

Traitement du Signal

cours de Master 2 Recherche

Pierre-Olivier Amblard

Contents

1	Introduction, signal?	2
1.1	Sondage sismique	2
1.2	Flux EUV du soleil	3
1.3	Signal de turbulence	4
1.4	Signal d'écho-location : la chauve-souris	5
1.5	Pour terminer et préparer la suite	5
2	Signaux déterministes	7
2.1	Zoologie	7
2.2	Signaux d'énergie finie, signaux de puissance moyenne finie	7
2.3	Langage commun, espaces vectoriels	8
2.4	Représentation fréquentielle : transformation de Fourier, en fréquences réduites, en Z	9
2.5	Quelques notions sur les filtres et la transformation des signaux	13
3	Signaux aléatoires	15
3.1	Définition	15
3.2	Stationnarité.	16
3.3	Analyse harmonique	16
3.4	Brownien et bruit blanc	18
3.5	Notion d'ergodisme, mesure des corrélation et DSP	19
4	Processus de markov, méthodes MCMC	21
4.1	Définitions, propriétés	21
4.2	Un exemple, le modèle AR	23
4.3	Méthode de Monte-Carlo pour l'intégration	24
4.4	Méthodes MCMC	27
5	Introduction aux processus ponctuels et à leurs produits dérivés	34
5.1	Processus ponctuels	34
5.2	Processus de Poisson	35
5.3	Bruits de grenaille	39

Ce document constitue des notes de cours de traitement du signal, cours dispensé en Master 2 recherche signal-image-parole-télécom.

Notations :

t est le temps continu

n le temps discret

ν la fréquence continue

λ fréquence continue réduite

m fréquence discrète

$\Gamma_{xy}(\tau)$ fonction d'intercorrélation entre x et y au retard τ

$\gamma_{xy}(\nu)$ densité spectrale de puissance d'interaction (interspectre) entre x et y à la fréquence ν

X ou $X(t)$ pour la variable aléatoire ou le signal aléatoire. Les lettres minuscules x et $x(t)$ sont réservés aux valeurs prises par la variable ou la fonction aléatoire.

1 Introduction, signal?

D'après le Larousse (ed. 1986) :

Signal n. m. (lat. *signalis* ; de *signum*, signe). Signe convenu pour avertir, annoncer, donner un ordre. — ... — En théorie de la communication, variation d'une grandeur de nature quelconque porteuse d'information. La définition est claire et signifie qu'un signal est une fonction mathématique d'un ensemble dans un autre codant une information. Par contre la définition n'indique rien sur les aspects de construction des signaux et de leur transfert, et sur les problèmes d'extraction de l'information. Le traitement du signal est la discipline qui s'intéresse à l'ensemble des aspects liés au signal: mise en forme de l'information, émission, réception, extraction de l'information utile.

Les techniques développées en traitement du signal reposent sur les mathématiques, mais également sur d'autres sciences qui sont à la source ou qui interviennent dans la transmission du signal. Par exemple, la transmission d'un signal radio FM utilise les ondes électromagnétiques, les potentiels d'action qui circulent dans les neurones sont des courants électriques générés par des effets biochimiques, ...

Dans un premier temps le traitement du signal essaie d'avoir un point de vue unificateur, oubliant les disciplines connexes liés au signal étudié, mais ce point de vue réducteur doit être modéré dans les traitements avancés et la richesse du traitement du signal sort en général de l'interaction avec les autres sciences.

Enfin, le traitement du signal participe à l'évolution des sciences et des techniques. Si les domaines les plus connus dans lesquels intervient le TS sont souvent de nature technologique (télévision, téléphonie, ...) il ne faut pas oublier que le TS est une discipline participant à l'avancée des connaissances scientifiques.

Sans être exhaustif voilà quelques exemples de domaines dans lesquels interviennent le traitement du signal :

- Sciences de l'observation (astronomie–lumière visible ou invisible–, géophysique interne–sondage sismique, sismologie, les signaux sont véhiculés par des ondes élastiques–, géophysique externe –environnement planétaire, sondage par ondes électromagnétiques et mesures effectuées par des RADAR–, télédétection –observation de la terre depuis l'espace–, physique fondamentale–étude de la turbulence par exemple–, neurosciences–compréhension du codage et du traitement de l'information par le cerveau–, observation par SONAR–ondes acoustiques–
- Télécommunication : domaine dans lequel le traitement du signal est fondamental (codage, lutte contre les interférences et les parasites, décodage, ...)
- ...

D'une façon générale, les signaux peuvent être issus de n'importe quel type de mesure, et le traiteur du signal revient alors à la définition du Petit Larousse.

Dans la suite, nous présentons quatre exemples de signaux issus de problématiques différentes.

1.1 Sondage sismique

La sismologie étudie les vibrations du sous-sol provoquées par des phénomènes naturels alors que la sismique provoque l'ébranlement et en étudie les conséquences. Le sondage sismique repose sur ce principe. Une source (explosion, vibroséisme, ...) provoque une onde élastique dans le sous-sol, onde qui va interagir avec son milieu de propagation. Une antenne munie de divers capteurs permet en un point plus éloigné de mesurer l'onde après son voyage sous terre. Les signaux reçus portent alors des informations relatives à la structure de leur milieu de propagation.

Un exemple de mesure issu d'une antenne sismique est montré sur la figure (1).

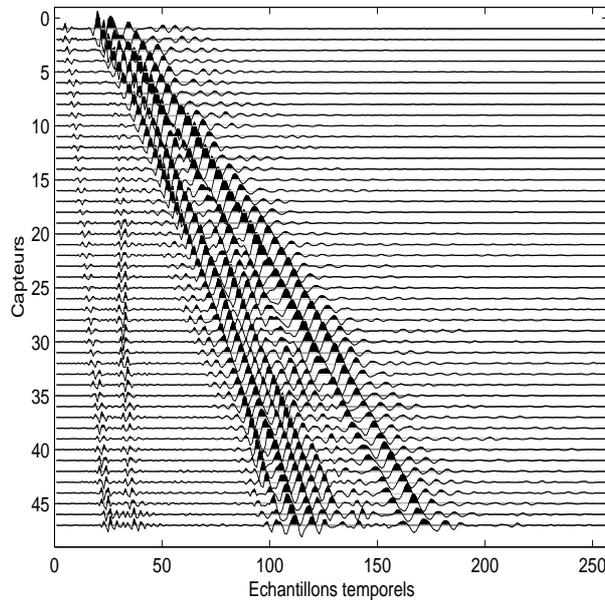


Figure 1: Exemple d'une trace sismique.

1.2 Flux EUV du soleil

L'étude du flux solaire est fondamentale pour des raisons évidentes de compréhension de notre étoile, mais également pour les problèmes climatiques qui deviennent critiques à notre époque. Différents appareils de mesure se situent sur divers satellites et mesurent le flux solaire dans diverses longueurs d'ondes. Un des buts ultime de ces mesures et de leur compréhension est la prédiction à court et moyen terme de l'activité solaire. Les images suivantes (2) sont des images du soleil dans certaines longueur d'ondes (extrême UltraViolet, images de l'atmosphère solaire, la brillance représente des zones à $60\text{-}70000\text{ K}$ pour 304 \AA , 10^6 pour 171 , $1.5 \cdot 10^6$ pour 195 et $2 \cdot 10^6$ pour 284 . Le plus chaud, le plus haut dans l'atmosphère solaire. Différentes fréquences renseignent

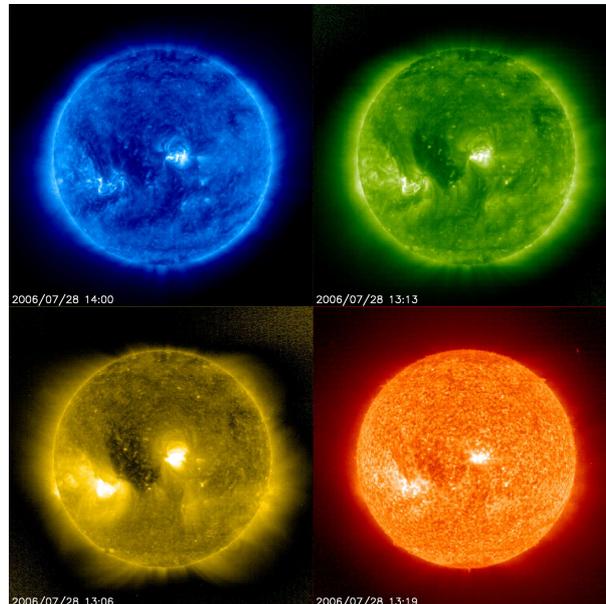


Figure 2: Images du soleil dans 4 bande fréquentielles différentes le 28 août 2006- EIT on board SOHO

donc sur différentes zones. Il est donc fondamental d'observer une plus grande gamme de fréquence. Or, les appareils nécessaires coûtent chers, et les missions spatiales ne sont pas légions. De plus, des idées issues de la physique montrent que le spectre entier est redondant et que la mesure de quelques raies spectrales permet de reconstruire l'ensemble du spectre. Trouver les raies pertinentes requiert l'utilisation de techniques de traitement des signaux. On peut alors s'intéresser à partir de plusieurs mesures du spectre solaire à décomposer ce spectre

en spectres élémentaires. Des résultats utilisant des techniques très récentes de traitement du signal sont montré sur la figure (3) D'autre part, peut-on reconstruire l'irradiance totale à partir de certaines autres mesures? La

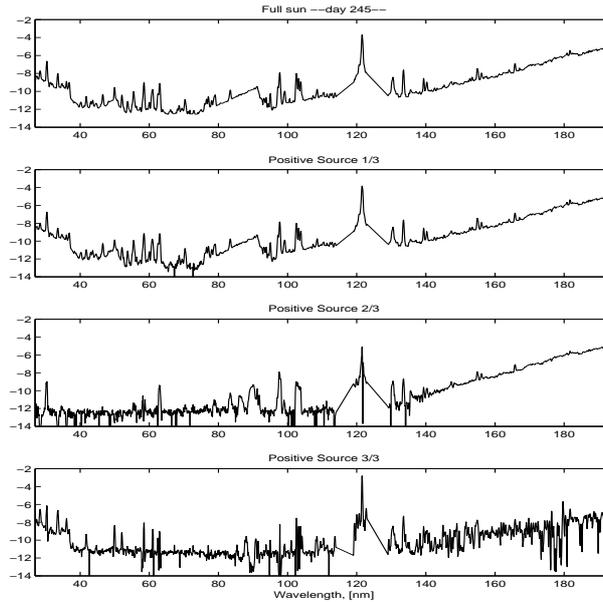


Figure 3: Décomposition en composantes élémentaires du spectre EUV solaire.

figure (4) suivante montre différents indices qui servent actuellement à l'étude de l'activité solaire. Le signal le plus bas correspond à l'irradiance totale mesurée sur terre et le signal du haut correspond au dénombrement des taches solaires. Les recherches très actuelles tentent de reconstruire l'irradiance à partir des autres mesures.

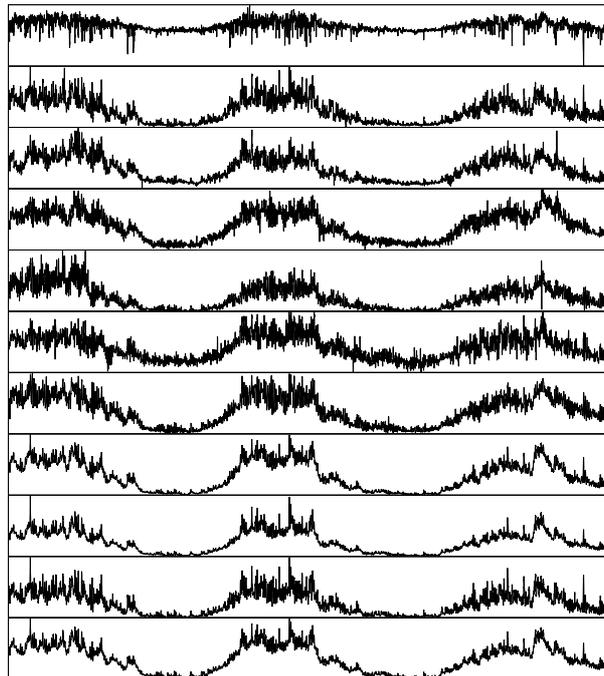


Figure 4: Différents indices de l'activité solaire sur les 26 dernières années.

1.3 Signal de turbulence

La turbulence dans les fluides reste aujourd'hui un défi pour les physiciens. Sa compréhension a une importance non seulement scientifique, mais également technologique. La turbulence est gérée mathématiquement par

l'équation de Navier-Stokes, équation aux dérivées partielles fortement non linéaire. Le terme non linéaire $\mathbf{v}\nabla\mathbf{v}$, où \mathbf{v} est le champ de vitesse, induit une complexité terrifiante. A côté des études mathématiques sur l'existence, la régularité de solutions, les physiciens ont une approche plus pragmatique. Les premiers résultats sur la turbulence sont dûs à A.N. Kolmogorov en 1941. Sous des hypothèses de stationnarité et d'homogénéité, il montre la loi des¹ $4/5$, et déduit par des arguments de dimension la forme de la densité spectrale de la vitesse dans un fluide turbulent. Pour contrer certaines remarques de Landau, il modifie sa théorie en 1962 avec Obukhov en prenant en compte la fluctuation de la dissipation d'énergie avec les échelles.

La figure suivante montre un exemple de signal issu d'un flot turbulent. Il s'agit de la vitesse longitudinale d'un flot d'air dans une soufflerie mesurée par un fil chaud. La figure suivante représente sa densité spectrale, c'est-à-dire la puissance du signal élatée sur l'ensemble des fréquences contenue dans le signal. La théorie de Kolmogorov prédit une dépendance en $1/f^{5/3}$ qui est bien vérifiée. Toutefois, la théorie est incomplète comme le prouve l'illustration suivante

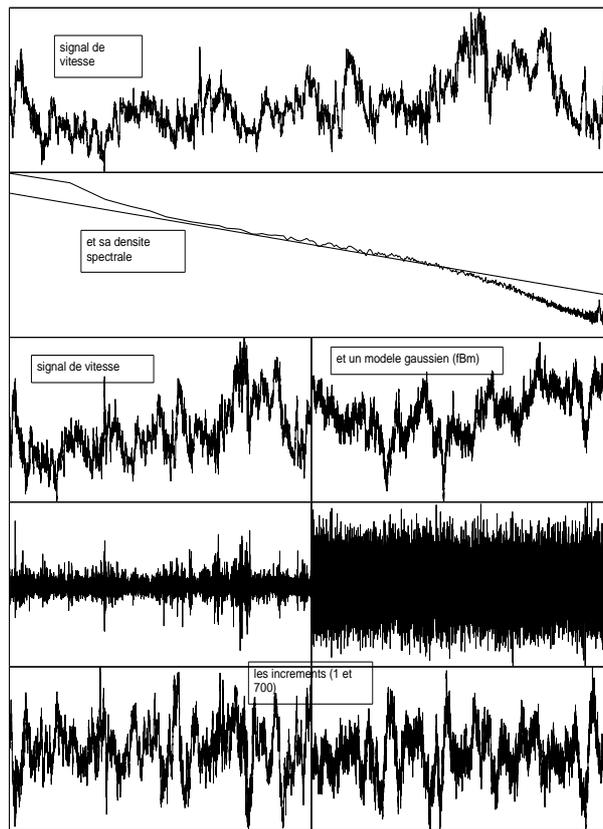


Figure 5: Signal de vitesse mesuré dans un écoulement turbulent et son spectre de puissance illustrant la pente en $5/3$. En bas comparaison du signal avec un modèle gaussien ayant la même densité spectrale de puissance. Les incréments de taille 1 montre bien l'invalidité de l'hypothèse gaussienne.

1.4 Signal d'écho-location : la chauve-souris

Pour terminer cette petite galerie d'exemple, regardons les signaux émis par une chauve-souris en phase d'attaque de proie. Ce signal est porté par des ondes acoustiques, la chauve-souris étant donc l'ancêtre de nos SONAR. Le spectre du signal peut-être trompeur quand à la structure intime du signal, ainsi que le révèle une analyse dite temps-fréquence (voir la figure (6)).

1.5 Pour terminer et préparer la suite

Ces quelques exemples divers ont montré la nécessité de développer des modèles précis de signaux et d'étudier leurs interactions avec les systèmes. Dans la suite, nous allons donc nous intéresser aux modèles de signaux, en

¹Cette loi lie le moment d'ordre 3 des incréments de vitesse de taille l à la dissipation d'énergie à l'échelle l . Sous certaines hypothèses, ce moment d'ordre trois est proportionnel à l'échelle, le coefficient de proportionnalité étant $-4/5$ de la dissipation d'énergie par unité de masse.

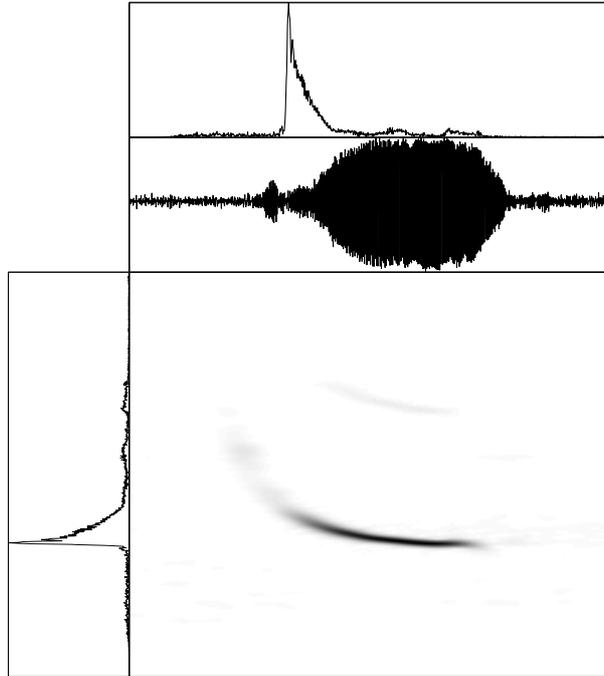


Figure 6: Signal d'écholocation d'une chauve souris dans ses représentations temps, fréquence et temps-fréquence.

présentant (ou en rappelant) les modèles de signaux déterministes, aléatoires. Nous étudierons les grandeurs qui révèlent des informations utiles (fonctions de corrélations, densités spectrales,...). Cette première partie du cours occupera 4 à 5 séances. Ensuite, nous examinerons des modèles plus complexes et très utiles dans de nombreux domaines : les signaux markoviens et les processus ponctuels.

2 Signaux déterministes

2.1 Zoologie

Un signal est défini comme une fonction mathématique d'un ensemble d'indexation \mathbb{X} vers un ensemble image \mathbb{Y} dans lequel le signal prend ses valeurs. L'ensemble \mathbb{X} peut être :

- \mathbb{R} ou un de ses intervalles, cas d'un signal à temps continu,
- \mathbb{Z} ou \mathbb{N} pour des signaux à temps discret par exemple,
- d'autres structures telle $\mathbb{N} \times \mathbb{N}$ ou encore un graphe

Dans ce cours nous nous limiterons aux signaux ne dépendant que d'une variable qui pourra être continue ou discrète.

L'ensemble Y image peut être :

- à une dimension, \mathbb{R} ou \mathbb{C} ,
- à n dimension pour des signaux multidimensionnels (exemple des enregistrements sismique)

Un signal x de la variable t sera donc écrit comme

$$\begin{aligned} x : \mathbb{X} &\longrightarrow \mathbb{Y} \\ t &\longmapsto x(t) \end{aligned}$$

Un signal est en général borné, et ce pour des raisons physiques.

Dans toute la suite, les variables t et n seront appelées variables temporelles, la première décrivant un espace d'indexation continu qui sera \mathbb{R} , alors que la deuxième sera réservée aux espaces d'indexation discrets, représenté par \mathbb{Z} . Bien que nous appelions ces variables des variables temporelles ne doit pas cacher le fait que dans certaines applications, les variables d'indexation ne seront pas le temps (exemple du flux EUV solaire). De plus, les signaux prendront leur valeurs dans \mathbb{C} .

Dans la suite, nous allons détailler deux grandes classes de signaux, les signaux d'énergie finie et les signaux de puissance moyenne finie. Nous donnerons ensuite une vision géométrique des signaux donnant ainsi un cadre commun, cadre qui permettra naturellement d'introduire d'autres représentations des signaux par changement de bases. Les représentations fréquentielles des signaux seront alors introduites. Nous reviendrons alors sur les notions d'énergie et de puissance. Enfin, nous examinerons les interactions entre les signaux et les systèmes.

2.2 Signaux d'énergie finie, signaux de puissance moyenne finie

On définit l'énergie d'un signal par

$$\begin{aligned} E_x &= \alpha \int_{\mathbb{R}} |x(t)|^2 dt \text{ pour le temps continu} \\ E_x &= \alpha \sum_{n \in \mathbb{Z}} |x(n)|^2 \text{ pour le temps discret} \end{aligned}$$

La constante α est placée ici pour dimensionner correctement l'énergie. Dans la suite, on posera $\alpha = 1$, mais il ne faut jamais oublier dans les applications que la somme du module carré du signal est proportionnel à l'énergie.

La puissance moyenne d'un signal est définie par

$$\begin{aligned} P_x &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T |x(t)|^2 dt \text{ pour le temps continu} \\ P_x &= \lim_{K \rightarrow +\infty} \frac{1}{2K+1} \sum_{n=-K}^K |x(n)|^2 \text{ pour le temps discret} \end{aligned}$$

L'ensemble des signaux pour lesquels l'énergie est finie est appelé espace des signaux d'énergie finie et est noté L^2 . Notons qu'un signal d'énergie finie a nécessairement une puissance moyenne nulle. L'espace des signaux de puissance moyenne finie est noté L^2_{loc} , et il contient évidemment L^2 . Un cas très particulier des signaux de puissance moyenne finie est l'ensemble des signaux périodiques de période T que l'on notera $L^2(T)$. Dans

cet espace la limite T tend vers l'infini est non nécessaire pour la définition de la puissance qui se restreint à $1/T \int_{[T]} |x(t)|^2 dt$, où $[T]$ correspond à un intervalle quelconque de longueur T .

Ecrire la définition de l'énergie présuppose que les intégrales (ou les sommes) existent. On définit donc l'espace $L^2(\mathbb{X})$ comme l'espace des signaux indexés par \mathbb{X} qui sont d'énergie finie. Notons que cet espace est un cas particulier des espaces $L^p(\mathbb{X})$ qui ont des relations non triviales les uns avec les autres.

2.3 Langage commun, espaces vectoriels

Rappelons qu'un ensemble est une espace vectoriel (ou linéaire – on dit linear space en anglais–) sur un corps de référence (\mathbb{R} ou \mathbb{C} pour nous) s'il est stable par addition et par multiplication par un scalaire du corps, c'est-à-dire pour tout x, y de l'ensemble et tout λ du corps, l'élément $x + \lambda y$ est à nouveau dans l'ensemble. Les éléments d'un espace vectoriel sont appelés vecteurs. On montre que les espace $L^2(\mathbb{C})$ et $L^2_{loc}(\mathbb{C})$ sont des espaces vectoriels sur le corps des complexes.

On introduit alors sur l'espace vectoriel un produit scalaire défini comme étant une forme bilinéaire symétrique ($\langle x|y \rangle = \langle y|x \rangle^*$) (ou sesquilinéaire dans le cas des complexes) définie-positive ($\langle x|x \rangle \geq 0$ avec égalité si et seulement si x est le vecteur nul). Rappelons que le produit scalaire permet en quelque sorte de comparer deux vecteurs. Par exemple dans l'espace usuel à 3 dimensions qui nous entoure avec le produit scalaire usuel on a $\langle x|y \rangle = |x||y| \cos \theta$, θ étant l'angle entre les directions portant les deux vecteurs. Ainsi le produit scalaire entre deux vecteurs distincts et non nuls est nul si les deux vecteurs sont orthogonaux. Nous garderons cette définition pour des espaces plus abstraits. Un espace vectoriels muni d'un produit scalaire est un espace de (pré)Hilbert. (Pour être de Hilbert, l'espace doit être complet, les suites de Cauchy convergent).

Le produit scalaire induit une norme sur l'espace vectoriel. Une norme est une application de l'espace dans \mathbb{R} définie-positive, vérifiant $\|\lambda x\| = |\lambda| \|x\|$ et l'inégalité $\|x + y\| \leq \|x\| + \|y\|$. Enfin d'une norme on déduit une distance entre deux vecteurs comme la norme de leur différence $\|x - y\| = d(x, y)$. Notons qu'une distance est une application symétrique, vérifiant l'inégalité triangulaire $d(x, y) \leq d(x, z) + d(z, y)$ et telle que $d(x, y) = 0$ si et seulement si $x = y$.

Rappelons l'inégalité de Schwartz : $|\langle x|y \rangle|^2 \leq \langle x|x \rangle \langle y|y \rangle$ avec égalité si les vecteurs sont proportionnels. Les espaces $L^2(\mathbb{C})$ et $L^2_{loc}(\mathbb{C})$ sont hilbertiens en les munissant des produits scalaires

$$\begin{aligned} \langle x|y \rangle &= \int x(t)y^*(t)dt \\ \langle x|y \rangle &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T x(t)y^*(t)dt \end{aligned}$$

Les espaces sont alors normés puis métriques en utilisant les normes et les distances induites. L'intérêt principal de cette formalisation est d'introduire des notions géométriques sur l'espace des signaux. En particulier nous allons brièvement présenter le théorème de la projection orthogonale et introduire la notion de bases.

Théorème de la projection orthogonale : Soit \mathcal{H} un espace de Hilbert et \mathcal{A} un sous espace vectoriel complet de \mathcal{H} . Soit $x \in \mathcal{H}/\mathcal{A}$. Alors $y_0 \in \mathcal{A}$ minimise $d(x, y_0)$ si et seulement si $\langle x - y_0|y \rangle = 0$ pour tout élément y de \mathcal{A} .

Autrement dit, le vecteur $x - y_0$ est orthogonal au sous-espace \mathcal{A} , c'est-à-dire y_0 est la projection orthogonale de x sur le sous-espace \mathcal{A} . Ce théorème est fondamentale en théorie de l'estimation dans laquelle on cherche à approcher au mieux un signal x par un signal d'une certaine classe de signaux. Dans la vision géométrique, le signal x appartient à un espace vectoriel (typiquement L^2) et la classe de signaux est un sous-espace vectoriel de L^2 . Obtenir la meilleure approximation de x dans cette classe revient alors à projeter orthogonalement x sur cette classe.

Bases : Une famille dénombrable e_1, \dots, e_n, \dots de vecteurs d'un espace de Hilbert est un système orthogonal si $\langle e_i|e_j \rangle = 0 \forall i \neq j$, orthonormé si de plus $\langle e_i|e_i \rangle = 1$, total $\langle x|e_j \rangle = 0 \forall j$ implique $x = 0$ (la seule fonction orthogonale à tous les éléments e est la fonction nulle).

Une famille e_1, \dots, e_n, \dots orthogonale constitue une base si tout x peut s'écrire $x = \sum_{i=1}^{+\infty} x_i e_i$ où $x_i = \langle x|e_i \rangle / \|e_i\|^2$. Attention, cette écriture signifie deux choses : les sommes partielles $\sum_{i=1}^N$ converge vers un fonction $f(x)$ et de plus $f(x) = x$. On peut montrer le résultat suivant : une famille e_1, \dots, e_n, \dots est une base orthogonale si et seulement si elle est totale, si et seulement si l'ensemble de ses combinaisons linéaires finies est dense dans l'espace considérée, si et seulement si $\|x\|^2 = \sum_{i=1}^{+\infty} |x_i|^2 \|e_i\|^2$ (égalité appelée égalité de Parseval).

Un autre point de vue (équivalent) est celui de l'approximation. Dans ce contexte, une famille de vecteurs est une base de l'espace si tout élément de l'espace peut être approché arbitrairement finement par une combinaison linéaire de ces vecteurs. Autrement dit pour tout x et tout $\varepsilon > 0$, on peut trouver un entier N et des nombre x_1, \dots, x_N tel que $d(x, \sum_{i=1}^N x_i e_i) \leq \varepsilon$.

Un exemple : on considère $L_P^2(T)$ l'ensemble des fonctions périodique de période T . L'ensemble des fonctions $e_k(t) = \exp 2i\pi kt/T, k \in \mathbb{Z}$ est une base orthogonale de $L_P^2(T)$. Cet exemple est celui des séries de Fourier. Le coefficient de Fourier est donné par $c_k = \langle x | e_k \rangle = T^{-1} \int_{[T]} x(t) \exp(-2i\pi kt/T) dt$ et on a $x(t) = \sum_{k \in \mathbb{Z}} c_k e_k(t)$ dans le sens $\|x(t) - \sum_{k=-N}^N c_k e_k(t)\| \rightarrow 0$ quand $N \rightarrow +\infty$. Cet exemple est important et possède une interprétation physique importante pour la suite. Il montre en effet que tout signal périodique se décompose comme une somme de sinusoides de fréquence multiple de la fréquence fondamentale $1/T$. De plus, l'égalité de Parseval s'écrit dans ce contexte $1/T \int_{[T]} |x(t)|^2 dt = \sum_{k \in \mathbb{Z}} |c_k|^2$. Or le terme $1/T \int_{[T]} |x(t)|^2 dt$ n'est autre que la puissance du signal x et l'égalité de Parseval montre comment la puissance du signal se répartit sur les composantes fréquentielles élémentaires. De plus cette discussion montre que définir le signal par sa représentation temporelle $x(t)$ est équivalent à définir le signal par sa décomposition en série de Fourier que nous appellerons représentation fréquentielle.

2.4 Représentation fréquentielle : transformation de Fourier, en fréquences réduites, en Z

Nous venons de voir comment la notion de fréquence peut être introduite naturellement dans l'étude des signaux périodiques. Généraliser cette représentation au cas des signaux non périodiques. Cette représentation se généralise vers les fonctions définies sur \mathbb{R} par

$$X(\nu) = \int x(t) e^{-2i\pi\nu t} dt = TF[x(t)]$$

où la fonction $X : \mathbb{R} \rightarrow \mathbb{C}$ est appelée transformée de Fourier de x . Cette définition est évidemment licite dans certaines conditions. Pour qu'une fonction ait une transformée de Fourier, il suffit qu'elle soit sommable, c'est à dire que $\int |x(t)| dt < +\infty$ et donc que la fonction appartienne à $L^1(\mathbb{C})$. Notons que la définition s'étend pour des fonctions de carré sommable, autrement dit les fonctions de $L^2(\mathbb{C})$. L'extension n'est pas immédiate puisque L^2 n'est inclu dans L^1 . Les subtilités mathématiques passées, nous pouvons rappeler les éléments fondamentaux concernant la transformée de Fourier.

On inverse la transformée à l'aide de

$$x(t) = \int X(\nu) e^{2i\pi\nu t} d\nu = TF^{-1}[X(\nu)]$$

La transformée de Fourier du produit de convolution est un produit simple,

$$z(t) = \int x(u) y(t-u) du \implies Z(\nu) = X(\nu) Y(\nu)$$

La transformation de Fourier échange dérivation et multiplication par un monôme :

$$\begin{aligned} TF[\dot{x}(t)] &= 2i\pi\nu X(\nu) \\ TF[-2i\pi t x(t)] &= \dot{X}(\nu) \end{aligned}$$

La transformation de Fourier échange translation et déphasage :

$$\begin{aligned} TF[x(t-\tau)] &= e^{-2i\pi\nu\tau} X(\nu) \\ TF[x(t)e^{2i\pi\nu_0 t}] &= X(\nu - \nu_0) \end{aligned}$$

LA TF d'une fonction (im)paire est (im)paire. La TF d'une fonction réelle et paire est réelle et paire.

Enfin, la transformation de Fourier est une transformation unitaire : non seulement elle conserve l'énergie (théorème de Parseval) mais elle conserve également les produits scalaires

$$\langle x(t) | y(t) \rangle = \int x(t) y(t)^* dt = \int X(\nu) Y(\nu)^* d\nu = \langle X(\nu) | Y(\nu) \rangle$$

Propriétés énergétiques : Nous avons déjà vu la définition des signaux d'énergie et de puissance finie. Notons que l'énergie du signal $x \in L^2$ n'est autre que sa norme L^2 alors que la puissance du signal $x \in L_{loc}^2$ est sa norme dans L_{loc}^2 . En utilisant les résultats sur les espaces vectoriels de Hilbert on voit que d'une manière générale le contenu énergétique (ou de puissance) est donc $\|x(t)\|^2 = \langle x | x \rangle$ où $\langle x | y \rangle = \int x(t) y^*(t) dt$ pour l'énergie et $\langle x | y \rangle = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T x(t) y^*(t) dt$ pour la puissance. Dans le cas de la puissance, il sera intéressant de considérer le signal $x_T(t) = x(t) \cdot \mathbf{1}_{[-T, T]}(t)$ égal à $x(t)$ sur l'intervalle $[-T, T]$ et 0 ailleurs.

Le langage des espaces vectoriels nous permet maintenant de comparer deux signaux entre eux à l'aide du produit scalaire. Pour comparer deux signaux il s'agit d'examiner les liens qui existent entre deux dates distinctes des deux signaux. On introduit alors la fonction d'intercorrélation (qui existe dans l'espace considéré en vertu de l'inégalité de Schwartz) par

$$\begin{aligned}\Gamma_{xy}(\tau) &= \langle x(t)|y(t-\tau) \rangle \\ \Gamma_{xy}(\tau) &= \int x(t)y^*(t-\tau)dt \text{ énergie finie} \\ \Gamma_{xy}(\tau) &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T x(t)y^*(t-\tau)dt \text{ puissance moyenne finie}\end{aligned}$$

dont les cas particuliers sont les fonctions de corrélation $\Gamma_{xx}(\tau)$. Notons que $\Gamma_{xx}(-\tau) = \Gamma_{xx}^*(\tau)$ et qu'en particulier la fonction d'autocorrélation d'un signal réel est paire. De plus la fonction de corrélation est maximale en zéro (inégalité de Schwartz ou sens physique).

Examinons les grandeurs similaires définies sur les représentations fréquentielles. Soient $X(\nu)$ la TF d'un signal d'énergie finie et $X_T(\nu)$ la TF du signal tronqué $x_T(t)$ lorsque x est de puissance moyenne finie. On définit alors la densité spectrale d'énergie ou de puissance d'interaction par

$$\begin{aligned}\gamma_{xy}(\nu) &= X(\nu)Y^*(\nu) \\ &= \lim_{T \rightarrow +\infty} \frac{1}{2T} X_T(\nu)Y^*(\nu)\end{aligned}$$

dont le cas particulier est la densité spectrale d'énergie ou de puissance $\gamma_{xx}(\nu) = |X(\nu)|^2$ ou $\lim_{T \rightarrow +\infty} \frac{1}{2T} |X_T(\nu)|^2$. Ces fonctions indiquent la répartition de l'énergie ou de la puissance le long des fréquences. L'énergie totale est donc la somme sur toutes les fréquences de la densité correspondante. Mais il existe un résultat plus complet encore mettant en lien fonction d'intercorrélation et densité spectrale d'interaction, c'est le théorème de Wiener-Khinchine qui stipule que ces paires de fonction sont transformées de Fourier l'une de l'autre, $\gamma_{xy}(\nu) = TF[\Gamma_{xy}(\tau)]$. Montrons le théorème de Wiener-Khinchine dans le cas de signaux de puissance moyenne finie. On doit effectuer la transformée de Fourier de l'intercorrélation:

$$\begin{aligned}TF[\Gamma_{xy}(\tau)] &= \int_{\tau} \left\{ \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T x(t)y^*(t-\tau)dt \right\} e^{-2i\pi\nu\tau} d\tau \\ &= \int_{\tau} \left\{ \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-\infty}^{+\infty} x_T(t)y_T^*(t-\tau)dt \right\} e^{-2i\pi\nu\tau} d\tau \\ &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-\infty}^{+\infty} x_T(t) \left\{ \int_{\tau} y_T^*(t-\tau)e^{-2i\pi\nu\tau} d\tau \right\} dt \\ &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-\infty}^{+\infty} x_T(t) \left\{ Y^*(\nu)e^{-2i\pi\nu t} \right\} dt \\ &= \lim_{T \rightarrow +\infty} \frac{1}{2T} X_T(\nu)Y^*(\nu)\end{aligned}$$

Notons que l'échange des limites et intégrales requiert des conditions que l'on suppose vérifiées ici.

Principe d'incertitude d'Heisenberg-Gabor : Un autre lien existe entre les propriétés énergétiques développées en temps ou en fréquence. Considérons en effet une mesure de la dispersion de l'énergie en temps, et une en fréquence par

$$\begin{aligned}\Delta t^2 &= \frac{1}{E_x} \int t^2 |x(t)|^2 dt \\ \Delta \nu^2 &= \frac{1}{E_x} \int \nu^2 |x(\nu)|^2 dt\end{aligned}$$

Ces grandeurs représentent la dispersion de l'énergie autour de l'axe t ou $\nu = 0$. Pour comprendre cela on peut voir $|x(t)|^2/E_x$ comme une densité de probabilité (centrée) et Δt^2 représente alors la variance. On a alors le résultat suivant appelé principe d'incertitude d'Heisenberg-Gabor:

$$\Delta t \Delta \nu \geq \frac{1}{4\pi}$$

avec égalité si et seulement si $x(t)$ est de forme gaussienne. Cette relation fondamentale stipule que si l'énergie est très concentrée dans un domaine (t ou ν) elle est alors forcément très dispersée dans l'autre. Le comble de cette assertion est le dirac dont la transformée de Fourier est la constante.

Pour démontrer l'inégalité, on utilise l'inégalité de Shwartz en remarquant que Δt^2 et $\Delta \nu^2$ sont les normes de deux signaux. En effet $\Delta t^2 = \langle tx(t)|tx(t) \rangle / E_x$ et $\Delta \nu^2 = \langle \nu X(\nu)|\nu X(\nu) \rangle / E_x$. En utilisant le fait que $TF[\dot{x}(t)] = 2i\pi\nu X(\nu)$ et que la TF conserve les produit scalaires on a $\Delta \nu^2 = \langle \dot{x}(t)/(2i\pi)|\dot{x}(t)/(2i\pi) \rangle / E_x$. Maintenant en utilisant l'inégalité de Schwartz on a

$$\begin{aligned}\Delta t^2 \Delta \nu^2 &= \frac{1}{4\pi^2 E_x^2} \langle tx(t)|tx(t) \rangle \langle \dot{x}(t)|\dot{x}(t) \rangle \\ &\geq \frac{1}{4\pi^2 E_x^2} \left| \langle tx(t)|\dot{x}(t) \rangle \right|^2\end{aligned}$$

avec égalité si les signaux sont proportionnels, c'est-à-dire si $\dot{x}(t) = ktx(t)$, une équation différentielle dont la solution est une gaussienne. Revenons au produit scalaire $\langle tx(t)|\dot{x}(t) \rangle$. Il s'écrit

$$\begin{aligned}\langle tx(t)|\dot{x}(t) \rangle &= \int tx(t)\dot{x}(t)^* dt \\ &= [t|x(t)|^2]_{-\infty}^{+\infty} - \int (x + t\dot{x}(t))x(t)^* dt\end{aligned}$$

où nous avons effectué une intégration par partie. Le premier terme est égal à zéro puisque nous supposons les signaux d'énergie finie et par suite,

$$\langle tx(t)|\dot{x}(t) \rangle = -E_x - \langle tx(t)|\dot{x}(t) \rangle^*$$

ou $\text{Re}(\langle tx(t)|\dot{x}(t) \rangle) = -E_x/2$. Mais le carré de la partie réelle d'un nombre complexe est toujours inférieur ou égal au module carré de ce nombre et nous avons donc finalement

$$\begin{aligned}\Delta t^2 \Delta \nu^2 &\geq \frac{1}{4\pi^2 E_x^2} \left| \langle tx(t)|\dot{x}(t) \rangle \right|^2 \\ &\geq \frac{1}{4\pi^2 E_x^2} \text{Re} \left(\langle tx(t)|\dot{x}(t) \rangle \right)^2 \\ &= \frac{1}{4\pi^2 E_x^2} \frac{E_x^2}{4}\end{aligned}$$

Signaux à temps discret, transformée en Z :

Pour les besoins des traitements numériques, la théorie du signal s'est tournée vers l'étude des signaux à temps discrets. Tout comme à temps continu, les signaux d'énergie finie et de puissance moyenne finie sont considérés, et l'on peut définir des fonctions de corrélations selon

$$\begin{aligned}\Gamma_{xy}(k) &= \sum_{n \in \mathbb{Z}} x_n y_{n-k}^* \text{ pour l'énergie finie} \\ \Gamma_{xy}(k) &= \lim_{K \rightarrow +\infty} \frac{1}{2K+1} \sum_{n=-K}^K x_n y_{n-k}^* \text{ pour la puissance moyenne finie}\end{aligned}$$

et l'on peut avoir les mêmes discussions qu'à temps continu. Par contre pour faire l'analyse harmonique, il faut introduire les outils adaptés au traitement des signaux à temps discret : il s'agit de la transformée en Z et de son cas particulier la transformée de Fourier en fréquences réduites.

Transformée en Z de séquences discrètes

Soit $x_n, n \in \mathbb{Z}$ une séquence discrète. Sa transformée en Z est définie par

$$X(Z) = \sum_{n \in \mathbb{Z}} x_n z^{-n}$$

où z est une variable complexe. Pour que la transformée existe, il faut que la série soit convergente, ce qui est en général assuré dans une couronne de convergence. La transformée s'inverse en intégrant le long d'un chemin entourant l'origine dans la couronne de convergence selon

$$x_n = \frac{1}{2i\pi} \oint X(z) z^n \frac{dz}{z}$$

La transformée en Z d'un signal retardé est $TZ(x_{n-k}) = z^{-k} X(z)$. La transformée en Z d'une convolution discrète est le produit des transformées en Z.

Si le cercle unité est dans le rayon de convergence, on peut alors poser $z = \exp 2i\pi\lambda$ et l'on définit alors la transformée de Fourier en fréquence réduite

$$\begin{aligned} X(\lambda) &= \sum_{n \in \mathbb{Z}} x_n e^{-2i\pi n\lambda} \\ x_n &= \int_{-1/2}^{1/2} X(\lambda) e^{2i\pi n\lambda} d\lambda \end{aligned}$$

Densités spectrales en Z On peut définir les densités spectrales d'énergie et de puissance d'interaction dans le cas discret en utilisant la transformée en Z . On va pour cela utiliser le théorème de WK appliqué à la TZ. On définit donc la DSEI des signaux x_n et y_n comme la TZ de leur fonction d'intercorrélacion, soit

$$\begin{aligned} \gamma_{xy}(z) &= \sum_k \Gamma_{xy}(k) z^{-k} \\ &= \sum_n \sum_k x_n y_{n-k}^* z^{-k} \\ &= \sum_n x_n \sum_l y_l^* z^{l-n} \\ &= \sum_n x_n z^{-n} \left(\sum_l y_l z^{l*} \right)^* \\ &= X(z) Y^* \left(\frac{1}{z^*} \right) \end{aligned}$$

La densité spectrale d'énergie s'en déduit. Elle est invariante par la transformation $z \mapsto 1/z^*$ qui montre que si la DSP a un pôle (resp. un zéro) d'angle θ et de module r est présent, alors nécessairement la DSP admet le pôle (resp. le zéro) de même angle et de module $1/r$.

Du monde continu vers le monde discret : échantillonnage

L'échantillonnage consiste à prélever à des instants discrets la valeur d'un signal. L'échantillonnage pourra être périodique, irrégulier et même aléatoire. Nous nous concentrons ici sur l'échantillonnage périodique. Soit $x(t)$ un signal. Le signal échantillonné à la cadence T_e s'écrit

$$x_e(t) = x(t) \cdot \sum_n \delta(t - nT_e)$$

où $\delta(t)$ est l'impulsion de Dirac. Rappelons que cette impulsion est une distribution qui satisfait $f(t) \cdot \delta(t - t_0) = f(t_0) \delta(t - t_0)$ et $f(t) \star \delta(t - t_0) = f(t - t_0)$ où \star est le produit de convolution. Le signal échantillonné s'écrit alors

$$x_e(t) = \sum_n x(nT_e) \delta(t - nT_e)$$

et l'on notera x_n la suite $x(nT_e)$. Il est intuitif de considérer que le signal $x(t)$ sera d'autant mieux représenté par la suite x_n que la période d'échantillonnage sera petite. Toutefois il existe des signaux pour lesquels une cadence petite mais non nulle permet de conserver toute l'information sur $x(t)$. Pour comprendre tout cela, examinons l'influence de l'échantillonnage sur la transformée de Fourier. La transformée de Fourier d'un peigne de Dirac $\sum_n \delta(t - nT_e)$ est un autre peigne de Dirac $(1/T_e) \sum_n \delta(\nu - n/T_e)$. On a donc

$$\begin{aligned} X_e(\nu) &= X(\nu) \star \frac{1}{T_e} \delta\left(\nu - \frac{n}{T_e}\right) \\ &= \sum_n X\left(\nu - \frac{n}{T_e}\right) \end{aligned}$$

Ainsi, la transformée de Fourier du signal échantillonné est donnée par la transformée de Fourier périodisée à la cadence $1/T_e$ de $x(t)$. Ceci est illustré sur les quatre premières planches de la figure (7)

On s'aperçoit en observant la transformée de Fourier de x_e que s'il n'y a pas de recouvrement entre les périodes, les motifs de base de la TF de $x(t)$ sont préservés. Dans ce cas, si l'on multiplie cette TF par une porte en fréquence on peut retrouver la TF de $x(t)$ et donc $x(t)$ par TF inverse. Comme la TF de x_e est obtenue en périodisant la TF de x , il n'y aura pas de recouvrement si $X(\nu) = 0$ pour $\nu > B \leq \nu_e/2 = 1/2T_e$. Dans ce cas

on retrouve $x(t)$ à partir de $x_e(t)$ par

$$\begin{aligned} x(t) &= TF^{-1} [X_e(\nu) \cdot T_e \mathbf{1}_{[-1/(2T_e), 1/(2T_e)]}(\nu)] \\ &= x_e(t) \star \text{sinc} \frac{\pi t}{T_e} \\ &= \sum_n x(nT_e) \text{sinc} \frac{\pi(t - nT_e)}{T_e} \end{aligned}$$

formule appelée formule d'interpolation de Shannon.

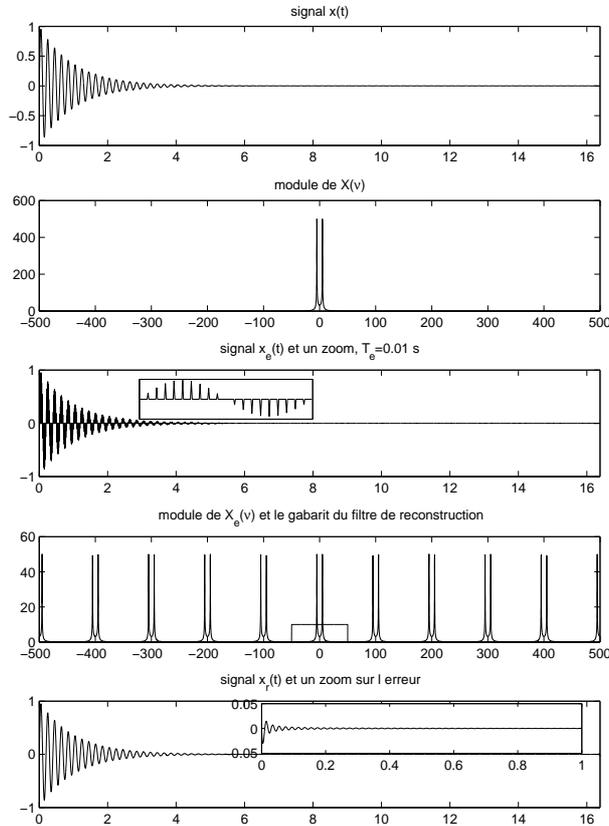


Figure 7: Illustration de l'échantillonnage

Si la condition de Shannon $\nu_e \geq 2\nu_{max}$, la périodisation provoque des recouvrements entre les motifs et on ne peut plus récupérer le signal initial. Ceci peut conduire à des reconstructions erronées induisant des erreurs énormes, comme illustré figure (8). En pratique, les bandes des signaux n'étant pas limitées, on applique préalablement à tout échantillonnage un filtre appelé anti-repliement qui limite la bande du signal et autorise un échantillonnage (toujours au-delà de deux fois la fréquence de coupure du filtre.)

2.5 Quelques notions sur les filtres et la transformation des signaux

Les filtres sont des opérateurs travaillant sur des espaces de signaux. Ils transforment un signal en un autre par $y(t) = (\mathcal{F}x)(t)$. Un filtre est linéaire si

$$\begin{aligned} (\mathcal{F}[x + y])(t) &= (\mathcal{F}x)(t) + (\mathcal{F}y)(t) \\ (\mathcal{F}[\lambda x])(t) &= \lambda(\mathcal{F}x)(t) \end{aligned}$$

Les transformées linéaires les plus générales peuvent s'écrire sous-forme intégrale selon

$$y(t) = (\mathcal{F}x)(t) = \int K(t, u)x(u)du$$

où K est appelé noyau de la transformation. Un filtre linéaire est dit invariant s'il commute avec l'opérateur de translation temporel : autrement dit, la sortie du filtre aujourd'hui sera la même que demain. Soit $T_\tau x$ le

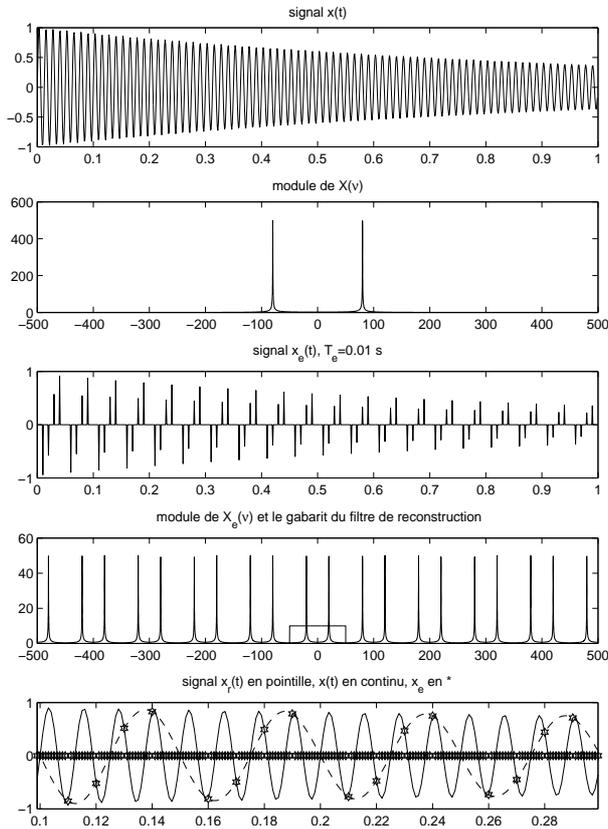


Figure 8: Illustration de l'échantillonnage

signal translaté de τ . Le filtre est invariant si

$$T_{\tau}(\mathcal{F}x)(t) = (\mathcal{F}T_{\tau}x)(t)$$

Cette égalité contraint le noyau à vérifier $K(t - \tau, u) = K(t, u + \tau)$ soit encore $K(t - \tau, u - \tau) = K(t, u)$ et ce pour tout retard τ . EN particulier cela est vérifié pour $\tau = -u$ et on obtient donc que $K(t, u) = K(t - u, 0)$. Donc le noyau n'est pas bi- mais monodimensionnel et on obtient pour \mathcal{F} un produit de convolution

$$y(t) = \int h(t - u)x(u)du$$

où l'on a noté $h(v) = K(v, 0)$. On montre facilement que le produit de convolution se transforme en un produit simple et la sortie du filtre linéaire invariant \mathcal{F} s'écrit dans le domaine des fréquences

$$Y(\nu) = H(\nu)X(\nu)$$

La fonction h est appelée réponse impulsionnelle et sa transformée de Fourier gain complexe du filtre.

3 Signaux aléatoires

De nombreux signaux issus de la physique, de la biologie, de l'économie et d'autres disciplines présentent un caractère très erratique qui les rend difficilement modélisables par des équations mathématiques usuelles. Pire, beaucoup de phénomènes sous-jacent aux signaux semblent dépendre des lois du hasard.

Nous devons alors étendre la théorie des signaux déterministes vers la théorie des signaux aléatoires, théorie dont le but est la définition de modèles et le développement d'analyse adéquats pour la description des signaux erratiques et non reproductibles.

3.1 Définition

Une fonction aléatoire est une variable aléatoire indexée par le temps. Nous ne referons pas la discussion temps continu-discret, mais les mêmes remarques peuvent être faites ici.

Soit donc un espace probabilisé (Ω, \mathcal{B}, P) où Ω est l'espace des épreuves, \mathcal{B} une tribu définie sur Ω et P une mesure de probabilité. Alors une fonction aléatoire est donnée par

$$\begin{aligned} X : (\Omega, \mathcal{B}, P) \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (\omega, t) &\longmapsto X(\omega, t) \end{aligned}$$

Ainsi, à t fixé, $X(t)$ est une variable usuelle. Pour décrire la structure probabiliste du signal, il faut réaliser des analyses multi-point ou multi-date, c'est-à-dire étudier les propriétés statistiques des k -uplets $(X(t_1), \dots, X(t_k))$. Le signal aléatoire est complètement décrit par la connaissance des mesures de probabilité conjointes de ces k -uplets, pour tout $k \in \mathbb{N}$.

Ainsi, on suppose connues les fonctions de répartitions

$$\Pr \{X(t_1) \leq x_1, \dots, X(t_k) \leq x_k\}, \forall k \in \mathbb{N}$$

et si elles sont différentiables, leurs densités de probabilité

$$p_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k; t_1, \dots, t_k), \forall k \in \mathbb{N}$$

Ces fonctions permettent l'évaluation de grandeurs moyennes

$$E[f_1(X(t_1)) \times \dots \times f_k(X(t_k))] = \int_{\mathbb{R}^k} f_1(x_1) \times \dots \times f_k(x_k) p_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k; t_1, \dots, t_k) dx_1 \dots dx_k$$

Par exemple, la valeur moyenne du signal aléatoire X est par définition

$$E[X(t)] = \int_{\mathbb{R}} x p_{X(t)}(x; t) dx$$

et le moment d'ordre 2 par

$$E[|X(t)|^2] = \int_{\mathbb{R}} |x|^2 p_{X(t)}(x; t) dx$$

La variance du signal est obtenue en travaillant sur le signal centré et l'on a

$$\begin{aligned} \text{Var}[X(t)] &= E[(X(t) - E[X(t)])^2] = \int_{\mathbb{R}} (x - E[X(t)])^2 p_{X(t)}(x; t) dx \\ &= E[|X(t)|^2] - |E[X(t)]|^2 \end{aligned}$$

La racine carrée de la variance définit l'écart type qui représente la taille caractéristique des fluctuations du signal autour de sa moyenne.

La valeur moyenne est donc *a priori* une grandeur dépendant du temps t . La fonction de corrélation est donnée par

$$E[X(t_1)X^*(t_2)] = \int_{\mathbb{R}^2} x_1 x_2^* p_{X(t_1), X(t_2)}(x_1, x_2; t_1, t_2) dx_1 dx_2$$

et est une fonction à deux dimensions. La fonction de covariance est obtenue en travaillant sur le signal centré $X(t) - E[X(t)]$.

On se limite en général à l'étude des deux premiers ordres $k = 1$ et $k = 2$ qui permettent à eux seuls de déjà bien caractériser le signal. A cet égard, il est important de rappeler que les écritures précédentes présupposent

évidemment que les intégrales (ou les sommes dans le cas discret) convergent. Nous introduisons alors la classe fondamentale des signaux aléatoires du second ordre définis comme les signaux de puissance finie, c'est-à-dire l'ensemble

$$\mathcal{H}^2 = \{X(t)/E[|X(t)|^2] < +\infty, \forall t\}$$

On montre facilement que cet espace est un espace vectoriel sur le corps de référence (\mathbb{R} ou \mathbb{C}), que l'on rend hilbertien en choisissant le produit scalaire $\langle X|Y \rangle = E[X(t)Y^*(t)]$. Dans ce formalisme, la fonction de corrélation peut être vue comme le produit scalaire $\langle X(t_1)|X(t_2) \rangle$ qui existe puisqu'en vertu de l'inégalité de Schwartz on a $|\langle X(t_1)|X(t_2) \rangle|^2 \leq \langle X(t_1)|X(t_1) \rangle \langle X(t_2)|X(t_2) \rangle = E[|X(t_1)|^2] E[|X(t_2)|^2]$ qui est une grandeur finie pour les signaux du second ordre.

Les covariances forment un ensemble stable par addition, multiplication et passage à la limite.

Continuité et dérivabilité. On dit que la fonction aléatoire est continue en moyenne quadratique en t si $E[|X(t+h) - X(t)|^2]$ tend vers 0 quand l'incrément h tend vers 0. Si cette propriété est vérifiée pour tout t , on dit que la fonction est continue en m.q. En développant l'espérance on obtient $E[|X(t+h) - X(t)|^2] = \Gamma_x(t+h, t+h) - \Gamma_x(t, t+h) - \Gamma_x(t+h, t) + \Gamma_x(t, t)$ et l'on voit que continuité en moyenne quadratique du processus est équivalent à continuité de $\Gamma_x(t, t)$.

On dit que la fonction aléatoire est dérivable en moyenne quadratique au point t si le taux d'accroissement au point t possède une limite finie en moyenne quadratique. On montre qu'une condition nécessaire et suffisante de dérivabilité en moyenne quadratique est l'existence de la dérivée seconde de Γ_x sur sa diagonale.

3.2 Stationnarité.

Un signal est stationnaire si ses propriétés statistiques sont invariantes par toute translation de l'origine des temps.

Examinons les conséquences de cette hypothèse sur les fonctions définissant la fonction aléatoire. Si $X(t)$ est stationnaire alors

$$\forall \tau \in \mathbb{R}, p_{X(t)}(x; t) = p_{X(t+\tau)}(x; t + \tau)$$

et donc en particulier pour $\tau = -t$. Donc on obtient, sous la condition de stationnarité

$$p_{X(t)}(x; t) = p_{X(0)}(x; 0)$$

montrant que les statistiques à une date sont indépendantes du temps. En particulier, cela implique que la moyenne est constante au cours du temps, ainsi que la variance.

En ce qui concerne l'analyse à deux dates, la stationnarité implique

$$\forall \tau \in \mathbb{R}, p_{X(t_1), X(t_2)}(x_1, x_2; t_1, t_2) = p_{X(t_1+\tau), X(t_2+\tau)}(x_1, x_2; t_1 + \tau, t_2 + \tau)$$

et donc en particulier pour $\tau = -t_2$, impliquant

$$p_{X(t_1), X(t_2)}(x_1, x_2; t_1, t_2) = p_{X(t_1-t_2), X(0)}(x_1, x_2; t_1 - t_2, 0)$$

Par suite, la stationnarité implique que les propriétés statistiques à deux dates ne dépendent que de l'écart entre ces deux dates. Ainsi la fonction de corrélation devient une fonction d'une variable τ que l'on convient d'appeler retard. On écrit alors

$$\Gamma_x(\tau) = E[x(t)x^*(t - \tau)]$$

exercice : Etude de la stationnarité au second ordre de $x(t) = A \cos(2\pi\nu_0 t + \varphi)$ où 1- φ est une v.a. uniformément répartie sur $[0, 2\pi]$ 2- A est une v.a. uniformément répartie sur $[0, 1]$. 3- A et φ sont des v.a. indépendantes uniformes.

3.3 Analyse harmonique

Réaliser l'analyse harmonique des signaux aléatoires n'est pas une chose aisée. Il faut en effet réaliser la transformation de Fourier (en supposant que l'on puisse) réalisation par réalisation. La transformée de Fourier serait alors également une fonction aléatoire dont il faudra étudier les statistiques.

La première difficulté réside dans le fait qu'un signal aléatoire n'a presque-sûrement pas de transformée de Fourier. En effet, un signal aléatoire n'est presque sûrement pas sommable ni de carré sommable. La définition

de la transformée de Fourier des signaux aléatoires doit donc s'effectuer dans le sens des distributions aléatoires ou en utilisant des artifices de construction comme le signal restreint à un intervalle.

On considère donc $x_T(\omega, t)$ le signal aléatoire égal à $x(\omega, t)\mathbf{1}_{[-T, T]}(t)$. Chaque réalisation est alors de carré sommable (à certaines conditions pas trop restrictives sur les lois de probabilité gérant le processus) et on peut définir

$$X_T(\omega, \nu) = \int x_T(\omega, t)e^{-2i\pi\nu t} dt$$

Supposons maintenant le signal $x(t)$ stationnaire, et examinons les propriétés statistiques de X_T au premier et au deuxième ordre. On a

$$\begin{aligned} E[X_T(\omega, \nu)] &= \int E[x_T(\omega, t)]e^{-2i\pi\nu t} dt \\ &= m \int_{-T}^T e^{-2i\pi\nu t} dt \end{aligned}$$

de sorte que $\lim_{T \rightarrow +\infty} E[X_T(\omega, \nu)] = m\delta(\nu)$. Calculons maintenant le moment d'ordre 2. On a

$$\begin{aligned} E[X_T(\nu)X_T^*(\nu)] &= \int \int E[x_T(\omega, t_1)x_T(\omega, t_2)^*]e^{-2i\pi\nu t_1}e^{2i\pi\nu t_2} dt_1 dt_2 \\ &= \int \int \mathbf{1}_{[-T, T]}(t_1)\mathbf{1}_{[-T, T]}(t_2)\Gamma_{xx}(t_1 - t_2)e^{-2i\pi\nu(t_1 - t_2)} dt_1 dt_2 \end{aligned}$$

En effectuant le changement de variables $u = t_1 - t_2, v = t_1$ on obtient

$$\begin{aligned} E[|X_T(\nu)|^2] &= \int \int \mathbf{1}_{[-T, T]}(v)\mathbf{1}_{[-T, T]}(v - u)\Gamma_{xx}(u)e^{-2i\pi\nu u} dudv \\ &= \int \Gamma_{\mathbf{1}\mathbf{1}}(u)\Gamma_{xx}(u)e^{-2i\pi\nu u} du \end{aligned}$$

où est la fonction de corrélation de l'indicatrice de $[-T, T]$. Cette fonction se calcule aisément et vaut $2T(1 - |u|/2T)\mathbf{1}_{[-2T, 2T]}(u)$. Soit $\gamma_{xx}(\nu)$ la transformée de Fourier de $\Gamma_{xx}(\tau)$. On obtient alors

$$\frac{1}{2T}E[|X_T(\nu)|^2] = \gamma_{xx}(\nu) \star TF[1 - \frac{|\tau|}{2T}\mathbf{1}_{[-2T, 2T]}(u)]$$

La fonction triangle $(1 - \frac{|\tau|}{2T})\mathbf{1}_{[-2T, 2T]}(u)$ tend à s'élargir de plus en plus au fur et à mesure que T grandit. On peut montrer rigoureusement que la limite de sa TF quand T tend vers l'infini est un dirac en zéro. On a alors le résultat suivant, théorème de Wiener Khintchine,

$$\lim_{T \rightarrow +\infty} \frac{1}{2T}E[|X_T(\nu)|^2] = \gamma_{xx}(\nu) = TF[\Gamma_{xx}(\tau)]$$

Cette définition de la densité spectrale de puissance γ_{xx} s'étend de la même manière aux densités spectrales de puissance d'interaction

$$\lim_{T \rightarrow +\infty} \frac{1}{2T}E[X_T(\nu)Y_T^*(\nu)] = \gamma_{xy}(\nu) = TF[\Gamma_{xy}(\tau)]$$

Transformation des grandeurs statistiques par filtrage linéaire, formule des interférences. Il est très utile d'examiner la transformation subie par filtrage linéaire invariant des grandeurs statistiques définissant le processus. Pour cela on considère une approche très générale qui conduit à la formule des interférences. Soient $h_1(t), h_2(t)$ deux réponses impulsionnelles de deux filtres linéaires invariants. Leurs gains complexes sont notés $H_i(\nu)$. Soient $x_1(t)$ et $x_2(t)$ deux signaux aléatoires qui attaquent h_1 et h_2 respectivement, donnant naissance aux sorties y_1 et y_2 respectivement. On a alors le résultat suivant, appelé formule des interférences

$$\gamma_{y_1 y_2}(\nu) = H_1(\nu)H_2^*(\nu)\gamma_{x_1 x_2}(\nu)$$

Ce résultat se montre par un calcul direct de l'intercorrélacion entre les sorties des deux filtres puis en prenant la transformée de Fourier. En particulier, la formule des interférences montre que la DSP de la sortie d'un filtre de gain complexe H est égale au module carré du gain multiplié par la DSP de l'entrée.

3.4 Brownien et bruit blanc

Nous allons illustrer les paragraphes précédents en introduisant et en étudiant le mouvement Brownien et son processus incrément. Le mouvement Brownien est une idéalisation mathématique des phénomènes de diffusion en physique. Il s'agit d'un processus à temps continu que l'on peut construire comme processus limite d'un processus à temps discret lorsque le pas de temps tend vers 0. Définissons ce processus.

$$\begin{aligned} B(0) &= 0 \text{ p.s.} \\ B(ndt) &= B((n-1)dt) + \varepsilon_n \end{aligned}$$

où les $\{\varepsilon_n\}_n$ sont des variables aléatoires i.i.d. (indépendantes et identiquement distribuées) prenant la valeur $\pm\eta$ avec probabilité 1/2. dt est le pas de temps qui sera amené à tendre vers 0. Le processus peut s'écrire également

$$B(ndt) = \sum_{k=1}^n \varepsilon_k$$

On montre alors facilement qu'il est centré est que sa variance est donnée par $n\eta^2$. Soit t une date quelconque et n la partie entière de t/dt . On a alors $B(t) = B(ndt)$. Calculons la fonction caractéristique de $B(t)$. Il vient

$$\begin{aligned} \varphi_B(u) &= E \left[e^{iuB(t)} \right] \\ &= E \left[e^{iu \sum_{k=1}^n \varepsilon_k} \right] \\ &= E \left[\prod_{k=1}^n e^{iu\varepsilon_k} \right] \\ &= \prod_{k=1}^n E \left[e^{iu\varepsilon_k} \right] \end{aligned}$$

la dernière égalité en vertu de l'indépendance des ε . Comme cette variable prend les valeurs $\pm\eta$ équiprobablement on a $E \left[e^{iu\varepsilon_n} \right] = \cos u\eta$, et par suite

$$\begin{aligned} \varphi_B(u) &= (\cos u\eta)^n \\ \log \varphi_B(u) &= n \log(\cos u\eta) \end{aligned}$$

Comme nous allons regarder le processus limite, la taille des sauts η ne peut pas être trop grande et doit même tendre vers 0 pour assurer la continuité. On développe alors le cos autour de 0, et en utilisant $\log(1-x) \approx -x$ on a

$$\begin{aligned} \log \varphi_B(u) &\approx n \log\left(1 - \frac{u^2\eta^2}{2}\right) \\ &\approx -n \frac{u^2\eta^2}{2} \end{aligned}$$

A t fixé, $dt \rightarrow 0$ est équivalent à $n \rightarrow +\infty$. Il semble alors que le développement précédent diverge. Il faut lever cette difficulté en remarquant que $\eta \rightarrow 0$ pour assurer la continuité. Le logarithme de la fonction caractéristique de $B(t)$ possède donc une limite finie si $n\eta^2$ est fini. Comme $n \approx t/dt$ on voit que η doit tendre vers zéro comme \sqrt{dt} . On choisit alors $\eta = \sigma\sqrt{dt}$ pour obtenir

$$\lim_{dt \rightarrow 0} \varphi_B(u) = e^{-\frac{\sigma^2 t u^2}{2}}$$

qui n'est rien d'autre que la fonction caractéristique d'une variable aléatoire gaussienne centrée de variance $\sigma^2 t$. La fonction de corrélation de $B(t)$ est aisée à obtenir. Calculons $E[B(t_1)B(t_2)]$. On a

$$E[B(t_1)B(t_2)] = \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} E[\varepsilon_{k_1}\varepsilon_{k_2}]$$

Supposons $t_1 > t_2$. Alors dans la double somme précédente seul le terme $E[\varepsilon_{n_2}^2] = n_2\eta^2 = \sigma^2 t_2$ reste puisque tous les autres sont nuls (indépendance et variables centrées). Si $t_1 < t_2$ on obtient $E[\varepsilon_{n_1}^2] = \sigma^2 t_1$ et dont $\Gamma_{BB}(t_1, t_2) = \sigma^2 \min(t_1, t_2)$.

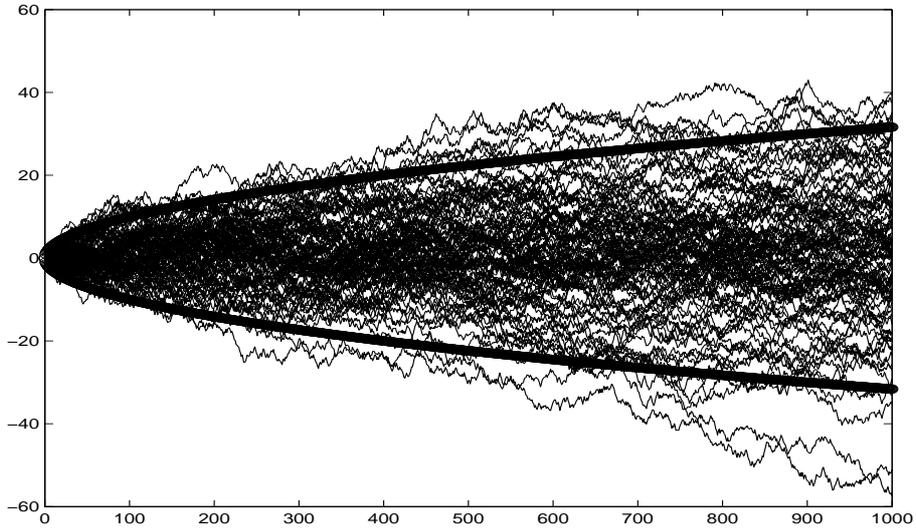


Figure 9: 100 réalisations du mouvement brownien, superposées à $\pm\sqrt{t}$

L'ensemble de ces résultats montre que le mouvement Brownien est un processus gaussien non stationnaire. La figure (9) l'illustre. 100 réalisations sont tracées et la loi \sqrt{t} est superposée. Le brownien est continu mais non dérivable. En effet $\lim E[(B(t+\varepsilon) - B(t))^2] = 0$ mais $\lim E[(B(t+\varepsilon) - B(t))^2/\varepsilon^2] = \infty$. Toutefois le processus incrément peut être défini par $X(t) = \varepsilon^{-1/2}(B(t+\varepsilon) - B(t))$. Sa variance est donnée par σ^2 et sa fonction de corrélation $E[X(t+\tau)X(t)]$ est nulle pour tout τ . Ce résultat n'est pas étonnant puisque par construction les incréments du Brownien sont indépendants (ce sont les ε_i du processus discret. Le processus incrément est appelé bruit blanc et sa fonction de corrélation est donnée par

$$\Gamma_{xx}(\tau) = E[X(t)X(t-\tau)] = \sigma^2\delta(\tau)$$

Sa densité spectrale de puissance est alors donnée par

$$\gamma_{xx}(\nu) = \sigma^2$$

qui montre que ce processus n'est pas physique, puisque sa puissance est infinie. Ceci vient du fait que la dérivée du Brownien n'existe pas. Le mot blanc vient de l'analogie avec la lumière blanche qui contient à égale puissance toutes les fréquences du visible.

3.5 Notion d'ergodisme, mesure des corrélation et DSP

La théorie des signaux aléatoires est construite pour modéliser des signaux naturels erratiques et gérés *a priori* par le hasard. Toutefois, dans les applications physiques où des signaux sont effectivement mesurés, nous n'avons accès qu'à une seule des réalisations du signal aléatoire. S'il s'agit alors de qualifier le signal aléatoire à partir de cette réalisation, le problème semble impossible, et la théorie des signaux aléatoires inutile. Cette théorie n'aura d'intérêt pratique que si elle permet de décrire les paramètres pertinents d'une réalisation particulière. L'ergodisme est la théorie, très difficile, qui justifie cela.

Lorsque l'on mesure une réalisation d'un signal aléatoire, les seules grandeurs moyennes que l'on peut calculer sont des moyennes temporelles du type

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T f(x(t))dt$$

où f est une fonction du signal $x(t)$. On dit alors que le signal aléatoire x est ergodique si les moyennes temporelles sont identiques presque-sûrement, c'est-à-dire l'ensemble des $\omega \in \Omega$ des réalisations pour lesquelles ce n'est pas vrai est de mesure nulle.

En particulier, si un signal est ergodique on a alors

$$\begin{aligned} \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T x(t)dt &= a \\ \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T x(t)x^*(t-\tau)dt &= g(\tau) \end{aligned}$$

Attention, à ce point il ne faut confondre les moyennes temporelles et les moyennes d'ensemble. Les deux points de vue se rejoignent toutefois si on analyse un signal ergodique et stationnaire. En effet on montre alors que

$$\begin{aligned}\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T x(t) dt &= E[x(t)] = m \\ \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T x(t)x^*(t-\tau) dt &= E[x(t)x^*(t-\tau)] = \Gamma_{xx}(\tau)\end{aligned}$$

En effet, pour la corrélation, on a

$$\begin{aligned}E \left[\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T x(t)x^*(t-\tau) dt \right] &= \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T \Gamma_{xx}(\tau) dt \\ &= \Gamma_{xx}(\tau).\end{aligned}$$

L'ergodisme permet en outre de mesurer les grandeurs statistiques d'un signal aléatoire stationnaire à partir d'une seule de ses réalisations. En effet, puisque sous les conditions d'ergodisme et de stationnarité il y a équivalence entre moyenne temporelle et fréquentielle, il est sensé de mesurer la fonction de corrélation du signal par

$$\widehat{\Gamma}_{xx}(\tau) = \frac{1}{2T} \int_{-T}^T x(t)x^*(t-\tau) dt$$

qui à la limite $T \rightarrow \infty$ donne la vraie fonction de corrélation. Le problème pratique liée au temps fini entre alors en ligne de compte, et on dira que $\widehat{\Gamma}_{xx}(\tau)$ est un estimateur de la fonction de corrélation, estimateur dont il faut étudier les caractéristiques statistiques. En effet, en tant que moyenne sur une durée finie, cette grandeur est aléatoire. On la qualifie alors en étudiant sa moyenne et sa variance. L'estimateur sera dit sans biais si sa valeur moyenne est précisément égale à la grandeur qu'il estime. La variance mesure alors la dispersion de l'estimateur autour de sa moyenne. Une petite variance est en générale souhaitable. Examinons biais et variance pour l'estimateur de la moyenne

$$\widehat{m} = \frac{1}{2T} \int_{-T}^T x(t) dt$$

On montre sans peine que $E[\widehat{m}] = m$ et que cet estimateur est sans biais. On montre également que

$$E[\widehat{m}^2] = \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|u|}{2T}\right) \Gamma_{xx}(u) du$$

Dans le cas où la moyenne est nulle cette expression est également celle de la variance. Nous nous plaçons dans ce cas. Si le signal est un bruit blanc, $\Gamma_{xx}(u) = \sigma^2 \delta(u)$ et on obtient

$$E[\widehat{m}^2] = \frac{\sigma^2}{2T}$$

Si le processus est à corrélation exponentielle $\Gamma_{xx}(u) = \sigma^2 \exp(-|u|/\tau)$ on obtient

$$E[\widehat{m}^2] = \frac{\sigma^2 \tau}{T} \left(1 - \frac{\tau}{2T} \left(1 - e^{-2T/\tau}\right)\right)$$

Cette variance est tracée comme une fonction de $\tau/2T$ sur la figure (10). On s'aperçoit, conformément aux hypothèses d'ergodisme, que la variance tend vers zero quand T tend vers l'infini, τ étant fixé. De plus, le comportement pour T grand est $\sigma^2 \tau/T$, résultat qui se dégrade à mesure que τ grandit. En regard du résultat pour le bruit blanc, cela montre que la corrélation rend plus difficile les tâches d'estimation. Ceci est logique et montre que les nouveaux échantillons d'un signal corrélé apportent moins d'information que ceux d'un signal décorrélé.

Pour terminer ce chapitre metionnons le fait que toutes les idées présentées ici s'étendent sans difficultés au cas des signaux à temps discrets. Toutefois, les grandeurs spectrales seront alors définies en utilisant les transformées en Z ou en fréquences réduites. Mais la modélisation aléatoire telle que présentée plus haut s'étend directement étant donnée sa nature discrète.

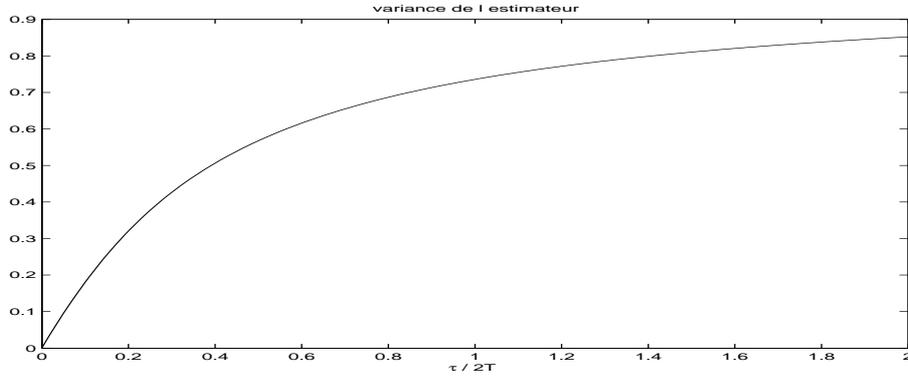


Figure 10: variance de l'estimateur de la moyenne pour du bruit exponentiel en fonction de $\tau/2T$.

4 Processus de markov, méthodes MCMC

Les processus de Markov sont des processus ayant une structure temporelle particulière et sont intimement liés à la théorie des systèmes. Ils sont également à la source de développements récents en traitement du signal qui utilisent ces processus dans des techniques numériques de calcul et d'optimisation appelées méthodes Monte-Carlo par chaînes de Markov.

Le terme processus de Markov est en général préféré pour le temps continu alors que la dénomination chaînes est plutôt employée pour le temps discret. Nous ne ferons pas la distinction ici. De même on doit distinguer les processus à états discrets des processus à états continus. Nous ne considérerons ici que les états continus.

4.1 Définitions, propriétés

Un processus est de Markov si le passé et le futur sont indépendants conditionnellement au présent. Soit x_k un processus à temps discret. Soit $\mathcal{P}_{x,k}$ un évènement lié au passé de x_k et $\mathcal{F}_{x,k}$ un évènement lié au futur de x_k . Alors le processus est de Markov si

$$\Pr(\mathcal{P}_{x,k}, \mathcal{F}_{x,k} | x_k) = \Pr(\mathcal{P}_{x,k} | x_k) \Pr(\mathcal{F}_{x,k} | x_k)$$

De cette définition très générale on tire la propriété fondamentale

$$p(x_k | x_{k-1}, x_{k-2}, \dots) = p(x_k | x_{k-1})$$

En effet,

$$\begin{aligned} p(x_k | x_{k-1}, x_{k-2}, \dots) &= \frac{p(x_k, x_{k-1}, x_{k-2}, \dots)}{p(x_{k-1}, x_{k-2}, \dots)} \\ &= \frac{p(x_k, x_{k-2}, \dots) | x_{k-1} p(x_{k-1})}{p(x_{k-1}, x_{k-2}, \dots)} \\ &= \frac{p(x_k | x_{k-1}) p(x_{k-2}, \dots) | x_{k-1} p(x_{k-1})}{p(x_{k-1}, x_{k-2}, \dots)} \\ &= \frac{p(x_k | x_{k-1}) p(x_{k-1}, x_{k-2}, \dots) | x_{k-1}}{p(x_{k-1}, x_{k-2}, \dots)} \end{aligned}$$

qui montre le résultat. Dans le cas d'états discrets, les grandeurs précédentes sont des probabilités alors qu'il s'agit de densités dans le cas de valeurs continues. Dans le cas discret, $p(x_k | x_{k-1})$ est une matrice appelée matrice de transition, alors que dans le cas continu $p(x_k | x_{k-1})$ est appelée densité de transition.

La densité de transition sert à la définition d'un opérateur dit de Markov qui envoie une densité de probabilité sur une autre densité de probabilité. Cet opérateur s'écrit

$$f_k(x) = \int p(x|y) f_{k-1}(y) dy$$

La densité de transition est également appelée noyau de transition. La densité de transition est telle que $\int p(x|y) dx = 1$ et traduit le fait que d'où que l'on parte à la date $k-1$ on est sûr d'arriver quelque part à la date k . **Quelques propriétés**

Composition La composition du noyau de transition permet de trouver la probabilité de passer d'un état à la date $k - 2$ vers un état à la date k . Il s'agit de l'équation de Chapman-Kolmogorov qui s'écrit

$$\begin{aligned} p(x_k|x_{k-2}) &= \int p(x_k, x_{k-1}|x_{k-2})dx_{k-1} \\ &= \int p(x_k|x_{k-1}, x_{k-2})p(x_{k-1}|x_{k-2})dx_{k-1} \\ &= \int p(x_k|x_{k-1})p(x_{k-1}|x_{k-2})dx_{k-1} \end{aligned}$$

et qui montre comment se composent les densités de transitions.

Mélange Le mélange de noyaux de transition est encore un noyau de transition. Par mélange, on entend

$$p(x|y) = \sum_{i=1}^n a_i p_i(x|y)$$

où les a_i sont tels que $\sum_i a_i = 1$ et les p_i sont n noyaux de transition.

Réversibilité Soit f une densité. On dit que le noyau est réversible par rapport à f si la probabilité de partir d'un ensemble A pour aller vers un ensemble B est la même que partir de B pour aller vers A , sachant que x a pour densité f , soit

$$\int_{x_k \in B} \int_{x_{k-1} \in A} p(x_k|x_{k-1})f(x_{k-1})dx_{k-1}dx_k = \int_{x_k \in A} \int_{x_{k-1} \in B} p(x_k|x_{k-1})f(x_{k-1})dx_{k-1}dx_k$$

Densité stationnaire ou invariante Une densité f est invariante pour le noyau de transition si

$$f(x_k) = \int p(x_k|x_{k-1})f(x_{k-1})dx_{k-1}$$

Si un noyau de transition est réversible par rapport à f alors f est une densité invariante. En effet, puisque pour tout ensemble A et B on a la condition de réversibilité

$$\int_{x_k \in B} \int_{x_{k-1} \in A} p(x_k|x_{k-1})f(x_{k-1})dx_{k-1}dx_k = \int_{x_k \in A} \int_{x_{k-1} \in B} p(x_k|x_{k-1})f(x_{k-1})dx_{k-1}dx_k$$

il suffit de choisir $B = \mathbb{R}$ pour obtenir

$$\begin{aligned} \int_{x_k \in \mathbb{R}} \int_{x_{k-1} \in A} p(x_k|x_{k-1})f(x_{k-1})dx_{k-1}dx_k &= \int_{x_{k-1} \in A} f(x_{k-1})dx_{k-1} \\ &= \int_{x_k \in A} \int_{x_{k-1} \in \mathbb{R}} p(x_k|x_{k-1})f(x_{k-1})dx_{k-1}dx_k \end{aligned}$$

La première ligne vient du fait que $\int_{\mathbb{R}} p(x|y)dx = 1$ (la probabilité d'arriver quelque part dans l'ensemble d'arrivée est 1) et la deuxième ligne est la réécriture de la condition de réversibilité. Puisque l'égalité est vraie pour tout A on montre que f est bien invariante.

Si des noyaux possèdent la même densité invariante, alors la composition et le mélange de ces noyaux ont également cette densité comme densité invariante

Homogénéité D'une manière générale, le noyau de transition peut dépendre explicitement du temps. Si ce n'est pas le cas, on dit que le processus de Markov est homogène.

Extension au cas multidimensionnel La définition des processus markoviens s'étend naturellement au cas des signaux multidimensionnels. Un signal de dimension n est un vecteur aléatoire indexé par le temps, soit encore un vecteur de signaux aléatoires. La description statistique se fait de la même façon que pour les signaux monodimensionnels mais cette fois en considérant les statistiques croisées entre toutes les composantes. Un signal multidimensionnel est de Markov si passé et futur sont indépendants conditionnellement au présent. Cette définition conduit alors de la même manière que pour le cas monodimensionnel à

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \dots) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$$

où une lettre grasse représente un vecteur.

Conservation du caractère Markovien par transformation.

Le caractère markovien est conservé par les bijections. D'une façon générale toute autre transformation détruit le caractère markovien. De plus un signal multidimensionnel peut être de markov sans que ses composantes le soit de manière individuelle, et inversement, deux signaux monodimensionnels markovien ne forment pas nécessairement un signal bidimensionnel markovien.

4.2 Un exemple, le modèle AR

Considérons des modèles de signaux suivants. Soit $\{\varepsilon_k\}_k$ une suite de variables aléatoires indépendantes et identiquement distribuées, et soit $f(x, k)$ une fonction de deux variables *a priori* quelconque. On définit le signal x_k par l'équation de récursion

$$x_k = f(x_{k-1}, k) + \varepsilon_k$$

Ce modèle définit un processus de Markov puisque x_k ne dépend de son passé qu'à travers l'instant précédent. Ce modèle est appelé modèle auto-régressif d'ordre 1 (cette terminologie est en général donnée au cas où la fonction f est linéaire en x et le modèle ci-dessus peut alors être appelé AR(1) non linéaire). Ce modèle est à mettre en parallèle de la définition discrète du mouvement brownien. Il en constitue une sorte de généralisation. On peut aisément déterminer la densité de transition du processus. Soit $p_\varepsilon(u)$ la densité de probabilité des variables ε . Alors la densité de x à l'instant k conditionnée par x à l'instant $k-1$ n'est autre que la densité de ε décentrée de la quantité $f(x_{k-1}, k)$ c'est-à-dire $p(x_k | x_{k-1} = y) = p_\varepsilon(x_k - f(y, k))$. La densité de l'état x à l'instant k suit alors la récursion

$$p_k(x) = \int_{\mathbb{R}} p_\varepsilon(x - f(y, k)) p_{k-1}(y) dy$$

L'existence d'une densité stationnaire n'est pas triviale à montrer et requiert des développements mathématiques dépassant ce cours.

Un cas plus simple à étudier est le modèle AR(1) linéaire, c'est-à-dire

$$x_k = ax_{k-1} + \varepsilon_k$$

On peut développer cette récursion pour obtenir

$$x_k = a^k x_0 + \sum_{i=0}^{k-1} a^i \varepsilon_{k-i}$$

si la condition initiale est en zéro et

$$x_k = a^{k+n} x_{-n} + \sum_{i=0}^{k-1+n} a^i \varepsilon_{k-i}$$

si la condition initiale est en $-n$. Ceci permet de rejeter la condition initiale en $-\infty$ en faisant tendre n l'infini. Ainsi, on s'aperçoit de suite que l'on doit avoir $|a| < 1$ pour que le signal existe. Alors, si n tend vers l'infini, la condition initiale est oubliée et

$$x_k = \sum_{i=0}^{+\infty} a^i \varepsilon_{k-i}$$

Comme les ε sont i.i.d., le signal x_k est stationnaire, au moins au second ordre. En effet, $E[x_k]$ est indépendant du temps et on calcule

$$\begin{aligned} E[x_k x_{k+n}] &= \sum_{i,j} a^i a^j E[\varepsilon_{k-i} \varepsilon_{k+n-j}] \\ &= \sum_{i,j} a^i a^j \sigma_\varepsilon^2 \delta_{i,j-n} \\ &= \sigma_\varepsilon^2 \sum_i a^i a^{i+n} \\ &= \frac{\sigma_\varepsilon^2}{1-a^2} a^n \end{aligned}$$

On obtient alors la puissance $E[x_k^2] = \sigma_\varepsilon^2 / (1-a^2)$. La densité stationnaire n'est pas aisée à obtenir. Supposons toutefois les ε gaussiens et centrés. Alors comme somme de variables gaussiennes le signal x_k est de densité gaussienne. On vérifie alors que la loi $\mathcal{N}(0, \sigma_\varepsilon^2 / (1-a^2))$ est une densité stationnaire.

Les modèles AR se généralisent à des ordres plus élevés selon

$$x_k = a_1 x_{k-1} + a_2 x_{k-2} + \dots + a_p x_{k-p} + \varepsilon_k$$

Ce signal n'est pas markovien car clairement x_k dépend de son passé *via* plusieurs dates antérieures. Toutefois, on construit à l'aide du modèle précédent un signal vectoriel markovien en introduisant le vecteur $\mathbf{x}_k = (x_k x_{k-1} x_{k-2} \dots x_{k-p})^T$. On a alors

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}\mathbf{x}_{k-1} + \boldsymbol{\varepsilon}_k \\ \mathbf{A} &= \begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_{p-1} & a_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & \dots & \dots & 0 & 1 & 0 \end{pmatrix} \\ \boldsymbol{\varepsilon}_k &= (\varepsilon_k 0 \dots 0) \end{aligned}$$

Clairement, le vecteur \mathbf{x}_k ne dépend de son passé que par \mathbf{x}_{k-1} . Le modèle AR(p) possède une représentation vectorielle qui est un processus de Markov. Ceci illustre également qu'un processus multidimensionnel peut être de Markov sans que ses composantes le soient.

4.3 Méthode de Monte-Carlo pour l'intégration

Dans de nombreuses disciplines scientifiques, et le traitement du signal n'échappe pas à la règle, des techniques théoriques sophistiquées reposent sur l'évaluation d'intégrales souvent impossible à calculer. Par exemple, calculer des espérances mathématiques de fonctions compliquées sur des densités connues ou des espérances simples sur des densités compliquées peut s'avérer impossible. Ce genre de situation apparaît en traitement du signal dans les domaines de l'estimation statistique, de la détection, du filtrage non linéaire optimal, de la classification, ... En présence de ces situations difficiles, le recours à des méthodes numériques est en général sinon indispensable au moins nécessaire pour résoudre les problèmes et acquérir des connaissances. Une des techniques d'intégration numérique est la méthode de Monte-Carlo.

Nous allons illustrer le propos en présentant le problème de l'intégration numérique. Nous développer quelques techniques pour illustrer la problématique, puis nous tourner vers les techniques utilisant les chaînes de Markov. Considérons un problème générique de calcul d'espérance mathématique. Soit $f(x)$ une densité de probabilité et h une fonction. On cherche à calculer $E_f[h(X)]$, c'est-à-dire à évaluer

$$E_f[h(X)] = \int h(x)f(x)dx$$

La méthode de Monte-Carlo consiste à approcher l'espérance précédente par une moyenne empirique utilisant N réalisations d'une variable aléatoire. Soit x_1, x_2, \dots, x_N N réalisations de la variable $X \sim f$ (\sim signifie distribué selon), alors

$$\bar{h}_N = \frac{1}{N} \sum_{i=1}^N h(x_i)$$

est un estimateur de $E_f[h(X)]$.

intermède : loi des grands nombres, théorème central limite

Loi faible des grands nombre : Si la suite x_i est i.i.d. et si $E[x] < +\infty$ alors pour tout $\varepsilon > 0$

$$\lim_{N \rightarrow +\infty} \Pr \left[\left(\frac{1}{N} \sum_{i=1}^N x_i - E[x] \right) > \varepsilon \right] = 0$$

il s'agit ici d'une convergence en probabilité.

Loi forte des grands nombre : Si la suite x_i est i.i.d. et si $E[x] < +\infty$ alors pour tout $\varepsilon > 0$ et tout $\delta > 0$, il existe N_0 tel que

$$\Pr \left[\left(\frac{1}{N} \sum_{i=1}^N x_i - E[x] \right) < \varepsilon \right] = 1 - \delta$$

pour tout $N \geq N_0$. Autrement dit la probabilité que la suite converge est 1, c'est-à-dire que l'on est en présence d'une convergence presque-sûre.

Théorème central limite : Si la suite x_i est i.i.d. et si $\text{Var}[x] < +\infty$ alors la variable

$$\frac{\frac{1}{N} \sum_{i=1}^N x_i - E[x]}{\sqrt{\text{Var}[x]}/\sqrt{N}}$$

tend en distribution vers une variable aléatoire gaussienne centrée normée.

retour à l'intégration Si on dispose d'un échantillon de N variables indépendantes distribuées selon f , alors la loi forte des grands nombre nous assure de la convergence presque-sûre de \bar{h}_N vers $E_f[h(X)]$ quand N tend vers l'infini. Ce résultat seul justifie la technique d'intégration Monte-Carlo. Elle ne dit toutefois rien sur des vitesses de convergence. Le recours à la loi faible des grands nombres ou au théorème central limite permet de qualifier la convergence et de donner des intervalles de confiance. Par exemple, si h est telle que $\text{Var}_f[h(X)]$ est finie, alors le théorème central limite indique que

$$\frac{\bar{h}_N - E_f[h(X)]}{\sqrt{\text{Var}[\bar{h}_N]}} \sim_{N \rightarrow +\infty} \mathcal{N}(0, 1)$$

Etant donnée l'indépendance des x_i , la variance $\text{Var}[\bar{h}_N]$ se calcule aisément selon

$$\begin{aligned} \text{Var}[\bar{h}_N] &= \frac{1}{N^2} \sum \text{Var}[h(x_i)] \\ &= \frac{1}{N} \text{Var}[h(x)] \\ &= \frac{1}{N} \left(\int h^2(x)f(x)dx - E_f[h(X)]^2 \right) \end{aligned}$$

qui, à l'instar de $E_f[h(X)]$, ne peut pas en général être explicitement calculée. Elle sera remplacée par son estimée empirique

$$v_N = \frac{1}{N^2} \sum_{i=1}^N h^2(x_i) - \frac{\bar{h}_N^2}{N}$$

exemple: Pour illustrer ce paragraphe, nous chercherons à calculer le moment du second ordre du mélange de gaussiennes

$$x \sim p\mathcal{N}(m_1, \sigma_1^2) + (1-p)\mathcal{N}(m_2, \sigma_2^2)$$

On obtient facilement $E[x^2] = p\sigma_1^2 + (1-p)\sigma_2^2 + pm_1^2 + (1-p)m_2^2$.

Pour obtenir un intervalle de confiance sur l'estimation, on utilise le théorème central limite (et on remplace la variance de l'estimée par son estimée empirique v_N) qui stipule que

$$Z_N = \frac{\bar{h}_N - E_f[h(X)]}{\sqrt{v_N}} \sim_{N \rightarrow +\infty} \mathcal{N}(0, 1)$$

L'intervalle de confiance à 95% est donné par $\Pr(Z_N \in [-z, z]) = 0.95$ donnant puisque Z_N est supposée centrée normée $z = 1.96$. L'intervalle de confiance est donc $\bar{h}_N \pm 1.96\sqrt{v_N}$.

Sur la figure (11) on montre le résultat de l'intégration pour $m_1 = 1, m_2 = 2, \sigma_1 = \sigma_2 = 1, p = .77$ donnant une valeur $E[x^2] = 2.69$. Sur la figure, on trace également la variance empirique qui montre la convergence.

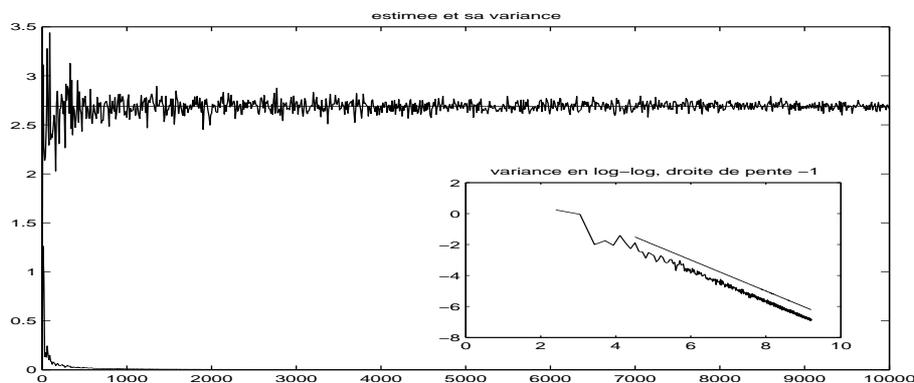


Figure 11: Intégration monte-carlo pour le calcul de $E[x^2]$ dans un mélange de gaussiennes.

Échantillonnage pondéré: Dans certains cas, le calcul Monte-Carlo précédent est soit mauvaise (variance trop élevé par exemple) soit impossible à réaliser si on ne sait pas simuler selon la loi f . Dans ce cas une astuce très

simple permet de contourner la difficulté. Cette astuce repose sur

$$\begin{aligned} E_f[h(X)] &= \int h(x)f(x)dx \\ &= \int \frac{h(x)f(x)}{g(x)}g(x)dx \\ &= E_g\left[\left(\frac{hf}{g}\right)(X)\right] \end{aligned}$$

à condition que le support de f soit inclus dans celui de g . La loi g est appelée loi instrumentale. Une approximation Monte-Carlo est alors obtenue à partir d'un N -échantillon simulé selon g par

$$\bar{h}_N = \frac{1}{N} \sum_{i=1}^N \frac{h(x_i)f(x_i)}{g(x_i)}$$

La variance de cet estimateur est alors donnée par

$$\begin{aligned} \text{Var}[\bar{h}_N] &= \frac{1}{N^2} \sum \text{Var}\left[\frac{h(x_i)f(x_i)}{g(x_i)}\right] \\ &= \frac{1}{N} \text{Var}\left[\frac{h(x)f(x)}{g(x)}\right] \\ &= \frac{1}{N} \left(\int \frac{h^2(x)f^2(x)}{g(x)}dx - E_f[h(X)]^2 \right) \end{aligned}$$

Le degré de liberté apporté par la loi instrumentale pose immédiatement la question du choix de cette loi. On peut chercher par exemple la loi qui minimise la variance précédente. Minimiser la variance revient à minimiser

$$\int \frac{h^2(x)f^2(x)}{g(x)}dx$$

sous la contrainte que le minimiseur est une densité. On minimise alors la fonctionnelle

$$\int \frac{h^2(x)f^2(x)}{g(x)}dx - \lambda \int g(x)dx$$

par rapport à g . Les équations d'Euler conduisent à

$$g^*(x) = \frac{|h(x)|f(x)}{\int |h(x)|f(x)dx}$$

En pratique on ne peut utiliser cette densité puisque le dénominateur est en général incalculable (c'est précisément l'intégrale que l'on cherche si $h(x) > 0$).

L'échantillonnage pondéré est illustré sur la figure (12). La loi instrumentale choisie est une gaussienne, la figure supérieure montrant le résultat si la variance est 4 alors que la figure inférieure montre le résultat pour une variance de 1. ON s'aperçoit que le choix de la loi instrumentale est crucial, et que l'on a intérêt à ce que les échantillons proposés par g scrutent correctement les zones de fortes probabilités de f .

Échantillonnage par acceptation-rejet : Une autre technique peut être utilisée pour générer des échantillons suivant une loi f . Il s'agit de la méthode d'acceptation et rejet. Supposons que sur le support de f on ait une autre loi g telle que $f(x) \leq Mg(x)$. Alors l'algorithme suivant

1. Générer $x \sim g$ et $u \sim \mathcal{U}([0, 1])$
2. Accepter $y = x$ si $u \leq f(x)/(Mg(x))$ sinon retourner en 1

génère une variable y distribuée suivant f .

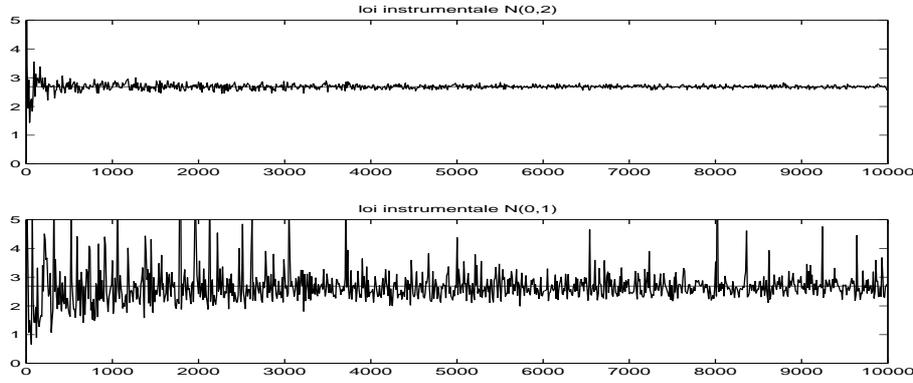


Figure 12: Intégration monte-carlo pour le calcul de $E[x^2]$ dans un mélange de gaussiennes par échantillonnage pondéré.

En effet, on cherche la loi de la variable $y|u \leq f(x)/(Mg(x))$. Or,

$$\begin{aligned}
 \Pr(y \leq y_0 | u \leq f(x)/(Mg(x))) &= \frac{\Pr(y \leq y_0, u \leq f(x)/(Mg(x)))}{\Pr(u \leq f(x)/(Mg(x)))} \\
 &= \frac{\Pr(x \leq y_0, u \leq f(x)/(Mg(x)))}{\Pr(u \leq f(x)/(Mg(x)))} \\
 &= \frac{\int_{-\infty}^{y_0} g(x) \left(\int_0^{f(x)/(Mg(x))} du \right) dx}{\int_{-\infty}^{+\infty} g(x) \left(\int_0^{f(x)/(Mg(x))} du \right) dx} \\
 &= \frac{\int_{-\infty}^{y_0} g(x) f(x) / (Mg(x)) dx}{\int_{-\infty}^{+\infty} g(x) f(x) / (Mg(x)) dx} \\
 &= \int_{-\infty}^{y_0} f(x) dx
 \end{aligned}$$

On remarque dans cette démonstration que le taux d'acceptation est $1/M$, qui est d'autant plus petit que M est grand. Sur la figure (13) on représente de haut en bas l'intégration utilisant un échantillonnage direct, un échantillonnage pondéré ($\mathcal{N}(0, 2)$) et l'échantillonnage par acceptation et rejet.

Il faut relativiser la meilleure qualité de l'acceptation-rejet par rapport aux autres. Dans l'exemple représenté, la valeur de M est environ 30, de sorte que le taux d'acceptation est de $1/30$. Pour obtenir la figure, l'algorithme AR a donc simulé 30 fois plus d'échantillons que les deux autres. Si nous avions utilisé le même nombre d'échantillon pour les autres, la variance serait divisée d'autant.

Les problèmes des méthodes précédentes résident dans le choix de la loi instrumentale et dans la probabilité d'acceptation qui peut être difficile à connaître. Pour pallier ces problèmes, des techniques alternatives ont été développées qui reposent sur les chaînes de Markov.

4.4 Méthodes MCMC

L'idée des méthodes Monte-Carlo par Chaînes de Markov (MCMC) consiste à utiliser les échantillons de la chaîne qui sont asymptotiquement distribués suivant la densité invariante de la chaîne. Précisément, s'il faut calculer

$$E_f[h(X)] = \int h(x)f(x)dx$$

à l'aide d'une approximation MCMC, il faut créer une chaîne de Markov de densité invariante f , et utiliser les échantillons à temps long x_t, x_{t+1}, \dots pour

$$\bar{h}_N = \frac{1}{N} \sum_{i=t}^{t+N-1} h(x_i)$$

Deux questions principales se posent :

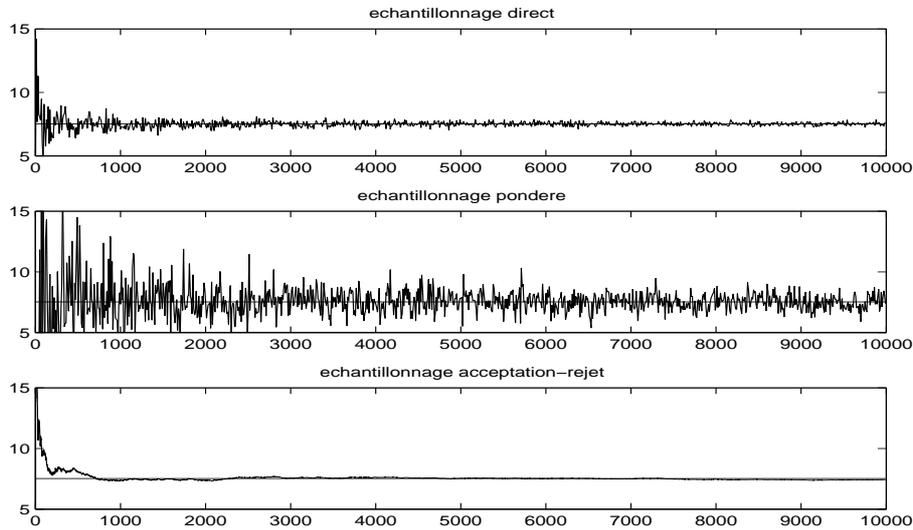


Figure 13: Intégration monte-carlo pour le calcul de $E[x^2]$ dans un mélange de gaussiennes par échantillonnage acceptation et rejet et comparaison avec l'échantillonnage direct et un échantillonnage pondéré de mauvaise qualité.

1. ayant à disposition une chaîne de Markov de densité invariante f , est-ce que l'estimateur \bar{h}_N converge (et comment?)?
2. comment construire une chaîne de Markov ayant pour distribution invariante une densité donnée f ?

Nous allons traiter ces deux points dans le sens inverse, sachant que les problèmes théoriques de convergence sont difficiles et ne seront qu'évoqués ici.

Construction de chaînes de Markov ayant une densité invariante prescrite.

La première chose à mentionner est que l'invariance d'une distribution est conservée par mélange et par composition. Cette remarque peut être utilisée pour combiner différents noyaux et avoir un algorithme plus efficace. Ceci étant dit, il y a deux grands types d'algorithmes, les algorithmes de Metropolis-Hastings, et l'algorithme de Gibbs.

1.- Metropolis-Hastings. Cet algorithme utilise une loi instrumentale qui est une densité conditionnelle $q(y|x)$. Cette densité doit être simulable pour tout x , et doit être soit connue analytiquement soit symétrique $q(y|x) = q(x|y)$. Alors l'algorithme de MH fournit une suite de v.a. x_t selon étant donnée x_t ,

1. Générer $y_0 \sim q(y|x_t)$
2. Choisir

$$x_{t+1} = \begin{cases} y_0 & \text{avec probabilité } \rho(x_t, y_0) \\ x_t & \text{avec probabilité } 1 - \rho(x_t, y_0) \end{cases}$$

où

$$\rho(x_t, y_0) = \min \left[\frac{f(y_0)q(x_t|y_0)}{f(x_t)q(y_0|x_t)}, 1 \right]$$

La suite de v.a. générée est bien une chaîne de Markov puisque conditionnellement au passé l'état courant ne dépend que de l'état à l'instant précédent. Cette chaîne a pour densité stationnaire la densité f .

Pour évaluer le noyau de la chaîne, on doit évaluer

$$\Pr [x_{t+1} \in A | x_t] = \Pr [y_0 \in A \text{ et on accepte } | x_t] + \Pr [x_t \in A \text{ et on rejette}]$$

La procédure d'acceptation et de réjection repose sur la comparaison d'une variable aléatoire u uniformément répartie sur $[0, 1]$ indépendante des autres à ρ . Donc

$$\begin{aligned} \Pr [x_{t+1} \in A | x_t] &= \int_A q(y|x_t) \left(\int_0^{\rho(x_t, y)} du \right) dy \\ &= \int_A q(y|x_t) \rho(x_t, y) dy \end{aligned}$$

Pour le deuxième terme on a

$$\begin{aligned}
\Pr [x_{t+1} = x_t \in A \text{ et on rejette } |x_t] &= \int_A \delta(y - x_t) dy \Pr (u > \rho(x_t, y)) \\
&= \int_A \delta(y - x_t) dy \int q(y|x_t) \int_{\rho(x_t, y)}^1 du dy \\
&= \int_A \delta(y - x_t) dy \int q(y|x_t) (1 - \rho(x_t, y)) dy
\end{aligned}$$

On déduit alors que la densité de transition est

$$k(x_2|x_1) = q(x_2|x_1)\rho(x_1, x_2) + \delta(x_2 - x_1) \int q(y|x_1)(1 - \rho(x_1, y)) dy$$

On vérifie bien que ce noyau est un noyau de transition puisque

$$\begin{aligned}
\int k(x_2|x_1) dx_2 &= \int q(x_2|x_1)\rho(x_1, x_2) dx_2 + \int q(y|x_1)(1 - \rho(x_1, y)) dy \int \delta(x_2 - x_1) dx_2 \\
&= 1
\end{aligned}$$

puisque $\int q(y|x_1) dy = 1$.

Montrons que ce noyau est f -réversible, ce qui assurera l'invariance de f . On veut montrer

$$k(x_2|x_1)f(x_1) = k(x_1|x_2)f(x_2)$$

Il vient

$$k(x_2|x_1)f(x_1) = q(x_2|x_1)\rho(x_1, x_2)f(x_1) + \left(1 - \int q(y|x_1)\rho(x_1, y) dy\right) \delta(x_2 - x_1)f(x_1)$$

Mais

$$\begin{aligned}
q(x_2|x_1)\rho(x_1, x_2)f(x_1) &= q(x_2|x_1) \min \left[\frac{f(x_2)q(x_1|x_2)}{f(x_1)q(x_2|x_1)}, 1 \right] f(x_1) \\
&= \min [f(x_2), f(x_1)q(x_2|x_1)] \\
&= q(x_1|x_2) \min \left[\frac{f(x_1)q(x_2|x_1)}{f(x_2)q(x_1|x_2)}, 1 \right] f(x_2) \\
&= q(x_1|x_2)\rho(x_2, x_1)f(x_2)
\end{aligned}$$

Quant au deuxième terme, on a évidemment (étant données les propriétés de la distribution de Dirac)

$$\left(1 - \int q(y|x_1)\rho(x_1, y) dy\right) \delta(x_2 - x_1)f(x_1) = \left(1 - \int q(y|x_2)\rho(x_2, y) dy\right) \delta(x_2 - x_1)f(x_2)$$

Les deux termes vérifiant la réversibilité, il en est de même pour leur somme.

Quelques exemples types de loi instrumentale :

- MH indépendant : dans ce cas, la loi instrumentale ne dépend pas de l'état précédent et $q(x|y) = q(x)$. Dans le cas où $f(x) \leq Mq(x)$ on pourrait utiliser la procédure d'acceptation et rejet, mais le MH est en général supérieur.
- MH avec marche aléatoire : l'idée est ici d'utiliser $x_{t+1} = x_t + z$ où z est une variable aléatoire de densité $q(z)$. Dans ce cas la probabilité d'acceptation s'écrit

$$\rho(x_t, y_0) = \min \left[\frac{f(y_0)q(x_t - y_0)}{f(x_t)q(y_0 - x_t)}, 1 \right]$$

qui prend une forme particulièrement simple si $q(z)$ est paire, puisqu'alors

$$\rho(x_t, y_0) = \min \left[\frac{f(y_0)}{f(x_t)}, 1 \right]$$

Cette dernière version a été proposée en 1955 par Metropolis, puis étendue en 1970 par Hastings.

Quelques illustrations :

On reprend ici les exemples des paragraphes précédents où il s'agit de calculer $E_f[h(x)]$ à l'aide de simulation Monte-Carlo. Evidemment ici, les échantillons sont simulés à l'aide d'une chaîne de Markov dont la loi invariante est la loi d'intérêt f . On utilise à nouveau le mélange de gaussiennes

$$x \sim p\mathcal{N}(m_1, \sigma_1^2) + (1-p)\mathcal{N}(m_2, \sigma_2^2)$$

On obtient facilement $E[x^2] = p\sigma_1^2 + (1-p)\sigma_2^2 + pm_1^2 + (1-p)m_2^2$.

- Utilisation du MH indépendant: la loi instrumentale $q(x|y) = q(x) = \mathcal{N}(m, \sigma^2)$. Dans un premier temps, on choisit $m = 0$ et $\sigma^2 = 100$ pour pouvoir scruter à peu près équitablement tous les états. Le résultat est montré figure (14). On s'aperçoit que le mode principal n'est pas assez visité. Pour favoriser le mode principal, on choisit $m = 10$ et les résultats apparaissent figure (15). Le choix de la loi est donc crucial, et doit permettre à tout les zones d'intérêt d'être visitées correctement par la chaîne de Markov. Par exemple, on montre les résultats désastreux de la loi $\mathcal{N}(0, 10)$ sur la figure (16). Pour être intermédiaire par rapport aux deux premières tentatives on utilise $\mathcal{N}(5, 100)$ sur la figure (17) qui donne les meilleurs résultats.
- On utilise une marche aléatoire avec des incréments gaussien $q(x|y) = \mathcal{N}(y, 10)$, le résultat étant montré figure (18). Nous avons effectué 10000 itérations, et il semble que la convergence ne soit pas encore très bonne.

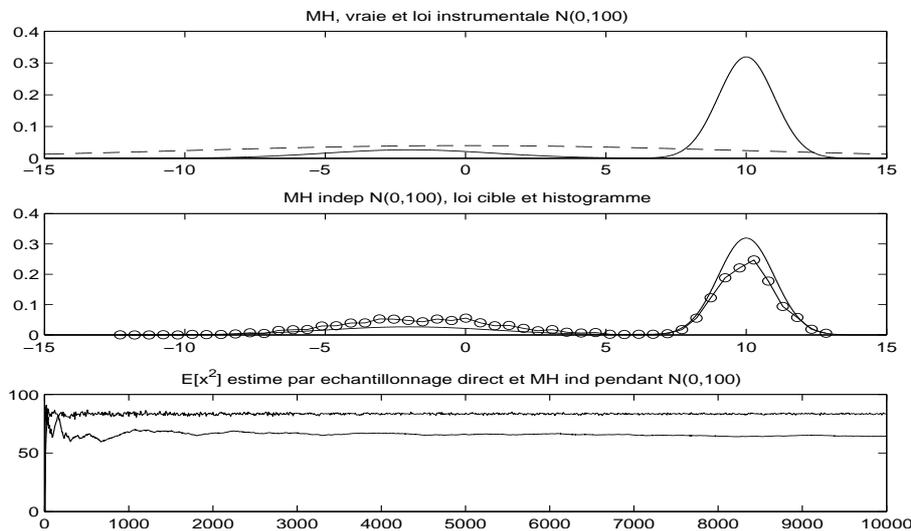


Figure 14: Simulation par MH indépendant avec comme loi instrumentale $\mathcal{N}(0, 100)$.

En pratique, comme dans la technique d'acceptation et rejet il faut que $f(x) \leq Mq(x|y)$.

Quelques variantes :

Le choix de de la forme de la probabilité d'acceptation est arbitraire. En effet, on vérifie aisément que

$$\rho(x_t, y_0) = \frac{g(y_0, x_t)}{f(x_t)q(y_0|x_t)}$$

où $g(y_0, x_t) = g(x_t, y_0)$ et assure $\rho(x_t, y_0) \in [0, 1]$ conduit également à un noyau f réversible

On peut mélanger deux algorithmes de MH pour en obtenir un autre.

On peut également mélanger des lois instrumentales pour obtenir un autre MH.

2.- Echantillonneur de Gibbs.

L'algorithme de MH illustré pour la simulation de variables monodimensionnelles est bien entendu valable dans le cas de variables multidimensionnelles. Ce cas est fondamental car les applications des méthodes MCMC concernent essentiellement des problèmes d'inférence où les grandeurs à simuler sont presque toujours multidimensionnelles. Un cas particulier de l'algorithme MH est appelé échantillonneur de Gibbs. Il fut mis au point au début des années 80, et on découvrit plus tard son lien avec les algorithmes MH.

On souhaite simuler suivant $f(x)$. Pour simplifier la présentation, on suppose que l'on peut scinder x en (x^1, x^2) . Alors l'échantillonneur de Gibbs est l'algorithme suivant :

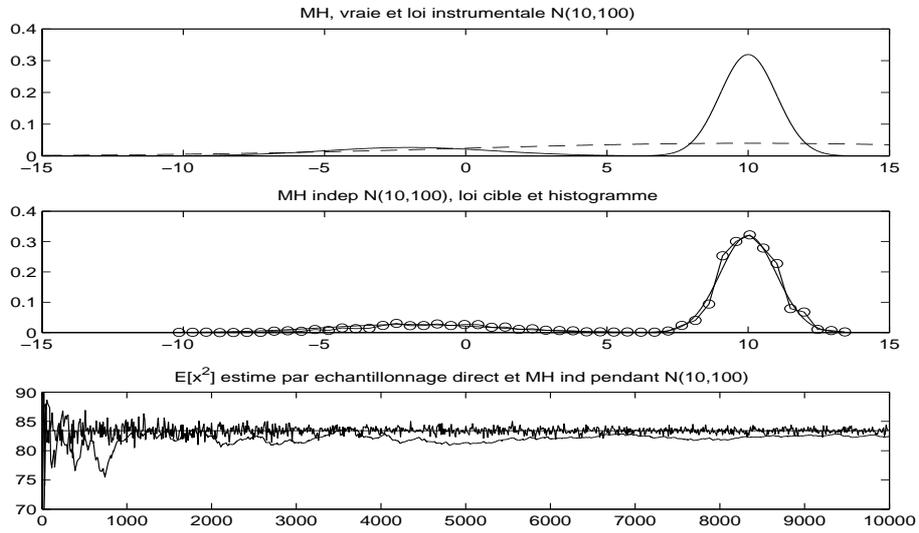


Figure 15: Simulation par MH indépendant avec comme loi instrumentale $\mathcal{N}(10, 100)$.

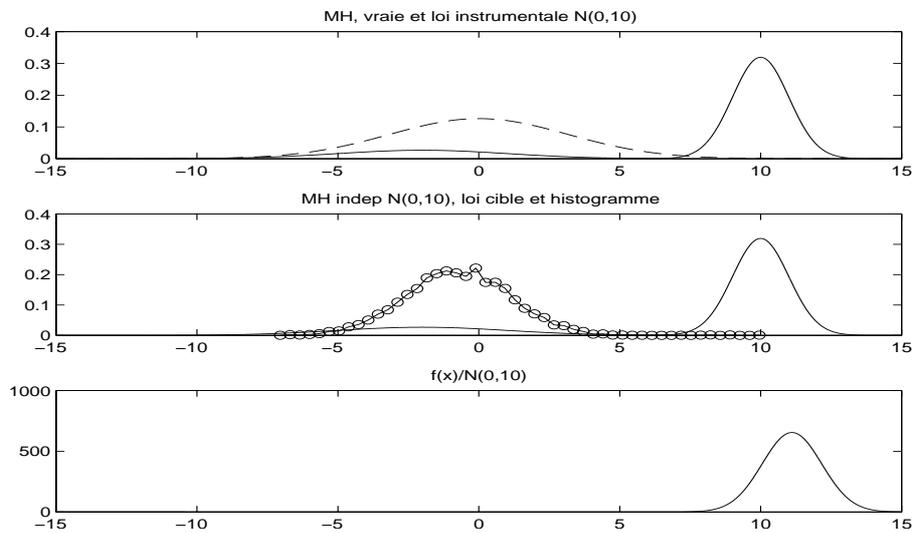


Figure 16: Simulation par MH indépendant avec comme loi instrumentale $\mathcal{N}(0, 10)$.

1. initialisation : $x_0 = (x_0^1, x_0^2)$ de manière déterministe ou aléatoire
2. itération $k \geq 1$:

$$\begin{cases} x_k^1 \sim f(x^1|x_{k-1}^2) \\ x_k^2 \sim f(x^2|x_k^1) \end{cases}$$

Il faut bien entendu pouvoir simuler suivant les densités conditionnelles. La chaîne de Markov ainsi définie a bien $f(x)$ comme densité invariante. Le noyau a pour forme

$$k(x^1, x^2|y^1, y^2) = f(x^1|y^2)f(x^2|x^1)$$

Il vient alors

$$\begin{aligned} \int k(x^1, x^2|y^1, y^2)f(y^1, y^2)dy^1 dy^2 &= \int \int f(x^1|y^2)f(x^2|x^1)f(y^1, y^2)dy^1 dy^2 \\ &= f(x^2|x^1) \int f(x^1|y^2)f(y^2)dy^2 \\ &= f(x^2|x^1)f(x^1) \\ &= f(x^1, x^2) \end{aligned}$$

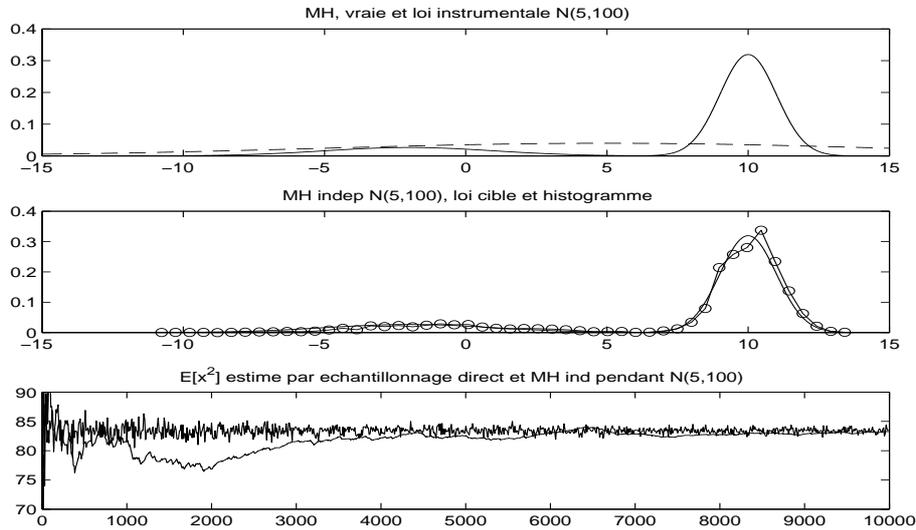


Figure 17: Simulation par MH indépendant avec comme loi instrumentale $\mathcal{N}(5, 100)$.

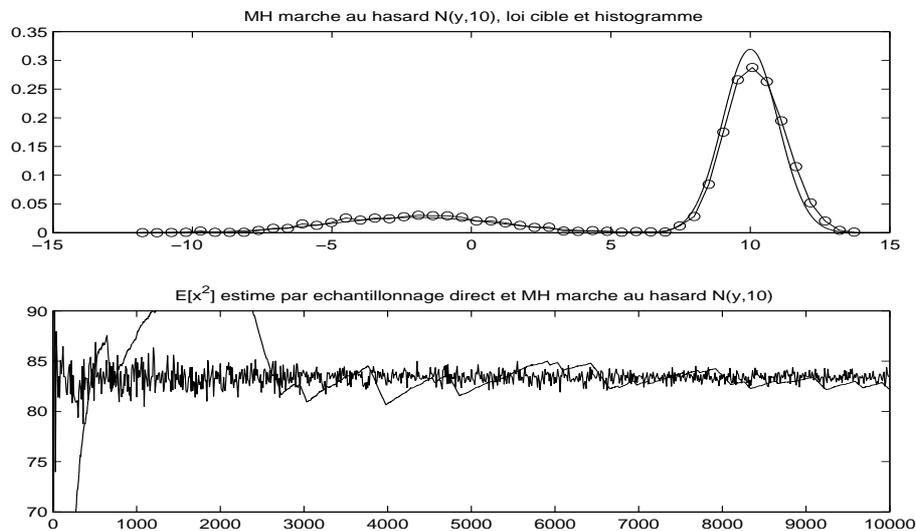


Figure 18: Simulation par MH marche au hasard avec comme loi instrumentale $\mathcal{N}(y, 10)$.

Pour illustrer le fonctionnement de l'algorithme de Gibbs, nous allons le détailler pour une loi normale bidimensionnelle

$$\mathcal{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

On montre alors facilement que $p(x_i|x_j) = p(x_i, x_j)/p(x_j)$ est la loi normale $\mathcal{N}(\rho x_j, 1 - \rho^2)$, où $i = 1, j = 2$ ou $i = 2, j = 1$. L'algorithme est illustré par la figure (19) pour $\rho = 0.7$. A gauche, on trace le nuage de points qui a bien la forme elliptique de la gaussienne, alors qu'à droite on superpose quelques itérations de la chaîne de Markov sur la densité cible pour illustrer la capacité d'exploration. Enfin, sur les données simulées, nous avons évalué le coefficients de corrélation et obtenu 0.69 pour 10000 échantillons.

Convergence.

Les problèmes de convergence sont de deux types. Le premier concerne la convergence de la densité des échantillons de la chaîne vers la densité invariante, et le deuxième concerne la convergence de l'utilisation que l'on fait de la densité, comme par exemple la convergence des moyennes des échantillons vers les intégrales que l'on souhaite évaluer.

Les propriétés de convergence sont très difficiles à étudier, et nous ne donnons ici que les résultats de base en leur donnant leur signification.

Irréductibilité : Soit f une mesure de probabilité. Une chaîne de Markov est f -irréductible si pour tout point x et tout ensemble A $f(A) > 0 \implies \exists n \in \mathbb{N}^*$ tel que $P^n(A|x) > 0$. Autrement dit, tout les ensembles de mesures

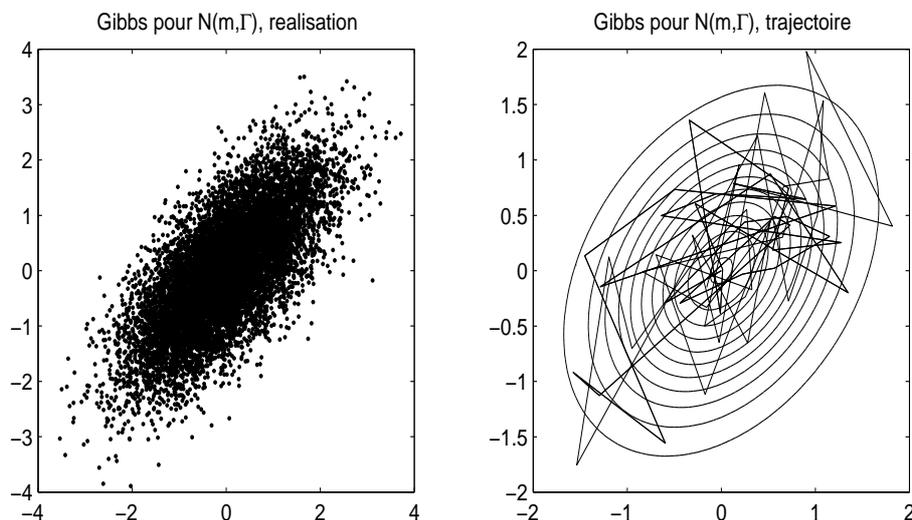


Figure 19: Simulation par Gibbs d'une gaussienne bidimensionnelle.

non-nulle pour f seront visités avec une probabilité non nulle et ce quelque soit le point de départ. On sent ici que la chaîne de Markov doit être capable de scruter l'ensemble de l'espace d'état.

Apériodicité : Il n'existe pas de partition (A_1, \dots, A_k) de l'espace des états tel que $P(x, A_{i+imodk}) = 1 \forall x \in A_i$. Autrement dit la chaîne de Markov n'a pas de comportement périodique.

Harris-récurrente : Soit x_i une chaîne de Markov de noyau P , de distribution invariante f et φ -irréductible. Elle est Harris-récurrente si pour tout ensemble A , $f(A) > 0 \implies \Pr(x_i \in A \text{ infinisment souvent}) = 1$. A nouveau la chaîne doit scruter les zones de probabilité non nulles souvent, la probabilité étant mesurée avec la mesure invariante.

Ergodicité : Une chaîne ayant une densité invariante et étant Harris-récurrente est dite ergodique

Convergence : Si la chaîne est φ -irréductible et f est la mesure invariante alors, la loi forte des grands nombres s'applique et $\Pr(h_N \rightarrow E_f[h(X)]) = 1$. Si de plus il y a apériodicité alors $\lim_{N \rightarrow +\infty} \|P^N(y|x_0) - f(y)\|_{VT} = 0$ où la norme en variation totale est définie par $\|f - g\|_{VT} = 1/2 \int |f(x) - g(x)| dx$.

Si de plus la convergence est exponentielle, $\|P^N(y|x_0) - f(y)\|_{VT} \leq c(x_0)\alpha^N$ alors le théorème central limite s'applique et $\sqrt{N}(h_N - E_f[h]) \rightarrow \mathcal{N}(0, \text{Var}h(x_0) + 2 \sum_{i=1}^{\infty} \text{Cov}(h(x_0), h(x_i)))$.

Le calcul de la variance est donc compliqué, mais cela est dû au fait que les échantillons ne sont pas indépendants. De plus, le fait qu'il ne reste qu'une somme est due au fait que la chaîne est supposé homogène (stationnaire).

Pour terminer il suffit de mentionner que les méthodes MCMC ont été illustrées ici sur des exemples très simples, et que leur puissance n'apparaît réellement que lorsqu'on les applique à des problèmes difficiles, typiquement en théorie de l'estimation lorsqu'il s'agit de connaître des moments de lois conditionnelles très complexes.

5 Introduction aux processus ponctuels et à leurs produits dérivés

Dans les processus et signaux que nous avons vu jusqu'ici, le caractère aléatoire existait dans l'amplitude des signaux. Certains problèmes nécessitent une approche différente pour modéliser des événements qui se réalisent à des instants aléatoires. Des exemples multiples existent dont :

1. la modélisation des files d'attente : l'instant d'arrivée d'un individu dans la file est aléatoire
2. les potentiels d'action, vecteur de l'information neuronale (voir la figure (20)).
3. l'arrivée de photons sur un photorécepteur
4. ...

D'une manière générale, tous les processus dans lesquels des changements apparaissent à des instants aléatoires peuvent être modélisés à partir des processus ponctuels. Nous allons ici donner quelques notions sur ces processus et voir la construction de quelques processus dérivés des processus ponctuels.

A titre d'illustration la figure (20) montre une suite de potentiels d'action générée par un modèle biophysique de neurones. Le nombre d'évènement est de 13 pour l'horizon temporel. La figure du bas présente une réalisation d'un processus de Poisson, processus qui est aux processus ponctuels ce que le processus gaussien blanc est aux processus usuels. Le taux d'évènement a été réglé de sorte à en moyenne avoir 13 événements sur l'horizon temporel considéré.

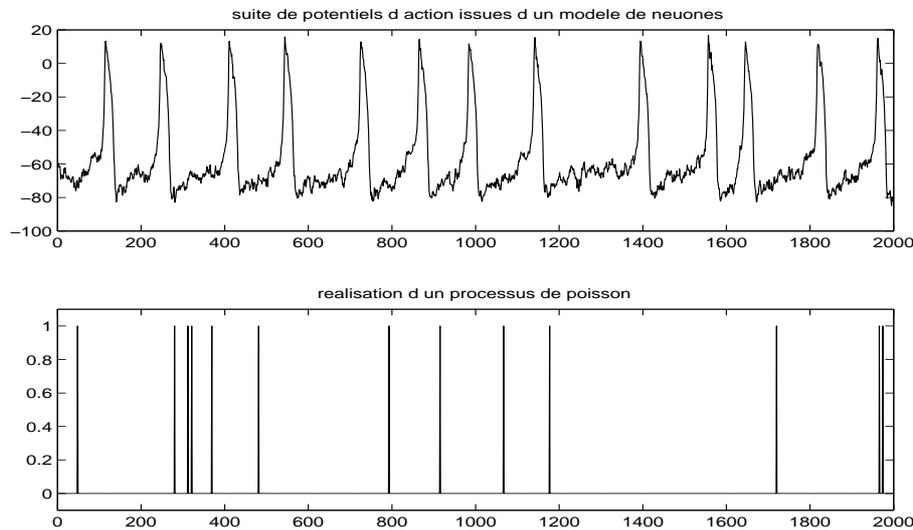


Figure 20: Illustration d'un processus reposant sur un processus ponctuel : les potentiels d'action arrive à des instants aléatoires.

5.1 Processus ponctuels

La description des processus ponctuels est assez difficile, et peut être faite de plusieurs manières : par le nombre de points par intervalle, par le temps entre les événements, comme un processus aléatoire au sens des distributions.

Processus au sens des distributions. Le processus que l'on veut décrire est l'apparition aléatoire d'évènements. Soit $\{t_i(\omega)\}$ cette liste de dates, en nombre fini ou infini. Le processus ponctuel pourra alors être écrit comme une somme de distributions de Dirac selon

$$x(t, \omega) = \sum_i \delta(t - t_i(\omega))$$

Cette forme sera particulièrement utile pour le filtrage du processus ponctuel et la définition de processus dérivés. Par exemple, on peut d'ores et déjà modifier la définition pour y inclure des amplitudes portées par les distributions de Dirac et écrire

$$x(t(\omega)) = \sum_i a_i(\omega) \delta(t - t_i(\omega))$$

les $a_i(\omega)$ pouvant être déterministes (indépendants de ω) ou aléatoires (par exemple l'énergie du photon arrivant à cette date, la taille de l'homme entrant dans une queue à cette date, ...)

Description par le nombre de points par intervalle. On peut décrire le processus par le nombre d'évènement apparaissant par intervalle. Soit $N(t, T; (\omega))$ le nombre aléatoire de dates apparaissant dans l'intervalle $[t, t + T[$. N est appelé la mesure de comptage. Si I est un intervalle, on a

$$N(I, \omega) = \sum_i \mathbf{1}_I(t_i(\omega))$$

où $\mathbf{1}_I$ est l'indicatrice de l'intervalle. Cette mesure est une mesure stochastique que l'on peut écrire également, pour des intervalles infinitésimaux, $N(dt) = dN(t) = \sum_i \delta_{t_i}(dt)$, lorsqu'on l'utilise dans l'intégrale stochastique du type

$$I = \int_0^T f(u) dN(u) = \sum_i f(t_i)$$

où les t_i sont les dates d'arrivée des évènements. Une spécification des processus ponctuels passe par la description statistique de la mesure de comptage.

Si l'on considère un instant origine t_0 , on associe à N un processus $N(t) = N(t_0) + N([t_0, t[)$. Le processus $N(t)$ est appelé le processus de comptage. D'autre part, les incréments infinitésimaux $dN(t) = N(t + dt) - N(t)$ correspondent évidemment à la mesure de comptage

. **Description par des intervalles de temps.** Au lieu de décrire les temps d'arrivée ou le nombre de ces dates par intervalles, on peut également décrire le processus par les intervalles de temps entre les dates d'arrivée. On distingue deux concepts.

Le temps de vie est le temps entre deux évènements du processus alors que le temps de survie est le temps entre une date quelconque et un évènement du processus.

5.2 Processus de Poisson

La caractérisation des processus vient par la caractérisation statistique du processus de comptage. La caractérisation la plus simple est la suivante :

1. Le nombre d'évènements dans deux intervalles disjoints sont indépendants statistiquement,
2. La probabilité d'apparition d'un évènement dans un intervalle de longueur dt est d'ordre dt ,

$$\Pr(N(t, t + dt) = 1) = \lambda(t)dt + o(dt)$$

3. La probabilité d'apparition de plus d'un évènement dans un intervalle de longueur dt est d'ordre dt^2 (c'est-à-dire nulle)
4. Le nombre d'évènements avant la date origine est nul

Soit $p_n(t) = \Pr(N(t_0, t) = n)$. On a alors

$$\begin{aligned} p_0(t + dt) &= \Pr(N(t_0, t) = n \text{ et } N(t, t + dt) = 0) \\ &= p_0(t)(1 - \lambda(t)dt) \end{aligned}$$

puisque les intervalles $[t_0, t[$ et $[t, t + dt[$ sont disjoints. p_0 suit donc l'équation différentielle $\dot{p}_0(t) + \lambda(t)p_0(t) = 0$ avec comme condition initiale $p_0(t_0) = 1$ puisque la probabilité d'avoir 0 évènement avant l'origine des temps est 1. Donc on a

$$p_0(t) = \exp\left(-\int_{t_0}^t \lambda(u)du\right)$$

Pour $n \geq 0$, il vient maintenant

$$\begin{aligned} p_{n+1}(t + dt) &= \Pr(N(t_0, t) = n \text{ et } N(t, t + dt) = 1) + \Pr(N(t_0, t) = n + 1 \text{ et } N(t, t + dt) = 0) \\ &= p_n(t)\lambda(t)dt + p_{n+1}(t)(1 - \lambda(t)dt) \end{aligned}$$

et p_{n+1} satisfait

$$\dot{p}_{n+1}(t) + \lambda(t)p_{n+1}(t) = \lambda(t)p_n(t)$$

avec $p_{n+1}(t_0) = 0$. La solution générale de l'équation différentielle est $p_{n+1}(t) = A \exp(-\int_{t_0}^t \lambda(u) du)$. Une solution particulière est donnée en utilisant la méthode de la variation des constantes et conduit à $A(t) \exp(-\int_{t_0}^t \lambda(u) du)$ où A satisfait $\dot{A}(t) = \lambda(t)p_n(t) \exp(+\int_{t_0}^t \lambda(u) du)$.

Pour $n = 0$, on a alors $\dot{A}(t) = \lambda(t)$ de sorte que

$$p_1(t) = A \exp(-\int_{t_0}^t \lambda(u) du) + \left(\int_{t_0}^t \lambda(u) du \right) \exp(-\int_{t_0}^t \lambda(u) du)$$

Appliquer la condition initiale conduit à $A = 0$, soit

$$p_1(t) = \left(\int_{t_0}^t \lambda(u) du \right) \exp(-\int_{t_0}^t \lambda(u) du)$$

On montre alors par récurrence de la même manière que

$$p_n(t) = \frac{\left(\int_{t_0}^t \lambda(u) du \right)^n}{n!} \exp(-\int_{t_0}^t \lambda(u) du)$$

On reconnaît ici la loi de Poisson de paramètre $\int_{t_0}^t \lambda(u) du$. La fonction $\lambda(t)$ est appelée le taux du processus ou intensité. Si $\lambda(t) = \lambda$, alors le paramètre de la loi est $(t - t_0)\lambda$. On dit alors que le processus de Poisson est homogène.

Quelques rappels sur la loi de Poisson. Une variable aléatoire x discrète suit la loi de Poisson de paramètre λ si

$$\Pr(x = k) = \frac{\lambda^k e^{-\lambda}}{k!} = \mathcal{P}(k, \lambda)$$

La fonction caractéristique est

$$\begin{aligned} \varphi(u) &= E[e^{iuk}] = \sum_{k \geq 0} e^{iuk} \frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda} \sum_{k \geq 0} e^{iuk} \frac{(e^{iu}\lambda)^k}{k!} = e^{-\lambda} e^{\exp(iu)\lambda} \end{aligned}$$

De là on montre facilement que les cumulants (les cumulants sont au logarithme de la fonction caractéristique ce que sont les moments à la fonction caractéristique) sont tous égaux à λ . En particulier, moyenne et variance sont égales à λ .

Temps de survie. Soit $t \leq t_0$ un instant quelconque, et $L(t)$ la variable aléatoire mesurant le temps entre t et le premier t_i du processus. Alors, dans la plupart des bouquins sur les processus de Poisson on trouve le résultat

$$P_L(l; t) dl = \lambda(t+l) \exp\left(-\int_t^{t+l} \lambda(u) du\right) dl$$

qui repose sur l'idée que entre t et $t+l$ il ne doit pas y avoir de point du processus et qu'il doit y en avoir un entre $t+l$ et $t+l+dl$. En fait, cf Picinbono, une petite erreur se glisse dans ce raisonnement. Pour qu'il soit valide il faut qu'il y ait au moins un point après t , ce qui n'est pas garanti pour une intensité non stationnaire. En fait chercher le temps de survie suppose qu'un point au moins existe après t . Donc, si on a les mêmes notations pour la probabilité conditionnelle

$$\begin{aligned} P_L(l; t) dl &= \frac{\lambda(t+l) \exp\left(-\int_t^{t+l} \lambda(u) du\right) \times \Pr(\text{au moins 1 après } t | 1 \text{ dans } [t+l, t+l+dl])}{\Pr(\text{au moins 1 après } t)} dl \\ P_L(l; t) &= \frac{\lambda(t+l) \exp\left(-\int_t^{t+l} \lambda(u) du\right)}{1 - \exp\left(-\int_t^{+\infty} \lambda(u) du\right)} \end{aligned}$$

évidemment pour $l \geq 0$. La probabilité conditionnelle au numérateur est égale à 1, et la probabilité du numérateur est égale à 1 moins la probabilité qu'il n'y ait pas d'évènement après t .

Temps de vie.

Le temps de vie est la durée entre deux évènements du processus. Soit t un évènement du processus, et L la variable aléatoire mesurant le temps entre cet évènement et le prochain. La variable L est donc conditionnée au fait qu'il y a un évènement en t et qu'il y en aura au moins un après. Par le même calcul que précédemment, on montre que

$$P_L(l; t) = \frac{\lambda(t+l) \exp\left(-\int_t^{t+l} \lambda(u) du\right)}{1 - \exp\left(-\int_t^{+\infty} \lambda(u) du\right)}$$

qui est la même distribution que le temps de survie.

Distribution conditionnelle des positions

On suppose que l'on observe n points $0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq T$ dans l'intervalle $[0, T]$. On cherche conditionnellement à cette connaissance la distribution des t_i , $P(t_1, \dots, t_n | N(T) = n)$. On a

$$P(t_1, \dots, t_n | N(T) = n) = \frac{P(t_1, \dots, t_n, N(T) = n)}{P(N(T) = n)}$$

$(t_1, \dots, t_n, N(T) = n)$ correspond à l'évènement (0 dans $[0, t_1[$, 1 dans $[t_1, t_1 + dt_1[$, 0 dans $[t_1 + dt_1, t_2[$, ..., 0 dans $[t_{n-1}, t_n[$, 1 dans $[t_n, t_n + dt_n[$, 0 dans $[t_n + dt_n, T[$. La probabilité d'avoir 1 évènement dans un intervalle infinitésimal est donné par $\int_t^{t+dt} \lambda(u) du = \lambda(t) dt$. Comme la longueur des intervalles où l'on ne doit pas voir d'évènement est T , la probabilité de la conjonction de ces évènements (indépendants par hypothèse) est égale à $\exp(-\int_0^T \lambda(u) du)$. On a donc

$$P(t_1, \dots, t_n | N(T) = n) dt_1 \dots dt_n = \frac{\lambda(t_1) dt_1 \dots \lambda(t_n) dt_n \exp(-\int_0^T \lambda(u) du)}{\exp(-\int_0^T \lambda(u) du) (\int_0^T \lambda(u) du)^n / n!}$$

et donc la densité conditionnelle s'écrit

$$P(t_1, \dots, t_n | N(T) = n) = \frac{n!}{(\int_0^T \lambda(u) du)^n} \prod_{i=1}^n \lambda(t_i)$$

On a vu au passage que la loi conjointe de la position des temps et qu'il y a n points dans l'intervalle $[0, T]$ est donnée par

$$P(t_1, \dots, t_n, N(T) = n) = \exp(-\int_0^T \lambda(u) du) \prod_{i=1}^n \lambda(t_i)$$

On vérifie bien que cette loi est une ddp en faisant bien attention au fait que $0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq T$. On écrit en effet

$$\begin{aligned} \sum_n \int P(t_1, \dots, t_n, N(T) = n) &= e^{-\int_0^T \lambda(u) du} + \sum_{n \geq 1} \int e^{-\int_0^T \lambda(u) du} \prod_{i=1}^n \lambda(t_i) dt_i \\ &= e^{-\int_0^T \lambda(u) du} (1 + I(T)) \quad \text{où} \\ I(t) &= \sum_{n \geq 1} \int_0^T dt_1 \lambda(t_1) \int_{t_1}^T dt_2 \lambda(t_2) \dots \int_{t_{n-2}}^T dt_{n-1} \lambda(t_{n-1}) \int_{t_{n-1}}^T dt_n \lambda(t_n) \end{aligned}$$

Soit $\Lambda(t)$ une primitive de $\lambda(t)$. Alors l'intégrale sur t_n vaut $\Lambda(T) - \Lambda(t_{n-1})$. On montre alors que l'intégrale sur t_i vaut $(\Lambda(T) - \Lambda(t_{i-1}))^{n-i+1} / (n-i+1)!$. L'assertion est valide pour $i = n$. On suppose vrai pour $i+1$ et on regarde ce qui se passe en i . L'intégrale en t_i s'écrit

$$\int_{t_{i-1}}^T dt_i \frac{(\Lambda(T) - \Lambda(t_i))^{n-i}}{(n-i)!} \lambda(t_i)$$

que l'on intègre par partie en dérivant la fraction et en intégrant $\lambda(t)$ en $\Lambda(t) - \Lambda(T)$. On obtient alors

$$\int_{t_{i-1}}^T dt_i \frac{(\Lambda(T) - \Lambda(t_i))^{n-i}}{(n-i)!} \lambda(t_i) = - \left[\frac{(\Lambda(T) - \Lambda(t_i))^{n-i+1}}{(n-i)!} \right]_{t_{i-1}}^T - \int_{t_{i-1}}^T dt_i \frac{(\Lambda(T) - \Lambda(t_i))^{n-i}}{(n-i-1)!} \lambda(t_i)$$

On remarque que l'intégrale de droite est la même que celle du membre de gauche, aux numérateurs (constants par rapport à t_i) près. On résoud alors en

$$\left(\frac{1}{(n-i)!} + \frac{1}{(n-i-1)!} \right) \int_{t_{i-1}}^T dt_i (\Lambda(T) - \Lambda(t_i))^{n-i} \lambda(t_i) = \frac{(\Lambda(T) - \Lambda(t_{i-1}))^{n-i+1}}{(n-i)!}$$

Comme

$$\left(\frac{1}{(n-i)!} + \frac{1}{(n-i-1)!} \right) = \frac{n-i+1}{(n-i)!}$$

on a bien le résultat excompté. Pour le calcul de $I(t)$ en posant $t_0 = 0$ on arrive donc a

$$\begin{aligned} I(t) = I(t) &= \sum_{n \geq 1} \frac{(\Lambda(T) - \Lambda(0))^n}{(n)!} \\ &= e^{-\int_0^T \lambda(u) du} - 1 \end{aligned}$$

de sorte que la normalisation à 1 de P est acquise.

On peut réécrire la vraisemblance selon

$$P(t_1, \dots, t_n, N(T) = n) = \exp \left(- \int_0^T \lambda(u) du + \sum_{i=1}^n \log \lambda(t_i) \right)$$

On introduit la mesure stochastique $dN(t) = \sum_i \delta_{t_i}(dt)$ pour écrire

$$P(t_1, \dots, t_n, N(T) = n) = \exp \left(- \int_0^T \lambda(u) du + \int_0^T \log \lambda(t) dN(t) \right)$$

où l'on voit apparaître l'intégrale stochastique du type

$$I = \int_0^T f(u) dN(u) = \sum_i f(t_i)$$

où les t_i sont les dates d'arrivée des évènements. Pour calculer les statistiques de I , on écrit

$$\begin{aligned} E[I] &= E \left[\sum_i f(t_i) \right] \\ &= E_N \left[E \left[\sum_i f(t_i) | N(T) \right] \right] \end{aligned}$$

Or, conditionnellement à $N(T) = n$ on a vu que la loi des t_i est donnée par

$$P(t_1, \dots, t_n | N(T) = n) = \frac{n!}{(\int_0^T \lambda(u) du)^n} \prod_{i=1}^n \lambda(t_i)$$

Ici, les temps sont ordonnés, et on remarque donc que cette loi est la même que la loi des statistiques d'ordre de n temps u_1, \dots, u_n tirés indépendamment les uns des autres suivant la même loi

$$\begin{aligned} P_U(u) &= \frac{\lambda(u)}{(\int_0^T \lambda(u) du)} \text{ si } 0 \leq u \leq T \\ &= 0 \text{ sinon} \end{aligned}$$

Donc calculer des statistiques de $\sum_i f(t_i)$ sur la loi des t_i conditionnée par le nombre d'évènements est équivalent à calculer les statistiques de $\sum_i f(u_i)$ sur la loi des u_i . Ainsi,

$$\begin{aligned}
E[I] &= E_N \left[\int \left(\sum_i f(u_i) \right) \frac{\prod_{i=1}^n \lambda(u_i) du_i}{\left(\int_0^T \lambda(u) du \right)^n} \right] \\
&= E_N \left[\sum_i \int f(u_i) \frac{\prod_{j=1}^n \lambda(u_j) du_j}{\left(\int_0^T \lambda(u) du \right)^n} \right] \\
&= E_N \left[\sum_i \frac{\int_0^T f(u) \lambda(u) du}{\int_0^T \lambda(u) du} \right] \\
&= \frac{\int_0^T f(u) \lambda(u) du}{\int_0^T \lambda(u) du} E_N[n] \\
&= \int_0^T f(u) \lambda(u) du
\end{aligned}$$

De plus, comme sachant le nombre de points, les t_i sont distribués comme les statistiques d'ordre d'un n-uplet de variables i.i.d. de loi $\lambda/\int \lambda$, on peut facilement générer une trace d'un processus sur un intervalle. On tire les n dates selon

$$\begin{aligned}
P_U(u) &= \frac{\lambda(u)}{\left(\int_0^T \lambda(u) du \right)} \text{ si } 0 \leq u \leq T \\
&= 0 \text{ sinon}
\end{aligned}$$

et on les ordonne. Ceci est particulièrement efficace dans le cas homogène puisqu'alors $P_U(u) = \mathbf{1}_{[0,T]}(u)/T$ est la loi uniforme sur $[0, T]$.

5.3 Bruits de grenaille

En général, les processus ponctuels ne sont pas observés directement, mais après avoir traversé un appareil de mesure, ou encore chaque évènement portant un message particulier. Un bruit de grenaille est la sortie d'un filtre linéaire invariant attaqué par un processus de Poisson. Soit $h(t)$ la réponse impulsionnelle du filtre. Le bruit de grenaille est défini par

$$\begin{aligned}
x(t) &= h(t) \star \sum_i \delta(t - t_i) \\
&= \sum_i h(t - t_i) \\
&= \int h(t - u) dN(u)
\end{aligned}$$

La moyenne du processus s'écrit

$$E[x(t)] = \int h(t - u) \lambda(u) du$$

de sorte que dans le cas d'un processus de Poisson inhomogène, la moyenne dépend du temps. Toutefois, si le processus de Poisson est homogène, alors $E[x(t)] = \lambda H(0)$, H étant la transformée de Fourier de h .

La fonction de corrélation de $x(t)$ s'écrit

$$E[x(t_1)x(t_2)] = \int \int h(t_1 - u_1) h(t_2 - u_2) E[dN(u_1)dN(u_2)]$$

le moment d'ordre 2 des incréments vaut $\lambda(u_1)\lambda(u_2)du_1du_2$ (indépendance des incréments) sauf si $u_1 = u_2$ auquel cas il vaut $\lambda(u_1)du_1$. Donc, la fonction de corrélation devient

$$E[x(t_1)x(t_2)] = \int \int h(t_1 - u_1) h(t_2 - u_2) \lambda(u_1)\lambda(u_2) du_1 du_2 + \int h(t_1 - u) h(t_2 - u) \lambda(u) du$$

de sorte que

$$\text{Cov}(x(t_1), x(t_2)) = \int h(t_1 - u) h(t_2 - u) \lambda(u) du$$

Dans le cas du processus de Poisson homogène on obtient

$$\text{Cov}(x(t_1), x(t_2)) = \lambda \int h(u)h(t_2 - t_1 + u)du$$

qui montre que le bruit de grenaille est alors stationnaire au second-ordre. Dans ce cas, on peut calculer la densité spectrale de puissance facilement. En effet, le résultat précédent montre que la fonction de corrélation est proportionnelle à la corrélation de h . On a alors

$$\gamma_x(\nu) = \lambda\gamma_h(\nu) = \lambda|H(\nu)|^2$$

On peut compliquer un peu ce modèle en ajoutant une amplitude aléatoire de sorte que

$$\begin{aligned} xx(t) &= h(t) \star \sum_i A_i \delta(t - t_i) \\ &= \sum_i A_i h(t - t_i) \\ &= \int h(t - u) dM(u) \end{aligned}$$

où la mesure de comptage vérifie maintenant $E[dM(t)] = E[A]\lambda(t)dt$ et $E[dM(t)] = E[A^2]\lambda(t)dt$ et garde ses propriétés d'indépendance d'un intervalle sur l'autre.