

Chapitre 1

Introduction aux signaux aléatoires

1.1 Historique

Le *traitement statistique du signal* (*statistical signal processing*) est un domaine scientifique, dont on peut faire remonter l'origine à la deuxième guerre mondiale, avec les travaux de Kolmogorov *Theory of filtering and prediction* (1941) et de Wiener *Extrapolation, interpolation and smoothing of stationary time series* (publié en 1949). Dans l'introduction de son livre, Norbert Wiener (1894-1964) écrit :

« This book represents an attempt to unite the theory and practice of two fields of work which are of vital importance in the present emergency, and which have a complete natural methodological unity, but which have up to the present drawn their inspiration from two entirely distinct traditions, and which are widely different in their vocabulary and the training of their personnel. These two fields are those of time series in statistics and of communication engineering. »

Le traitement statistique du signal est issu des statistiques et des sciences de l'ingénieur en communication. Il trouve aujourd'hui des applications dans toutes les disciplines scientifiques et sa théorie est étroitement liée à celle de l'information et à différentes branches de la physique et des mathématiques. L'étude et l'observation dans la nature des systèmes de communication, c'est-à-dire d'échanges ou de stockages d'informations, utilisent les statistiques, dont l'objet est de rechercher et de quantifier des grandeurs invariantes, cachées dans le fatras apparent de phénomènes attribués au hasard.

Les premiers succès, permettant de quantifier et d'énoncer des propriétés d'invariance pour toute une classe de phénomènes attribués aux effets du hasard, remontent au XVII^e siècle, avec le calcul combinatoire développé par Pierre Fermat (1601–1665) et Blaise Pascal (1623–1662) vers 1654. C'est la théorie des jeux de hasard, qui modélise les situations où le résultat d'une expérience peut prendre une valeur quelconque dans un ensemble fini de valeurs possibles, chaque valeur particulière ayant autant de chance qu'une autre d'être obtenue. Par exemple lancer un dé, jouer à la roulette, au loto, distribuer des cartes entre plusieurs joueurs après les avoir battues, sont des jeux de hasard.

Dans ces situations, le hasard est « pur », c'est-à-dire qu'il y a équiprobabilité entre les différentes issues possibles. Par exemple pour un lancer de dé, les équations décrivant le mouvement restent invariantes par changement de coordonnées qui conserve le cube. Ou encore dans le jeu de la roulette, les équations décrivant le mouvement du plateau tournant restent invariantes par changement de coordonnées qui conserve les 36 cases du plateau. Symétrie et équiprobabilité sont deux facettes d'une même idéalisation de la réalité.

En 1812, avec la parution du *Traité analytique des probabilités*, Pierre-Simon Laplace (1749–1827) étend le calcul des probabilités à des situations plus générales, où l'ensemble des résultats possibles décrit une surface S ou un volume V ou plus généralement un hypervolume dans un espace de dimension n . Voici, comment il introduit la théorie des hasards dans son livre *La théorie analytique des probabilités*, publié en 1819 :

« La théorie des hasards consiste à réduire tous les événements du même genre à un certain nombre de cas également possibles et à déterminer le nombre de cas favorables à l'événement dont on cherche la probabilité. Le rapport de ce nombre à celui de tous les cas possibles est la mesure de cette probabilité, qui n'est ainsi qu'une fraction dont le numérateur est le nombre de cas favorables, et dont le dénominateur est le nombre de tous les cas possibles. »

Ainsi, par exemple quand l'ensemble des résultats possibles est une surface S ou un volume V , la probabilité d'un événement E est définie par le rapport :

$$P(E) = \frac{\iint_{\text{cas favorables à } E} dx dy}{\iint_S dx dy}$$

ou

$$P(E) = \frac{\iiint_{\text{cas favorables à } E} dx dy dz}{\iiint_V dx dy dz}.$$

Cette définition de la probabilité d'un événement E offre une interprétation géométrique au calcul des probabilités. Pour un hypervolume V donné, le tirage d'un point dans V est l'effet du « pur » hasard, quand chaque point de V a la même probabilité d'être choisi. Cela se traduit par une invariance par translation : la probabilité pour que la Fortune choisisse un point dans un sous-ensemble E de V ne dépend que du volume de E et pas de sa position relative dans V . Pour que l'interprétation géométrique du calcul analytique des probabilités ne soit pas fortuite, il faut que les lois de la théorie des probabilités se conservent par changement de coordonnées. Ce qui a pour conséquence que, plus généralement, la probabilité d'un événement E doit être définie par :

$$P(E) = \frac{\iint_{\text{cas favorables à } E} \varphi(x, y) dx dy}{\iint_S dx dy}$$

ou

$$P(E) = \frac{\iiint_{\text{cas favorables à } E} \varphi(x, y, z) dx dy dz}{\iiint_V dx dy dz},$$

où la fonction φ traduit un changement de coordonnées. Il n'en demeure pas moins que pour une bonne compréhension d'un phénomène attribué au hasard, il est toujours très utile de trouver le système de coordonnées dans lequel φ est constante, c'est-à-dire les grandeurs indépendantes sur lesquelles porte l'effet du pur hasard. Les symétries associées au phénomène y apparaîtront beaucoup plus clairement. Voici ce qu'écrit J. Harthong [12] à ce sujet :

« La caractéristique du pur hasard *est* l'équiprobabilité. Si des épreuves ne sont pas équiprobables, c'est que certaines sont favorisées et donc qu'une cause, qui n'est pas le hasard, agit secrètement. Il ne s'agit pas là d'un choix philosophique : si une étude statistique établit que le cancer du pancréas est plus fréquent dans la haute vallée de l'Isère que partout ailleurs, on en déduira qu'une cause, agissant dans cette haute vallée et non ailleurs (ou agissant plus fortement dans cette haute vallée qu'ailleurs) est responsable de cette sur-fréquence. Même si après de longues recherches cette cause reste introuvable, il n'y aura pas de querelle philosophique sur la question de savoir si la Fortune *est* ou *n'est pas* aveugle ; on ne parlera du hasard que si la fréquence est distribuée uniformément, car c'est là le sens même du mot hasard. »

Dans *La théorie analytique des probabilités*, Laplace donne aussi les premiers éléments de la théorie des erreurs en démontrant que la méthode des moindres carrés, introduite empiriquement par Carl Friedrich Gauss (1777–1855) et Adrien Marie Le Gendre (1752–1833), est optimale. Il relie le calcul des probabilités aux méthodes des statisticiens. Au XVIII-ème siècle, ces derniers ont observé empiriquement la loi des grands nombres : la proportion de personnes mortes pendant une année d'une maladie particulière fluctue peu d'une année sur l'autre. Plus généralement, les moyennes statistiques tendent à ne plus beaucoup fluctuer dès que le nombre de mesures est assez grand. Le théorème de la limite centrale, qui permet de donner un intervalle de confiance pour l'estimation statistique d'une moyenne et qui est dû à Abraham De Moivre (1667–1754), remonte à 1718.

Tout au long du XIX-ème siècle et jusqu'au début du XX-ème siècle, le calcul des probabilités s'est étendu à presque tous les domaines scientifiques en particulier en physique, avec la mécanique statistique et la théorie cinétique des gaz de Ludwig Boltzmann (1844–1906), la théorie du mouvement brownien¹, la mécanique quantique (1924–25), la théorie de l'information (\simeq 1945–50) ...

Le calcul des probabilités et les statistiques sont aujourd'hui très présents en sciences de l'ingénieur, en sciences humaines, en économie, en biologie et en médecine. C'est dire leur importance !

¹ Albert Einstein (1879–1955) donne en 1905 une explication physique à l'agitation incessante de grains de pollen en suspension dans l'eau, phénomène microscopique observé en 1827 par le botaniste anglais Robert Brown (1773–1858).

Au fur et à mesure que le calcul des probabilités s'installait dans tous les domaines scientifiques, le support intuitif en terme de jeux de hasard (nombre de cas favorables sur nombre de cas possibles) devenait de plus en plus critiquable et douteux, en particulier avec l'étude des fonctions aléatoires qui commence vers 1930. En effet, quand la collection de tous les résultats possibles d'une expérience est un ensemble de fonctions, par exemple l'ensemble \mathbb{R}^N des fonctions de \mathbb{N} dans \mathbb{R} , la notion d'hypervolume dans un tel ensemble n'est ni intuitive, ni naturelle.

Pour éliminer des paradoxes, qui apparaissaient quand on étendait à des ensembles infinis des notions claires et naturelles quand elles étaient définies dans des ensembles finis, il a fallu reconstruire la théorie des probabilités. Ce fut l'objet d'importantes recherches dans les années 1925–35. Plusieurs théories controversées virent le jour. Citons celle de Von Mises qui, fondée sur la loi empirique des grands nombres², avait l'avantage d'être intuitive mais le défaut majeur d'être une théorie pauvre en déductions, c'est-à-dire permettant de déduire logiquement très peu de résultats à partir des axiomes de base. C'est finalement le point de vue d'Andréï Nikolaïevitch Kolmogorov (1903–1987), exposé en 1933 dans *Fondement axiomatique des probabilités* qui (n'ayant pas ce défaut) fut adopté et qui perdure aujourd'hui. Cette nouvelle théorie des probabilités trouve ses fondements dans la théorie de la mesure et de l'intégration. Elle permet de déduire des axiomes de base (peu nombreux) tous les résultats empiriques observés par les statisticiens et elle offre un cadre consistant dans lequel peuvent être étudiés avec les mêmes outils les jeux de hasard (espaces probabilisés finis, variables aléatoires) et les fonctions aléatoires (collections infinies de variables aléatoires, mouvement brownien, etc.).

Le support intuitif en termes de jeux de hasard de la théorie des probabilités est souvent mal adapté aux problèmes de communications (c'est-à-dire d'échanges ou de stockage d'informations) et il est préférable de l'interpréter en terme d'informations, de quantité d'information qu'apporte le résultat d'une expérience aléatoire ou de l'information que l'on peut tirer de l'observation d'une grandeur associée à une expérience aléatoire pour estimer une autre grandeur associée à la même expérience ... Voici ce qu'écrit Léon Brillouin (1881–1969) dans l'introduction de son livre *La science et la théorie de l'information* (1959) pour définir le concept d'information :

« Envisageons un problème qui comporte un certain nombre de réponses possibles lorsque l'on ne possède pas d'informations particulières sur la situation présente. Si l'on parvient à obtenir quelques informations sur le problème, le nombre des réponses possibles se trouve diminué et une information totale peut même conduire à une seule réponse possible. L'information est une fonction du rapport des réponses possibles après et avant qu'on l'ait reçue. »

Les fonctions aléatoires se rencontrent dans la nature en différentes situations. Nous allons en voir quatre exemples types avant d'en donner une définition

²Dans la théorie des probabilités de Von Mises, la probabilité d'un événement E est définie par

$$P(E) = \lim_{N \rightarrow +\infty} \frac{n_E}{N},$$

où n_E est le nombre de fois que l'événement E est réalisé parmi un nombre N de réalisations indépendantes de l'expérience aléatoire.

formelle.

1.2 Exemples de signaux aléatoires

1.2.1 Marche aléatoire ou suite infinie de piles ou faces

Introduction

Une *marche* (ou *promenade*) *aléatoire* à une dimension est le mouvement d'un point se déplaçant par sauts successifs le long d'un axe, soit vers l'avant, soit vers l'arrière, le choix entre ces deux possibilités étant fait au hasard. Notons $0 < t_1 < \dots < t_n < \dots$ les instants auxquels se produisent les sauts successifs $\alpha_0, \alpha_1, \dots, \alpha_n, \dots$ et pour $n \geq 1$ notons X_n la position du point à l'instant t_n , de sorte que pour $n \geq 1$ la relation

$$\alpha_n = X_{n+1} - X_n \quad (1.1)$$

est satisfaite, et enfin notons X_0 la position du point à l'instant initial t_0 . Nous supposerons, sans perte de généralité, que $X_0 = 0$ et $t_0 = 0$. D'autre part, nous n'étudierons pas la cinématique entre les positions X_0, X_1, \dots et nous ferons comme si le point se déplaçait à vitesse constante entre deux positions successives. À chaque saut, le sens (avant ou arrière) dépend du résultat d'un tirage au sort, par exemple en lançant une pièce : pile on avance et face on recule. Dans le cas le plus général, les longueurs $|\alpha_n|$ des sauts et les durées $t_{n+1} - t_n$ entre deux pas successifs varient (elles sont modélisées par des variables aléatoires). Dans le cas contraire, quand

$$\exists \alpha > 0, \exists \Delta > 0, \forall n \in \mathbb{N}, \quad \alpha_n = \pm \alpha \quad \text{et} \quad t_{n+1} - t_n = \Delta,$$

la marche aléatoire est dite *uniforme*, on peut alors supposer sans perte de généralité que $\alpha = 1$ et $\Delta = 1$.

Il est possible d'étendre la définition de marches aléatoires dans des espaces à plusieurs dimensions, par exemple en dimension 2 ou 3. Dans ce cas, à chaque saut, l'ensemble des directions possibles n'est pas réduit à deux éléments, mais est constitué d'un cercle (en dimension 2) ou d'une sphère (en dimension 3), sur lequel est donnée une distribution de probabilité, avec aussi une distribution de probabilité pour la longueur du saut. Une marche aléatoire en dimension trois est un modèle mathématique du *mouvement brownien* : l'agitation incessante de grains de pollen en suspension dans l'eau est due aux chocs avec des molécules dont la répartition des vitesses (direction et intensité) est aléatoire. Son étude mathématique est sensiblement plus difficile que celle de la marche aléatoire uniforme à une dimension, mais cette dernière permet déjà de voir beaucoup de propriétés du mouvement brownien.

La marche aléatoire uniforme à une dimension permet également de modéliser une suite, aussi longue soit-elle, de lancers d'une pièce et de donner ainsi une interprétation géométrique au jeu de pile ou face. Voici un exemple de graphes d'une marche aléatoire, obtenus par simulation :

```
> piece := 2*rand(0..1)-1:
```

```
> X[0]:=0:for n to 512 do X[n]:=X[n-1]+piece() od:
> plot([seq([n,X[n]],n=0..512)]);
```

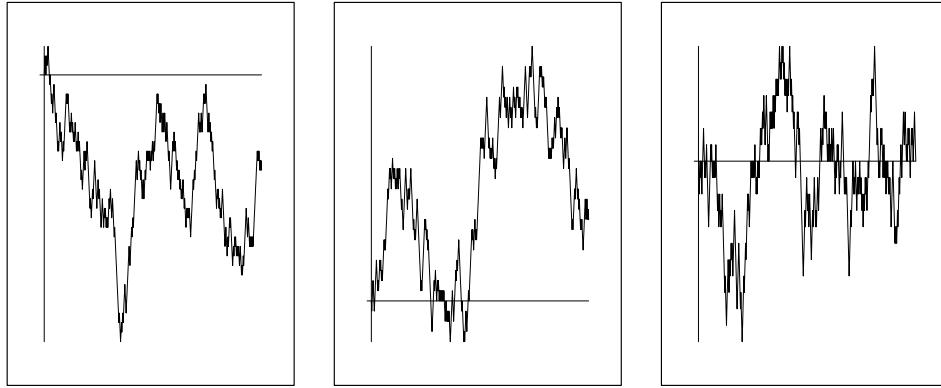


Figure 1.1: Différentes trajectoires d'une marche aléatoire à une dimension.

L'espace des épreuves pour une marche aléatoire uniforme à une dimension et à n pas (c'est-à-dire constituée de n sauts) est l'ensemble de toutes les trajectoires possibles. Chaque trajectoire étant parfaitement déterminée par les n sens (avant ou arrière) associés, on peut la représenter par un n -uplet de signes + ou -. Autrement dit, l'espace des épreuves Ω est isomorphe à celui du jeu de pile ou face à n lancers, son cardinal vaut 2^n . Pour $x \in \mathbb{Z}$, recherchons la probabilité qu'après n sauts le point soit en la position x . Lors des n sauts, le point en a fait p en avant et q en arrière avec la condition $p+q = n$. L'événement $X_n = x$ se produit si et seulement si $p-q = x$. L'ordre dans lequel les sauts ont eu lieu n'intervient pas, seuls comptent les nombres p et q . Nous en déduisons la probabilité

$$P(X_n = x) = \begin{cases} 2^{-n} C_n^{(n+x)/2} & \text{si } n \text{ et } x \text{ ont la même parité} \\ 0 & \text{sinon,} \end{cases} \quad (1.2)$$

avec la convention

$$C_n^p = 0 \quad \text{si} \quad p < 0 \quad \text{ou} \quad p > n.$$

Intéressons-nous maintenant à l'événement $X_m = x$ pour $0 < m < n$. Cet événement se produit si et seulement s'il existe deux nombres entiers p et q satisfaisant aux conditions :

$$\begin{cases} p+q = m \\ p-q = x. \end{cases}$$

Les nombres p et q correspondent aux nombres de sauts en avant et en arrière pendant les m premiers pas et la trajectoire du point après le m -ème saut n'influe pas. Il y a donc

$$2^{n-m} C_m^{(m+x)/2}$$

trajectoires différentes qui vérifient l'événement $X_m = x$ quand m et x ont la même parité et 0 sinon. On en déduit la probabilité

$$P(X_m = x) = \begin{cases} 2^{-m} C_m^{(m+x)/2} & \text{si } m \text{ et } x \text{ ont la même parité} \\ 0 & \text{sinon.} \end{cases}$$

Le résultat ne dépend pas de n : quand on s'intéresse à un événement qui se produit à l'instant m ou avant, il est inutile de tenir compte de ce qui se passe après cet instant.

Le problème du retour à zéro

D'après le calcul précédent, la probabilité de l'événement $X_{2\ell} = 0$ (retour à zéro après 2ℓ sauts) vaut³

$$P(X_{2\ell} = 0) = 2^{-2\ell} \frac{2\ell!}{(\ell!)^2} \quad \left(\simeq \frac{1}{\sqrt{\pi\ell}} \quad \text{pour } \ell \gg 1 \right).$$

Nous nous intéressons maintenant au problème du premier retour à zéro et nous recherchons la probabilité du premier retour à zéro en 2ℓ pas :

$$R_{2\ell} : X_{2\ell} = 0 \quad \text{et} \quad X_m \neq 0 \quad \text{si} \quad 0 < m < 2\ell. \quad (1.3)$$

Suivons un raisonnement géométrique donné par J. Harthong [12] et qu'il attribue à Désiré André (1887). L'ensemble des trajectoires à 2ℓ pas qui satisfont aux conditions de la relation (1.3) se scinde en deux sous-ensembles : celui des trajectoires strictement positives entre les instants 1 et $2\ell - 1$, noté P_ℓ , et celui des trajectoires strictement négatives entre ces instants, noté N_ℓ . Puisque l'on peut passer d'une trajectoire de N_ℓ à une trajectoire de P_ℓ par une symétrie par rapport à la droite $X = 0$, ces sous-ensembles ont le même cardinal. Une trajectoire $X \in P_\ell$ strictement positive ne peut s'annuler pour la première fois en 2ℓ que si elle satisfait aux conditions $X_{2\ell-1} = 1$ et $X_{2\ell-2} = 2$. Le cardinal de P_ℓ est donc égal à celui de l'ensemble $E_{\ell-1}^1$ des trajectoires qui vérifient $X_0 = 0$, $X_1 = 1$, $X_2 > 0$, ..., $X_{2\ell-3} > 0$, $X_{2\ell-2} = 2$. Pour calculer le cardinal de $E_{\ell-1}^1$, introduisons les événements :

- $A_\ell^r : X_0 = 0, X_1 = +1, X_{2\ell} = 2r,$
- $B_\ell^r : X_0 = 0, X_1 = +1, X_{2\ell} = 2r \text{ et } \exists j \text{ tel que } (1 < j < 2\ell \text{ et } X_j = 0),$
- $C_\ell^r : X_0 = 0, X_1 = -1, X_{2\ell} = 2r,$
- $D_\ell^r : X_0 = 0, X_{2\ell} = 2r,$
- $E_\ell^r : X_0 = 0, X_1 = +1, X_{2\ell} = 2r \text{ et } \forall j (1 < j < 2\ell \implies X_j > 0).$

Supposons $r \geq 0$. Puisqu'une trajectoire dans A_ℓ^r ne peut, commençant par un pas en avant, atteindre la position $2r$ à l'instant 2ℓ qu'en s'y rendant par une trajectoire toujours strictement positive ou bien en passant par zéro, il vient

$$A_\ell^r = E_\ell^r \cup B_\ell^r.$$

³En utilisant la formule de Stirling :

$$n! \simeq \sqrt{2\pi n} n^n e^{-n} \quad \text{pour } n \gg 1.$$

D'autre part, à toute trajectoire de B_ℓ^r peut être associée une trajectoire de C_ℓ^r , en appliquant une symétrie par rapport à l'axe des abscisses au début de la trajectoire, c'est-à-dire sur le tronçon compris entre l'instant 0 et celui du premier retour à zéro. Cette correspondance définit une bijection entre les sous-ensembles B_ℓ^r et C_ℓ^r (c'est le principe de symétrie de Désiré André). Nous venons d'établir la relation

$$\begin{aligned}\text{card}(E_\ell^r) &= \text{card}(A_\ell^r) - \text{card}(C_\ell^r) \\ &= C_{2\ell-1}^{\ell+r-1} - C_{2\ell-1}^{\ell+r} = C_{2\ell}^{\ell+r} \cdot \frac{r}{\ell},\end{aligned}$$

où la deuxième égalité s'obtient par un raisonnement semblable à celui qui nous a permis de démontrer la relation (1.2). Il ne reste plus qu'à utiliser ces relations dans le cas qui nous intéresse ($E_{\ell-1}^1$) pour trouver que la probabilité d'un premier retour à zéro à l'instant 2ℓ vaut

$$P(R_{2\ell}) = \frac{2 \text{card}(E_{\ell-1}^1)}{2^{2\ell}} = 2^{-2\ell} \frac{(2\ell)!}{(\ell!)^2} \frac{1}{2\ell-1}.$$

On voit que cette probabilité est $2\ell-1$ fois plus petite que celle d'un retour à zéro (pas nécessairement le premier) à l'instant 2ℓ .

1.2.2 File d'attente à un péage d'autoroute

Intéressons-nous aux queues qui peuvent se former à un péage d'autoroute. Une fréquence moyenne élevée de formations de queues longues a un impact économique important. En effet, si la probabilité qu'il se forme des queues de trois heures est de l'ordre de 5%, cela signifie que toutes les 60 heures environ, une file d'attente de trois heures apparaît ! De quoi décourager de nombreux usagers !

Supposons qu'il n'y ait qu'un lieu de paiement (s'il y en avait plusieurs le problème ne serait pas essentiellement différent) et prenons comme unité de temps la durée moyenne de paiement au péage (de l'ordre de quelques dizaines de secondes). Il est assez évident *a priori* que la probabilité de formation de queues de longueur fixée doit être une fonction décroissante du nombre moyen de véhicules par unité de temps.

Choisissons une origine des temps arbitraire et notons Q_n la longueur de la queue à l'instant n (dans notre système d'unité). Cette quantité est égale à la différence entre le nombre de véhicules déjà arrivés au péage et le nombre des véhicules l'ayant franchi. Notons également X_n le nombre de véhicules qui arrivent au lieu de paiement dans l'intervalle de temps $[n, n+1]$. Recherchons la loi de probabilité de X_n , sachant qu'en moyenne ce nombre vaut $a < 1$.

Remarquons que l'inégalité $a < 1$, entraîne qu'en « moyenne » il ne se forme pas de queue ; en revanche, si l'inégalité est inversée $a > 1$, il y aura à coup sûr formation d'une queue qui s'allongera indéfiniment, puisqu'en moyenne il arrivera plus de véhicules au péage qu'il n'en repartira. Toutefois, même avec une moyenne satisfaisant à la condition $a < 1$, la probabilité qu'il se forme une file d'attente, éventuellement longue, n'est pas nulle. En effet, il peut arriver que plusieurs véhicules se présentent au péage en une seule unité de temps,