

UNIVERSITÉ LIBRE DE BRUXELLES

Faculté des Sciences

Département de Mathématique

BRUSSELS SUMMER OF
PROBABILITY AND STATISTICS

Ecole d'été organisée par CHRISTOPHE LEY et YVIK SWAN

Cours donnés par MAUDE GATHY, MAXIME GHEYSENS, ANN KIEFER,
CHRISTOPHE LEY, YVIK SWAN et GERMAIN VAN BEVER

Avec la participation exceptionnelle du Professeur FRANCIS BUEKENHOUT

Année académique 2008-2009

Août 2009

Table des matières

1	Avant-propos	2
2	Processus de Poisson	3
2.1	Définitions	3
2.2	Propriétés	5
2.3	Homogénéité	8
3	Profondeur de demi-espace	12
3.1	Introduction	12
3.2	Boules (de glace) et demi-espaces	14
3.3	Demi-espaces et profondeur	15
3.4	Pour aller plus loin	19
4	Le théorème du vote	19
4.1	Enoncé et démonstration	19
4.2	Application à la ruine d'une compagnie d'assurance	21
4.3	Application au théorème classique du vote	22
5	Le théorème de Gauss	23
5.1	Introduction	23
5.2	Loi gaussienne et maximum de vraisemblance	24
5.3	Preuve du théorème de Gauss	25
6	Méthode probabiliste en théorie des graphes	27
6.1	Généralités sur la méthode probabiliste	27
6.2	Modèles probabilistes en théorie des graphes	28
6.3	Nombres de Ramsey	31
6.4	Pour aller plus loin	33

1 Avant-propos

Le concept de “*mini-école d’été de probabilités*” est né, sous l’impulsion de Christophe Ley, durant l’été 2008. La première édition fut relativement confidentielle : deux enseignants (C. Ley et Y. Swan), et trois étudiants, tous issus de la première année du bachelor en sciences mathématiques de l’ULB. Malgré le nombre peu élevé de participants, notre première expérience se termina sur une conclusion encourageante : la joie de la découverte et du partage des mathématiques compense largement la perte de quelques jours de vacances au soleil.

Forts de ce premier succès, nous avons décidé de mettre en place une deuxième édition de notre école d’été. Avec l’ambition née de l’expérience, nous avons ouvert la porte à un plus large public, en invitant non seulement des étudiants du bachelor de l’ULB, mais aussi des étudiants en provenance du bachelor wiskunde de la VUB. (Nous avons même été rejoints par deux étudiants en fin de cycle du secondaire, plus motivés à l’idée d’en découdre avec des concepts mathématiques avancés que d’aller bronzer à la plage.) Afin de gérer au mieux ce nouvel arrivage d’étudiants – et satisfaire leur appétit insatiable pour les belles mathématiques – nous avons également renforcé l’équipe d’enseignement, en faisant appel à Maude Gathy, Maxime Gheysens, Ann Kiefer et Germain Van Bever. Leur enthousiasme et leur pédagogie ont comblé les esprits de nos apprenti-mathématiciens.

Enfin, on ne peut parler de cette deuxième édition de la *Brussels Summer School* sans mentionner le cadeau que nous a offert, par sa présence, le Professeur Francis Buekenhout. Par sa maîtrise et son talent, il nous a donné à tous, enseignants comme participants, une belle leçon de mathématiques *et* de pédagogie. En effet, il nous a démontré, en à peine plus de deux heures, que même les matières les plus ardues, techniques et absconses (la théorie des immeubles de Tits!!) peuvent être enseignées de façon agréable et compréhensible, sans pour autant voir leur sel dilué dans l’océan de la vulgarisation. Nous avons donc maintenant la preuve que «*a mathematician’s understanding of a theory is not to be considered complete until he has made it so clear that he can explain it to the first man whom he meets on the street*».

2 Processus de Poisson

Parmi les distributions de probabilité les plus connues, on trouve la *Poisson*, la *géométrique* et l'*uniforme*. Ces trois lois allient simplicité et élégance mathématique, tout en étant d'application dans de nombreux modèles. Bien qu'elles servent d'outils dans la description de phénomènes fort différents, nous allons maintenant voir que toutes les trois dérivent d'un même concept : le processus de Poisson.

2.1 DÉFINITIONS

Imaginons une expérience qui, à intervalles réguliers mais aléatoires, donne lieu à un événement. Supposons, pour simplifier, que le nombre d'occurrences par unité de temps – le taux – est constant. Supposons finalement que l'occurrence – ou non – d'un événement à un moment donné n'influe pas sur l'avenir du processus.

Exemple 2.1. *On peut penser, par exemple, aux visites sur une page web, au passage de voitures sur une portion d'autoroute, à la désintégration d'une particule radioactive, etc.*

Quel outil permettra de décrire tout l'historique d'un tel processus ? Quelle information permettra de reconstruire l'entièreté du déroulement de l'expérience ? La réponse à cette question est triple : on peut considérer soit les instants auxquels les événements ont lieu, soit le temps écoulé entre deux occurrences successives, soit le nombre d'occurrences ayant lieu dans n'importe quel intervalle de temps. Ces trois concepts résument clairement, chacun à sa manière, les propriétés du processus. De plus, un peu de réflexion nous aide rapidement à conclure que la donnée de n'importe quel des trois permet de déterminer les deux autres.

Il nous faut donc définir (i) une suite de variables aléatoires T_1, T_2, \dots enregistrant les instants auxquels les événements successifs ont lieu, (ii) une suite de variables aléatoires W_1, W_2, \dots enregistrant le temps écoulé entre deux événements consécutifs et (iii) une famille de variables aléatoires $N(t)$ enregistrant, pour chaque instant $t > 0$, le nombre d'événements ayant eu lieu durant les t premières unités de temps. Remarquons que, pour tout $k \geq 1$, $W_k = T_k - T_{k-1}$ (avec $T_0 = 0$) et que $N(t) = k$ (presque sûrement) lorsque $T_k = t$: la donnée de la suite T_k permet donc de reconstruire les W_k et la famille $N(t)$; il en va bien évidemment de même dans chacun des deux autres sens.

Une fois dégagées les variables aléatoires qui gouvernent le processus, il nous faut compléter notre modèle en choisissant une distribution de probabilité pour chacune des quantités d'entre elles. La seule obligation figurant dans notre cahier des charges est de choisir ces distributions de façon à obtenir un modèle *réaliste* et *calculable*. Faut-il pour cela choisir les T_k i.i.d. Gaussiens, exponentiels, ou autre chose ? Faut-il

prendre les W_k i.i.d. de loi Gamma ? Faut-il forcer la famille des $N(t)$ à être de même distribution binomiale pour tout t ? Chaque choix possible donnera, bien entendu, naissance à un processus aux propriétés différentes, certains étant plus intéressants que d'autres. Cependant, réflexion faite, il existe deux distributions qui s'imposent au mathématicien : l'exponentielle pour W_k et la Poisson pour $N(t)$.

Pourquoi ? Pour $N(t)$, l'explication est claire : on veut une loi de dénombrement ($N(t)$ compte le nombre d'occurrences) non bornée (on n'a pas imposé de nombre maximum) : on n'a donc qu'une seule loi en tête, la loi de Poisson. Le choix d'une loi exponentielle pour les W_k repose sur une argumentation un peu plus subtile : on veut un processus qui, après chaque occurrence, recommence "comme neuf" afin d'être homogène dans le temps *et* indépendant du passé. Il faut donc que la quantité de temps entre deux arrivées successives ne dépende pas de l'historique du processus : il nous faut une loi sans mémoire. Par conséquent, il faut que les W_k soient i.i.d. de loi exponentielle. Ce choix fixe, bien entendu, la loi des T_k qui ne sont maintenant plus qu'une somme de k exponentielles i.i.d. : les T_k suivent une loi Erlang.

Nous allons maintenant voir qu'outre leur vraisemblance par rapport au modèle que nous cherchons à décrire, ce choix de distributions allie aux deux qualités susmentionnées une troisième beaucoup plus importante à nos yeux : l'élégance.

Commençons par quelques définitions.

Définition 2.1. Soit (Ω, \mathcal{A}, P) un espace probabilisé. Soit I un ensemble dénombrable. Un processus stochastique en temps continu $\{X_t, t \geq 0\}$ à valeurs dans I est une famille de variables aléatoires $X_t : \Omega \rightarrow I$.

Définition 2.2. Un processus sera dit continu à droite (on note c.a.d.) si, pour tout $\omega \in \Omega$ et pour tout $t \geq 0$, il existe $\epsilon > 0$ tel que

$$X_s(\omega) = X_t(\omega) \text{ pour tout } t \leq s \leq t + \epsilon.$$

Remarque : L'absence de l'accent grave sur l'abréviation c.a.d. est destinée à faire une distinction entre « continu à droite » et « c'est-à-dire », mot non-mathématique habituellement abrégé c.à.d.

Définition 2.3. On associe à tout processus c.a.d. à espace d'états dénombrable les deux processus suivants :

- une suite d'instant d'arrivées $0 = T_0 < T_1 < T_2 < \dots < T_j < \dots$, où $\{T_j = s\}$ signifie que la j ème occurrence a eu lieu à l'instant s .
- un processus de comptage $\{N(t), t \geq 0\}$ qui, comme son nom l'indique, compte le nombre d'occurrences après l'instant 0 et jusqu'à l'instant t .

Remarquez que le processus de comptage $N(t)$ est donc lui-même un processus c.a.d. à valeurs dans \mathbb{N} tel que $N(0) = 0$. De plus, pour tout ω , la fonction $N(\cdot)(\omega) : \mathbb{R} \rightarrow \mathbb{N}$ est croissante surjective.

Définition 2.4. *Un processus stochastique continu $\{X_t, t \geq 0\}$ est dit à incréments indépendants si, pour toute suite d'intervalles disjoints $(a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)$, les variables aléatoires*

$$X_{b_1} - X_{a_1}, X_{b_2} - X_{a_2}, \dots, X_{b_n} - X_{a_n}$$

sont indépendantes.

En particulier, un processus de comptage est à incréments indépendants si, pour tout s, t , les variables $N(t)$ et $N(t+s) - N(t)$ sont indépendantes, i.e. le nombre d'événements ayant eu lieu durant les t premières unités de temps n'influence pas le nombre d'événements qui aura lieu dans les s unités de temps suivantes.

Définition 2.5. *Un processus stochastique continu $\{X_t, t \geq 0\}$ est dit à incréments stationnaires si $X_{t+h} - X_s$ a la même distribution que $X_t - X_s$ pour tout $s < t$ et pour tout $h > 0$.*

En particulier, un processus de comptage est à incréments stationnaires si on a $N(t+s) - N(t) \stackrel{\mathcal{L}}{=} N(s)$, i.e. la loi du nombre d'événements ne dépend que de la longueur de l'intervalle de temps considéré.

Nous sommes maintenant en mesure de définir le concept de *processus de Poisson*.

Définition 2.6. *Un processus de Poisson homogène de taux λ est un processus c.a.d. à incréments indépendants et stationnaires tel que (i) $P[N(h) = 0] = 1 - \lambda h + o(h)$ et (ii) $P[N(h) = 1] = \lambda h + o(h)$, pour un certain $\lambda > 0$.*

Notons l'extrême dépouillement de cette définition qui ne caractérise que le comportement *infinitésimal* du processus, en demandant (i) que la probabilité d'une occurrence dans un petit intervalle de temps soit proportionnelle à la longueur de cet intervalle et (ii) que la probabilité qu'il y ait deux occurrences dans un petit intervalle de temps soit négligeable. Il est remarquable que ceci suffise pour garantir un processus possédant toutes les qualités annoncées plus haut. Comment une telle prouesse est-elle possible? C'est ce que nous allons nous atteler à découvrir dans la section suivante.

2.2 PROPRIÉTÉS

Explorons maintenant les conséquences de la Définition 2.6.

Théorème 2.1. *Soit $\{N(t), t \geq 0\}$ un processus de Poisson. Alors, pour tout $s < t$, le nombre d'occurrences dans l'intervalle de temps $[s, t]$ suit une loi de Poisson de paramètre $\lambda(t-s)$. On a, en particulier,*

$$P[N(t) = k] = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k = 0, 1, 2, \dots \quad (1)$$

Avant de prouver ce théorème, commençons par un argument simple qui permet de comprendre pourquoi la loi de Poisson est la seule à pouvoir satisfaire nos exigences. En fait, ceci n'a (presque) rien à voir avec l'habituelle explication "la Poisson est une variable aléatoire adaptée au dénombrement de bla bla", mais découle tout simplement de ce que la *fonction* $x \mapsto e^x$ est la seule compatible avec l'hypothèse d'indépendance et de stationnarité!

En effet, par indépendance, nous obtenons

$$P[N(t+s) = 0] = P[N(s) = 0] P[N(t+s) - N(s) = 0],$$

de même que par stationnarité, nous savons que

$$P[N(t+s) - N(s) = 0] = P[N(t) = 0].$$

Par conséquent, la fonction $f(t) = P[N(t) = 0]$ doit satisfaire à l'équation fonctionnelle $f(t+s) = f(t)f(s)$ dont les seules solutions sont $f(t) \equiv 0$ et $f(t) = e^{-\gamma t}$ pour un certain $\gamma \in \mathbb{R}$ (cf Section 5 pour une résolution). La fonction $f(t)$ n'étant pas identiquement nulle, la troisième contrainte de la définition nous permet de conclure que $\gamma = \lambda$ et donc

$$P[N(t) = 0] = e^{-\lambda t}.$$

On conclut donc à la *nécessité* d'une loi de type Poisson, et un peu de réflexion permet de voir que l'implication inverse est valide également!

Preuve. Soit $t > 0$. Définissons, pour tout $n \geq 0$,

$$P_n(t) = P[N(t) = n].$$

Affirmation 1 : $P_0(t) = e^{-\lambda t}$.

Preuve : On a

$$P_0(t+h) = P[N(t+h) = 0] = P[N(t+h) - N(t) = 0] P_0(t) = P[N(h) = 0] P_0(t),$$

et donc

$$P_0(t+h) = P_0(t)(1 - \lambda h + o(h)).$$

On en déduit

$$\frac{P_0(t+h) - P_0(t)}{h} = -\lambda P_0(t) + \frac{o(h)}{h},$$

et, en laissant h tendre vers 0 dans cette dernière équation, on obtient l'EDO

$$P_0'(t) = -\lambda P_0(t).$$

Etant donnée la condition $P_0(0) = 1$, cette dernière se résout facilement pour donner

$$P_0(t) = e^{-\lambda t}.$$

Affirmation 2 : $P_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$.

Preuve : On a

$$P_n(t+h) = P_n(t)(1-\lambda h) + \lambda h P_{n-1}(t) + o(h).$$

Donc

$$\frac{P_n(t+h) - P_n(t)}{h} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(h)}{h},$$

ce qui, après avoir pris la limite pour $h \rightarrow 0$, nous donne

$$P'_n(t) = -\lambda P_n(t) + \lambda P_{n-1}(t),$$

avec comme condition de bord $P_n(0) = 0$ pour tout $n \geq 1$. Une récurrence sur n permet de résoudre ces équations pour obtenir (1).

Affirmation 3 : Conclusion.

Preuve : Notons Y le nombre d'occurrences dans l'intervalle $[s, t]$. Etant donné que

$$P[Y = k] = P[N(t-s) = k] = P_k(t-s),$$

on obtient le résultat. ■

Remarque. Il y a en moyenne λt occurrences dans un intervalle de temps de longueur t . Il y a donc, en moyenne, λ occurrences par unité de temps. Ceci explique pourquoi on parle de processus de Poisson de *taux* λ .

Théorème 2.2. Soit $\{N(t), t \geq 0\}$ un processus de Poisson de taux $\lambda > 0$. Soient $0 = T_0 < T_1 < T_2 < \dots$ les instants d'arrivées, et notons pour $k \geq 1$

$$W_k = T_k - T_{k-1}$$

les longueurs des intervalles de temps entre deux arrivées successives. Alors

- (i) les variables aléatoires W_k sont mutuellement indépendantes, identiquement distribuées et de loi exponentielle de moyenne $1/\lambda$;
- (ii) les variables aléatoires T_k suivent une loi Erlang de paramètres (λ, k) .

Preuve : Etant donné que les incréments du processus sont indépendants, tout événement se déroulant après l'instant T_n est indépendant de T_i pour tout $i = 1, 2, \dots, n-1$. Ceci permet de prouver l'indépendance mutuelle des W_k . De plus, pour tout $s \geq 0$ et pour tout $n \geq 1$, on a

$$P[W_n > s] = P[N(T_{n-1} + s) - N(T_{n-1}) = 0] = P[N(s) = 0] = e^{-\lambda s}.$$

La première affirmation découle de cette dernière égalité ; la deuxième est une conséquence immédiate de la première. ■

De façon générale, on peut prouver les équivalences suivantes.

Théorème 2.3. *Soit $\{X_t, t \geq 0\}$ un processus croissant, continu à droite et à valeurs dans \mathbb{N} tel que $X_0 = 0$. Soient $0 = T_0 < T_1 < T_2 < \dots$ les instants où le processus change d'état et soient les variables aléatoires $W_k = T_k - T_{k-1}$ pour tout $k \geq 1$. Alors les trois propriétés suivantes sont équivalentes :*

- (a) *les longueurs W_1, W_2, \dots sont indépendantes de loi exponentielle de paramètre λ et $X_{T_n} = n$ pour tout n ;*
 (b) *les incréments de $\{X_t\}$ sont indépendants et*

$$P[X_{t+h} - X_t = 0] = 1 - \lambda h + o(h), \quad P[X_{t+h} - X_t = 1] = \lambda h + o(h);$$

- (c) *les incréments de $\{X_t\}$ sont stationnaires et indépendants et, pour tout t , X_t suit une loi de Poisson de paramètre λt .*

Tout processus continu satisfaisant à l'une des trois propriétés ci-dessus est appelé Processus de Poisson.

2.3 HOMOGÉNÉITÉ

En introduisant le processus de Poisson, nous avons annoncé, de façon peut-être un peu effrontée, qu'il représentait un sommet d'élégance mathématique. L'objectif de cette section est de convaincre notre cher lecteur de l'exactitude de cette appréciation.

Théorème 2.4. *Soit $\{N(t), t \geq 0\}$ un processus de Poisson de taux λ , et soient T_1, T_2, \dots les instants de saut du processus. Pour tout $s, t \geq 0$, conditionnellement à l'événement $\{N(t+s) - N(s) = 1\}$, l'instant auquel cet unique saut a lieu est distribué uniformément sur l'intervalle $[s, s+t]$.*

Ce théorème implique, en particulier, que sachant qu'un événement aura lieu dans un intervalle de temps donné, l'instant auquel cet événement se manifestera est uniformément distribué sur cet intervalle de temps.

Preuve : Remarquons simplement que

$$P[N(t+u) - N(t) = 0 \mid N(t+s) - N(t) = 1] = P[N(u) = 0 \mid N(s) = 1].$$

Un calcul standard (qui demande tout de même un minimum d'attention) permet alors de déduire le résultat. ■

Théorème 2.5. Soient U_1, U_2, \dots des variables aléatoires i.i.d. de loi uniforme sur l'intervalle $[0, t]$. Alors, conditionnellement à l'événement $\{N(t) = k\}$, on a

$$T_1 \stackrel{\mathcal{L}}{=} U_{(1)}$$

où $U_{(1)} = \min\{U_1, U_2, \dots, U_k\}$.

Cette affirmation découle d'arguments similaires à ceux qui ont permis d'établir le théorème précédent – sa preuve est laissée comme exercice. De façon plus générale on prouve le résultat suivant.

Théorème 2.6. Soient U_1, U_2, \dots des variables aléatoires i.i.d. de loi uniforme sur $[0, t]$. Notons $U_{(j)}$ la j ème statistique d'ordre. Alors, conditionnellement à l'événement $\{N(t) = k\}$, on a

$$(T_1, T_2, \dots, T_k) \stackrel{\mathcal{L}}{=} (U_{(1)}, U_{(2)}, \dots, U_{(k)}).$$

La preuve de ce résultat se fait via un argument de localisation : le comportement à long terme du processus étant complexe, on se simplifie la vie en ne regardant que ce qui se passe durant un court intervalle de temps.

Preuve : Il nous faut prouver que, conditionnellement à l'événement $\{N(t) = k\}$, la densité jointe des variables aléatoires T_1, \dots, T_k est donnée par

$$f(x_1, x_2, \dots, x_k) = \frac{k!}{t^n} \mathbf{1}_{\{0 < x_1 < x_2 < \dots < x_k < t\}}.$$

Remarquons d'abord que

$$f(x_1, x_2, \dots, x_k) = 0$$

dès que $\mathbf{1}_{\{0 < x_1 < x_2 < \dots < x_k < t\}} = 0$ – les deux fonctions sont donc égales lorsque la condition de l'indicatrice n'est pas satisfaite.

Considérons maintenant un k -uplet de réels n'annulant pas l'indicatrice, et prenons $\delta > 0$. Notons A_δ l'événement $\{x_1 < T_1 < x_1 + \delta, x_2 < T_2 < x_2 + \delta, \dots, x_k < T_k < x_k + \delta\}$. On a alors

$$\begin{aligned} \mathbb{P}[A_\delta \mid N(t) = k] &= \mathbb{P}[\forall j \in \{1, \dots, k\}, T_j \in [x_j, x_j + \delta] \mid N(t) = k] \\ &= \frac{\mathbb{P}[N(t) = k \cap \forall j \in \{1, \dots, k\}, T_j \in [x_j, x_j + \delta]]}{\mathbb{P}[N(t) = k]}. \end{aligned}$$

En faisant appel à l'indépendance des incréments, on voit (un petit dessin pourrait s'avérer utile) que le numérateur de cette dernière expression se factorise en

$$\mathbb{P}[N(x_1) = 0] \mathbb{P}[N(x_1 + \delta) - N(x_1) = 1] \cdots \mathbb{P}[N(t) - N(x_k + \delta) = 0].$$

La stationnarité des incréments permet de calculer cette quantité aisément, ce qui nous donne

$$\mathbb{P}[N(t) = k \cap \forall j \in \{1, \dots, k\}, T_j \in [x_j, x_j + \delta]] = (\lambda\delta)^k e^{-\lambda t},$$

et donc

$$\mathbb{P}[A_\delta | N(t) = k] = \frac{\delta^k k!}{t^k}.$$

En prenant la limite pour $\delta \rightarrow 0$, on obtient finalement

$$f(x_1, x_2, \dots, x_k) = \lim_{\delta \rightarrow 0} \frac{\mathbb{P}[A_\delta | N(t) = k]}{\delta^k} = \frac{k!}{t^k}.$$

Le théorème est donc prouvé. ■

Nous concluons cette section par la résolution de l'exercice suivant.

Problème : *Des voitures entrent dans un tronçon d'autoroute de longueur a suivant un processus de Poisson de taux λ . En supposant que la i ème voiture roule à vitesse constante V_i , où $(V_i)_{i \geq 1}$ est une suite de variables aléatoires indépendantes et identiquement distribuées, quelle est la distribution du nombre de voitures présentes dans ce tronçon à l'instant t ?*

Malgré sa trivialité apparente, cette question est en réalité une version particulière d'un problème beaucoup plus riche, aux ramifications profondes et multidisciplinaires, à savoir

A quel point peut-on changer les différentes hypothèses apparaissant dans la définition du processus de Poisson tout en conservant des propriétés intéressantes ?

Répondre à ce problème ouvre la voie à l'étude des processus Markoviens, du mouvement Brownien, des processus de Lévy et à bien d'autres réjouissances que nous n'aurons malheureusement pas le temps d'aborder ici. Nous nous contenterons seulement de développer les outils nécessaires à la résolution du problème.

Introduisons le processus

$$N_a(t) = \sum_{i=1}^{N(t)} \mathbf{1}_{\{V_i(t-T_i) < a\}},$$

qui compte le nombre de voitures présentes dans le tronçon à l'instant t (on pose, bien entendu, $N_a(t) = 0$ lorsque $N(t) = 0$). Il nous faut calculer $\mathbb{P}[N_a(t) = k]$, pour $t > 0$ et $k \in \mathbb{N}$.

Soit F la fonction de répartition commune des V_i . Remarquons, dans un premier temps, que si l'on remplace les T_i par des uniformes, la question se simplifie fortement.

En effet, en notant U_i , $i \geq 1$, des variables aléatoires i.i.d. de loi uniforme sur $[0, t]$, on a, pour tout $k \geq 1$,

$$\mathbb{P} \left[\sum_{i=1}^{N(t)} 1_{\{V_i(t-U_i) < a\}} = k \right] = \sum_{n=1}^{\infty} \mathbb{P} \left[\sum_{i=1}^n 1_{\{V_i(t-U_i) < a\}} = k \right] \mathbb{P}[N(t) = n].$$

Or, clairement,

$$\mathbb{P} \left[\sum_{i=1}^n 1_{\{V_i(t-U_i) < a\}} = k \right] = \binom{n}{k} A_t^k (1 - A_t)^{n-k},$$

avec

$$A_t = \frac{1}{t} \int_0^t F(a/u) du.$$

Une simple application des règles de calcul habituelles nous donne alors

$$\mathbb{P} \left[\sum_{i=1}^{N(t)} 1_{\{V_i(t-U_i) < a\}} = k \right] = e^{-\lambda A_t} \frac{(\lambda A_t)^k}{k!}.$$

On en déduit que la version simplifiée du processus $N_a(t)$ est encore un processus de Poisson, bien que son taux ne soit plus constant mais dépendant du temps.

Qu'en est-il pour notre processus originel ? Ce problème est réglé par le (très joli) résultat suivant.

Lemme 2.1. *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d., indépendantes des temps d'arrivées $(T_i)_{i \geq 1}$ d'un processus de Poisson de taux λ . Alors, pour toute fonction $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, on a*

$$S(t) = \sum_{i=1}^{N(t)} g(T_i, X_i) \stackrel{\mathcal{L}}{=} \sum_{i=1}^{N(t)} g(tU_i, X_i), \quad (2)$$

avec $(U_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. de loi uniforme sur $[0, 1]$, indépendantes des X_i et des T_i .

Preuve : Grâce au Théorème 2.6, nous savons que, pour tout $n \in \mathbb{N}$ et pour tout $x \in \mathbb{R}$,

$$\mathbb{P} \left[\sum_{i=1}^{N(t)} g(T_i, X_i) \leq x \mid N(t) = n \right] = \mathbb{P} \left[\sum_{i=1}^n g(tU_{(i)}, X_i) \leq x \right]$$

avec $U_{(i)}$ la i ème statistique d'ordre de l'échantillon (U_1, \dots, U_n) . Etant donné qu'on somme sur tous les indices, on peut alors utiliser le fait que les X_i sont i.i.d. pour écrire

$$\mathbb{P} \left[\sum_{i=1}^n g(tU_{(i)}, X_i) \leq x \right] = \mathbb{P} \left[\sum_{i=1}^n g(tU_{(i)}, X_{\pi(i)}) \leq x \right], \quad (3)$$

cette égalité étant valide pour toute permutation $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$. La classification par ordre croissant des U_i n'étant en fait qu'une permutation particulière, nous pouvons appliquer (3) à la permutation π_0 définie par $U_{(i)} = U_{\pi_0(i)}$ pour obtenir

$$\mathbb{P} \left[\sum_{i=1}^n g(tU_{(i)}, X_i) \leq x \right] = \mathbb{P} \left[\sum_{i=1}^n g(tU_{\pi_0(i)}, X_{\pi_0(i)}) \leq x \right] = \mathbb{P} \left[\sum_{i=1}^n g(tU_i, X_i) \leq x \right].$$

(Pour justifier cette dernière rigoureusement, il faudrait faire appel à la notion d'*espérance conditionnelle*, un concept qu'il serait beaucoup trop long d'établir dans le cadre de ces notes. Nous vous demandons donc de nous faire confiance.) L'affirmation découle de cette dernière identité. ■

Nous concluons cette discussion en remarquant que (2) implique, en particulier, l'égalité en loi

$$N_a(t) \stackrel{\mathcal{L}}{=} \sum_{i=1}^{N(t)} 1_{\{V_i(t-U_i) < a\}}.$$

On a, par conséquent, résolu le problème posé au début de cette section.

Solution : *Le processus $N_a(t)$ est un processus de Poisson, de taux λA_t .*

3 Profondeur de demi-espace

3.1 INTRODUCTION

L'apprenti-étudiant attentif et passionné sait que tout ce qui sort de la bouche des assistants n'est que vérité. Cette vision réductrice ne tient bien sûr pas compte des réponses approximatives ou affirmations floues de ces derniers. Il est cependant certaines vérités auxquelles il n'est possible que d'adhérer. Un exemple de telle lapalissade est : "*La médiane, c'est cool!*". Elle est, en effet, facile à calculer, à interpréter et surtout robuste comme un cheval de trait (l'adjectif "robuste" signifie bien sûr peu sensible à des variations des observations et à la présence de points aberrants). Elle présente donc de nombreux attraits, dont la portée ne dépasse pas l'entendement du passant lambda. Il y a toutefois un problème majeur qui se dresse sur le chemin du passant mu (mathématicien cette fois) mû par l'attrait de la nouveauté : comment peut-on définir une médiane dans le cas de données multidimensionnelles ?

Un réflexe naturel en cas de plongée dans le monde froid de la nouveauté est de se raccrocher à ce que l'on connaît. On pourrait donc être tenté, dans un premier temps, d'adapter la définition de la médiane comme préimage de $1/2$ ($F^{-1}(1/2)$) pour la fonction de distribution en l'étendant au cas multidimensionnel. On voit rapidement que ce n'est pas une bonne idée : la médiane ainsi choisie n'est plus

unique et plutôt mal définie! Pour s'en convaincre, il suffit de considérer l'ensemble $E = \{(-1, -1), (1, 1)\}$, pour lequel tout point du type (x, y) avec $-1 < x < 1$ et $-1 < y < 1$ est tel que $F(x, y) = 1/2$: ce simple exemple suffit donc pour faire capoter cette première approche.

Une autre généralisation "naturelle" serait de définir la médiane composante par composante, comme l'a fait Hayford en 1902, i.e. on note

$$\text{Med}_C(X_1, \dots, X_n) = (\text{Med}(X_{11}, \dots, X_{n1}), \dots, \text{Med}(X_{1p}, \dots, X_{np}))',$$

la médiane composante par composante de $X_1, \dots, X_n \in \mathbb{R}^p$. A nouveau, un exemple suffira pour se convaincre que là non plus, on n'a pas trouvé la définition idéale. En effet, considérons l'ensemble

$$\{x_1, \dots, x_5\} = \{(1, 1); (0, 1); (-1, 1); (-1, -1); (1, -1)\},$$

et prenons la matrice

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Cette matrice est de déterminant non nul, et

$$\text{Med}_C(Ax_1, \dots, Ax_5) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 1 \end{pmatrix} = A(\text{Med}_C(x_1, \dots, x_5)).$$

Pire encore, si on prend la matrice de rotation

$$B = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix},$$

$$\text{Med}_C(Bx_1, \dots, Bx_5) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \neq \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} = B(\text{Med}_C(x_1, \dots, x_5)).$$

On voit donc qu'en définissant une médiane composante par composante, l'objet qui en résulte est *fortement dépendant du choix du système de coordonnées*! La médiane étant un paramètre de position, il paraît pourtant naturel de demander que tout changement de système de coordonnées opère sur la médiane le même changement que sur les observations elles-mêmes (comme c'est le cas pour la moyenne).

Après avoir tenté de généraliser la médiane en (i) inversant la fonction de répartition et (ii) inversant les fonctions de répartition des marginales, on se rend compte qu'il faut peut-être être plus subtil. Quelle autre propriété caractérise la médiane univariée? Le lecteur attentif se souviendra que la médiane univariée est l'unique point minimisant la quantité $\sum_i |X_i - \mu|$ (en μ)! Une troisième approche (due à Weber [11]) est par conséquent de définir la médiane de n observations X_1, \dots, X_n dans

\mathbb{R}^p comme le point $\hat{\mu}$ minimisant la quantité $\sum_i \|X_i - \mu\|$. Cette médiane a plusieurs avantages : en plus d'être bien définie, elle est unique *et* équivariante (dans le sens décrit ci-dessus) par transformation orthogonale (cf par exemple un des articles les plus accessibles et les plus courts de l'histoire des *Annals of Statistics*, [5]). Il semblerait donc qu'on se rapproche d'une définition intéressante. Malheureusement nos ardeurs sont rapidement calmées, car cette médiane n'est pas invariante par transformation affine, ce qui est, on l'a vu, très gênant...

Malgré que ces trois définitions ne soient pas utilisables, nos échecs ont tout de même permis de mettre en lumière une propriété naturelle qu'une médiane multivariée devrait posséder, à savoir l'équivariance.

Définition 3.1. Soient X_1, \dots, X_n des observations dans \mathbb{R}^p , A une matrice $p \times p$ inversible et b un vecteur fixé. Une médiane M fondée sur les observations X_1, \dots, X_n est affine-équivariante si $M(AX_1 + b, \dots, AX_n + b) = AM(X_1, \dots, X_n) + b$ (c'est-à-dire si elle se comporte comme elle devrait). Si A est orthogonale, on parlera d'équivariance orthogonale.

3.2 BOULES (DE GLACE) ET DEMI-ESPACES

C'est en 1929 qu'Hotelling (certes inconsciemment) introduit la propriété permettant une généralisation du concept de médiane au cadre multivarié. Il faudra attendre 1975 et Tukey pour une définition propre de médiane de demi-espaces. Cette propriété a, comme souvent en statistique, été motivée par un jeu économique, dont l'énoncé est le suivant.

Le problème du marchand de glaces Imaginons deux marchands de glaces (ou de votre friandise préférée) souhaitant chacun placer une échoppe le long d'une plage (idéalisée ici par une droite). Sur cette droite, les clients arrivent selon une loi de distribution F . On suppose que chaque client choisit l'échoppe la plus proche lors de son arrivée sur la plage. Si les deux échoppes sont à égale distance, il choisira au hasard. On supposera également que les emplacements des deux échoppes sont distincts. Le problème pour les vendeurs est alors de placer leur échoppe de manière à maximiser leur profit personnel. Le jeu se passera comme suit : A , sympathique bonhomme bonhomme (et oui c'est un marchand de glace) place son échoppe en premier, à un endroit x . Le vendeur B , marchand retors, véreux et intelligent place alors la sienne à un endroit y de manière à maximiser ses profits. Comment doit se placer A de manière à maximiser son profit tout en minimisant celui de B ?

Commençons par formaliser les concepts introduits dans ce problème. Notons, dans un premier temps, que si $x < y$ alors A obtient les clients arrivant dans l'inter-

valle $(-\infty, \frac{1}{2}(x+y))$ et partage les clients arrivant en $\frac{1}{2}(x+y)$. Les clients arrivant plus loin que $\frac{1}{2}(x+y)$ vont chez le marchand B . Le profit du vendeur A est alors

$$P_A = \frac{1}{2} \left[F\left(\frac{x+y}{2}\right) + F\left(\left(\frac{x+y}{2}\right)^-\right) \right].$$

Une formule similaire est valable si $x > y$.

Si A place son échoppe en x , le seul projet de B est alors de trouver l'emplacement optimal y afin de contrecarrer A . Il lui faut donc choisir y arbitrairement proche de x , d'un côté ou de l'autre, afin de maximiser son profit. L'infimum des revenus en x , qui maximise donc le profit de y est alors

$$\min [F(x), 1 - F(x^-)],$$

i.e. $F(x)$ si B se place à la droite de A , $1 - F(x^-)$ s'il se place à gauche. A cherche alors l'emplacement x maximisant cette quantité (celui-ci n'ayant aucun contrôle sur B , on suppose que ce dernier placera toujours son échoppe optimalement), ce qui nous mène à la médiane, vu que cette dernière quantité est effectivement maximisée en la médiane de la distribution.

On remarquera ici l'approche de la médiane en tant que solution d'un problème de "minimax" (on cherche à maximiser (en x) une expression qui est le minimum de plusieurs termes en x). La généralisation de cette approche au cas multivarié est assez immédiate (imaginez une plage de dimension 2 pour une bonne représentation) : si x est fixé et y est arbitrairement proche, le minimum des revenus en x est donné par le minimum des probabilités des demi-espaces contenant x . On cherche alors à maximiser cette quantité en fonction de l'emplacement x , et c'est ce faisant qu'on va définir ce qui s'avérera dans la suite être une "bonne" médiane multivariée.

3.3 DEMI-ESPACES ET PROFONDEUR

L'objectif de cette section est d'aider A à contrecarrer les plans de B , c'est-à-dire de lui montrer à A qu'il est possible pour lui de maximiser ses profits d'une certaine manière.

Commençons, pour ce faire, par quelques définitions. (Afin d'éclairer les différentes notations, il est fortement conseillé d'accompagner chacune des notions que nous allons maintenant introduire par un petit dessin bidimensionnel!) Définissons

$$H_{\theta,u} = \{x \in \mathbb{R}^p : u'(x - \theta) \geq 0\},$$

le demi-espace fermé, de bord orthogonal à u , passant par θ et contenant $\theta + u$ (afin de retrouver les notations usuelles, on place maintenant notre marchand A en le

point θ , B se trouvant alors dans la direction u depuis θ). Définissons la quantité d'intérêt (qui ne se limite alors plus à une distribution de proba mais à toute mesure sur \mathbb{R}^p) :

Définition 3.2. La *profondeur de demi-espace* de $\theta \in \mathbb{R}^p$ par rapport à μ est

$$D_H(\theta) = \inf_{u \in S^{p-1}} \mu(H_{\theta,u}).$$

De manière à être tout à fait général, cette notion est définie pour tout mesure sur \mathbb{R}^p . Il nous faut bien sûr avoir en tête le cas $\mu = \mathbb{P}$. L'objectif de cette introduction est de prouver le

Théorème 3.1. Pour un certain $\theta^* \in \mathbb{R}^p$, pour une mesure de probabilité \mathbb{P} ,

$$\sup_{\theta} D_H(\theta) = D_H(\theta^*).$$

Remarquons que ce théorème n'est en général pas valable pour des mesures quelconques. Afin de s'éviter une démonstration longue et pleine de subtilités, il va falloir ruser (l'apanage de tout mathématicien) et, comme souvent, jouer avec des ensembles. Définissons

Définition 3.3. La *zone de profondeur* α , notée D_α est

$$D_\alpha = \{\theta \in \mathbb{R}^p \mid D_H(\theta) \geq \alpha\}.$$

Nous nous proposons d'explorer quelques propriétés de la profondeur de demi-espace, et ce afin de démontrer le théorème 3.1.

Proposition 3.1. Pour tout $\alpha > 0$,

$$D_\alpha = \bigcap \{H \mid H \text{ est un demi-espace ouvert tel que } \mu(H^c) < \alpha\}.$$

Preuve: Clairement,

$$\begin{aligned} \theta \in D_\alpha &\Leftrightarrow \inf_{u \in S^{p-1}} \mu(H_{\theta,u}) \geq \alpha \\ &\Leftrightarrow \forall u \in S^{p-1}, \mu(H_{\theta,u}) \geq \alpha. \end{aligned}$$

Nous affirmons que ceci est équivalent à la condition

$$[H \text{ est un demi-espace ouvert tel que } \mu(H^c) < \alpha \Rightarrow \theta \in H].$$

En effet, soit H un demi-espace ouvert tel que $\mu(H^c) < \alpha$. Supposons que $\theta \notin H$. Alors on peut trouver un u tel que $H_{\theta,u} \subset H^c$ et donc $\mu(H_{\theta,u}) \leq \mu(H^c) < \alpha$, une

contradiction. Réciproquement, soit $u \in S^{p-1}$. Supposons que $\mu(H_{\theta,u}) < \alpha$. Posons $H = H_{\theta,u}^c$. H est ouvert, tel que $\mu(H^c) = \mu(H_{\theta,u}) < \alpha$ et tel que $\theta \notin H$. Une contradiction également.

Enfin, l'équivalence suivante est triviale.

$$\begin{aligned} & [H \text{ est demi-espace ouvert tel que } \mu(H^c) < \alpha \Rightarrow \theta \in H] \\ \Leftrightarrow & \theta \in \bigcap \{H \mid H \text{ est un demi-espace ouvert tel que } \mu(H^c) < \alpha\}. \end{aligned}$$

■

Corollaire 3.1. *Pour tout α ,*

$$D_\alpha \subset \bigcap \{H \mid H \text{ est un demi-espace fermé tel que } \mu(H^c) < \alpha\}.$$

En général, on ne peut pas prédire si D_α sera ouvert ou fermé comme en témoignent les exemples suivants : si $\mu = \delta_a$, pour $a \in \mathbb{R}^p$, la profondeur de a est 1, la profondeur de tout autre point est nulle. Ainsi, $D_{1/2} = \{a\}$ est fermé. D'autre part, prenons $p = 2$, $A = \mathbb{R} \times (-1, 1)$. Soit $\mu(B) = m_2(B \cap A)$, où m_2 désigne la mesure de Lebesgue dans \mathbb{R}^2 . La profondeur de θ est infinie si $\theta \in A$, 0 sinon. Ainsi $D_{1/2} = A$, un ouvert. Notre marchand de glace sera sauvé par la finitude des mesures de probabilité (exercice : essayez d'expliquer cela à votre fournisseur de crèmes glacées).

Dans le cas d'une mesure bornée (qui peut être ramenée à une mesure de probabilité trivialement), nous avons les propriétés suivantes.

Proposition 3.2. *Pour tout α ,*

$$D_\alpha = \bigcap \{H \mid H \text{ est un demi-espace fermé tel que } \mathbb{P}(H^c) < \alpha\}.$$

Preuve: L'inclusion

$$D_\alpha \subset \bigcap \{H \mid H \text{ est un demi-espace fermé tel que } \mathbb{P}(H^c) < \alpha\}$$

est exactement le corollaire 3.1. Soit $\theta \in \mathfrak{S}$, où

$$\mathfrak{S} = \bigcap \{H \mid H \text{ est un demi-espace fermé tel que } \mathbb{P}(H^c) < \alpha\}.$$

Supposons par l'absurde que $\theta \notin D_\alpha$. Il existe alors $u \in S^{p-1}$ tel que $\alpha_0 = \mathbb{P}(H_{\theta,u}) < \alpha$. En effet, si ce n'était pas le cas, alors $D_H(\theta) \geq \alpha$. Définissons la suite de demi-espaces ouverts

$$H_{\theta,u}^{(n)} = \left\{ x \in \mathbb{R}^p \mid u'(x - \theta) > -\frac{1}{n} \right\}.$$

Claiement, $H_{\theta,u}^j \subset H_{\theta,u}^k$ pour $j < k$. De plus,

$$H_{\theta,u} = \bigcap_{n=1}^{\infty} H_{\theta,u}^{(n)} = \lim_{k \rightarrow \infty} \bigcap_{n=1}^k H_{\theta,u}^{(n)} = \lim_{k \rightarrow \infty} H_{\theta,u}^{(k)}.$$

Puisque la mesure est bornée, il nous est permis d'utiliser le théorème de la convergence dominée. Et ainsi,

$$\mathbb{P}(H_{\theta,u}) = \lim_{n \rightarrow \infty} \mathbb{P}(H_{\theta,u}^{(n)}).$$

Il existe donc un certain naturel N tel que $\mathbb{P}(H_{\theta,u}^{(N)}) < \alpha$, $H_{\theta,u}^{(N)}$ étant un ouvert. Posons alors

$$H := (H_{\theta,u}^{(N)})^c,$$

demi-espace fermé tel que son complémentaire a une probabilité strictement inférieure à α et tel que $\theta \notin H$. Une contradiction. ■

Corollaire 3.2. *Pour tout $0 \leq \alpha \leq 1$, D_α est fermé.*

Proposition 3.3. *Pour tout $0 < \alpha \leq 1$, D_α est borné.*

Preuve: Notons $B(n) = \{x \in \mathbb{R}^p \mid \|x\| \leq n\}$ la boule de rayon n centrée en l'origine. Trivialement,

$$1 = \mathbb{P}(\mathbb{R}^p) = \mathbb{P}(\cup_n B(n)) = \lim_{n \rightarrow \infty} \mathbb{P}(B(n)).$$

Ainsi, il existe N tel que $\mathbb{P}(B(N)) > 1 - \alpha$. Soit $\theta \notin B(N)$. Il existe u tel que $\mathbb{P}(H_{\theta,u}) < \alpha$, c'est-à-dire $D_H(\theta) < \alpha$, ou encore, $\theta \notin D_\alpha$. ■

Corollaire 3.3. *Pour tout $0 < \alpha \leq 1$, D_α est compact.*

La preuve du théorème ne requiert maintenant qu'une quantité minimale de courage (dont vous êtes certainement doté) :

Preuve du théorème 3.1 : Posons

$$\alpha_* := \sup_{\theta} D_H(\theta) \leq 1.$$

Pour tout $0 < \alpha < \alpha_*$, D_α est un ensemble non vide et compact. Considérons les ensembles $D_n = D_{\alpha_*(1-\frac{1}{n})}$, $n = 2, 3, 4, \dots$. Soient $\theta_n \in D_n$. Les D_n étant emboîtés, pour tout n , on a que

$$\theta_n \in D_{\alpha_* - \frac{\alpha_*}{2}} = D_{\frac{\alpha_*}{2}},$$

compact. Ainsi, il existe une sous-suite (θ_{n_k}) telle que

$$\theta_{n_k} \rightarrow \theta_* \in D_{\frac{\alpha_*}{2}}.$$

Clairement, $\theta_* \in D_\alpha$ pour tout $0 < \alpha < \alpha_*$, puisque $\theta_{n_k} \in D_n$ pour k grand, n fixé. Ainsi,

$$D_H(\theta_*) \geq \alpha,$$

pour tout $0 < \alpha < \alpha_*$. Et donc,

$$D_H(\theta_*) \geq \alpha_* = \sup_{\theta} D_H(\theta) \Rightarrow D_H(\theta_*) = \alpha_*$$

■

3.4 POUR ALLER PLUS LOIN

La présente introduction n'est certes que la partie immergée de l'iceberg qu'est la profondeur. D'autres définitions (cherchant à quantifier la centralité d'un point dans un jeu d'observations) sont nombreuses et dignes d'intérêt. Voici quelques articles permettant de rentrer en douceur dans le monde de la profondeur de position : Small [?] décrit les différents types de médiane multivariée d'intérêt (dont bien sûr la médiane de demi-espace). L'article fondateur de Tukey [10] est digne d'attention, mais l'approche adoptée par Rousseeuw et Ruts [9] est plus complète et mathématique. La structure de cette introduction est issue de ce dernier. L'autre notion phare en profondeur de position est la médiane simpliciale introduite par Liu [?]. Celle-ci étudie les différentes propriétés de cette profondeur à l'aide de preuves remarquablement ingénieuses. Serfling et Zuo [?] sont incontournables lorsqu'il s'agit de profondeur. Ce dernier article aborde de manière générale et abstraite les différents propriétés de classes de fonctions de profondeur.

4 Le théorème du vote

4.1 ENONCÉ ET DÉMONSTRATION

Ce chapitre est consacré à un résultat élégant et simple appelé le *théorème du vote* (en anglais "ballot theorem"). Nous allons, dans la suite, énoncer et démontrer ce théorème, et présenter deux applications qui en découlent. La première sert à déterminer la probabilité de ruine d'une compagnie d'assurance dans un modèle classique en théorie du risque, alors que la seconde est un résultat démontré à la fin du 19ème siècle, appelé le *théorème classique du vote*. Lors d'une élection opposant deux candidats, nous déterminerons la probabilité qu'un candidat mène durant tout le dépouillement des votes, sachant que ce candidat a gagné l'élection. Finalement, une variante du théorème du vote sera présentée.

Théorème 4.1 (Théorème du vote). Soit X_1, X_2, \dots une suite de variables aléatoires i.i.d. à valeurs dans \mathbb{N} . Notons les sommes partielles correspondantes $S_n = \sum_{i=1}^n X_i$ pour tout $n \geq 1$, avec $S_0 = 0$. Alors, pour $n \geq 1$ et $0 \leq k \leq n$,

$$P[S_1 < 1, S_2 < 2, \dots, S_n < n \mid S_n = k] = (n - k)/n. \quad (4)$$

Graphiquement, cette égalité signifie qu'étant donné que $S_n = k$, la probabilité que les sommes partielles S_1, \dots, S_n soient restées sous la première bissectrice est égale à $(n - k)/n$.

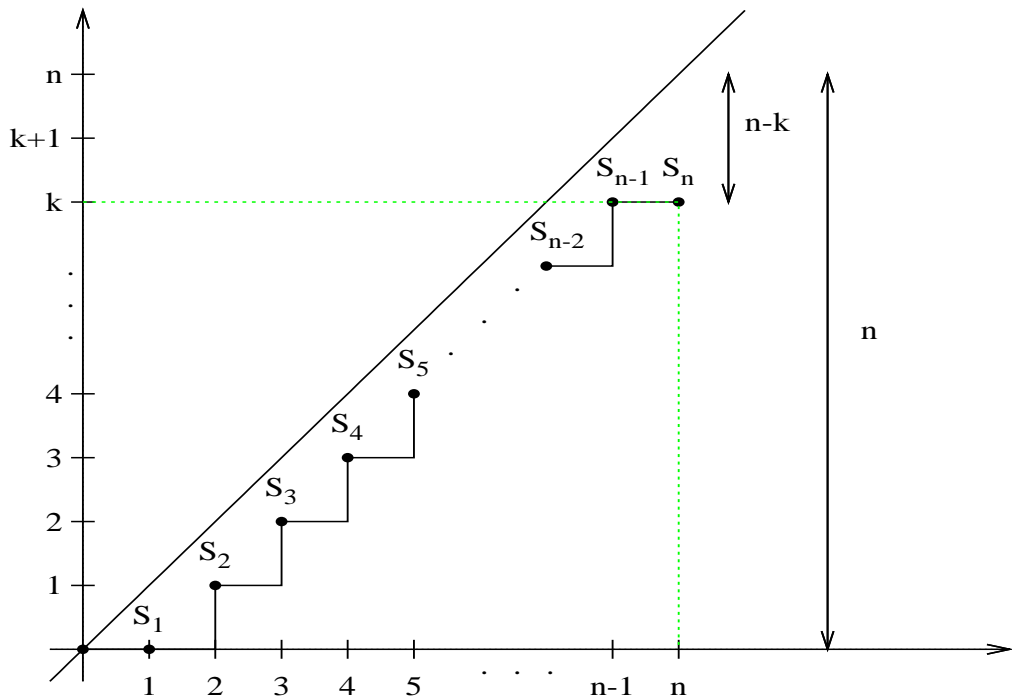


Figure 1. Représentation graphique du théorème du vote.

Preuve : Pour $k = n$, l'égalité (4) est trivialement satisfaite. Pour $0 \leq k < n$, nous allons démontrer le théorème du vote par induction sur n :

1. Pour $n = 1$, nous avons $P[S_1 < 1 \mid S_1 = 0] = 1$, d'où (4) tient pour $0 \leq k < 1$.
2. Supposons que l'égalité (4) est vraie pour $n \leq l - 1$ et montrons qu'elle reste satisfaite pour $n = l$. Pour $0 \leq k < l$, conditionner par rapport à la valeur de S_{l-1} nous donne

$$P[S_1 < 1, \dots, S_l < l \mid S_l = k]$$

$$\begin{aligned}
&= \sum_{j=0}^k \mathbb{P}[S_1 < 1, \dots, S_l < l \mid S_l = k, S_{l-1} = j] \mathbb{P}[S_{l-1} = j \mid S_l = k] \\
&= \sum_{j=0}^k \mathbb{P}[S_1 < 1, \dots, S_{l-1} < l-1 \mid S_{l-1} = j] \mathbb{P}[S_{l-1} = j \mid S_l = k] \\
&= \sum_{j=0}^k \frac{(l-1-j)}{(l-1)} \mathbb{P}[S_{l-1} = j \mid S_l = k] \\
&= \sum_{j=0}^k \mathbb{P}[S_{l-1} = j \mid S_l = k] - \frac{1}{l-1} \sum_{j=0}^k j \mathbb{P}[S_{l-1} = j \mid S_l = k] \\
&= 1 - \frac{1}{l-1} \mathbb{E}[S_{l-1} \mid S_l = k].
\end{aligned}$$

Comme les variables aléatoires X_n sont indépendantes et que $\mathbb{E}[S_l \mid S_l = k] = k$, il en découle que

$$\begin{aligned}
\mathbb{E}[S_{l-1} \mid S_l = k] &= \mathbb{E}[X_1 + \dots + X_{l-1} \mid S_l = k] = (l-1)\mathbb{E}[X_1 \mid S_l = k] \\
&= \frac{(l-1)}{l} \mathbb{E}[S_l \mid S_l = k] = \frac{(l-1)k}{l},
\end{aligned}$$

ce qui conclut la preuve. ■

4.2 APPLICATION À LA RUINE D'UNE COMPAGNIE D'ASSURANCE

Considérons une compagnie d'assurance établissant son bilan à la fin de périodes successives notées $1, 2, \dots$. Cette compagnie d'assurance possède un capital initial nul et reçoit des primes de ses assurés. Ceux-ci les paient de façon continue à un taux $c = 1$ par unité de temps de sorte qu'à l'instant t , c'est-à-dire à la fin de la t -ème période, ceux-ci ont versé le montant t à la compagnie d'assurance. La première bissectrice correspond donc ici au capital de la compagnie constitué du capital initial et de l'accumulation des primes reçues. De plus, nous supposons que la compagnie d'assurance doit rembourser les montants des sinistres survenus sur chaque période. Nous notons par X_n le montant des sinistres relatifs à la n -ème période et payé à la fin de celle-ci, c'est-à-dire en n . Ces variables aléatoires sont supposées indépendantes, identiquement distribuées et positives.

La compagnie d'assurance est ruinée lorsqu'elle n'a plus assez d'argent pour faire face aux paiements des sinistres. Par conséquent, l'instant de ruine, noté T , est défini par

$$T = \inf\{t \in \mathbb{N} \mid t \leq S_t\}.$$

Il en découle que la probabilité que la ruine survienne après l'instant n est donnée par $P[T > n] = P[S_r < r, r = 1, \dots, n]$. Dès lors, grâce au théorème du vote, nous avons

$$\begin{aligned} P[S_r < r, r = 1, \dots, n] &= \sum_{k=0}^{n-1} P[S_r < r, r = 1, \dots, n \mid S_n = k] P[S_n = k] \\ &= \sum_{k=0}^{n-1} \frac{(n-k)}{n} P[S_n = k] = \frac{1}{n} \sum_{k=0}^{n-1} (n-k) P[S_n = k] \\ &= E\left[\left(\frac{n-S_n}{n}\right)^+\right], \end{aligned}$$

où $x^+ := \max\{x, 0\}$.

4.3 APPLICATION AU THÉORÈME CLASSIQUE DU VOTE

Ci-après, nous énonçons le théorème dit *classique du vote* :

Théorème 4.2 (Théorème classique du vote). *Si lors d'un vote, un candidat A remporte a votes et un candidat B remporte b votes, et que $a > \mu b$ (où μ est un naturel strictement positif), alors la probabilité que durant tout le compte des votes, ceux remportés par A soient toujours strictement supérieurs à μ fois ceux remportés par B est*

$$\frac{a - \mu b}{a + b},$$

à condition que tous les résultats de vote possibles soient équiprobables et indépendants.

Ce résultat fut démontré indépendamment par Bertrand et par André pour $\mu = 1$ en 1887. Ce théorème fut énoncé sans démonstration pour $\mu \geq 1$ par Barbier la même année. En 1924 finalement, une première démonstration fut apportée.

Nous allons examiner le dépouillement successif des bulletins de vote. Nous définissons deux suites de variables aléatoires $X_i^A, i = 1, \dots, a+b$, et $X_i^B, i = 1, \dots, a+b$, rendant compte du résultat du i -ème bulletin de vote :

$$X_i^A = \begin{cases} 1 & \text{si le vote } i \text{ est en faveur du candidat A,} \\ 0 & \text{si le vote } i \text{ est en défaveur du candidat A,} \end{cases}$$

et

$$X_i^B = \begin{cases} 1 & \text{si le vote } i \text{ est en faveur du candidat B,} \\ 0 & \text{si le vote } i \text{ est en défaveur du candidat B.} \end{cases}$$

Notons par S_i^A et S_i^B le nombre de votes obtenus par A et B jusqu'au i -ème dépouillement. Ceux-ci sont donnés par $S_i^A = X_1^A + \dots + X_i^A$ et $S_i^B = X_1^B + \dots + X_i^B$, et vérifient $S_i^A + S_i^B = i$. La probabilité d'intérêt est alors donnée par

$$\begin{aligned} & \mathbb{P}[S_i^A > \mu S_i^B, i = 1, \dots, a + b \mid S_{a+b}^A = a] \\ &= \mathbb{P}[i - S_i^B > \mu S_i^B, i = 1, \dots, a + b \mid S_{a+b}^B = b] \\ &= \mathbb{P}[i > (\mu + 1)S_i^B, i = 1, \dots, a + b \mid S_{a+b}^B = b] \\ &= \mathbb{P}[i > \tilde{S}_i^B, i = 1, \dots, a + b \mid \tilde{S}_{a+b}^B = (\mu + 1)b], \end{aligned}$$

où $\tilde{S}_i^B = \tilde{X}_1^B + \dots + \tilde{X}_i^B$ avec $\tilde{X}_j^B = (\mu + 1)X_j^B$, $j = 1, \dots, a + b$.

Finalement, en utilisant le théorème du vote pour les variables \tilde{X}_j^B , nous obtenons

$$\mathbb{P}[S_i^A > \mu S_i^B, i = 1, \dots, a + b \mid S_{a+b}^A = a] = \frac{(a + b) - (\mu + 1)b}{a + b} = \frac{a - \mu b}{a + b}.$$

5 Le théorème de Gauss

Nul n'est censé ignorer la loi (normale)!

5.1 INTRODUCTION

La loi normale, encore appelée loi gaussienne en l'honneur du célèbre mathématicien Carl Friedrich Gauss, est sans conteste la distribution de probabilité la mieux connue de par le monde. Qui, en effet, ne connaît pas la célèbre “cloche de Gauss” ? Que ce soit en finance, en pharmacologie, en géographie, en **ingénierie** ou en médecine, cette loi est utilisée dans de nombreuses applications pratiques des outils statistiques et sert de base à presque tous les modèles d'utilisation courante.

L'universalité de cette loi est bien souvent justifiée par le *théorème central limite*, qui nous enseigne que toute somme (normalisée) de variables aléatoires dont le comportement n'est pas trop pathologique aura une loi asymptotiquement gaussienne. Ceci implique en particulier que tout phénomène qui peut être décrit comme résultant d'un grand nombre d'interactions indépendantes donnera naissance à une distribution approximativement gaussienne !

Il se trouve toutefois que le *TCL* n'est pas la seule propriété remarquable de la loi normale. L'objectif de cette section est de vous présenter une autre caractéristique qui distingue cette dernière des autres distributions de probabilité.

5.2 LOI GAUSSIENNE ET MAXIMUM DE VRAISEMBLANCE

Lorsque l'on étudie un phénomène aléatoire dont on connaît la distribution, il est souvent utile de chercher à estimer les paramètres de ladite distribution. Par exemple, on sait que la distribution des cotes d'examen des étudiants – du moins dans des sections avec suffisamment d'étudiants – est approximativement normale ; on a toutefois besoin d'*estimer* la moyenne μ et l'écart-type σ . Bien souvent, on estime ces paramètres en cherchant les valeurs de μ et de σ qui maximisent la vraisemblance de l'échantillon observé. On appelle les estimateurs ainsi obtenus les *estimateurs du maximum de vraisemblance*.

Une propriété célèbre des modèles gaussiens est l'égalité entre l'estimateur du maximum de vraisemblance pour la moyenne et la moyenne empirique. Plus précisément, il est (très) bien connu que si x_1, \dots, x_n est un échantillon aléatoire k -dimensionnel issu d'une population gaussienne de moyenne μ et de variance-covariance Σ , alors l'estimateur du maximum de vraisemblance pour μ est donné par

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

et ce indépendamment du fait que Σ soit connu ou nécessite une estimation.

Cette section sera consacrée à la preuve d'un résultat bien moins connu mais non moins intéressant pour autant, à savoir que la propriété que nous venons d'énoncer constitue, en fait, une équivalence : la loi gaussienne est la seule dont le MLE (pour μ) coïncide tout le temps avec la moyenne empirique !

Théorème 5.1. *Soit une population de variables aléatoires absolument continues telle que pour toute valeur du paramètre de position $\mu \in \mathbb{R}$ la fonction de densité correspondante peut s'écrire $f(x - \mu)$ au point $x \in \mathbb{R}$. Considérons un échantillon aléatoire de taille $n \geq 3$ extrait de cette population, et qui satisfasse aux deux conditions suivantes :*

- (i) *$f(x)$ est dérivable en x et sa dérivée $f'(x)$ est continue en au moins un point $x \in \mathbb{R}$;*
- (ii) *pour tout échantillon x_1, \dots, x_n , la moyenne empirique \bar{x} est solution de l'équation de vraisemblance pour le paramètre de position μ , i.e. $MLE(\mu) = \bar{x}$. Alors $f(x)$ n'est autre que $\phi(x; \sigma^2)$, la fonction de densité d'une loi $\mathcal{N}(0, \sigma^2)$ pour un certain $\sigma^2 > 0$.*

La première formulation du Théorème 5.1 date de 1809 et a été publiée – bien que sous une forme très différente – par Gauss dans *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Ce résultat est alors resté quelque peu oublié, jusqu'à ce qu'en 1963 une compagnie d'édition anglaise réédite le *Theoria etc.*

dans une traduction anglaise. Depuis lors, plusieurs statisticiens de grande renommée se sont intéressés à cette caractérisation de la distribution gaussienne, et ont étendu le résultat originel en relaxant certaines des hypothèses qui étaient nécessaires pour Gauss. A ce jour, la forme la moins restrictive de ce résultat est celle donnée ci-dessus, qui est due à Adelchi Azzalini (le créateur de la loi skew-normale) et à Marc G. Genton.

5.3 PREUVE DU THÉORÈME DE GAUSS

La log-vraisemblance associée à un échantillon x_1, \dots, x_n issu d'une population de fonction de densité $f(x-\mu)$ est donnée par $\ell_{f;\mu}(x_1, \dots, x_n) := \sum_{i=1}^n \log(f(x_i-\mu))$, et l'équation de vraisemblance pour μ est donc donnée par

$$\sum_{i=1}^n \frac{d}{d\mu} \log(f(x_i - \mu)) = 0,$$

ce qui, en posant $g(x) = \frac{d}{dx} \log(f(x))$ (changement de variables permis car $f(x) > 0$ sur son domaine), devient $\sum_{i=1}^n g(x_i - \mu) = 0$.

Une densité f satisfera donc à la condition (ii) du théorème lorsque

$$\sum_{i=1}^n g(x_i - \bar{x}) = 0 \tag{5}$$

pour tout échantillon x_1, \dots, x_n de taille $n \geq 3$. Par conséquent, prouver le théorème revient à montrer que $f(x) = \phi(x; \sigma^2)$ est la seule fonction dont la dérivée du logarithme satisfait à (5).

Par où faut-il commencer pour résoudre une telle équation? Souvent, afin de mieux comprendre la structure d'un problème, il est bon de procéder par exemples. Nous allons maintenant voir que dans le présent cas, cette stratégie va s'avérer payante : en prenant des exemples d'échantillons bien choisis, nous allons obtenir des conditions sur la fonction g qui mèneront à sa détermination exacte.

Considérons dans un premier temps l'échantillon $x_1 = x_2 = \dots = x_n = u$ pour une constante réelle u donnée. L'équation (5) devient alors

$$ng(0) = 0,$$

dont on déduit déjà que $g(0) = 0$ nécessairement. Choisissons dans un second temps un échantillon de la forme $2u, 0, u, \dots, u$. L'équation (5) devient alors

$$g(u) + g(-u) + (n-2)g(0) = g(u) + g(-u) = 0,$$

ce qui implique que g est une fonction impaire. Choisissons finalement un échantillon de la forme $u, v, -(u+v), 0, \dots, 0$, avec u et v deux réels donnés. La moyenne empirique de cet échantillon étant nulle, il s'ensuit que $g(u) + g(v) + g(-(u+v)) = 0$, expression qui, de par la parité de g , se simplifie en

$$g(u) + g(v) = g(u+v). \quad (6)$$

Toute solution de (5) doit donc satisfaire à (6), qui n'est autre que la célèbre *équation fonctionnelle de Cauchy*. Prouver le théorème revient donc à résoudre cette dernière.

Pour ce faire, remarquons d'abord que si g est solution de (6), alors $g(n) = ng(1)$ pour tout $n \in \mathbb{N}$ (n'oublions pas que $g(0)=0$). Ce résultat s'étendant aisément au cas des entiers, on a donc

$$g(z) = zg(1) \quad \forall z \in \mathbb{Z}.$$

Considérons maintenant $q = \frac{m}{n}$, avec $m, n \in \mathbb{Z}$. A nouveau, il est clair que $g(m/n) = \sum_{i=1}^m g(1/n) = mg(1/n)$. Or, comme $1 = n/n$, on déduit allègrement que $g(1) = g(n/n) = ng(1/n)$, et donc

$$g(q) = g(m/n) = mg(1/n) = mg(1)/n = qg(1).$$

Ces considérations nous poussent à conjecturer que $g(x) = xg(1)$ pour tout réel x , conjecture qu'il est aisé de valider via un argument de continuité. En effet, la condition (i) du théorème exige que la dérivée de la densité f soit continue en au moins un point. La fonction g sera donc elle aussi continue en au moins un point. Or nous pouvons, sans perte de généralité, supposer que ce point est l'origine et donc, étant donné que pour tout $x_0 \in \mathbb{R}$ on a

$$\lim_{x \rightarrow x_0} g(x) - g(x_0) = \lim_{x \rightarrow x_0} [g(x) - g(x_0)] = \lim_{x \rightarrow x_0} g(x - x_0) = g(0) = 0,$$

on peut conclure que

$$\lim_{x \rightarrow x_0} g(x) = g(x_0)$$

pour tout x_0 réel.

La solution générale de (6) est donc de la forme $g(x) = g(1)x = cx$, avec $c \in \mathbb{R}$ une constante. On en déduit, en intégrant, que

$$f(x) = de^{cx^2/2} \quad \forall x \in \mathbb{R},$$

avec d et c des constantes réelles. Comme f doit être intégrable sur tout \mathbb{R} , il suit que $c < 0$; de plus, comme son intégrale doit valoir 1, il suit que $f(x) = \phi(x; -1/c)$, ou encore, en posant $\sigma^2 = -1/c$,

$$f(x) = \phi(x; \sigma^2).$$

Le théorème est donc prouvé. ■

Bien évidemment, ce résultat peut être étendu au cas k -dimensionnel, cas (un lecteur averti comprendra ici l'usage de k comme indicateur de dimension plutôt que p) laissé comme exercice intéressant à celui qui en a envie. Les techniques restent grosso modo les mêmes.

6 Méthode probabiliste en théorie des graphes

6.1 GÉNÉRALITÉS SUR LA MÉTHODE PROBABILISTE

Parmi les succès les plus inattendus de la théorie des probabilités on trouve la *méthode probabiliste*, terme général donné à tout résultat déterministe (c'est-à-dire non probabiliste) obtenu à l'aide des probabilités. Outre l'intérêt intrinsèque de cette méthode, les preuves probabilistes s'avèrent souvent plus élégantes et plus intuitives que les preuves élémentaires (qui d'ailleurs pour certains résultats n'existent pas encore!). La méthode probabiliste s'applique à quasiment toutes les disciplines mathématiques : analyse, algèbre, théorie des nombres, combinatoire, géométrie, etc. Nous nous intéresserons ici à son application à la théorie des graphes, qui est un des domaines où elle s'applique le mieux, mais il faut garder à l'esprit que la théorie des graphes nous servira plutôt de prétexte à l'étude de la méthode probabiliste. Par conséquent, nous encourageons vivement le lecteur d'être plus admiratif envers les idées mises en œuvre qu'aux résultats auxquels elles mèneront.

Avant d'entrer dans le vif du sujet, signalons quelques résultats élégants qui peuvent être obtenus par la méthode probabiliste.

Théorème 6.1 (Théorème d'approximation de Weierstrass). *Toute fonction continue de $[0, 1]$ dans \mathbb{R} peut être uniformément approchée par des polynômes.*

Weierstrass démontra ce théorème vers 1885 par des méthodes purement analytiques. En 1912, Sergeï Natanovitch Bernstein donna une preuve probabiliste élégante et constructive de ce théorème.

Théorème 6.2 (Théorème fondamental de l'algèbre). *Tout polynôme non constant à coefficients complexes admet au moins une racine.*

Ce célèbre théorème possède de nombreuses preuves (algébriques, analytiques, topologiques, ...). Étonnamment, une preuve probabiliste existe également, basée sur les propriétés des mouvements browniens (cf. D. Revuz et M. Yor, *Continuous Martingales and Brownian Motion*).

Théorème 6.3 (B.J. Green et T. Tao, 2003). *Il existe des progressions arithmétiques de nombres premiers de longueur arbitraire.*

Ce théorème, déjà conjecturé plus de deux siècles auparavant, est une des plus importantes avancées réalisées dans notre connaissance des nombres premiers, dont les mystères font rêver les mathématiciens depuis plusieurs millénaires. Aucune preuve non probabiliste n'est connue à ce jour.

Le point de départ de la méthode probabiliste est toujours le même : transformer un ensemble Ω d'objets que nous souhaitons étudier en un espace probabilisé (Ω, \mathcal{A}, P) ou en plusieurs espaces probabilisés $(\Omega_n, \mathcal{A}_n, P_n)$, où n sera un paramètre des objets étudiés (par exemple l'ordre des graphes). L'idée consiste ensuite à travailler avec un objet aléatoire $\omega \in \Omega$, grâce auquel nous pourrions allègrement obtenir des résultats d'existence. Par exemple, si τ désigne une propriété des objets de Ω et que nous pouvons montrer que $P[\tau(\omega)] > 0$, alors nous savons qu'il doit exister au moins un ω possédant la propriété τ . Ou encore : si N est une variable aléatoire sur Ω à valeurs dans \mathbb{R} et que $E[N] = \mu$, alors nous savons qu'il doit exister ω_1 et ω_2 tels que $N(\omega_1) \leq \mu \leq N(\omega_2)$. Un autre type de résultats tout aussi important consiste à montrer que « la plupart » des objets étudiés partagent une certaine propriété. Par exemple, si les Ω_n sont croissants, nous pourrions dire que « presque tous les objets possèdent la propriété τ » si $P_n[\tau(\omega)] \rightarrow 1$ quand $n \rightarrow \infty$. Toutes ces idées théoriques risquent de paraître vagues et incompréhensibles pour le moment, mais elles seront mieux comprises une fois que nous les aurons appliquées (ce qui sera fait dans la suite).

6.2 MODÈLES PROBABILISTES EN THÉORIE DES GRAPHES

Commençons par quelques notations. Pour tout naturel n , nous noterons $N = \binom{n}{2}$; de plus, il sera également commode d'écrire $[n]$ pour l'ensemble $\{1, \dots, n\}$.

Nous considérerons des graphes construits sur les sommets $[n]$. Un graphe *vide* (c'est-à-dire ne contenant aucune arête) sur $[n]$ sera noté \mathcal{E}_n . Un graphe *complet* (c'est-à-dire dont toute paire de sommets est adjacente) sur $[n]$ sera noté \mathcal{K}_n ; il possède N arêtes et 2^N sous-graphes (dont, bien sûr, \mathcal{E}_n). Le nombre d'arêtes d'un graphe H sera noté $e(H)$, et le sous-graphe de G engendré par l'ensemble de sommets α sera noté $G[\alpha]$. Finalement, des sommets sont dits *indépendants* si aucune arête n'en relie deux d'entre eux (autrement dit, si le sous-graphe engendré est vide).

Comme d'habitude en probabilités, nous réserverons l'usage de la lettre p pour un réel compris entre 0 et 1 et la lettre q désignera la probabilité complémentaire à p , c'est-à-dire $1 - p$.

Il s'avèrera aussi utile, dans la suite, de donner des approximations pour certaines factorielles, aussi utiliserons-nous les bornes suivantes (qui sont une version de la formule de Stirling) :

$$\sqrt{2\pi s} \left(\frac{s}{e}\right)^s \leq s! \leq e^{\frac{1}{12s}} \sqrt{2\pi s} \left(\frac{s}{e}\right)^s, \quad \forall s \in \mathbb{N}. \quad (7)$$

Un des premiers résultats en théorie des graphes semble être un théorème de T. Szele daté de 1943, que nous prouverons plus loin. Néanmoins, la théorie probabiliste des graphes a véritablement démarré sous l'impulsion du mathématicien hongrois Pál Erdős qui surprit tout le monde en résolvant un problème de la théorie de Ramsey en 1947 (voir section 6.3) en ayant recours à des moyens probabilistes. L'étude des graphes aléatoires pour eux-mêmes débuta vraiment en 1959 avec plusieurs articles d'Erdős et Rényi, et depuis ne cesse d'engranger de nouveaux résultats. Plusieurs modèles probabilistes existent pour la théorie des graphes. Nous en utiliserons deux, les plus courants et utiles.

Soit M compris entre 0 et N . Nous définissons l'espace $\mathcal{G}(n, M)$ comme l'ensemble des sous-graphes de \mathcal{K}_n possédant M arêtes (un exercice facile révèle directement que ces sous-graphes sont au nombre de $\binom{N}{M}$) muni de la distribution uniforme. Nous noterons G_M (ou, si nécessaire, $G_{n,M}$) un graphe aléatoire de $\mathcal{G}(n, M)$. Par conséquent, si H est un graphe construit sur $[n]$ à M arêtes,

$$P_M[G_M = H] = \binom{N}{M}^{-1}.$$

Soit ensuite p un réel entre 0 et 1. Pour construire un graphe aléatoire, nous pouvons considérer toutes les N arêtes existantes de \mathcal{K}_n et déterminer pour chacune avec probabilité p si nous la choisissons pour notre graphe ou pas, et ceci indépendamment des autres arêtes. Ceci forme donc un espace probabilisé $\mathcal{G}(n, p)$ dont nous noterons G_p (ou $G_{n,p}$) un graphe aléatoire, de telle sorte que pour un sous-graphe H de \mathcal{K}_n , nous ayons

$$P_p[G_p = H] = p^{e(H)} q^{N-e(H)}.$$

Afin de se familiariser avec la méthode, nous commençons par énoncer quelques résultats simples qui nous seront utiles pour la suite. Fixons un naturel n et un naturel s compris entre 0 et n . Pour tout graphe G , notons $X_s(G)$ le nombre de ses sous-graphes complets d'ordre s et $X'_s(G)$ le nombre d'ensembles de sommets indépendants de taille s . Si G est un graphe aléatoire, X_s et X'_s deviennent des variables aléatoires. Déterminons leur espérance dans les deux modèles que nous avons introduits.

Pour ce faire, utilisons une technique courante en probabilité : décomposons nos variables aléatoires comme sommes d'indicatrices. Nommons \mathcal{T} l'ensemble des sous-ensembles de $[n]$ de taille s et, pour chaque $\alpha \in \mathcal{T}$, Y_α l'indicatrice du sous-graphe complet sur les sommets de α et Y'_α celle du sous-graphe vide sur les sommets de α :

$$Y_\alpha(G) = \begin{cases} 1 & \text{si } G[\alpha] \text{ est complet,} \\ 0 & \text{sinon,} \end{cases}$$

$$Y'_\alpha(G) = \begin{cases} 1 & \text{si } G[\alpha] \text{ est vide,} \\ 0 & \text{sinon.} \end{cases}$$

Nous avons ainsi

$$\begin{aligned} X_s(G) &= \sum_{\alpha \in \mathcal{T}} Y_\alpha(G), \\ \mathbb{E}[X_s] &= \sum_{\alpha \in \mathcal{T}} \mathbb{E}[Y_\alpha(G)] = \sum_{\alpha \in \mathcal{T}} \mathbb{P}[G[\alpha] = \mathcal{K}_\alpha] \end{aligned}$$

et des égalités similaires pour X'_s et Y'_α .

Intéressons-nous d'abord à l'espace $\mathcal{G}(n, M)$. En notant $S = \binom{s}{2}$, nous trouvons facilement

$$\mathbb{E}_M[Y_\alpha] = \mathbb{P}_M[G_M[\alpha] = \mathcal{K}_\alpha] = \binom{N-S}{M-S} \binom{N}{M}^{-1},$$

puisque une fois les S arêtes de \mathcal{K}_α fixées, nous pouvons choisir librement les $M-S$ autres arêtes parmi les $N-S$ arêtes restantes. Le même raisonnement fournit

$$\mathbb{E}_M[Y'_\alpha] = \mathbb{P}_M[G_M[\alpha] = \mathcal{E}_\alpha] = \binom{N-S}{M} \binom{N}{M}^{-1}.$$

Pour $\mathcal{G}(n, p)$, nous obtenons de la même manière

$$\mathbb{E}_p[Y_\alpha] = \mathbb{P}_p[G_p[\alpha] = \mathcal{K}_\alpha] = p^S$$

et

$$\mathbb{E}_p[Y'_\alpha] = \mathbb{P}_p[G_p[\alpha] = \mathcal{E}_\alpha] = q^S.$$

Enfin, comme $|\mathcal{T}| = \binom{n}{s}$, nous pouvons conclure

$$\begin{aligned} \mathbb{E}_M[X_s] &= \binom{n}{s} \binom{N-S}{M-S} \binom{N}{M}^{-1}, & \mathbb{E}_M[X'_s] &= \binom{n}{s} \binom{N-S}{M} \binom{N}{M}^{-1}, \\ \mathbb{E}_p[X_s] &= \binom{n}{s} p^S, & \mathbb{E}_p[X'_s] &= \binom{n}{s} q^S. \end{aligned} \quad (8)$$

Ces simples résultats nous permettront de prouver un résultat important de la théorie des graphes.

6.3 NOMBRES DE RAMSEY

Rassemblez six personnes : il y aura toujours un groupe de trois personnes se connaissant ou un groupe de trois personnes ne se connaissant pas mutuellement. Ce résultat de prime abord anodin est un cas spécial d'une riche branche de la théorie des graphes et de la mathématique, la *théorie de Ramsey*. En quelques mots, la théorie de Ramsey s'intéresse au fait « qu'un désordre complet est impossible »¹, c'est-à-dire qu'aussi arbitraires que soient certains objets, ils contiennent des sous-objets structurés, qui peuvent être aussi larges que souhaité pour autant que les objets de départ le soient suffisamment. Nous nous intéresserons aux premiers résultats de la théorie de Ramsey, qui établissent qu'à partir d'un certain ordre, les arêtes d'un graphe complet ne peuvent être partitionnées sans obtenir d'autres graphes complets.

Afin d'assouplir la terminologie et de faciliter la visualisation, nous appellerons *coloration rouge-bleu* une partition des arêtes d'un graphe en deux ensembles. Nous dirons alors naturellement qu'une arête est rouge ou bleue selon la partition à laquelle elle appartient. De même, un sous-graphe sera dit rouge (respectivement bleu) si toutes ses arêtes le sont et nous parlerons également de sous-graphe *monochromatique* pour signifier qu'il est d'une unique couleur. Prenons garde à ne pas confondre ces termes avec d'autres colorations plus classiques en théorie des graphes, comme celle consistant à colorier les sommets d'un graphe sans que deux sommets adjacents ne soient de la même couleur. En particulier, nous ne posons aucune condition sur notre coloration, sinon que toute arête doit posséder une et une seule couleur.

Définition 6.1. *Soient s et t deux naturels non nuls. Le nombre de Ramsey relatif à s et t , noté $R(s, t)$, est la plus petite valeur de n telle que n'importe quelle coloration rouge-bleu de \mathcal{K}_n contienne un \mathcal{K}_s rouge ou un \mathcal{K}_t bleu. Les nombres de Ramsey $R(s, s)$ sont dits diagonaux et notés $R(s)$.*

Les nombres de Ramsey diagonaux $R(s)$ peuvent également être interprétés comme la plus petite valeur de n telle que tout graphe d'ordre n possède un sous-graphe trivial (c'est-à-dire complet ou vide) d'ordre s .

Il est évident grâce à la définition que les nombres de Ramsey sont symétriques : $R(s, t) = R(t, s)$. Il est de même aisé de voir que

$$\begin{aligned} R(1, t) &= R(t, 1) = 1, \\ R(2, t) &= R(t, 2) = t. \end{aligned}$$

Ainsi, $R(1) = 1$ et $R(2) = 2$. Le résultat anodin du début de cette section est la traduction du fait que $R(3) = 6$ et il est possible de montrer que $R(4) = 18$. Ramsey prouva que ces nombres sont tous finis. Néanmoins, les nombres de Ramsey,

¹T. S. Motzkin

surtout diagonaux, sont très difficiles à calculer et nous ignorons à l'heure actuelle la valeur exacte de $R(5)$ (quoique nous sachions qu'elle est comprise entre 43 et 49). Une citation célèbre de Joel Spencer nous rapporte l'opinion d'Erdős sur la question :

Erdős nous demande d'imaginer une force extraterrestre, largement plus puissante que nous, atterrissant sur la Terre et exigeant de nous la valeur de $R(5)$ sinon la Terre sera détruite. En ce cas, affirme-t-il, nous devrions rassembler tous nos ordinateurs et tous nos mathématiciens et tenter de trouver la valeur. Mais supposons qu'au lieu de cela ils nous demandent la valeur de $R(6)$. En ce cas, pense-t-il, nous devrions essayer de détruire les extraterrestres².

Cette difficulté explique en partie la surprise et l'enthousiasme des mathématiciens lorsqu'Erdős fournit grâce à des méthodes probabilistes des bornes inférieures exponentielles pour les nombres de Ramsey, diagonaux ou non.

Théorème 6.4 (Erdős, 1947). 1. Si s et n ($3 \leq s \leq n$) sont tels que

$$\binom{n}{s} < 2^{\binom{s}{2}-1},$$

alors $R(s) \geq n + 1$. En particulier,

$$R(s) > \frac{1}{e\sqrt{2}} s 2^{s/2}.$$

2. Si s, t et n ($3 \leq s \leq t \leq n$) et p sont tels que

$$\binom{n}{s} p^{\binom{s}{2}} + \binom{n}{t} q^{\binom{t}{2}} < 1, \quad (9)$$

alors $R(s, t) \geq n + 1$.

Preuve : Nous allons démontrer ces résultats grâce à l'argument de l'espérance, par lequel nous prouverons l'existence d'une coloration ne contenant pas de graphe complet monochromatique d'ordre s ou t .

1. Colorions chaque arête de \mathcal{K}_n aléatoirement et indépendamment en rouge ou en bleu. Ceci revient tout simplement à considérer un graphe aléatoire G de $\mathcal{G}(n, \frac{1}{2})$, en identifiant par exemple les arêtes rouges avec celles de G et les arêtes bleues avec ses non-arêtes. Dans ce cas, \mathcal{K}_n contiendra un \mathcal{K}_s rouge si G contient un sous-graphe complet d'ordre s et un \mathcal{K}_s bleu si G contient un

²Joel H. Spencer, Ten Lectures on the Probabilistic Method, Society for Industrial and Applied Mathematics, p. 5

ensemble de sommets indépendants de taille s . Or, en reprenant les notations de la section précédente,

$$\begin{aligned} E_{\frac{1}{2}}[X_s + X'_s] &= 2 \binom{n}{s} 2^{-\binom{s}{2}} && \text{(par (8))} \\ &< 1. && \text{(par hypothèse)} \end{aligned}$$

Nous pouvons donc en déduire l'existence d'un graphe G d'ordre n tel que $X_s(G) + X'_s(G) = 0$, c'est-à-dire que G ne contient ni sous-graphe complet d'ordre s ni ensemble de sommets indépendants de taille s . Ainsi, $R(s) \geq n + 1$. Pour obtenir la seconde inégalité, nous pouvons utiliser les approximations de Stirling (voir (7)) avec $n = \lfloor s2^{s/2}/(e\sqrt{2}) \rfloor$ pour calculer

$$\binom{n}{s} 2^{-\binom{s}{2}+1} < \frac{n^2}{s!} 2^{-\binom{s}{2}+1} < \frac{(e\sqrt{2})^{-s} s^s 2^{\frac{s^2}{2}}}{\sqrt{2\pi s} (\frac{s}{e})^s} 2^{-\binom{s}{2}+1} = \frac{2}{\sqrt{2\pi s}} < 1.$$

2. Le raisonnement est essentiellement le même en ce cas puisque l'espérance de $(X_s + X'_t)(G_{n,p})$ vaut exactement le membre de gauche de (9). ■

Il n'est pas difficile de se convaincre qu'il est possible de trouver n et p afin de satisfaire à l'inégalité (9), mais une formule explicite pour n nécessite quelques calculs sans intérêt d'un point de vue probabiliste. Nous n'allons pas approfondir la question ici.

6.4 POUR ALLER PLUS LOIN

Pour découvrir d'autres applications de la méthode probabiliste en mathématique discrète, un bon début serait de feuilleter [1], qui contient en outre la démonstration de Bernstein du théorème d'approximation de Weierstrass et celle de l'existence de graphes ayant un tour de taille (ou calibre) et un nombre chromatique arbitrairement grands, célèbre résultat d'Erdős. Il illustre également l'usage de la variance dans la méthode probabiliste. Pour poursuivre dans cette voie, [6] est également une lecture de choix, quoique légèrement plus ardue et plus éclectique.

Le sujet spécifique des graphes aléatoires peut être abordé dans [3, chapitre VII]. Le chapitre VI de ce même ouvrage donne également un bon aperçu de la richesse de la théorie de Ramsey. Enfin, pour une étude plus approfondie et complète des graphes aléatoires, [4] semble être incontournable. Ces deux ouvrages exposent notamment d'importants résultats sur les propriétés de presque tout graphe.

Tous ces ouvrages sont disponibles à la Bibliothèque des sciences et techniques de l'Université libre de Bruxelles.

Remerciements

Nous remercions Maude Gathy, Maxime Gheysens et Germain Van Bever pour leur apport rédactionnel au présent document. Nous remercions également Yves Dominicy d'avoir accepté la tâche ingrate de correcteur orthographique. Toute erreur subsistant dans ce document nous incombe, bien entendu. Nous sommes également très reconnaissants envers le Professeur Francis Buekenhout pour sa participation à l'école d'été et pour ses notes de cours (voir annexe). Ses talents de pédagogue ont permis à un public hétéroclite — et pour la plupart inexpérimenté — d'avoir une bonne idée d'un sujet aussi complexe que la théorie des immeubles. Nous remercions également Charlotte Bouckaert pour les très jolies photos qu'elle a prises lors de l'école d'été et dont un échantillon est inclus ci-après. Last but not least, nous tenons à remercier tous les étudiants qui sont venus pour assister à ces mini-cours et qui ont, par leur présence et leur enthousiasme, contribué à ce que cette deuxième édition de la Brussels Summer School se déroule dans une atmosphère des plus plaisante.

Références

- [1] Noga ALON and Joel SPENCER, *The Probabilistic Method*, with an appendix by Paul ERDÖS, Wiley-Interscience, coll. Series in Discrete Mathematics and Optimization, New York, 1992
- [2] Azzalini, A. and Genton, M.G. (2007), *On Gauss's characterization of the normal distribution*, Bernoulli 13, pp. 169-174
- [3] Béla BOLLOBÁS, *Modern Graph Theory*, Springer-Verlag, coll. Graduate Texts in Mathematics, vol. 184, New York, 1998
- [4] Béla BOLLOBÁS, *Random Graphs*, 2nd edition, Cambridge University Press, coll. Cambridge studies in advanced mathematics, vol. 73, Cambridge, 2001
- [5] Ducharme, G.R. and Milasevic, P. (1987). *Uniqueness of the spatial median*, Annals of Statistics 15, No. 3, pp. 1332-1333
- [6] Paul ERDÖS and Joel SPENCER, *Probabilistic Methods in Combinatorics*, Academic Press, coll. Probability and Mathematical Statistics, London, 1974
- [7] Hotelling, H. (1929). *Stability in competition*. Economic Journal 39, pp. 41-57
- [8] Gauss, C.F. (1809), *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*, Hamburg : Perthes et Besser. Traduction anglaise par C.H. Davis, réimprimée par Dover, New York, en 1963
- [9] Rousseeuw, P.J. and Ruts, I. (1999). *The depth function of a population distribution*. Metrika 49, pp. 213-244

-
- [10] Tukey, J.W. (1975). *Mathematics and the picturing of data*, Proceedings of the International Congress of Mathematicians (Vancouver, B.C., 1974), Vol. 2, pp. 523-531
- [11] Weber, A. (1909). *Über den Standort der Industrien*, Tübingen. Traduction anglaise par C.J. Friedrich, réimprimée par l'Université de Chicago, en 1929