

Analyse de données et apprentissage

Analyse spectrale de signaux aléatoires

El-Hadi Djermoune

11 février 2026

Université de Lorraine

Faculté des Sciences et Technologies

M1 Ingénierie de Systèmes Complexes

CRAN UL CNRS – département BioSiS

Bureau: 422, Bât. Henri Poincaré 4^e étage

email: el-hadi.djermoune@univ-lorraine.fr

Présentation du cours

Rappels

Périodogramme simple

Variantes du périodogrammes

Corrélogramme

Que faut-il attendre de ce cours ?

C'est un cours sur l'analyse de données avec une introduction à l'apprentissage machine.

Volume horaire : 6×2h CM & 6×2h TD (MISC81.1) + 16h TP (MISC803.1)

Intervenants : El-Hadi Djermoune (CM, TD, TP), Paul Catala (TD, TP)

Tables des matières

- Analyse spectrale non paramétrique (suite de MPA...)
- Régression linéaire simple et multiple
- Estimation fréquentielle par la méthode de Prony
- Analyse en composantes principales
- Introduction à l'apprentissage automatique
- Classification (k-NN) et clustering (k-means)

Présentation du cours

Rappels

Périodogramme simple

Variantes du périodogrammes

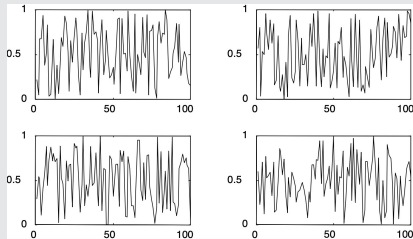
Corrélogramme

Suite de variables aléatoires

Définition

Une **suite de variables aléatoires** (suite aléatoire; *sequence of random variables*) est un ensemble de v.a. (discrètes ou continues) fonctions de $n \in \mathbb{Z}$ noté $X(n)$ ou X_n .

Exemples de trajectoires



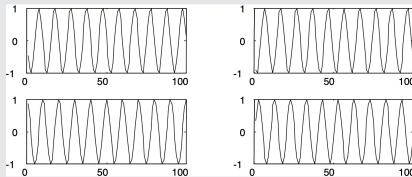
Quatre trajectoires d'une suite de v.a. i.i.d.

- Les morceaux de trajectoires comprennent $N = 100$ échantillons.
- Les v.a. de la suite sont **i.i.d.** selon une loi uniforme entre 0 et 1.

Définition

Une **suite de variables aléatoires** (suite aléatoire; *sequence of random variables*) est un ensemble de v.a. (discrètes ou continues) fonctions de $n \in \mathbb{Z}$ noté $X(n)$ ou X_n .

Exemples de trajectoires



Quatre trajectoires d'une sinusoïde à phase aléatoire

- Les morceaux de trajectoires comprennent $N = 100$ échantillons (signal à temps discret représenté en continu pour faciliter la lecture) d'un signal sinusoïdal $x(n) = A \cos(2\pi f_1 n T_e + \phi)$.
- La phase ϕ est une v.a. distribuée uniformément entre 0 et 2π .
- En supposant que $T_e = 1$ seconde, que valent A et f_1 ?

Caractérisation d'une suite de variables aléatoires

- La description d'une suite de N variables aléatoires est donnée par la densité de probabilité conjointe :

$$p_X(x_1, x_2, \dots, x_N),$$

- Un cas particulier est celui où la loi est **identique** pour chaque v.a. constituant la suite : **suite identiquement distribuée**.
- Un autre cas particulier est celui où toutes les v.a. de la suite sont **indépendantes** : **suite indépendante**.
- De façon générale, la description complète reste toutefois complexe et une caractérisation plus simple, se limitant aux moments statistiques d'ordre 2 (**moyenne statistique** et **fonction d'autocorrélation**), est souvent utilisée.

Moyenne statistique et variance

En général, moyenne statistique et variance dépendent de l'instant considéré. Ces deux fonctions du temps sont respectivement définies par les expressions suivantes :

$$m_X(n) = \mathbb{E} \{X(n)\},$$

$$\sigma_X^2(n) = \mathbb{E} \{[X^*(n) - m_X^*(n)] \cdot [X(n) - m_X(n)]\},$$

où la notation $(\cdot)^*$ signifie complexe conjugué.

Fonctions de corrélation statistiques

1. La **fonction d'autocorrélation statistique** décrit la corrélation entre deux v.a. $X(n_1)$ et $X(n_2)$. Elle dépend des instants n_1 et n_2 :

$$r_{XX}(n_1, n_2) \triangleq \mathbb{E} \{X^*(n_1)X(n_2)\}.$$

2. La **fonction d'autocovariance statistique** exprime la covariance entre deux v.a. $X(n_1)$ et $X(n_2)$:

$$c_{XX}(n_1, n_2) \triangleq \mathbb{E} \{[X^*(n_1) - m_X^*(n_1)][X(n_2) - m_X(n_2)]\}.$$

Moments temporels

- Les moments temporels sont **différents** des moments statistiques.
- Ils sont définis par rapport à **une** trajectoire **$x(n)$** de la suite aléatoire $X(n)$.

Moyenne temporelle

La moyenne temporelle (*time average*) est définie par

$$\overline{x(n)} \triangleq \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(n).$$

Autocorrélation temporelle

$$\overline{r_{xx}(n)}(k) \triangleq \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x^*(n)x(n+k).$$

- $\bar{x}$ ne dépend pas de n mais de l'expérience (trajectoire) considérée $\Rightarrow$ la moyenne temporelle est elle-même une v.a.
- $\bar{x}$ est définie pour une durée d'observation tendant vers l'infini $\Rightarrow$ seule une estimée $\hat{\bar{x}}$ est accessible à partir d'un jeu de données réelles.

Stationnarité

1. Une suite aléatoire est **stationnaire au sens strict** (*strict-sense stationary*) si **toutes** ses caractéristiques probabilistes sont invariantes pour tout changement de l'origine du temps. C'est le cas pour toute suite i.i.d.
2. Une suite aléatoire $X(n)$ est **stationnaire au sens faible** (*wide-sense stationary*) si :
 - sa moyenne est indépendante de n , i.e. $m_X(n) = m_X$;
 - sa fonction d'autocovariance ne dépend que du délai (*time lag*) $k = n_2 - n_1$:

$$c_{XX}(n, n + k) = c_{XX}(k) = r_{XX}(k) - m_X^2.$$

Il est facile de vérifier que la variance, qui correspond à la valeur de la fonction d'autocovariance $c_{XX}(0) = \sigma_X^2$, ne dépend pas non plus de n .

Ergodicité

Une suite aléatoire $X(n)$ est **ergodique à l'ordre k** si ses moments temporels d'ordre k sont indépendants de la trajectoire considérée. Une suite stationnaire et ergodique vérifie donc :

1. $m_X(n) = m_X = \overline{x(n)}$,
2. $r_{XX}(n, n + k) = r_{XX}(k) = \overline{r_{XX}(n)}(k)$.

Définition (DSP)

1. La DSP d'un signal aléatoire **stationnaire au sens large** est la transformée de Fourier de sa **fonction d'autocorrélation** :

$$S_{XX}(f) = \sum_{k=-\infty}^{\infty} r_{XX}(k) e^{-j2\pi f k} \iff r_{XX}(k) = \int_{-\frac{1}{2}}^{+\frac{1}{2}} S_{XX}(f) e^{j2\pi f k} df$$

2. Si la **suite est ergodique**, la DSP peut être obtenue par la **moy. temp.** :

$$S_{XX}(f) = \lim_{N \rightarrow \infty} \mathbb{E} \left\{ \frac{1}{2N+1} \left| \sum_{n=-N}^N x(n) e^{-j2\pi f n} \right|^2 \right\}.$$

- La DSP représente la **répartition de la puissance** par fréquence.
- La puiss. moy. dans une bande de fréquences s'obtient par intégration.
- La **puissance moyenne** d'une suite aléatoire stationnaire vérifie :

$$P_X \triangleq \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N |x(n)|^2 = r_{XX}(0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{XX}(f) df.$$

- On définit l'analyse spectrale comme toute méthode permettant de **caractériser le contenu spectral** d'un signal particulier à l'aide d'un **nombre fini d'observations**.
- Pour les signaux déterministes, cela revient à calculer la transformée de Fourier du signal mesuré.
- Pour les **signaux aléatoires**, l'analyse spectrale est basée sur l'**estimation** la **densité spectrale de puissance** à partir d'une seule trajectoire.
- Dans la suite du chapitre, les signaux aléatoires sont supposés **stationnaires et ergodiques**.

Présentation du cours

Rappels

Périodogramme simple

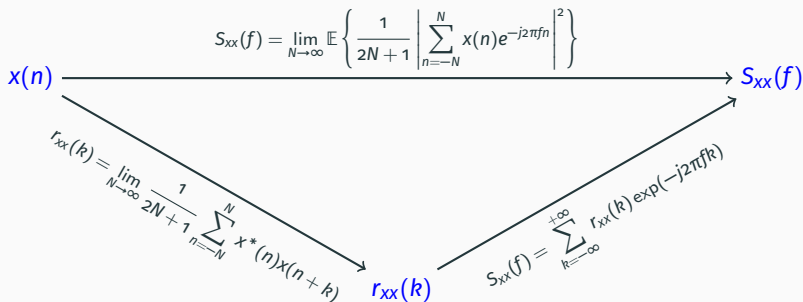
Variantes du périodogrammes

Corrélogramme

Estimateurs de la DSP : périodogramme et corrélogramme

Soit $x(n)$ une **trajectoire** de la suite $X(n)$; elle correspond aux **données mesurées**.

Il y a deux façons de calculer la DSP, soit directement à partir des échantillons $x(n)$, soit à partir de la fonction d'autocorrélation $r_{xx}(k)$:



- le chemin direct aboutit au **périodogramme**
- le chemin indirect correspond au **corrélogramme**

Périodogramme simple

L'**estimateur spectral** du **périodogramme** repose sur la définition suivante de la DSP :

$$S_{xx}(f) = \lim_{N \rightarrow \infty} \mathbb{E} \left\{ \frac{1}{2N+1} \left| \sum_{n=-N}^N x(n) e^{-j2\pi f n} \right|^2 \right\}. \quad (1)$$

Utilisation concrète

- on suppose que la suite de v.a. $X(n)$ est **stationnaire et ergodique** ;
- les données sont mesurées à partir de l'instant $n = 0$: **la somme dans (1) commence à $n = 0$** ;
- les données sont en nombre fini ($= N$) : **N est fini, on supprime la limite** ;
- $x(n)$ représente une seule trajectoire de $X(n)$: **on supprime l'opérateur d'espérance**.

Périodogramme

Le périodogramme est défini par :

$$\hat{S}_P(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-j2\pi f n} \right|^2. \quad (2)$$

Lien avec la fonction d'autocorrélation temporelle

Le périodogramme peut être réécrit de la façon suivante :

$$\begin{aligned}\hat{S}_P(f) &= \frac{1}{N} \left(\sum_{n=0}^{N-1} x(n) e^{-j2\pi f n} \right)^* \cdot \left(\sum_{m=0}^{N-1} x(m) e^{-j2\pi f m} \right) \\&= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} x^*(n) x(m) e^{-j2\pi f (n-m)} \\&= \sum_{k=-(N-1)}^{N-1} \left(\frac{1}{N} \sum_{n=0}^{N-|k|-1} x^*(n) x(n+k) \right) e^{-j2\pi f k}, \quad \text{en posant } k = n - m \\&= \sum_{k=-(N-1)}^{N-1} \hat{r}_{xx}^{(1)} e^{-j2\pi f k}.\end{aligned}$$

Estimateur biaisé de l'autocorrélation

Le périodogramme **est égal** à la transformée de Fourier de de la fonction d'autocorrélation calculée selon l'**estimateur biaisé** défini par :

$$\hat{r}_{xx}^{(1)}(k) = \frac{1}{N} \sum_{n=0}^{N-|k|-1} x^*(n) x(n+k). \quad (3)$$

Biais et variance du périodogramme

Définition

On appelle **biais d'un estimateur** la moyenne statistique de l'écart entre la valeur estimée $\hat{\theta}$ et la vraie valeur θ :

$$\text{biais}(\hat{\theta}) = \mathbb{E} \{ \hat{\theta} \} - \theta. \quad (4)$$

Biais du périodogramme

Le périodogramme est une suite aléatoire. Sa moyenne statistique s'écrit :

$$\begin{aligned} \mathbb{E} \{ \hat{S}_P(f) \} &= \sum_{k=-(N-1)}^{N-1} \mathbb{E} \{ \hat{r}_{xx}^{(1)}(k) \} e^{-j2\pi f k} = \sum_{k=-(N-1)}^{N-1} \frac{1}{N} \sum_{n=0}^{N-|k|-1} \mathbb{E} \{ x^*(n) x(n+k) \} e^{-j2\pi f k} \\ &= \sum_{k=-(N-1)}^{N-1} \frac{1}{N} \sum_{n=0}^{N-|k|-1} r_{xx}(k) e^{-j2\pi f k} \\ &= \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N} \right) r_{xx}(k) e^{-j2\pi f k}. \end{aligned}$$

Biais du périodogramme

$$\mathbb{E} \{ \hat{S}_P(f) \} - S_{xx}(f) = \sum_{k=-(N-1)}^{N-1} \frac{|k|}{N} r_{xx}(k) e^{-j2\pi f k} \neq 0 \text{ (en général)}. \quad (5)$$

- en général, le périodogramme est un estimateur biaisé de la DSP ;
- il est asymptotiquement (i.e. quand $N \rightarrow \infty$) non biaisé ;
- il est non biaisé dans le cas d'un bruit blanc (pour lequel $r_{xx}(k) = \sigma_b^2 \delta(k)$).

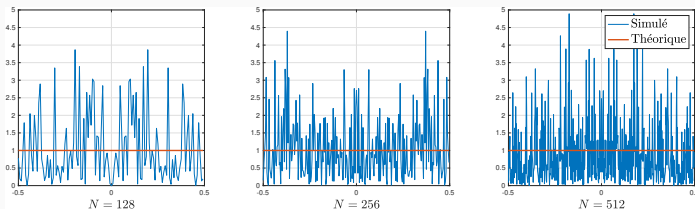
Biais et variance du périodogramme

Variance du périodogramme

La variance du périodogramme n'est pas facile à calculer. Il en existe néanmoins une approximation valable pour de grandes valeurs de N :

$$\text{var} \{ \hat{S}_P(f) \} \approx S_{xx}^2(f) \cdot \left[1 + \left(\frac{\sin 2\pi f N}{N \sin 2\pi f} \right)^2 \right] \geq S_{xx}^2(f). \quad (6)$$

- la variance du périodogramme est médiocre : elle est de même ordre de grandeur que la DSP théorique ;
- de plus, elle ne tend pas vers 0 quand N augmente.



DSP d'un bruit blanc de variance 1 - périodogramme simple

Conclusion

- Le périodogramme n'est un pas un bon estimateur de la DSP car sa variance est élevée.
- Ce problème de variance s'explique par la suppression de l'espérance mathématique dans l'équation (1).
- Plutôt que la suppression pure et simple de l'espérance, les périodogrammes moyennés consistent à la remplacer par une moyenne empirique.

Présentation du cours

Rappels

Périodogramme simple

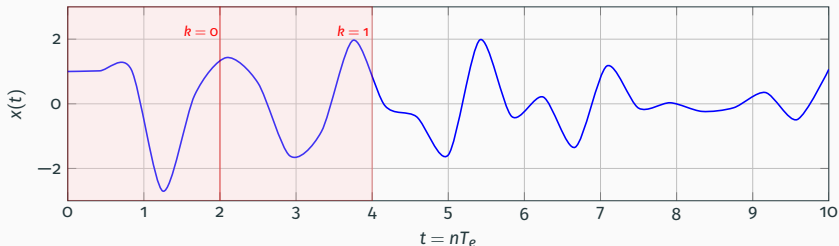
Variantes du périodogrammes

Corrélogramme

Périodogrammes moyennés

- **But :** améliorer (réduire) la variance du périodogramme simple.
- L'amélioration de la variance va s'accompagner d'une **dégradation de la résolution** spectrale $\Rightarrow$ **compromis variance/résolution**.
- **Estimateur de Bartlett (1948) :**
 - le signal de longueur N est découpé en $K_B = \frac{N}{M}$ tranches **disjointes**, chacune de taille M ;
 - on calcule le périodogramme de chacun des K_B segments de signal;
 - l'estimateur de Bartlett correspond à la moyenne des K_B périodogrammes.
- **Estimateur de Welch (1967) :**
 - le signal de longueur N est découpé en K_W tranches **non nécessairement disjointes**, chacune de taille M . Si R est le nombre d'échantillons communs à deux tranches consécutives de signal, on obtient alors $K_W = 1 + \lfloor \frac{N-M}{M-R} \rfloor > \frac{N}{M}$ segments;
 - chacun des segments est pondéré par une **fenêtre d'apodisation**; ▶ page 30
 - on calcule le périodogramme de chacune des K_W tranches ainsi obtenues;
 - l'estimateur de Welch correspond à la moyenne des K_W périodogrammes.
- Dans les 2 cas, la résolution spectrale passe de $\frac{1}{N}$ à $\frac{1}{M} \geq \frac{1}{N}$.

Estimateur de Bartlett



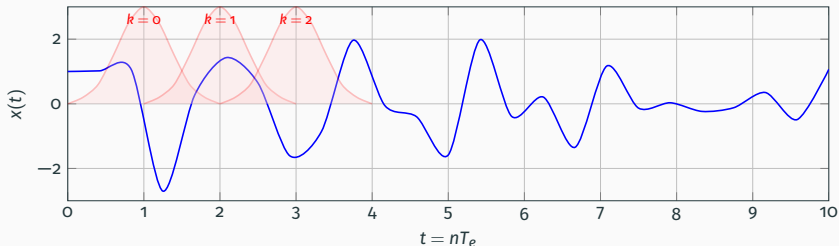
Bartlett

- Principe :

$$\hat{S}_P^{(k)}(f) = \frac{1}{M} \left| \sum_{m=0}^{M-1} x(M \cdot k + m) e^{-j2\pi f m} \right|^2, \quad k = 0, \dots, K_B - 1$$
$$\hat{S}_B(f) = \frac{1}{K_B} \sum_{k=0}^{K_B-1} \hat{S}_P^{(k)}(f) \quad (7)$$

- Résultat :** le biais n'est pas modifié mais la variance diminue d'un facteur K_B .

Estimateur de Welch



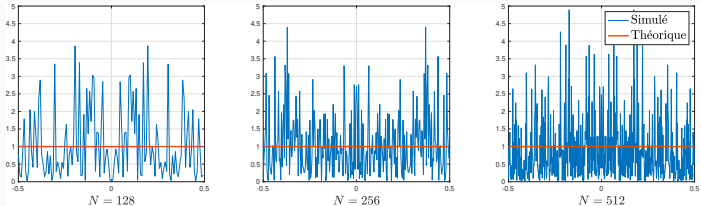
Welch

- Principe (typiquement, le recouvrement est de 50% $\Rightarrow R = \frac{M}{2}$) :

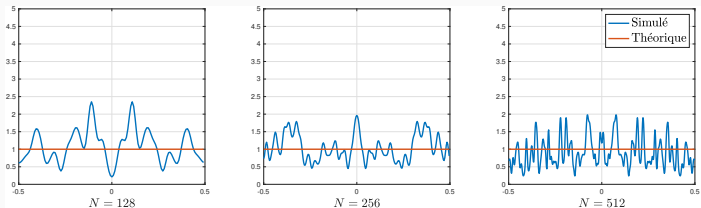
$$\hat{S}_P^{(k)}(f) = \frac{1}{\|w\|^2} \left| \sum_{m=0}^{M-1} w(m) x((M-R) \cdot k + m) e^{-j2\pi f m} \right|^2, \quad k = 0, \dots, K_W - 1$$
$$\hat{S}_W(f) = \frac{1}{K_W} \sum_{k=0}^{K_W-1} \hat{S}_P^{(k)}(f) \quad (8)$$

- Résultat :** la variance diminue d'un facteur plus grand que K_B (grâce à l'utilisation de la fenêtre $w(k)$).

Exemple : DSP d'un bruit blanc



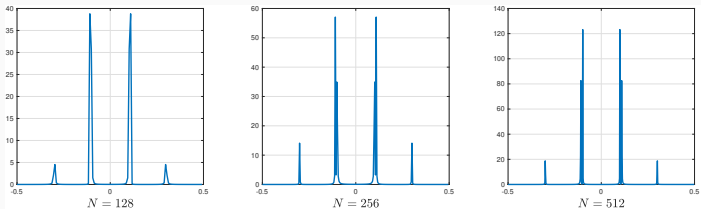
DSP d'un bruit blanc de variance 1 - périodogramme simple



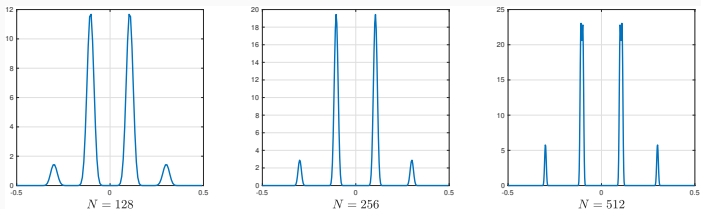
DSP d'un bruit blanc de variance 1 - périodogramme de Welch

$M = N/4$, $R = M/2$ (overlapping de 50%) $\Rightarrow K_W = 7$, fenêtre de Hamming

Exemple : DSP d'un signal sinusoïdal



DSP d'un signal sinusoïdal ($f \in \{100, 110, 300\}$ Hz, $f_e = 1000$ Hz) - périodogramme simple



DSP d'un signal sinusoïdal ($f \in \{100, 110, 300\}$ Hz, $f_e = 1000$ Hz) - périodogramme de Welch
 $M = N/4$, $R = M/2$ (overlapping de 50%) $\Rightarrow K_W = 7$, fenêtre de Hamming

Présentation du cours

Rappels

Périodogramme simple

Variantes du périodogrammes

Corrélogramme

L'estimateur spectral du **corrélogramme** repose sur la définition suivante de la DSP :

$$S_{xx}(f) = \sum_{k=-\infty}^{+\infty} r_{xx}(k) \exp(-j2\pi f k), \quad \text{avec} \quad r_{xx}(k) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x^*(n)x(n+k)$$

Corrélogramme simple

Dans un cas pratique (N fini), le corrélogramme simple s'exprime par :

$$\hat{S}_c(f) = \sum_{k=-M}^M \hat{r}_{xx}(k) \exp(-j2\pi f k), \quad \text{avec} \quad \hat{r}_{xx}(k) = \frac{1}{L_k} \sum_{n=0}^{N-|k|-1} x^*(n)x(n+k) \quad (9)$$

dans lequel $L_k = N$ ou $N - |k|$, et $\hat{r}_{xx}(k) = \hat{r}_{xx}^*(-k)$.

Estimation de l'autocorrélation

• estimateur biaisé :

$$\hat{r}_{xx}^{(1)}(k) = \frac{1}{N} \sum_{n=0}^{N-|k|-1} x^*(n)x(n+k) \quad (10)$$

• estimateur non biaisé :

$$\hat{r}_{xx}^{(2)}(k) = \frac{1}{N-|k|} \sum_{n=0}^{N-|k|-1} x^*(n)x(n+k) \quad (11)$$

• Estimateur non biaisé

- biais nul
- variance (Jenkins et Watts) :

$$\text{var} \left\{ \hat{r}_{xx}^{(2)}(k) \right\} \approx \frac{N}{(N-|k|)^2} \sum_{n=-\infty}^{\infty} r_{xx}^2(n) + r_{xx}(n+k)r_x(n-k).$$

⇒ La variance de $\hat{r}_{xx}^{(2)}(k)$ croît lorsque k s'approche de N .

⇒ Lorsque $N \rightarrow \infty$ la variance tend vers 0.

• Estimateur biaisé

- biais non nul pour N fini mais converge vers 0 quand $N \rightarrow \infty$
- variance (Jenkins et Watts) :

$$\text{var} \left\{ \hat{r}_{xx}^{(1)}(k) \right\} \approx \frac{1}{N} \sum_{n=-\infty}^{\infty} r_{xx}^2(n) + r_{xx}(n+k)r_x(n-k).$$

⇒ La variance de $\hat{r}_{xx}^{(1)}(k)$ est plus faible que celle de $\hat{r}_{xx}^{(2)}(k)$ pour k proche de N .

⇒ Lorsque $N \rightarrow \infty$ la variance tend vers 0.

Corrélogramme de Blackman-Tukey

De la même manière que le périodogramme de Welch, Blackman et Tukey proposent de pondérer les autocorrélations estimées par une fenêtre d'apodisation pour obtenir :

$$\hat{S}_{BT}(f) = \sum_{k=-M}^M w(k) \hat{r}_{xx}(k) \exp(-j2\pi f k) \quad (12)$$

dans lequel $w(k)$ est une fenêtre ayant les propriétés suivantes :

- $w(k) \in [0, 1], \forall k$;
- $w(k) = w(-k)$;
- $w(k) = 0$, pour $|k| > M$.

Exemples de fenêtres (centrées) : $N = 512$, $M = 21$

