LEPL1108 Mathématiques discrètes et probabilité

Simon Desmidt

Année académique 2022-2023 - Q1

UCLouvain

Table des matières

- Remarque : ce cours demande souvent, pour les démonstrations, de passer par l'absurde.

Graphes

1.1 Définitions

- Un graphe est un triple :
  - N: un ensemble fini de nœuds
  - R : un ensemble fini d'arêtes
  - I : une relation d'incidence $\subset N \times R$. Cette relation est généralement sous-entendue.

On écrit : $I = \{(i_1, \alpha_1), (i_1, \alpha_2), \ldots\}$ ou $i_1 I \alpha_1$ $(// (i_1, \alpha_1) \in I)$

- $\alpha \in R$ est une boucle si ses deux extrémités sont un même nœud.
- $|N|$ est l'ordre du graphe, c'est le nombre de nœuds.
- Le degré d'un nœud n, $deg(n)$ est le nombres d'arêtes incidentes au nœud n, les boucles comptant double.
- Un graphe est simple si il n'a ni boucle, ni nœuds reliés par des arêtes multiples.

$\alpha = \{i, j\}$ identifie alors l'unique arête telle que $(i, \alpha) \in I$ et $(j, \alpha) \in I$.

- Un parcours P de longueur k dans G est une suite alternée $P \coloneqq (i_0, \alpha_1, i_1, \ldots, \alpha_k, i_k)$ de $k+1$ nœuds et k arêtes de G.
- Un parcours est fermé[1] si $i_0 = i_k$, et ouvert sinon.
- Un cycle est un parcours dont tous les nœuds et arêtes sont distincts, à l'exception du premier et du dernier nœud.
- Une piste est un parcours dont les arêtes sont distinctes.
- Un circuit est une piste fermée.
- Un circuit/piste eulérien(ne) sur un graphe sans nœud isolé est un circuit/piste qui passe par toutes les arêtes de ce graphe.
- Un graphe simple $G = (N, R)$ est appelé graphe biparti s'il existe une partition de N en deux classes N_0, N_1 telles que si $\{i, j\} \in R$, alors $i \in N_0$ et $j \in N_1$.

[1] Un graphe fermé n'a jamais de nœud de degré impair.

- Un graphe simple G d'ordre ≥ 2 est biparti ssi G ne possède aucun cycle de longueur impaire.
- Un graphe est dit connexe s'il existe un parcours reliant toute paire de nœuds.
- Si le graphe $G = (N, R)$ n'est pas connexe, il peut être formé de composantes connexes. C'est un ensemble de graphes $\{G_1 = (N_1, R_1), \ldots, G_n = (N_n, R_n)\}$ tel que :
  - $\{N_1, \ldots, N_n\}$ forme une partition de N.
  - R_i contient toutes les arêtes incidentes à des nœuds de N_i.
  - Les ensembles R_i sont disjoints.
  - Chaque graphe G_i est connexe.

$\rightarrow$ Remarque : il n'existe pas de graphe connexe contenant un nombre impair de nœuds de degré impair.

- Un chemin est un parcours ouvert dont tous les nœuds sont distincts.
- Un graphe simple possède un chemin hamiltonien s'il possède un parcours passant par chacun de ses nœuds 1! fois.
- Un graphe simple possède un cycle hamiltonien si il possède un cycle passant par chacun de ses nœuds.
- Un graphe hamiltonien est un graphe simple possédant un cycle hamiltonien.
- Un graphe orienté est une paire de :
  - N un ensemble fini de nœuds $(i, j, \ldots)$.
  - $R \subseteq (N \times N)$ un ensemble d'arêtes (ayant une direction).
- Un graphe orienté pondéré est un graphe orienté auquel on adjoint une fonction de coût $c : R \to \mathbb{R}$.
- Les graphes simples $G = (N, R)$ et $G' = (N', R')$ sont isomorphes si :
  - Il existe une bijection $f : N \to N'$
  - $\{i, j\} \in R \iff \{f(i), f(j)\} \in R'$

$\rightarrow$ Remarque : la relation d'isomorphisme est une relation d'équivalence sur les graphes. il est facile de vérifier si f définit un isomorphisme, mais il reste difficile de décider si deux graphes sont isomorphes.

1.2 Théorème des pistes eulériennes

Le graphe G sans nœud isolé possède une piste eulérienne ssi il est connexe et contient au maximum deux nœuds de degré impair.[2]

[2] Voir section 4.1.

1.3 Algorithme de Dijkstra – CM 1, slides 22-28

- In : $G = (N, R)$ un graphe orienté pondéré $^{\geq 0}$ et $s \in N$.
- Out : le coût du plus court chemin de s vers tout $n \in N$.
- Init : soit $N_s \coloneqq \{s\}, \overline{N}_s \coloneqq N - N_s,\ d(i) \coloneqq c(s,i)\ \forall i \in N$.
- Loop : tant que $\overline{N}_s \neq \emptyset$:
  - $v \coloneqq$ élément de $\overline{N}_s$ qui minimise $d(v)$
  - $N_s \coloneqq N_s \cup \{v\}$ et $\overline{N}_s = N - N_s$
  - $\forall i \in \overline{N}_s\ : d(i) \coloneqq \min(d(i),\ d(v) + c(v,i))$
- Return : d fournit les coûts recherchés pour chaque nœud n.

1.4 Algorithme de Bellman-Ford

→ Remarque : la différence entre Dijkstra et BF est la condition ≥ 0 sur la pondération des arêtes.

- In : $G = (N, R)$ un graphe orienté pondéré sans cycle négatif et $s \in N$.
- Out : le coût du plus court chemin de s vers tout $n \in N$.
- Init : $d(s) \coloneqq 0,\ d(i) \coloneqq \infty\ \forall i \in N - \{s\}$.
- Loop, répétée $|N| - 1$ fois :
  - $\forall (i,j) \in R$: Si $d(i) + c((i,j)) < d(j)$, alors $d(j) \coloneqq d(i) + c((i,j))$.
- Return : d fournit les coûts recherchés.

1.5 Arbres

Un arbre est un graphe connexe qui ne possède pas de cycle. Les propriétés suivantes sont équivalentes :

- G est connexe et sans cycle.
- G est connexe et $|R| = |N| - 1$.
- G est sans cycle et $|R| = |N| - 1$.
- G est sans cycle et lui ajouter une arête crée un et un seul cycle.
- G est connexe et supprimer une arête quelconque le déconnecte.
- Deux nœuds distincts de G sont reliés par un et un seul chemin (et G est sans boucles).

1.5.1 Arbres sous-tendants

L'arbre $G' = (N', R')$ est un arbre sous-tendant de $G = (N, R)$ si $N = N'$ et $R' \subseteq R$.

G est connexe ssi G possède un arbre sous-tendant.

A est un arbre sous-tendant de poids minimum de G ssi A sous-tend G et tout arbre A' sous-tendant G est tel que $c(A) \leq c(A')$.

1.5.2 Algorithme de Kruskal

Pour un graphe $G = (N, R)$ connexe.

- Init : $R_{\text{ord}} \coloneqq \text{trier}(R)$, $R' \coloneqq \emptyset$
- Loop : tant que $|R'| < |N| - 1$:
  - $(\alpha, R_{\text{ord}}) \coloneqq R_{\text{ord}}$
  - Si $(N, R' \cup \{\alpha\})$ est sans cycle, alors $R' \coloneqq R' \cup \{\alpha\}$
- Return : (N, R') est un arbre sous-tendant G de poids minimum.

== On met toutes les arêtes avec un poids le moins élevé, sauf quand ça crée un cycle, jusqu'à avoir un arbre.

1.6 Représentation matricielle

1.6.1 Matrice d'adjacence

La matrice adjacente A est de genre $|N| \times |N|$:

- $a_{i,j} \coloneqq$ nombre d'arêtes reliant i et j si $i \neq j$
- $a_{i,i} \coloneqq$ deux fois le nombre de boucles sur i

1.6.2 Matrice d'incidence

La matrice d'incidence M d'un graphe $G = (N, R)$ est la matrice dont l'entrée $m_{i,\alpha}$ est telle que :

- $m_{i,\alpha} \coloneqq 2$ si α est une boucle sur i.
- $m_{i,\alpha} \coloneqq 1$ si α relie i à un autre nœud.
- $M_{i,\alpha} \coloneqq 0$ si α n'est pas incidente au nœud i.

Dans la matrice d'incidence d'un graphe, la somme des entrées d'une colonne vaut 2 et la somme des entrées d'une ligne vaut le degré du nœud correspondant.

1.6.3 Propriétés

- $\sum_{j \in N} a_{i,j} = \sum_{j \in N} a_{j,i} = \deg(i)$
- $\sum_{(i,j) \in N^2} a_{i,j} = 2\,|R|$
- $\sum_{i \in N} \deg(i) = 2\,|R|$
- Le nombre de nœuds de degré impair d'un graphe est pair.
- Si G est un graphe simple, le nombre de parcours de longueur k entre ses nœuds i et j est donné par $\left(A^k\right)_{i,j}$.

Structures algébriques

2.1 Définitions

- Une structure algébrique est une paire :
  - A, un ensemble
  - $\circ$, un opérateur sur A, i.e. une fonction $\circ\ : A \times A \to A$
- Représentation par table de Cayley[a] :

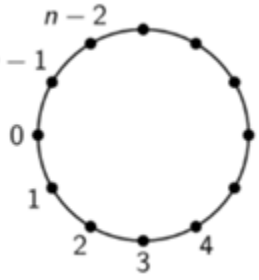
$\circ$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

[a] Possible uniquement dans le cas fini.

- Un monoïde est une structure $(A, \circ)$ telle que :
  - $\circ$ est associatif : $a \circ (b \circ c) = (a \circ b) \circ c = a \circ b \circ c$
  - $\circ$ admet un neutre, noté 1_A ou $\epsilon \in A$, tel que $\epsilon \circ a = a \circ \epsilon = a,\ \forall a \in A$.
- Sous-structure : $B \subseteq A$ est un sous-monoïde de $(A,\ \circ)$ si :
  - $\forall x, y \in B$, on a $x \circ y \in B$ (stabilité)
  - $1_A \in B$
- Soit $(A, \circ)$ un monoïde et $(a, b) \in A$. a est l'inverse de b si $a \circ b = b \circ a = \epsilon$.
- Un groupe est un monoïde $(G, \cdot)$ tel que tous les éléments de G possèdent un inverse par rapport à $\cdot$.
- Un groupe $(G, \cdot)$ est commutatif si $a \cdot b = b \cdot a,\ \forall a, b \in G$.
- $H \subseteq G$ est un sous-groupe de $(G, \cdot)$ si :
  - $\forall x, y \in H$, on a $x \cdot y \in H$
  - $1_G \in H$
  - Si $x \in H$, alors $x^{-1} \in H$
- Soit H un sous-groupe de G. La classe latérale de $a \in G$ modulo$H$ est $aH = \{ax | x \in H\}$. Quand H est clair dans le contexte, on écrit $[a]$ pour aH.
- Si G est commutatif, alors $(ax)\,(by) = (ab)\,(xy)$. En supposant que $x, y \in H$, $a' = ax$ et $b' = by$, cela implique que $a' \in aH$ et $b' \in bH \Rightarrow (a'b') \in (ab)\,H$. Le groupe quotient G/H est le groupe commutatif formé de
  - L'ensemble des classes latérales de G modulo H
  - L'opérateur défini par $[a]\,[b] = [ab]$

2.2 Théorèmes

- Le neutre d'un monoïde est unique : Soient ϵ, ϵ' neutres pour $(A, \circ)$. Alors $\epsilon \circ \epsilon' = \epsilon'$ et $\epsilon \circ \epsilon' = \epsilon \Longrightarrow \epsilon = \epsilon'$.
- Si b possède un inverse, alors cet inverse est unique : Soient a_1, a_2 tels que $a_1 \circ b = b \circ a_2 = 1_A$. Alors $a_1 = a_1 \circ 1_A = a_1 \circ (b \circ a_2) = (a_1 \circ b) \circ a_2 = 1_A \circ a_2 = a_2$.
- On peut simplifier les équations dans les groupes : si $ac = cb$, alors $a = b$: Soit $d = c^{-1}$. Alors $ac = bc \Longrightarrow a = a\epsilon = acd = bcd = b\epsilon = b$.
- Les inverses dans un groupe sont symétriques : Si $ab = \epsilon$, alors $a = b^{-1}$ et $b = a^{-1}$: $ab = \epsilon \Longrightarrow bab = b = babb^{-1} \Longrightarrow ba = \epsilon$.
- Il existe une bijection entre aH et H, et les classes latérales distinctes modulo H forment une partition de G.[1]
- Les classes latérales distinctes modulo H forment une partition de G. [2]
- Théorème de Lagrange : soient G un groupe fini et H un sous-groupe de G tels que $|G| = n$ et $|H| = m$, alors m divise n : les classes latérales de H sont toutes de taille m et forment une partition de G, donc $|G| = k\,|H|$ où k est le nombre de classes latérales de $H \Rightarrow |G/H| = |G| \,/\, |H|$.



2.3 Groupes

2.3.1 Sous-groupes des entiers

Soient $a \in \mathbb{Z}$ et $n \in \mathbb{N}^{>1}$. $a \; mod \; n$ est le reste de la division de a par n.

Soit $\mathbb{Z}_n = \{0, \ldots, n-1\}$ muni de l'addition modulo n : $a + b = r \; mod \; n$.

$\mathbb{Z}_n$ est isomorphe à $\mathbb{Z}/n\mathbb{Z}$: considérer $f \; : \; \mathbb{Z}_n \to \mathbb{Z}/n\mathbb{Z}$ telle que $f\,(a) = [a]$.

2.3.2 Sous-groupes finis

Soit H un sous-ensemble fini non-vide d'un groupe G. Si H est stable, alors H est un sous-groupe de G : Soit $a \in H$ et $f \; : H \to H : x \to ax$.

- Présence du neutre : comme f est bijective, $\exists b \in H : f\,(b) = a = ab \Rightarrow b = \epsilon$.
- Présence de l'inverse $\exists b \in H \; : f\,(b) = \epsilon \Rightarrow b = a^{-1}$.

[1] Démonstration 1.
[2] section 4.2.

2.3.3 Groupes cycliques

Soit G un groupe fini et $g \in G$. Le sous-groupe $\langle g \rangle$ engendré par g est $\langle g \rangle = \{g, g^2, g^3, \ldots\}$

$\rightarrow$ Remarque : si $\circ$ est $+$, alors g^2 est $2g$.

Il existe m : $g^m = \epsilon$, vu que $\langle g \rangle$ est un sous-groupe de G. L'ordre de g est la plus petite valeur de m telle que $g^m = \epsilon$, i.e. $|\langle g \rangle|$. Pour tout $g \in G$, l'ordre de g divise $|G|$.

Le groupe G est cyclique si il existe g tel que $\langle g \rangle = G$. Si $G = \langle g \rangle$ d'ordre fini n, alors $g^n = \epsilon$.

Soit G cyclique et $|G| = n$. G possède un unique sous-groupe H d'ordre m pour tout m divisant n et H est cyclique.

2.3.4 Résoudre une équation modulo

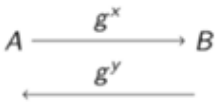
$$ax^2 + bx + c = 0 \bmod n \tag{2.1}$$

- Multiplier pour avoir une équation du type $x^2 + b'x + c' = 0 mod n$.
- Factoriser : $(x + \alpha)^2 + \beta = 0 \bmod n$.
- Isoler : $(x^2 + \alpha)^2 = -\beta \bmod n$.
- Chercher les valeurs de $x + \alpha$.
- Isoler x.

2.4 Protocole de Diffie-Hoffman

Soit $G = \langle g \rangle$ un groupe cyclique d'ordre n.

- A choisit $x \leftarrow \{0, 1, \ldots, n-1\}$
- B choisit $y \leftarrow \{0, 1, \ldots, n-1\}$



- Après échange de g^x et g^y, chacun calcule g^{xy}.

On peut facilement retrouver x et y avec un logarithme discret lorsque $g < 3072\ bits$.

2.4.1 Problème décisionnel de Diffie-Hoffman

Soit $G = \langle g \rangle$ un groupe cyclique d'ordre n. Soit $h = g^x$ un élément de G pioché au hasard.

$(G \times G)$ est un groupe, et $\langle (g, h) \rangle$ en est un sous-groupe. Peut-on distinguer un élément aléatoire de $(G \times G)$ d'un élément aléatoire de $\langle (g, h) \rangle$?

2.5 Anneau et corps

2.5.1 Anneau

Un anneau est une structure $(A, +, \cdot)$ telle que :

- $(A, +)$ est un groupe commutatif, on écrit 0 pour son neutre.
- $(A, \cdot)$ est un monoïde, on écrit 1 pour son neutre.
- La multiplication se distribue sur l'addition, à gauche et à droite.

Un anneau est commutatif si la multiplication est commutative.

2.5.2 Propriétés des anneaux

- 0 est absorbant : $\forall r,\ 0 \cdot r = r \cdot 0 = 0$
- Si A est un anneau et $0 = 1$, alors $A = \{0\}$
- Si $(A, +, \cdot)$ est un anneau tel que $0 \neq 1$, alors $(A, \cdot)$ n'est pas un groupe.

2.5.3 Corps

Un corps est un anneau $(A, +, \cdot)$ tel que $(A \setminus \{0\}, \cdot)$ est un groupe.

Un corps est commutatif si $(A \setminus \{0\}, \cdot)$ est un groupe commutatif.

2.5.4 Propriétés

- $a \in A$ est un diviseur de 0 si il existe $b \neq 0 : ab = 0$.
- Soit A un anneau. Si $a \in A$ est un diviseur de 0, alors a n'est pas inversible.
- Soit A un anneau fini. $a \in A$ est un diviseur de $0 \iff a \in A$ n'est pas inversible.

2.5.5 Anneau $\mathbb{Z}_n$

$\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ muni de l'addition modulaire et de la multiplication modulaire forme un anneau commutatif.

- Si n est premier, alors $\mathbb{Z}_n$ est un corps, noté $\mathbb{F}_n$.
- Si n est composé, alors $\mathbb{Z}_n$ n'est pas un corps.
- $a \in \mathbb{Z}_n$ est inversible ssi $\text{pgcd}\,(a, n) = 1$.

2.5.6 Anneau des polynômes

On peut définir les opération sur les polynômes au départ de celles sur un anneau sous-jacent.

Soit A un anneau, et $A[X]$ l'anneau des polynômes sur A. Les opérations sont héritées des polynômes classiques et de l'anneau sous-jacent.

2.5.7 Racines des polynômes

Si $\mathbb{F}$ est un corps, tout polynôme $p \in \mathbb{F}[X]$ de degré $n > 0$ possède au plus n racines.

2.6 Codes de Reed-Solomon

Un code est défini par une fonction $c : (\mathbb{F}_q)^k \to (\mathbb{F}_q)^n$:

- q est la taille de l'alphabet qu'on emploie pour décrire le message.
- Chaque élément de l'alphabet est appelé symbole.
- k est la dimension des messages que l'on veut encoder.
- n est la taille de bloc que l'on va transmettre $(n > k)$.

Idée : couper le message en séquences de k symboles, appliquer c et obtenir les $n > k$ symboles à transmettre sur le canal. On pourra corriger jusqu'à $(n - k)$ symboles.

On peut définir un code de Reed-Solomon $c : (\mathbb{F}_q)^k \to (\mathbb{F}_q)^n$:

- Interpréter un message $a_0, \ldots, a_{k-1} \in \mathbb{F}_q$ comme un polynôme $a = \sum_{i=0}^{k-1} a_i X^i \in \mathbb{F}_q[X]$
- Choisir n points distincts $x_1, \ldots x_n \in \mathbb{F}_q$.
- Calculer le mode-code $a(x_1), \ldots a(x_n)$.

À partir de n'importe quels k symboles parmi n, on peut reconstruire le message original par interprétation polynomiale. Cela fonctionne car tout ensemble de k points définit un unique polynôme de degré au plus $k - 1$.[3]

[3] Si $k = 3$ et $n = 5$, on a une parabole. On peut perdre les informations de 2 des 5 points, car 3 points forment toujours une et une seule parabole (=information).

Dénombrement et probabilités

3.1 Dénombrer

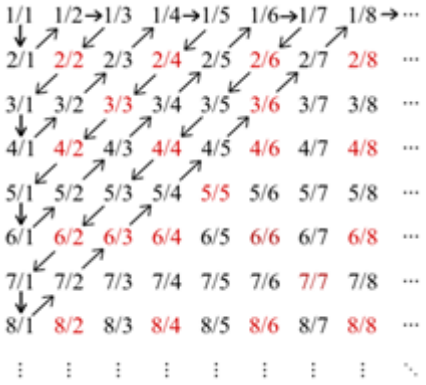
- Équipotence :

A et B sont équipotents ssi il existe une bijection de A vers B. L'équipotence est notée $A \approx B$.
A est fini si $A \approx \{1, \ldots, n\}$ pour $n \in \mathbb{N}$. n est alors le cardinal de A et est noté $|A|$.
A est infini dénombrable si $A \approx \mathbb{N}$.

- $\mathbb{Z} \approx \mathbb{N}$:

Il existe une bijection $f : \mathbb{N} \to \mathbb{Z} : x \to \begin{cases} (x+1)/2 \text{ si } x \text{ impair} \\ -x/2 \text{ si } x \text{ pair} \end{cases}$.

- $\mathbb{Q} \approx \mathbb{N}$:



Si on supprime tous les éléments rouges (fractions pouvant être réduites), on a un bijection de $\mathbb{Q} \to \mathbb{N}$. On peut ajouter une partie symétrique à gauche avec les rationnels négatifs.

Puisqu'on a deux bijection $\mathbb{Q} \to \mathbb{N}$ et $\mathbb{Z} \to \mathbb{N}$, on a également la bijection $\mathbb{Q} \to \mathbb{Z}$.

- $\mathbb{R} \not\approx \mathbb{N}$:

Voir section 4.3.
Protocole pour prouver $A \approx \mathbb{N}$:

1. Grouper les éléments de A selon une caractéristique commune et ordonner les groupes.
2. Indexer les éléments d'un même groupe ($= g(a)$).
3. Pour un certain groupe, calculer le nombre total d'éléments dans les groupes précédents ($= h(a)$).
4. $f : A \to \mathbb{N} : a \to g(a) + h(a)$

3.2 Probabilités : définitions

Soit $f : A \to \mathbb{R}^{\geq 0} : a \to 1$. On définit $|A| \coloneqq \sum_{a \in A} f(a)$.

La fonction f donne une mesure unitaire à chaque élément de l'ensemble. On peut donner d'autres mesures à ces éléments (coût d'un graphe : $c(G) = \sum_{r \in R} c(r)$,...). La mesure pourrait donc aussi être la probabilité d'un événement.

Soit Ω l'univers, i.e. l'ensemble des résultats possibles et ω le résultat. Il existe une bijection $p : \Omega \to \mathbb{R}^{\geq 0} : \omega \to p(\omega)$. Avec p la fonction de probabilité.

Soit un événement B, i.e. un ensemble de résultats b possibles. On définit $P(B) \coloneqq \sum_{b \in B} p(b)$.

3.2.1 Espace probabilisé

Un espace probabilisé est composé de :

- Un univers Ω.
- Un ensemble $\mathcal{A}$ de sous-ensembles de Ω reprenant tous les événements auxquels on peut s'intéresser.
- Une mesure de probabilité $P : \mathcal{A} \to [0, 1]$ qui assigne une probabilité à chaque événement.

Exigences sur la manière de définir $\mathcal{A}$:

- Si $A \in \mathcal{A}$, alors $\Omega - A \in \mathcal{A}$.
- Si $A, B \in \mathcal{A}$, alors $A \cup B \in \mathcal{A}$.
- $\Omega \in \mathcal{A}$.

3.2.2 σ-algèbre

On demande que l'ensemble des événements forme une σ-algèbre sur Ω :
$\mathcal{A} \subseteq \mathcal{P}(\Omega)$[1] est une σ-alèbre sur Ω si :

- $\Omega \in \mathcal{A}$.
- $A \in \mathcal{A} \Longrightarrow \Omega - A \in \mathcal{A}$.
- Si $A_1, A_2, \ldots$ est un ensemble dénombrable d'éléments de $\mathcal{A}$, alors $A = A_1 \cup A_2 \cup \ldots \in \mathcal{A}$.

Conséquences :

- $\emptyset \in \mathcal{A}$ car $\Omega \in \mathcal{A}$ et $\emptyset = \Omega - \Omega$.
- Si $A, B \in \mathcal{A}$, alors $A \cap B = \overline{\overline{A} \cup \overline{B}} \in \mathcal{A}$, avec $\overline{X} = \Omega - X$.
- Si $A, B \in \mathcal{A}$, alors $A - B = A \cap \overline{B} \in \mathcal{A}$.

[1] $\mathcal{P}(\cdot)$ est le power set, i.e. l'ensemble des sous-ensembles possibles.

3.2.3 Mesure de probabilité

Une mesure de probabilité $P : \mathcal{A} \to \mathbb{R}^{\geq 0}$ qui assigne une probabilité à chaque événement d'une σ-algèbre doit satisfaire les 3 axiomes de Kolmogorov :

- $0 \leq P(A) \leq 1 \; \forall A \in \mathcal{A}$
- $P(\Omega) = 1$
- Si $A_1, A_2, \ldots$ est une famille dénombrable d'événements disjoints ($A_i \cap A_j = \emptyset \; \forall i \neq j$), alors $P(A_1 \cup A_2 \cup \ldots) = \sum_i P(A_i)$

3.2.4 Mesure de probabilité uniforme

Soit Ω un ensemble fini, $\mathcal{A} = \mathcal{P}(\Omega)$, et $P(\{\omega\}) = \frac{1}{|\Omega|} \; \forall \omega \in \Omega$. Alors $P(A) = \frac{|A|}{|\Omega|}$ avec $A \in \mathcal{A}$. Il s'agit de la mesure de probabilité uniforme sur Ω.

3.2.5 Mesure positive

Une mesure positive $\mu : \mathcal{A} \to \mathbb{R}^{\geq 0}$ qui assigne un nombre réel (ou infini) à chaque événement d'une σ-algèbre doit satisfaire les 2 axiomes suivants :

- $\mu(\emptyset) = 0$.
- Si $A_1, A_2, \ldots$ est une famille dénombrable d'événements disjoints, $\mu(A_1 \cup A_2 \cup \ldots) = \sum_i \mu(A_i)$.

3.3 Somme et union

3.3.1 Règle de la somme

- $A \cap B = \emptyset \implies |A \cup B| = |A| + |B|$
- $A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$ (// Axiome 3 de Kolmogorov).

3.3.2 Règle du produit

- $|A \times B| = |A| \cdot |B|$

3.3.3 Principe d'inclusion et d'exclusion

- $|A \cup B| = |A| + |B| - |A \cap B|$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$

Généralisation :
Soit $R_n := \{1, \ldots, n\}$ et $S_I := \cap_{i \in I} S_i$

$$\left| \bigcup_{i \in R_n} S_i \right| = \sum_{r=1}^{n} \left((-1)^{r-1} \sum_{I \subset R_n,\ |I|=r} |S_I| \right) \tag{3.1}$$

$$P\left(\bigcup_{i \in R_n} S_i \right) = \sum_{r=1}^{n} \left((-1)^{r-1} \sum_{I \subset R_n,\ |I|=r} P(S_I) \right) \tag{3.2}$$

3.4 Nombres binomiaux

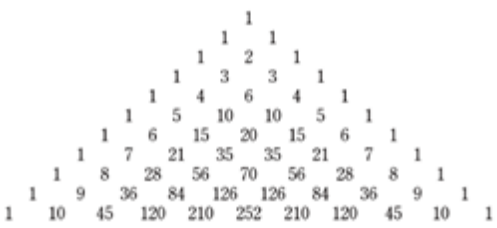
$B(n,k)$ est le nombre de sous-ensembles de cardinal k d'un ensemble de cardinal n. On le note $B(n,k) = \binom{n}{k} = C_n^k$.

Il s'agit par exemple du nombre d'ordres possibles dans lesquels on peut placer k croix sur une grille de n cases.

$$\boxed{B(n,k) = \frac{n!}{k!\,(n-k)!}} \tag{3.3}$$

3.4.1 Propriétés

- Par symétrie, $B(n,k) = B(n, n-k)$. Chaque sous-ensemble de taille k définit un unique sous-ensemble de taille $n-k$, et inversement.
- Soient A un n-ensemble, $\mathcal{A}_k$ l'ensemble de ses k-sous-ensembles et $\mathcal{A}_{n-k}$ l'ensemble de ses $(n-k)$-sous-ensembles. La fonction $f : \mathcal{A}_k \to \mathcal{A}_{n-k} : x \to A \backslash x$ est une bijection.
- Par récurrence, $B(n,k) = B(n-1, k-1) + B(n-1,k)$. Les sous-ensembles de A de taille k sont exclusivement, soit des sous-ensembles de taille k de $A \backslash \{x\}$, soit des sous-ensembles de taille $k-1$ de $A \backslash \{x\}$ auxquels on ajoute x.
- Triangle de Pascal :



- $\sum_{k=0}^{n} B(n,k) = 2^n$

3.4.2 Mot caractéristique

"01011" est le mot caractéristique du sous-ensemble si le

- i-ème symbole est "1" si l'élément a_i est dans le sous-ensemble.
- i-ème symbole est "0" si l'élément a_i n'est pas dans le sous-ensemble.

Toute séquence de n "0" et "1" définit un unique sous-ensemble 2^n sous-ensembles.

→ Remarque : $B(n,k)$ est le nombre de mots binaires de longueur n et de poids (nombre de "1") k.

3.4.3 Binôme de Newton

$$(x+y)^n = \sum_{k=0}^{n} B(n,k)\, x^k y^{n-k} \Longrightarrow \sum_{k=0}^{n} (-1)^k B(n,k) = 0 \; si \; n \geq 1 \tag{3.4}$$

3.4.4 Matrice des nombres binomiaux

Si on définit une matrice triangulaire dont les lignes sont celles du triangle de Pascal, on obtient un inverse très similaire :

$$\begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ 1 & 2 & 1 & & \\ 1 & 3 & 3 & 1 & \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ 1 & -2 & 1 & & \\ -1 & 3 & -3 & 1 & \\ 1 & -4 & 6 & -4 & 1 \end{pmatrix} = I_5$$

3.4.5 k-sélections d'un n-ensemble

[2]

Chaque élément du n-ensemble peut apparaitre plusieurs fois, on peut donc avoir $k > n$.

Une sélection sur l'ensemble A est une fonction $s : A \to \mathbb{N}$. Une k-sélection est telle que $\sum_{a \in A} s(a) = k$. Si $A \coloneqq \{a_1, \ldots, a_n\}$, on note $s \coloneqq a_1^{(k_1)} \ldots a_n^{(k_n)}$ où $k_i = s(a_i)$ et on note le nombre de k-sélections d'un n-ensemble $B^*(n,k)$.

On peut représenter la k-sélection s par son mot caractéristique : $11\ldots101\ldots10\ldots01\ldots1$. On écrit k_1 "1", puis un "0", puis k_2"1",... jusque k_n "1".

- Chaque sélection donne un mot contenant k "1" et $n-1$ "0".
- Chaque mot contenant k "1" et $n-1$ "0" donne une sélection.
- Chaque mot de ce type définit aussi un k-sous-ensemble d'un $(k+n-1)$-ensemble.

$$\Longrightarrow B^*(n,k) \coloneqq B(k+n-1,k) \tag{3.5}$$

3.5 Fonctions : injections et surjections

3.5.1 Dénombrer les fonctions

Une relation R de A vers B est un ensemble $R \subset A \times B$.

Une fonction de A vers B est un triple (A,B,R), avec R une relation de A vers B, telle que $\forall a \in A, \exists 1!\, b \in B : aRb$.

[2] Avec répétitions.

3.5.2 Injections

Soient A et B finis, $|A| = n,\ |B| = k$. On a

- Le nombre $\left(A^B\right)$ de fonctions de B vers A est noté n^k, car $\left|A^B\right| = |A|^{|B|} = n^k$.
- Le nombre $\text{In}\,(n \leftarrow k)$ d'injections de B vers A est $[n]_k \coloneqq n\,(n-1)\ldots(n-k+1)$.
- Le nombre de bijections de B vers A (si $k = n$) est $n!$.
- Le nombre de bijections de A vers A (ou permutations de A) est $n!$.

3.5.3 Surjections

Soient A et B finis, $|A| = n,\ |B| = k$, avec $k \leq n$. On a

- Le nombre $\text{Sur}\,(n \to k)$ de surjections de A vers B est $\sum_{r=0}^{k} (-1)^r\, B\,(k,r)\,(k-r)^n$
- Soient $S = \{\text{ fonctions de } A \to B\} = B^A$ et $S_i \subset S$ tel que aucun élément de A n'est envoyé sur b_i. On a $\text{Sur}\,(n \to k) = |S| - |S_1 \cup \ldots \cup S_k|$

$$|S| = k^n,\ \left|\bigcup_{i\in\{1,\ldots k\}} S_i\right| = \sum_{r=1}^{k}\left((-1)^{r-1}\sum_{I\subset\{1,\ldots,k\},\ |I|=r}|S_I|\right) \tag{3.6}$$

Avec $S_I = \bigcap_{i\in I} S_i$. Par développement, on a finalement

$$\boxed{\text{Sur}\,(n \to k) = \sum_{r=0}^{k} (-1)^r\, B\,(k,r)\,(k-r)^n} \tag{3.7}$$

3.5.4 Dérangements

Un dérangement sur A est une permutation f de A sans point fixe (sans qu'un a ne reste à sa place) : $\forall a \in A\ :\ f\,(a) \neq a$.
Le nombre de dérangements d'un n-ensemble est

$$d_n \coloneqq n! \sum_{r=0}^{n} \frac{(-1)^r}{r!} \tag{3.8}$$

On a donc $\lim_{n\to\infty} \frac{d_n}{n!} = \frac{1}{e}$.

On peut définir le nombre de k-partitions d'un ensemble de cardinal n : $S\,(n,k) \coloneqq \frac{1}{k!}\text{Sur}\,(n \to k)$.

3.5.5 Nombres de Stirling

On a par récurrence $S\,(n,k) = S\,(n-1\,,k-1) + k\,S\,(n-1,k)$. Toute répartition de l'ensemble A de cardinal n en k blocs est :

- Soit une partition de $A \backslash \{x\}$ en $k-1$ blocs, à laquelle on ajoute le bloc x (premier terme de la somme).
- Soit une partition de $A \backslash \{x\}$ en k blocs, modifiée en ajoutant x à l'un des blocs (second terme de la somme).

3.5.6 Nombres de Bell

Soit b_n le nombre de partitions distinctes sur n éléments. $b_n \coloneqq \sum_{i=0}^{n} S(n,i)$. Par récurrence, $b_n = \sum_{i=0}^{n-1} B(n-1,i)\, b_i$.

3.6 Approximation de Stirling

On peut évaluer les factorielles par la relation suivante :

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \Longrightarrow \lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1 \tag{3.9}$$

On a donc des bornes pour les factorielles et les nombres binomiaux :

$$n! \geq \left(\frac{n}{e}\right)^n,\ \left(\frac{n}{k}\right)^k \leq B(n,k) \leq \left(\frac{ne}{k}\right)^k \tag{3.10}$$

3.7 Lois et variables aléatoires discrètes

Pour des combinaisons équiprobables, la probabilité $= \frac{\text{nombre de cas favorables}}{\text{nombre de cas totaux}}$. Pour un lancer de dés, la probabilité de gagner k fois parmi n manches est

$$P = \frac{B(n,k)}{2^n} \tag{3.11}$$

3.7.1 Épreuve et schéma de Bernoulli

On définit un espace probabilisé (épreuve de Bernoulli):

- $\Omega = \{E, S\}$
- E pour échec et S pour succès
- La σ-algèbre des événements est $P(\Omega) = \{\emptyset, \{E\}, \{S\}, \Omega\}$
- Si la probabilité de succès est p, la probabilité d'échec est $q = 1 - p$

Pour un espace probabilisé $\Omega = \{E, S\}^n$ (schéma de Bernoulli), la σ-algèbre des événements est $\mathcal{P}(\Omega)$ et la probabilité d'un mot à k succès et $n-k$ échecs est $P = p^k q^{n-k}$.

3.7.2 Événements indépendants

Dans un espace probabilisé $(\Omega, \mathcal{A}, P)$ quelconque, deux événements A, B sont indépendants si $P(A \cap B) = P(A)\, P(B)$. On note l'indépendance $A \perp\!\!\!\perp B$.

Trois événements A, B, C sont indépendants si :

- $P(A \cap B) = P(A)\, P(B)$
- $P(B \cap C) = P(B)\, P(C)$

- $P(A \cap C) = P(A)\,P(C)$
- $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$

!! Des événements peuvent être indépendants deux à deux, sans l'être à trois.

3.7.3 Variables aléatoires

Soient un espace probabilisé $(\Omega, \mathcal{A}, P)$ et un espace $(\Omega', \mathcal{A}')$ avec une σ-algèbre d'événements, mais pas de mesure de probabilité. Une variable aléatoire est une fonction $X : \Omega \to \Omega'$ telle que pour tout événement $A' \subseteq \Omega'$, la préimage $X^{-1}(A')$ est un événement de Ω. La variable aléatoire de l'espace Ω' attribue une probabilité $P'(A')$ à chaque événement A' de Ω' : $P'(A') = P(X^{-1}(A))$.

$P'(A')$ peut également s'écrire $P(X \in A')$, $P(\{\omega \in \Omega : X(\omega) \in A'\})$.

3.7.4 Épreuve d'un schéma de Bernoulli

Soit la variable aléatoire $X_k : \Omega \to \Omega'$ où

- $\Omega = \{E, S\}^n$ muni d'un schéma de Bernoulli, avec probabilité p de succès.
- $\Omega' = \{E, S\}$, épreuve de Bernoulli de probabilité p de succès.

$\Rightarrow$ X_k indique le résultat de la $k$ème épreuve pour $1 \leq k \leq n$.

Ces notations sont alors équivalentes :

- P("La $k$ème épreuve est un succès") $= p$
- $P(X_k = S) = p$
- $P'(S) = p$

3.7.5 Loi binomiale

Soit la variable aléatoire $N : \Omega \to \Omega'$, où

- $\Omega = \{E, S\}^n$ muni d'un schéma de Bernoulli, avec probabilité p de succès.
- $\Omega' = \{E, S\}$, épreuve de Bernoulli de probabilité p de succès.
- N est le nombre de succès.

$$\boxed{P(N = k) = B(n, k)\,p^k q^{n-k}} \tag{3.12}$$

On dit que N est une variable aléatoire binomiale, ou encore qu'elle suit une loi/distribution binomiale.

3.7.6 Variable indicatrice d'un événement

On considère un espace probabilisé quelconque $(\Omega, \mathcal{A}, P)$ et un événement quelconque $A \in \mathcal{A}$. Alors la fonction caractéristique de A est définie par $1_A : \Omega \to \{0,1\} : \omega \to 1$ ssi $\omega \in A$, et s'appelle aussi variable (aléatoire) indicatrice de l'événement A. On a $P(1_A = A) = P(A)$.

3.7.7 Variables aléatoires indépendantes

Soient trois espaces probabilisés $(\Omega, \mathcal{A}, P)$, $(\Omega_0, \mathcal{A}_0, P_0)$, $(\Omega_1, \mathcal{A}_1, P_1)$, avec deux variables aléatoires $X_0 : \Omega \to \Omega_0$ et $X_1 : \Omega \to \Omega_1$. Ces deux variables aléatoires sont indépendantes si pour tous événements $A_0 \subseteq \Omega_0$, $A_1 \subseteq \Omega_1$, on a que $X_0^{-1}(A_0)$ et $X_1^{-1}(A_1)$ sont des événements indépendants dans Ω, i.e. si tout événement de la forme $X_0 \in A_0$ est indépendant de tout événement de la forme $X_1 \in A_1$.

Les variables aléatoires sont donc indépendantes si

$$P\left(X_0^{-1}(A_0) \cap X_1^{-1}(A_1)\right) = P\left(X_0^{-1}(A_0)\right) P\left(X_1^{-1}(A_1)\right) = P_0(A_0) P_1(A_1) \tag{3.13}$$

3.7.8 Produits d'espaces probabilisés

Soient deux espaces probabilisés $(\Omega_0, \mathcal{A}_0, P_0)$ et $(\Omega_1, \mathcal{A}_1, P_1)$. Leur produits est l'espace probabilisé $(\Omega, \mathcal{A}, P)$ défini tel que :

- $\Omega = \Omega_0 \times \Omega_1$
- La σ-algèbre des événements A comprend tous les produits cartésiens d'événements $A_0 \times A_1$, et toutes leurs unions dénombrables, compléments, intersection dénombrables.
- P est définie par $P(A_0 \times A_1) = P_0(A_0) P_1(A_1)$ et étendue aux autres événements par les axiomes de Kolmogorov.

Propriété : soient espaces probabilisés $(\Omega_0, \mathcal{A}_0, P_0)$ et $(\Omega_1, \mathcal{A}_1, P_1)$ et leur produit sur $\Omega_0 \times \Omega_1$. Alors les projections sur la première composante, $X_0 : \Omega_0 \times \Omega_1 \to \Omega_0$ et sur la deuxième composante $X_1 : \Omega_0 \times \Omega_1 \to \Omega_1$ sont des variables aléatoires indépendantes.

3.7.9 Mesure de probabilité conditionnelle

Soient un espace probabilisé $(\Omega, \mathcal{A}, P)$ et un événement $A \subseteq \Omega$ de probabilité non nulle. Il est possible de définir une mesure de probabilité sur A :

- L'univers est A
- Les événements sont les événements de $\Omega \cap A$
- La probabilité de l'événement $B \subseteq \Omega$, notée $P(B|A)$, est $\frac{P(A \cap B)}{P(A)}$

3.7.10 Indépendance et probabilité conditionnelle

Dans un espace probabilisé $(\Omega, \mathcal{A}, P)$ quelconque, soient deux événements A, B de probabilités non nulles. Ils sont indépendants ssi $P(B) = P(B|A)$ ssi $P(A) = P(A|B)$.

3.7.11 Schéma de Bernoulli infini

On peut imaginer un nombre infini d'épreuves indépendantes. L'univers est $\Omega \{S, E\}^{\mathbb{N}}$. On définit des événements de base de la forme $X_i \in A_i$, $i = 1, \ldots, n$, avec n non borné fini pour tout ensemble $A_k \subseteq \{E, S\}$, avec X_k indiquant le résultat de la $k$ème épreuve.

On génère tous les événements par unions dénombrables, compléments et intersections dénombrables. On définit $P(X_i \in A_i) = \prod_{i=1}^{n} P(X_i \in A_i)$.

3.7.12 Loi géométrique

Soit un schéma de Bernoulli infini, avec p probabilité de succès, $q = 1 - p$ probabilité d'échec. Le nombre d'essais infructueux avant d'obtenir un premier succès est $X : \Omega \to \mathbb{N} \cup \{+\infty\}$. La probabilité de k échecs suivis d'un succès est $q^k p = P(EE \ldots ES) = P(X = k)$.

On a une série géométrique infinie si $k = +\infty : 1 + q + q^2 + \ldots = \frac{1}{1-q}$, donc la probabilité de ne jamais avoir de succès est $1 - p \cdot \frac{1}{1-q} = 0$. L'événement n'est pas impossible, mais il a une probabilité nulle. L'ensemble vide n'est donc pas le seul ensemble à avoir une probabilité nulle.

3.8 Lois et variables aléatoires réelles

3.8.1 L'aiguille de Buffon

Imaginons la situation où nous devons lancer une aiguille de longueur 1 sur un parquet de lattes de largeur 2. Quelle serait la probabilité pour l'aiguille de se trouver à cheval sur deux lattes ?

Hypothèses de calcul :

- Les arêtes du plancher sont des droites de coordonnées $x = 0,\ x = \pm 2$, etc.
- L'aiguille est caractérisée par la coordonnée en x de sa tête, et l'angle θ de l'aiguille avec l'horizontale.
- On suppose la coordonnée x dans l'intervalle $[0, 1]$, les autres cas pouvant s'y réduire par symétrie.
- Toute configuration (x, θ) de l'aiguille dans $[0, 1] \times [0, \pi]$ a la même probabilité de subvenir.

On définit la probabilité comme :

$$P = \frac{\text{Surface des cas } (x, \theta) \text{ favorables}}{\text{Surface des cas } (x, \theta) \text{ totaux}} \tag{3.14}$$

Dans notre cas, $P = 1/\pi$, voir section 4.4.

3.8.2 Mesure de probabilité uniforme – cas discret

Pour la mesure de probabilité uniforme sur un ensemble fini Ω, nous avons besoin d'un univers Ω, d'une σ-algèbre $\mathcal{P}(\Omega)$, et pour tout événement A, nous avons $P(A) = \frac{|A|}{|\Omega|}$.

3.8.3 Mesure de probabilité uniforme – cas continu

Dans $\mathbb{R}^n$, on définit une σ-algèbre $\mathcal{B}$ de parties mesurables, dite σ-algèbre de Borel-Lebesgue :

- Les produits cartésiens d'intervalles sont mesurables.
- Les unions finies/dénombrables et les compléments d'ensembles mesurables sont mesurables.

On peut maintenant définir une mesure μ: $\mathcal{B} \to \mathbb{R} \cup \{\infty\}$:

- $\mu([a_1, b_1] \times \ldots \times [a_n, b_n]) = |b_1 - a_1| \ldots |b_n - a_n|$
- Si A est une union finie ou dénombrables de mesurables A_i disjoints deux à deux, alors $\mu(A) = \sum_i \mu(A_i)$

De cette mesure viennent les notions de longueur, surface,…

Pour la mesure de probabilité uniforme sur un ensemble mesurable $\Omega \subseteq \mathbb{R}^n$ de mesure $\mu(\Omega)$ finie, nous avons besoin d'un univers Ω, d'une σ-algèhbre des événements : mesurables de $\mathbb{R}^n$ intersectés avec Ω, et pour tout événement $A \subseteq \Omega$: $P(A) = \frac{\mu(A)}{\mu(\Omega)} \left(= \frac{\text{surface}_A}{\text{surface}_\Omega} \right)$.

3.8.4 Lien entre les cas discret et continu

Considérons la mesure de probabilité uniforme sur $\Omega = [0, 1]$. Pour $x \in [0, 1]$, on peut écrire $x = 0.b_1 b_2 b_3 \ldots$, où $b_i \in \{0, 1\}$ est le $i$ème bit du développement binaire de x. Chaque b_i prend la valeur $b_i = 0$ ou $b_i = 1$, avec probabilité $1/2$. $\forall i \neq j,\ b_i$ et b_j sont indépendants.

Donc la suite $b_1 b_2 b_3 \ldots$ forme un schéma de Bernoulli infini avec une probabilité de succès $1/2$.

3.8.5 Produits d'espaces de probabilité – cas uniforme discret

Les projections $\Omega \to \Omega_0$ et $\Omega \to \Omega_1$ sur chaque composante sont des variables aléatoires indépendantes. Le produit de deux ensembles finis munis de la mesure de probabilité uniforme est un ensemble fini avec la mesure de probabilité uniforme.

3.8.6 Produits d'espaces de probabilité – cas uniforme continu

Le produit de deux ensembles $\Omega_0 \subseteq \mathbb{R}^m$ et $\Omega_1 \in \mathbb{R}^n$ chacun muni de la mesure de probabilité uniforme est l'ensemble $\Omega = \Omega_0 \times \Omega_1 \subseteq \mathbb{R}^{m+n}$ muni de la mesure de probabilité uniforme. En effet,

$$P(A_0 \times A_1) = P_0(A_0) P_1(A_1) = \frac{\mu(A_0)}{\mu(\Omega_0)} \frac{\mu(A_1)}{\mu(\Omega_1)} = \frac{\mu(A_0 \times A_1)}{\mu(\Omega_0 \times \Omega_1)} \tag{3.15}$$

→ Remarque : les deux projections $\Omega \to \Omega_0,\ \Omega \to \Omega_1$ sur les deux composantes sont des variables aléatoires indépendantes.

3.8.7 Approximation de la binomiale par une loi de Poisson

Pour rappel, la loi binomiale est $P(N=k) = B(n,k)\, p^k q^{n-k}$. On note $N \sim Bin(n,p)$.

Pour un $n \gg 1$ et un $p \ll 1$, on peut approximer la binomiale par

$$P(N=k) \approx \frac{\mu^k}{k!} e^{-\mu} \tag{3.16}$$

On appelle cette variable aléatoire entière une variable aléatoire de Poisson de paramètre μ, notée $N \sim \text{Po}(\mu)$.

Pour un événement $A \subseteq \mathbb{N}$, $|P(N \in A) - P(N' \in A)| < \min(p,\ np^2)$
En pratique, l'approximation n'est bonne que si p est petit : l'erreur absolue reste toujours petite, mais l'erreur relative peut devenir grande.

Supposons que dans chaque très petit intervalle de temps Δt choisi arbitrairement, on ait un succès ("arrivée") avec probabilité $p = \lambda \Delta t$, et deux succès ou plus avec une probabilité $\mathcal{O}(\Delta t^2)$ négligeable. On considère un intervalle de temps T, découpé en $n = T/\Delta t$ intervalles de temps.
Lorsque $\Delta t \to 0$, le nombre d'arrivées N dans un intervalle de temps $[0,T]$ est

$$P(N=k) = \frac{(\lambda T)^k}{k!} e^{-\lambda T} \tag{3.17}$$

Ici, la loi de Poisson est exacte, car l'erreur commise est 0, puisque l'on choisit arbitrairement le Δt.
Par le point suivant lorsque $\Delta t \to 0$, pour un temps τ entre deux arrivées, $P(\tau \in [t + \Delta t\,[\,) = \lambda e^{-\lambda t} \Delta t$.
Il s'agit de la loi exponentielle : $\tau \sim \text{Exp}(\lambda)$.

3.8.8 Variable aléatoire géométrique – limite $p \ll 1$

Soit un schéma de Bernoulli infini, avec p la probabilité de succès, et $q = 1-p$ la probabilité d'échec. Soit K le nombre d'essais infructueux avant d'obtenir un premier succès. La probabilité de k échecs suivis d'un succès est $q^k p$. On note $K \sim Geo(p)$.
Si $p \to 0$ et $k \to \infty$, avec pk constant, on a $q^k p = (1-p)^{\frac{kp}{p}}\ p \to pe^{-kp}$.

3.8.9 Variable aléatoire réelle sans mémoire

Une variable aléatoire réelle positive X est dite sans mémoire si $P(X \geq s+t | X \geq t) = P(X \geq s)\ \forall s,t \in \mathbb{R}^+$. Cela signifie que le temps déjà attendu n'aide en rien à prédire le temps encore à attendre. La loi exponentielle est le seul exemple de variable aléatoire réelle continue sans mémoire.

$$P(X \geq t) = \int_t^\infty \lambda e^{-\lambda r} dr = e^{-\lambda T} \tag{3.18}$$

3.8.10 Variable aléatoire entière sans mémoire

Une variable aléatoire entière positive X est dite sans mémoire si $P\left(X \geq s+t \middle| X \geq t\right) = P(X \geq s)\ \forall s, t \in \mathbb{N}$.
L'interprétation est la même que dans le cas réel.

$$P\left(X \geq t\right) = q^t \tag{3.19}$$

3.8.11 Temps d'attente indépendants

Dans un schéma de Bernoulli infini, soit X_1 le nombre d'échecs avant le premier succès, X_2 entre le premier et le second,… Les temps d'attentes sont tous de même loi géométrique, et tous indépendants entre eux. Les variables X_i sont indépendantes et identiquement distribuées (i.d.d.).

3.8.12 Fonction de répartition – cas réel

Pour définir une mesure de probabilité sur $\mathbb{R}$, il suffit de la définir sur tous les intervalles. La fonction de répartition de cette mesure est définie par

$$F\left(x\right) = P\left(\left]\infty, x\right]\right) = P\left(X \leq x\right) \tag{3.20}$$

Propriétés :

- $\lim_{x \to -\infty} F\left(x\right) = 0$
- $\lim_{x \to +\infty} F\left(x\right) = 1$
- $F\left(x\right)$ est croissante
- Pour $y < x$, $F\left(x\right) - F\left(y\right) = P\left(\left]y, x\right[\right)$
- $\forall x, \sup_{y<x} F\left(y\right) = P\left(\left]-\infty, x\right[\right) = P\left(X < x\right)$
- $\forall x,\ F\left(x\right)$ est continue à droite : $\inf_{y>x} F\left(y\right) = F\left(x\right)$

Donc connaître F permet de calculer la probabilité de tout intervalle, et donc de tout événement.

3.8.13 Densité de probabilité – cas réel

Pour certains $F\left(x\right)$, il existe une fonction $\mathbb{R} \to \mathbb{R}$ telle que $F\left(x\right) = \int_{-\infty}^{x} f\left(t\right) dt$. On dit que f est une fonction de densité de probabilité, et la variable aléatoire est dite continue.

La probabilité d'un événement $A \subseteq \mathbb{R}$ est $\int_A f\left(x\right) dx$.

3.8.14 Probabilités individuelles – cas entier

Dans certains cas, la mesure de probabilité est concentrée sur un ensemble dénombrable $S \subseteq \mathbb{R}$, i.e. $P\left(S\right) = 1$. La mesure de probabilité est discrète. Chaque élément $s \in S$ est de probabilité $p_s = P\left(\{s\}\right)$, et $\sum_{s \in S} p_s = 1$.

La fonction caractéristique F est alors constante par morceaux, et discontinue en chaque $s \in S$. Il n'existe donc pas de fonction de densité de probabilité.

3.9 Caractéristiques des variables aléatoire réelles

3.9.1 Espérance d'une variable aléatoire réelle

Si la variable discrète X prend des valeurs dans un ensemble dénombrable $S \subseteq \mathbb{R}$, l'espérance (ou moyenne, valeur attendue ou espérée) de X est définie par

$$\mathbb{E}X = \sum_{s \in S} s p_s \tag{3.21}$$

Dans le cas continu de densité de probabilité f, elle est définie par

$$\mathbb{E}X = \int_{\mathbb{R}} x f(x)\, dx \tag{3.22}$$

En toute généralité,

$$\mathbb{E}X = \lim_{\Delta x \to 0} \sum_{x = k\Delta x,\ k \in \mathbb{Z}} x P\left(X \in]x - \Delta x/2,\ x + \Delta x/2]\right) \tag{3.23}$$

Si X est positive l'espérance s'interprète comme la surface entre la fonction de répartition et la droite horizontale $y = 1$, i.e.

$$\mathbb{E}X = \int 1 - F(x)\, dx \tag{3.24}$$

3.9.2 Espérance d'une variable aléatoire réelle géométrique

Soit la variable aléatoire géométrique X le nombre attendu d'échecs avant un succès dans un schéma de Bernoulli.

$$\mathbb{E}X = \sum_k k q^k p = \frac{1}{p} - 1 \tag{3.25}$$

3.9.3 Fonction de répartition d'une fonction de variable réelle

La fonction d'une variable aléatoire réelle est une variable aléatoire réelle.
Supposons h: $\mathbb{R} \to \mathbb{R}$ et $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle. Alors on peut définir la variable $Y = h(X) : \Omega \to \mathbb{R}$.

La fonction de répartition de Y est

$$F_Y(y) = P(Y \leq y) = P\left(X \in h^{-1}(]-\infty, y[)\right) = P\left(\{\omega \in \Omega |\ h(X(\omega)) \leq y\}\right) \tag{3.26}$$

3.9.4 Densité de répartition d'une fonction de variable réelle

Supposons $h : \mathbb{R} \to \mathbb{R}$ et $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle continue de densité f. La variable aléatoire $Y = h(X) : \Omega \to \mathbb{R}$ n'est pas forcément continue, mais si h est une bijection dérivable, alors la densité existe et est $g(y) = \frac{f(x)}{|h'(x)|}$ pour $y = h(x)$.

3.9.5 Espérance d'une fonction de variable aléatoire réelle

Supposons $h : \mathbb{R} \to \mathbb{R}$ et $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle. Alors on peut définir la variable $Y = h(X) : \Omega \to \mathbb{R}$

Si Y est discrète sur un ensemble S_Y :

$$\mathbb{E}Y = \sum_{y \in S_Y} yP(Y = y) \tag{3.27}$$

Et si Y est continue et de fonction de densité g :

$$\mathbb{E}Y = \int yg(y)dy \tag{3.28}$$

Si X est discrète sur un ensemble S_X :

$$\mathbb{E}Y = \sum_{x \in S_X} h(x)P(X = x) \tag{3.29}$$

Et si Y est continue et de fonction de densité f :

$$\mathbb{E}Y = \int h(x)f(x)dx \tag{3.30}$$

3.9.6 Mesure de probabilité conjointe

Si on s'intéresse à deux variables aléatoires (= vecteur aléatoire) $X : \Omega \to \mathbb{R}$ et $Y : \Omega \to \mathbb{R}$, on peut regarder la mesure de probabilité créée sur $\mathbb{R}^2$ par le vecteur aléatoire $(X, Y) : \Omega \to \mathbb{R}^2 : \omega \to (X(\omega), Y(\omega))$. C'est la mesure de probabilité conjointe de X et Y.

On peut la caractériser par la fonction de répartition $F(x, y) = P(X \leq x, Y \leq y)$.

Indépendance : $X \perp\!\!\!\perp Y$ ssi $F(x, y) = F_X(x)F_Y(y)$ pour tous x, y, avec F_I la fonction de répartition de I.

Cas discret :

S'il y a $S \subseteq \mathbb{R}^2$ dénombrable de probabilité 1, alors chaque $(x, y) \in S$ a une probabilité $p_{xy} = P(X = x, Y = y)$. Dans ce cas, la mesure de probabilité sur X (resp. Y) est aussi discrète, avec $p_x = \sum_{y:(x,y)\in S} p_{xy}$.

On peut aussi définir une probabilité conditionnelle :

$$p_{y|x} = \frac{p_{xy}}{p_x} = P(Y = y|X = x) \tag{3.31}$$

de sorte que $p_{xy} = p_x p_{y|x}$.

Indépendance : $X \perp\!\!\!\perp Y$ ssi $p_{xy} = p_x p_y, \forall (x, y) \in S$.

Cas continu

S'il y a une fonction de densité de probabilité $f : \mathbb{R}^2 \to \mathbb{R}$ telle que $F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x', y')dx'dy'$, alors chaque $(x, y) \in S$ a une probabilité $p_{xy} = P(X = x, Y = y)$. Dans ce cas, la mesure de probabilité sur X (resp. Y) est aussi continue, avec $f_X = \int_{\mathbb{R}} f(x, y)dy$.

On peut aussi définir une probabilité conditionnelle :

$$f(y|x) = \frac{f(x, y)}{f_X(x)} \tag{3.32}$$

de sorte que $f(x, y) = f_X(x)f(y|x)$.
Indépendance : $X \perp\!\!\!\perp Y$ ssi $f(x, y) = f_X(x)f_Y(y), \forall (x, y)$.

3.9.7 Composition de vecteurs aléatoires réels

Soit $(X, y) : \Omega \to \mathbb{R}^2$ un vecteur aléatoire réel, et $h : \mathbb{R}^2 \to \mathbb{R}$. On peut composer ces fonction et obtenir une nouvelle variable aléatoire réelle $Z : \Omega \to \mathbb{R} : \omega \to h(X(\omega), Y(\omega))$. On note $Z = h(X, Y)$.

Si (X, Y) est une variable aléatoire discrète sur $S \subset \mathbb{R}^2$, alors

$$\mathbb{E}Z = \sum_{(x,y)\in S} p_{xy} h(x, y) \tag{3.33}$$

Si (X, Y) est une variable continue sur $\mathbb{R}^2$ de densité de probabilité f, alors

$$\mathbb{E}Z = \int_{\mathbb{R}} h(x, y) f(x, y) dxdy \tag{3.34}$$

3.9.8 Linéarité de l'espérance

La combinaison linéaire de variables aléatoires reste une variable aléatoire. On en déduit une espérance :

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y) \tag{3.35}$$

3.9.9 Espérance d'une binomiale

Soit un schéma de Bernoulli à n épreuves, de probabilité de succès p. Soit X_i la variable indicatrice d'un succès à l'épreuve i, donc $X_i = 1$ si succès et 0 sinon. C'est une "indicatrice de Bernoulli". Son espérance est $\mathbb{E}X_i = p$.
Alors $X = \sum_i X_i$ est la variable binomiale, et son espérance est

$$\mathbb{E}X = np \tag{3.36}$$

3.9.10 Espérance d'une Poisson

Une variable X de Poisson de paramètre μ peut être obtenue comme la limite d'une binomiale $Bin(n, p)$ pour $\mu = pn$, avec $p \to 0, n \to \infty$. On a donc

$$\mathbb{E}X = \mu \tag{3.37}$$

Cela donne une interprétation au paramètre μ.

3.9.11 Moments et variance d'une variable aléatoire réelle

Soit une variable aléatoire réelle $X : \Omega \to \mathbb{R}$. Considérons ses puissances.

- Le moment d'ordre k de X est $\mathbb{E}(X^k)$.
- La variance est $\text{Var}(X) = \sigma_X^2 = \mathbb{E}((X - \mathbb{E}X)^2) = \mathbb{E}(X^2) - (\mathbb{E})^2$.
- La déviation standard (ou écart-type) est $\sigma_X = \sqrt{\text{Var}(X)}$

La déviation standard quantifie les fluctuations typiques autour de l'espérance.

3.9.12 Corrélation des variables aléatoires

Soit $Cov(X, Y)$ la covariance des variables aléatoires X, Y. Elle est définie par

$$Cov(X, Y) = \mathbb{E}(X - \mathbb{E}X) \cdot \mathbb{E}(Y - \mathbb{E}Y) \tag{3.38}$$

ou

$$Cov(X, Y) = \mathbb{E}XY - \mathbb{E}X \cdot \mathbb{E}Y \tag{3.39}$$

Soit ρ la corrélation entre les variables aléatoires X et Y.

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)} \tag{3.40}$$

Propriétés :

- $\rho(X, Y) = 1$ ssi $Y = aX + b$, pour $a > 0$
- $\rho(X, Y) = -1$ ssi $Y = aX + b$, pour $a < 0$
- $\rho(X, Y) = \rho(aX + bY)\ \forall a > 0, b$
- On peut trouver $a, b \in \mathbb{R}$ et une variable aléatoire ϵ d'espérance nulle, décorrélée de X et de variance $1 - \rho^2)\text{Var}(Y)$ telle que $Y = aX + b + \epsilon$
- C'est la meilleure relation linéaire entre X et Y : toute relation $Y = cX + d + \delta$ a une variance $\text{Var}(\delta) \geq \text{Var}(\epsilon)$

3.9.13 Corrélation de deux variables aléatoires réelles centrées réduites

Une variable aléatoire est centrée réduite lorsque son espérance est nulle et sa variance est l'unité :

$$\begin{cases} X' = (X - \mathbb{E}X)/\sigma(X) \\ Y' = (Y - \mathbb{E}X)/\sigma(Y) \end{cases} \tag{3.41}$$

Alors

$$\begin{cases} \mathbb{E}X'Y' = Cov(X', Y') = \rho(X', Y') = \rho(X, Y) \\ Y' = \rho X' + \epsilon' \end{cases} \tag{3.42}$$

avec ϵ' une variable aléatoire d'espérance nulle, de variance $1 - \rho^2$ et décorrélée de X.

$\rightarrow$ Remarque $\rho X'$ est la projection orthogonale de Y' le long de X' qui crée l'erreur de plus petite norme.

On déduit de cela que dans la relation $Y = aX + b + \epsilon$, $a = \rho(X, Y)\sigma(Y)/\sigma(X)$.

3.9.14 Corrélation et indépendance

Si les variables aléatoires X et Y sont indépendantes, le coefficient de corrélation, et donc la covariance également, est nulle.

3.9.15 Variance d'une somme de variables aléatoires indépendantes

Soient deux variables aléatoires $X : \Omega \to \mathbb{R}$ et $Y : \Omega \to \mathbb{R}$ indépendantes. La variance de leur combinaison linéaire est

$$\boxed{\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)} \tag{3.43}$$

- Pour une variable binomiale, on a une variance npq.
- Pour une loi de Poisson, on a une variance μ.

3.9.16 Espérance conditionnelle

$$p_{y|x} = P(Y = y|X = x) \tag{3.44}$$

L'espérance conditionnelle de Y sachant que $X = x$ est $\mathbb{E}(Y|X = x) = \sum_y y p_{y|x}$, et est un nombre réel qui dépend de x.

$$\mathbb{E}(Y|X = x) = \sum_y y p_{y|x} \tag{3.45}$$

La fonction $x \to \mathbb{E}(Y|X = x)$ est une variable aléatoire réelle et l'espérance conditionnelle de Y sachant X est $\mathbb{E}(Y|X)$. L'espérance totale de Y est donc $\mathbb{E}Y = \mathbb{E}(\mathbb{E}(Y|X))$. Autrement dit

$$\mathbb{E}Y = \sum_x p_x \mathbb{E}(Y|X = x) \tag{3.46}$$

Dans le cas continu, on a alors

$$\begin{cases} \mathbb{E}(Y|X = x) = \int_y y f(y|x) dy \\ \mathbb{E}Y = \int_x f_X(x) \mathbb{E}(Y|X = x) dx \end{cases} \tag{3.47}$$

3.9.17 Variance conditionnelle

La variance conditionnelle de Y sachant que $X = x$ est

- Cas continu :

$$\text{Var}(Y|X = x) = \sum_y (y - \mathbb{E}(Y|X = x))^2 p_{y|x} \tag{3.48}$$

$$\text{Var}(Y|X = x) = \int_y (y - \mathbb{E}(Y|X = x))^2 f(y|x) dy \tag{3.49}$$

Loi de la variance totale :

$$\boxed{\text{Var}Y = \mathbb{E}(\text{Var}(Y|X)) + \text{Var}(\mathbb{E}(Y|X))} \tag{3.50}$$

Pour la variable Y^2, la formule devient :

$$\mathbb{E}(Y^2) = \mathbb{E}(\mathbb{E}(Y^2|X)) = \mathbb{E}(\text{Var}(Y|X)) + \text{Var}(\mathbb{E}(Y|X)) + \mathbb{E}(Y)^2 \tag{3.51}$$

3.10 Théorèmes de concentration

3.10.1 Concentration autour de la moyenne

- Inégalité de Markov :

Pour une variable aléatoire réelle positive X, on a

$$P(X \geq a\mathbb{E}(X)) \leq \frac{1}{a}, \forall a > 0 \tag{3.52}$$

- Inégalité de Bienaymé-Tchebychev :

Pour une variable aléatoire réelle X, on a

$$P(|X - \mathbb{E}(X)| \geq a\sigma(X)) \leq \frac{1}{a^2}, \forall a > 0 \tag{3.53}$$

3.10.2 Loi des grands nombres

Etant donnée une suite X_i de variales i.i.d. d'espérance μ, les moyennes empiriques partielles convergent vers l'espérance :

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| \geq \epsilon\right) \to 0 \ \forall \epsilon > 0 \tag{3.54}$$

Cela signifie que la moyenne empirique sur un grand nombre d'expériences indépendantes converge vers l'espérance.

Si les X_i sont des variables indicatrices d'un événement de probabilité p, alors $X = \sum_{i=1}^{n} X_i$ est une binomiale $Bin(n, p)$, et $P(|\frac{X}{n} - p| \geq \epsilon) \to 0 \forall \epsilon > 0$

Cela signifie que la fréquence empirique de réalisation d'un événement sur un grand nombres d'expériences i.i.d. converge vers la probabilité de cet événement.

Cas général :

Soit la variance de chaque X_i finie.

- La variance de $\frac{\sum_{n=1}^{n} X_i}{n}$ est $\frac{\text{Var}X_1}{n}$.
- L'écart-type de $\frac{\sum_{n=1}^{n} X_i}{n}$ est $\frac{\sigma(X_1)}{\sqrt{n}}$

Fixons un certain ϵ. A mesure que n augmente et que la variance de la moyenne empirique diminue, on obtient que $\epsilon \geq a\frac{\sigma(X_1)}{\sqrt{n}}$, pour un a aussi grand qu'on veut. Alors

$$P\left(\left|\frac{\sum_(i = 1)^n X_i}{n} - \mathbb{E}(X_1)\right| \geq \epsilon\right) \leq \frac{1}{a^2} \tag{3.55}$$

Cette valeur converge bien vers 0.

3.10.3 Théorème central limite

Soit une suite X_i de variables i.i.d. d'espérance μ et de variance finie σ^2. Les moyennes partielles normalisées convergent vers une variable aléatoire universelle :

Soit $Z_n = \frac{\sum_{i=1}^{n} X_i}{n}$ la moyenne partielle des n premiers termes, d'espérance μ et de variance σ^2/n. On définit la variable centrée réduite

$$P(Z'_n \leq z) \to \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \tag{3.56}$$

A la limite, on a une variable aléatoire x de densité de probabilité $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

3.10.4 Loi normale

Une variable aléatoire Z normale (ou gaussienne) de moyenne μ et de variance σ^2 est définie par une densité de probabilité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.57}$$

Elle se note $Z \sim N(\mu, \sigma^2)$.

On peut également réduire et centrer la variable aléatoire normale X: $Z = \frac{X-\mu}{\sigma}$. Cela permet de ne pas devoir calculer les valeurs des probabilités, car elles se trouvent dans des tables.

$\to$ Remarque : il n'existe pas de formule explicite pour la fonction de répartition.

3.10.5 Convergence

Il est possible de borner l'erreur commise :

$$\left| P(Z'_n \leq z) - \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right| < \frac{\rho^3}{\sigma^3\sqrt{n}} = \mathcal{O}(1/\sqrt{n}) \tag{3.58}$$

avec $\rho^3 = \mathbb{E}(|X_i - \mathbb{E}X_i|^3)$

Indicatrice de Bernoulli :

Pour l'indicatrice de Bernoulli d'espérance p, on a $\sigma^2 = pq$ et $\rho^3 = pq(q^2 + p^2) \leq pq$. L'erreur est donc inférieure à $1/\sqrt{npq}$.

La binomiale est donc bien approximée par une normale si $npq >> 1$.

$\to$ Une binomiale s'approxime par une Poisson d'espérance $\mu = np$ si $p << 1$, $\forall n$, et par une normale d'espérance $\mu = np$ et de variance $\sigma^2 = npq >> 1$, $\forall p$. DONC si $p << 1$ et $\mu = np >> 1$, alors la binomiale s'approxime à la fois par une Poisson et par la normale.

3.10.6 Méthodes de Monte-Carlo

Les méthodes de Monte-Carlo consistent en une évaluation numérique de grandeurs par des méthodes probabilistes. L'erreur est tolérée si elle est bornée. Pour une apporoche à n points, l'erreur décroit en $1/\sqrt{n}$.

Si on approche l'intégration de fonctions à une dimension par une méthode de Monte-Carlo, la méthode des rectangles, qui décroit en $1/n$ est plus efficace, mais pour une fonction à d dimensions, Monte-Carlo a toujours une erreur qui décroit en $1/\sqrt{n}$. Elle est donc plus efficace que la méode des rectangles lorsque $d > 2$.

3.11 Probabilités bayésiennes

3.11.1 Théorème de Bayes

Rappel des probabilités conditionnelles : $P(A|B) = P(A \cap B)/P(B)$

Le théorème de Bayes est

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{3.59}$$

avec $P(A)$ la probabilité a priori de l'événement A, et $P(A|B)$ la probabilité a posteriori, i.e. corrigée par la connaissance que B est réalisé.

Si $\Omega = A_1 \cup A_2 \cup ... \cup A_k$, k événements disjoints, alors

- $P(B) = \sum_i P(A_i)P(B|A_i)$
- $P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_i P(A_i)P(B|A_i)} = \frac{P(A_i \cap B)}{P(B)}$

$$P(A) = P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B}) \tag{3.60}$$

3.11.2 Problème d'inférence

Soit une urne contenant des boules noires et blanches, dont une proportion p inconnue de boules blanches. Soit b le nombre de boules blanches lorsque l'on tire n boules en les replaçant. Calculer la probabilité que la suivante soit blanche :

Comme on n'a aucune informtion sur p, supposons que toutes les proportions sont équiprobables : mesure uniforme sur $[0, 1]$. On considère donc la valeur de p comme une variable aléatoire.

$\rightarrow$ Remarque : c'est comme si on avait 101 urnes de proportion $p = 0\%, 1\%, 2\%, ...$ et on choisit une urne uniformément au hasard.

Calculs :

- Evénement A_x a priori : $p \in [x, x + dx]$ ($P(A) = dx$
- Evénement B : b boules blanches parmi n
- $P(B|A_x) = \binom{n}{b} x^b(1-x)^{n-b}$
- $P(B) = \int_0^1 \binom{n}{b} x^b(1-x)^{n-b}dx$

La probabilité a posteriori est alors

$$P(A_x|B) = \frac{\binom{n}{b} x^b (1-x)^{n-b} dx}{\int_0^1 \binom{n}{b} x^b (1-x)^{n-b} dx} \tag{3.61}$$

Règle de succession de Laplace :

La règle de succession de Laplace pour estimer la probabilité de succès d'un schéma de Bernoulli avec b succès observés sur n essais est

Sachant A_x, la probabilité que la prochaine boule soit blanche est x.
Sachant B, la probabilité que la prochaine boule soit blanche est

$$\mathbb{E}(p|B) = \int pP\left(p \in [x, x+dx]|B\right) = \frac{\int_0^1 x \binom{n}{b} x^b (1-x)^{n-b} dx}{\int_0^1 \binom{n}{b} x^b (1-x)^{n-b} dx} = \frac{b+1}{n+2} \tag{3.62}$$

En comparaison, un raisonnement fréquentiste propose une probabilité $\frac{b}{n}$, ce qui n'apporte pas de contradiction pour un n suffisamment grand.

Inférence statistique :
L'inférence statistique signifie estimer la mesure de probabilité d'un phénomène du monde réel à partir d'observations.

- Point de vue bayésien : On met une probabilité a priori sur tout ce qu'on ne connait pas, i.e. sur l'ensemble des mesures de probabilités possibles, et on la met à jour à chaque observation empirique, en utilisant le théorème de Bayes.
- On met une probabilité sur des événements qu'on peut observer de manière répétée et i.i.d., et on estime ces probabilités grâce à la loi des grands nombres et au théorème central limite.

3.12 Fonctions génératrices

3.12.1 Fonction génératrice

Soit $(f_n)_{n=0}^{\infty}$ une suite de réels. La série formelle

$$\boxed{f(x) = \sum_{n=0}^{\infty} f_n x^n} \tag{3.63}$$

est la fonction génératrice de cette suite.

Opérations :

- Somme : $s(x) = f(x) + g(x)$ telle que $s_n = f_n + g_n$

- Multiplication : $p(x) = f(x) \cdot g(x)$ telle que $p_n = \sum_{i=0}^{n} f_i \cdot g_{n-i}$

L'addition est commutative, associative, admet une neutre et est inversible. Elle forme donc un groupe. La multiplication est associative et admet un neutre. Elle forme donc un monoïde. De plus, elle se distribue sur l'addition. $\Longrightarrow$ Ces opérations forment un anneau commutatif.

Inversion :

Toutes les fonctions génératrices non nulles ne sont pas inversibles dans l'anneau des fonctions génératrices. Cet anneau n'est donc pas un corps.
Condition d'inversion : la fonction génératrice $f(x)$ est inversible ssi $f_0 \neq 0$.

$\rightarrow$ Remarque : $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$

Dérivation :

On définit l'opération de dérivation sur les fonctions génératrices comme

$$\boxed{f'(x) = \sum_{n=0}^{\infty} (n+1) f_{n+1} x^n} \tag{3.64}$$

- L'opération de multiplication est vérifiée : $(fg)' = f'g + fg'$
- Développement de Taylor : $f_n = \frac{1}{n!} \left. \frac{d^n f(x)}{dx^n} \right|_0$

3.12.2 Applications

$\rightarrow$ Remarque : $f(x) = \frac{x}{(1-x)^2}$ est la fonction génératrice des naturels. En effet, $f_n = n$ $\forall n \in \mathbb{N}$. Cette formule est indispensable dans ce chapitre!

Dénombrement :

On peut retrouver des fonctions de dénombrement vues précédemment en écrivant le problème sous la forme de fonctions génératrices.

Récurrence :

En remplaçant la formule de récurrence des éléments f_n dans la fonction génératrice, on peut retrouver les termes de la suite.

1. Transformer $f_k \rightarrow \sum_{n \geq 0} f_k x^n \rightarrow f(x) - f_0 - f_1 x - f_2 x^2 - ..., k-1$ termes venant de la CI.
2. isoler $f(x)$.
3. Décomposer en fonctions simples (inverses de premiers degrés).
4. $\sum f_n x^n = \sum (...)_1 x^n + \sum (...)_2 x^n$.
5. $f_n = (...)_1 + (...)_2$.

Fonction génératrice exponentielle :

Soit $(f_n)_{n=0}^{\infty}$ une suite de réels. La série formelle

$$f(x) = \sum_{n=0}^{\infty} \frac{f_n}{n!} x^n \tag{3.65}$$

est la fonction génératrice exponentielle de cette suite.

L'addition fonctionne toujours, mais le produit est modifié : $p = f \cdot g$ telle que $p_n = \sum_{k=0}^{n} B(n,k) f_k g_{n)k}$. La dérivée est par contre plus simple : $p = f'$ telle que $p_n = f_{n+1}$.

$$f_n = \frac{d^n f(x)}{dx^n} \tag{3.66}$$

quand f est la fonction génératrice exponentielle de $(f^n)_{n=0}^{\infty}$.

Fonction génératrice des moments :

Soit X une variable aléatoire. La fonction génératrice des moments est

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}[\sum_{n=0}^{\infty} \frac{1}{n!}(tX)^n] = \sum_{n=0}^{\infty} \frac{\mathbb{E}[X^n]}{n!} t^n \tag{3.67}$$

où $\mathbb{E}[X^n]$ est le moment d'ordre n de X.

Si X et Y sont indépendantes, pour un $c \in \mathbb{R}$:

- $M_{X+Y}(t) = M_X(t) MY_(t)$
- $M_{cX}(t) = M_X(ct)$

Lien avec les sections précédentes :

- Soit $X \sim$ Bernoulli(p) : $M_x(t) = 1 + (e^t - 1)p$.

- Soit $Y \sim Bin(n,p)$: $M_Y(t) = (1 + (e^t - 1)p)^n = e^{\mu(e^t - 1)}$, pour $\mu = np, n \to \infty, p \to 0$.

- Soit $Z \sim Po(\mu)$: $M_Z(t) = e^{\mu(e^t - 1)}$.

- Soit $X \sim N(\mu, \sigma^2)$: $M_X(t) = e^{t\mu + t^2\sigma^2/2}$.

3.13 Fonctions de hachage

Soit $H : \{0,1\}^* \to \{0,1\}^n$ la fonction de hachage. L'exposant * signifie que le mot binaire d'entrée est de longueur arbitraire, mais finie. La fonction de hachage envoie une chaîne de bits de n'importe quelle longueur vers une chaine de bits de longueur fixée n. Elle se comporte comme une fonction aléatoire, mais elle est déterministe.

3.13.1 Collisions

Soit $H : \{0,1\}^* \to \{0,1\}^n$ modélisée par une fonction aléatoire. Il est nécessaire d'évaluer H $2^{n/2}$ fois pour trouver une collision, i.e. $x_0, x_1 : H(x_0) = H(x,1), x_0 \neq x_1$.

Si on cherche $P(\text{collisions} > 1) \approx 1 - e^{-\frac{k(k-1)}{2*2^n}}$ = 1/2, l'exponentielle vaut $1/2$, l'exposant vaut donc $\log(2)$ et donc $k = 1.2 * 2^{n/2}$.

Conséquence, si $n = 256$ (comme souvent), il est pratiquement impossible de trouver une collision.

Conclusion :

Les fonctions de hachage donnent une emprunte digitale unique sur 32 bytes pour un nombre pratiquement illimité de documents.

3.13.2 Génération de nombres aléatoires

Une fonction de hachaqge permet de générer des séquences de bits pseudoaléatoires de longueur quelconque :

1. Choisir une graine (seed) s.
2. Sortir $H(s,0), H(s,1), ...$

$\to$ $H(s,i)$ est la fonction de hachage sur la concaténation de s et i.

Echantillonnage de distributions :

Supposons que l'on dispose de bits aléatoires : $X_i \sim \text{Ber}(1/2)$.
Comment piocher un entier uniformément $[0,b[$? On choisit n tel que $2^{n-1} \geq b < 2^n$.

Méthode 1 :

- Piocher $n\,bits \sim \text{Ber}(1/2)$, en faire un entier $X : 0 \leq X < 2^n$. Si $X < b$, sortir X, sinon recommencer.
- X est uniformément distribué sur $[0,2^n[$, et l'est donc aussi sur $[0,b[$.
- $P(X < b) = \frac{b}{2^n} \leq 1/2$ à chaque itération $\Rightarrow P(i \text{ itérations échouent}) \geq 2^{-i}$.
- En moyenne, $\frac{2^n}{b} n \geq 2n$ bits aléatoires nécessaires.

Méthode 2 :

- Piocher $n+m$ bits $\sim \text{Ber}(1/2)$, en faire un entier $X : 0 \leq X < 2^{n+m}$. Sortir $Y = X \mod b$.
- Soit $2^{n+m} = qb + r$ avec $r < b$. On a $q > 2^m$ puisque $b < 2^n$. $P(Y < 2^{n+m} \; mod \; b) = \frac{q+1}{2^{n+m}}$ et $P(2^{n+m} \; mod \; b \leq Y < b) = \frac{q}{2^{n+m}}$.
- $n+m$ bits aléatoires utilisés garantissent $\frac{P(Y=x_0)}{P(Y=x_1)} \leq 1 + 2^{-m} \forall x_0, x_1$.

Comment piocher uniformément une permutation de $0, 1, ..., n-1$?

Dans une urne contenant n boules, en les piochant une à une et en mettant la $i$ème boule piochée en position $n-i-1$ de la permutation en cours de formation.

- Mélange de Fisher-Yates :

$a = [0, 1, 2, ..., n-1]$

for j in range $(n-1, 0, -1)$:

Pick random $i \in [0, j]$

Swap $a[i]$ and $a[j]$

return a

Chacune des $n!$ permutations est le résultat d'une unique séquence de choix. On a besoin de $n-1$ entiers uniformément distribués sur des intervalles quelconques.

Conclusion :

Possibilité d'échantilloner, de manière aussi approchée que souhaité, n'importe quelle distribution sur base de bits uniformément aléatoires.

Démonstrations

4.1 Démonstration 1

Il existe une bijection entre aH et H :
Soit $f : H \to aH$ tel que $f(x) = ax$

- f est injective : vu que a est inversible, $ax_1 = ax_2 \Longrightarrow x_1 = x_2$
- f est surjective : par définition de aH, tout $y \in aH$ est dans l'image de f.

$\Rightarrow$ f est bijective.

4.2 Démonstration 2

Les classes latérales distinctes modulo H forment une partition de G.
Soient $a_1H, a_2H, \ldots$ les classes latérales modulo H.

- Tout élément de G est dans une classe latérale. Si $g \in G$, alors $g \in gH$ puisque $\epsilon \in H$.
- Les classes latérales aH et bH sont disjointes ou identiques : soient $a, b \in G$ et $x \in bH$.

On montre que $x \in aH \Longleftrightarrow b^{-1}a \in H$:

$\Longrightarrow$ $x = ah_1 = bh_2$ avec $h_1, h_2 \in H$. On a $ah_1 = bh_2 \Longrightarrow a = bh_2h_1^{-1} \Longrightarrow b^{-1}a = h_2h_1^{-1} \in H$

$\Longleftarrow$ $b^{-1}a \in H$ et $x = bh_2$. On a $b^{-1}a \in H \Longrightarrow a^{-1}b \in H$ et $x = bh_1 = (aa^{-1})\,bh_1 = a\left((a^{-1}b)\,h_1\right) \in aH$

4.3 Démonstration 3

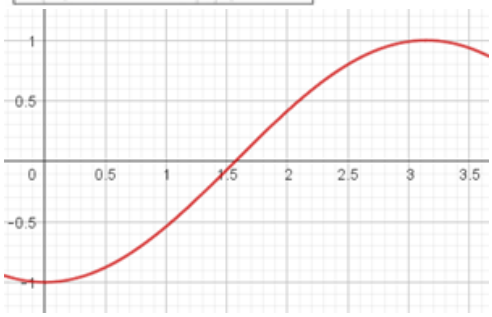
Si $[0, 1[\not\approx \mathbb{N}$, $\mathbb{R} \not\approx \mathbb{N}$.
Imaginons une bijection $f : \mathbb{N}^{\geq 1} \to [0, 1[$.

- Soit $f(i) = a_i = 0.a_{i1}a_{i2}a_{i3}\ldots$
- On construit $b = 0.b_1b_2b_3\ldots$ ne pouvant se trouver dans Im (f).
- Soit $b_i = \begin{cases} 1 \; si \; a_{ii} = 0 \\ 0 \; sinon \end{cases}$ On veille à ce que $a_{ii} \neq b_i$.

- $\nexists i \ : f(i) = b$: si $f(i) = b$, alors $a_i = b$ et en particulier $a_{ii} = b_i$

Ceci contredit notre hypothèse d'existence de f bijective[1] et on sait que $\mathbb{N} \approx \mathbb{N}^{\geq 1}$, donc $\mathbb{N} \not\approx [0, 1[$.

4.4 Démonstration 4



Soit x la coordonnée en x de la tête de l'aiguille, et θ l'angle formée par celle-ci avec l'horizontale. La coordonnée de la pointe de l'aiguille est $x + \cos\theta$. L'aiguille ne sera à cheval sur deux lattes que si $x + \cos\theta < 0$ pour $x > 0$. $\Longleftrightarrow x \leq -\cos\theta$. Cette condition est vérifiée sur l'intervalle en θ $\left[\frac{\pi}{2}, \pi\right]$. La surface sous cette courbe est une intégrale :

$$\int_{\pi/2}^{\pi} -\cos\theta d\theta = 1 \tag{4.1}$$

La surface des cas totaux est le rectangle de hauteur 1 et de longueur π. Sa surface vaut donc π.

On trouve finalement un probabilité pour l'aiguille de tomber à cheval sur deux lattes de $P = 1/\pi$.

[1] f est injective, mais pas surjective.