CHAPITRE 1

Rappel de probabilités et de statistique

MTH3302 - Méthodes probabilistes et statistiques pour l'I.A.

Jonathan Jalbert – Automne 2021

Ce chapitre constitue une révision expéditive du cours MTH2302x - Probabilités et statistique pour ingénieurs. Les encadrés intitulés *exemples* correspondent à des exemple que nous compléterons en classe. Les encadrés intitulés *exercices* sont des exemples que nous effectuerons en classe que si le temps nous le permet. Dans le cas contraire, je vous suggère de tenter ces exercices par vous-même.

À la fin de ce chapitre, vous devriez être en mesure de :

— Calculer la probabilité d'événements pour des expériences aléatoires, ce qui inclut les probabilités conjointes, conditionnelles et marginales.

— Utiliser le théorème de Bayes pour inverser le conditionnement.

— Reconnaître les variables aléatoires classiques.

— Estimer les paramètres d'une distribution par la méthode du maximum de la vraisemblance.

1.1 Probabilités

Lorsqu'on veut analyser une **expérience aléatoire**, c'est-à-dire une expérience dont le résultat ne dépend que du hasard, la première étape consiste à dresser la liste de tous les résultats possibles. Les événements dont on espère calculer la probabilité correspondent à des sous-ensembles de l'ensemble des résultats possibles. La théorie des probabilités fait donc appel aux notions et aux résultats de la théorie des ensembles. Pour ce cours, je suppose que les étudiants sont familiers avec les notions élémentaires de la théorie des ensembles, telles les opérations d'union, d'intersection et de complémentations.

Considérons l'expérience aléatoire $\mathcal{E}$ où Ω représente l'ensemble des résultats possibles de cette expérience. L'ensemble Ω est parfois appelé l'**ensemble fondamental** de l'expérience $\mathcal{E}$. On souhaite associer à un événement $A \subset \Omega$ un nombre que l'on dénotera $\mathbb{P}(A)$ et que l'on appellera la **probabilité** de l'événement A. La définition suivante constitue le point de départ de la théorie mathématique des probabilités :

Définition (Axiomes de Kolmogorov)**.** *Une **probabilité** sur Ω, ou une mesure de probabilité sur Ω, est une fonction $\mathbb{P}$ qui satisfait les trois conditions suivantes :*

1. $0 \leq \mathbb{P}(A) \leq 1$ *pour* $A \subset \Omega$.
2. $\mathbb{P}(\emptyset) = 0$ *et* $\mathbb{P}(\Omega) = 1$.
3. *Si* $A_1, A_2, \ldots$ *sont des événements mutuellement exclusifs, alors*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Rappelons que des événements sont dits mutuellement exclusifs s'ils sont disjoints deux à deux, c'est-à-dire si pour tout choix de $i \neq j$, $A_i \cap A_j = \emptyset$.

À l'aide des axiomes de Kolmogorov et des propriétés des ensembles, on peut montrer les identités suivantes qui sont très utiles en pratique :

(i) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, où A^c dénote le complément de A.

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

Exercice 1

Comment interprétez-vous la notion de probabilité ? Quel sens physique lui donnez-vous ?

1.1.1 Le cas discret

Nous sommes dans le cas discret lorsque l'ensemble des résultats possibles Ω d'une expérience aléatoire $\mathcal{E}$ est ou bien un ensemble fini, ou bien un ensemble infini dénombrable. L'ensemble fondamental est alors composé d'une suite dénombrables de résultats possibles, disons $\Omega = \{\omega_1, \omega_2, \ldots\}$. Selon les axiomes de Kolmogorov, on a que

$$\mathbb{P}(A) = \sum_{\{i:\omega_i \in A\}} \mathbb{P}(\omega_i) \quad \text{pour tout } A \subset \Omega$$

Exemple 1

Jack, Fabrizio, Olaf et Sven se réunissent dans un pub de Southampton pour jouer au Poker. Pour décider qui sera le brasseur, ils tirent chacun à tour de rôle une pièce de monnaie et le premier qui obtient pile sera le brasseur. L'ordre du tirage est le suivant : Fabrizio, Jack, Olaf puis Sven. Quelles sont leurs probabilités respectives d'être le brasseur ?

Rép. : 8/15, 4/15, 2/15 et 1/15.

Remarque. *L'exemple précédent utilise le résultat classique concernant les séries géométriques :*

$$\sum_{j=0}^{\infty} r^j = \frac{1}{1-r} \quad \textit{pour tout} \quad -1 < r < 1.$$

1.1.2 Le cas où l'ensemble fondamental est fini et les résultats équiprobables

Si Ω est une ensemble fini et si les résultats possibles ont tous la même probabilité, alors pour tout événement A on a

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

Exemple 2

Durant leur partie de Poker à Southampton où deux billets pour un voyage à bord du Titanic sont en jeu, Jack reçoit d'emblée une main pleine, c'est-à-dire un triple accompagné d'un double. Quelle est la probabilité d'une telle main ?

Rép. : $\frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}}$.

1.1.3 Le cas non dénombrable

Dans le cas où l'ensemble fondamental d'une l'expérience aléatoire est non dénombrable, la probabilité associée à un événement particulier est nulle. Si on suppose pour le moment que l'ensemble des résultats possibles sont uniformes (analogue d'équiprobable dans le cas discret), la probabilité d'un événement A est donné par

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

Exemple 3

Jack et Fabrizio jouent aux fléchettes à bord du Titanic. Supposons que la cible est un disque de 461 mm diamètre et que les fléchettes atterrissent au hasard et de façon uniforme sur la cible. Quelle est la probabilité qu'une fléchette atteigne le *bullseye* de diamètre 13 mm ?

Rép. : $(13/461)^2$.

1.2 Probabilité conditionnelle

La notion de probabilité conditionnelle est l'une des plus importante en probabilité. Elle permet de mettre à jour l'information que l'on possède sur un événement sachant qu'un autre événement s'est produit.

Définition. *Soit l'expérience aléatoire $\mathcal{E}$ avec l'ensemble des résultats possibles Ω. Soit A et B des événements tels que $\mathbb{P}(B) > 0$. La* ***probabilité conditionnelle*** *de A sachant B, dénotée $\mathbb{P}(A \mid B)$, est définie par*

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

La probabilité conditionnelle demeure une mesure de probabilité, c'est-à-dire une quantité qui satisfait les axiomes de Kolmogorov. En particulier, si $\mathbb{P}(B) > 0$, alors

(i) $0 \leq \mathbb{P}(A \mid B) \leq 1$ pour tout événement A.

(ii) $\mathbb{P}(\emptyset \mid B) = 0$ et $\mathbb{P}(\Omega \mid B) = 1$.

(iii) Si $A_1, A_2, \ldots$ sont des événements mutuellement exclusifs, alors

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i \mid B).$$

(iv) $\mathbb{P}(A \mid B) = 1 - \mathbb{P}(A^c \mid B)$.

(v) $\mathbb{P}(C \cup D \mid B) = \mathbb{P}(C \mid B) + \mathbb{P}(D \mid B) - \mathbb{P}(C \cap D \mid B)$.

Exemple 4: (suite de l'exemple 1)

Sachant que Fabrizio n'obtient pas pile à son premier essai, quelle est la probabilité que Jack soit le brasseur ?

Rép. : 8/15.

La définition de probabilité conditionnelle s'applique toujours, même lorsque l'ensemble des résultats possibles n'est pas fini.

Exercice 2: (suite de l'exemple 3)

Sachant que la fléchette de Fabrizio est à l'intérieur de la demi-bulle de 32 mm de diamètre, quelle est la probabilité qu'elle soit dans le bullseye de 13 mm de diamètre ?

Rép. : $(13/32)^2$.

1.3 La règle de multiplication

La plupart du temps, les probabilités $\mathbb{P}(B)$ et $\mathbb{P}(A \mid B)$ sont données ou bien faciles à calculer. C'est plutôt la probabilité $\mathbb{P}(A \cap B)$ qui est inconnue et difficile à calculer. En manipulant la définition de la probabilité conditionnelle, on peut obtenir l'équation suivante pour $\mathbb{P}(A \cap B)$:

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

L'importance de cette formule est telle qu'elle porte le nom de **règle de multiplication**. Puisque $(A \cap B)$ et $(B \cap A)$ correspondent au même ensemble, on a également que

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

Exemple 5

Un panier contient cinq boules rouges et trois boules bleues. On tire deux boules au hasard et sans remise du panier. Quelle est la probabilité de tirer deux boules rouges ?

Rép. : 5/14.

La règle de multiplication se généralise pour plus de deux événements. Considérons d'abord le cas de trois événements A, B et C. En appliquant une première fois la règle de multiplication, on a que

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

En appliquant une deuxième fois la règle de multiplication au terme $\mathbb{P}(A \cap B)$, on trouve

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

Dans le cas général, on a que

$$\mathbb{P}(E_1 \cap \ldots \cap E_n) = \mathbb{P}(E_1)\,\mathbb{P}(E_2 \mid E_1)\,\mathbb{P}(E_3 \mid E_1 \cap E_2)\ \ldots\ \mathbb{P}(E_n \mid E_1 \cap \ldots E_{n-1}).$$

Exemple 6

Un panier contient cinq boules bleues, six boules blanches et sept boules rouges. On tire quatre boules au hasard et sans remise à partir du panier. Quelle est la probabilité d'obtenir quatre boules de la même couleur ?

Rép. : $\frac{\binom{5}{4}+\binom{6}{4}+\binom{7}{4}}{\binom{18}{4}}$.

1.4 La loi des probabilités totales

La loi des probabilités totales permet de calculer la probabilité d'un événement, disons A, en le scindant à l'aide d'une partition de Ω. Rappelons d'abord ce qu'est une partition.

Définition. *Si $E_1, E_2, \ldots, E_n$ sont des événements mutuellement exclusifs deux à deux,* i.e. *$E_i \cap E_j = \emptyset$ pour tout $i \neq j$ et s'ils sont exhaustifs,* i.e.

$$E_1 \cup E_2 \cup \ldots \cup E_n = \Omega,$$

alors on dit qu'ils forment une ***partition*** *de l'ensemble Ω.*

Théorème (Loi des probabilités totales)**.** *Si les ensembles $E_1, E_2, \ldots, E_n$ forment une partition de Ω, alors pour tout événement $A \subset \Omega$ on a que*

$$\mathbb{P}(A) = \sum_{j=1}^{n} \mathbb{P}(A \mid E_j)\ \mathbb{P}(E_j).$$

L'exemple suivant permet d'illustrer la loi des probabilités totales.

Exemple 7

Dans l'entrepôt d'une usine de fabrication de iPhone, 50% des iPhones sont fabriqués par la machine I, 30% par la machine II et 20 % par la machine III. La proportion de iPhones défectueux est de

- 3% lorsqu'ils sont fabriqués par la machine I ;
- 5% lorsqu'ils sont fabriqués par la machine II ;
- 8% lorsqu'ils sont fabriqués par la machine III.

Quelle est la proportion de iPhone défectueux dans l'entrepôt de cette usine ? Autrement dit, si on choisit un iPhone au hasard dans l'entrepôt, quelle est la probabilité qu'il soit défectueux ? La probabilité qu'il soit défectueux après trois ans est de 100% mais ça c'est un autre problème...

Rép. : $4,6\%$.

Exercice 3

On lance une pièce de monnaie jusqu'à ce qu'on obtienne pile. Puis on lance un dé un nombre de fois égal au nombre de lancers de la pièce de monnaie. Quelle est la probabilité d'obtenir au moins un six avec le dé ?

Rép. : 2/7.

1.5 Le théorème de Bayes

Le théorème de Bayes permet d'*interchanger* le conditionnement d'une probabilité conditionnelle. Par exemple, sachant que le iPhone sélectionné au hasard dans l'entrepôt est défectueux dans le cas de l'exemple 8, on pourra calculer la probabilité qu'il ait été fabriqué par la machine I.

Théorème (Théorème de Bayes)**.** *Soit $E_1, E_2, \ldots, E_n$ des événements formant une partition de Ω. Soit A un événement tel que $\mathbb{P}(A) > 0$. Alors pour tout $i \in \{1, \ldots, n\}$, on a que*

$$\begin{aligned}\mathbb{P}(E_i \mid A) &= \frac{\mathbb{P}(A \mid E_i)\ \mathbb{P}(E_i)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A \mid E_i)\ \mathbb{P}(E_i)}{\sum_{j=1}^{n} \mathbb{P}(A \mid E_j)\ \mathbb{P}(E_j)}.\end{aligned}$$

Exemple 8

(Suite de l'exemple 8) Si le iPhone sélectionné au hasard dans l'entrepôt est défectueux, quelle est la probabilité qu'il ait été fabriqué par la machine I ?

Rép. : 32.6%.

1.6 Indépendance

Parfois, le fait de savoir qu'un événement A s'est réalisé ne change en rien la probabilité que l'événement B se réalise. Par exemple, si on lance un dé deux fois et que l'on définit :

— A : obtenir un six au premier lancer ;

— B : obtenir un six au deuxième lancer.

Que A se réalise ou non, la probabilité de B demeure la même. On a donc que $\mathbb{P}(B \mid A) = \mathbb{P}(B)$. De la même manière, on a que $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. Les événements A et B sont donc *indépendants*. Voici la définition formelle de l'indépendance :

Définition. *Les événements A et B sont dits **indépendants** si on a*

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

Exercice 4

Montrer que si A et B sont indépendants au sens de la définition précédente, alors

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

et

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

L'indépendance pour trois événements se définie ainsi :

Définition (Indépendance de trois événements)**.** *Les événements A, B et C sont dits indépendants (ou mutuellement indépendant) si les quatre conditions suivantes sont satisfaites :*

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

Lorsque seulement les trois premières condition sont satisfaites, on dit que les événements sont indépendants deux à deux. Des événements peuvent être indépendants deux à deux sans toutefois être mutuellement indépendants, comme l'illustre l'exemple suivant.

Exercice 5

On lance une pièce de monnaie à deux reprises et on pose :

A **:** obtenir pile au premier lancer,

B **:** obtenir pile au deuxième lancer,

C **:** obtenir deux résultats identiques.

Les événements sont indépendants deux à deux mais ils ne sont pas mutuellement indépendants. Pouvez-vous le montrer ?

La définition de l'indépendance peut être généralisée à une collection d'événements.

Définition (Indépendance pour une collection d'événements)**.** *Soit $\mathcal{B} = \{B_j : j \in J\}$ une collection quelconque d'événements. Ces événements sont dits **mutuellement indépendants** si pour chaque entier positif n et pour chaque choix de $j_1, j_2, \ldots, j_n \in J$ avec $j_h \neq j_\ell$ pour tout $h \neq \ell$, on a*

$$\mathbb{P}\left(\bigcap_{k=1}^{n} B_{j_k}\right) = \prod_{k=1}^{n} \mathbb{P}(B_{j_k}).$$

À l'aide des opérations ensemblistes usuelles, on peut montrer le principe de la préservation de l'indépendance. Ici, nous nous contentons d'énoncer les résultats les plus utiles en pratique :

— Si A et B sont indépendants, alors A^c et B sont indépendants.
— Si A et B sont indépendants, alors A^c et B^c sont indépendants.
— Si A, B et C sont indépendants, alors $A \cap B$ et C sont indépendants.
— Si A, B, C et D sont indépendants, alors $A \cap B$ et $C \cup D$ sont indépendants.
— Si $B_1, B_2, \ldots, B_\ell$ sont indépendants, alors $B_1^c, B_2^c, \ldots, B_\ell^c$ sont indépendants.

Exemple 9

Soit les événements mutuellement indépendants A, B, C et D tels que

$$\mathbb{P}(A) = 0,40 \quad \mathbb{P}(B) = 0,50 \quad \mathbb{P}(C) = 0,40 \quad \mathbb{P}(D) = 0,30.$$

Quelle est la probabilité qu'au moins un de ces événements se réalise ?

Rép. : 0.874.

1.7 Variables aléatoires et distributions

Une variable aléatoire est une quantité dont la valeur numérique dépend du résultat d'une expérience aléatoire. De façon formelle, étant donné une expérience aléatoire $\mathcal{E}$ avec l'ensemble fondamental Ω et une mesure de probabilité $\mathbb{P}$, une variable aléatoire, disons Y est une fonction définie sur Ω et à valeurs dans $\mathbb{R}$. L'exemple suivant illustre le concept de variables aléatoires.

Exemple 10

On lance une pièce de monnaie jusqu'à ce qu'on obtienne pile pour la première fois. Soit la variable aléatoire Y, le nombre de lancer nécessaires pour obtenir le premier pile. Voici les résultats possibles de l'expérience aléatoire :

ω	$\mathbb{P}(\omega)$
(P)	$1/2$
(F, P)	$1/4$
(F, F, P)	$1/8$
$\vdots$	$\vdots$

La variable aléatoire Y peut donc prendre une valeur dans l'ensemble $\{1, 2, \ldots\}$. On a que

$$\mathbb{P}(Y = 1) = 1/2, \quad \mathbb{P}(Y = 2) = 1/4, \quad \mathbb{P}(Y = 3) = 1/8, \quad \ldots$$

En général, on a que $P(Y = y) = 2^{-y}$ pour $y \in \{1, 2, \ldots\}$.

Remarque (au sujet de la notation)**.** *En théorie des probabilités, on utilise habituellement les lettres majuscules près du début de l'alphabet* ($A, B, C, \ldots$) *pour dénoter des événements et les lettres majuscules près de la fin de l'alphabet* ($Z, Y, X, \ldots$) *pour dénoter des variables aléatoires. Les lettres minuscules sont réservée aux résultats particuliers de la variable aléatoire.*

Étant donné une variable aléatoire Y, on s'intéresse généralement à l'ensemble des probabilités $\mathbb{P}(Y \in B)$ où $B \subset \mathbb{R}$. En particulier, on s'intéresse à

$$\mathbb{P}(Y \in B) = \mathbb{P}\left\{Y^{-1}(B)\right\} = \mathbb{P}\left[\{\omega \in \Omega : Y(\omega) \in B\}\right].$$

Cette fonction s'appelle la distribution de probabilité de la variable aléatoire Y. On dit aussi la loi de probabilité de la variable aléatoire Y.

1.7.1 Variables aléatoires discrètes

Une variable aléatoire Y est dite discrète si l'ensemble de ses valeurs possible est un ensemble dénombrable. La distribution d'une variable aléatoire discrète prend la forme suivante :

$$\mathbb{P}(Y \in B) = \sum_{y \in B} p(y),$$

pour une certaine fonction $p(y)$. Cette fonction s'appelle la **fonction de masse** de la variable aléatoire Y. Bien sûr, on a que

— $p(y) \geq 0$ pour tout $y \in \mathbb{R}$;

— $\sum_{y \in \mathbb{R}} p(y) = 1$.

L'espérance, ou la moyenne, de la variable aléatoire discrète Y avec fonction de masse $p(y)$ est définie par l'équation suivante :

$$\mathbb{E}(Y) = \sum_{y \in \mathbb{R}} y\, p(y).$$

L'espérance correspond à la somme pondérée des valeurs possibles de Y. La pondération est donnée par la fonction de masse. L'espérance est une mesure de tendance centrale de la variable aléatoire. Si on répétait une infinité de fois l'expérience aléatoire que l'on effectuait la moyenne arithmétique de tous les résultats obtenus, on obtiendrait l'espérance. L'espérance d'une fonction d'une variable aléatoire $g(Y)$ est tout simplement la somme des valeurs transformées pondérées par leurs probabilités respectives :

$$\mathbb{E}\{g(Y)\} = \sum_{y \in \mathbb{R}} g(y)\, p(y).$$

Rappelons que la variance de la variable aléatoire Y peut s'écrire à l'aide de l'espérance :

$$\begin{aligned}\mathbb{V}ar(Y) &= \mathbb{E}\left[\{Y - \mathbb{E}(Y)\}^2\right] \\ &= \mathbb{E}(Y^2) - \{\mathbb{E}(Y)\}^2 .\end{aligned}$$

La première formule est utile pour interpréter la variance. Il s'agit de l'espérance du carré de la distance entre Y et sa moyenne. Il s'agit d'une mesure de dispersion de la variable aléatoire autour de la tendance centrale. La deuxième équation est utile pour calculer la variance.

Exemple 11

On lance deux dés bien balancés. Soit Y, la somme des deux dés. Quelle est la fonction de masse de la variable aléatoire Y ? Quelles sont l'espérance et la variance ?

Rép. : $p(y) = \frac{6-|7-y|}{36}$ pour $y \in \{2, 3, \ldots, 12\}$. $\mathbb{E}(Y) = 7$.

Introduisons ici quelques lois de probabilités discrètes classiques. Lorsque vous reconnaîtrez ces lois, vous pourrez directement utiliser leur fonction de masse, leur espérance et leur variance sans avoir à les recalculer.

La loi de Bernoulli

Considérons ici une expérience aléatoire qui ne peut prendre que deux résultats possibles, disons le résultat *succès* qui survient avec la probabilité p ou le résultat *échec* qui survient avec la probabilité $(1 - p)$. Notez que $0 \leq p \leq 1$. Une telle expérience aléatoire est appelée *épreuve de Bernoulli* en l'honneur du mathématicien Jacob Bernoulli (1654–1705). La loi de Bernoulli consiste à faire correspondre le succès à la valeur 1 et l'échec à 0. La fonction de masse s'écrit de la façon suivante :

$$p(y) = \begin{cases} (1-p) & \text{si } y = 0; \\ p & \text{si } y = 1; \\ 0 & \text{sinon.} \end{cases}$$

On note $Y \sim \mathcal{B}ernoulli(p)$ pour exprimer que la variable Y est distribuée selon la loi de Bernoulli de paramètres p. L'espérance et la variance d'une telle loi sont données par

$$\mathbb{E}(Y) = p \qquad \text{et} \qquad \mathbb{V}ar(Y) = p(1-p).$$

La loi binomiale

Considérons une succession de $n \in \mathbb{N}^*$ épreuves de Bernoulli avec probabilité de succès p effectuées de façon indépendante. La variable aléatoire

$$Y = \text{ le nombre de succès parmi les } n \text{ épreuves}$$

est une loi binomiale de paramètres (n, p). La fonction de masse est la suivante :

$$p(y) = \begin{cases} \binom{n}{y}\, p^y\, (1-p)^{(n-y)} & \text{si } y \in \{0, 1, \ldots, n\}, \\ 0 & \text{sinon}\,; \end{cases}$$

où

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}.$$

On note $Y \sim \mathcal{B}inomiale(n, p)$ pour exprimer que la variable Y est distribuée selon la loi binomiale de paramètres (n, p). L'espérance et la variance d'une telle loi sont données par

$$\mathbb{E}(Y) = np \qquad \text{et} \qquad \mathbb{V}ar(Y) = np(1-p).$$

La loi géométrique

On considère une série d'épreuves de Bernoulli, avec probabilité de succès p, indépendantes les unes des autres. Soit la variable aléatoire

$$Y = \text{le nombre d'épreuves nécessaires pour obtenir le premier succès.}$$

Cette variable aléatoire est appelée loi géométrique. Sa fonction de masse est la suivante :

$$p(y) = \begin{cases} (1-p)^{y-1}\, p & \text{si } x \in \{1, 2, 3, \ldots\}, \\ 0 & \text{sinon.} \end{cases}$$

On note $Y \sim \mathcal{G}eometrique(p)$ pour exprimer que la variable Y est distribuée selon la loi géométrique de paramètres p. L'espérance et la variance d'une telle loi sont données par

$$\mathbb{E}(Y) = \frac{1}{p} \qquad \text{et} \qquad \mathbb{V}ar(Y) = \frac{(1-p)}{p^2}.$$

La loi de Poisson

La loi de Poisson, de paramètre $\nu > 0$, nommée ainsi en l'honneur du mathématicien français Siméon Denis Poisson (1781–1840), est la loi avec la fonction de masse suivante :

$$p(y) = \begin{cases} \frac{e^{-\nu} \nu^y}{y!} & \text{si } y \in \{0, 1, 2, \ldots\}, \\ 0 & \text{sinon.} \end{cases}$$

On note $Y \sim \mathcal{P}oisson(\nu)$ pour exprimer que la variable Y est distribuée selon la loi de Poisson de paramètres ν. L'espérance et la variance d'une telle loi sont données par

$$\mathbb{E}(Y) = \nu \qquad \text{et} \qquad \mathbb{V}ar(Y) = \nu.$$

1.7.2 Variables aléatoire continues

Une variable aléatoire Y est dite absolument continue si l'ensemble de ses valeurs possibles est un ensemble non dénombrable. La distribution d'une variable continue est de la forme

$$\mathbb{P}(Y \in B) = \int_B f(y)\, dy$$

pour une certaine fonction $f(y)$. La fonction $f(y)$ s'appelle alors la **densité de probabilité** de Y. La densité possède les deux caractéristiques suivantes :

— $f(y) \geq 0$ pour tout $y \in \mathbb{R}$,

— $\displaystyle\int_{-\infty}^{\infty} f(y)\, dy = 1$.

Remarque. *Dans le cas continu, on a toujours que*

$$\mathbb{P}(a < Y < b) = \mathbb{P}(a < Y \leq b) = \mathbb{P}(a \leq Y < b) = \mathbb{P}(a \leq Y \leq b).$$

L'espérance, ou la moyenne, de la variable aléatoire continue Y avec la densité $f(y)$ est définie par l'équation suivante :

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y\, f(y)\, dy$$

Cette définition correspond à l'analogue continu de l'espérance du cas discret.

Introduisons ici quelques lois de probabilités continues classiques.

La loi uniforme continue

La loi uniforme continue sur l'intervalle (a, b) est la loi avec la densité suivante :

$$f(y) = \begin{cases} \frac{1}{b-a} & \text{si } a < y < b, \\ 0 & \text{sinon.} \end{cases}$$

On note $Y \sim \mathcal{U}niforme(a, b)$ pour exprimer que la variable Y est distribuée selon la loi de uniforme sur l'intervalle (a, b). L'espérance et la variance d'une telle loi sont données par

$$\mathbb{E}(Y) = \frac{a+b}{2} \qquad \text{et} \qquad \mathbb{V}ar(Y) = \frac{(b-a)^2}{12}.$$

La loi exponentielle

La loi exponentielle avec paramètre $\lambda > 0$ est la loi avec la densité suivante :

$$f(y) = \begin{cases} \lambda\, e^{-\lambda y} & \text{si } y > 0, \\ 0 & \text{sinon.} \end{cases}$$

On note $Y \sim \mathcal{E}xponentielle(\lambda)$ pour exprimer que la variable Y est distribuée selon la loi de Poisson de paramètres λ. L'espérance et la variance d'une telle loi sont données par

$$\mathbb{E}(Y) = \frac{1}{\lambda} \qquad \text{et} \qquad \mathbb{V}ar(Y) = \frac{1}{\lambda^2}.$$

La loi gamma

Fixons $\alpha > 0$ et $\lambda > 0$. La loi gamma de paramètre (α, λ) est la loi de densité :

$$f(y) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} & \text{si } y > 0, \\ 0 & \text{sinon}\,; \end{cases}$$

où $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u}\, du$ est la fonction gamma de Euler. On note $Y \sim \mathcal{G}amma(\alpha, \beta)$ pour exprimer que la variable Y est distribuée selon la loi gamma de paramètres (α, β). L'espérance et la variance d'une telle loi sont données par

$$\mathbb{E}(Y) = \frac{\alpha}{\lambda} \qquad \text{et} \qquad \mathbb{V}ar(Y) = \frac{\alpha}{\lambda^2}.$$

La loi gamma-inverse

Fixons $\alpha > 0$ et $\lambda > 0$. La loi gamma-inverse de paramètre (α, λ) est la loi de densité :

$$f(y) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp\left(-\frac{\lambda}{y}\right) & \text{si } y > 0, \\ 0 & \text{sinon}\,; \end{cases}$$

On note $Y \sim \mathcal{G}amma\mathcal{I}nverse(\alpha, \beta)$ pour exprimer que la variable Y est distribuée selon la loi gamma-inverse de paramètres (α, β). L'espérance et la variance d'une telle loi sont données par

$$\mathbb{E}(Y) = \frac{\lambda}{\alpha - 1} \text{ pour } \alpha > 1 \qquad \text{et} \qquad \mathbb{V}ar(Y) = \frac{\lambda^2}{(\alpha-1)^2(\alpha-2)} \text{ pour } \alpha > 2.$$

La loi bêta

La loi bêta sur l'intervalle $(0, 1)$ de paramètres $\alpha > 0$ et $\beta > 0$ est la loi de probabilité avec la densité suivante :

$$f(y) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1} & \text{si } 0 < y < 1, \\ 0 & \text{sinon.} \end{cases}$$

Lorsque $\alpha = \beta$, la loi bêta est symétrique. On note $Y \sim \mathcal{B}eta(\alpha, \beta)$ pour exprimer que la variable Y est distribuée selon la loi bêta de paramètres (α, β). L'espérance et la variance d'une telle loi sont données par

$$\mathbb{E}(Y) = \frac{\alpha}{\alpha + \beta} \qquad \text{et} \qquad \mathbb{V}ar(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta-1)}.$$

La loi normale

La loi normale définie sur $\mathbb{R}$ de paramètres $\mu \in \mathbb{R}$ et $\sigma^2 > 0$ est la loi de probabilité avec la densité suivante :

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}.$$

On note $Y \sim \mathcal{N}(\mu, \sigma^2)$ pour exprimer que la variable Y est distribuée selon la loi normale de paramètres (μ, σ^2). L'espérance et la variance de $Y \sim \mathcal{N}(\mu, \sigma^2)$ sont données par :

$$\mathbb{E}(Y) = \mu \qquad \text{et} \qquad \mathbb{V}ar(Y) = \sigma^2.$$

La loi de Student

La loi de Student (décentrée) définie sur $\mathbb{R}$ de paramètres $\mu \in \mathbb{R}$, $\sigma > 0$ et $\nu > 0$ est la loi de probabilité avec la densité suivante :

$$f(y) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})} \left\{1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right\}^{-\frac{(\nu+1)}{2}}.$$

On note $Y \sim t_\nu(\mu, \sigma)$ pour exprimer que la variable Y est distribuée selon la loi de Student de paramètres (μ, σ, ν). L'espérance et la variance de $Y \sim t_\nu(\mu, \sigma)$ sont données par :

$$E(Y) = \mu \text{ si } \nu > 1 \qquad \text{et} \qquad V(Y) = \frac{\nu}{\nu - 2}\sigma^2 \text{ si } \nu > 2.$$

1.8 Vecteur aléatoires

Un vecteur aléatoire de dimension n est un vecteur

$$\boldsymbol{Y} = (Y_1, \dots, Y_n)$$

dont chaque composante Y_j est une variable aléatoire. Dans cette section, la théorie sera présentée pour le cas $n = 2$. Le cas $n > 2$ se traite essentiellement de la même façon. Pour alléger la notation, nous allons dénoter le vecteur aléatoire par (X, Y) plutôt que par (Y_1, Y_2). Voici un exemple élémentaire d'un vecteur aléatoire.

Exemple 12

Une urne contient 7 boules bleues, 8 boules vertes et 10 boules rouges. On fait 5 tirage au hasard et sans remise. On considère les variables aléatoires suivantes :

— X : Le nombre de boules bleues parmi les boules tirées.

— Y : Le nombre de boules vertes parmi les boules tirées.

Plutôt que d'étudier séparément la distribution de X et la distribution de Y, nous étudierons plutôt la **distribution conjointe** de X et Y afin de comprendre et de décrire le lien entre X et Y. À l'instar du cas unidimensionnel, nous allons étudier séparément le cas discret et le cas continu.

1.8.1 Le cas discret

La distribution conjointe de X et Y est discrète si l'ensemble des valeurs possibles du point aléatoire (X, Y) est ou bien un ensemble fini, ou bien un ensemble infini dénombrable. Dans ce cas, la distribution conjointe de X et Y est décrite par la **fonction de masse conjointe** définie par l'équation suivante :

$$p(x, y) = \mathbb{P}\{(X, Y) = (x, y)\} = \mathbb{P}\{(X = x) \cap (Y = y)\}$$

La fonction de masse conjointe satisfait les propriétés suivantes :
- $p(x, y) \geq 0$ pour tout $(x, y) \in \mathbb{R}^2$,
- $\displaystyle\sum_{(x,y)\in\mathbb{R}^2} p(x, y) = 1$.

De plus, pour tout $B \subset \mathbb{R}^2$ on a que

$$\mathbb{P}\{(X, Y) \in B\} = \sum_{(x,y)\in B} p(x, y).$$

Lois marginales

Considérons le scénario suivant : on connaît la fonction de masse conjointe $p(x, y)$ des variables aléatoires (X, Y) et on s'intéresse uniquement à la variable aléatoire Y. On aimerait donc connaître la fonction de masse $p_Y(y)$ de la variable aléatoire Y. Cette fonction $p_Y(y)$ peut-être calculée à partir de la fonction de masse conjointe $p(x, y)$ de a façon suivante :

$$p_Y(y) = \mathbb{P}(Y = y) = \sum_{x\in\mathbb{R}} p(x, y).$$

De la même façon , on a que

$$p_X(x) = \mathbb{P}(X = x) = \sum_{y\in\mathbb{R}} p(x, y).$$

Dans le contexte multidimensionnel, la distribution de Y est appelée la **distribution marginale** de Y et la fonction de masse de Y et appelée la **fonction de masse marginale** de Y.

Lois conditionnelles

Considérons le scénario de l'exemple 15. On a fait les 5 tirages et on vous dit apprend qu'il y a exactement 2 boules bleues qui ont été tirées. Autrement dit $X = 2$. Étant donné cette information, quelle est la probabilité qu'on ait tirée 0 boule verte, 1 boule verte, 2 boules vertes, etc. On peut calculer ces probabilité à l'aide de la fonction de masse conjointe de (X, Y) et de la fonction de masse marginale de X de la façon suivante :

$$\mathbb{P}(Y = y \mid X = 2) = \frac{P\{(X = 2) \cap (Y = y)\}}{\mathbb{P}(X = 2)} = \frac{p(2, y)}{p_X(2)}.$$

Vue comme une fonction de y, la probabilité conditionnelle $\mathbb{P}(Y = y \mid X = 2)$ s'appelle la **fonction de masse conditionnelle** de Y sachant que $X = 2$ et elle est dénotée $p_{Y|X=2}(y)$. Plus généralement, si $X = x$, alors la fonction de masse conditionnelle de Y sachant que $X = x$ est dénotée $p_{Y|X=x}(y)$ et est définie par l'équation

$$p_{Y|X=x}(y) = \frac{p(x, y)}{p_X(x)}.$$

1.8.2 Le cas absolument continu

On dit que la distribution conjointe des variables X et Y est absolument continue s'il existe une fonction $f(x, y)$ telle que

$$\mathbb{P}\{(X, Y) \in B\} = \int\int_B f(x, y)\, dx\, dy$$

pour tout ensemble $B \subset \mathbb{R}^2$ pour lequel cette intégrale est bien définie. La probabilité $\mathbb{P}\{(X, Y) \in B\}$ correspond donc au volume sous le graphe de la fonction $f(x, y)$ au dessus de la région B. La fonction $f(x, y)$ s'appelle la densité conjointe de X et Y et elle satisfait les conditions suivantes :

— $f(x, y) \geq 0$ pour tout $(x, y) \in \mathbb{R}^2$,

— $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$.

Exemple 13

On tire deux nombres au hasard sur l'intervalle $(0, 1)$ et on considère les variables aléatoires suivantes :

— X : Le plus petit des nombres tirés.

— Y : Le plus grand des nombres tirés.

Lois marginales

Les **densités marginales** de X et Y correspondent aux analogues continus des fonctions de masse marginales. Elles sont définies par les équations suivantes :

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy,$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx,$$

Lois conditionnelles

La **densité conditionnelle** est l'analogue continu de la fonction de masse conditionnelle. La densité conditionnelle de Y sachant que $X = x$ est définie par l'équation suivante :

$$f_{(Y|X=x)}(y) = \frac{f(x,y)}{f_X(x)}.$$

1.9 Indépendance de variables aléatoires

Nous avons vu précédemment que deux événements sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$. Alors la définition suivante est tout à fait naturelle.

Définition. *Les variables aléatoires X et Y sont dites **indépendantes** si pour tout choix de $C \subset \mathbb{R}$ et de $D \subset \mathbb{R}$, les événements $(X \in C)$ et $(Y \in D)$ sont indépendants.*

Dans le cas discret, les variables X et Y sont indépendantes si et seulement si leur fonction de masse conjointe est égale au produite de leurs fonctions de masse marginales :

$$p(x,y) = p_X(x) \times p_Y(y).$$

Dans le cas continu, les variables X et Y sont indépendantes si et seulement si leur densité conjointe est égale au produite de leurs densités marginales :

$$f(x,y) = f_X(x) \times f_Y(y).$$

Le théorème suivant est très utile en pratique.

Théorème. *Si X et Y sont des variables aléatoires indépendantes et si $h(x)$ et $g(y)$ sont des fonctions réelles, alors les variables aléatoires $h(X)$ et $g(Y)$ sont indépendantes.*

1.10 Calcul de probabilités et d'espérances par conditionnement

1.10.1 Calcul de probabilités par conditionnement

Cette section n'est qu'une application directe de la loi des probabilités totales pour les variables aléatoires. Soit un événement A et la variable aléatoire Y. Si Y est discrète avec la fonction de masse $p_Y(y)$, alors

$$\begin{aligned}\mathbb{P}(A) &= \sum_{y\in\mathbb{R}} \mathbb{P}(A \mid Y = y) \times \mathbb{P}(Y = y), \\ &= \sum_{y\in\mathbb{R}} \mathbb{P}(A \mid Y = y) \times p_Y(y).\end{aligned}$$

Si Y est une variable aléatoire continue avec la densité $f(y)$, alors

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid Y = y) \times f_Y(y)\, dy.$$

1.10.2 Calcul de l'espérance par conditionnement

Supposons que nous sommes en présence de deux variables aléatoires X et Y et que l'on veuille calculer l'espérance de Y. Si on connaît la distribution marginale de Y, on peut procéder directement. On peut aussi procéder par conditionnement :

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} \mathbb{E}(Y \mid X = x) \times f_X(x)\, dx$$

si X est continue ou

$$\mathbb{E}(Y) = \sum_{x\in\mathbb{R}} \mathbb{E}(Y \mid X = x) \times p_X(x)\, dx$$

si X est discrète.

Remarque. *On utilise souvent la notation $\mathbb{E}(Y \mid X)$pour représenter la fonction $g(X)$, où $g(x)$ correspond à $\mathbb{E}(Y \mid X = x)$. Cette notation permet d'écrire l'espérance par conditionnement sous la forme compacte suivante :*

$$\mathbb{E}(Y) = \mathbb{E}\left\{\mathbb{E}(Y \mid X)\right\}.$$

1.10.3 Calcul de la variance par conditionnement

Puisque l'on peut écrire la variance de Y sous la forme suivante :

$$\mathbb{V}ar(Y) = \mathbb{E}(Y^2) - \left\{\mathbb{E}(Y)\right\}^2,$$

on peut montrer que la variance de Y s'exprime aussi sous la forme suivante :

$$\mathbb{V}ar(Y) = \mathbb{E}\left\{\mathbb{V}ar(Y \mid X)\right\} + \mathbb{V}ar\left\{\mathbb{E}(Y \mid X)\right\},$$

où $\mathbb{V}ar(Y \mid X)$ correspond à la fonction $h(x) = \mathbb{V}ar(Y \mid X = x)$.

1.11 Estimation des paramètres d'un modèle statistique

Soit une population d'intérêt où l'on s'intéresse à une caractéristique de celle-ci. Un recensement permettant de recueillir la valeur de cette caractéristique pour chacun des individus de la population est généralement fastidieux ou carrément impossible à réaliser. La procédure classique en statistique consiste à recueillir un échantillon de la population pour ensuite **inférer** les valeurs possibles de la caractéristique à la population.

Exemple 14

Supposons que l'on veuille estimer la proportion de fumeurs parmi les étudiants de Polytechnique. La population ici correspond aux étudiants de Polytechnique. La caractéristique étudiée est la proportion de fumeurs. Recenser tous les étudiants pourrait s'avérer fastidieux. La théorie statistique permet d'estimer la proportion de fumeurs dans la population à partir d'un échantillon de celle-ci.

La plupart du temps, on suppose que la caractéristique étudiée est distribuée selon une certaine loi de probabilité. Les paramètres de cette loi sont inconnus et doivent être estimés à l'aide d'un échantillon aléatoire.

Définition. *Soit une certaine loi de probabilité. Un ensemble de réalisations indépendantes de cette loi de probabilité constitue un* ***échantillon aléatoire****. On dénote l'échantillon aléatoire de taille n observé par $(y_1, y_2, \ldots, y_n)$ et les variables aléatoires correspondantes par $(Y_1, Y_2, \ldots, Y_n)$.*

Exemple 15: (suite)

Posons la variable aléatoire

$$Y_j = \begin{cases} 1 & \text{si le } j^e \text{ étudiant est fumeur,} \\ 0 & \text{si le } j^e \text{ étudiant est non-fumeur.} \end{cases}$$

Dans ce cas, on peut supposer que $Y \sim \mathcal{B}ernoulli(p)$ où p correspond à la proportion de fumeurs dans la population. Il suffit d'estimer p à l'aide d'un échantillon aléatoire.

1.11.1 Estimation par maximum de la vraisemblance

La méthode du maximum de la vraisemblance est une méthode très performante pour obtenir des estimations des paramètres inconnus. Comme son nom l'indique, cette méthode consiste à utiliser les valeurs qui maximisent la vraisemblance de l'échantillon observé comme estimations des paramètres. Soit le paramètre générique inconnu θ. La fonction de vraisemblance est définie de la façon suivante :

$$L(\theta) = \begin{cases} \prod_{i=1}^{n} p_{(Y_i|\theta)}(y_i) & \text{si le modèle statistique est discret,} \\ \prod_{i=1}^{n} f_{(Y_i|\theta)}(y_i) & \text{si le modèle statistique est continu.} \end{cases}$$

L'estimateur du maximum de la vraisemblance, dénoté ici par $\hat{\theta}$, correspond au point θ qui maximise $L(\theta)$.

La plupart du temps, il est plus facile de maximiser la log-vraisemblance dénotée $\mathcal{L}(\theta)$:

$$\mathcal{L}(\theta) = \ln\{L(\theta)\}.$$

La fonction logarithme étant une fonction monotone croissante sur les réels positifs, le point $\hat{\theta}$ qui maximise $L(\theta)$ est aussi celui qui maximise $\mathcal{L}(\theta)$.

Remarque. *Voici quelques propriétés élémentaires de la fonction logarithme :*

(i) $\ln(ab) = \ln(a) + \ln(b)$

(ii) $\ln(a/b) = \ln(a) - \ln(b)$

(iii) $\ln(a^b) = b\ln(a)$

(iv) $\ln(e) = 1$

(v) $\frac{d}{du}\ln h(u) = \frac{h'(u)}{h(u)}$

Exemple 16: (suite)

Si on recueille un échantillon de taille n d'étudiants de Polytechnique, la fonction de vraisemblance s'écrit ainsi :

$$L(\theta) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i}.$$

La log-vraisemblance prend la forme suivante :

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} y_i \ln(\theta) + \left(n - \sum_{i=1}^{n} y_i\right)\ln(1-\theta).$$

On dérive pour trouver le point critique :

$$\frac{d\mathcal{L}(\theta)}{d\theta} = \frac{\sum_{i=1}^{n} y_i}{\theta} - \frac{n - \sum_{i=1}^{n} y_i}{1-\theta}.$$

On trouve que l'estimateur du maximum de la vraisemblance est $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$, qui correspond à la moyenne de l'échantillon.

1.12 Exercices

1. On tire une carte à la fois d'un jeu de carte ordinaire de 52 cartes, comportant 4 *couleurs* et 13 *valeurs*.
   a) Quelle est la probabilité d'obtenir un as au premier tirage ?
   b) Quelle est la probabilité d'obtenir un as au 13^e tirage sachant qu'exactement 2 as ont été obtenus dans les 12 premiers tirages ?
   c) Quelle est la probabilité d'obtenir un as au 13^e tirage sachant que l'on a obtenu trois rois, un 10 et deux 7 parmi les 6 premiers tirages ?
   d) Quelle est la probabilité d'obtenir un as au 13^e tirage ?
   e) Quelle est la probabilité que le premier as survienne au 13^e tirage ?
   f) Quelle est la probabilité que le premier as survienne au 13^e tirage sachant que l'on a obtenu trois rois, un 10 et deux 7 parmi les 6 premiers tirages ?

2. On lance un dé à deux reprises et on considère les événements suivants :
   A **:** la somme des lancers est de 5,
   B **:** la somme des lancers est de 7,
   C **:** obtenir un quatre au deuxième lancer.
   Répondez aux questions suivantes :
   a) Les événements A et C sont-ils indépendants ?
   b) Les événements B et C sont-ils indépendants ?

3. Paul fait un examen à choix multiples, avec 4 choix de réponse pour chaque question. Pour chaque question, on suppose que la probabilité que Paul connaisse la réponse est de 2/3. S'il ne connaît pas la réponse, il en choisit une au hasard. Si Paul a obtenu la bonne réponse à la question 1, quelle est la probabilité qu'il connaissait vraiment la réponse ?

4. Un chapeau contient 3 cartes de tailles identiques. Les deux côtés de la carte A sont rouge et les deux côtés de la carte B sont noir. La carte C a un côté rouge et un côté noir. On tire une carte au hasard et on la dépose à plat sur la table. Le coté visible de la carte est rouge. Quelle est la probabilité qu'il s'agisse de la carte A ?

5. Lequel des scénarios suivants offre la meilleure probabilité de survie ?
   (i) Prendre 10 fois un avion dont la probabilité d'écrasement est de $1/10$ par vol.
   (ii) Prendre 100 fois un avion dont la probabilité d'écrasement est de $1/100$ par vol.

6. Un panier contient 4 boules rouge et une boule noire. On fait des tirages sans remise jusqu'à ce qu'on obtienne la boule noire. On considère la variable aléatoire Y : le nombre de tirages nécessaires pour obtenir la boule noire.
   a) Quelle est la fonction de masse de Y ?
   b) Quelle est l'espérance de Y ?

7. Refaites le numéro précédent avec cette fois-ci des tirages avec remise.

8. Un panier contient 10 boules rouges et 5 boules noires. On fait 4 tirages sans remise et on considère la variable aléatoire suivante Y : le nombre de boules noires parmi les 4 boules tirées. Quelle est la fonction de masse de Y ?

9. On suppose que Y est une variable aléatoire continue avec densité exponentielle de paramètre $\lambda = 3$.
   a) Calculer $\mathbb{P}\left(1/8 < Y \leq 1/2\right)$.
   b) Calculer $\mathbb{P}\left(Y > 1\right)$.

10. La densité d'une certaine variable aléatoire Y est donnée par

$$f(y) = \begin{cases} \frac{c}{y^5} & \text{si } y > 1, \\ 0 & \text{sinon.} \end{cases}$$

   a) Calculez la constante c.
   b) Calculez $\mathbb{P}(2 < Y < 3)$.
   c) Calculez l'espérance de Y.

11. On fait m tirages sans remise à partir d'un panier contenant n boules numérotées de 1 à n. Soit Y : le minimum des m valeurs obtenues. Calculer la fonction de masse de Y

12. On lance un dé à six faces plusieurs fois. Soit X : le nombre de lancers nécessaires pour obtenir un 6 et Y : le nombre de lancers nécessaires pour obtenir un 5.
    a) Calculez l'espérance de X.
    b) Calculez $E(X \mid Y = 1)$.
    c) Calculez $E(X \mid Y = 5)$.

13. Vous êtes invité à une fête. Supposons que les heures d'arrivée des invités sont indépendantes et distribuées selon la loi uniforme sur l'intervalle $(0, 1)$. Supposons qu'à part vous, le nombre d'invités qui se présenteront est distribué selon une loi de Poisson de moyenne 10.
    a) Quelle est l'espérance du nombre d'invités qui arriveront avant vous ?
    b) Quelle est la probabilité que vous êtes la n^e personne à arriver ?

14. On obtient un échantillon aléatoire de taille n, disons $Y_1, \ldots, Y_n$ de la densité $f(y)$. Obtenez l'estimateur de maximum de la vraisemblance du paramètre θ pour chacune des densités $f(y)$ suivantes :
    a)
    $$f(y) = \begin{cases} \frac{\theta^4 y^3}{6} e^{-\theta y} & \text{si } y \geq 0, \\ 0 & \text{sinon.} \end{cases}$$
    b)
    $$f(y) = \begin{cases} \theta y^{\theta - 1} & \text{si } 0 \leq y \leq 1, \\ 0 & \text{sinon.} \end{cases}$$
    c)
    $$f(y) = \begin{cases} \frac{3y^2}{\theta^3} e^{-\left(\frac{y}{\theta}\right)^3} & \text{si } y \geq 0, \\ 0 & \text{sinon.} \end{cases}$$
    d)
    $$f(y) = \begin{cases} \frac{y}{\theta^2} e^{-\frac{y^2}{2\theta^2}} & \text{si } y \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Bibliographie

[1] Sheldon M. Ross. *Introduction to Probability Models*. Academic Press, tenth edition, 2009.

[2] Sheldon M. Ross. *A First Course in Probability*. Pearson Education, ninth edition, 2014.