X}(\Omega)$ est inclus dans $[0, +\infty[$.

Alors, pour tout λ dans $]0, +\infty[$, la probabilité $\mathbb{P}([\mathrm{X} \geqslant \lambda])$ vérifie la majoration

$$\mathbb{P}([\mathrm{X} \geqslant \lambda]) \leqslant \frac{\mathbb{E}(\mathrm{X})}{\lambda}.$$

Démonstration de l'inégalité de Markov. Soit $\lambda > 0$. On considère l'événement $\mathrm{A} = [\mathrm{X} \geqslant \lambda]$ et la variable aléatoire
$$\mathrm{Y} = \mathrm{X} - \lambda \mathbb{1}_{\mathrm{A}}.$$

Vérifions que cette variable aléatoire est à valeurs positives. Prenons ω dans Ω. Si ω est dans A, on trouve alors

$$\mathrm{Y}(\omega) = \mathrm{X}(\omega) - \lambda \geqslant 0.$$

Si ω n'est pas dans A, on trouve alors

$$\mathrm{Y}(\omega) = \mathrm{X}(\omega) \geqslant 0.$$

La variable aléatoire Y est donc positive. On en déduit que son espérance est positive. Or cette espérance s'écrit

$$\mathbb{E}(\mathrm{Y}) = \mathbb{E}(\mathrm{X}) - \lambda\mathbb{P}(\mathrm{A}) = \mathbb{E}(\mathrm{X}) - \lambda\mathbb{P}([\mathrm{X} \geqslant \lambda]).$$

Le fait que cette espérance soit positive et que λ soit strictement positif donne l'inégalité attendue. ♡

Cette inégalité n'est pas un attendu du programme mais je la mentionne car c'est une conséquence directe des propriétés des paragraphes précédentes et elle sert souvent, notamment pour démontrer la formule qui suit. De plus, le recours à une fonction indicatrice est une méthode qu'il est utile d'avoir vue.

Inégalité de Bienaymé-Tchebychev. Soit X une variable aléatoire, dont on note m l'espérance. Elle vérifie alors l'inégalité

$$\forall a > 0, \qquad \mathbb{P}(|\mathrm{X} - m| \geqslant a) \leqslant \frac{\mathbb{V}(\mathrm{X})}{a^2}.$$

Démonstration de l'inégalité de Bienaymé-Tchebychev. On applique l'inégalité de Markov en remplaçant X par $(\mathrm{X} - m)^2$ et en remplaçant λ par a^2. ♡

Cette inégalité signifie que la variance est un *indicateur de dispersion.* Plus la variance est petite, moins il est probable de s'écarter de la moyenne.

3 Couples de variables aléatoires

3.1 Loi du couple

On considère deux variables aléatoires X_1 et X_2 sur un même espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Chacune de ces deux variables aléatoires possède sa propre loi mais on souhaite généralement savoir comment ces deux variables interagissent entre elles.

Définition. La *loi conjointe* du couple de variables aléatoires $\mathrm{C} = (\mathrm{X}_1, \mathrm{X}_2)$ est la probabilité $\mathbb{P}^{\mathrm{C}}$ sur l'ensemble produit $\mathrm{X}_1(\Omega) \times \mathrm{X}_2(\Omega)$ définie par

$$\forall (x_1, x_2) \in \mathrm{X}_1(\Omega) \times \mathrm{X}_2(\Omega), \qquad \mathbb{P}^{\mathrm{C}}((x_1, x_2)) = \mathbb{P}([\mathrm{X}_1 = x_1] \cap [\mathrm{X}_2 = x_2]).$$

Dans la pratique, on n'utilise pas la notation $\mathbb{P}^{\mathrm{C}}$ et on se contente de donner les nombres $\mathbb{P}([\mathrm{X}_1 = x_1] \cap [\mathrm{X}_2 = x_2])$.

Exemple 1. Reprenons l'exemple 1 des paragraphes 2.1 et 2.2 (les variables aléatoires X_1 et X_2 sont les valeurs affichées par deux dés à six faces).

Dans la modélisation choisie, on obtient

$$\forall (x_1, x_2) \in [\![1, 6]\!]^2, \qquad \mathbb{P}([\mathrm{X}_1 = x_1] \cap [\mathrm{X}_2 = x_2]) = \frac{1}{36}.$$

On dit que le couple $(\mathrm{X}_1, \mathrm{X}_2)$ suit une loi uniforme sur le produit $[\![1, 6]\!] \times [\![1, 6]\!]$.

Exemple 2. On se place dans le même cadre et on pose

$$\mathrm{X} = \min(\mathrm{X}_1, \mathrm{X}_2) \qquad \text{et} \qquad \mathrm{Y} = \max(\mathrm{X}_1, \mathrm{X}_2).$$

On remarque que X et Y ont tous deux $[\![1, 6]\!]$ pour univers image. Prenons donc x et y dans $[\![1, 6]\!]$.

Notons déjà que dans le cas $x > y$, l'événement $[\mathrm{X} = x] \cap [\mathrm{Y} = y]$ est impossible car l'inégalité $\mathrm{X} \leqslant \mathrm{Y}$ est certaine. La probabilité de cet événement est donc nulle.

Faisons maintenant l'hypothèse $x = y$. On obtient alors l'égalité

$$[\mathrm{X} = x] \cap [\mathrm{Y} = y] = [\mathrm{X}_1 = x] \cap [\mathrm{X}_2 = x]$$

donc cet événement a une probabilité égale à $1/36$.

Faisons enfin l'hypothèse $x < y$. On obtient alors l'égalité

$$[\mathrm{X} = x] \cap [\mathrm{Y} = y] = ([\mathrm{X}_1 = x] \cap [\mathrm{X}_2 = y]) \cup ([\mathrm{X}_1 = y] \cap [\mathrm{X}_2 = x]).$$

Les événements $[\mathrm{X}_1 = x] \cap [\mathrm{X}_2 = y]$ et $[\mathrm{X}_1 = y] \cap [\mathrm{X}_2 = x]$ sont incompatibles donc

$$\mathbb{P}([\mathrm{X} = x] \cap [\mathrm{Y} = y]) = \mathbb{P}([\mathrm{X}_1 = x] \cap [\mathrm{X}_2 = y]) + \mathbb{P}([\mathrm{X}_1 = y] \cap [\mathrm{X}_2 = x]) = \frac{2}{36} = \frac{1}{18}.$$

La loi du couple (X, Y) est donc résumée par

$$\forall (x,y) \in [\![1,6]\!] \times [\![1,6]\!], \qquad \mathbb{P}([\mathrm{X} = x] \cap [\mathrm{Y} = y]) = \begin{cases} 0 & \text{si } x > y \\ 1/36 & \text{si } x = y \\ 1/18 & \text{si } x < y. \end{cases}$$

Souvent, l'expérience aléatoire étudiée fait intervenir une temporalité dans laquelle X_2 se décide après X_1. Dans un tel cas, on connaît pour X_2 ses lois conditionnelles relativement aux événements de la forme $[\mathrm{X}_1 = x_1]$ et on peut alors reconstituer la loi du couple. On va en voir un exemple.

Exercice résolu. Un sac contient n boules numérotées de 1 à n. On en pioche une et sa valeur est notée X_1. On retire alors du sac toutes les boules de numéro strictement supérieur à X_1 et on garde toutes les autres (en particulier, on remet la première boule piochée). On pioche ensuite une deuxième boule et sa valeur est notée X_2.

On demande la loi conjointe du couple $(\mathrm{X}_1, \mathrm{X}_2)$.

Solution. On modélise ce problème en décidant que la variable aléatoire X_1 suit la loi uniforme sur $[\![1,n]\!]$ et que pour tout x_1 dans $[\![1,n]\!]$, la loi de X_2 sachant $[\mathrm{X}_1 = x_1]$ est la loi uniforme sur l'intervalle entier $[\![1,x_1]\!]$.

On obtient alors $\mathrm{X}_1(\Omega) = [\![1,n]\!]$ et $\mathrm{X}_2(\Omega) = [\![1,n]\!]$ et

$$\forall (x_1,x_2) \in [\![1,n]\!] \times [\![1,n]\!], \qquad \mathbb{P}([\mathrm{X}_1 = x_1] \cap [\mathrm{X}_2 = x_2]) = \mathbb{P}([\mathrm{X}_1 = x_1])\mathbb{P}_{[\mathrm{X}_1 = x_1]}([\mathrm{X}_2 = x_2]) = \begin{cases} \frac{1}{n} \times \frac{1}{x_1} & \text{si } x_2 \leqslant x_1 \\ 0 & \text{si } x_2 > x_1. \end{cases}$$

3.2 Lois marginales

La loi conjointe d'un couple $(\mathrm{X}_1, \mathrm{X}_2)$ permet de reconstituer les lois individuelles des deux variables aléatoires X_1 et X_2. Ces lois individuelles sont aussi appelées *lois marginales*. Pour obtenir la loi de X_2, on applique la formule des probabilités totales au système complet d'événements constitué des événements de la forme $[\mathrm{X}_1 = x_1]$.

$$\mathbb{P}([\mathrm{X}_2 = x_2]) = \sum_{x_1 \in \mathrm{X}_1(\Omega)} \mathbb{P}([\mathrm{X}_1 = x_1] \cap [\mathrm{X}_2 = x_2]).$$

Exemple. Plaçons-nous dans le cadre de l'exercice résolu du paragraphe précédent. On applique la formule des probabilités totales avec le système complet d'événéments $\{[\mathrm{X}_1 = x_1]\ ;\ x_1 \in [\![1,n]\!]\}$.

$$\forall x_2 \in [\![1,n]\!], \qquad \mathbb{P}([\mathrm{X}_2 = x_2]) = \sum_{x_1=1}^{n} \mathbb{P}([\mathrm{X}_1 = x_1] \cap [\mathrm{X}_2 = x_2]) = \sum_{x_1=x_2}^{n} \frac{1}{n} \times \frac{1}{x_1}.$$

3.3 Formule du transfert

On considère un couple $(\mathrm{X}_1, \mathrm{X}_2)$ de variables aléatoires définies sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ainsi qu'une fonction f de deux variables telle que la variable $\mathrm{Y} = f(\mathrm{X}_1, \mathrm{X}_2)$ soit bien définie.

Il est alors possible de calculer l'espérance de Y sans déterminer sa loi

$$\mathbb{E}(f(\mathrm{X}_1, \mathrm{X}_2)) = \sum_{x_1 \in \mathrm{X}_1(\Omega)} \sum_{x_2 \in \mathrm{X}_2(\Omega)} f(x_1,x_2)\mathbb{P}([\mathrm{X}_1 = x_1] \cap [\mathrm{X}_2 = x_2]).$$

Exemple. Plaçons-nous dans le cadre de l'exercice résolu du paragraphe 3.1 et calculons l'espérance de la variable aléatoire X_1X_2.

$$\begin{aligned}\mathbb{E}(X_1X_2) &= \sum_{x_1=1}^{n}\sum_{x_2=1}^{n} x_1x_2\mathbb{P}([X_1 = x_1] \cap [X_2 = x_2]) = \sum_{x_1=1}^{n}\sum_{x_2=1}^{x_1} x_1x_2 \times \frac{1}{nx_1} = \sum_{x_1=1}^{n}\sum_{x_2=1}^{x_1} \frac{x_2}{n} \\ &= \frac{1}{n}\sum_{x_1=1}^{n} \frac{x_1(x_1+1)}{2} = \frac{1}{2n}\sum_{x_1=1}^{n}(x_1^2 + x_1) = \frac{1}{2n}\left(\frac{n(n+1)(2n+1)}{6} + \frac{n(n+1)}{2}\right) \\ &= \frac{n+1}{2}\left(\frac{2n+1}{6} + \frac{1}{2}\right) = \frac{(n+1)(n+2)}{6}.\end{aligned}$$

3.4 Covariance

La *covariance* d'un couple (X_1, X_2) de variables aléatoires est le nombre

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}\big((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))\big).$$

On peut remarquer l'identité

$$\mathrm{Cov}(X, X) = \mathbb{V}(X)$$

ainsi qu'une variante de l'identité de Huygens

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2).$$

La covariance est une forme bilinéaire symétrique positive [6] sur l'espace vectoriel $\mathbb{R}^\Omega$. À ce titre, elle possède une inégalité de Cauchy-Schwarz

$$|\mathrm{Cov}(X_1, X_2)|^2 \leqslant V(X_1)V(X_2).$$

Il existe un cas d'égalité mais je n'en parlerai pas ici.

Propriété. Pour tout couple (X_1, X_2) de variables aléatoires, la variance de la somme $X_1 + X_2$ est donnée par

$$\mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2) + 2\,\mathrm{Cov}(X_1, X_2).$$

Cette formule s'obtient en développant directement $\mathrm{Cov}(X_1+X_2, X_1+X_2)$ et en exploitant la linéarité de l'espérance. Cette formule est analogue à la formule classique des produits scalaires

$$||x_1 + x_2||^2 = ||x_1||^2 + ||x_2||^2 + 2(x_1|x_2)$$

et pour cause : sa démonstration est la même.

3.5 Couple de variables aléatoires indépendantes

Dire que deux variables aléatoires X_1 et X_2 sont *indépendantes* signifie que pour tout élément x_1 de $X_1(\Omega)$ et tout élément x_2 de $X_2(\Omega)$, les événements $[X_1 = x_1]$ et $[X_2 = x_2]$ sont indépendants. En formule, cela s'écrit

$$\forall x_1 \in X_1(\Omega), \quad \forall x_2 \in X_2(\Omega), \qquad \mathbb{P}([X_1 = x_1] \cap [X_2 = x_2]) = \mathbb{P}([X_1 = x_1])\mathbb{P}([X_2 = x_2]).$$

On peut remarquer que si x_1 est pris ailleurs que dans $X_1(\Omega)$ ou x_2 est pris en dehors de $X_2(\Omega)$, alors la formule ci-dessus est automatiquement vraie car les deux membres de l'égalité sont nuls. La définition ci-dessus peut donc se réécrire

$$\forall (x_1, x_2) \in \mathbb{R}^2, \qquad \mathbb{P}([X_1 = x_1] \cap [X_2 = x_2]) = \mathbb{P}([X_1 = x_1])\mathbb{P}([X_2 = x_2]).$$

Exemple. Reprenons les variables aléatoires du paragraphe 3.1. On observe que les variables aléatoires X_1 et X_2 sont indépendantes mais que les variables X et Y ne le sont pas car, par exemple

$$\mathbb{P}([X = 6] \cap [Y = 1]) = 0 \neq \mathbb{P}([X = 6])\mathbb{P}([Y = 1]).$$

Propriété. Soit (X_1, X_2) un couple de variables aléatoires indépendantes. Alors, pour tout couple (A_1, A_2) de parties de $\mathbb{R}$, les événements $[X_1 \in A_1]$ et $[X_2 \in A_2]$ sont indépendants.

6. Mais pas définie positive.

Propriété. Soient f_1 et f_2 deux fonctions définies respectivement sur $X_1(\Omega)$ et $X_2(\Omega)$. Si X_1 et X_2 sont indépendantes, alors les variables aléatoires $f_1(X_1)$ et $f_2(X_2)$ sont alors indépendantes.

Théorème. Si X_1 et X_2 sont des variables aléatoires indépendantes, alors l'espérance de leur produit est donnée par

$$\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2).$$

En d'autres termes, la covariance de X_1 et X_2 est nulle dans ce cas.

Démonstration du théorème. Appliquons la formule du transfert

$$\mathbb{E}(X_1X_2) = \sum_{x_1\in X_1(\Omega)}\sum_{x_2\in X_2(\Omega)} x_1x_2\mathbb{P}([X_1 = x_1]\cap[X_2 = x_2]).$$

Utilisons l'indépendance des deux variables

$$\mathbb{E}(X_1X_2) = \sum_{x_1\in X_1(\Omega)}\sum_{x_2\in X_2(\Omega)} x_1x_2\mathbb{P}([X_1 = x_1])\mathbb{P}([X_2 = x_2]).$$

On reconnaît alors le produit de deux sommes.

$$\mathbb{E}(X_1X_2) = \left(\sum_{x_1\in X_1(\Omega)} x_1\mathbb{P}([X_1 = x_1])\right)\left(\sum_{x_2\in X_2(\Omega)} x_2\mathbb{P}([X_2 = x_2])\right) = \mathbb{E}(X_1)\mathbb{E}(X_2).$$

Le théorème est démontré. ♡

Corollaire. On suppose que X_1 et X_2 sont indépendantes. La variance de leur somme est alors donnée par

$$\mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2).$$

Cette formule peut être considérée comme une sorte de formule de Pythagore, dans la mesure où une covariance nulle ressemble à une orthogonalité.

Attention. Comme on peut s'en douter, la réciproque est fausse : l'égalité $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)$ peut être vraie sans que les variables aléatoires ne soient indépendantes. On va le voir dans le contre-exemple qui suit.

Contre-exemple de la réciproque. On se donne deux boîtes, appelées A et B, ainsi que deux boules. Chaque boule est placée dans une boîte choisie au hasard de manière équiprobable (et indépendante).

On note X le nombre de boîtes vides à la fin de l'expérience et on note Y le nombre de boules dans la boîte A. Les univers images sont

$$X(\Omega) = [\![0, 1]\!] \qquad \text{et} \qquad Y(\Omega) = [\![0, 2]\!]$$

et la loi conjointe est donnée par

X \ Y	0	1	2
0	0	1/2	0
1	1/4	0	1/4

À partir de ces données, on trouve les lois marginales de X et Y

x	0	1
$\mathbb{P}([X = x])$	1/2	1/2

et

y	0	1	2
$\mathbb{P}([Y = y])$	1/4	1/2	1/4

On trouve alors les valeurs

$$\mathbb{E}(X) = \frac{1}{2}, \qquad \mathbb{E}(Y) = 1, \qquad \mathbb{E}(XY) = 1 \times 2 \times \frac{1}{4} = \frac{1}{2} = \mathbb{E}(X)\mathbb{E}(Y)$$

mais les variables X et Y ne sont pas indépendantes

$$\mathbb{P}([X = 0]\cap[Y = 0]) = 0 \qquad \text{mais} \qquad \mathbb{P}([X = 0])\mathbb{P}([Y = 0]) = \frac{1}{8}.$$

3.6 Couple de variables aléatoires binomiales indépendantes

On fixe p dans $]0,1[$ et on prend deux entiers strictement positifs n et m. On se donne deux variables aléatoires X et Y sur un même espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. On suppose que les lois de X et Y sont respectivement $\mathcal{B}(n,p)$ et $\mathcal{B}(m,p)$. On suppose enfin que ces deux variables aléatoires sont indépendantes.

Sous ces conditions, la variable aléatoire S égale à $\mathrm{X}+\mathrm{Y}$ suit la loi binomiale $\mathcal{B}(n+m,p)$.

Démonstration. Remarquons déjà que l'univers image $\mathrm{S}(\Omega)$ est inclus dans $[\![0,n+m]\!]$. Prenons maintenant un entier k dans $[\![0,n+m]\!]$.

Appliquons la formule des probabilités totales avec le système complet d'événements $\{[\mathrm{X}=x]\ ;x\in[\![0,n]\!]\}$.

$$\mathbb{P}([\mathrm{S}=k]) = \sum_{x=0}^{n} \mathbb{P}([\mathrm{S}=k]\cap[\mathrm{X}=x]).$$

Remarquons maintenant l'égalité

$$[\mathrm{S}=k]\cap[\mathrm{X}=x] = [\mathrm{X}+\mathrm{Y}=k]\cap[\mathrm{X}=x] = [\mathrm{X}=x]\cap[\mathrm{Y}=k-x].$$

En exploitant l'indépendance de X et Y, on obtient maintenant

$$\mathbb{P}([\mathrm{S}=k]) = \sum_{x=0}^{n} \mathbb{P}([\mathrm{X}=x]\cap[\mathrm{Y}=k-x]) = \sum_{x=0}^{n} \mathbb{P}([\mathrm{X}=x])\mathbb{P}([\mathrm{Y}=k-x]).$$

Si l'indice x dépasse strictement k, alors le nombre $\mathbb{P}([\mathrm{Y}=k-x])$ est nul.

De même, si l'indice x dépasse strictement n, alors le nombre $\mathbb{P}([\mathrm{X}=x])$ est nul. On peut donc réécrire la somme ci-dessus sous la forme

$$\mathbb{P}([\mathrm{S}=k]) = \sum_{x=0}^{k} \mathbb{P}([\mathrm{X}=x])\mathbb{P}([\mathrm{Y}=k-x]) = \sum_{x=0}^{k} \binom{n}{x} p^x(1-p)^{n-x}\binom{m}{k-x}p^{k-x}(1-p)^{m-k+x} = \sum_{x=0}^{k}\binom{n}{x}\binom{m}{k-x}p^k(1-p)^{n+m-k}.$$

On peut mettre en facteur ce qui ne dépend pas de l'indice x. On conclut enfin en exploitant la formule de Vandermonde [7].

$$\mathbb{P}([\mathrm{S}=k]) = p^k(1-p)^{n+m-k}\sum_{x=0}^{k}\binom{n}{x}\binom{m}{k-x} = p^k(1-p)^{n+m-k}\binom{n+m}{k}.$$

On constate qu'effectivement, la loi de S est la loi binomiale $\mathcal{B}(n+m,p)$. ♡

4 Vecteurs aléatoires

4.1 Loi conjointe

Dans ce paragraphe, on fixe un entier n supérieur ou égal à 2. On considère un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et on se donne n variables $\mathrm{X}_1, \dots, \mathrm{X}_n$ sur cet espace probabilisé.

La liste de fonctions $\overrightarrow{\mathrm{X}} = (\mathrm{X}_1, \dots, \mathrm{X}_n)$ est appelée un *vecteur aléatoire* à valeurs dans $\mathbb{R}^n$. La *loi conjointe* de ce vecteur aléatoire est la probabilité $\mathbb{P}^{\overrightarrow{\mathrm{X}}}$ définie sur le produit $\mathrm{X}_1(\Omega)\times\cdots\times\mathrm{X}_n(\Omega)$ par

$$\mathbb{P}^{\overrightarrow{\mathrm{X}}}((x_1,\dots,x_n)) = \mathbb{P}([\mathrm{X}_1=x_1]\cap\ldots\cap[\mathrm{X}_n=x_n]).$$

Comme on le voit, cette définition ne fait que généraliser la définition de la loi des couples. Dans la pratique, on s'intéressera principalement aux variables aléatoires indépendantes.

4.2 Indépendance

On considère n variables aléatoires $\mathrm{X}_1, \dots, \mathrm{X}_n$ sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Dire que ces variables aléatoires sont *mutuellement indépendantes* signifie qu'elles vérifient la propriété

$$\forall (x_1,\dots,x_n)\in\mathrm{X}_1(\Omega)\times\cdots\times\mathrm{X}_n(\Omega), \qquad \mathbb{P}([\mathrm{X}_1=x_1]\cap\ldots\cap[\mathrm{X}_n=x_n]) = \prod_{k=1}^{n}\mathbb{P}([\mathrm{X}_k=x_k]).$$

7. Cette identité n'est pas une formule du cours mais un exercice classique de dénombrement, que nous aurons l'occasion de voir en classe.

L'indépendance est en général une hypothèse de modélisation pour des expériences où les issues des différentes phases sont réputées ne pas avoir d'influence les unes sur les autres. Typiquement, lors d'une série de tirages à pile ou face, les valeurs obtenues aux tirages successifs définissent des variables aléatoires qu'il est usuel de supposer mutuellement indépendantes.

Propriétés.

1. Si $X_1, \dots, X_n$ sont mutuellement indépendantes, alors, pour toutes parties $A_1, \dots, A_n$ de $X_1(\Omega), \dots, X_n(\Omega)$ respectivement, les événements $[X_1 \in A_1], \dots, [X_n \in A_n]$ sont mutuellement indépendants.
2. Si $X_1, \dots, X_n$ sont mutuellement indépendantes, alors toute sous-famille de $(X_1, \dots, X_n)$ est constituée de variables aléatoires mutuellement indépendantes.
3. Si $X_1, \dots, X_n$ sont mutuellement indépendantes, alors, pour tout k dans $[\![1, n-1]\!]$, pour toute fonction f convenable de k variables et toute fonction g convenable de $n-k$ variables, les variables aléatoires $f(X_1, \dots, X_k)$ et $g(X_{k+1}, \dots, X_n)$ sont indépendantes.
4. Si $X_1, \dots, X_n$ sont mutuellement indépendantes, alors la variance de leur somme est donnée par
$$\mathbb{V}(X_1 + \cdots + X_n) = \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n).$$

Attention. Il serait faux de croire que si des variables aléatoires sont deux à deux indépendantes, alors elles sont mutuellement indépendantes.

Considérons une fois de plus les variables aléatoires X_1 et X_2 qui modélisent les résultats de deux dés lancés indépendamment et notons X_3 le reste de la division de $X_1 + X_2$ par 2. En d'autres termes, la variable aléatoire X_3 est la fonction indicatrice de l'événement [La somme $X_1 + X_2$ est impaire.].

On peut alors vérifier que les variables aléatoires X_1, X_2, X_3 sont deux à deux indépendantes mais qu'elles ne le sont pas mutuellement. Par exemple,
$$\mathbb{P}([X_1 = 1] \cap [X_2 = 1] \cap [X_3 = 1]) = 0 \qquad \text{mais} \qquad \mathbb{P}([X_1 = 1])\mathbb{P}([X_2 = 1])\mathbb{P}([X_3 = 1]) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{2} \neq 0.$$

4.3 Cas des variables de Bernoulli

Soient n dans $\mathbb{N}^*$ et p dans $]0, 1[$. On considère des variables aléatoires $X_1, \dots, X_n$ sur un même espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. On suppose que ces variables sont mutuellement indépendantes et qu'elles suivent toutes la loi de Bernoulli $\mathcal{B}(p)$.

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

Démonstration. Pour tout entier r compris entre 1 et n, on pose
$$S_r = X_1 + \cdots + X_r$$
et on note β_r l'assertion « La loi de S_r est $\mathcal{B}(r, p)$. ».

On sait déjà que l'assertion β_1 est vraie car la loi binomiale $\mathcal{B}(1, p)$ est identique à la loi de Bernoulli $\mathcal{B}(p)$.

Prenons maintenant un entier r dans $[\![1, n-1]\!]$ pour lequel l'assertion β_r est vraie.

On sait que les variables aléatoires $X_1, \dots, X_r, X_{r+1}$ sont mutuellement indépendantes. On en déduit que les variables aléatoires $X_1 + \cdots + X_r$ et X_{r+1} sont indépendantes. En d'autres termes, les variables aléatoires S_r et X_{r+1} sont indépendantes.

Ces deux variables aléatoires suivent respectivement les lois $\mathcal{B}(r, p)$ et $\mathcal{B}(1, p)$ donc leur somme suit la loi binomiale $\mathcal{B}(r+1, p)$ d'après la propriété démontrée au paragraphe 3.6. En d'autres termes, la variable aléatoire S_{r+1} suit la loi binomiale $\mathcal{B}(r+1, p)$. L'assertion β_{r+1} est donc vraie.

Par récurrence finie, les assertions $\beta_1, \dots, \beta_n$ sont toutes vraies. La propriété annoncée est démontrée. ♡

Cette démonstration justifie le fait que la loi binomiale $\mathcal{B}(n, p)$ modélise le nombre de succès lors d'une série de n expériences identiques et indépendantes ayant chacune une probabilité de succès égale à p.

Corollaire. On retrouve les valeurs
$$\mathbb{E}(S_r) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = p + \cdots + p = np$$
et
$$\mathbb{V}(S_r) = \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n) = p(1-p) + \cdots + p(1-p) = np(1-p).$$

Le premier calcul utilise simplement la linéarité de l'espérance. Le deuxième calcul utilise l'indépendance mutuelle des X_k.

4.4 Complément : un exemple d'application de l'inégalité de Bienaymé-Tchebychev

On considère n variables aléatoires indépendantes de même loi, notées $X_1, \ldots, X_n$. On note m leur valeur commune et σ leur écart-type commun.

Alors, pour tout $a > 0$, l'inégalité de Bienaymé-Tchebychev appliquée à la variable aléatoire

$$M_n = \frac{X_1 + \cdots + X_n}{n}$$

donne

$$\mathbb{P}(|M_n - m| \geqslant a) \leqslant \frac{\sigma^2}{a^2} \times \frac{1}{n}.$$

Une interprétation de cette majoration est que si l'on effectue un grand nombre de mesures d'une même grandeur, alors leur moyenne (appelée moyenne *empirique*) est probablement proche de l'espérance de cette grandeur (qui est leur moyenne *théorique*).

5 Complément : espérance conditionnelle

Le contenu de ce paragraphe n'est pas à notre programme mais j'ai dans l'idée que ce sera un dépassement de programme récurrent dans les épreuves de concours.

5.1 Espérance conditionnelle

On considère une variable aléatoire X définie sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ainsi qu'un événement A de probabilité non nulle.

La variable aléatoire X possède alors une loi conditionnelle sachant A, qui est sa loi relativement à la probabilité $\mathbb{P}_A$ sur Ω. À ce titre, cette variable aléatoire possède une espérance relative à cette probabilité, qui est définie par

$$\mathbb{E}(X|A) = \sum_{x \in X(\Omega)} x \mathbb{P}_A([X = x]) = \sum_{x \in X(\Omega)} x \mathbb{P}([X = x]|A).$$

Exemple. Plaçons-nous dans le cadre de l'exercice résolu du paragraphe 3.1. Prenons x_1 dans $[\![1, n]\!]$. On sait que la loi de X_2 sachant $[X_1 = x_1]$ est la loi uniforme sur l'intervalle entier $[\![1, x_1]\!]$; son espérance est donc

$$\mathbb{E}(X_2|[X_1 = x_1]) = \frac{1 + x_1}{2}.$$

5.2 Formule de l'espérance totale

On considère une variable aléatoire X définie sur un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ainsi qu'un système complet d'événements $(A_1, \ldots, A_n)$.

La *formule de l'espérance totale* s'écrit

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X|A_i)\mathbb{P}(A_i),$$

où il s'entend que le produit $\mathbb{E}(X|A_i)\mathbb{P}(A_i)$ est nul si le nombre $\mathbb{P}(A_i)$ est nul, même si l'espérance conditionnelle n'est pas définie dans ce cas.

Démonstration de la formule de l'espérance totale. Notons $x_1, \ldots, x_r$ les valeurs prises par X. La somme de droite dans la formule s'écrit

$$\sum_{i=1}^{n} \mathbb{E}(X|A_i)\mathbb{P}(A_i) = \sum_{i=1}^{n} \sum_{k=1}^{r} x_k \mathbb{P}_{A_i}([X = x_k])\mathbb{P}(A_i).$$

On peut permuter les deux symboles de sommation.

$$\sum_{i=1}^{n} \mathbb{E}(X|A_i)\mathbb{P}(A_i) = \sum_{k=1}^{r} \sum_{i=1}^{n} x_k \mathbb{P}_{A_i}([X = x_k])\mathbb{P}(A_i).$$

Le facteur x_k ne dépend pas de l'indice i donc on peut le mettre en facteur.

$$\sum_{i=1}^{n} \mathbb{E}(\mathrm{X}|\mathrm{A}_i)\mathbb{P}(\mathrm{A}_i) = \sum_{k=1}^{r} x_k \sum_{i=1}^{n} \mathbb{P}_{\mathrm{A}_i}([\mathrm{X} = x_k])\mathbb{P}(\mathrm{A}_i).$$

La formule des probabilités totales donne alors

$$\sum_{i=1}^{n} \mathbb{E}(\mathrm{X}|\mathrm{A}_i)\mathbb{P}(\mathrm{A}_i) = \sum_{k=1}^{r} x_k \sum_{i=1}^{n} \mathbb{P}_{\mathrm{A}_i}([\mathrm{X} = x_k])\mathbb{P}(\mathrm{A}_i) = \sum_{k=1}^{r} x_k \mathbb{P}([\mathrm{X} = x_k]) = \mathbb{E}(\mathrm{X}).$$

La formule de l'espérance totale est démontrée. ♡

Exemple. Reprenons le cadre de l'exercice résolu du paragraphe 3.1 avec le système complet d'événements

$$\{[\mathrm{X}_1 = x_1]\ ;\ 1 \leqslant x_1 \leqslant n\}$$

La formule de l'espérance totale donne

$$\mathbb{E}(\mathrm{X}_2) = \sum_{x_1=1}^{n} \mathbb{E}(\mathrm{X}_2|[\mathrm{X}_1 = x_1])\mathbb{P}([\mathrm{X}_1 = x_1]) = \sum_{x_1=1}^{n} \frac{x_1+1}{2} \times \frac{1}{n} = \frac{1}{2n}\left(\frac{n(n+1)}{2} + n\right) = \frac{n+3}{4}.$$

Dans le même cadre, calculons à nouveau l'espérance de $\mathrm{X}_1\mathrm{X}_2$.

$$\mathbb{E}(\mathrm{X}_1\mathrm{X}_2) = \sum_{x_1=1}^{n} \mathbb{E}(\mathrm{X}_1\mathrm{X}_2|[\mathrm{X}_1 = x_1])\mathbb{P}([\mathrm{X}_1 = x_1]).$$

Sachant $[\mathrm{X}_1 = x_1]$, la variable aléatoire $\mathrm{X}_1\mathrm{X}_2$ vaut $x_1\mathrm{X}_2$ et on peut appliquer la linéarité de l'intégrale.

$$\mathbb{E}(\mathrm{X}_1\mathrm{X}_2) = \sum_{x_1=1}^{n} \mathbb{E}(x_1\mathrm{X}_2|[\mathrm{X}_1 = x_1])\mathbb{P}([\mathrm{X}_1 = x_1]) = \sum_{x_1=1}^{n} x_1\mathbb{E}(\mathrm{X}_2|[\mathrm{X}_1 = x_1])\mathbb{P}([\mathrm{X}_1 = x_1]) = \sum_{x_1=1}^{n} x_1 \times \frac{x_1+1}{2} \times \frac{1}{n}.$$

La fin du calcul est alors identique à celle du calcul du paragraphe 3.3.

5.3 Démonstration de la formule de transfert

Démonstration de la formule de transfert. Considérons une variable aléatoire X, de valeurs $x_1, \dots, x_r$, et une fonction f définie sur l'ensemble $\mathrm{X}(\Omega)$. Appliquons maintenant la formule de l'espérance totale avec le système complet d'événements $\{[\mathrm{X} = x_1], \dots, [\mathrm{X} = x_r]\}$.

$$\mathbb{E}(f(\mathrm{X})) = \sum_{k=1}^{r} \mathbb{E}(f(\mathrm{X})|[\mathrm{X} = x_k])\mathbb{P}([\mathrm{X} = x_k]).$$

Remarquons maintenant que sachant $[\mathrm{X} = x_k]$, la variable aléatoire $f(\mathrm{X})$ est constante, égale à $f(x_k)$. Son espérance conditionnelle sachant $[\mathrm{X} = x_k]$ vaut donc $f(x_k)$. On obtient donc

$$\mathbb{E}(f(\mathrm{X})) = \sum_{k=1}^{r} f(x_k)\mathbb{P}([\mathrm{X} = x_k]),$$

ce qui est la formule de transfert dans le cas où il y a une seule variable aléatoire.

La formule du transfert pour les couples de variables aléatoires, vue au paragraphe 3.3, se démontre pareillement, en prenant pour système complet d'événements l'ensemble des événements de la forme $[\mathrm{X} = x_i] \cap [\mathrm{Y} = y_j]$.

On pourrait d'ailleurs imaginer une formule de transfert relative aux vecteurs aléatoires de taille quelconque.

6 Exercices

6.1 Probabilités sur un ensemble fini

Exercice 1. (*) On lance n fois une pièce.

Pour tout $k \in [\![1, n]\!]$, on considère les événements P_k = [On a fait `pile` au k-ième lancer.] et F_k = [On a fait `face` au k-ième lancer.]. Exprimer à l'aide des événements de la forme P_k et F_k les événements suivants :

A = [On a fait `pile` à tous les lancers.] ;
B = [On a fait `pile` à au moins un lancer.] ;
C = [On a fait `pile` à au moins un lancer, mais pas au cours des deux premiers.] ;
D = [On a fait `pile` pour la première fois à l'avant-dernier lancer.] ;
E = [On n'a jamais obtenu la succession `(pile, face)`.] ;
F = [On a fait `pile` exactement une fois.] ;
G = [On a fait `pile` exactement deux fois.].

Exercice 2. (*) On tire au hasard 8 cartes d'un jeu de 32 cartes classique. Déterminer la probabilité de chacun des événements suivants :

A = [On pioche les 4 rois.] ; B = [On pioche deux carrés.] ;
C = [On pioche au moins un carré.] ; D = [On pioche un seul carré.] ;
E = [On pioche tous les $\heartsuit$.] ; F = [On pioche quatre $\heartsuit$ et quatre $\diamondsuit$.].

Exercice 3. (*) Au cours d'un jeu de hasard, 20 boules sont piochées simultanément dans un sac contenant 70 boules numérotés de 1 à 70. Le participant doit cocher entre 5 et 10 cases sur le bulletin.

a. Quelle est la plus probable des éventualités suivantes : *avoir* 3 *numéros gagnants en cochant* 5 *cases* ou *avoir* 6 *numéros gagnants en cochant* 10 *cases* ?

b. Plus généralement, on pioche n boules parmi les N boules numérotées d'une urne et on coche a cases du bulletin de jeu. Pour tout $k \in [\![0, a]\!]$, calculer la probabilité d'avoir exactement k numéros gagnants.

Exercice 4. (*) Un militaire est affecté chaque année au hasard dans l'une des quatre villes $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}$. Quand il est dans une ville donnée une certaine année, la probabilité qu'il y soit encore l'année suivante vaut p, les trois autres destinations étant équiprobables.

On suppose qu'il est initialement affecté dans la ville A.

Pour tout entier n, on note A_n l'événement [Le militaire est affecté à la ville A pour l'année n.] et sa probabilité est notée a_n. On définit de même des événements $\mathrm{B}_n, \mathrm{C}_n, \mathrm{D}_n$ et leur probabilité b_n, c_n, d_n.

a. Que valent a_0, b_0, c_0, d_0 ?

b. Que vaut $a_n + b_n + c_n + d_n$?

c. Pour tout entier n, exprimer a_{n+1} en fonction de a_n, b_n, c_n, d_n, puis en fonction de a_n seulement.

d. En déduire une expression de a_n en fonction de n. Faire de même avec b_n, c_n, d_n.

e. Qu'observe-t-on quand n tend vers $+\infty$? Interpréter.

Exercice 5. (**) On range n boules distinctes dans n boîtes. On note π la probabilité qu'exactement une boîte soit vide.

Exprimer π_n puis en trouver un équivalent quand n tend vers $+\infty$.

6.2 Variables aléatoires

Exercice 6. (*) Soit X une variable aléatoire suivant la loi $\mathcal{B}(n,p)$. Calculer l'espérance de $\dfrac{1}{\mathrm{X}+1}$.

Exercice 7. (**) Pour tout $n \in \mathbb{N}^*$, on pose $\alpha_n = \sum\limits_{k=0}^{n-1} \left(\mathrm{e}^{k/n} - 1\right)$.

Pour tout $n \in \mathbb{N}^*$, on considère une variable aléatoire X_n dont la loi est donnée par

$$\mathrm{X}_n(\Omega) = \left\{ \frac{k}{n} \ ; \ k \in [\![0, n-1]\!] \right\}, \qquad \forall k \in [\![0, n-1]\!], \quad \mathbb{P}\left(\mathrm{X}_n = \frac{k}{n}\right) = \frac{\mathrm{e}^{k/n} - 1}{\alpha_n}.$$

a. Trouver un équivalent de α_n quand n tend vers $+\infty$.

b. On note F_n la fonction de répartition de X_n. Exprimer $\mathrm{F}_n(x)$ pour tout x réel.

c. Montrer que la suite de fonctions $(\mathrm{F}_n)_{n \geqslant 1}$ converge simplement sur $\mathbb{R}$ vers une fonction f continue, à exprimer explicitement.

d. Montrer que la convergence est uniforme.

6.3 Couples de variables aléatoires

Exercice 8. Pour tout cet exercice, on fixe un entier n supérieur ou égal à 3. Un sac contient $(n-2)$ boules vertes et deux boules jaunes. On extrait de ce sac les boules une par une **sans remise**. On note X la variable aléatoire égale au numéro du tirage où la première boule jaune est extraite et Y la variable aléatoire égale au numéro du tirage où la deuxième boule jaune est extraite.

Pour tout $i \in [\![1, n]\!]$, on considère l'événement J_i =[On pioche une boule jaune au tirage numéro i.]

a. Donner $\mathrm{X}(\Omega)$ et $\mathrm{Y}(\Omega)$, les ensembles des valeurs prises par X et Y, en justifiant rapidement les valeurs extrêmes.

b. Soit $k \in \mathrm{X}(\Omega)$. Exprimer l'événement $[\mathrm{X} = k]$ à l'aide des événements J_i.

c. Pour tout $k \in \mathrm{X}(\Omega)$, montrer l'égalité $\mathbb{P}(\mathrm{X} = k) = \dfrac{2(n-k)}{n(n-1)}$.

d. Montrer que l'espérance de X vaut $\dfrac{n+1}{3}$.

e. Calculer la variance de X.

f. Pour tout couple (k, ℓ) d'indices vérifiant $1 \leqslant k < \ell \leqslant n$, exprimer l'événement $[\mathrm{X} = k] \cap [\mathrm{Y} = \ell]$ à l'aide d'événements de la forme J_i, et en déduire la loi du couple (X, Y).

g. Pour tout $\ell \in [\![2, n]\!]$, montrer l'égalité

$$\mathbb{P}(\mathrm{Y} = \ell) = \frac{2(\ell - 1)}{n(n-1)}.$$

h. Calculer $\mathrm{Cov}(\mathrm{X}, \mathrm{Y})$.

Exercice 9. On lance deux dés équilibrés à n faces. Les résultats sont modélisés par deux variables aléatoires U_1 et U_2, supposées indépendantes.
On pose $\mathrm{X} = \min(\mathrm{U}_1, \mathrm{U}_2)$ et $\mathrm{Y} = \max(\mathrm{U}_1, \mathrm{U}_2)$.

a. Déterminer la loi et l'espérance de X.

b. Calculer $\mathrm{X} + \mathrm{Y}$ en fonction de U_1 et U_2. En déduire l'espérance de Y.

c. Calculer de même XY et en déduire la covariance de (X, Y).

6.4 Vecteurs aléatoires

Exercice 10. On fixe $n \in \mathbb{N}^*$ et on considère une urne qui contient $2n$ boules. Dans cette urne, il y a n boules numérotées de 1 à n et n boules non numérotées. On prélève au hasard et **simultanément** n boules dans l'urne et on considère la variable aléatoire S_n égale à la somme des numéros piochés (les boules non numérotées ne comptent pas).

Pour tout $i \in [\![1, n]\!]$, on note B_i la variable aléatoire égale à 1 si la boule portant le numéro i a été piochée et 0 sinon.

On prendra pour univers des possibles l'ensemble Ω des groupes de n boules piochées dans l'urne (en d'autres termes, $\Omega = \wp_n(\mathrm{U})$, où U est l'ensemble des boules de l'urne).

a. Montrer que B_i suit la loi $\mathcal{B}(1/2)$ pour tout $i \in [\![1, n]\!]$.

b. Exprimer S_n à l'aide des B_i et en déduire $\mathbb{E}(\mathrm{S}_n)$.

c. Pour tout couple (i, j) d'indices distincts de $[\![1, n]\!]$, déterminer la loi de $\mathrm{B}_i\mathrm{B}_j$.
En déduire l'égalité $\mathrm{Cov}(\mathrm{B}_i, \mathrm{B}_j) = \dfrac{-1}{4(2n-1)}$.

d. Démontrer l'égalité $\displaystyle\sum_{1 \leqslant i<j \leqslant n} ij = \frac{n(n^2-1)(3n+2)}{24}$.

e. Démontrer l'égalité $\mathbb{V}(\mathrm{S}_n) = \displaystyle\sum_{i=1}^{n} i^2 \mathbb{V}(\mathrm{B}_i) + 2 \sum_{1 \leqslant i<j \leqslant n} ij \,\mathrm{Cov}(\mathrm{B}_i, \mathrm{B}_j)$.

En déduire le calcul de $\mathbb{V}(\mathrm{S}_n)$.

Exercice 11. (**) Soit $k \in \mathbb{N}^*$. On considères des variables aléatoires $\mathrm{U}_1, \ldots, \mathrm{U}_k$ indépendantes, de loi uniforme sur l'ensemble $\{-1; 1\}$. On note

$$\mathrm{S}_k = \sum_{i=1}^{k} \mathrm{U}_i.$$

Pour tout $\lambda \in \mathbb{R}$, on pose $\varphi(\lambda) = \ln\left(\mathbb{E}(\mathrm{e}^{\lambda \mathrm{U}_1})\right)$.

a. Pour tout $\lambda \in \mathbb{R}$, prouver l'inégalité $\varphi(\lambda) \leqslant \lambda^2/2$.

b. Soit $t \in \mathbb{R}$. Pour tout $\lambda > 0$, prouver l'inégalité $\mathbb{P}(\mathrm{S}_k \geqslant t) \leqslant \exp(k\varphi(\lambda) - \lambda t)$.

c. En déduire l'inégalité de Hoeffding

$$\forall t > 0, \quad \mathbb{P}(\mathrm{S}_k \geqslant t) \leqslant \exp\left(-\frac{t^2}{2k}\right).$$
