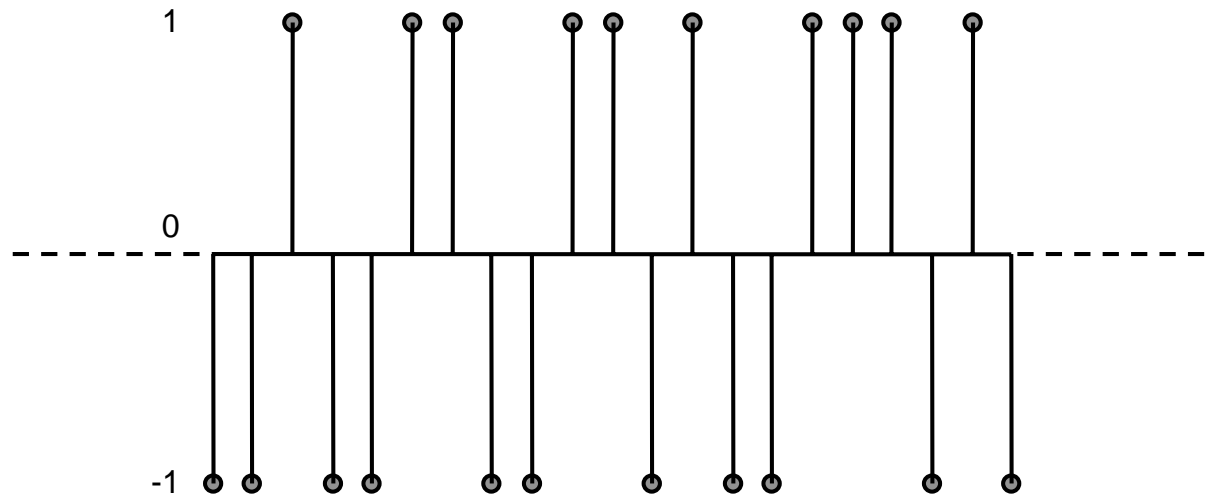


- La quasi totalité des signaux extraits de phénomènes réels (biomédicaux par exemple) présentent un aspect aléatoire. Ceci veut dire qu'il est impossible de prévoir la forme exacte des signaux obtenus, et que ceux-ci ne peuvent être décrits de façon analytique.
- Dans le cas de ces signaux, il est impossible d'appliquer directement certains concepts. En particulier, la notion de TF d'un signal aléatoire a peu de sens: la somme des termes peut ne pas être finie, et on obtient une grandeur aléatoire.

- Exemple: le processus de Bernoulli

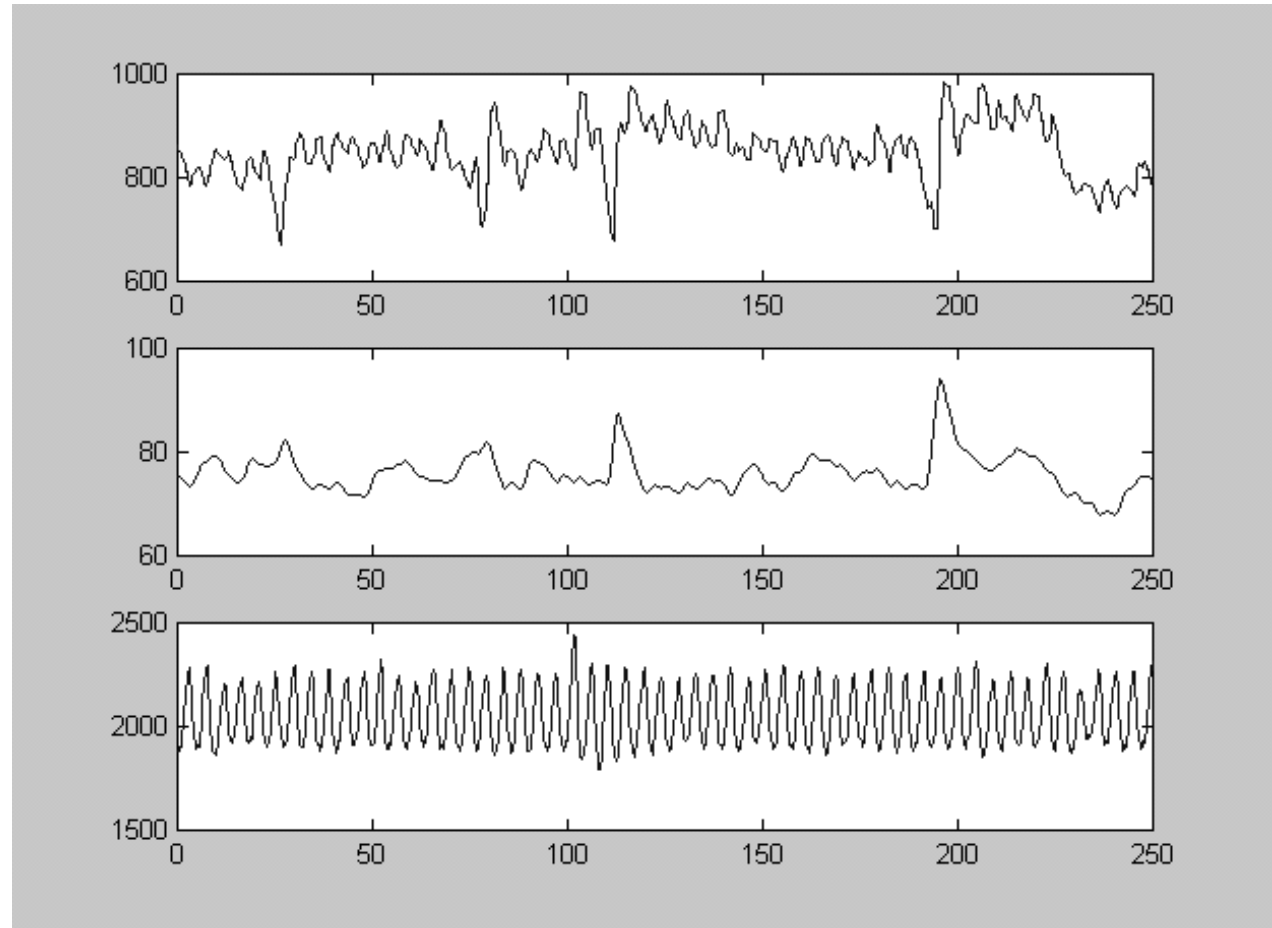


$$x(k) = \begin{cases} 1 & \text{avec probabilité } 1/2 \\ -1 & \text{avec probabilité } 1/2 \end{cases}$$

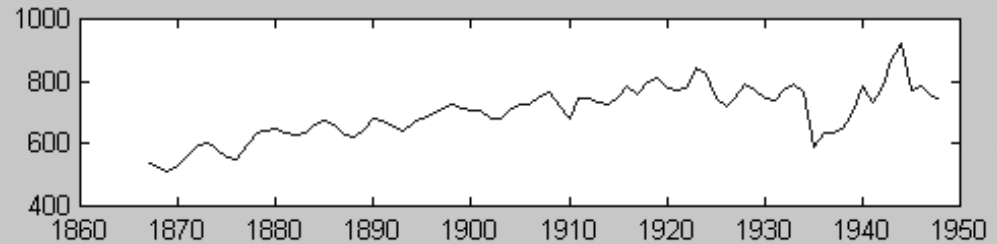
Intervalles entre
battements cardiaques (ms)

Pression artérielle
moyenne (mm Hg)

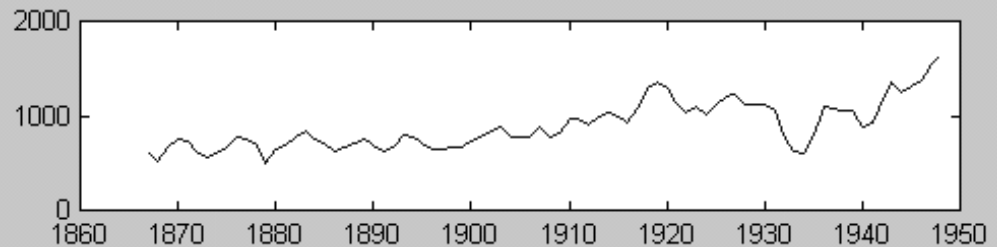
Volume respiratoire
(unités arbitraires)



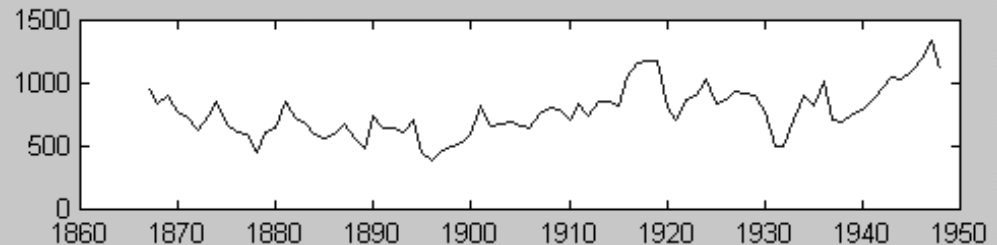
Nombre de porcs



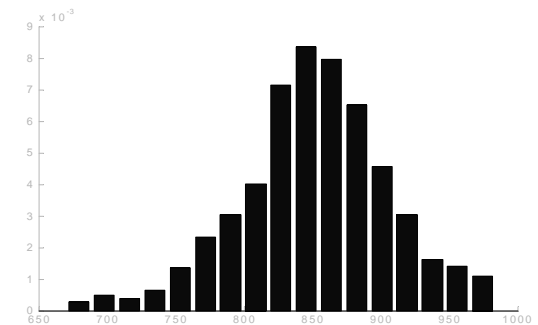
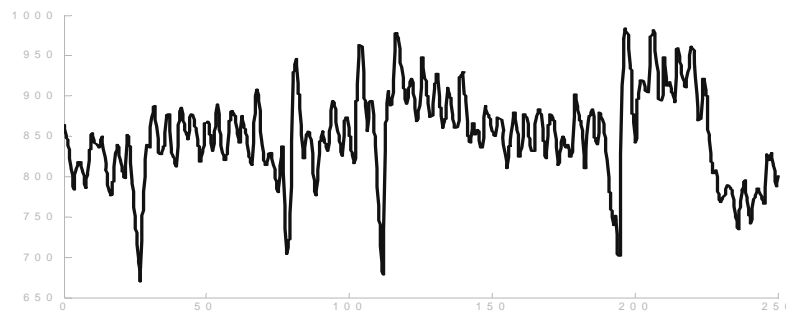
Prix du maïs



Production de maïs



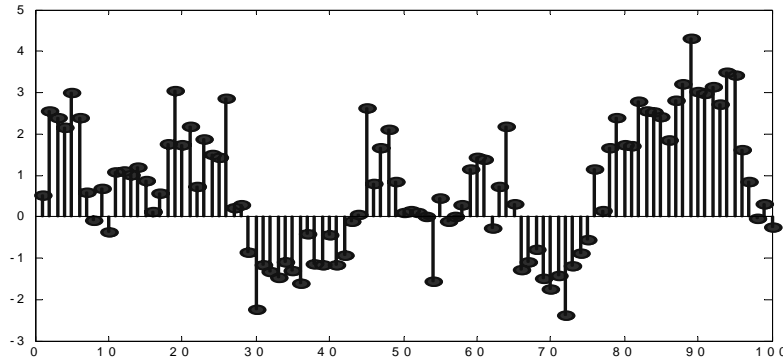
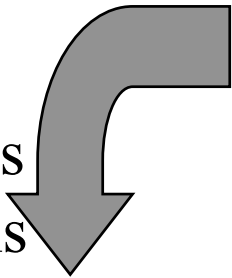
- Nous avons déjà mis ceci en pratique. Si on considère que les échantillons du signal sont des réalisations d'une même variable aléatoire, on peut estimer la densité de probabilité de celle-ci:



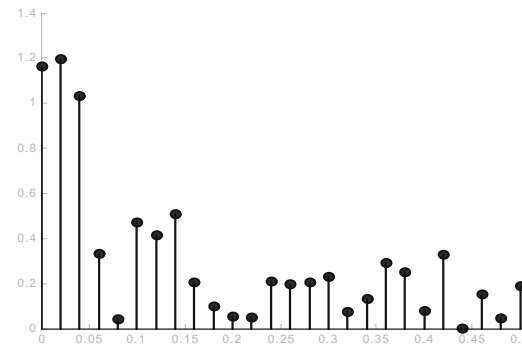
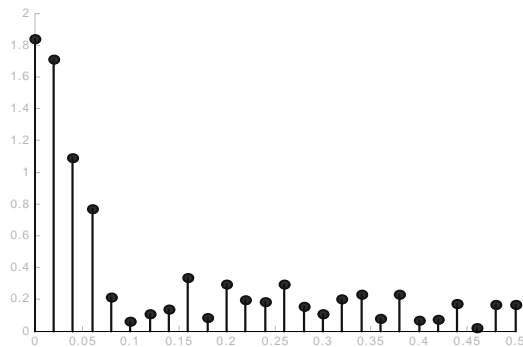
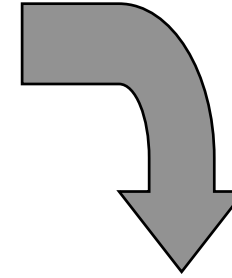
Et également la valeur moyenne et la variance. Mais ceci ne décrit pas la *relation* entre échantillons.

- Le problème avec Fourier peut s'illustrer:

TF sur
50 premiers
échantillons



TF sur
50 derniers
échantillons

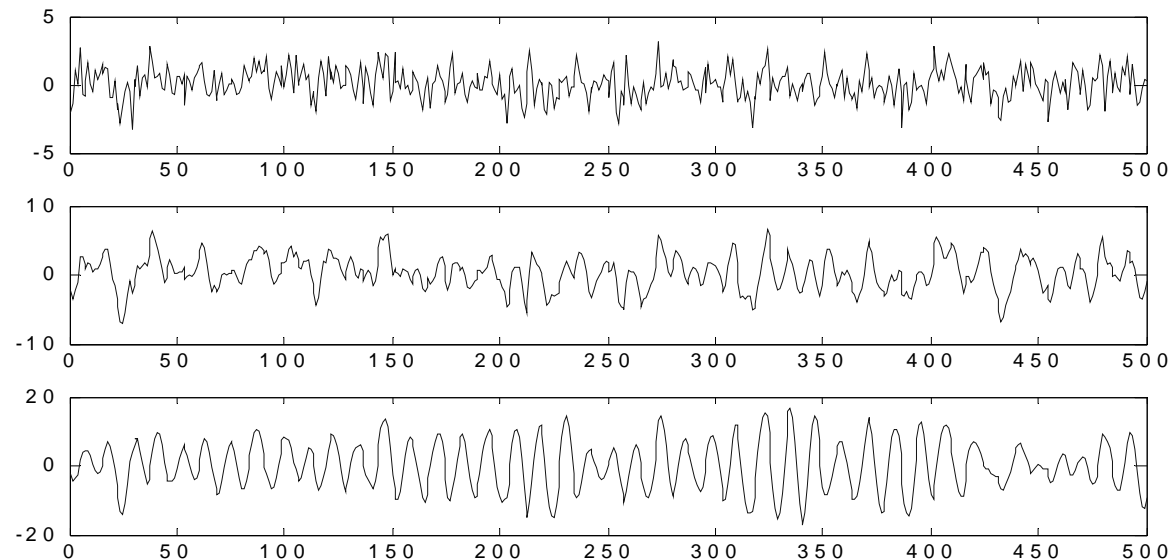



- On voit que les transformées sont semblables, mais pas identiques. Ce ne serait pas le cas avec un signal déterministe.
- L'idée de base pour contrecarrer cet effet aléatoire est la suivante:

On va extraire d'un signal stochastique un signal (ensemble de valeurs) qui décrit la structure du signal, et qui n'est pas stochastique: c'est ce qu'on appelle sa fonction d'autocorrélation.

- Qu'entend-on par structure?

plus
de
structure



- En fait, on a de plus en plus de *relation* entre les différents échantillons du signal

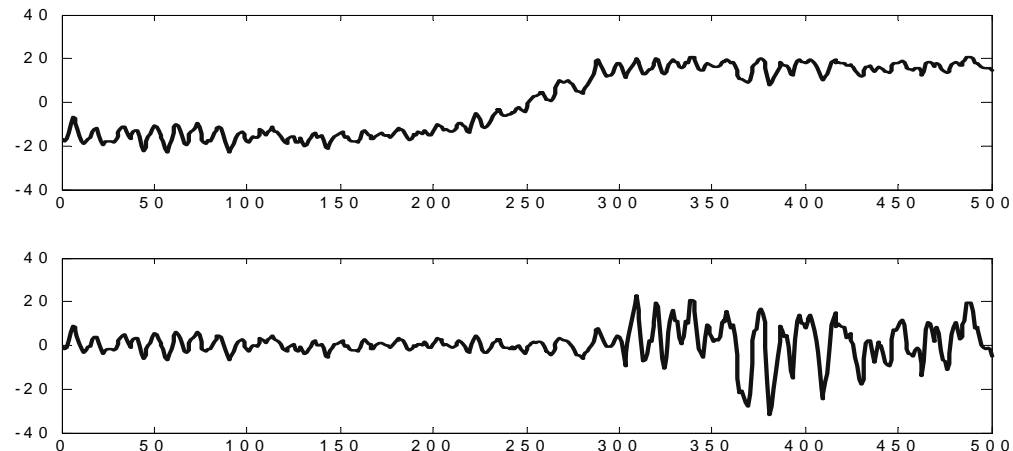
- Ca tombe bien, nous avons déjà vu un moyen de mesurer la relation entre deux variables aléatoires: la corrélation.
- On mesure donc la corrélation entre deux échantillons séparés de k indices, c'est-à-dire:

$$R_{xx}(n,k) = \mathbf{E}[x(n)x(n+k)]$$

- Si le signal est *stationnaire*, cette fonction ne dépend que de k :

$$R_{xx}(k) = \mathbf{E}[x(n)x(n+k)]$$

- En gros, un signal stationnaire est un signal dont les caractéristiques ne changent pas au cours du temps. Les signaux de la page 8 sont stationnaires, mais des signaux tels que:



ne le sont pas.

- Appelons m_x la moyenne des échantillons du signal x . On a les propriétés suivantes pour la fonction d'autocorrélation:
 - Valeur en zéro:

$$\begin{aligned} R_{xx}(0) &= E[x(k)^2] = E[(x(k) - m_x)^2] + m_x^2 \\ &= \sigma_x^2 + m_x^2 \quad \sigma_x^2 \text{ variance de } x \end{aligned}$$

De plus $|R_{xx}(k)| \leq R_{xx}(0)$. Normal, la relation entre deux échantillons est la plus grande quand ce sont les mêmes!

- Valeur limite:

$$R_{xx}(k) \rightarrow m_x^2 \text{ lorsque } k \rightarrow \infty$$

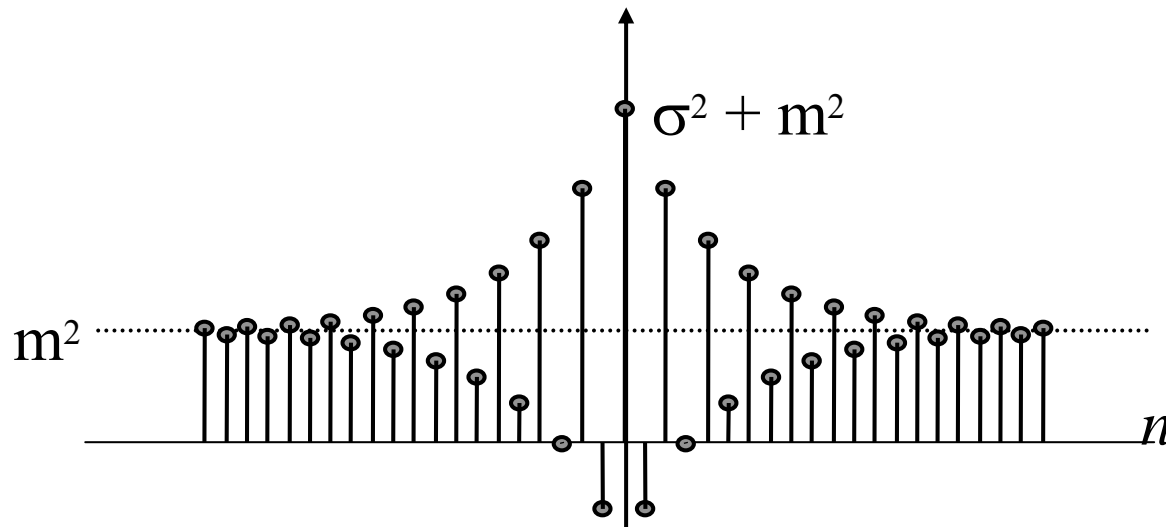
Normal. Si deux échantillons sont éloignés dans le temps, la seule relation entre eux est leur valeur moyenne. Le carré vient du fait qu'on a un produit.

- Souvent d'ailleurs on préfère travailler avec des signaux à valeur moyenne nulle (on la soustrait), ou on prend la fonction d'autocovariance:

$$C_{xx}(k) = \mathbf{E}[(x(n)-m_x)(x(n+k)-m_x)]$$

- Enfin, la fonction d'autocorrélation est une fonction paire:

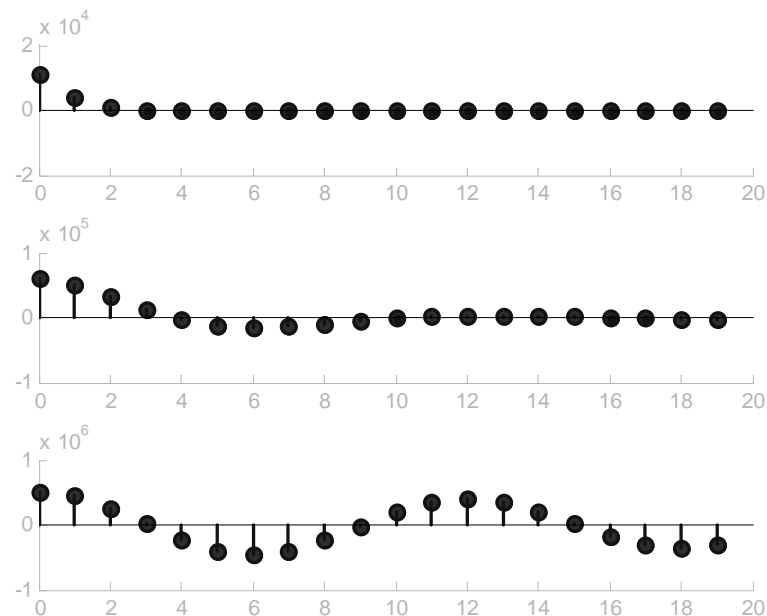
$$R_{xx}(-k) = \mathbf{E}[x(n)x(n-k)] = \mathbf{E}[x(n-k)x(n)] = R_{xx}(k)$$



On la représente donc souvent seulement pour $k \geq 0$.

- Pour un signal avec plus de structure, la fonction d'autocorrélation décroît moins vite (autocorrélation des signaux de la page 8):

plus
de
structure



- Si le signal x est périodique de période T , alors $x(n+T) = x(n)$. Mais ceci donne:

$$R_{xx}(T) = \mathbf{E}[x(n)x(n+T)] = \mathbf{E}[x(n)^2] = R_{xx}(0)$$

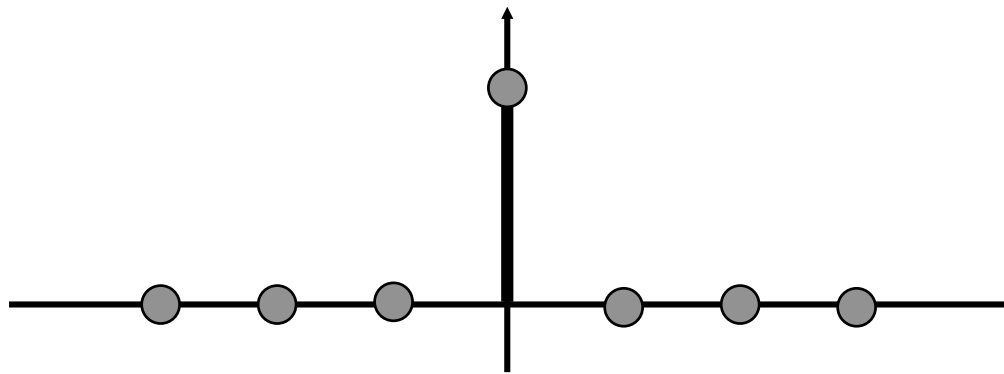
$$R_{xx}(T+1) = \mathbf{E}[x(n)x(n+T+1)] = \mathbf{E}[x(n)x(n+1)] = R_{xx}(1)$$

.....

Sa fonction d'autocorrélation est donc aussi périodique de période T . En particulier, la fonction d'autocorrélation d'un sinus ou un cosinus est un cosinus à la même fréquence.

- Un bruit blanc est caractérisé par une fonction d'autocorrélation de la forme:

$$R_{xx}(k) = \sigma_x^2 d(k)$$



- Ses échantillons sont donc tous décorrélés.

- Un problème supplémentaire est que l'on ne dispose en pratique que d'un nombre fini N d'échantillons $x(n)$, $n = 0, \dots, N-1$. On va devoir estimer la moyenne et la fonction d'autocorrélation du signal.
- L'estimateur de choix de la moyenne est bien sûr:

$$\hat{m}_x = \frac{1}{N} \sum_{k=0}^{N-1} x(k)$$

- Cet estimateur est non biaisé (pas d'erreur systématique) et consistant (s'améliore quand N croît).

- Un premier estimateur de l'autocorrélation est:

$$r_x^*(k) = \frac{1}{N - |k|} \sum_{n=0}^{N-|k|-1} x(n)x(n+k)$$

- Cet estimateur est non biaisé (pas d'erreur systématique) et consistant (s'améliore quand N croît), mais il a une variance importante pour k proche de N , car le dénominateur devient petit.

- Pourquoi la valeur absolue $|\cdot|$ dans la formule? Comme la fonction d'autocorrélation est paire, on ne l'estime que pour $k \geq 0$, mais avec la valeur absolue la formule est valable pour tout k .
- D'où vient ce $N-|k|$? C'est le nombre de produits $x(n)x(n+k)$ qu'on peut faire avec N échantillons.
- Notez que la plus grande valeur de k pour laquelle on peut estimer est $k = N-1$, et que pour cette valeur *on n'a qu'un produit* pour calculer la moyenne.

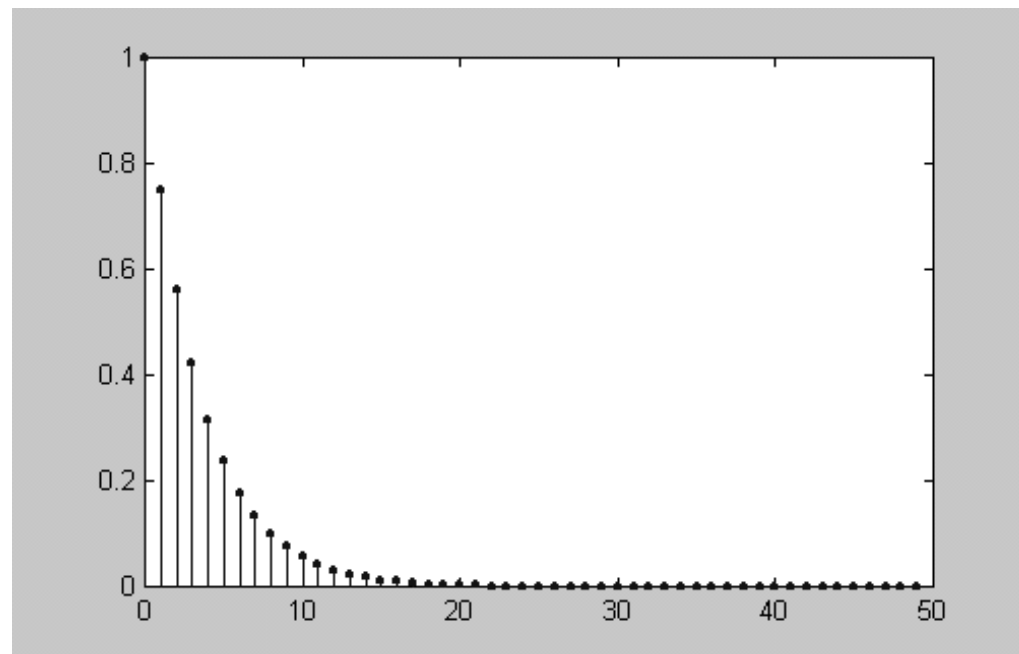
- Un deuxième estimateur de l'autocorrélation est:

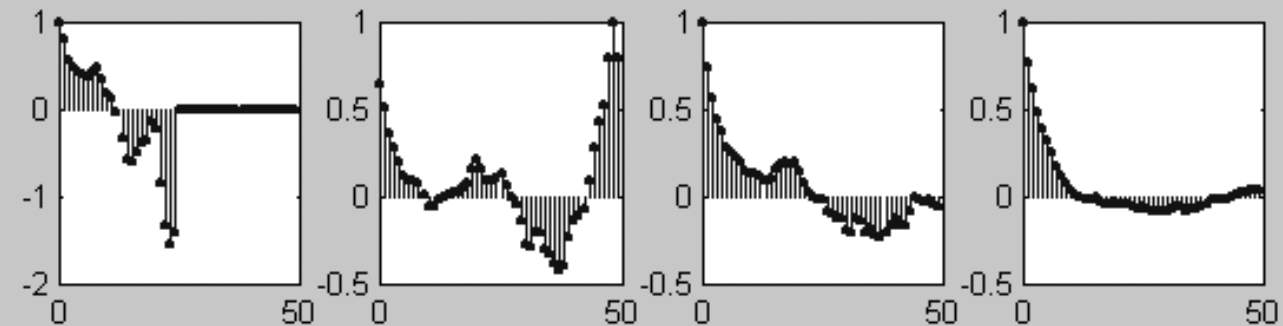
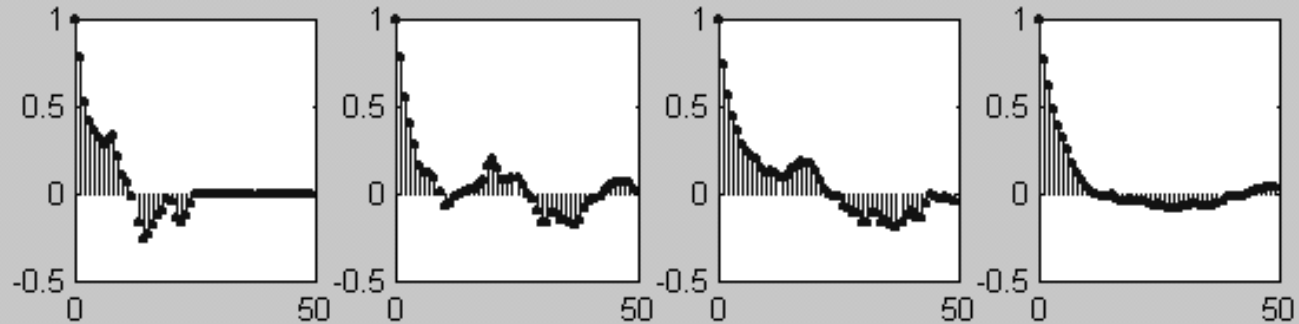
$$r_x(k) = \frac{1}{N} \sum_{n=0}^{N-|k|-1} x(n)x(n+k)$$

- Cet estimateur est biaisé (erreur systématique) mais consistant (s'améliore quand N croît), et il a une variance plus faible pour k proche de N , car on divise la somme par une plus grande valeur. En fait, c'est le plus utilisé.

- Exemple: estimation de l'autocorrélation sur un signal de fonction d'autocorrélation théorique:

$$R(k) = a^{|k|}$$
$$a = 0.75$$



$N = 25$ $N = 50$ $N = 200$ $N = 2000$ Estimateur
non biaiséEstimateur
biaisé

- On suit le même cheminement pour quantifier la relation entre les échantillons de deux signaux. S'ils sont conjointement stationnaires on obtient:

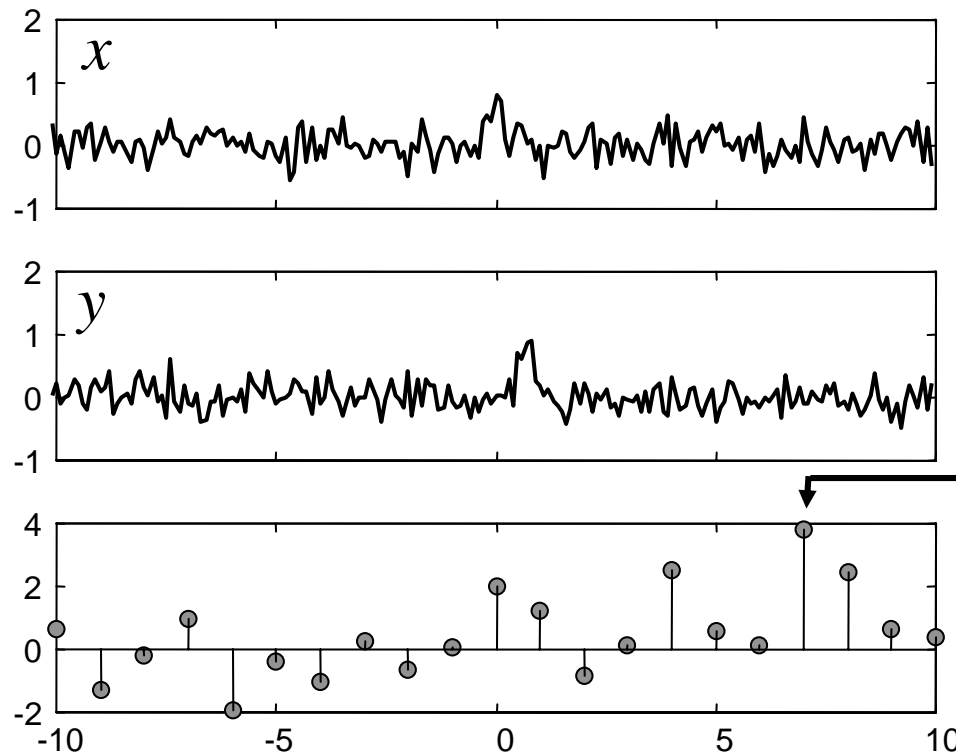
$$R_{xy}(k) = \mathbf{E}[x(n) y(n+k)]$$

- Notons que cette fonction d'intercorrélacion n'est pas paire. On a:

$$R_{xy}(-k) = \mathbf{E}[x(n) y(n-k)] = \mathbf{E}[y(n-k) x(n)] = R_{yx}(k)$$

- La fonction d'intercorrélation est par exemple utilisée pour estimer le retard entre deux signaux:

2 signaux
perturbés,
écart 7 éch.



Maximum de
 $R_{xy}(n)$

- En effet, si on obtient une grande valeur pour $R_{xy}(k)$, ceci veut dire qu'il y a une forte relation entre les échantillons $x(n)$ et $y(n+k)$. Les signaux x et y sont donc proches, mais avec un retard de k échantillons pour y .
- Il est donc logique que la fonction d'intercorrélacion ne soit pas paire: y en retard sur x n'est pas la même chose que x en retard sur y .

- On peut de même construire des estimateurs de l'intercorrélacion de deux signaux. L'estimateur biaisé est:

$$r_{xy}(n) = \begin{cases} \frac{1}{K} \sum_{k=0}^{K-n-1} x(k)y(k+n) & \text{pour } 0 \leq n < K \\ \frac{1}{K} \sum_{k=-n}^{K-1} x(k)y(k+n) & \text{pour } -K < n \leq 0 \end{cases}$$

qui tient compte de la non symétrie de l'intercorrélacion

- On peut normaliser la fonction d'autocorrélation pour qu'elle prenne des valeurs entre -1 et 1 (on rend donc ce mesure indépendant de l'échelle du signal):

$$\rho_{xx}(k) = \frac{R_{xx}(k)}{R_{xx}(0)}$$

- Pareil pour l'intercorrélation:

$$\rho_{xy}(k) = \frac{R_{xy}(k)}{\sqrt{R_{xx}(0)R_{yy}(0)}}$$