Guy Melançon

INF 162 – Probabilités pour l'informatique

Licence Informatique

20 octobre 2010

Département informatique
UFR Mathématiques Informatique
Université Bordeaux I
Année académique 2010 - 2011

Table des matières

Chapitre 1
Préambule

The capricious gods that were previously invoked to explain the lack of predictability in the world have been replaced by mathematical, statistical and computer-based models that allow us to understand and manipulate uncertain events.

D. Hand, H. Mannila, P. Smyth (2001). Principles of Data Mining, MIT Press.

Ce cours développe la théorie des probabilités et adopte un point de vue pragmatique : il s'agit de pouvoir utiliser des éléments de cette théorie à des fins informatiques. Les probabilités interviennent souvent dès lors que l'on veut analyser le comportement d'un algorithme ou comprendre comment se distribuent certains objets de l'informatique en fonction de paramètres de forme : les permutations, les arbres, les graphes, etc.

Un autre aspect présent dans ce cours est la simulation : les processus probabilistes prennent tout leur sens en informatique dès lors que l'on peut leur donner corp et les "faire vivre" en machine. La simulation apparaît comme une alternative, parfois obligée, à une analyse mathématique (analytique) en proposant plutôt une imitation d'un système. Elle met en oeuvre un modèle (agençant mathématique et informatique), est souvent conceptuellement plus simple et permet d'obtenir des informations quantitatives là où l'analyse mathématique échoue.

Les simulations sont aujourd'hui très populaires et sont pratiquées dans de nombreux domaines scientifiques, technologiques, industriels et économiques. Une simulation génère un volume important de données décrivant un système (à travers le modèle que l'on en a formulé). Contrairement à une étude analytique, les inférences liées à la simulation (les conclusions tirées des valeurs observées) sont de nature statistiques et peuvent être interprétées de façon équivoque. Les résultats obtenus sont intimement liés aux postulats de départ et une légère variation de ceux-ci peut modifier radicalement les résultats obtenus (selon la sensibilité du modèle). La mise au point de simulations requière la génération de valeurs possibles de certains paramètres dont on suppose souvent qu'ils suivent une loi de probabilité, ou de certains processus que l'on modélisent à l'aide de processus

stochastiques dont on maîtrise la portée dans le modèle. Le cours nous donnera l'occasion de mettre un pied sur ce terrain en soulignant la nécessité de bien maîtriser les aspects stochastiques de ces ingrédients d'une simulation.

Le cours est en grande partie basé sur les ouvrages de Richard Isaac [Isa05] et de Pierre Brémaud [Bré09] – la bibliothèque des sciences regorgent d'ouvrages qui présentant les notions de ce cours. Le logiciel opensource R [1], ou le site web SMEL [2] sont conseillés pour mettre en pratique les notions probabilistes (et statistiques, le cas échéant). Nous serons amenés à utiliser ces outils lorsqu'il s'agira de manipuler/simuler les lois de probabilités. L'utilisation de tableurs pourra même être envisagée [BMPS98] [BPSD07].

1. The R Project for Statistical Computing, `http://www.r-project.org/`.
2. "Statistiques Médicale en Ligne", `http://mistis.inrialpes.fr/software/SMEL/`.

Bibliographie

[BMPS98] Anne Brygoo, Michelle Morcrette, Odile Paliès, and Michèle Soria. *Initiation à la programmation par Word et Excel, Principes et macros.* Vuibert, 1998.

[BPSD07] Anne Brygoo, Maryse Pelletier, Michèle Soria, and Séverine Dubuisson. *Programmation et Algorithmique en VBA pour Excel.* Dunod, 2007.

[Bré09] Pierrre Brémaud. *Initiation aux probabilités et aux chaînes de Markov.* Springer, 2009.

[Isa05] Richard Isaac. *Une initiation aux probabilités.* Vuibert, 2005.

Chapitre 2
Rappels

2.1 Calcul, algèbre et combinatoire

Cette section rappelle quelques identités remarquables que l'on retrouve fréquemment en calcul de probabilité. Il est primordial d'être à l'aise avec ces manipulations algébriques élémentaires afin de conduire correctement le calcul qui suit du raisonnement probabiliste.

Exercice 2.1 Identités remarquables *Montrez (par récurrence) :*

$$\sum_{k=0}^{n} 2k+1 = (n+1)^2$$

$$\sum_{k=0}^{n} k = \frac{n(n+1)}{2}$$

$$\sum_{k=0}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

Série géométrique. Soit une valeur q avec $0 \leq q < 1$ alors la puissance q^n tend vers 0 lorsque $n \to \infty$. Mais plus encore, on peut faire la somme de toutes ces valeurs q^n et on obtient :

$$\lim_{n\to\infty} 1 + q + \cdots + q^n = \sum_{k\geq 0} q^n = \frac{1}{1-q}$$

En effet, désignons par $S_n = 1 + q + \cdots + q^n = \sum_{k=0}^{n} q^k$. On a alors $S_{n+1} = S_n + q^{n+1}$ et $S_{n+1} = 1 + qS_n$. Les deux membres droit de ces égalités coincident, d'où l'on tire $S_n = \frac{1-q^{n+1}}{1-q}$. En prenant la limite, et puisque $q^{n+1} \to 0$ lorsque $n \to \infty$, on obtient bien l'identité annoncée.

Exercice 2.2 *Calculez S_n et la série géométrique pour $q = 1/2$, $q = 1/3$, $q = 1/4$.*

Série harmonique. Montrez que la série :

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverge. Comparez-là à la série :

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \dots$$

qui ne converge pas (puisqu'en regroupant ses termes on trouve $1/2 + 1/2 + 1/2 + \ldots$).

En revanche, la suite $\sum_{n=1}^{p} \frac{1}{n} - \log(n)$ converge. Sa limite, notée γ est appelée la constante d'Euler.

Exercice 2.3 *Ecrivez un court programme qui évalue la constante d'Euler pour différentes valeurs de n (et en précision croissante).*

Coefficients binomiaux. Les coefficients binomiaux sont au centre de l'étude de la loi de probabilité du même nom. Ils permettent d'énumérer les sous-ensembles d'un ensemble donné (voir la prochaine section). On peut les définir de manière algébrique, comme suit. On définit le coefficient $\binom{n}{k}$ pour tout $n \geq 0$ et tout $0 \leq k \leq n$ en posant $\binom{n}{0} = \binom{n}{n} = 1$ et $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$. On peut écrire ces nombres dans un grand tableau triangulaire, chaque ligne débutant et se terminant par 1. La récurrence donne une façon facile de remplir les case de ce tableau : le nombre de la ligne n en position k s'obtient des nombres de la ligne précédente en position $k-1$ et k.

Exercice 2.4 Triangle de Pascal *Calculez les valeurs des 10 premières lignes du tableau de coefficient binomiaux.*

Exercice 2.5 Identité du binôme *L'identité du binôme généralise l'identité remarquable $(x+y)^2 = x^2 + 2xy + y^2$. Utilisez la récurrence définissant les coefficients binomiaux pour établir l'identité :*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Ecrivez l'identité particulière que l'on obtient lorsque $x = p$ et $y = 1 - p$.

2.2 Dénombrement

Le calcul des probabilités discrètes, dont il sera beaucoup question dans le cours, requière souvent que l'on puisse énumérer tous les objets d'un événement, sous-ensemble de l'espace de probabilités (l'ensemble

de toutes les observations possibles). Ces opérations de dénombrement s'appuient le plus souvent sur l'utilisation de coefficients qui calculent des configurations typiques :

- le *produit cartésien* de deux ensembles E, F noté $E \times F$: il est formé de la liste des paires ordonnées (e, f) avec $e \in E, f \in F$; son cardinal est $|E| \cdot |F|$ (où $|E|$ désigne le cardinal de l'ensemble E) ; plus généralement, le produit cartésien $E_1 \times \cdots \times E_k$ est formé des k-uplets $(e_1, \ldots, e_k)$ avec $e_i \in E_i$ et contient $|E_1| \cdots |E_k|$ éléments.
- le nombre de suites ordonnées de n éléments distincts (on parle aussi de *permutations* de ces éléments) est $n! = n(n-1)(n-2)\cdots 2 \cdot 1$. On convient que $0! = 1$;
- le nombre d'*arrangements* de k éléments parmi n : ce sont des suites ordonnées de k éléments distincts choisis parmi n (à la différence d'une permutation, on ne prend pas tous les n éléments) ; le nombre d'arrangements k éléments parmi n est $\frac{n!}{(n-k)!}$;
- le *coefficient binomial* $\binom{n}{k}$ calculant le nombre de *sous-ensembles* à k éléments parmi n.

Dans chacun des exercices suivants, rapportez-vous à des situations modélisées par un produit cartésien, une permutation, un arrangement ou un sous-ensemble [1].

Exercice 2.6 Morpion *On peut choisir de mettre ou non une croix dans chacune des cases d'un carré* 3×3 *(morpion). Combien y a-t-il de façons distinctes de procéder ?*

Exercice 2.7 Lancers de dé *On lance trois fois de suite un dé numéroté de 1 à 6 et on note les triplets ainsi obtenus. Combien y a-t-il de tels triplets ? (en tenant compte de l'ordre).*

Exercice 2.8 Football *Pour constituer une équipe de football, on a le choix entre 20 postulants. En supposant que chaque joueur est polyvalent (il peut jouer à tous les postes), combien peut-on constituer d'équipes*[2] *différentes ?*

Si parmi les 20 postulants, 17 sont joueurs de champ et 3 sont gardiens. Combien d'équipes distinctes peut-on alors constituer ?

Solution Tous les postes sont différents, on peut donc les ordonner et les énumérer 1, 2, …, 11. Il s'agit donc de choisir un joueur pour le premier poste – on a 20 candidats, puis un joueur pour le poste 2 – il reste 19 candidats, etc. Le nombre total d'équipe que l'on peut former est donc $20 \times 19 \times \cdots \times 10$.

Si on distingue le poste de gardien pour lequel on n'a que 3 candidats, on réduit ce nombre à $3\times$ (on choisit le gardien) $17 \times 16 \times \cdots \times 8$.

1. Les exercices proposés ici sont tirés du web (`http://pagesperso-orange.fr/gilles.costantini`)
2. Par équipe, il faut entendre "qui joue à quel poste", chaque poste étant distinct.

Exercice 2.9 Chiffres *Avec les nombres de 1 à 6, on veut constituer un nombre de 3 chiffres distincts. Combien de nombres distincts peut-on réaliser ?*

Sans répétitions, combien de nombres de 3 chiffres peut-on former à l'aide des six chiffres 2, 3 ,5, 6, 7, 9 ? Combien de ces nombres sont :

- *inférieurs à 500 ?*
- *impairs ?*
- *pairs ?*
- *multiples de 5 ?*

Solution On peut fabriquer 6^3 nombres de trois chiffres (on est autorisé à répéter le même chiffre), peu importe que l'on utilise les chiffres $\{1,\dots,6\}$ ou $\{2,3,5,6,7,9\}$. Si un chiffre ne peut être utilisé qu'une fois, alors on a plus que $6\times 5\times 4$ possibilités.

Si on doit fabriquer un nombre inférieur à 500, alors on ne peut plus utiliser que les chiffres $2,3$ en position des centaines, faisant tomber le nombre de possibilités à $2\times 5\times 4$.

Exercice 2.10 Boîtes et craies *On dispose de trois boîtes et de cinq craies de couleur bleue, rouge, jaune, verte et orange. 1) De combien de façons distinctes peut-on ranger les cinq craies dans les trois boîtes ? 2) Même question en laissant l'une des boîtes vides. 3) Même question si la bleue et la rouge sont rangées ensembles. 4) Même question si la bleue et la rouge sont rangées ensembles, mais seules.*

Exercice 2.11 Mots et alphabets *Combien de mots de 6 lettres peut-on écrire en utilisant 3 lettres D et 3 lettres H ?*

Combien de mots de 4 lettres peut-on former avec les 26 lettres de l'alphabet : a) En admettant les répétitions de lettres ? b) Sans lettres répétées ? Quelle est la probabilité qu'un mot de 4 lettres n'aie pas de lettres répétées ?

- *Combien de mots de 5 lettres peut-on faire avec les 26 lettres de l'alphabet ?*
- *Combien de ces mots ne comportent que des lettres distinctes ?*
- *Combien de ces mots comportent exactement 4 lettres distinctes (et donc une lettre répétée) ?*

Exercice 2.12 Triangles *Quel est le nombre de triangles que l'on peut former avec 10 points distincts (on supposera que trois points distincts ne se trouvent jamais sur une même droite, pour éviter les triangles dégénérés ...) ?*

Exercice 2.13 Chemins discrets *On considère les chemins sur le carré* 3×3 *vu comme un grille reliant les points (0, 0), (0, 1), (1, 0), ..., (2, 2). Combien y a t-il de chemins qui vont de (0,0) à (2,2) ? Plus généralement, combien de chemins vont de* $(0,0)$ *à* (m,n) *?*

Exercice 2.14 Trafic aérien *En hiver une compagnie aérienne dessert 6 villes. Quel est le nombre de lignes en service (elle propose tous les vols reliant deux de ces villes) ? En été, la compagnie a 45 lignes en service. Quel est le nombre de villes desservies ?*

Exercice 2.15 Loto *Au loto, quelle est le nombre de grilles qui ont k numéros gagnants ? (pour* $k = 3, 4, 5, 6$ *sur une grille de taille 49).*

Exercice 2.16 Jeu de cartes *Dans un jeu de 32 cartes, on choisit 5 cartes au hasard (ces 5 cartes s'appellent une "main").*

- *Quel est le nombre total de mains que l'on peut obtenir ?*
- *Combien de mains contiennent exactement 4 as ?*
- *Combien de mains contiennent exactement 3 as et 2 rois ?*
- *Combien de mains contiennent au moins 3 rois ?*
- *Combien de mains contiennent au moins un as ?*

Exercice 2.17 Boules et urnes *Une urne contient 49 boules numérotées de 1 à 49. On tire successivement 6 boules, sans remise. On appelle "tirage" cet ensemble de 6 numéros obtenus (sans tenir compte de l'ordre).*

- *Combien y a-t-il de tirages au total ?*
- *Combien y a-t-il de tirages qui contiennent 3 numéros pairs et 3 numéros impairs ?*
- *Combien y a-t-il de tirages qui contiennent au moins 5 numéros pairs ? (C'est-à-dire 5 numéros pairs ou 6 numéros pairs)*

Chapitre 3
Probabilités : propriétés élémentaires

Ce chapitre présente la théorie des probabilités, discrètes et continues. Il aborde d'abord les probabilités discrètes : celles qui décrivent des phénomènes dont les "possibilités de jeu" sont finies. Il est utile d'avoir une vision ensembliste des probabilités pour raisonner et calculer. Calculer une probabilités c'est d'abord déterminer l'espace que l'on observe, ou celui dans lequel se déroule une expérience. Il s'agit de décrire quelles sont toutes ces choses qui peuvent être observées (ou qui peuvent effectivement arriver) : tous les états possibles que peut prendre un système, ou toutes les configurations que l'on peut observer après que ce soit produit certains événements, etc. Pour raisonner, on se ramène souvent à des modèles très simples (états 0, 1 ; des configurations décrits sous forme de k-uplets de valeurs simples, etc.) qui seuls autorisent une solution numérique exacte ou approchée.

Ainsi, les exemples illustrant le calcul des probabilités s'appuie souvent sur les "jeux de hasard", les dés, les cartes, le tirage aléatoire de boules de couleurs dans des urnes, etc. Le travail consiste ensuite à savoir relever une situation concrète vers l'un de ces modèles simplistes. Par exemple, par souci de performance ou de robustesse des structures de données en informatique, on peut vouloir calculer la probabilité que deux clés de hachage (attribuée lors du stockage dans une table de hachage) entrent en collision. Du point de vue des probabilités, on peut de manière équivalente se poser la question : quelle est la probabilité que deux personnes dans une assemblée de N personnes aient la même date de naissance ? Du point de vue méthodologique, si l'on sait résoudre l'un des problèmes, on aura la clé pour résoudre l'autre.

De même, les exemples illustrant la théorie des processus stochastiques s'inspirent souvent des jeux de hasard. Du point de vue méthodologique, on peut penser à un hamster qui change de pièce dans un labyrinthe pour y manger des graines (un système changeant d'état), à des puces sautant d'un chien à un autre (une entreprise gagnant des parts de marché sur sa concurrente), ...

Développons maintenant la discussion autour d'un exemple. Imaginons que nous observions un système et qu'il s'agisse de comprendre comment il évolue. Un dispositif nous permet de savoir à tout moment s'il est actif ou passif (s'il émet ou non, une information, une onde, etc., par exemple). On peut donc imaginer effectuer une série de N observations du système. L'ensemble de ces observations, chacune étant codée par un 0 ou par un 1, nous donne une liste ordonnée de N valeurs. Les informaticiens parleront d'un vecteur de N bits ; les mathématiciens d'un élément du produit cartésien $\{0,1\}^N$. A priori, nous ne connaissons rien du système observé et faisons l'hypothèse que toutes les possibilités d'observations (tous les vecteurs de bits ; tous les N-uplets du produit cartésien) sont possibles. Observez au passage que l'expérience consistant à jouer N fois à pile ou face peut être décrite exactement de la même manière et que l'analyse que nous nous apprêtons à faire s'applique tout aussi bien aux observations du système qu'au jeu de hasard.

Imaginez que vous observiez le N-uplet uniquement formé de 0. En d'autres mots, à chaque fois que vous avez noté l'état du système il était inactif. Bien que ce N-uplets soient aussi probables que toutes les autres séquences d'observations, vous êtes autorisés à vous questionner. C'est bien le caractère aléatoire de votre expérience qu'il s'agit de comprendre : votre ignorance du comportement du système vous autorise à modéliser la situation en ces termes, vous avez des raisons de croire que vos observations doivent vous livrer une suite au caractère erratique plutôt qu'une séquence trop ordonnée, régulière ou périodique.

Ainsi, comprendre le système en termes probabilistes, c'est pouvoir répondre à des questions comme : quelle est la probabilité d'observer une séquence comptant k valeurs à 1 ? quelle est la probabilité d'observer une séquence comptant au moins k valeurs à 1 ? des séquences comptant au moins k valeurs à 1 consécutives ? etc.

Chacune de ces questions portent sur un *évènement* : c'est un sous-ensemble de l'ensemble de toutes les *épreuves* ou *observations* possibles. Introduisons quelques notations. On désignera par Ω l'espace de probabilité, c'est-à-dire l'ensemble de toutes les épreuves $\omega \in \Omega$. C'est, dans l'exemple que nous décrivions à l'instant, l'ensemble de tous les éléments du produit cartésien. Un *évènement* A est un sous-ensemble de Ω (les N-uplets comptant k valeurs à 1, ou les N-uplets comptant au moins k valeurs à 1, par exemple). La *probabilité* que A se réalise sera notée $P(A)$. La théorie des probabilités apporte un formalisme rigoureux à cette formulation des choses. Dans certains cas, il faut porter attention aux évènements que l'on peut considérer : puisque ce sont des ensembles contenus dans Ω, qu'ils peuvent être inclus les uns dans les autres, s'intersecter, etc. On pourrait aussi avoir à considérer le cas où Ω est infini, où le nombre d'évènements $A, B, C \ldots$ est lui aussi infini, etc. [1]

1. La théorie introduit la notion de tribus et de boréliens, puisqu'il faut s'assurer de pouvoir considérer l'union et l'intersection des évènements que l'on observe, de la cal-

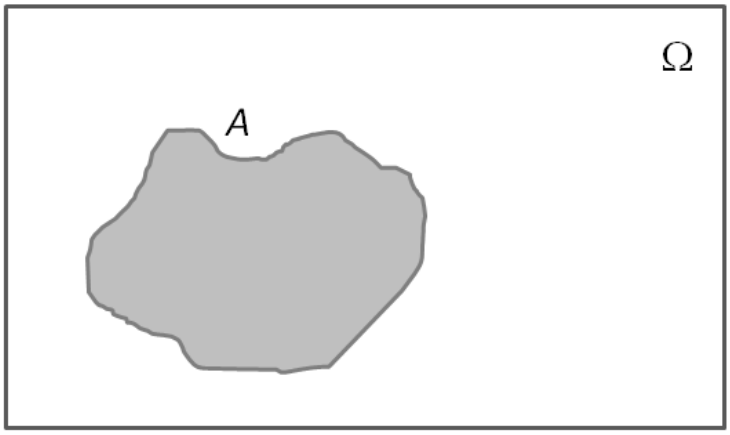


Fig. 3.1 Un évènement représenté classiquement à l'aide d'un diagramme de Venn. Intuitivement, sa probabilité est proportionnelle à son aire (relative au rectangle englobant représentant Ω.

Classiquement et en s'inspirant de la théorie des ensembles, on représente un évènement par un sous-ensemble A du référentiel Ω comme à la figure 3.4. Il faut lire cette illustration en pensant à la probabilité $P(A)$ comme étant égale à la part de l'aire de l'ensemble A dans le rectangle Ω. Cette façon de voir les choses peut être utile pour motiver ou justifier certains résultats ou définitions.

Concluons par un dernier exemple illustrant bien l'utilité des exercices parfois "ludiques" des cours de probabilités. *Q. Dans une assemblée de N personnes, quelle la probabilité que deux d'entre elles soient nées le même jour ?* A première vue, voilà un exercice de lycée sans lien évident avec les objectifs d'une licence en informatique. Mais pourtant, la question n'est étrangère à une question beaucoup plus informatique : étant donné une fonction de hachage $h : U \to [1, m]$ quelle est la probabilité d'une collision ? (deux éléments de U qui seraient associés à une même clé). On pense ici à une fonction définit sur un domaine assez grand – tous les mots possibles sur un alphabet, beaucoup plus grand que m. On exige que la fonction soit uniforme, c'est-à-dire que le nombre de clé associé à un entier $i \in [1, m]$ ne dépende pas de i et soit $|h^{-1}(i) = \frac{|U|}{m}|$.

Le coup des anniversaires dans une soirée est un cas particulier : la clé de hachage associée à une personne est la date du jour de sa naissance

culabilité des probabilité lorsque l'on fait des unions et/ou des intersections finies ou infinies. Ces calculs font aussi appel à la théorie de l'intégration (la théorie de la mesure), incontournable lorsqu'il s'agit d'utiliser les probabilités pour décrire le monde des particules en physique, par exemple.

(jour/mois), et on fait l'hypothèse que le nombre de personnes nés un jour donné est à peu près toujours le même peu importe le jour – ce qui semble raisonnable.

Exercice 3.1 Anniversaire et hachage *Résolvez les deux problèmes précédents. Précisez à chaque fois ce que doit être Ω et ce que sont les évènements à considérer.*

- *Dans une assemblée de N personnes, quelle la probabilité que deux d'entre elles soient nées le même jour ?*
- *Etant donné une fonction de hachage $h : U \to [1, m]$ quelle est la probabilité d'une collision ? (Deux éléments de U sont associés à une même clé.)*

3.1 Propriétés élémentaires

Plus formellement, un espace de probabilité est formé de trois ingrédients. Un ensemble Ω rassemblant les *épreuves* (les observations possibles), un ensemble d'*évènements* $\mathcal{U} = \{A, B, C, \ldots\}$, où $A, B, C, \ldots$ sont des sous-ensembles de Ω, et une *mesure de probabilité* P.

Les probabilités s'intéressent beaucoup à la théorie de la mesure (les conditions que doit satisfaire P, la façon de la calculer selon les propriétés de l'espace Ω et l'algèbre des évènements, entraînant une formalisation de l'algèbre des évènements $\mathcal{U}$). L'ensemble $\mathcal{U}$ est souvent infini, et il faut prendre des précautions lorsque l'on effectue une union, ou une intersection infinie d'éléments de $\mathcal{U}$. L'ensemble $\mathcal{U}$ est souvent appelé une *tribu*. Nous n'aborderons toutefois pas (directement) ces subtilités mathématiques.

Toujours en raisonnant de manière ensembliste (et à partir de la figure 3.4), si A et B sont deux évènements disjoints ($A \cap B = \emptyset$) alors on a $P(A \cup B) = P(A) + P(B)$, puisque l'aire totale de l'ensemble $A \cup B$ est la somme des aires de A et de B.

L'ensemble $\Omega \setminus A$ est le complément de A, qu'on note A^c. On a donc $P(A^c) = 1 - P(A)$.

Exercice 3.2 Décrire un espace de probabilités

Dans une fabrique de processeurs, on prélève toutes les heures les trois derniers processeurs produits. Ceux-ci sont classés dans deux catégories : fonctionnel, codé 1 et défectueux, codé 0.

1. *Décrivez l'espace associé à cette expérience aléatoire (par quoi peut-on représenter l'expérience (les états des trois processeurs p_1, p_2, p_3 sélectionnés)*
   , ?)
2. *Décrivez (en termes ensemblistes) les évènements suivants :*

- $A =$ *"le premier processeur est défectueux"*
- $B =$ *"le dernier est fonctionnel"*
- $C =$ *"deux processeurs sont défectueux"*
- $D =$ *"au moins deux processeurs sont fonctionnels".*

Exercice 3.3 Le problème du chevalier de Méré *Lequel de ces deux événements est le plus probable : "Obtenir au moins une fois un six en quatre lancers de dés", ou "Obtenir au moins un double six en vingt quatre lancers de deux dés".*

Solution On note par $\Omega = \{1, 2, \ldots, 6\}$ l'ensemble des valeurs possibles prises par un dé (lors d'un lancer unique), pour lequel on a $P(\omega) = 1/6$ ($\omega \in \Omega$).

L'espace de probabilité qui décrit l'expérience "4 lancers de dés" est Ω^4. L'évènement $A =$ "avoir au moins un six" est complémentaire de l'évènement $A^c =$ "ne jamais avoir de six". Or, il y a 5^4 séquences $(\omega_1, \omega_2, \omega_3, \omega_4)$ dont aucune composante ne comporte un 6. On a donc $P(A) = 1 - P(A^c) = 1 - (5/6)^4$.

L'espace de probabilité qui décrit l'expérience "24 lancers de 2 dés" est $(\Omega \times \Omega)^{24}$. L'évènement $B =$ "avoir au moins un double six" est complémentaire de l'évènement $B^c =$ "ne jamais avoir de double six". Or, il y a 35 paires de dés qui ne correspondent pas à un double 6, et donc 35^{24} séquences $(\omega_{1,1}, \omega_{2,1}), (\omega_{2,1}, \omega_{2,2}), \ldots, (\omega_{24,1}, \omega_{24,2})$ dont aucune composante n'est un double six. On a donc $P(B) = 1 - P(B^c) = 1 - (35/36)^{24}$.

Reste à comparer ces deux valeurs, et on trouve $P(A) > P(B)$. Question subsidiaire : combien de lancers additionnels de double dés faut-il pour que les probabilités s'inversent ?

Exercice 3.4 Le bon choix *Vous jouez à un jeu télévisé pour gagner un PC de dernière génération. Le PC se trouve derrière une porte parmi trois A, B ou C ; les deux autres portes cachent des babioles. Après avoir fait votre choix, le présentateur télé ouvre une porte derrière laquelle se trouve un première babiole.*

Il vous propose soit de rester sur votre premier choix, soit d'échanger et d'opter plutôt pour la porte qu'il n'a pas ouverte ... Avez-vous avantage à rester sur votre premier choix ou à changer ?

Etudiez d'abord vos chances de gagner en faisant l'hypothèse que vous changez de porte quoiqu'il arrive. Puis étudiez le cas où vous ne changez pas de porte.

Exercice 3.5 Urnes et boules

Une urne contient 8 boules blanches et 6 boules noires, chaque boule ayant la même probabilité d'être tirée.

1. *On tire simultanément 5 boules. Quelle est la probabilité d'obtenir :*
   - *3 blanches et 2 noires ?*
   - *des boules de couleurs différentes ?*

2. *On tire successivement 5 boules avec remise de chaque boule tirée. Quelle est la probabilité d'avoir :*
   - *3 blanches puis 2 noires ?*
   - *3 blanches et 2 noires dans un ordre quelconque ?*

Exercice 3.6 Jeux de données

On considère des jeux de données permettant de tester la robustesse de l'implémentation d'un algorithme, dont 8 sont jugés d'une complexité "élevée" et 6 sont jugés de complexité "moyenne". On choisit au hasard des jeux de données qu'on soumet à l'algorithme, chaque jeu de données ayant la même probabilité d'être choisi.

1. *On tire simultanément 5 jeux de données. Quelle est la probabilité d'obtenir :*
   - *3 jeux de complexité élevée et 2 de complexité moyenne ?*
   - *des jeux de données de complexités différentes (pas tous de complexité élevée ou tous de complexité moyenne) ?*
2. *On tire successivement 5 jeu de données, en s'autorisant à choisir le même jeu de données plus d'une fois. Quelle est la probabilité d'avoir :*
   - *3 jeux de complexité élevée puis 2 de complexité moyenne (dans cet ordre) ?*
   - *3 jeux de complexité élevée puis 2 de complexité moyenne (peu importe l'ordre dans lequel ils sont apparus) ?*

Exercice 3.7 Tirage aléatoire en machine *Explorez la (les) librairie(s) mises à disposition dans divers langages de programmation (C, Java, python, etc.).*

A minima, le langage vous permet de tirer au hasard (et uniformément, c'est-à-dire que le tirage est équiprobable) un nombre réel de l'intervalle [0, 1).

Utilisez cette méthode de la librairie pour écrire un court programme qui permet de simuler un évènement de Bernouilli avec probabilité p (le lancer d'un pièce au jeu de pile ou face, où pile est obtenu avec probabilité p).

En répétant ce tirage aléatoire un grand nombre de fois, et en cumulant les fréquences des résultats, vous prendrez soin de vérifier empiriquement que votre tirage aléatoire suit bien la probabilité prescrite p.

Solution On peut simuler le tirage de Bernouilli de paramètre $0 < p < 1$ si on dispose d'un générateur de nombre réel : une fonction retournant un nombre réel de l'intervalle $[0, 1)$. C'est ce que fait la fonction python `random` (de la librairie `random`). Après avoir défini la fonction Bernouilli, on l'appelle m fois (avec m de plus en plus grand, pour constater empiriquement que la fréquence de succès s'approche de plus en plus de la probabilité théorique p).

Exercice 3.8 Rademacher *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi de* Rademacher $R(p)$*, c'est-à-dire $P(X_n = 1) = p$ et $P(X_n = -1) = 1 - p$ avec $0 < p < 1$. Ecrivez un court programme*

```
from random import *

def Bernouilli(p):
    if random() < p:
        return 1
    else:
        return 0

def freqEmpirique(p, m):
    s = 0
    for i in range(m):
        s += Bernouilli(p)
    return s/m
```

Fig. 3.2 Code python réalisant un tirage de Bernouilli de paramètre p.

qui permet de simuler la loi de Rademacher (que l'on peut comparer à un lancer à pile ou face mais à valeur 1 ou -1 avec probabilité p).

Exercice 3.9 Tirage aléatoire sur l'intervalle $[a,b)$ *Utilisez le tirage aléatoire d'un réel de l'intervalle* $[0,1)$ *pour construire une méthode pour tirer aléatoirement et uniformément un réel de l'intervalle* $[a,b)$ *où* $a < b$ *sont deux nombres réels ou entiers quelconques.*

Exercice 3.10 Histogramme *Un histogramme permet de rendre compte de la distribution de valeurs dans un échantillon numérique. Plus précisément, supposons que nous ayons une suite de nombres* $\mathcal{S} = x_1, x_2, \ldots$ *entiers ou réels. On peut alors fabriquer des couples* (x, n_x) *où* $x \in E$ *et* n_x *indique le nombre de valeurs* x_n *qui coincide avec* x*. Le nombre* n_x *est appelé la* fréquence *de* x *dans la suite* $\mathcal{S}$*.*

Pour visualiser les fréquences on peut placer les couples (x, n_x) *dans le plan. Cette visualisation peut être toutefois irrégulière et difficile à interpréter. On peut alors regrouper les valeurs* x_n *par intervalles disjoints et cumuler les fréquences des éléments appartenant à un même intervalle.*

Reprenez le tirage aléatoire de nombre réels de l'intervalle $[a,b)$ *et construisez un histogramme rendant compte des valeurs obtenues au cours d'un grand nombre de tirages. Vous trouverez un utilitaire permettant de visualiser l'histogramme*[2]*.*

Exercice 3.11 Simuler un lancer de dé *Selon les possibilités offertes par la (les) librairie(s) du langage utilisé, vous pouvez tirer au hasard un nombre réel de l'intervalle* $[0,1)$ *ou un entier naturel (borné par la taille des mots mémoire et le format de stockage interne).*

2. On peut penser à `gnuplot`, par exemple.

A partir de l'un (ou des deux) de ces moyens de base, fabriquer une méthode permettant de simuler le lancer d'un dé équiprobable à six faces. Vous prendrez soin de vérifier empiriquement que votre tirage aléatoire est bien équiprobable.

Simulez l'expérience du chevalier de Méré et constater le résultat théorique de l'exercice 3.3.

Solution

```
# reproduit l'expérience de Méré
# 'simple' ou 'double'
def mere(type):
    if type == 'simple':
        for i in range(4):
            if lancerDe() == 6:
                return 1
        return 0
    else:
        for i in range(24):
            l1 = lancerDe()
            l2 = lancerDe()
            if l1 == 6 and l2 == 6:
                return 1
        return 0
```

Fig. 3.3 Code python simulant l'expérience de Méré. On peut, comme pour le tirage de Bernouilli (Fig. 3.2), calculer empiriquement la fréquence de succès de l'expérience.

Exercice 3.12 Générer des entiers *En machine, les entiers sont décrits par des vecteurs de bits de taille N (= 32 ou 64). On peut donc penser générer aléatoirement des entiers en jouant à pile ou face N fois avec $p = 1/2$ et en prenant le résultat du jeu comme résultat. Une suite générée ainsi forme t-elle une suite aléatoire d'entiers ?*

Ecrivez un court programme qui tire au hasard un nombre entiers de l'ensemble $\{0, 2^N - 1\}$ de cette façon ($N = 32$ ou 64 selon l'architecture de la machine utilisée).

Solution On suppose disposer d'une fonction `Bernouilli(p)` qui permet de simuler un évènement de Bernouilli de paramètre p. Il suffit donc de répéter N tirages de Bernouilli avec $p = 1/2$, et de faire la conversion de la base 2 à la base 10.

Si, par exemple, on utilise cette procédure pour générer un milliers d'entiers ($N = 10$), on doit pouvoir observer le caractère uniforme de la génération. On dresse un histogramme des nombres générés et on le trace :

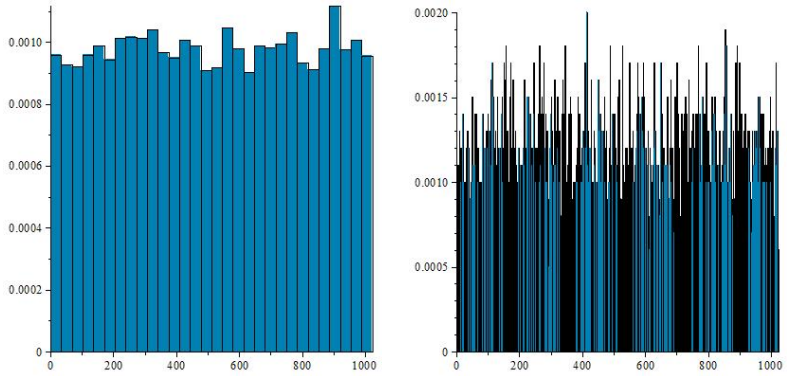


Fig. 3.4 Ces deux histogrammes illustrent la répartition des entiers générés par la procédure. Le caractère uniforme de la génération aléatoire est révélé par l'allure "plate" de l'histogramme. Remarquez que l'on perd ce caractère plat et permet de voir l'aspect chaotique du tirage aléatoire dès lors que l'histogramme contient trop de classes.

```
# retourne un entier choisi uniformément aléatoirement
# en fabriquant un vecteur de bits de taille N
# (donné en paramètre)
def randomVectInt(N):
    t = [0] * N
    for i in range(N):
        if random.random() < 0.5:
            t[i] = 0
        else:
            t[i] = 1
    s = 0
    b = 1
    for i in range(N):
        s += t[i] * b
        b *= 2
    return s
```

Fig. 3.5 Code python générant des entiers aléatoires en base 2, en simulant une loi binomiale de paramètre $p = 1/2$ (puis en retournant l'entier converti en base 10).

Exercice 3.13 Probabilités et modèles physiques *Une corde dans un cercle est un segment de droite le coupant en deux points de sa circonférence. On choisit "au hasard" une corde d'un cercle de rayon unitaire. Quelle est la probabilité qu'elle soit moins longue que le côté d'un triangle équilatéral inscrit dans le cercle*

Nous allons voir que la réponse que l'on peut donner à cette question dépend de ce que l'on entend par "choisit au hasard".

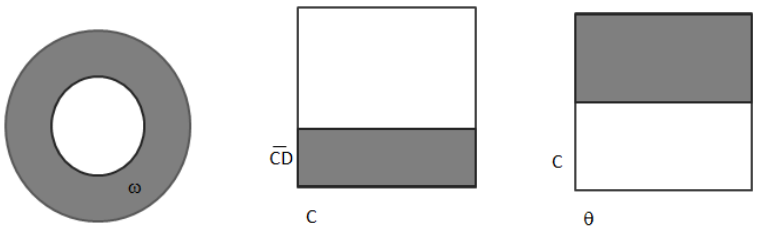


Fig. 3.6 On choisit une corde aléatoirement en choisissant au hasard un point du cercle. Les trois figures illustrent les trois espaces de probabilités considérés pour donner une solution. Chacune décrit ce que l'on entend par "choisir une corde au hasard" (un point ω dans le cercle ; un couple de points $(C,D) \in [0,2\pi] \times [0,2\pi]$ sur la circonférence du cercle ; un rayon déterminé par un angle $\theta \in [0,2\pi]$ et un point $C \in [0,1]$ sur ce rayon). La zone grisée correspond à chaque fois à l'évènement A.

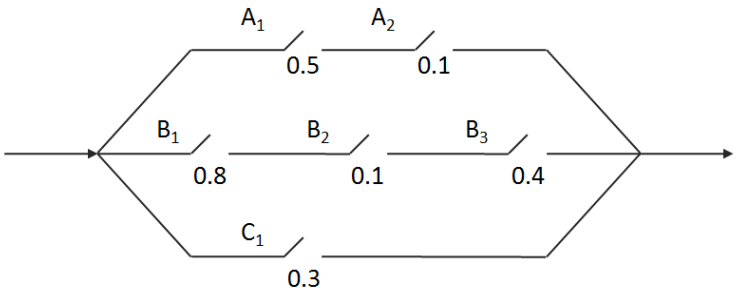
Solution 1. *Pour toute corde, il existe un unique rayon qui la croise à angle droit. Ainsi, à tout point ω d'un cercle $\mathcal{C}$ centré en O, on peut faire correspondre une unique corde perpendiculaire au rayon passant par ω. On peut donc interpréter "choisir une corde au hasard" comme étant "choisir un point ω du cercle et tracer l'unique corde passant par le rayon $O\omega$. Cette corde est de longueur supérieure au côté du triangle à condition que ce point se situe hors du cercle $\mathcal{C}'$ centré en O et de rayon $1/2$. Calculez la probabilité cherchée selon ce modèle.*

Solution 2. *Toute corde est déterminée par deux points C, D de la circonférence du cercle $\mathcal{C}$. Observez que le côté du triangle est lui-même une corde de longueur exactement $\sqrt{3}$ qui correspond à une portion de la circonférence de longueur $2\pi/3$. Etant donné un point C, les points D qui sont à une distance inférieure à $2\pi/3$ de C sur la circonférence permettent de construire une corde satisfaisant la condition. Choisir une corde au hasard correspond ici au choix d'un couple ordonné de points C, D sur la circonférence. Calculez la probabilité cherchée selon ce modèle.*

Solution 3. *Toute corde est déterminée par un point C se situant sur un rayon du cercle, lui-même déterminée par un angle $\theta \in [0,2\pi)$. La corde correspond donc à un couple (θ, C) avec $C \in [0,1)$, qui satisfait la condition dès lors que $C > 1/2$. Choisir une corde au hasard correspond ici au choix d'un couple ordonné de points C, θ. Calculez la probabilité cherchée selon ce modèle.*

Exercice 3.14 *On considère l'ensemble de relais de la figure 3.1. Les nombres indiquent les probabilités que le relais soient "ouverts" (ça ne circule pas). Les relais sont indépendants (formalisez cette notion). Calculez la probabilité pour qu'au moins une branche soit "fermée" (et que ça circule).*

Exercice 3.15 *Simulez par programmation le tirage aléatoire d'une corde selon les trois modèles de l'exercice précédent.*



Exercice 3.16 Guerre de processus *Imaginons deux threads s'exécutant simultanément, chacun réclamant au système une portion mémoire à différents moments. On suppose que les threads t_a et t_b demande à chaque fois une portion mémoire de taille fixe et qu'à la fin de leur exécution t_a aura demandé m portions mémoires, alors que t_b en aura demandé n, avec $m \geq n$. On s'intéresse à la probabilité que le thread t_a ait constamment plus de mémoire à sa disposition que le threads t_b.*

Nous allons modéliser ce problème de la façon suivante : les demandes successives des threads sont décrites par une séquence de a et de b (un mot). Une séquence qui maintient l'avantage à t_a (plus de mémoire accordée à t_a à tout moment) est un mot dont les préfixes comptent toujours plus de a que de b.

Soient $n, m \geq 0$, avec $m \geq n$. On note $C(m,n)$ l'ensemble des mots u de $\{a,b\}^*$ tels que $|u|_a = m$ et $|u|_b = n$, i.e. u est constitué de m lettres a et de n lettres b.

Soit $p : \{a,b\}^* \to Z$ l'application définie par $p(u) = |u|_a - |u|_b$. On propose de calculer la probabilité de l'évènement

$$D = \{u \in C(m,n) | \forall v \neq \epsilon \text{ prefixe de } u, p(v) > 0\}.$$

- *Soit $C_b(m,n)$ l'ensemble des mots de $C(m,n)$ commençant par b. Décrire l'évènement $E = C(m,n) \backslash D$ et montrer que $P(E) = 2P(C_b(m,n))$.*
- *En déduire la probabilité que le threads t_a dispose toujours de plus de mémoire que le threads t_b.*

Exercice 3.17 Quidam et casino *(Cet exercice est emprunté à et adapté de [Bré09, Chap.,3.1].) Au casino, un croupier choisit deux nombres $m, n \in [0, N]$, avec $m < n$ disons, qu'ils ne vous dévoilent pas. Il écrit ces deux nombres sur des bouts de papier qu'il place secrètement sous deux chapeaux. Il vous demande ensuite de choisir l'un des chapeaux et vous dévoile le nombre qui y était caché. Vous devez ensuite parier en répondant à la question : le nombre resté sous l'autre chapeau est-il plus grand que celui que l'on vient de vous dévoiler ?*

A priori, à choisir sans stratégie particulière mais seulement "au hasard" (en jouant à pile ou face, par exemple), vous n'avez qu'une chance sur deux de gagner. Existe t-il une stratégie de réponse qui vous garantit que vous avez plus de chances de donner la bonne réponse ? C'est ce que nous allons voir ...

Choisissez un nombre Z dans $[0,N]$ au hasard et selon une distribution de probabilité quelconque. Si Z est strictement plus grand que le nombre dévoilé, pariez que le nombre resté secret est plus grand que celui qui a été dévoilé. Montrez que cette stratégie est avantageuse.

Notons X le nombre qui vous a été dévoilé. Il s'agit de calculer la probabilité $P(X=m,Z>m)+P(X=n,Z<n)$ (ce sont les deux cas gagnants). Le croupier choisit l'un des deux chapeaux de manière équiprobable et les variables X et Z sont indépendantes, par conséquent :

$$\begin{aligned} & P(X=m,Z>m)+P(X=n,Z<n) \\ = \quad & P(X=m)P(Z>m)+P(X=n)P(Z\leq n) \\ = \quad & \tfrac{1}{2}\sum_{k=m+1}^{N} P(Z=k)+\tfrac{1}{2}\sum_{k=0}^{n} P(Z=k) \\ = 1/2+\sum_{k=m+1}^{n} & = 1/2+P(m+1\leq Z\leq n)\geq 1/2 \end{aligned}$$

Chapitre 4
Indépendance et probabilités conditionnelles

Les situations plus "réalistes" exigent d'étudier les *probabilités conditionnelles*. Des exemples illustreront bien ce propos. Imaginons que vous observiez le système de l'exemple en page 15 décrit à la section précédente et que vous constater que le système est actif environ deux fois sur trois. La question de connaître la probabilité d'observer le système en état actif quatre fois de suite ne se pose plus de la même manière. En termes probabilistes, vous cherchez à connaître la probabilité d'observer une séquence comptant quatre valeurs à 1 consécutives, sachant que le nombre total de valeurs à 1 est $2N/3$. Deux évènements sont parties prenantes ici. L'ensemble A des séquences comptant 4 valeurs à 1 consécutives, et l'ensemble B des séquences comptant $2N/3$ valeurs à 1. Et l'on cherche à calculer la probabilité de A *sachant* que l'on est dans la situation décrite par B.

Pensez à ces ensembles sous forme graphique. Puisque l'on suppose que l'évènement B s'est produit, on cherche à calculer combien d'éléments de A sont susceptibles de se produire. En termes ensemblistes, les évènements que l'on pourra observer sont ceux de $A \cap B$. La probabilité cherchée est ici $\frac{|A \cap B|}{|B|}$, ce qui est aussi égal à $\frac{P(A \cap B)}{P(B)}$.

C'est ce qu'on appelle la *probabilité conditionnelle de A sachant B*, et qu'on note $P(A|B)$:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{4.1}$$

Notez au passage qu'on a

$$P(A \cap B) = P(A|B)P(B).$$

Evènements indépendants L'évènement A ne dépend pas de l'évènement B si on a $P(A|B) = P(A)$: le fait qu'on suppose que B se soit produit n'influe pas sur la probabilité d'observer A. On en déduit $P(A \cap B) =$

$P(A)P(B)$ et que par conséquent, on a aussi $P(B|A) = P(B)$. On dit alors que ces évènements sont *mutuellement indépendants*.

Exercice 4.1 Poker *Vous jouez au poker, version Texas Hold'em*[1]*. Le croupier distribue les cartes et vous recevez vos deux cartes : un as et un valet de pique (!).*

1. *Quelle est la probabilité que vous obteniez une quinte flush royale ? (une série de cinq cartes dans la même couleur allant de l'as au 10).*
2. *Quelle est la probabilité que vous obteniez un full formé de trois as et de deux valets ?*

Exercice 4.2 Bits *Des bits sont envoyés à la suite sur un canal, et sont égaux soit à 0, soit à 1 avec même probabilité.*

1. *Vous enregistrés deux bits à la suite* b_0, b_1.
   - *Quelle est la probabilité que les deux bits aient des valeurs différentes (l'un vaut 0, l'autre vaut 1) ?*
   - *Quelle est la probabilité que le second bits soit égal à 1 si le premier est égal à 0 ?*
2. *On note maintenant la valeur de quatre bits envoyés à la suite.*
   - *Trouvez la probabilité d'avoir au moins deux bits égaux à 1 parmi les quatre.*
   - *Trouvez la probabilité d'avoir au moins deux bits égaux à 1 tout en étant certain (sachant que) que l'un d'entre eux l'est.*
   - *Trouvez la probabilité que tous les bits soient effectivement égaux à 1 sachant qu'au moins deux d'entre eux sont effectivement égaux à 1.*
   - *Trouvez la probabilité d'avoir au moins deux bits égaux à 1 tout en étant certain (sachant que) que l'un d'entre eux l'est et que le quatrième bit est à 0.*

Exercice 4.3 *D'après le tableau suivant, dans quel(s) cas les évènements* A *et* B *sont-ils indépendants ?*

	$P(A)$	$P(B)$	$P(A \cup B)$
cas I	*0.1*	*0.9*	*0.91*
cas II	*0.4*	*0.6*	*0.76*
cas III	*0.5*	*0.3*	*0.73*

Exercice 4.4 Simuler des probabilités conditionnelles *On considère un espace de probabilité* Ω *sur lequel on définit deux évènements* A *et* B*. On répète* N *fois une expérience* $\omega \in \Omega$ *et on note le nombre de fois où se produisent :*

1. Dans cette version du jeu, chaque joueur reçoit d'abord deux cartes. Puis cinq cartes sont découvertes au centre de la table. Le joueur qui "gagne" – on ignore ici l'effet des mises – est celui qui fabrique la meilleure main de cinq cartes en composant avec les siennes et les cartes communes à tous les joueurs.

– $n(A) =$ *nombre de réalisations de* A
– $n(B) =$ *nombre de réalisations de* B
– $n(A \cap B) =$ *nombre de réalisations simultanées de* A *et* B

puis on calcule les fréquences qui approxime les probabilités théoriques $P(A) \sim \frac{n(A)}{N}$, $P(B) \sim \frac{n(B)}{N}$, $P(A \cap B) \sim \frac{n(A \cap B)}{N}$ *et* $P(A|B) \sim \frac{n(A \cap B)}{n(A)}$.

(a) Simuler la probabilité conditionnelle d'obtenir, en lançant deux dés, une somme supérieure à 6 sachant qu'au moins un des deux dés est supérieur ou égal à 3.

(b) Simuler la probabilité conditionnelle d'obtenir, en jouant 100 *fois à pile ou face, la probabilité d'avoir pile 52 fois sachant que vous avez eu pile aux deux premiers lancers.*

Exercice 4.5 $P(A \cap B \cap C)$

Soient A, B, C *des évènements quelconques, tels que* A *et* $A \cap B$ *sont de probabilités non nulles. Utilisez la formule de probabilité conditionnelle pour déduire l'identité* $P(A)P(B|A)P(C|A \cap B) = P(A \cap B \cap C)$.

Exercice 4.6 Emission de bits *Sur une séquence* S *de* n *bits, on définit une procédure d'ajout d'un* $(n+1)$*ème bit* b*. La séquence* S *est reçue (sur un canal), puis on construit une séquence* S' *qu'on renvoie.*

Si $|S|_1$ *(le nombre de bits égaux à 1 dans* S*) est impair alors* $b = 1$ *sinon* $b = 0$*. On obtient ainsi une séquence de bit* S'*. L'émission des bits n'étant pas sécurisée, il se peut que la séquence* S' *ne soit pas cohérente. On suppose qu'un bit 1 de la séquence* S' *est émis avec la probabilité* p*. Dans quelle mesure l'ajout du* $(n+1)$*ème bit permet-t-il de contrôler que la séquence* S *n'a pas été altérée ?*

Exercice 4.7 *Montrer :*

$$P(A|B \cap C) = P(A|C) \cdot \frac{P(B|A \cap C)}{P(B|C)}$$

4.1 Formule de Bayes

Les probabilités conditionnelles ne sont pas symétriques, c'est-à-dire qu'il n'y a a priori aucune raison pour qu'on ait $P(A|B) = P(B|A)$.

Prenons encore une fois un exemple. Supposons que l'on mette au point un algorithme d'analyse du langage naturel qui détermine (de manière probabiliste) si une page web qu'il analyse est écrit en anglais "natif". Les concepteurs de l'algorithme sont assez confiant et estiment que l'algorithme dit juste 95% du temps. On estime par ailleurs que 55% des pages écrites en anglais et publiées sur le web sont en anglais natif. On peut donc chercher à évaluer la robustesse de l'algorithme en calculant la probabilité qu'il affirme qu'un page est en anglais natif alors que ce n'est pas le cas

(*faux positif*), ou qu'une page soit écrite en anglais natif alors que l'algorithme affirme le contraire (*faux négatif*). De la même manière, on peut considérer les cas où l'algorithme dit juste (*vrai positif* : l'algorithme affirme que la page est en anglais natif et c'est bien le cas). ; ou le cas où la page n'est pas en anglais natif comme l'affirme l'algorithme (*vrai négatif*).

Soit Ω l'ensemble des pages web écrites en anglais, et désignons par A l'évènement l'algorithme affirme qu'une page est en anglais natif, et B l'évènement "une page est écrite en anglais natif". La probabilité d'observer un faux positif est donc $P(A|B^c)$. La probabilité d'observer un faux négatif est $P(B|A^c)$.

Dans l'exemple, il est vraisemblable d'évaluer les probabilités $P(A|B)$ et $P(A|B^c)$ en faisant tourner l'algorithme sur un échantillon de pages web dont on contrôle *a priori* la qualité de l'anglais. (Il nous faut aussi pouvoir évaluer la probabilité $P(B)$). Mais qu'en est-il de $P(B|A)$? Il se trouve que ces deux probabilités sont liés par la **formule de Bayes** :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{4.2}$$

que l'on déduit simplement de la définition des probabilités $P(A|B)$ et $P(B|A)$.

Ainsi, on peut faire jouer la dépendance d'une variable sur l'autre à condition de pouvoir connaître au moins en partie (et dans l'un des deux sens) les probabilités conditionnelles.

Selon que l'on connaît la valeur $P(A)$ ou pas, on pourra utiliser la formule (4.2) ou une forme étendue :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} \tag{4.3}$$

Pour montrer la formule (4.3), il suffit d'utiliser l'identité (4.1). On trouve :

$$\begin{aligned} P(B|A) &= \frac{P(B \cap A)}{P(A)} \\ &= \frac{P(A \cap B)P(B)}{P(A)P(B)} \\ &= \frac{P(A|B)P(B)}{P(A)} \\ &= \frac{P(A|B)P(B)}{P(A \cap B) + P(A \cap B^c)} \\ &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} \end{aligned}$$

Exercice 4.8 Langue naturelle ...

On met au point une heuristique (un algorithme) qui détecte si une page web écrite en anglais est rédigée par un natif *(quelqu'un dont l'anglais est la langue maternelle). On évalue à 55% le pourcentage de pages sur le web qui sont écrites en anglais par des natifs. L'heuristique réussit à détecter correctement que la page est écrite par un natif dans 95% des cas lorsque la page est effectivement écrite par un natif. Elle affirme cependant incorrectement que la page est écrite par un natif alors que ce n'est pas le cas avec probabilité 1%.*

Quelle est la probabilité qu'une page soit écrite par un natif lorsque l'heuristique l'affirme ?

Exercice 4.9 Processeurs en panne *Un contrôle systématique est effectuée sur un ensemble de processeurs dont 15% présentent une panne non apparente. Ce dépistage est débuté par un test qui donne 95% de résultats positifs pour les processeurs défectueux, et 10% de résultats positifs pour les processeurs non défectueux. Quelle est la probabilité (conditionnelle) qu'un processeur pris au hasard présente la panne sachant que le test a donné un résultat positif ?*

Exercice 4.10 Bayes encore *Montrez une version généralisée de la formule de Bayes qui implique trois évènements, A, B, C disjoints et complémentaires, c'est-à-dire que $A \cup B \cup C = \Omega$ et $A \cap B \cap C = \emptyset$. Soit un autre évènement D. Exprimez $P(A|D)$ en fonction des probabilités conditionnelles $P(B|D)$, $P(C|D)$ (et des probabilités $P(A), P(B), P(C)$, etc.).*

Exercice 4.11 Barrettes mémoire *Une usine fabrique des barrettes mémoire à l'aide de trois machines A, B et C. La machine A assure 20% de la production et 5% des barrettes fabriquées par A sont défectueuses. La machine B assure 30% de la production et 4% des barrettes fabriquées par B sont défectueuses. La machine C assure 50% de la production et 1% des barrettes fabriquées par C sont défectueuses.*

1. *On choisit au hasard une barrette. Calculer les probabilités :*
   - *pour que la barrette soit défectueuse et produite par A,*
   - *pour que la barrette soit défectueuse et produite par B,*
   - *pour que la barrette soit défectueuse et produite par C.*
2. *Calculer les probabilités pour qu'une barrette défectueuse :*
   - *provienne de A,*
   - *provienne de B,*
   - *provienne de C.*

Exercice 4.12 Règles des causes totales *Soient A, B, C, D des évènements tels que B, C et D sont mutuellement exclusifs et tels que $B \cup C \cup D = \Omega$. Montrez qu'on a :*

$$P(A) = P(A|B)P(B) + P(A|C)P(C) + P(A|D)P(D).$$

Proposez une généralisation de cette règle des causes totales.

Exercice 4.13 Paquets perdus *L'envoi d'un paquet du serveur S_1 au serveur S_2 sur internet passe par deux routeurs intermédiaires R_1 et R_2. La probabilité que le paquet se perde au niveau S_1, R_1 ou R_2 est p. On constate, au niveau du serveur S_2 la perte du paquet. Quelle est la probabilité qu'il ait été perdu au niveau de S_1 ? de R_1 ? de R_2 ?*

Solution On sait que la probabilité qu'un paquet se perde sur chaque segment du chemin $S_1 \to R_1 \to R_2 \to S_2$ est p. Notons par $P(S_1 \to R_1)$, $P(R_1 \to R_2)$, $P(R_2 \to S_2)$ les probabilités que la perte se produise sur un des segments du chemin. On peut calculer la probabilité que le paquet se perde (constaté au niveau du serveur S_2) : soit le paquet est perdu sur le premier segment, soit sur le deuxième, soit sur le troisième. Cette probabilité est donc égale à $P(S_1 \to R_1) + P(R_1 \to R_2) + P(R_2 \to S_2) = p + (1-p)p + (1-p)^2 p$.

Les probabilités conditionnelles cherchées sont donc $P(\text{perte}|S_1 \to R_1) = P(\text{perte})/P(S_1 \to R_1) = p/(p+(1-p)p+(1-p)^2 p)$. Les autres valeurs se calculent de manière similaire $P(\text{perte}|S_1 \to R_1) = P(\text{perte})/P(S_1 \to R_1) = p/(p+(1-p)p+(1-p)^2 p)$., $P(\text{perte}|R_1 \to R_2) = (1-p)p/(p+(1-p)p+(1-p)^2 p)$, $P(\text{perte})/P(R_2 \to S_2) = (1-p)^2 p/(p+(1-p)p+(1-p)^2 p)$.

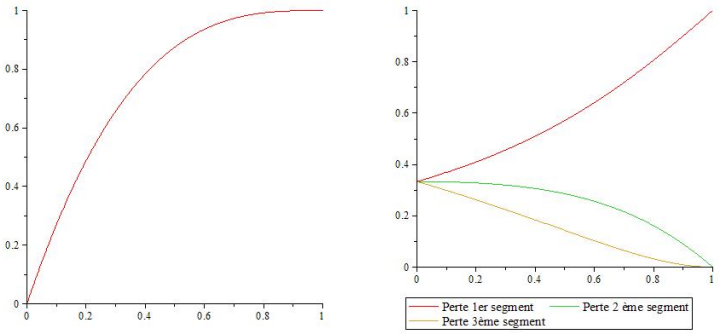


Fig. 4.1 La courbe de gauche indique comment évolue la probabilité de perdre un paquet en fonction de p (la probabilité de perdre un paquet sur l'un des segments. Les courbes de la figures de droite donne l'évolution des probabilités conditionnelles en fonction de p. Les probabilités conditionnelles concernant les deux derniers segments décroissent alors que la probabilité conditionnelle concernant le premier segment tend vers 1.

Il est intéressant de voir comment évolue ces probabilités conditionnelles en fonction de la valeur de p. A l'évidence, plus p augmente, plus il est raisonnable d'attribuer la perte d'un paquet au segment $S_1 \to R_1$. On

observe alors naturellement une décroissance des probabilités conditionnelles impliquant les deux derniers segments du chemin [2].

Exercice 4.14 Paquets endommagés *Un routeur reçoit l'essentiel des paquets qu'il fait transiter depuis deux autres routeurs R_1 et R_2. On constate qu'environ un paquet sur 200 est endommagé*[3] *lorsque ceux-ci proviennent de R_1, alors que seulement un paquet sur 1000 est endommagé lorsqu'ils proviennent de R_2. On considère un lot de paquets reçus de l'un de ces deux routeurs R_1 ou R_2. Le premier paquet inspecté n'est pas endommagé. Quelle est la probabilité pour que le second paquet ne soit pas, lui non plus, endommagé ?*

Exercice 4.15 Moteurs de recherche *On met au point un algorithme qui lance un même requête sur trois moteurs de recherche M_1, M_2 et M_3. La précision de ces moteurs de recherche est évaluée en fonction de la pertinence des réponses qu'ils retournent, et de manière probabiliste : M_1 renvoie une réponse pertinente 3 fois sur 4, M_2 1 fois sur 4 et M_3 1 fois sur 2. Après avoir lancé ses requêtes, l'algorithme retient deux des réponses comme étant pertinentes et rejette la troisième. Quelle est la probabilité que la réponses rejetée proviennent de M_1 ? de M_2 ? de M_3 ?*

Exercice 4.16 *On tire au hasard deux points sur le segment $[0,1)$ indépendamment l'un de l'autre. Le plus petit nombre est supérieur à 1/3. Quelle est la probabilité pour que le plus grand soit supérieur à 3/4 ?*

Exercice 4.17 Spam *Vous venez d'installer un module de détection de courriers indésirables (spam) dans votre client de courrier électronique. Le module réussit à identifier les courriers indésirables dans 99% des cas. Son taux de* faux positifs *est toutefois de 2%*[4]*. Les statistiques officielles indiquent que 10% du courriers électroniques reçus est indésirable. Quelle est la probabilité qu'un message soit effectivement indésirable lorsque le module indique que c'est le cas ?*

---

2. Les expressions pour les probabilités conditionnelles convergent vers 1/3 lorsque $p \to 0$. Elles sont donc non continues en ce point puisque ces probabilités sont évidemment nulles si $p = 0$.

3. On suppose disposer d'un procédé pour déterminer si les données d'un paquet ont subi un dommage.

4. Ce sont les cas où le module annonce qu'un message est indésirable alors qu'il ne l'est pas.

Chapitre 5
Variables aléatoires, espérance et variance

Connaître la probabilité d'un évènement A, c'est pouvoir calculer la "taille" de l'ensemble sous-jacent à A. Les choses sont parfois moins "directes", et décrites de manière composite : dans un casino, vous misez 2 € et vous lancez trois dés. Si vous avez un triple vous empochez 5 € (vous gagnez 3 €), si vous avez un double vous empochez 3 € (vous gagnez 1 €). Sinon, vous perdez votre mise (vous gagnez -2 €).

Naturellement, on s'intéresse à la probabilité de gain. Quelle est la probabilité de perdre sa mise, de gagner 1 € ou 3 € ? Les gains dépendent évidemment des évènements possibles lors du lancer des dés, mais les probabilités cherchées sont formulées en fonction du gain, et non pas en fonction des évènements eux-mêmes. En d'autres termes, les probabilités dépendent toujours des évènements sur $\Omega = \{1, \ldots, 6\}^3$ mais les probabilités sont formulées en fonction du gain possible du joueur.

Il faut voir le gain comme une fonction $X : \Omega \to \mathbb{N}$ dont le domaine est l'espace où sont lancés les dés et à valeurs dans l'ensemble des entiers. Cette fonction est une *variable aléatoire*. Ainsi, si le lancer de dés est $\omega = (3, 3, 3)$ (un triple de 3) alors le gain du joueur est de $X(\omega) = 3$.

On peut ainsi se poser la question de connaître la probabilité d'avoir un gain de 3 € qu'on écrit $P(X = 3)$. La réponse se fait en calculant l'évènement $A = X^{-1}(3)$ dans Ω, c'est l'ensemble de tous les lancers de dés qui donne un gain de 3 €. Dans cet exemple, on a $A = \{(1, 1, 1), \ldots, (6, 6, 6)\}$ et la probabilité est donc $P(X = 3) = |A|/|\Omega| = 1/36$.

Exercice 5.1 Casino *Reprenez l'exemple du joueur de casino qui lance les dés. Calculez les probabilités $P(X = 1)$ et $P(X = -2)$.*

Exercice 5.2 Système actif/passif *On effectue une séquence de N observations d'un système qui se trouve dans l'un de deux états 0 ou 1, avec probabilité p d'être actif (valeur observée = 1). On considère la variable aléatoire qui donne le nombre de fois où le système est actif (on observe la valeur 1).*

Précisez ce qu'est Ω, définissez la variable aléatoire X. Calculez la probabilité $P(X = k)$.

Exercice 5.3 Phénotypes et génotypes *Certains caractères héréditaires sont portés par paires de gènes, un gène pouvant prendre deux formes (allèles), dominantes ou récessives (couleur des yeux, caractère de la peau des petits pois, etc.). Ces gènes sont donc groupés par paires et se présentent dans l'organisme sous trois* génotypes *possibles : AA, Aa, aa (on ne distingue pas Aa et aA). Le phénotype suit de la dominance du caractère A sur a. L'organisme porteur montrera le phénotype qui suit du gène dominant A mais pourra tout de même transmettre la gène récessif a. Ainsi, une personne aux yeux marron peut très bien avoir des enfants aux yeux bleus.*

Chaque parent transmet à un descendant l'un des deux gènes qu'il porte. On suppose que le gène transmis par le parent est choisi au hasard parmi les deux possibles.

On définit deux variables aléatoires X_1, X_2 qui prennent les valeurs des phénotypes AA, Aa ou aa des deux parents et Y_1, Y_2 qui prennent les valeurs du gène A ou a transmis par chacun des deux parents à leur descendant.

Sachant que le gène transmis par le parent est choisi au hasard parmi les gènes de son génotype, et sachant que le gène transmis par l'un des parents ne dépend pas de celui transmis par l'autre, donnez les probabilités conditionnelles des variables Y_i sachant X_j.

On suppose que les génotypes AA, Aa, aa sont répartis dans la population avec probabilité $p, 2r, q$ (avec $p+2r+q=1$). Exprimez p, q et r en fonction des probabilité sur les Y_i. (Vous utiliserez la règle des causes totales.)

Exercice 5.4 Guêpe et probabilité conditionnelle *Deux pièces A et B sont reliées entre elles par une porte ouverte. Seule la pièce B possède une issue vers l'extérieur. Une guêpe initialement dans la pièce A voudrait sortir à l'air libre. Son trajet obéit aux règles suivantes :*

- *Lorsqu'elle est en A au temps $t = n$, alors au temps $t = n+1$, elle reste en A avec une probabilité égale à 1/3 ou elle passe en B avec une probabilité égale à 2/3,*
- *Lorsqu'elle est en B au temps $t = n$, alors au temps $t = n+1$, elle retourne en A avec une probabilité égale à 1/4, ou elle reste en B avec une probabilité égale à 1/2, ou elle sort à l'air libre avec une probabilité égale à 1/4.*

Au temps $t = 0$, la guêpe est dans la pièce A. Lorsqu'elle est sortie, elle ne revient plus.

1. *Calculez explicitement les distributions de probabilités des variables X_0 et X_1,*
2. *Exprimez $P(X_{n+1} = A)$ et $P(X_{n+1} = B)$ en fonction de $P(X_n = A)$ et $P(X_n = B)$ (avec des notations évidentes),*
3. *Vérifiez que la suite $\frac{6}{10}P(X_n = A) - \frac{3}{10}P(X_n = B)$ est constante,*
4. *Vérifiez que la suite $\frac{4}{10}P(X_n = A) + \frac{3}{10}P(X_n = B)$ est géométrique de raison $\frac{5}{6}$,*
5. *En déduire l'expression de $P(X_n = A)$ et $P(X_n = B)$,*

6. *Montrer que pour* $n \geq 2$, $P(X_n = S) = \frac{1}{4}\sum_{k=1}^{n-1} P(X_{n-1} = B)$. *En déduire* $P(X_n = S)$.

Exercice 5.5 *Soit* $a \in (0, 1/2)$ *un nombre réel.*

Dans une bourse de valeurs, un titre donné peut monter, *rester* stable *ou* baisser.

Dans un modèle mathématique, on considère que :

- *le premier jour le titre est stable*
- *si un jour* n *le titre monte, le jour* $n+1$ *; il montera avec probabilité* $1-2a$, *restera stable avec probabilité* a *et baissera avec probabilité* a
- *si un jour* n *le titre est stable, le jour* $n+1$ *il montera avec probabilité* a, *restera stable avec probabilité* $1-2a$ *et baissera avec probabilité* a
- *si un jour* n *le titre baisse, le jour* $n+1$ *il montera avec probabilité* a, *restera stable avec probabilité* a *et baissera avec la probabilité* $1-2a$

On note M_n *(resp.* S_n, *resp.* B_n*) l'événement "le titre donné monte" (resp. reste stable, resp. baisse) le jour* n. *On pose* $p_n = P(M_n)$, $q_n = P(S_n)$ *et* $r_n = P(B_n)$.

a) Que vaut $p_n + q_n + r_n$ *? En déduire l'expression de* r_n *en fonction de* p_n *et* q_n,

b) Expliciter p_{n+1} *(resp.* q_{n+1}*) en fonction de* p_n, q_n, r_n,

c) En déduire p_n, q_n *puis* r_n,

d) Donner la limite de ces trois suites et interpréter le résultat.

5.1 Distribution de probabilité

Une variable aléatoire calcule donc une valeur associée à toute épreuve $\omega \in \Omega$ ou à tout évènement $A \subset \Omega$. Elle peut prendre ses valeurs dans tout ensemble, bien que nous considérerons le plus souvent des variables aléatoires à valeurs dans $\mathbb{N}$ ou $\mathbb{R}$.

La distribution de probabilité d'une variable aléatoire décrit la probabilité associée à chacune des valeurs qu'elle peut prendre. certaines distributions de probabilités ont un caractère générique.

5.1.0.1 Distribution binomiale

Un évènement aléatoire à valeur 0 ou 1, avec probabilité p de succès (valeur = 1) est appelé un *évènement de Bernouilli* de probabilité p. L'évènement de Bernouilli par excellence est le tirage d'une pièce à pile ou face. Observez l'état 0 ou 1 d'un système est aussi modélisé par un évènement de Bernouilli.

On considère une suite de variables aléatoires $X_1, \ldots, X_N$ où les variables X_i correspondent à des évènements de Bernouilli *indépendants* de probabilité p. On considère la variable aléatoire $S = X_1 + \cdots X_N$.

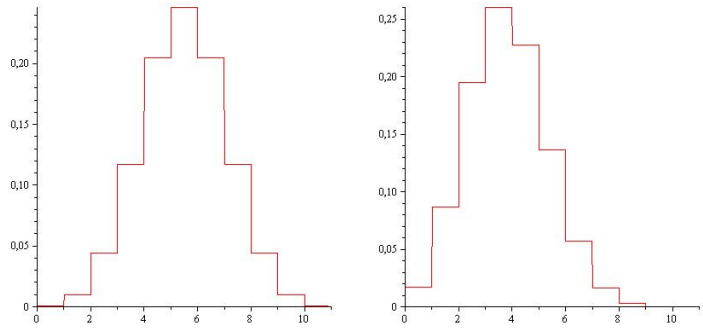


Fig. 5.1 La distribution binomiale peut être décrite à l'aide d'un histogramme ou d'une courbe brisée. Observez sa symétrie lorsque $p = 1/2$ (gauche), qui est perdue lorsque $p \neq 1/2$ (droite avec $p = 1/3$).

Exercice 5.6 *Donnez la distribution de probabilité de S. Cette distribution est la* distribution binomiale *de paramètres (N, p) que l'on a déjà vu à l'exercice 5.2.*

Une variable aléatoire est elle-même une fonction $X : \Omega \to \mathbb{N}$ (ou $\mathbb{R}$), que l'on peut composer avec une seconde fonction $f : \mathbb{R} \to \mathbb{R}$. On peut, par exemple, considérer la fonction $f : x \mapsto x^2$ et considérer une nouvelle variable aléatoire $f(X)$ qui calcule le carré de la valeur retournée par X. Il faut souvent s'assurer de conditions satisfaite par la fonction f, mais nous aurons le plus souvent affaire à des cas simples où la fonction est continue et monotone croissante.

5.2 Espérance

Dès lors que l'on considère une variable aléatoire, on s'intéresse aux valeurs qu'elle peut prendre et aux probabilités qu'elle atteigne effectivement ces valeurs. On s'intéresse naturellement à la valeur moyenne de la variable, qui donne une première indication de son "comportement". Dans l'exemple du joueur au casino, on peut effectivement se demander de combien sera le gain du joueur en moyenne. En réalité, ce sont les gérants du casino qui doivent prendre soin de calculer cette valeur, puisque le joueur par nature espère pouvoir échapper à la loi des grands nombres !

Chacun sait calculer la moyenne de la variable aléatoire qui donne le résultat d'un lancer de dé. (Cette variable est horriblement simple, puisqu'elle est définie sur l'ensemble $\Omega = \{1, \ldots, 6\}$ et retourne la valeur du dé lancer). Vous faites la somme des valeurs possibles divisée par le cardinal de l'ensemble, et vous trouvez 3,5. Ce cas parait facile car il est particulier : les valeurs prises par la variable sont équiprobables. Vous procédez de même si vous calculez la moyenne de votre semestre, *lorsque toutes les matières ont même coefficient.* Et si les coefficients varient ? Vous en tenez compte dans le calcul, vous effectuez alors une somme pondérée – ce raisonnement vaut aussi pour calculer la moyenne d'une variable aléatoire.

Prenons un autre exemple, celui où la variable aléatoire Y retourne la somme de deux dés lancés sur le tapis. L'espace de probabilités est $\Omega = \{(i, j)|i, j \in \{1, \ldots, 6\}\}$. Elle prend donc ses valeurs dans $\{2, \ldots, 12\}$. On a $P(Y = 2) = 1/36$, $P(Y = 3) = 2/36$, $P(Y = 4) = 3/36, \ldots$ La moyenne de la variable doit tenir compte des valeurs prises par la variable, mais aussi des probabilités d'atteindre chacune de ces valeurs. On définit l'espérance d'un variable aléatoire X :

$$\mathbb{E}(X) = \sum_k k \cdot P(X = k) \tag{5.1}$$

où la somme parcourt l'ensemble des valeurs prises par la variable (nous n'avons considéré que des valeurs entières jusqu'à maintenant et ça nous suffira pendant un moment – les choses ne sont pas foncièrement différente si on passe aux variables à valeurs dans R).

Exercice 5.7 Casino – encore *Quel est l'espérance du gain du joueur de dés au casino ?*

Exercice 5.8 Une formule simple *Soit $X : \Omega \to \mathbb{N}$ une variable aléatoire, montrez que $\mathbb{E}(X) = \sum_{n \geq 0} P(X \geq n)$.*

Définition 1. Fonction indicatrice Sur tout espace de probabilité Ω, et pour tout évènement $A \subset \Omega$, on peut construire une variable aléatoire $1_A : \Omega \to 0, 1$ qui vaut $1_A(\omega) = 1 \iff \omega \in A$. En d'autres mots la variable indique si une épreuve ω appartient à A. Ainsi, on a $P(1_A = 1) = P(A)$.

Exercice 5.9 *Observez $\mathbb{E}(1_A) = P(A)$.*

Exercice 5.10 Loi binomiale *Calculer l'espérance d'une variable suivant la loi binomiale de paramètres (N, p).*

Exercice 5.11 Analyse d'algorithme *On se propose d'analyser la complexité (temps de calcul) en moyenne d'un algorithme. On considère l'algorithme ci-dessous pour trouver le plus grand élément dans une liste non vide L de n entiers :*

```
début
```

```
(1) M := L[1]
(2) pour j := 2 à n faire
(3)   si L[j] > M alors M := L[j]
fin
```

Lors du traitement d'une liste de taille n :
- *Quel est le nombre de comparaisons effectuées (ligne 3) ?*
- *Quel est le nombre d'affectations (lignes 1 et 3) dans le cas le plus favorable ? le plus défavorable ?*

Dans la suite, on suppose que les éléments de la liste sont deux-à-deux distincts et que les $n!$ permutations possibles ont la même probabilité d'apparaître dans la liste. Soit $p(n,k)$ la probabilité pour que, lors du déroulement de l'algorithme sur une liste (aléatoire) L de taille n, k affectations soient nécessaires.
- *Montrez la récurrence suivante :*

$$p(n,k) = \frac{1}{n}p(n-1,k-1) + \frac{n-1}{n}p(n-1,k)$$

pour $n \geq 2$ et $1 \leq k \leq n$. Il faut distinguer deux cas de figures et raisonner selon que le n^{me} élément est maximal ou non.
- *Soit $\mathbb{E}_n$ l'espérance mathématique (la moyenne) du nombre d'affectations effectuées lors du traitement d'une liste de taille n. Déduire de la récurrence ci-dessus la récurrence sur $\mathbb{E}_n$*

$$\mathbb{E}_n = \frac{1}{n} + \mathbb{E}_{n-1},$$

dont on peut déduire $\mathbb{E}_n = \sum_{k=1}^{n} \frac{1}{k}$.
- *Montrez que l'ordre de grandeur de $\mathbb{E}_n$ se compare à $\log n$ lorsque $n \to \infty$.*

Exercice 5.12 Simuler un loi binomiale *Soit la variable aléatoire $S = X_1 + \cdots + X_n$ où les X_i correspondent à des évènements de Bernouilli avec probabilité p. Donnez un procédé permettant de simuler cette variable aléatoire. En d'autres mots vous tirez au hasard des entiers variant de 0 à N, non pas de manière uniforme mais de façon à ce que la probabilité d'obtenir l'entier k soit égale à $P(S = k)$.*

5.2.0.2 Loi géométrique

Une variable aléatoire X suit une loi géométrique de paramètre p si $P(X = k) = (1-p)^k p$. Cette loi a une interprétation assez intuitive :
- Vous jouez à pile ou face, et vous lancez la pièce tant que vous n'avez pas obtenu pile (avec probabilité p d'obtenir pile). La variable qui indique le nombre de fois où vous avez lancé la pièce avant d'obtenir pile suit une loi géométrique.

– Vous observez un signal (0 ou 1), et vous poursuivez l'observation en notant à quel moment (combien de signaux reçus) vous avez vu avant d'observer la valeur 1.

Exercice 5.13 *Calculez l'espérance et la variance de la loi géométrique.*

Solution Ce calcul fait appel, sans surprise, à la manipulation de séries géométriques (voir section 2.1). Désignons par X la variable de loi géométrique de paramètre p. La série $s(q) = \sum_{i\geq 0} q^i = \frac{1}{1-q}$ peut être vue comme une fonction réelle de la variable $q \in [0,1)$, que l'on peut donc dériver $s'(q) = \sum_{i\geq 1} iq^{i-1} = \frac{-1}{(1-q)^2}$, $s''(q) = \frac{1}{(1-q)^3}$.

$$\mathbb{E}(X) = \sum_{k\geq 1} k(1-p)^{k-1}p = ps'(1-p) = \frac{1}{p}$$

Le même type de raisonnement (en prenant soin de remarquer que $k^2 = k(k-1)+k$ permet de trouver $\mathbb{V}(X) = \frac{1-p}{p^2}$.

Exercice 5.14 Le collectionneur *Imaginez que vous collectionnez des objets offerts à chaque achat au supermarché. La collection compte 9 objets différents et vous vous mettez en tête de continuer à faire vos courses à ce supermarché pour faire vos courses, chaque semaine disons, jusqu'à ce que vous ayez la collection complète. Pendant combien de semaines aurez-vous à faire vos courses avant d'avoir obtenu la collection complète (en moyenne) ?*

Modélisez la situation ainsi. Vous considérez les variables $X_1, \ldots, X_9$ qui indique le temps d'attente (en nombre de semaines) avant d'obtenir une $i$ème carte. On a nécessairement $X_1 = 1$ puisque la première carte démarre la collection. Qu'en est-il de X_2 ? ..., puis X_9 ?

Exercice 5.15 Simuler un loi géométrique *Simulez une loi géométrique. En d'autres mots, donnez un procédé qui permet de choisir un entier au hasard tel que l'entier k soit choisi avec probabilité $(1-p)^k p$. On peut procéder de deux manières différentes. Rappelons d'abord que $\sum_{k=0}^{\infty}(1-p)^k p = 1$.*

– *On choisit au hasard un nombre réel $r \in [0,1)$. On calcule l'entier i tel que $i = \inf_i \sum_{k=0}^{i}(1-p)^k p > r$. C'est l'entier choisit au hasard par la loi géométrique.*
– *On choisit un nombre réel $r \in [0,1)$. On calcule $i = \lfloor \frac{\log U}{\log 1-p} \rfloor + 1$. C'est l'entier retourné par la loi géométrique.*
– *Montrer que la variable $X = \lfloor \frac{\log U}{\log 1-p} \rfloor + 1$ suit bien une loi géométrique (de paramètre p). En effet, puisque $P(X = k) = P((1-p)^{k-1} < U \leq (1-p)^k)$.*

5.2.0.3 Distribution de Poisson

Imaginons des évènements de Bernouilli qui se produisent de manière indépendante et qui se répètent à des temps réguliers. On peut penser à la réception de paquets provenant de N canaux distincts, à la détection de noyaux d'hélium émis par des atomes d'uranium, etc. Cette façon de voir les choses modélise bien la réalité sur un temps très court. Un canal émettra, ou non, un seul paquet ; un atome d'uranium émettra, ou non, un seul noyau d'hélium.

Comme précédemment, on peut considérer la variable $S_N = X_1 + \cdots + X_N$ qui compte le nombre de succès observés sur un intervalle de temps très court, la difficulté étant qu'on ne connaît pas le nombre N d'évènements – puisqu'on observe que le nombre k de succès. En revanche, on suppose pouvoir évaluer le nombre moyen λ de succès observés sur un intervalle de temps.

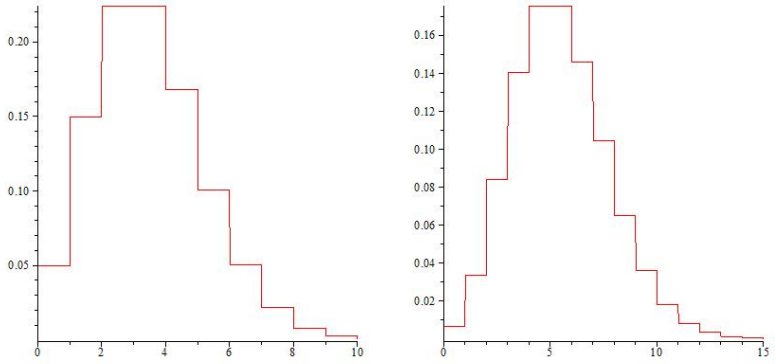


Fig. 5.2 La distribution de Poisson suit une courbe centrée en λ, qui décroît exponentiellement et tend asymptotiquement vers 0 lorsque $k \to +\infty$ (gauche $\lambda = 2$, droite $\lambda = 5$).

Or, on sait que la variable S_N suit une loi binomiale de paramètres (N, p_N), et par conséquent on a $\mathbb{E}(S_N) = Np_N$, et donc, $p_N = \lambda/N$.

Cela étant posé, la probabilité de n'observer aucun succès est égale à $P(S_N = 0) = (1 - p_N)^N = (1 - \lambda/N)^N$. On a donc :

$$\lim_{N\to\infty} P(S_N = 0) = e^{-\lambda}$$

et, pour $N > k$:

$$\begin{aligned}\frac{P(S_N = k+1)}{P(S_N = k)} &= \frac{\binom{N}{k+1} p^{k+1}(1-p)^{N-k-1}}{\binom{N}{k} p^k (1-p)^{N-k}} \\ &= \frac{N-k}{k+1}\frac{p}{1-p} \\ &= \frac{\lambda}{k+1}\end{aligned}$$

L'unique solution satisfaisant ces conditions nous donne :

$$\lim_{N\to\infty} P(S_N = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

La distribution $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ où λ est un nombre réel positif est appelé la *distribution de Poisson* de paramètre λ.

Exercice 5.16 Dimensionner un tableau *On doit prédimensionner un tableau, qui est une ressource pour un ensemble de processus : un processus qui s'exécute a besoin d'une entrée dans le tableau. Si aucune entrée n'est disponible alors il est mis en attente dans une file. On sait qu'en moyenne λ processus s'exécutent en même temps. Comment dimensionner le tableau pour que la probabilité d'avoir à mettre (au moins) un processus en attente soi au plus de 10%.*

Solution On suppose donc que le nombre de processus à traiter (à chaque moment) est donné par une loi de Poisson de paramètre λ. Ce nombre, k, doit être tel que la probabilité de dépasser la taille du tableau, N soit au plus de 10%. En d'autres mots, on doit avoir $\sum_{k>N} P(X = k) < 10\%$, ou encore $\sum_{k\leq N} P(X = k) \geq 90\%$. La solution numérique exige de recourir à une table de la loi de Poisson (ou encore de simuler la loi avec le paramètre λ). Ainsi, lorsque $\lambda = 10$ un tableau de taille $N = 14$ suffit ; avec $\lambda = 5$, on peut prendre $N = 8$.

Exercice 5.17 Tickets *Le nombre de tickets postés sur un serveur du service informatique au cours de la première heure de la journée est environ de 10. Utilisez la loi de Poisson pour calculer la probabilité qu'en un jour normal, il n'y ait aucun ticket qui soit posté pendant les 90 première minute de la première heure.*

Solution Nous allons considérer trois variables aléatoires X, Y, Z, chacune suivant une loi de Poisson de paramètres différents. La variable X est de paramètre $\lambda = 10$ où l'unité de temps sous-jacente est l'heure. La variable Y est de paramètre $\lambda = 15$ et l'unité de temps sous-jacente est de 90 minutes. finalement, Z est de paramètre $\lambda = 5$ et l'unité de temps sous-jacente est la demi-heure. Chacune de ces variables modélise le phénomène étudiée, mais sur des intervalles de temps différents.

La valeur cherchée est $P(Y=0)$. On a $P(Y=0) = \sum_{k\geq 0} P(Y=0|X=k)P(X=k) = P(Y=0|X=0)P(X=0)$ puisque l'intersection $Y_0 \cap X = k$ est vide sauf si $X=0$. On a donc :

$$P(Y=0) = P(Y=0|X=0)P(X=0) = P(Z=0) = e^{-5} \sim 0.00673794$$

Exercice 5.18 Clients *Les clients arrivent dans une banque à un rythme moyen de 40 clients par heure qu'on modélise par une loi de Poisson de paramètre 40 (avec l'unité de temps qui est l'heure). Sachant que 30 client sont arrivés à la première heure, quelle est la probabilité pour qu'il y en ait 60 qui arrivent dans les premières 90 minutes ?*

5.3 Variance

La moyenne donne une première indication du comportement d'une variable aléatoire (ou de la distribution de ses valeurs). La variance vient compléter cette information et nous renseigne sur l'étalement des valeurs autour de la moyenne : les valeurs sont-elles en moyenne proches ou éloignées de la moyenne ?

Deux variables aléatoires de même moyenne peuvent différer au niveau de la variance. Pensez par exemple à une variable aléatoire de Rademacher de paramètre $p = 1/2$, et dont l'espérance est nulle (voir l'exercice 3.8). Une variable qui vaut 10^6 ou 10^{-6} avec même probabilité a elle aussi une espérance nulle, mais varie sur un tout autre domaine.

La traduction littérale de la définition que nous avons donnée de la variance nous amène à :

$$\mathbb{V}(X) = \mathbb{E}(\big(X - \mathbb{E}(X)\big)^2) \tag{5.2}$$

La variance de la variable X sera aussi parfois notée σ_X^2. L'*écart-type* est la racine carrée de la variance, et est notée $\sigma_X = \sqrt{\sigma_X^2}$.

Exercice 5.19 Casino – encore *Quel est la variance du gain du joueur de dés au casino ?*

Exercice 5.20 *Soit X une variable aléatoire. Observez que la variable $X - \mathbb{E}(X)$ est de même variance que X.*

Exercice 5.21 Loi binomiale *Calculer la variance d'une variable suivant la loi binomiale de paramètres (N,p).*

Exercice 5.22 Poisson *Calculez l'espérance et la variance de la distribution de Poisson de paramètre λ.*

Exercice 5.23 Algorithme *Un algorithme reçoit en entrée une liste d'entiers positifs. Sa complexité est fonction du nombre d'entiers pairs p de la liste et vaut $f(p) = p^2+1$. Décrivez le comportement de l'algorithme (sa complexité moyenne et sa variance en temps d'exécution).*

5.4 Somme et produit de variables aléatoires

Soient deux variables aléatoires indépendantes X_1, X_2 à valeurs dans $\mathbb{R}$. Comme ces deux variables sont indépendantes, on a :

$$P(X_1 + X_2 = z) = \sum_{x+y=z} P(X_1 = x)P(X_2 = y)$$

$$P(X_1X_2) = \sum_{xy=z} P(X_1 = x)P(X_2 = y)$$

Par conséquent, on peut montrer :

Exercice 5.24

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) \tag{5.3}$$

$$\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2) \tag{5.4}$$

lorsque les variables X_1 et X_2 sont indépendantes.

De même, si $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes à valeurs dans $\mathbb{R}$, alors :

$$\sigma^2_{X_1+\cdots+X_n} = \sigma^2_{X_1} + \cdots + \sigma^2_{X_n} \tag{5.5}$$

Exercice 5.25 Démonstration *Il suffit de démontrer l'identité (5.5) lorsque $n = 2$.*

Exercice 5.26 Scalaire *Soit une variable aléatoire X à valeurs dans $\mathbb{R}$ et $\alpha \in \mathbb{R}$ positif. On considère la variable aléatoire αX. Montrez :*

$$\mathbb{E}(\alpha X) = \alpha\mathbb{E}(X), \quad \mathbb{V}(\alpha X) = \alpha^2\mathbb{V}(X)$$

Comparez ce résultat aux identités (5.3) et (5.5).

Exercice 5.27 *Soit X une variable aléatoire correspondant à un évènement de Bernouilli de paramètre p. Calculez son espérance et sa variance. En déduire directement le calcul de l'espérance et de la variance d'une variable aléatoire suivant une loi binomiale de paramètres N et p.*

Solution La variable X prend deux valeurs 0 ou 1 avec probabilité $1-p$ et p respectivement. Son espérance est donc $\mathbb{E}(X) = 0 \cdot (16p) + 1 \cdot p = p$. Sa variance est $(0 - \mathbb{E}(X))^2 \cdot (1-p) + (1 - \mathbb{E}(X))^2 \cdot p = p(1-p)$.

Une variable $S_N = X_1 + \cdots + X_N$ de loi binomiale de paramètres N et p est une somme de N variables X_i ($i = 1, \ldots, N$) de Bernouilli indépendantes de même paramètre p. On déduit le calcul de l'espérance et de la variance de S_N des équations (5.3) et (5.5).

Chapitre 6
Lois des grands nombres

L'espérance (la moyenne) nous donne déjà une première information sur le comportement d'une variable aléatoire. La variance nous informe sur la distribution des valeurs de la variable autour de sa valeur moyenne. La loi des grands nombres vient préciser ces informations, en terme probabilistes.

Théorème 6.0.1 Inégalité de Markov *Soit $Z : \Omega \to \mathbb{R}$ une variable aléatoire non négative. Alors :*

$$P(Z \geq a) \leq \frac{\mathbb{E}(Z)}{a} \tag{6.1}$$

Inégalité de Chebyshev *En particulier, si $X : \Omega \to \mathbb{R}$ est une variable aléatoire alors pour tout $\epsilon > 0$, on a :*

$$P(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2} \tag{6.2}$$

La preuve est relativement simple. On considère la variable aléatoire indicatrice $1_{Z \geq a}$. Cette variable vaut $1_{Z \geq a}(\omega) = 1$ exactement lorsque $Z(\omega) \geq a$. Par conséquent, $\forall \omega \in \Omega, Z(\omega) \geq a 1_{Z \geq a}$. Par suite, puisque l'espérance est une fonction monotone, $\mathbb{E}(Z) \geq a\mathbb{E}(1_{Z \geq a}) = aP(Z \geq a)$.

Pour obtenir la seconde identité, on applique (6.1) à la variable aléatoire $|X - \mathbb{E}(X)|^2$ et $a = \epsilon^2$. □

La loi des grands nombres établit un lien entre probabilité et fréquence empirique. Dans le jeu de pile ou face, avec probabilité $1/2$ d'avoir pile ou face, la loi nous assure que :

$$\lim_{n \to \infty} P(\frac{S_n}{n} - \frac{1}{2}) = 1$$

le nombre de fois où l'on observe "pile" (la valeur de la variable S_n) tend vers $1/2$ si l'on joue un grand nombre de fois. On le déduit de l'identité de Chebyshev. En effet, soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes

et identiquement distribuées. Considérons la variables aléatoire $(X_1+\cdots+X_n)/n$, dont l'espérance est égale à $\mathbb{E}(X_1)$ (pourquoi ?) et la variance est égale à $nV(X_1)$ (pourquoi ?). L'identité de Chebyshev nous donne :

$$P(\big|\frac{X_1+\cdots+X_n}{n}-\mathbb{E}(X_1)\big|\geq\epsilon)\leq\frac{\sigma_{X_1}^2}{n\epsilon^2} \tag{6.3}$$

puisque $\mathbb{E}(\frac{X_1+\cdots+X_n}{n})=\mathbb{E}(X_1)$ et $\mathbb{V}(\frac{X_1+\cdots+X_n}{n})=\frac{\mathbb{V}(X_1)}{n}$ en vertu des identités (5.3) et (5.5) et de l'exercice 5.26. Par suite :

$$\lim_{n\to\infty}P(\big|\frac{X_1+\cdots+X_n}{n}-\mathbb{E}(X_1)\big|\geq\epsilon)=0. \tag{6.4}$$

C'est en ce sens que la moyenne empirique tend vers la moyenne probabiliste. Cette identité est la *loi faible des grands nombres.*

La *loi forte des grands nombres* affirme que :

$$\lim_{n\to\infty}\frac{X_1+\cdots+X_n}{n}=\mathbb{E}(X_1) \tag{6.5}$$

selon laquelle la variable aléatoire du membre gauche tend vers la variable constante de valeur $\mathbb{E}(X_1)$ – encore faudrait-il préciser ici ce que nous entendons par là puisqu'il s'agit de convergence en probabilité, mais nous ne nous étendrons pas sur ce résultat.

Exercice 6.1 *On modélise le nombre de téléchargements d'un fichier sur une heure par une variable aléatoire d'espérance 50.*

- *Estimer la probabilité que le nombre de téléchargements dépasse 75 ?*
- *On sait de plus que la variance du nombre de téléchargements est de 25. Estimer la probabilité que le nombre de téléchargements sur une heure soit compris entre 40 et 60 ?*

Exercice 6.2 *Dans une population de 30 000 000 individus, la proportion d'individus présentant le caractère C est $p=0.4$. On prélève un échantillon de taille $N=1600$ et on note X le nombre d'individus de l'échantillon présentant le caractère C.*

- *Minorer la probabilité des événements :*

$$0.30\leq\frac{X}{N}\leq 0.50\quad 0.35\leq\frac{X}{N}\leq 0.45\quad 0.38\leq\frac{X}{N}\leq 0.42$$

- *Calculer la longueur minimale L de la fourchette telle que :*

$$P(0.40-\frac{L}{2}\leq\frac{X}{N}\leq 0.40+\frac{L}{2})\geq 0.95$$

Exercice 6.3 *On effectue des tirages avec remise dans une urne contenant deux boules blanches et quatre boules bleues. Quelqu'un affirme que, bien évidemment, vous aurez tiré une boule blanche une fois sur trois. Combien de tirages faut-il effectuer pour que cette affirmation soit correcte : que*

l'on s'écarte de la valeur prédite (1/3) d'au plus 0.2 avec probabilité au plus 1/100 ?

Exercice 6.4 *Un fournisseur d'accès à Internet met en place un point local d'accès, qui dessert 5000 abonnés. A instant donné, chaque abonné a une probabilité égale à 20% d'être connecté. Les comportements des abonnés sont supposés indépendants les uns des autres.*

- *On note X la variable aléatoire égale au nombre d'abonnés connectés à un instant t. Quelle est la loi de X ? Quelle est son espérance, son écart-type ?*
- *Le fournisseur d'accès souhaite savoir combien de connexions simultanées le point d'accès doit pouvoir gérer pour que sa probabilité d'être saturé soit inférieure à 2,5%.*

Exercice 6.5 *On lance une pièce équilibrée. En utilisant l'inégalité de Chebychev (6.2), estimer le nombre de lancers nécessaires pour que la fréquence de Pile observé au jeu de Pile ou Face soit comprise entre 0.4 et 0.6 avec une probabilité au moins égale à 0.9.*

6.1 Simulation et méthodes Monte Carlo

Les méthodes de Monte Carlo sont une alternative souvent efficace aux méthodes numériques (pour calculer ou approximer des fonctions). Elles sont aussi conceptuellement faciles à appréhender et à implémenter. Prenons un exemple pour illustrer notre propos.

Considérez la figure 3.4 et imaginons que l'on souhaite calculer l'aire de l'ensemble A. La difficulté vient de ce que le contour de l'ensemble peut difficilement être décrit par une fonction analytique (un cercle, une ellipse, etc.). Du point de vue informatique, on peut imaginer que la courbe est donnée sous la forme d'une ligne brisée – cela pourrait nous permettre de réduire le problème par un calcul, fastidieux, d'aires de petits triangles. On peut procéder différemment.

Il est facile de calculer l'aire du rectangle englobant Ω. On peut ensuite raisonner de manière probabiliste : si on choisit "au hasard" un point (x, y) dans le rectangle, la probabilité qu'il appartienne à l'ensemble A est donnée par le ratio des aires (aire A / aire Ω). On peut donc procéder comme suit. Convenons que le coin inférieur gauche du rectangle est à l'origine, et que ses dimensions sont $a \times b$ (base $\times$ hauteur).

- On effectue un tirage aléatoire uniforme de points dans $[0, a] \times [0, b]$;
- On teste l'appartenance des points à l'ensemble A ;
- On calcule le ratio (nombre de points dans A) / nombre total de points.

Les tirages aléatoires des coordonnées (x, y) se font de manière indépendantes et selon des lois uniforme sur chaque intervalle. Plus le nombre

de tirage est grand, plus l'évaluation de l'aire sera précise. La précision de cette approximation dépend aussi de la précision du contour de l'ensemble A.

Formalisons l'approche que nous venons d'introduire. La méthode de Monte Carlo revient en réalité à évaluer l'espérance mathématique d'une variable aléatoire. Dans le cas qui nous préoccupe, la variable aléatoire X (associée à l'ensemble A) est défini sur l'ensemble des points dans $[0,a] \times [0,b]$ et vaut 1 ou 0 selon qu'un point (x,y) est ou non dans l'ensemble A. On peut alors considérer des variables aléatoires $X_1, \ldots, X_N$ (les tirages aléatoires des n points dans le rectangle) et de considérer l'approximation :

$$E(X) = \frac{1}{N}(X_1 + \cdots + X_N).$$

Puisque l'on approxime une quantité à l'aide de variables aléatoires, la loi des grands nombres nous permet de maîtriser l'erreur commise. Ainsi, on voit que l'erreur dépend fortement de la variance de la loi de probabilités des variables utilisées. On peut donc avoir intérêt à utiliser une loi de variance la plus petite possible. On verra plus loin comment estimer l'erreur commise à l'aide d'une loi gaussienne (section **??**).

Chapitre 7
Probabilités et simulation, génération aléatoire

Vous utiliserez pour ces exercices l'environnement ou le langage de programmation et la librairie de votre choix. Des exemples typiques sont :

- Java et la librairie/classe java.util.Random ou encore java.math (méthode random()) ;
- le langage C++ comprend aussi les librairies nécessaires à la simulation de loi de probabilités ;
- un tableur (Excel/OpenOffice) permet aussi de faire certaines choses [BMPS98] [BPSD07].

Exercice 7.1 Loi uniforme sur $[0,1]$ **– Tirage aléatoire d'entiers** *Ecrivez un court programme effectuant un tirage aléatoire uniforme dans l'intervalle réel* $[0,1]$.

Effectuez un tirage à plusieurs reprises. La suite calculée est-elle invariablement la même ? Y a t-il moyen d'obtenir la même suite lors de deux exécutions de ce même programme (pensez à positionné le germe – seed, en anglais – de la suite).

Exercice 7.2 Pile ou face *Utilisez le tirage uniforme dans* $[0,1]$ *pour simuler un tirage de Bernouilli avec probabilité* $0 < p < 1$.

Exercice 7.3 Rademacher *Utilisez le tirage uniforme dans* $[0,1]$ *pour simuler une loi de Rademacher, qui vaut 1 avec probabilité* p *et -1 avec probabilité* $1-p$.

Exercice 7.4 Tirage aléatoire d'entiers *Ecrivez un court programme qui tire au hasard des entiers dans un intervalle* $[a,b]$. *Assurez-vous du caractère uniforme de la loi sous-jacente.*

Relancez votre programme à plusieurs reprises. La suite calculée est-elle invariablement la même ? Y a t-il moyen d'obtenir la même suite lors de deux exécutions de ce même programme (pensez à positionné le germe – seed, en anglais – de la suite).

Exercice 7.5 Bandit manchot *Ecrivez un programme qui simule le jeu du bandit manchot décrit au début de la section 5. Observez le gain du jouer converger vers l'espérance mathématique prédit par la théorie.*

Exercice 7.6 Nombres et chapeaux *Ecrivez un programme qui simule le jeu du casino où le croupier dissimule deux nombres sous des chapeaux (exercice 3.17).*

7.1 Génération aléatoire

Nous avons vu à la section précédente comment simuler une loi de probabilité. Les besoins de la simulation peuvent parfois être un peu différent et exiger de générer aléatoirement une structure de données. On peut par exemple avoir à générer aléatoirement un ensemble de k éléments parmi N, ou une permutation de N éléments.

Imaginez que vous ayez à mettre au point une simulation pour tester et valider un système de gestion du trafic urbain, tenter de valider un scénario d'équilibrage des charges, etc. Vous allez vouloir générer un afflux du trafic en certains points du réseau au départ de la simulation. En d'autres mots, vous allez par exemple vouloir sélectionner, au hasard, k points du réseau parmi N. Ou encore, vous souhaiterez générer des entiers n_1, n_2, ..., n_k dont la somme fait N. Ce problème n'est pas anodin. En mettant une procédure ad hoc pour construire une telle structure, on est pas assuré de pouvoir générer toutes les structures possibles, et même si c'est le cas, on n'est pas assuré de le faire de manière équiprobable.

Exercice 7.7 Générer des entiers *En machine, les entiers sont décrits par des vecteurs de bits de taille N (= 32 ou 64). On peut donc penser générer aléatoirement des entiers en jouant à pile ou face N fois et en prenant le résultat du jeu comme résultat. Une suite générée ainsi forme t-elle une suite aléatoire d'entiers ?*

Ecrivez un court programme qui tire au hasard un nombre entiers de l'ensemble $\{0, 2^N - 1\}$ de cette façon ($N = 32$ ou 64 selon l'architecture de la machine utilisée).

Exercice 7.8 Générer des ensembles *Etant donné une liste ordonnée de N éléments distincts e_1, ..., e_N on peut coder un sous-ensemble F de $E = \{e_1, \ldots, e_N\}$ par un vecteur $(b_1, \ldots, b_N)$ de N bits 0 ou 1 où $b_i = 1$ si et seulement si l'élément $e_i \in F$. Proposez une procédure pour générer aléatoirement de manière équiprobable un sous-ensemble. Implémentez-là.*

Exercice 7.9 Graphe aléatoire *Un* graphe simple *G est donné par un ensemble de* sommets *V et d'*arêtes *E où $E \subset V \times V$. Une arête $e \in E$ est une paire de sommets $\{u, v\}$. Erdös et Renyi ont proposé un modèle de graphe*

aléatoire qui consiste à effectuer pour chaque paire $\{u, v\}$ un tirage de Bernouilli avec probabilité p.

Soit Ω l'ensemble de tous les graphes simples sur un ensemble V et considérons la variable aléatoire $X : \Omega \to \mathbb{N}$ donnant le nombre d'arêtes d'un graphe G.

- *Quelle est la distribution de la variable X.*
- *Quel est le nombre moyen d'arêtes d'un graphe aléatoire selon le modèle Erdös-Renyi ?*
- *Donnez une implémentation de la procédure de tirage aléatoire d'un graphe.*

Exercice 7.10 Générer un arbre étiqueté *Cet exercice se penche sur la génération aléatoire uniforme d'un arbre étiquetés. Un tel arbre $T = (V, E =$ est un graphe connexe non-orienté, sans cycle, sur l'ensemble $V = \{1, \ldots, N\}$. Ces structures sont très présentes en informatiques et dans des problèmes divers de calculs de flots, de couverture de graphes, etc.*

Travailler directement sur la structure de graphe pour tenter de générer une telle structure n'est pas simple. En effet, comment déterminer les arêtes $e \in E$ à ajouter ? comment s'assurer que l'ajout d'une arête ne provoque l'apparition de cycle , comment s'assurer de générer un graphe connexe ?

La solution passe par un codage de la structure en un objet plus simple, au moins pour ce qui concerne la génération aléatoire. On peut coder un arbre par une séquence de Prüfer ; c'est une séquence $S = (s_0, \ldots, s_{N-2})$ de $N - 2$ entiers choisis parmi $\{1, \ldots, N\}$. On l'obtient d'un arbre étiqueté en appliquant la méthode suivante :

- *(à répéter tant qu'il reste plus de deux sommets dans l'arbre courant T)*
- *identifier la feuille v de l'arbre courant ayant le numéro minimum ;*
- *ajouter à la suite S le seul sommet s* adjacent *à v dans l'arbre T courant ;*
- *enlever de l'arbre T courant le sommet v et l'arête* incidente *à v.*

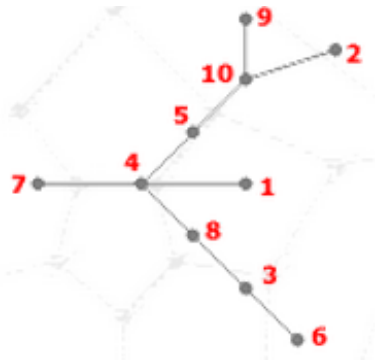


Fig. 7.1 Un arbre étiqueté.

Simulez l'algorithme de codage d'un arbre à partir de l'arbre de la figure 7.1.

A l'inverse, on peut construire l'arbre correspondant à une séquence de Prüfer en suivant le procédé suivant :

- *S est donnée, on initialise I à $\{1, \ldots, N\}$.*
- *A répéter tant qu'il reste des éléments dans S et plus de deux éléments dans I*
- *identifier le plus petit élément i de I n'apparaissant pas dans la suite S ;*
- *relier par une arête de T le sommet i avec le sommet s correspondant au premier élément de la suite S ;*
- *enlever i de I et s de S.*
- *Les deux éléments qui restent dans I à la fin de l'algorithme constituent les extrémités de la dernière arête à ajouter à T.*

Calculez l'arbre associée à la séquence de Prüfer $S = 4, 10, 3, 8, 4, 4, 5, 10$.

On transforme donc ainsi le problème de la génération aléatoire d'un arbre étiqueté en celui de la génération aléatoire d'une séquence de $N - 2$ entiers indépendants. Ecrivez et implémentez maintenant un algorithme qui génère aléatoirement et de manière uniforme un arbre étiquetés sur $\{1, \ldots, N\}$.

Exercice 7.11 Générer aléatoirement une permutation *Une permutation est une liste de N entiers distincts choisis parmi $\{1, \ldots, N\}$. En d'autres mots, c'est une liste non ordonnée de ces entiers, "dans n'importe quel ordre". Générer aléatoirement et uniformément une telle liste peut être utile pour tester un algorithme. Pensez aux algorithmes de tri : en faisant tourner l'algorithme sur un bon nombre de permutations choisies au hasard, on peut évaluer empiriquement ses performances. Pensez à un programme qui doit traiter des tâches dans un ordre qui n'est pas déterminé au départ et qui peut varier aléatoirement : il faut bien pouvoir simuler ces ordres d'arrivée des tâches pour le tester.*

Comment donc générer une permutation en s'assurant que chacune ait la même probabilité d'apparaître ? On pourrait penser à une méthode naïve : on choisit au hasard N entiers dans $\{1, \ldots, N\}$, en s'assurant que chacun n'apparait qu'une seule fois (si on tire un entier déjà dans la séquence, on le jette et on recommence). Mais est-on bien certain dans ces conditions d'avoir un tirage uniforme ?

Essayez. Implémentez cet algorithme naïf et observez la distribution empirique des permutations (pour de petites valeurs de N).

Le tirage est bien uniforme, essentiellement parce que les tirages des entiers sont indépendants. L'inconvénient de ce procédé est sa complexité. On peut en effet être amener à rejeter souvent des tirages d'entiers. En effet, la probabilité de tirer à la $i$ème position un nombre déjà présent dans la liste est i/N, et croit donc à mesure que l'on construit la permutation. Sauriez-vous montrez que la complexité de cet algorithme est de l'ordre de $N \log N$?

Exercice 7.12 Permutations et tables d'inversion *Nous allons maintenant voir un autre algorithme, plus efficace, pour générer une permutation.*

Une permutation peut être codée par une table d'inversion *(on parle aussi de son code de Lehmer), qui décrit la position relative des entiers dans la liste. La table associée à la permutation identité* $1, 2, \ldots, N$ *est* $(0, 0, \ldots, 0)$ *puisque tout est bien en ordre. Etant donnée une permutation* $\sigma = \sigma_1\sigma_2\ldots\sigma_N$*, le nombre d'inversions* b_i *de* σ_i *est le nombre d'entiers* σ_j *qui se trouve à sa gauche (*$j < i$*) et qui sont plus grands (*$\sigma_j > \sigma_i$*).*

Donnez la table d'inversion de la permutation $N, \ldots, 2, 1$*. Donnez la table d'inversion de* $1, 3, 5, \ldots, N-1, 2, 4, \ldots, N$ *(pour* N *pair par exemple).*

Une suite $b = b_1, b_2, \ldots, b_{N-1}$ *d'entiers* $0 \leq b_i \leq N - i$ *détermine une unique permutation. On commence par considérer la séquence réduite à l'entier* N *et on ajoute à la suite les entiers* $N-1, N-2, \ldots, 2, 1$ *en plaçant* i *à la position* b_i *(la position à l'extrême gauche est la position 0).*

Donnez la permutation ($N = 5$*) associée à la table d'inversion* $3, 1, 1, 1$*. Que peut-on dire des permutations dont la table d'inversion commence par* $N - 1$ *?*

La génération aléatoire uniforme sur $\{1, \ldots, N\}$ *d'une permutation peut donc se réduire à une série de* $N-1$ *tirages aléatoires indépendants d'entiers (chacun sur un intervalle différents).*

Exercice 7.13 Partitions d'ensemble *Une partition de l'ensemble* $V = \{1, 2, \ldots, N\}$ *est un découpage en sous-ensemble* $V_1, V_2, \ldots, V_k \subset V$ *disjoints – c'est-à-dire que* $V_i \cap V_j = \emptyset$ *pout toute paire d'indices* $i \neq j$ *– et tels que* $V_1 \cup V_2 \cup \cdots \cup V_k = V$*. La partition est souvent notée* $\pi = \{V_1, ^l dots, V_k\}$ *et* k *est son nombre de* blocs.

Le nombre de partitions en k *d'un ensemble à* N *éléments est égal au* nombre de Stirling $S(N, k)$*. La somme* $\sum_{k=1}^{N} S(N, k) = B_N$ *est égal au* nombre de Bell*. On peut calculer les nombre de Stirling à l'aide la récurrence :*

$$S(N, k) = S(N-1, k-1) + kS(N-1, k) \tag{7.1}$$

soumis aux conditions initiales $S(N, 1) = S(N, N) = 1$*. La récurrence s'interprète facilement : soit l'élément* N *est seul dans un bloc. Sinon, suffit de construire une partition sur* $\{1, \ldots, N-1\}$ *et de choisir au hasard un bloc (parmi les* k *blocs) dans lequel on l'insère. Les nombres* $S(N, k)/B_N$ *nous donnes donc la distribution de probabilité de la variable rendant compte du nombre de blocs d'une partition.*

On peut générer une partition en reprenant le schéma suggéré par la récurrence (7.1). Cela dit, il faut avoir préalablement choisi le nombre k *de blocs qu'aura cette partition. Cela exige donc de choisir au hasard le nombre* k*,* mais selon la distribution des valeurs $S(N, k)$*.*

Précisez cet algorithme et implémentez-le.

Chapitre 8
Variable aléatoire réelle, densité de probabilité

Nous avons jusqu'à maintenant vu des exemples et résultats concernant les variables aléatoires discrètes, c'est-à-dire qui prennent des valeurs entières dans $\mathbb{N}$ ou $\mathbb{Z}$. Il est toutefois possible, et parfois nécessaire, d'étendre la notion de probabilité au cas des variables prenant une valeur dans $\mathbb{R}$. Imaginons qu'il s'agissent d'étudier la variation de température d'un processus, ou la variation de prix d'une action sur un marché, etc. Le domaine des valeurs ne peut alors être restreint à un ensemble fini ou aux entiers naturels.

On se trouve alors face à une difficulté, puisque la probabilité d'observer une valeur particulière est, en réalité, nulle. En effet, quelle est la probabilité d'observer la valeur d'une action à un prix donné ? d'observer une température précise ? Il est plus naturel de chercher à connaître la probabilité que la valeur observée se trouve dans un intervalle de valeurs. Cela nous amène à considérer la notion de *densité de probabilité*.

8.0.0.4 densité de probabilité

Soit $(\Omega, \mathcal{U}, P)$ un espace de probabilité continu et X une variable aléatoire sur $\mathbb{R}$. La probabilité des évènements sur Ω exige alors nombre de précautions que nous passerons sous silence. Soulignons toutefois que la mesure de probabilité se calcule à l'aide d'une *intégrale*, en exigeant d'abord l'existence d'une *densité de probabilité*, c'est-à-dire une fonction f_X définie sur $\mathbb{R}$ et telle que :

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1.$$

On peut voir cette équation comme le pendant continu de la condition exigeant que la somme des probabilités des épreuves $\omega \in \Omega$ soit égales à 1.

La densité f_X nous permet alors de calculer les probabilités associées à la variable X :

$$P(X \leq b) = \int_{-\infty}^{b} f_X(x) \tag{8.1}$$

Cette définition permet de calculer, par exemple, la probabilité $P(a \leq X \leq b) = \int_a^b f_X(x)$ ou encore $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$. La notion d'espérance mathématique doit, elle aussi, être revue :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx.$$

8.0.0.5 Fonction de répartition

La fonction de répartition associée à une densité de probabilité f_X est la fonction F_X donnée par $F_X(a) = P(X \leq b) = \int_{-\infty}^{b} f_X(x)$.

La fonction de densité et la fonction de répartition sont évidemment liées par la relation entre $F'_X = f_X$, c'est-à-dire qu'on obtient f_X en dérivant F_X, puisqu'on intègre f_X pour obtenir F_X.

8.1 Lois continues classiques

8.1.0.6 Loi de distribution uniforme

La *loi de distribution uniforme* sur l'intervalle $[a,b)$ est une variable aléatoire réelle X de densité de probabilité $f_X(x) = \frac{1}{b-a} 1_{[a,b)}$ où $1_{[a,b)}$ est la fonction indicatrice associé à l'intervalle $[a,b)$. On dit alors que X suit une loi uniforme sur l'intervalle $[a,b)$.

En réalité, on connaît déjà cette loi. C'est celle qui sous-tend l'appel à la fonction random() qui retourne un nombre réel aléatoire de l'intervalle $[0,1)$. Cette fonction fait ce qu'on attend d'elle : elle choisit au hasard un nombre réel parmi tous les nombres réels de l'intervalle [0, 1). Cet énoncé n'est pas anodin n et pourrait soulever quelques interrogations : comment peut-on faire un tirage aléatoire d'un nombre réel en machine alors qu'on ne dispose que d'un nombre fini de représentations en machine ? La question n'est en effet pas simple, et nous la passerons sous silence. Nous considérerons que cette loi uniforme est une brique de base pour toutes nos constructions informatiques (et mathématiques).

Notez au passage que la variable aléatoire $X_{a,b}$ uniforme sur $[a,b)$ se déduit de la variable $X_{0,1}$ uniforme sur $[0,1)$ par multiplication $X_{a,b} =$

$bX_{0,1} - a$ (c'est bien ainsi que l'on peut procéder pour implémenter la variable aléatoire uniforme sur $[a, b)$).

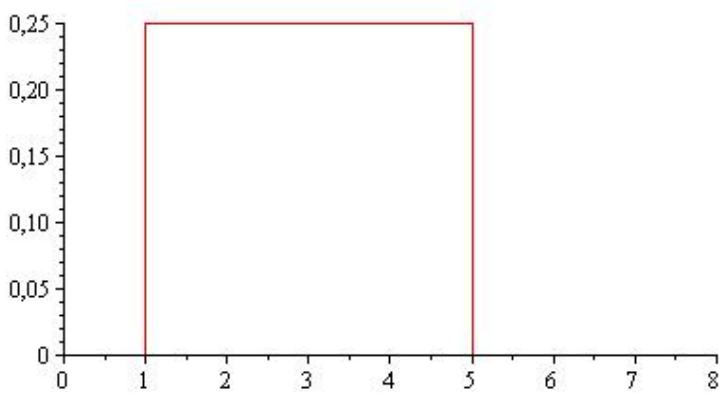


Fig. 8.1 La loi uniforme correspond à une distribution constante sur tout l'intervalle $[a, b)$ (avec ici $a = 1, b = 5$), l'aire sous la courbe étant égale à 1.

Exercice 8.1 *Vérifiez que $f_X(x) = \frac{1}{b-a} 1_{[a,b)}$ est bien une densité. Calculez sa fonction de répartition (qui donne $P(X \leq x)$). Calculez l'espérance mathématique et la variance de la loi uniforme.*

Exercice 8.2 *Nous avons vu à l'exercice 3.9 comment simuler la loi de distribution uniforme sur l'intervalle $[a, b)$.*

Remarque 8.1.1 *Nous avons eu affaire à des cas de variables aléatoires continues non-uniformes sans le souligner explicitement. Considérons en effet le cas d'une variable aléatoire discrète X à valeurs dans un ensemble fini V, et notons $p_v = P(X = v)$. La variables X est alors complètement définie par la suite des probabilités $(p_v)_{v \in V}$. Pour simuler la variable X il nous faut choisir un état $v \in V$ en respectant les probabilités d'obtenir chacun des états $v \in V$.*

On le fait en tirant aléatoirement un nombre réel $p \in [0, 1)$ et en déterminant à partir de ce p l'un des états comme suit : on découpe l'intervalle $[0, 1)$ en $|V|$ intervalles chacun de longueur p_v. Le nombre p tombe alors dans un unique intervalle et détermine ainsi l'état $v \in V$ (la probabilité de tomber en plein sur une extrémité d'un intervalle est nulle).

Méthodes à rejet Supposons que l'on sache simuler une loi uniforme sur un ensemble A : on sait choisir aléatoirement un élément de A de manière équiprobable à l'aide d'un algorithme $\mathcal{A}$. Alors on peut utiliser cet

algorithme pour simuler une loi uniforme sur un sous-ensemble $B \subset A$. Il suffit de générer des éléments de A jusqu'à tomber sur un élément de B.

Cette méthode simule bien une variable aléatoire X de loi uniforme sur l'ensemble B. Il nous faut montrer que la probabilité de tirer aléatoirement un élément $b \in B$ est égale à $1/|B|$. Or, on a probabilité $1/|A|$ de tomber sur $b \in B$ dès le premier tirage, puis probabilité $(1-|B|/|A|)1/|A|$ d'y arriver au second tirage, etc. on a probabilité de réussir dès le premier tirage, puis probabilité $(|A|-|B|)/|A| \cdot |B|/|A|$ de réussir au second tirage, etc. Ainsi, $P(X=b) = \sum_{k\geq 1}(1-|B|/|A|)^{k-1}1/|A| = 1/|B|$ montrant ainsi que le tirage aléatoire est bien uniforme sur B.

Exercice 8.3 *Ecrivez un court programme qui permet de générer aléatoirement un points dans un disque de rayon R donné.*

Remarque 8.1.2 *Nous avons déjà implicitement utilisée soulignée l'usage de cette méthode lors du tirage aléatoire de permutations (exercice 7.11, page 54). La méthode "naïve" que nous présentions tire au hasard un entier d'un sous-ensemble $A \subset \{1,\dots,N\}$ en utilisant une méthode à rejet.*

Cette méthode se généralise au cas des lois continues. Simuler une variable aléatoire X de densité f_X revient à tirer un point au hasard sous le graphe de f_X et retourner l'abcisse de ce point. On peut donc, au passage, avoir à utiliser la méthode à rejet pour tirer au hasard (uniformément) un point sous le graphe de f_X.

Répartition inverse La dernière remarque n'est pas étrangère à un procédé général. Soit une variable aléatoire X admettant une fonction de densité de probabilité f_X. On définit la fonction de répartition (associée à X) en posant $F(x) = \int_{-\infty}^{x} f_X(t)dt$. On a donc $F(x) = P(X \leq x)$.

On pose $F^{-1}(t) = \inf\{x, F(x) \geq t\}$. (Il nous fait prendre quelques précautions avec F^{-1} puisque F n'est pas nécessairement injective ; elle est toutefois monotone.

Proposition 1. *Si U est une variable aléatoire de loi uniforme sur l'intervalle $[0,1)$, alors la variable $F^{-1}(U)$ suit la même loi que X.*

En d'autres mots, il suffit de pouvoir simuler la loi uniforme sur $[0,1)$ pour pouvoir simuler n'importe quelle autre variable aléatoire, à condition de pouvoir calculer l'inverse F^{-1} de la fonction de répartition. Remarquez que c'est exactement ce qui est proposé à la remarque 8.1.1, où la fonction de répartition est alors une fonction en escalier.