Quatre espaces de probabilité importants

1 L'espace Ω est fini ou dénombrable.

Dans ce cas on suppose habituellement que la tribu des évènements $\mathcal{A}$ est $\mathcal{P}(\Omega)$, l'ensemble de toutes les parties de Ω. Par exemple, si Ω est formé de 2 éléments notés a et b, alors $\mathcal{P}(\Omega)$ est constitué des 4 sous ensembles suivants: l'ensemble vide $\emptyset$, les deux singletons $\{a\}$ et $\{b\}$ et $\Omega = \{a,b\}$ lui même. Plus généralement, on a le fait suivant:

Proposition 2.1 Si un ensemble Ω a un nombre fini N d'éléments, alors l'ensemble des parties de Ω: $\mathcal{P}(\Omega)$ a 2^N éléments.

Démonstration On procède par récurrence sur N. C'est trivial pour $N = 1$ ou 0. Si c'est vrai pour N, considérons

$$\Omega = \{a_1, \ldots ,a_N ,a_{N+1}\} \quad \text{et} \quad \Omega' = \{a_1, \ldots ,a_N\}.$$

Les parties de Ω se partagent en deux catégories:

Catégorie 1: celles qui ne contiennent pas a_{N+1}.

Catégorie 2: celles qui contiennent a_{N+1}.

Il est clair que la catégorie 1 est égale à $\mathcal{P}(\Omega')$ et que la catégorie 2 est en bijection avec $\mathcal{P}(\Omega')$, la bijection étant obtenue en ajoutant a_{N+1} aux éléments de $\mathcal{P}(\Omega')$. Comme d'après l'hypothèse de récurrence $\mathcal{P}(\Omega')$ a 2^N éléments, on en conclut que $\mathcal{P}(\Omega)$ a $2^N + 2^N = 2^{N+1}$ éléments, et la récurrence est étendue.

Proposition 2.2 Si Ω est infini dénombrable, alors $\mathcal{P}(\Omega)$ est infini non dénombrable.

Démonstration La démonstration est analogue à la démonstration de Cantor. Sans perte de généralité on suppose Ω égal à l'ensemble $\mathbf{N}$ des entiers positifs ou nuls. Si $X \subset \mathbf{N}$, on lui associe la fonction indicatrice $\mathbf{1}_X$ définie sur $\mathbf{N}$ et à valeurs 0 ou 1 par $\mathbf{1}_X(k) = 1$ si $k \in X$ et $\mathbf{1}_X(k) = 0$ si $k \notin X$. Remarquons aussi qu'inversement, si une fonction f définie sur $\mathbf{N}$ est à valeurs 0 ou 1, alors c'est une indicatrice d'ensemble, c'est-à-dire qu'il existe X tel que $f = \mathbf{1}_X$: il s'agit de $X = \{k \in \mathbf{N}; f(k) = 1\}$.

Montrons alors la proposition par l'absurde en supposant que $\mathcal{P}(\mathbf{N})$ soit dénombrable, c'est-à-dire qu'il existe une application bijective $n \mapsto X_n$ de $\mathbf{N}$ sur $\mathcal{P}(\mathbf{N})$. Alors la fonction f définie sur $\mathbf{N}$ et à valeurs 0 ou 1 par

$$f(k) = 1 - \mathbf{1}_{X_k}(k)$$

est l'indicateur de quelque sous ensemble X_n de $\mathbf{N}$ et donc pour tout k de $\mathbf{N}$ on a

$$\mathbf{1}_{X_n}(k) = 1 - \mathbf{1}_{X_k}(k),$$

ce qui est une contradiction si $k = n$.

Les probabilités sont alors décrites par le résultat suivant

Proposition 2.3 Soit Ω un ensemble fini ou dénombrable. Soit $x \mapsto p_x$ une application de Ω dans les réels ≥ 0 telle que

$$\sum_{x\in\Omega} p_x = 1.$$

Pour tout $A \subset \Omega$, notons alors

$$P(A) = \sum_{x\in A} p_x.$$

Alors $(\Omega,\mathcal{P}(\Omega),P)$ est un espace de probabilité. Inversement, toute probabilité P sur $(\Omega,\mathcal{P}(\Omega))$ est du type précédent, avec $p_x = P(\{x\})$.

Remarque Si Ω est fini, la proposition est évidente. Si Ω est dénombrable, les sommes ci dessus quand A est dénombrable ont la signification suivante: puisque A est dénombrable, on peut numéroter ses éléments, c'est-à-dire qu'il existe une application bijective $n \mapsto x_n$ de $\mathbf{N}$ sur A. $P(A)$ est alors défini rigoureusement comme la somme de la série $\sum_{n=0}^{\infty} p_{x_n}$. Toutefois, ce nombre ne dépend que de A, et non de la numérotation particulière de A choisie par $n \mapsto x_n$, grČce au théorème suivant sur les séries, que nous admettrons, ainsi que la proposition elle même:

Théorème 2.4 Si la série $\sum_{n=0}^{\infty} u_n$ est absolument convergente de somme S, et si $n \mapsto \sigma(n)$ est une bijection de $\mathbf{N}$ sur lui même, alors $\sum_{n=0}^{\infty} u_{\sigma(n)}$ est aussi absolument convergente et de somme S.

Exercices sur 2.1.

1. Soit $\lambda > 0$. Soit P la probabilité définie sur $(\mathbf{N},\mathcal{P}(\mathbf{N}))$ par

   $$P(\{n\}) = e^{-\lambda}\frac{\lambda^n}{n!}.$$

   Soit A l'ensemble des nombres pairs. Calculer $P(A)$. Soit N un entier, montrer que

   $$P(\{0,1,\ldots,N\}) = 1 - \int_0^{\lambda} e^{-t}\frac{t^N}{N!}dt$$

   (Méthode: considérer les deux membres comme des fonctions de λ dont on montrera qu'elles ont même valeur pour $\lambda = 0$ et même dérivée).
2. Soit P la probabilité définie sur $(\mathbf{N}^*,\mathcal{P}(\mathbf{N}^*))$ par $P(\{n\}) = 2^{-n}$. Calculer la probabilité de tirer un nombre $n > 3$; un nombre n multiple de 3; un nombre dont le reste est 3 si on le divise par 4.

2 Le cas équiprobable.

Considérons le cas particulier de la Proposition 2.3 où Ω a un nombre fini $N = |\Omega|$ d'éléments et où tous les p_x sont égaux (et donc égaux à $1/N$.) Dans ce cas, si $A \subset \Omega$ on a

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Pour exploiter cette égalité, il est nécessaire de posséder quelques principes généraux de dénombrement d'ensembles et de fonctions contenus dans les deux prochains théorèmes. Si E et F sont des ensembles, on note par $E \times F$ leur produit cartésien, c'est-à-dire l'ensemble des couples (x,y) tels que $x \in E$ et $y \in F$. On note par F^E l'ensemble des fonctions f définies sur E et à valeurs dans F. Si E est fini et est de taille $n = |E|$ et si k est un entier avec $0 \leq k \leq n$ on note par $\mathcal{P}_k(E)$ l'ensemble des parties de E de taille k.

Théorème 2.5

1. Si E et F sont des ensembles finis, alors $|E \times F| = |E| \times |F|$. Plus généralement, si $F_1, \ldots, F_n$ sont des ensembles finis: $|F_1 \times \cdots \times F_n| = |F_1| \times \cdots \times |F_n|$. Ensuite $|F^E| = |F|^{|E|}$. Enfin, si $p = |F| \geq n = |E|$, le nombre de fonctions *injectives* de E vers F est $p(p-1)(p-2)\cdots(p-n+1)$. En particulier, le nombre de fonctions bijectives de E vers E, appelées permutations de E, est égal à $n!$
2. Si E est fini et est de taille $n = |E|$ et si k est un entier avec $0 \leq k \leq n$ alors

$$|\mathcal{P}_k(E)| = C_n^k = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!}.$$

Démonstration

1. La première formule est évidente : si $e_1, \ldots, e_n$ et $f_1, \ldots, f_p$ sont les éléments de E et F, le nombre de couples (e_i, f_j) est np. L'extension à n facteurs est immédiate également. Cette extension est ensuite appliquée au cas particulier où tous les ensembles F_j sont égaux au même ensemble F. Si $|E| = n$, il y a alors bijection entre F^E et $F \times \cdots \times F$ (n fois). D'où $|F^E| = |F| \times \cdots \times |F| = |F|^n = |F|^{|E|}$. Quant au nombre de fonctions injectives, la formule donnée se justifie facilement: on identifie E à $(1,2,\ldots,n)$, et l'image de 1 peut occuper p positions, l'image de 2 peut occuper une des $p-1$ positions restantes, l'image de 3 une des $p-2$ positions restantes, etc. Faire $E = F$ pour le nombre de permutations de E (on rappelle que si $|E| = |F|$ avec E fini, alors une fonction f de E vers F est injective si et seulement si elle est surjective).
2. Rappelons pour cette partie la formule de Pascal:

Proposition 2.6 Si k est un entier avec $1 \leq k \leq n$ on a

$$C_n^{k-1} + C_n^k = C_{n+1}^k.$$

Démonstration

$$C_n^{k-1}+C_n^k=\frac{n!}{(k-1)!(n-k)!}\left[\frac{1}{n-k+1}+\frac{1}{k}\right]=\frac{(n+1)!}{k)!(n+1-k)!}=C_{n+1}^k.$$

Pour prouver 2) on observe que c'est trivial pour $k=0$, puis on fixe $k>0$ et on montre 2) par récurrence sur n. C'est trivial pour $n=k$. Supposons enfin 2) vrai pour n et supposons que E ait $n+1$ éléments, qu'on prend égaux à $1,2,\ldots,n+1$ sans perte de généralité. Soit aussi E' l'ensemble des n premiers entiers. On partage alors les éléments de $\mathcal{P}_k(E)$ en deux catégories:

Catégorie 1: ceux qui ne contiennent pas $n+1$.

Catégorie 2: ceux qui contiennent $n+1$.

La catégorie 1 est égale à $\mathcal{P}_k(E')$ et a donc C_n^k éléments par l'hypothèse de récurrence. La catégorie 2 est en bijection avec $\mathcal{P}_{k-1}(E')$ (enlever $n+1$ à un membre de la catégorie 2 pour avoir un élément de $\mathcal{P}_{k-1}(E')$) et donc par l'hypothèse de récurrence a C_n^{k-1} éléments. La formule de Pascal montre alors que $\mathcal{P}_k(E)$ a C_{n+1}^k éléments et la récurrence est étendue.

Voici un exemple d'application du théorème précédent.

Proposition Anniversaires. n personnes sont réunies. Quelle est la probabilité que au moins deux d'entre elles aient le même anniversaire?

On formalise le problème en le simplifiant un peu: on ignore d'abord le problème du 29 février, et on postule donc que l'espace des observablesest $\Omega=F^E$ où E est l'ensemble des personnes et où F est l'ensemble des $p=365$ jours de l'année: on observe donc la fonction $f\in\Omega$ qui à chaque personne associe son anniversaire. On postule ensuite qu'on est dans le cas équiprobable, ce qui n'est qu'une approximation: il y a plus d'enfants conçus au printemps et en été qu'en novembre sous nos climats. Finalement, il est plus facile de calculer la probabilité du complémentaire A^c de l'évènement A "deux personnes au moins ont le même anniversaire", car c'est la probabilité que la fonction f soit injective. D'après le théorème 2.5 1), c'est

$$P(A^c)=\frac{1}{365^n}365(365-1)\cdots(365-n+1)=\prod_{k=1}^{n-1}(1-\frac{k}{365})=\exp\sum_{k=1}^{n-1}\log(1-\frac{k}{365}).$$

Si n n'est pas grand, une évaluation approximative de cette somme se fait en remplaçant $\log(1-x)$ par $-x$ et en utilisant la somme d'une progression arithmétique étudiée en Terminale

$$\sum_{k=1}^{n-1}k=\frac{1}{2}n(n-1)\sim n/2,$$

qui conduit à l'approximation $P(A^c)\sim\exp(-n/730)$. Pour voir par exemple pour quel n on a $P(A^c)\sim 1/2$ on prend $n\sim\sqrt{730\log 2}\sim 23$. Pour un calcul plus sérieux, on peut utiliser l'encadrement pour $0 < x < 1$:

$$-x - \frac{x}{2(1-x)} < \log(1-x) < -x - \frac{x}{2};$$

La majoration de droite se déduit du développement en série entière, celle de gauche se montre en étudiant la fonction $x + \frac{x}{2(1-x)} + \log(1-x)$. On a aussi besoin de la somme des premiers carrés:

$$\sum_{k=1}^{n-1} k^2 = \frac{1}{6}n(2n-1)(n-1) \sim n^3/3,$$

qui s'établit par récurrence. Si $x \leq (n-1)/365$, alors $-1/(1-x) \geq -365/(365-n+1)$. D'où l'encadrement :

$$-\frac{n(n-1)}{2}\frac{1}{365} - \frac{n(n-1)(2n+1)}{6}\frac{1}{2\times 365^2}\frac{365}{365-n+1} <$$

$$\log P(A^c) < -\frac{n(n-1)}{2}\frac{1}{365} - \frac{n(n-1)(2n+1)}{6}\frac{1}{2\times 365^2}.$$

Par exemple, si $n = 35$ on trouve $P(A^c) = 0{,}186\ldots$ Si 35 personnes sont réunies, la probabilité que deux d'entre elles au moins aient le même anniversaire est donc $0{,}813\ldots$

Le prochain théorème sert en particulier à résoudre le problème plus difficile du calcul du nombre de fonctions *surjectives*.

Théorème 2.7 (Principe d'inclusion-exclusion) Soit E un ensemble fini et soit f et g des fonctions réelles définies sur $\mathcal{P}(E)$ satisfaisant pour tout $A \subset E$:

$$f(A) = \sum_{B\subset A} g(B).$$

Alors pour tout $A \subset E$:

$$g(A) = \sum_{B\subset A} (-1)^{|A\setminus B|} f(B).$$

Démonstration Si $C \subset A \subset E$ notons

$$F(A,C) = \sum_{C\subset B\subset A} (-1)^{|A\setminus B|}.$$

Si $|A\setminus C| = n$, puisque il y a C_n^k parties de $A\setminus C$ de taille k on peut écrire $F(A,C) = \sum_{k=0}^{n}(-1)^k C_n^k$, qui est à son tour $(1+(-1))^n$ à cause de la formule du binôme de Pascal $(a+b)^n = \sum_{k=0}^{n} a^{n-k}b^k C_n^k$. Donc si $n > 0$, c'est-à-dire si $C \neq A$, on a

$F(A,C)=0$. Si $n=0$, c'est-à-dire si $C=A$ on a $F(A,C)=1$. Calculons alors le second membre de l'égalité à démontrer:

$$\sum_{B\subset A}(-1)^{|A\setminus B|}f(B)=\sum_{B\subset A}(-1)^{|A\setminus B|}\sum_{C\subset B}g(C)=$$

$$\sum_{C\subset B}g(C)\sum_{C\subset B\subset A}(-1)^{|A\setminus B|}=\sum_{C\subset B}g(C)F(A,C)=g(A).$$

La première égalité exploite le lien entre f et g, la seconde inverse les sommations par rapport aux indices de sommation B et C, la troisième résulte de la définition de $F(A,C)$, la quatrième du calcul de F précédent et fournit le résultat voulu.

Voici deux applications.

Proposition Nombre de fonctions surjectives. Si $|E|=n\geq|F|=p$, quel est le nombre de fonctions surjectives de E vers F?

Pour répondre on applique le théorème précédent aux fonctions f et g définies sur $\mathcal{P}(F)$ ainsi: si $A\subset F$, $f(A)=|A|^n$ est le nombre de fonctions de E vers F dont l'image est contenue dans A (on pourrait donc dire tout aussi bien les fonctions de E vers A); et $g(A)$ est le nombre de fonctions de E vers F dont l'image est exactement égale à A (on pourrait dire les fonctions de E vers A qui sont surjectives). On veut donc calculer $g(F)$.

Les hypothèses du théorème sont remplies, on a bien en effet $f(A)=\sum_{B\subset A}g(B)$. Par conséquent

$$g(F)=\sum_{B\subset F}(-1)^{|F\setminus B|}|B|^n=\sum_{k=0}^{p}C_p^k(-1)^{p-k}k^n.$$

Proposition Problème des rencontres. Si E a n éléments, combien y a-t-il de permutations σ de E sans point fixe, c'est-à-dire telles que pour tout $j\in E$ on ait $\sigma(j)\neq j$?.

On applique le théorème précédent aux fonctions f et g définies sur $\mathcal{P}(E)$ ainsi: si $A\subset E$, $f(A)=|A|!$ est le nombre de permutations de E telles que pour tout $j\in A^c$ on ait $\sigma(j)=j$, et $g(A)$ est le nombre de permutations de E telles que pour tout $j\in A^c$ on ait $\sigma(j)=j$ et pour tout $j\in A$ on ait $\sigma(j)\neq j$. On veut donc calculer $g(E)$.

Les hypothèses du théorème sont remplies, on a bien en effet $f(A)=\sum_{B\subset A}g(B)$. Par conséquent

$$g(E)=\sum_{B\subset E}(-1)^{|F\setminus B|}|B|!=$$

$$\sum_{k=0}^{n}C_n^k(-1)^{n-k}k!=$$

$$n!\sum_{k=0}^{n}(-1)^{n-k}\frac{1}{(n-k)!}=$$

$$n!\sum_{k=0}^{n}(-1)^{k}\frac{1}{k!}.$$

Si Ω est l'ensemble des permutations de E et si il est muni de la probabilité équiprobable, la probabilité pour qu'une permutation aléatoire soit sans point fixe est donc

$$\sum_{k=0}^{n}(-1)^{k}\frac{1}{k!},$$

soit approximativement $e^{-1}=0{,}367...$ si $n>6$.

Exercices sur 2.2.

1. Soit des entiers tels que $2\leq a\leq b\leq c$. On tire de façon équiprobable une partie de taille a de l'ensemble des $b+c$ entiers >0. Calculer la probabilité pour que 0 d'entre eux soient $>a$; pour que 2 d'entre eux exactement soient $>a$.
2. Deux dés non pipés sont marqués sur leurs six faces 1,2,2,3,3,4 et 1,3,4,5,6,8 respectivement. On jette une fois ces deux dés et on note par A_k l'évènement "la somme des points i du premier dé et des points j du second est k". Calculer pour $k=2,3,\ldots,12$ le nombre $P(A_k)$.
3. 12 méchantes fées se penchent sur le berceau des quintuplés et attribuent chacune au hasard à un enfant un défaut. Quel est la probabilité qu'il y ait au moins un enfant parfait?

3 Le schéma Succès-Echec.

Le schéma Succès-Echec fini. Si une expérience a deux issues, arbitrairement notées succès (S) et échec (E) et si on la répète n fois, ce qu'on observe est une suite de longueur n de S et de E. Pour modéliser cela, on introduit l'espace des observables$\Omega=\{E,S\}^n$ formé des 2^n suites $\omega=(\omega_1,\ldots,\omega_n)$ où les ω_j sont égaux à E ou S. On munit Ω de la tribu $\mathcal{P}(\Omega)$. Quant à la probabilité, on se fixe un nombre p tel que $0<p<1$ qui est la probabilité d'un succès si on n'effectue qu'une fois l'expérience. Introduisons alors l'importante quantité $X(w)$ définie ainsi: si $\omega=(\omega_1,\ldots,\omega_n)\in\Omega$ alors $X(w)$ désigne le nombre de succès que comprend la suite ω. Par exemple, $X(SSES)=3$, $X(EEEE)=0$. Pour $\omega\in\Omega$ tel que $X(\omega)=k$ on définit alors $P(\{\omega\})=p^k(1-p)^{n-k}$; Comme tout évènement $A\in\mathcal{P}(\Omega)$ est réunion de singletons $\{\omega\}$ deux à deux disjoints, cela suffit à définir $P(A)$ et donc la probablité P sur $(\Omega,\mathcal{P}(\Omega))$.

Parmi ces évènements, les plus importants sont les $\{X=k\}$ (ceci est une sténographie que nous utiliserons souvent pour écrire brièvement l'évènement $\{\omega\in$

$\Omega\ ;\ X(\omega)=k\}$). Voici leur probabilité:

Proposition 2.9 Pour le schéma Succès Echec finiassocié à la probabilité p d'un succès, si X est le nombre de succès en n expériences, alors

$$P(X=k)=C_n^k p^k(1-p)^{n-k}.$$

Démonstration Notons $A=\{\omega\in\Omega\ ;\ X(\omega)=k\}$. Définissons l'application de A dans $\mathcal{P}_k(\{1,2,\ldots,n\})$ par $\omega\mapsto\{j\ ;\ \omega_j=S\}$. Il est clair que c'est une bijection; donc d'après le Théorème 2.5 b), $|A|=C_n^k$. Enfin puisque tous les $\{\omega\}$ contenus dans A ont la même probabilité $p^k(1-p)^{n-k}$ on obtient

$$P(A)=\sum_{\omega\in A}P(\{\omega\})=|A|p^k(1-p)^{n-k}=C_n^k p^k(1-p)^{n-k}.$$

Le schéma Succès-Echec infini. Il s'agit ensuite de modéliser le cas où on veut effectuer un nombre arbitraire d'expériences: par exemple on peut vouloir répéter les essais jusqu'à ce qu'apparaisse 4 succès consécutifs. Une telle modélisation est impossible avec le schéma fini ci dessus, et on prend alors pour espace Ω des observablesl'ensemble $\{E,S\}^{\mathbf{N}^*}$ des suites infinies de S et de E, en notant par $\mathbf{N}^*$ l'ensemble des entiers >0. Il est clair que Ω est en bijection avec les parties de $\mathbf{N}^*$, et donc d'après la proposition 2.2 Ω n'est pas dénombrable. Cela cause une sérieuse difficulté en ce qui concerne la construction de l'espace de probabilité correspondant. On construit la tribu $\mathcal{A}$ et la probabilité P par un procédé d'approximation que nous décrivons maintenant.

Fixons l'entier n et définissons $\Omega'=\{E,S\}^{\{1,\ldots,n\}}$ et $\Omega''=\{E,S\}^{\{n+1,n+2,\ldots\}}$, de sorte que $\Omega=\Omega'\times\Omega''$, et définissons la tribu suivante de parties de Ω :

$$\mathcal{A}_n=\{A\times\Omega''\ ;\ A\in\mathcal{P}(\Omega')\}.$$

Intuitivement, les évènements de $\mathcal{A}_n$ sont les évènements ne dépendant que de ce qui s'est passé jusqu'à l'instant n. En particulier, nous avons $\mathcal{A}_n\subset\mathcal{A}_{n+1}$.

Si $\omega'=(\omega_1,\ldots,\omega_n)\in\Omega'$ comprend k succès, définissons la probabilité $P_n(\{\omega'\}\times\Omega'')=p^k(1-p)^{n-k}$. Cela permet donc de définir la probabilité P_n sur $\mathcal{A}_n$. L'espace de probabilité $(\Omega,\mathcal{A}_n,P_n)$ est presque identique à l'espace du schéma Succès Echec fini décrit ci dessus.

Maintenant, notons

$$\mathcal{A}'=\cup_{n\geq 1}\mathcal{A}_n.$$

La famille $\mathcal{A}'$ n'est pas une tribu, car ce n'est pas fermé pour la réunion dénombrable. Voici un contre exemple. Soit A_n l'ensemble des suites ω infinies comprenant au moins un succès à l'instant n ou avant. Alors A_n est dans $\mathcal{A}_n$ et donc dans $\mathcal{A}'$. Pourtant $A=\cup_{n\geq 1}A_n$ n'est pas dans $\mathcal{A}'$. En effet A est l'ensemble des suites ω infinies comprenant au moins un succès. Mais il n'existe pourtant aucun n tel que $A\in\mathcal{A}_n$, et donc $A\notin\mathcal{A}'$. Réaliser cette chose subtile fait progresser dans la compréhension de la théorie. On définit alors la tribu $\mathcal{A}$ sur Ω comme la plus petite tribu contenant $\mathcal{A}'$.

Pour définir enfin la probabilité P sur $\mathcal{A}$, on fait l'observation essentielle suivante: on a non seulement $\mathcal{A}_n \subset \mathcal{A}_{n+1}$, mais de plus la restriction de P_{n+1} au sous ensemble $\mathcal{A}_n$ de $\mathcal{A}_{n+1}$, qui était le domaine de définition de P_{n+1}, coincide avec P_n. Par conséquent, il existe une fonction universelle P' définie sur $\mathcal{A}'$ telle que pour tout $A \in \mathcal{A}'$ on ait $P'(A) = P_n(A)$ pour tous les n tels que $A \in \mathcal{A}_n$. A partir de ce point, les choses cessent d'être élémentaires, et nous sommes obligés d'admettre le théorème suivant, dont la démonstration est donnée en troisième année d'université:

Théorème 2.10 Il existe une et une seule probabilité P sur $\mathcal{A}$ telle que pour tout $A \in \mathcal{A}'$ on ait $P(A) = P'(A)$.

On peut ainsi démontrer l'idée intuitive qu'un évènement de probabilité strictement positive, même petite, finit toujours par arriver. Plus précisément, si A est l'ensemble des $\omega \in \Omega$ comprenant au moins un succès, alors $P(A) = 1$. En effet, si B_n est l'ensemble des $\omega \in \Omega$ comprenant au moins un succès avant l'instant n ou à l'instant n, alors $A = \cup_{n \geq 1} B_n$ et $B_n \subset B_{n+1}$. Par continuité monotone (Th. 1.3, (2)) on a donc $\lim P(B_n) = P(A)$. Comme $P(B^c) = (1-p)^n$ tend vers 0, on a le résultat. Plus généralement on peut montrer que toute séquence a finie donnée à l'avance (par exemple SSEESSEESSEESSEE, ou le codage en binaire d'une fable de La Fontaine) finira par arriver. Plus précisément:

Théorème 2.11 Soit $a = (a_1, \ldots, a_n) \in \{E,S\}^n$ une suite fixée de longueur n de succès et d'échecs, et soit

$$A = \{\omega \in \Omega\,;\ \text{il existe } N \geq 0 \text{ avec } \omega_{N+1} = a_1, \ldots, \omega_{N+n} = a_n\}.$$

Alors $P(A) = 1$.

Démonstration Soit k le nombre de S dans la suite a. Notons

$$A_N = \{\omega \in \Omega\,;\ \omega_{N+1} = a_1, \ldots, \omega_{N+n} = a_n\}.$$

Alors $P(A_N) = p^k(1-p)^{n-k}$ par définition de P. Introduisons $B_m = \cup_{j=0}^{m-1} A_{jn}$. Alors $B_m \subset B_{m+1}$ et

$$A = \cup_{N \geq 0} A_N \supset B = \cup_{m \geq 0} B_m.$$

On a de plus

$$P(B_m^c) = P(\cap_{j=0}^{m-1} A_{jn}^c) = (1 - p^k(1-p)^{n-k})^m \to_{m \to \infty} 0.$$

Par continuité monotone, on a donc $P(B^c) = 0$. D'où $1 = P(B) \leq P(A) = 1$.

4 Le cas où $\Omega = \mathbb{R}$.

Ce cas est naturellement le plus important de tous. La tribu mise sur $\mathbb{R}$ est la tribu de Borel $\mathcal{B}$ définie à la section 1 comme la plus petite tribu contenant les intervalles (ouverts, fermés, semi ouverts, demi droites) Parmi ses éléments, les boréliens, les seuls qu'on aura concrètement à manipuler sont les réunions d'intervalles.

Pour décrire les probabilités sur $(\mathbb{R},\mathcal{B})$, introduisons une définition importante:

Définition Soit F une fonction de $\mathbb{R}$ dans $\mathbb{R}$. On dit que F est une *fonction de répartition* si elle satisfait aux trois propriétés suivantes:

- F est croissante (au sens large);
- $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$;
- F est continue à droite en tout point x, c'est-à-dire $\lim_{h\searrow 0} F(x+h) = F(x)$.

On a alors le théorème fondamental suivant:

Théorème 2.12 Soit P une probabilité sur $(\mathbb{R},\mathcal{B})$. Soit F_P la fonction réelle définie par

$$F_P(x) = P(]-\infty,x]).$$

Alors F_P est une fonction de répartition. Inversement, si F est une fonction de répartition, alors il existe une et une seule probabilité sur $(\mathbb{R},\mathcal{B})$ telle que $F_P = F$.

Démonstration Si $x < y$, alors $A =]-\infty,x] \subset B =]-\infty,y]$, et donc $F_P(x) = P(A) \leq P(B) = F_P(y)$. Ensuite, si (x_n) tend vers $-\infty$ en décroissant et si $A_n =]-\infty,x_n]$, alors $A_n \supset A_{n+1}$ et $\cap_{n\geq 1} A_n = \emptyset$; par continuité monotone $P(A_n)$ tend vers 0. Donc $\lim_{n\to\infty} F_P(x_n) = 0$. Comme ceci est vrai quelle que soit la suite (x_n) tendant vers $-\infty$ en décroissant, on en déduit $\lim_{x\to-\infty} F_P(x) = 0$. De même, si (y_n) tend vers ∞ en croissant et si $B_n =]-\infty,y_n]$, alors $B_n \subset B_{n+1}$ et $\cup_{n\geq 1} B_n = \mathbb{R}$; par continuité monotone $P(B_n)$ tend vers $P(\mathbb{R}) = 1$ et on a $\lim_{y\to+\infty} F_P(y) = 1$.

Enfin, si $h_n \searrow 0$, soit $C_n =]\infty,x+h_n]$. Alors $C_n \supset C_{n+1}$ et $\cap_{n\geq 1} C_n =]\infty,x]$. Par continuité monotone on a donc $\lim_{n\to+\infty} F(x+h_n) = F_P(x)$, d'où la continuité à droite annoncée de la fonction F_P.

Nous admettrons la réciproque, qui est la partie difficile.

Commentaires: Ce résultat est assez rassurant: bien qu'on connaisse mal la tribu $\mathcal{B}$, et donc les probabilités définies dessus, il y a en fait bijection entre l'ensemble de toutes les probabilités sur $\mathbb{R}$ et l'ensemble moins abstrait de toutes les fonctions de répartition. Mais la démonstration complète est réservée à la 3 ème année.

La fonction de répartition permet de calculer les probabilités de tous les intervalles. Pour simplifier, adoptons la notation pour la limite à gauche en x de la fonction croissante F:

$$F(x-0) = \lim_{h\nearrow 0} F(x+h).$$

Proposition 2.13 Soit F la fonction de répartition d'une probabilité P sur $(\mathbb{R},\mathcal{B})$. Alors

- $P(]-\infty,x[) = F(x-0)$, $P(]x,+\infty[) = 1-F(x)$, $P([x,+\infty[) = 1-F(x-0)$.

– Pour $a \leq b$, $P(]a,b]) = F(b) - F(a)$, $P([a,b[) = F(b-0) - F(a-0)$.
– $P(]a,b[) = F(b-0) - F(a)$, $P([a,b]) = F(b) - F(a-0)$ et en particulier

$$P(\{a\}) = F(a) - F(a-0).$$

Démonstration La première égalité s'obtient en considérant $A_n =]-\infty,x+h_n]$, où h_n est < 0 et croŇt vers 0. Alors $A_n \subset A_{n+1}$ et $\cup_{n \geq 1} A_n =]-\infty,x[$. Par convergence monotone l'égalité s'ensuit. Les deux suivantes s'obtiennent par passage au complémentaire. La suivante découle de l'égalité

$$]-\infty,b] =]-\infty,a] \cup]a,b],$$

et du fait que au second membre les deux ensembles sont disjoints. De même

$$]-\infty,b[=]-\infty,a[\cup [a,b[$$

fournit l'égalité suivante grâce à la première égalité de la liste. Laissons les dernières en exercice.

Donnons maintenant des exemples de fonctions de répartition

Définition **Fonctions de répartition à densité.** Soit f une fonction positive définie sur $\mathbb{R}$ qui ait des discontinuités au plus en un nombre fini de points $a_1 < a_2 < \cdots < a_N$ et qui soit telle que les intégrales $\int_{a_i}^{a_{i+1}} f(x)dx$ convergent et satisfassent

$$\sum_{i=0}^{N} \int_{a_i}^{a_{i+1}} f(x)dx = 1,$$

avec la convention $a_0 = -\infty$ et $a_{N+1} = +\infty$.

On définit alors la fonction F par $F(x) = \int_{-\infty}^{x} f(t)dt$. Il est clair que F est une fonction de répartition. Ici, elle est de plus continue et, d'après le théorème fondamental du calcul intégral, elle satisfait $F'(x) = f(x)$ pour tout $x \notin \{a_1, \ldots, a_N\}$. La fonction f s'appelle alors la *densité* de la fonction de répartition F.

Par exemple $f_1(x) = \frac{1}{2}e^{-|x|}$, $f_2(x) = \frac{1}{\pi}\frac{1}{1+x}$, $f_3(x) = 0$ si $x \leq 0$ et $f_3(x) = \frac{1}{\sqrt{\pi x}}e^{-x}$ si $x > 0$, qu'il est plus rapide de définir par

$$f_3(x) = \frac{1}{\sqrt{\pi x}} e^{-x} \mathbf{1}_{]0,\infty[}(x),$$

où $\mathbf{1}_E(x) = 1$ si $x \in E$ et $\mathbf{1}_E(x) = 0$ sinon: la fonction $\mathbf{1}_E$ s'appellera désormais *l'indicateur* de l'ensemble E. Dernier exemple:

$$f_4(x) = \mathbf{1}_{[0,1]}(x).$$

Dans ces exemples, $N=0$ pour f_1 et f_2, $N=1$ pour f_3 et $N=2$ pour f_4.

Il est important de ne pas confondre les deux fonctions F et f. Pour les exemples ci dessus de densités, les fonctions de répartition correspondantes seront respectivement

$$F_1(x)=\frac{1}{2}e^x \text{ pour } x\leq 0,\, F_1(x)=1-\frac{1}{2}e^{-x},$$

$$F_2(x)=\frac{1}{2}+\frac{1}{\pi}\arctan x,$$

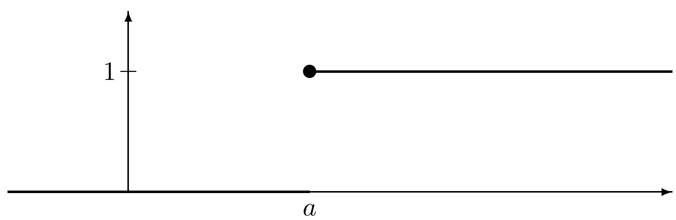
$$F_4(x)=0 \text{ pour } x\leq 0,\, F_4(x)=x \text{ pour } 0\leq x\leq 1,\, F_4(x)=1 \text{ pour } 1\leq x,$$

($F_3(x)$ ne peut s'exprimer de façon élémentaire).

Définition La probabilité δ_a de Dirac. Si a est un réel, il s'agit de la probabilité sur $\mathbb{R}$ définie par $\delta_a(A)=0$ si $a\notin A$, et $\delta_a(A)=1$ si $a\in A$. Appliquant ceci à $A=]-\infty,x]$, on obtient la fonction de répartition

$$F_{\delta_a}(x)=0 \text{ pour } x<a,\, F_{\delta_a}(x)=1 \text{ pour } a\leq x.$$

Voici son graphe



Si $a=0$, cette fonction s'appelle l'échelon de Heaviside. Les travaux de 1894 de cet ingénieur électricien sont à la source de la théorie moderne des distributions. Cette théorie permet par exemple de donner un sens à la dérivation de la fonction ci dessus: c'est la probabilité de Dirac δ_a qui jouerait alors le rôle de la dérivée.

Définition Probabilité discrète sur un nombre fini de points. Soit N un entier >0, soit $a_1<a_2<\cdots<a_N$ des réels et soit $p_1,\ldots,p_N$ des nombres positifs tels que $p_1+\cdots+p_N=1$. On considère la probabilité sur $\mathbb{R}$ définie par

$$P=p_1\delta_{a_1}+\cdots+p_N\delta_{a_N}.$$

En d'autres termes, si A est un borélien:

$$P(A)=p_1\delta_{a_1}(A)+\cdots+p_N\delta_{a_N}(A)=\sum_{j;a_j\in A}p_j.$$

En particulier, si $A=]-\infty,x]$, on obtient la fonction de répartition

$$F_P(x)=\sum_{j;a_j\leq x}p_j,$$

dont le graphe est celui d'une fonction en escalier croissante, où le saut en a_j est égal à p_j. Ce cas revient un peu au cas où Ω n'avait qu'un nombre fini de points, puisqu'ici P est concentrée sur $\{a_1,\ldots,a_N\}$.

Si on remplace la suite finie précédente par un ensemble dénombrable de $\mathbb{R}$, l'extension est facile.

Définition Probabilité discrète. On s'intéresse à l'ensemble dénombrable formé des points d'une suite (a_n) telle que $a_1 < a_2 < \cdots < a_n < \cdots$ et soit p_n des nombres positifs tels que $\sum_1^\infty p_n = 1$. On formera la probabilité P définie pour tout Borélien A par

$$P(A) = \sum_1^\infty p_n \delta_{a_n}(A),$$

dont la fonction de répartition est en escalier croissante vec une infinité de points de discontinuités.

Définition Type mixte. On rencontre un peu rarement des fonctions de répartition de la forme $F = \lambda G + (1-\lambda)H$ où G est une fonction de répartition à densité, comme vu à l'exemple 1, où H est une fonction de répartition d'une probabilité discrète, comme vu aux exemples 2, 3 ou 4, et où $0 < \lambda < 1$. Si H a une discontinuité en a de saut p, alors F a une discontinuité en a de saut $(1-\lambda)p$.

Exercices sur 2.4.

1. Calculer la densité des fonctions de répartition suivantes:
   $F_1(x) = 0$ si $x \leq 0$ et $F_1(x) = 1 - \exp(-x)$ si $x > 0$;
   $F_2(x) = 0$ si $x \leq 1$ et $F_2(x) = 1 - \frac{1}{x^a}$ si $x > 1$ (avec a>0).
2. Calculer la fonction de répartition de la densité suivante:
   $f(x) = 1/2$ si $-2 < x < -1$, $f(x) = 1/2$ si $1 < x < 2$, et 0 ailleurs.
3. On note par $[x]$ la partie entière du nombre réel x, c'est-à-dire l'entier n tel que $n \leq x < n+1$. Par exemple $[\sqrt{2}] = 1$, $[-\sqrt{2}] = -2$, $[3] = 3$. On considère la probabilité discrète de fonction de répartition $F(x) = 0$ si $x < 0$ et $F(x) = 1 - \frac{1}{2^{[x]+1}}$ si $x \geq 0$. Tracer le graphe de F. Calculer les probabilités des évènements suivants:
   $A_1 = \{0\}$, $A_2 = \{1,2\}$, $A_3 = \{4,5,\ldots\}$.