Chapitre 6 : Les estimateurs usuels

Nous avons au chapitre précédent que le but des techniques d'estimation est d'utiliser les observations X (aléatoires) pour extraire de l'information sur la grandeur d'intérêtY. Suivant que Y soit déterministe ou aléatoire, on fera appel à différentes techniques. Ainsi :

- Pour estimer une variable certaine, dans le cas le plus général, on souhaite que l'estimateur soit non biaisé, on cherche donc l'estimateur à minimum de variance dont le plus connue est le maximum de vraisemblance (MV) basée sur l'emploi de la probabilité des valeurs observées *p(X/Y)*. Si on autorise un biais, on cherche alors à minimiser l'erreur quadratique moyenne.
- Lorsque Y la variable à estimer est aléatoire, on aura recours aux estimateurs Bayesiens. Plusieurs cas de figures sont alors possibles :

- Lorsqu'on suppose *p(Y)* connue, on utilisera en fonction de la fonction de coût à minimiser, différents estimateurs. Nous nous restreindrons, dans le cadre de ce chapitre, au maximum à postériori (MAP).

- Si seules sont connues les moments d'ordre 1 et 2 de *p(Y)* et *p(X),* on utilise alors souvent l'estimateur linéaire non-biaisé à variance minimale. Le filtre de Wiener en est un exemple.

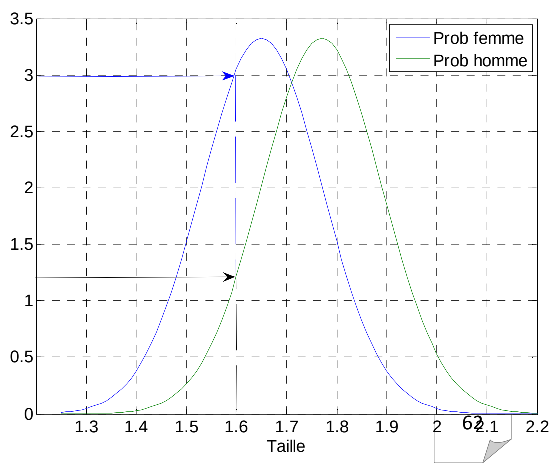
- Dans le cas où on ne possède aucune information statistique sur X et Y et que la seule dont on dispose est que X est une mesure bruitée de *fct(Y),* on adoptera l'estimateur des moindres carrés pour estimer Y qui s'affranchit de tout cadre probabiliste.

1. Estimateur du maximum de vraisemblance

On utilise cette estimateur lorsque que l'on cherche à déterminer une variable Y déterministe mais inconnue dépendant de mesures aléatoires (observations) X :$(x_1, \ldots, x_n)$.

Nous avons vu, au chapitre précédent qu'il était possible de déterminer les moments statistiques en se basant sur la loi des grands nombres. Le maximum de vraisemblance est une technique plus générale pour construire un estimateur (dont la variance est la plus petite possible) des ses moments lorsque l'on dispose de la probabilité de la loi ayant engendré les mesures aléatoires.

Exemple : Supposons que lors d'une enquête policière, un suspect inconnu (femme ou homme) mesure environ 1.60 m, on aura plutôt tendance à rechercher une femme tandis que s'il mesure environ 1.80 m, on recherchera plutôt un homme. La notion de maximum de vraisemblance permet de formaliser cette intuition. On peut modéliser la distribution des tailles (en mètres) féminines par une loi gaussienne d'espérance μ_F = 1, 65 et d'écart type σ_F = 0.12 et celle des tailles masculines par une loi gaussienne d'espérance μ_H = 1, 75 et d'´ecart type σ_H = 0.13.



Les densités de ces deux lois sont représentées sur la figure ci-contre. Lorsque l'on connait la taille *x* d'un suspect, on pourra supposer que ce suspect est une femme si la densité des tailles féminines prise en x est supérieure à celle des tailles masculines et vice et versa.

On peut estimer le genre de cet individu en choisissant X ∈ {F,H} qui maximise la vraisemblance

$$X \rightarrow \quad \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{(Y-\mu x)^2}{2\sigma x^2}}$$

Définition : On suppose que pour toute réalisation $x = (x_1, \ldots, x_n)$ de l'´echantillon X de densité $p(x/y)$, il existe une unique valeur y qui maximise la vraisemblance de la réalisation x. La vraisemblance s'écrit :

$$L(x/y) = \prod_{i=1}^{n} p(x_i/y)$$

Ainsi, le principe de la vraisemblance revient à déterminer la valeur du paramètre y et ce en fonction des observations $(x_1, \ldots, x_n)$; qui assure la plus grande probabilité d'apparition de ces observations $(x_1, \ldots, x_n)$.

Comme la fonction logarithme est strictement croissante, il revient au même de maximiser la vraisemblance Y → L(x/y) que maximiser la log-vraisemblance Y → ln(L(x/y))

*Exemple 1*On a observé un échantillon indépendant de taille n : $x = (x_1, \ldots, x_n)$ issu d'une loi de Poisson de moyenne μ. Trouvons l'estimateur du maximum de vraisemblance de μ.

Rappelons que pour une loi de Poisson $\quad p_X(x) = \frac{\mu^x e^{-\mu}}{x!}, \qquad m = \mu, \qquad \sigma^2 = \mu$

D'abord, on applique le log à $p(x_1, \ldots, x_{n,}, \mu)$, puis on dérive par rapport à μ pour trouver le maximum.

$$\mathrm{Ln}(L(x_1, \ldots, x_n | \mu)) = \mathrm{Ln}\,(p(x_1, \ldots, x_n | \mu)) = -n\mu + \ln(\mu).\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!)$$

$$\frac{d\mathrm{Ln}(L(x_1, \ldots, x_n, \mu))}{d\mu} = -n + \frac{1}{\mu}\sum_{i=1}^{n} x_i \text{ d'où } \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$ soit la moyenne arithmétique de l'échantillon.

*Exemple 2*On réalise un ensemble de k mesures regroupées dans le vecteur r. Chaque mesure estde la forme $r_i = a + n_i$, où a est un paramètre déterministe inconnu et n_i sont des échantillonsindépendants et identiquement distribués (iid) selon une loi gaussienne $N(0,\sigma^2)$

Puisque les bruits n_i sont indépendants, on peut écrire :

$$L(r|a) = \prod_{i=1}^{k} p(r_i|a) = \prod_{i=1}^{k} \frac{1}{\sigma\ \sqrt{2\pi}} e^{-\frac{(ri-a)^2}{2\sigma^2}} = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^k} \exp\left(-\frac{1}{2\sigma^2}\sum_{i}^{k}(r_i - a)^2\right)$$

Alors : $Ln\left(L(r|a)\right) = -k\ln\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2\sigma^2}\sum_{i}^{k}(r_i - a)^2$

On dérive par rapport à a, on trouve : $\frac{dLn\left(L(r|a)\right)}{da} = -\frac{1}{\sigma^2}\sum_{i}^{k}(r_i - a) = -\frac{1}{\sigma^2}\sum_{i}^{k} r_i + \frac{1}{\sigma^2}k.a$

$$\Rightarrow a = \frac{1}{k}\sum_{i}^{k} r_i$$

2. Estimateurs Bayesiens

On fait appel aux estimateurs Bayésiens pour estimer une variable aléatoire notée y aléatoire possédant une loi à priori et dépendant des observations x de dimension n : $(x_1, \ldots, x_n)$. La stratégie bayesienne consiste à estimer y de façon à minimiser l'écart entre la valeur estimé $\hat{y}$et la vraie valeur y à travers une fonction de coût $C(y; \hat{y})$ associée à chaque estimation, l'estimation optimale étant alors celleassurant la minimisation de ce coût.Les choix les plus classiques de fonctions de coût sont alors les suivants :

1. Coût quadratique : $C(y; \hat{y}) = (y-\hat{y})^2 \Rightarrow$(estimateur moyenne conditionnelle a posteriori MMSE : Minimisation de l'erreur quadratique moyenne)

2. Coût en valeur absolue : $C(y; \hat{y}) = |y- \hat{y}| \Rightarrow$(estimateur médiane à posteriori)

3. Coût uniforme : $C(y; \hat{y}) = 1 - 1_\Delta(y- \hat{y}) \Rightarrow$(estimateur maximum à posteriori MAP)

*<u>Estimateur MAP</u>*L'estimateur du maximum a posteriori est obtenu en considérant les fonctions de coûtégales à un, partout sauf sur un petit voisinage des valeurs de y où $p(y/x)$ est maximaleet où la fonction de coût vaut zéro.On obtient l'estimateur du maximum a posteriori enfaisant tendre vers 0 la taille de ce voisinage. Minimiser le coût, revient à minimiser le risque de Bayes $B(\hat{y}/x) = 1 - p(\hat{y}/x)$.

Minimiser $B(\hat{y}/|x)$ par rapport $\hat{a}$ revient donc à maximiser la densité de probabilité $p(\hat{y}/x)$. L'estimateur qui maximise $p(\hat{y}/x)$ est appelé l'estimateur du maximum à posteriori. Grâce à la formule de Bayes, la densité de probabilité conditionnelle peut se formuler :

$$p(y/x) = \frac{p(x/y)\,p(y)}{p(x)}$$

Ainsi, maximiser $p(y/x)$ revient à maximiser $p(x/y)p(x)$ par rapport à y. On ne considère pas $p(x)$ puisque cette probabilité ne dépend pas de y.

En outre, la fonction logarithme est monotone, cela revient à maximiser $\ln\big(p(x/y)\big) + \ln\big(p(y)\big)$

<u>Exemple</u> : Soit une séquence de N observations supposées indépendantes :$y_i = a + n_i$; dans laquelle a suit une loi normale $N(m; \sigma_a)$ et n_i $N(0; \sigma_n)$ pour tout i = 1,……N.

L'observation $y$étant constituée par la collections des échantillons y_i. L'hypothèse d'indépendance et de normalité de ces derniers garantit que la densité $p(y/a)$ se met sous la forme :

$$p(y/a) = \prod_{i=1}^{N} p(y_i/a) = \frac{1}{\left(\sigma_n\sqrt{2\pi}\right)^N}\exp\left(-\frac{1}{2\sigma_n^2}\sum_i^N (y_i - a)^2\right) \text{ et } p(a) = \frac{1}{\sigma_a\sqrt{2\pi}}\exp\left(-\frac{(a-m)^2}{2\sigma_a}\right)$$

On en déduit que :

$$\ln\big(p(y/a)\big) + \ln\big(p(a)\big) = -N\ln\left(\sigma_n\sqrt{2\pi}\right) - \frac{1}{2\sigma_n^2}\sum_i^k (y_i - a)^2 - \sigma_a\sqrt{2\pi} - \frac{1}{2\sigma_a^2}(m-a)^2$$

En dérivant cette expression par rapport à a, on obtient :

$$-\frac{1}{\sigma_n^{\ 2}}\sum_{i=1}^{Nk}(y_i-a)-\frac{(m-a)}{\sigma_a^{\ 2}}=-\frac{1}{\sigma_n^{\ 2}}\sum_{i=1}^{N}y_i+\frac{1}{\sigma_n^{\ 2}}N.a-\frac{(m-a)}{\sigma_a^{\ 2}}$$

Si on pose $b=N\sigma_a^{\ 2}/\sigma_n^{\ 2}$ alors $\hat{a}=\frac{1}{1+1/b}\hat{\mu}_y+\frac{1}{1+b}m$

Cet estimateur $\hat{a}$au sens du MAP réalise une moyenne pondérée entre cette moyenne empirique μ_y et la moyenne apriori m, sa valeur tendant vers l'une ou l'autre selon que le nombre des observations et/ou le rapport signal-sur-bruit est plus fort ou plus faible :

$$\lim_{b\to\infty}\hat{a}=\hat{\mu}_y \qquad \lim_{b\to 0}\hat{a}=m$$

_Remarque :_L'estimateur de Bayes est en principe le plus intéressant car il utilise l'information maximale. Mais celle-ci (en particulier la densité a priori *p(a)*n'est pas toujours disponible, et il faut alors recourir à des estimateurs sous-optimaux. Remarquons, en particulier, que s'il n'y a pas d'information a priori sur a et que tous les a sont donc équiprobables (ce qui revient à dire que *p(a)* est uniforme), alors on revient à l'estimateur du maximum de vraisemblance.

3. Estimateur linéaire à variance minimale

Lorsque les probabilités précédentes $p(y/x)$et $p(y)$ ne sont pas connues mais que l'on dispose des moments d'ordre 1 et 2 (moyennes supposées centrées des covariances R_{yx} et R_{xx}), on utilise alors souvent l'estimateur linéaire non-biaisé à variance minimale. On recherche une estimée $\hat{y}$ de y qui soit une fonction linéaire des observations, c'est-à-dire:

$$\hat{y}=\sum_{i=1}^{n}h_i x_i=h^T x$$

Dans le cas de l'estimation en moyenne quadratique, le filtre h doit alors être déterminé de telle sorte que la variance de l'erreur d'estimation E{$(y-\hat{y})^2$} soit minimale :

$$\mathrm{E}\left\{(y-\hat{y})^2\right\}=\mathrm{E}\left\{\left(y-h^T x\right)^T\left(y-h^T x\right)\right\}=E\left\{y\,y^T\right\}-E\left\{y^T h^T x\right\}-E\left\{h\,x^T y\right\}+E\left\{h\,x^T h^T x\right\}$$

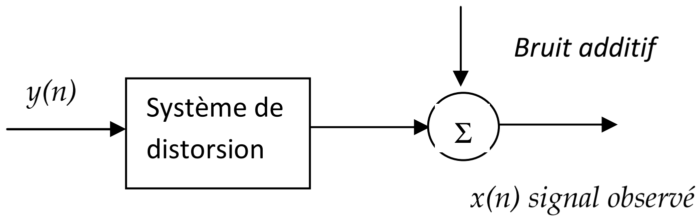
On commence par dériver par rapport à h puis on annule la dérivée, ce qui nousdonne

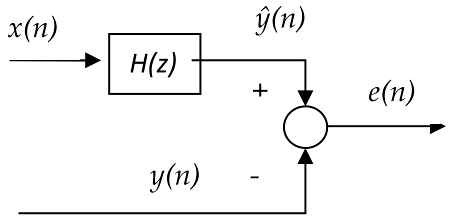
$$-E\left\{y^T x\right\}-E\left\{x^T y\right\}+2h\ E\left\{x^T x\right\}=0\Rightarrow h=E\left\{y^T x\right\}/E\left\{x^T x\right\}=R_{yx}/R_{xx}$$

Application : **_Filtre de Wiener_**

Dans de nombreuses applications, les signaux temporels sont entachées d'un bruit que l'on souhaite supprimer ou du moins réduire. Comme le signal utile aléatoire occupe les mêmes bandes de fréquences que le signal parasite, on ne peut recourir au filtrage classique. Le filtre de Wiener apporte une solution à ce problème lorsque le processus est stationnaire. Les filtres de Wiener sont dits optimum au sens du critère de l'erreur quadratique moyenne entre leur sortie et une sortie désirée. .

La figure suivante illustre un problème courant d'estimation linéaire. *y(n)* correspond au signal aléatoire stationnaire qui nous intéresse mais n'est pas directement accessible. Seul *x(n)* l'est, il est obtenu après passage de *y(n)* dans un système linéaire suivi de l'addition d'un bruit aléatoire stationnaire *bb(n)*.





Le problème qui se pose est comment retrouver *y(n)*à partir de *x(n)*. Une solution consiste à filtrer *x(n)* de telle sorte que la sortie $\hat{y}(n)$ soit la plus proche possible de*y(n)*. On peut mesurer la qualité de l'estimation par*e(n)* défini par : $e(n) = y(n) - \hat{y}(n)$

On cherche donc un filtre qui minimisera l'erreur. Il est pratique de chercher à minimiser $e^2(n)$ car c'est une fonction quadratique facilement dérivable. Par ailleurs, étant donné que les signaux sont aléatoires, la fonction coût à minimiser est l'erreur quadratique moyenne (MSE) définie par : $\xi(n) = E(e^2(n))$

Si on suppose que le filtre recherché H est un filtre RIF de longueur N, on peut en calculer les coefficients par résolution d'un système linéaire d'équations. $\mathrm{h} = [b_0 \quad b_1 \quad \dots \quad b_{N-1}]^T$

Le signal estimé $\hat{y}(n)$ peut alors s'écrire : $\hat{y}(n) = \sum_{i=0}^{N-1} b_i x(n-i)$

Rappelons que l'on cherche à minimiser $\xi(n) = E\{e^2(n)\} = E\left\{\left(y(n) - \sum_{i=0}^{N-1} b_i x(n-i)\right)^2\right\}$

Pour en obtenir le minimum, il suffit de chercher de dériver et d'annuler la fonction coût par rapport aux variables $b_i$de la réponse impulsionnelle du filtre. La dérivée de la fonction coût par rapport au $j^{\text{ème}}$ point de la réponse impulsionnelle est donnée par :

$$\frac{\partial \xi}{\partial b_j} = E\left\{\frac{\partial}{\partial b_j}\{e^2(n)\}\right\} = E\left\{2e(n)\frac{\partial e(n)}{\partial b_j}\right\} = E\left\{2e(n)\frac{\partial}{\partial b_j}\{-b_j\, x(n-j)\}\right\}$$

$$\frac{\partial \xi}{\partial b_j} = -E\{2e(n)\, x(n-j)\} = -E\left\{2\left(y(n) - \sum_{i=0}^{N-1} b_i x(n-i)\right) x(n-j)\right\}$$

En faisant l'hypothèse que les signaux *x(n)* et *y(n)* sont stationnaires, et en annulant la dérivée, on trouve : $R_{yx}(j) = \sum_{i=0}^{N-1} b_i R_{xx}(j-i)$

Ce qui pour les différentes valeurs de *j*, nous donne le système d'équations suivantes à résoudre :

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \ldots\ldots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \ldots\ldots & R_{xx}(N-2) \\ & & \ldots\ldots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \ldots\ldots & R_{xx}(0) \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix} = \begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}$$

Il faut noter que l'obtention des coefficients du filtre repose sur la connaissance de la fonction d'autocorrélation du signal d'entrée et de l'intercorrélation entre les signaux d'entrée et de sortie désirée.

Application **:** Réduction de bruit (pour la transmission vocale, les images, etc.)

Exemple 1 : On suppose que l'observation $x(n)=y(n)+bb(n)$et que le bruit additif $bb(n)$ est centré et non corrélé au signal. Simplifions les équations de Wiener-Hopf en conséquences:

$R_{yx}(k)=E\{ y(n)\ [\ y(n-k)+bb(n-k)\ ]\ \}= R_{yy}(k)$
$R_{xx}(k)=E\{(y(n)+bb(n))(y(n-k)+bb(n-k))\}= R_{yy}(k)+R_{bb}(k)$

Soit le système à résoudre suivant :

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \ldots\ldots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \ldots\ldots & R_{xx}(N-2) \\ & & \ldots\ldots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \ldots\ldots & R_{xx}(0) \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix} = \begin{bmatrix} R_{xx}(0)-R_{bb}(0) \\ R_{xx}(1)-R_{bb}(1) \\ . \\ R_{xx}(N-1)-R_{bb}(N-1) \end{bmatrix}$$

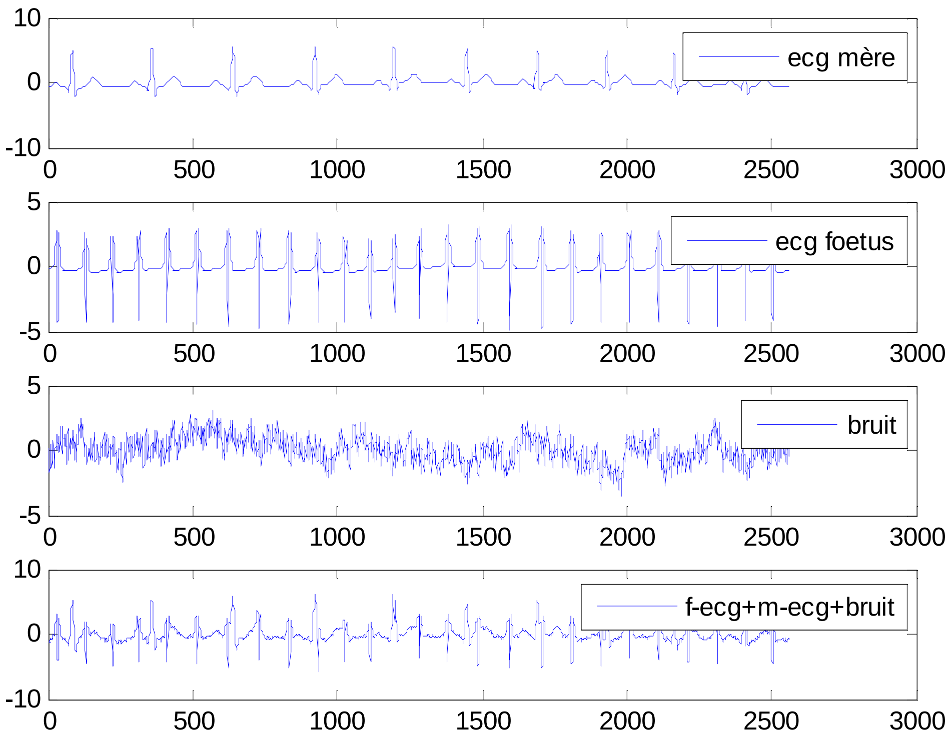
Exemple 2:

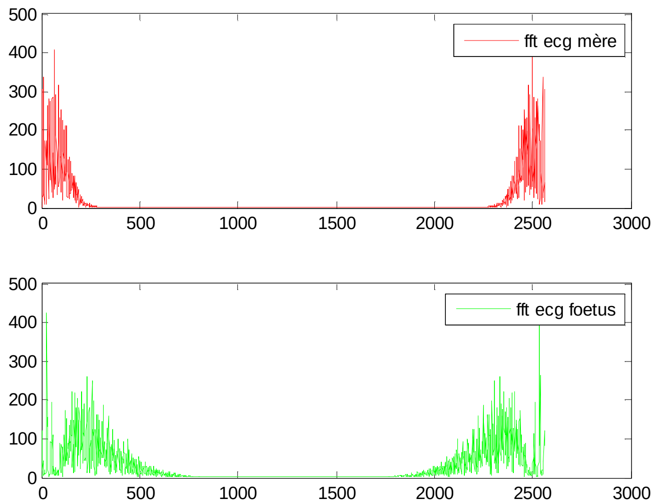
On suppose que l'observation $x(n) = y(n) + bb(n)$

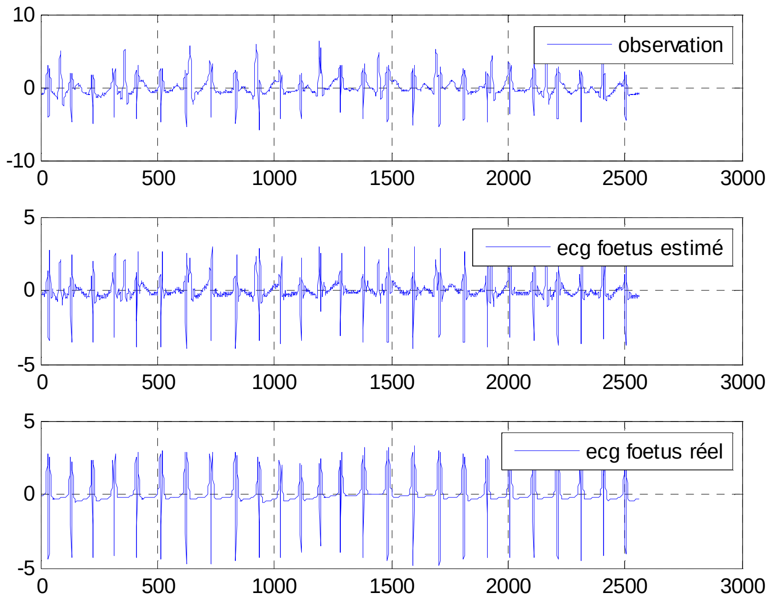
Le signal a estimer $y(n)$a pour la fonction d'autocorrélation $R_y(k)=\alpha^{|k|}$ $0<\alpha<1$. Il est décorrélé du bruit blanc $bb(n)$ de variance σ_b^2. Cherchons h(n)tel que $H(z) = b_0 + b_1 z^{-1}$.

$$\begin{bmatrix} 1+\sigma_b^2 & \alpha \\ \alpha & 1+\sigma_b^2 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha \end{bmatrix} \qquad H(z) = \frac{1}{\left(1+\sigma_b^2\right)^2 - \alpha^2}\left[(1+\sigma_b^2-\alpha^2)+\alpha\sigma_b^2 z^{-1}\right]$$

Exemple 3: Comme illustration du filtrage de Wiener, on prend souvent la mesure de l'activité cardiaque d'un fœtus à l'aide d'un électrocardiogramme (ECG) pris au niveau de l'abdomen de la mère (le signal *x(n))* et qui va naturellement être perturbé par l'ECG de celle-ci auquel se rajoute le bruit thermique des électrodes et des équipements électroniques. Pour retrouver l'ecg du foetus, on réalise une deuxième mesure fournissant l'ECG de la mère (le signal bb). On peut employer alors le filtre de Wiener pour estimer le signal *y(n)* représentant l'ECG du foetus.







On peut observer que les TF des deux signaux ecg (mère et foetus) occupent la même plage de fréquences. A partir des auto-corrélations de l'ECG mesuré sur l'abdomen *x(n)* et celui de la mère *bb(n)*, on retrouve les paramètres b_i du filtre qu'on appliquera à x(n) pour obtenir une estimée de y(n) soit l'ECG foetal.

*Remarque*Quand les fonctions d'auto et d'intercorrélation ne sont pas connues (cas le plus courant), alors on va approcher le filtre optimal de Wiener en utilisant une boucle de retour et un algorithme de minimisation: c'est ce que l'on appelle **le filtrage adaptatif**. Dans ce cas, on remplacera la connaissance des fonctions de corrélation par une phase d'apprentissage permettant de modifier itérativement la réponse impulsionnelle du filtre.

4. Estimateurs au sens des moindres carrés

Dans le filtrage de Wiener, le critère d'optimalité est stochastique : on désire minimiser la moyenne de l'erreur au carré. Cela requiert des statistique de second ordre des processus (moyennes et covariances). Uneautre approche consiste à minimiser, non plus la moyenne stochastique, mais temporelle de cette erreur, c'est la méthode des moindre carrés. C'est la moins contraignante, aucun aspect probabiliste n'entre en jeu.

$\xi = \sum_{n=L}^{N} e^2(n)$ où L et N sont des indices qui représentent l'intervalle dans lequel la minimisation se fait.

On suppose connues les données [x(0), x(1), ..., x(N-1)} et en adoptant à nouveau un filtre FIR de longueur L, on a

$e(n) = y(n) - \sum_{i=0}^{L-1} b_i\, x(n-i)$,on cherche à minimiser $\xi = \sum_{n=L}^{N} e^2(n) = \sum_{n=L}^{N}\left(y(n) - \sum_{i=0}^{L-1} b_i\, x(n-i)\right)^2$

Le signal d'entrée prend alors la forme suivante : $\begin{bmatrix} y(L) & y(L+1) & \ldots\ldots & y(N) \\ y(L-1) & y(L) & \ldots\ldots & y(N-1) \\ & & \ldots\ldots & \\ y(1) & y(2) & \ldots\ldots & y(N-L+1) \end{bmatrix}.$

On procède comme pour le filtre de Wiener, on dérive et annule les dérivée par rapport à b_i on obtient:

$$\frac{\partial \xi}{\partial b_j} = 2\sum_{n=L}^{N} \frac{\partial e(n)}{\partial b_j} e(n) = -2\sum_{n=L}^{N} x(n-i)e(n) = 0 \Rightarrow \sum_{n=L}^{N} x(n-i)\left[y(n) - \sum_{k=0}^{L-1} b_k\, x(n-k)\right] = 0, \quad i = 0, 2, \ldots\ldots, L-1$$

d'où $\sum_{k=0}^{L-1} b_k \left[\sum_{n=L}^{N} x(n-i)x(n-k)\right] = \sum_{n=L}^{N} x(n-i)y(n) \Rightarrow \sum_{k=0}^{L-1} b_k\, \overline{R_{xx}(k-i)} = \overline{R_{yx}(i)} \qquad i = 0, 2, \ldots\ldots, L-1$

donc un système d'équations $\begin{bmatrix} \overline{R_{xx}(0)} & \overline{R_{xx}(1)} & \ldots\ldots & \overline{R_{xx}(L-1)} \\ \overline{R_{xx}(1)} & \overline{R_{xx}(0)} & \ldots\ldots & \overline{R_{xx}(L-2)} \\ & & \ldots\ldots & \\ \overline{R_{xx}(L-1)} & \overline{R_{xx}(L-2)} & \ldots\ldots & \overline{R_{xx}(0)} \end{bmatrix} . \begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix} = \begin{bmatrix} \overline{R_{yx}(0)} \\ \overline{R_{yx}(1)} \\ . \\ \overline{R_{yx}(L-1)} \end{bmatrix}$

Où R_{xx} et R_{yx} sont respectivement des autocorrélations et intercorrélations temporelles .

<u>*Remarque*</u> : Le filtrage des moindres carrés se place en déterministe. Alors que le filtre de Wiener utilise la méthode des moindres carrés en stochastique.

TD n°6 : Estimateurs usuels

1. Soient ($X_1, X_2, \ldots X_N$), N variables aléatoires indépendantes de même loi de densité :

$$p(x) = \beta e^{\beta(\alpha - x)} U(x - \alpha)$$

- α étant connu, déterminer l'estimateur du maximum de vraisemblance $w=1/\beta$ notée $\hat{w}$.
- Vérifier qu'il est sans biais et convergent

2. Soient ($X_1, X_2, \ldots X_N$), N variables aléatoires indépendantes suivant une loi de probabilité définie par :

$$p(x;\theta) = \frac{1}{(r-1)!} \frac{x^{r-1}}{\theta^r} e^{-x/\theta} \qquad \text{avec} \quad \text{moy} = r\theta \qquad \text{et} \qquad \text{var} = r\theta^2$$

- Dans quel cas utilise-t-on le maximum de vraisemblance ?
- Déterminer $\hat{\theta}$ l'estimateur du maximum de vraisemblance de θ.
- Calculer le biais
- Calculer la variance de $\hat{\theta}$.
- Cet estimateur est-il consistant ?

3. La même information binaire $\theta \in \{0,1\}$ est transmise 2 fois consécutives vers un récepteur à travers un canal de transmission. Ces 2 informations sont perturbées par un bruit supposé Gaussien centré de variance σ^2. Le message reçu s'écrit alors $z=(z_1,z_2)$ où $z_i=\theta+e_i$ et $e_i \in N(0,\sigma^2)$. Le problème consiste à trouver le symbole émis à partir du message reçu z.
- Déterminer l'estimateur du maximum de vraisemblance du paramètre θ.

On suppose que l'on dispose d'une information à priori sur les bits '0' et '1' qui se traduit par p(0)=p(1)=0.5.
- Déterminer l'estimateur du maximum à posteriori de θ.
- Même question pour p(0)=q et p(1)=1-q

4. Dans une usine de fabrication de composants électroniques, on prélève au hasard des composants jusqu'à tomber sur un qui serait défaillant. Cette expérience suit une loi géométrique de paramètre p :

$$p(x/\theta) = p(1-p)^x \quad \text{avec} \qquad \mu_x = \frac{(1-p)}{p} \qquad \sigma_x^2 = \frac{(1-p)}{p^2}$$

- Déterminer $\hat{\theta}$ l'estimateur du maximum de vraisemblance de $\theta = (1-p)/p$
- Montrer que le biais est nul et calculer la variance de cet estimateur
- Cet estimateur est-il consistant ?
- Dans quelle situation utilise-t-on l'estimateur linéaire à variance minimale?

5. Soit une variable aléatoire x suivant la probabilité de Rayleigh définie par $p(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ et soient ($x_1, x_2, \ldots x_N$) un échantillon i.i.d de cette loi.

$$\text{avec} \qquad \mu_x = \sigma\sqrt{\frac{\pi}{2}} \qquad \sigma_x^2 = \frac{4-\pi}{2}\sigma^2$$

- Montrer que x est forcément supérieure ou égale à 0.
- Montrer que $\widehat{\sigma^2}$ l'estimateur du maximum de vraisemblance de σ^2 vaut $\widehat{\sigma^2} = \frac{1}{2N}\sum_{i=1}^{N} x_i^2$
- Montrer qu'il est non biaisé.
- Dans quel cas considère-t-on l'estimation par MV et MAP équivalente?

6. On considère un processus de la forme X(n) = θ + W(n) où W(n) est un processus gaussien de moyenne 0 et de variance 1 tel que W(n) et W(j) sont indépendants si n≠j. On suppose que θ suit une loi N $(0,\sigma^2)$ indépendante de W(n) n∈Z.
- Caractériser le filtre de Wiener permettant d'estimer θ à partir de $X_n, X_{n-1}, \ldots, X_{n-p+1}$.
- Calculer les coefficients du filtre.

7. On considère un processus de la forme X(n) = b(n) + αb(n-1) + W(n) où W(n) est un processus gaussien de moyenne 0 et de variance σ^2 tel que W(n) et W(j) sont indépendants si n≠j. On suppose que b(n) est une variable aléatoire uniforme à valeurs dans {−1,1}, indépendante de(W(n))n∈Z et que de même, b(n) et b(j) sont indépendants si n ≠ j (P(b(n) = 1) = P(b(n) = −1) =1/2.
- Construire un filtre qui estime b(n) .

8. On désire dé-bruiter un signal de parole z(n) corrompu par du bruit additif b(n) indépendant du signal sonore et ce, par filtrage de Wiener. On suppose connues quelques valeurs d'autocorrélation pour les deux signaux telles que :
$R_{ZZ}[0] = 1.5$; $R_{ZZ}[1] = 0.5$; $R_{ZZ}[2] = 0.25$; $R_{ZZ}[3] = 0.125$; $R_{ZZ}[4] = 0.0625$
$R_{bb}[0] = 1$; $R_{bb}[1] = 0.25$; $R_{bb}[2] = 0.0625$; $R_{bb}[3] = 0.015625$
-Pourquoi ne peut-on pas employer un filtre adapté ou un filtre moyenneur ?
-Donner les équations de Wiener-Hopf permettant d'estimer z(n).
-Déterminer le filtre de Wiener d'ordre 2 permettant de retrouver le signal utile $\hat{z}(n)$.
-Exprimer H(z).

9. On considère un problème d'estimation de bruit b(n).
Le signal observé est $x(n) = s(n) + b(n) - b(n-1)$.
On suppose que le signal s(n) est centré avec $R_{ss}(n)=0.8^{|n|}$et qu'il est décorrélé du bruit dont l'autocorrélation est $R_{bb}(n)=0.8\delta(n)$.

- Déterminer les moyennes statistiques de s*(n)* et *b(n).*
- Quand à-t-on recours au filtre de Wiener?
- Donner les équations de Wiener-Hopf permettant d'estimer *b(n)*
- Déterminer le filtre de Wiener d'ordre 2 permettant de retrouver le signal utile $\hat{b}(n)$.
- Exprimer $\hat{b}(n)$

Solutions

1. $\hat{\omega} = \frac{1}{n}\sum_{i=1}^{n} x_i - \alpha = \bar{x}_i - \alpha \quad p(x) = \frac{1}{\omega} e^{\frac{1}{\omega}(\alpha - x)}$ ⇒ loi exponentielle avec y= α-x et $\mu_y=\omega$
$\sigma_y^2= \omega^2$, b=0 $\sigma^2= \omega^2/n$

2. Variable déterministe, $\theta = \frac{1}{r\,N}\sum_{i=1}^{N} x_i$, b=0, $\sigma^2=\theta^2/Nr$, consistant.

3. $\hat{\theta} = \frac{1}{2}(z_1 + z_2)$ pour que θ=0, $z_1+z_2<1$ θ uniforme MAP⇔MV, θ=0 si $z_1 + z_2 < 1 + 2\sigma^2 Ln((1-q)/q))$

4. $\hat{\theta} = \frac{\sum_1^N Xi}{N}$ $\quad b_{\hat{\theta}} = 0 \quad \sigma^2_{b_{\hat{\theta}}} = 0$

5. b=0 la v.a. à estimer est uniforme

6. $R_{xx}(0)=1+\sigma^2$ $R_{xx}(k>0)= \sigma^2$ $R_{\theta x}(k)=\sigma^2$ $b_i=\sigma^2/(N\sigma^2+1)$

7. $R_{xx}(1:3)=(1+\alpha^2+\sigma^2, \alpha, 0)$ $R_{bx}(1:3) =(1, 0, 0)$

8. μz=0, μb=0, signal (parole) aléatoire,

9. μs=0, μb=0 , Sb(f)=0.8. Quand signal utile et bruit occupent même plage de fréquences.
$b_1=1/3$ et $b_2=-1/3$.

<u>**Exercices supplémentaires**</u>

1. Déterminer l'estimateur à vraisemblance maximale des paramètres θ des distributions :
- $p(x ;\theta) = (1+\theta)\, x^{\theta}$ 0≤x≤1 et θ>-1
- $p(x ;\theta) = \theta e^{-\theta x}$ x≥0 et θ>0
- $p(x ;\theta) = \alpha\theta x^{\alpha-1} exp(-\theta x^{\alpha})$ x≥0 et θ>0

2. La durée T séparant deux arrivées successives de requêtes à un serveur suit une loi exponentielle de paramètre θ et de densité de probabilité : $p(t/\theta) = \theta e^{-\theta t}$ avec t>0. On suppose que le paramètre θ est stochastique et suit une loi exponentielle de paramètre λ connu, de densité $p(\theta) = \lambda e^{-\lambda\theta}$.

- Montrer que la densité a posteriori de θ est proportionnelle à $\theta e^{-\theta(t+\lambda)}$
- On observe un échantillon de n durée $t_1, \dots, t_N$.

Déterminer à partir des observations ti les estimateurs du maximum a posteriori de θ.

3. Soient $(X_1, X_2, \dots X_N)$, N variables aléatoires indépendantes suivant des lois de poisson de paramètre λ.

- Déterminer l'estimateur du maximum de vraisemblance du paramètre λ.
- Est-il sans biais, efficace ?

4. La distribution de Pareto est un type particulier de loi de puissance qui a des applications en sciences physiques et sociales. elle est définie par :

$$p(x/\alpha;\theta) = \theta\alpha^{\theta} x^{-\theta-1} \qquad x > \alpha \qquad \text{avec} \qquad \mu_x = \frac{\alpha\theta}{\theta-1} \qquad \sigma_x^2 = \frac{\theta\alpha^2}{(\theta-1)^2(\theta-2)}$$

- On suppose α connue, déterminer alors $\hat{\theta}$ l'estimateur du maximum de vraisemblance de θ
- Dans quelle situation utilise-t-on l'estimateur MAP?

5. Soit une variable aléatoire x suivant la probabilité de Rayleigh définie par $p(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ et soient $(x_1, x_2, \dots x_N)$ un échantillon i.i.d de cette loi

$$\text{avec} \qquad \mu_x = \sigma\sqrt{\frac{\pi}{2}} \qquad \sigma_x^2 = \frac{4-\pi}{2}\sigma^2$$

On considère l'estimateur suivant de σ^2 qui vaut $\widehat{\sigma^2} = \frac{1}{2\pi N^2}\left(\sum_{i=1}^{N} x_i\right)^2$

- Etudier son biais.
- Etudier sa variance.
- Est-il consistant?

6. On considère un processus de la forme X(n) = S(n) − 2S(n-1) + S(n-2) + W(n) où W(n) est un processus gaussien de moyenne 0 et de variance σ^2 tel que W(n) et W(j) sont indépendants si n≠j. On suppose que S(n) suit une loi N (0,1) indépendante de (W(n)) n∈Z et que de même, S(n) et S(j) sont indépendants si n ≠ j.

- Donner l'équation de Wiener–Hopf permettant de calculer les coefficients du filtre(anti causal) de Wiener d'ordre 3 permettant d'estimer S(n-2) à partir de X(n) , X(n-1) et X(n-2) .

-Vérifier que si σ = 0, la solution est : S(n-2) = -(X(n) + 3X(n-1) + X(n-2))/5

7. On considère une observation x(n)=s(n)+w(n) où w(n) est un bruit blanc de variance σ^2 indépendant de x(n). On suppose aussi que le signal utile x(n) est centrée et réduit et que les x(n) sont indépendants.

- Dans quel cas, utilise-t-on le filtre de Wiener ?
- Déterminer le filtre de Wiener d'ordre 3 permettant de retrouver le signal utile $\hat{s}(n)$.
- Donner l'expression de $\hat{s}(n)$ en fonction de x(n).

8. On considère un problème d'estimation d'un signal θ bruité.

Le signal observé est $x(n) = \theta + b(n)$.

On suppose que θ suit une loi uniforme sur $[-\theta_0, \theta_0]$ et qu'elle est décorrélée du bruit *b(n)* qui possède une DSP qui vaut σ^2.

- Que signifie θ et b(n) sont décorrélés ? quel est le lien avec l'indépendance?
- Calculer la moyenne et la variance de *b(n)*. Est-il SSL?
- Déterminer le filtre de Wiener d'ordre N permettant de retrouver le signal utile θ .
- Exprimer H(z) puis θ.

9. On considère un problème d'estimation d'un signal s(n) bruité..
Le signal observé est $x(n) = \alpha s(n) + b(n)$.
On suppose que s(n) et b(n) sont SSL et décorréles et que le bruit possède une DSP $S_{bb}(f)=0.25$.

- Rappeler les conditions d'applications du filtrage de Wiener
- Déterminer le filtre de Wiener d'ordre 2 permettant de retrouver le signal utile $\hat{s}(n)$. On supposera que $R_{ss}(k)=2*0.5^{|k|}$
- Exprimer H(z) puis $\hat{s}(n)$

10. On considère un problème d'estimation d'un signal s(n) bruité et ayant subi un écho.
Le signal observé est $x(n) = s(n) + 0.5s(n-1) + b(n)$.
On suppose connue l'autocorrélation du signal utile et qu'elle a pour expression $0.5^{|k|}$ et l'on suppose que le signal utile est décorrélé du bruit dont l'autocorrélation est $R_{bb}(k)=0.25^{|k|}$.

1. Déterminer les moyennes statistiques de x*(n)* et *b(n)*
2. Donner les équations de Wiener-Hopf permettant d'estimer *s(n)*
3. Déterminer le filtre de Wiener d'ordre 2 permettant de retrouver le signal utile $\hat{s}(n)$.
4. Exprimer $\hat{s}(n)$ et commenter

Solutions

1. $\hat{\theta} = \dfrac{-N}{\sum_{i=1}^{n} Ln(x_i)} - 1$ $\hat{\theta} = N / \sum_{i=1}^{n} x_i$ $\hat{\theta} = N / \sum_{i=1}^{n} x_i^{\alpha}$

2. $p(t/\theta)p(\theta) = \lambda\theta e^{-\theta(t+\lambda)}$ $\hat{\theta}_{MAP} = N / \left(\lambda + \sum_{i=1}^{N} t_i\right)$

3. b=0 , $\sigma^2=\lambda/n$ (=0 $n\to\infty$) consistant

4. $p=1/(1+\theta)$ $\hat{\theta} = \dfrac{\sum_{i=1}^{n} x_i}{N}$ $b_\theta=0$, $\sigma_\theta^2=\sigma^2/n$

5. et 6. Solutions données dans énoncés respectifs

7. Estimer un signal aléatoire stationnaire (le débruiter).
$R_{xx}(0)=1+\sigma^2$, $R_{xx}(k>0)= 0$, $R_{Sx}(k)=1$, $R_{sx}(k>0)= 0$ $\hat{s}(n) = \dfrac{1}{1+\sigma^2} x(n)$

8. E{ θ . b(n)} = μ_θ . μ_b Indépendants⇒Décorrélés $\mu_b=0$ et $\sigma_b^2=\sigma^2$ $b_i=\sigma^2/(N\sigma^2+1)$ voir exo 6

9. Signaux : utile et observation stationnaires et conjointement stationnaires connus

$R_{xx}(k)=\alpha^2 R_{ss}(k)+ R_{bb}(k)$ $R_{sx}(k)=\alpha R_{ss}(k)$

10. même que 7

TP n°6 : Estimation du MV et Filtrage de Wiener

But : A travers deux exemples concrets, étudier deux techniques d'estimation : le maximum de vraisemblance et l'estimateur linéaire à variance minimale

Exercice 1 : La même information binaire $X \in \{0,1\}$ est transmise 2 fois consécutives vers un récepteur à travers un canal de transmission. Ces 2 informations sont perturbées par un bruit B supposé Gaussien centré de variance σ^2. Le message reçu s'écrit alors $R=(R_1,R_2)$ où $R_i=X+B_i$. Le problème consiste à trouver le symbole émis à partir du message reçu R. Pour considérer que le signal émis est 0, il faudrait que $p(R_1,R_2;0)<p(R_1,R_2;1)$, soit après développement (voir TD), on aboutit à la condition $R_1+R_2<1$ pour 0.

```
N=10; var=0.25; M=2*N;
B=sqrt(var)*randn(M,1);X=2*rand(N,1);X=floor(X);
%Signal émis
for i=1:N
    E(2*i-1,1)=X(i);    E(2*i,1)=X(i);
end
%Signal Reçu
R=E+B;
%Estimation par maximum de vraisemblance
for i=1:N
if (R(2*i-1,1)+R(2*i,1))>1    Z(i)=1 ;else Z(i)=0;
end
end
figure; subplot(311);stem(X);legend ('Signal original');
subplot(312);stem(R);legend ('Signal reçu');
subplot(313);stem(Z);legend ('Signal estimé par MV')
```

1. Expliquer brièvement le programme.
2. Augmenter la variance du bruit et commenter pour différentes valeurs de N.

Exercice 2: L'extraction d'ECG du fœtus se fait à partir de plusieurs électrodes placées en différents endroits du ventre de la mère ; mais les enregistrements obtenus sont des mélanges de l'ECG du fœtus noté (FECG) et celui de la mère noté (MECG) auquel se rajoute le bruit thermique des électrodes et des équipements électroniques d'où l'emploi du filtre de Wiener.

```
load mecg1.dat; load fecg1.dat; load noise1.dat; fe=256;obs=mecg1+fecg1+0.15*noise1;
figure; subplot(411);plot(mecg1);legend ('ecg mère');
subplot(412);plot(fecg1);legend ('ecg foetus');
subplot(413);plot(noise1);legend ('bruit');
subplot(414);plot(obs);legend ('observation=fecg+mecg+bruit');
figure; subplot(211); plot(abs(fft(mecg1)),'r'); legend ('fft ecg mère');
subplot(212);  plot(abs(fft(fecg1)),'g');  legend ('fft ecg foetus');
% Détermination de l'ECG du foetus avec comme donnée l'observation et l'ECG de la
mère
xm=mecg1-mean(mecg1);obs=obs-mean(obs);
%===== Détermination des coefficients du filtre
Ncoeff= 20; N=length(obs); Rxx= xcorr(obs);Rxx=Rxx(N:N+Ncoeff-1);
Rbb= xcorr(xm); Rbb=Rbb(N:N+Ncoeff-1); C = toeplitz(Rxx); D= Rxx-Rbb; b = C\D;
%=====estimation du Signal
fecg1_est=filter(b,1,obs);
%===== Affichages
figure; subplot(311); plot(obs); legend ('observation');grid
subplot(312); plot(fecg1_est);legend ('ecg foetus estimé'); grid
subplot(313); plot(fecg1);legend ('ecg foetus réel'); grid
```

1. Pourquoi a-t-on utilisé l'instruction `D= Rxx-Rbb;` ?
2. Quelles lignes du programmes faut-il changer pour retrouver l'ECG de la mère.

Exercice 3

```
clc;clear all; close all;
%création d'un signal sinusoidal
nom_fich = uigetfile('*.wav', 'Selectionner le fichier son');
N1=1000; N2=10000;
% Lire, écouter et afficher le son complet
[S,fe}=wavread(nom_fich,[N1 N2});
sound(S,fe); pause(3);figure; plot(S);
N=length(S); d=3000; alpha=0.5; sigma=0.5; BB=sigma*randn(N,1);
%Bruitage du signal + echo
for t=d+1:N
  XB(t,1)=S(t,1)+alpha*S(t-d,1);
end
sound(XB,fe);pause(3);
XB=XB+BB; sound(XB,fe);pause(3);
figure;subplot(3,1,1);plot(1:N,S,'g',1:N,XB,'b'),
legend('signal utile', 'signal bruité');
% Détermination du filtre de Wiener (RIF :bi)
M=20;    %ordre du modèle
M1=M+1;  % taille de la matrice de Toeplitz
% Résolution de la matrice Ryy.b=Ryx
Ryy=xcorr(XB); R=toeplitz(Ryy(N:N+M1-1));
Ryx=xcorr(XB,S); r=Ryx(N:N+M1-1);
b=R\r     %Détermination des paramètre bi du filtre
% Application du filtre au signal bruité
X_filt=2*filter(b,1,XB);sound(X_filt,fe);
subplot(3,1,2);plot(1:N,XB,'b',(1:N),X_filt,'r'),
legend('signal bruité', 'signal filtré','Location','SouthWest');
subplot(3,1,3);plot(1:N,S,'g',(1:N),X_filt,'r'),
legend('signal utile', 'signal filtré','Location','SouthWest');
% Visualisation de la DSP du filtre
fe=1;L=512;
[H,f}=freqz(b,1,L,fe);
figure; plot(f,abs(H.^2),'r'); legend('Filter de Wiener')
```

1. Que contient le signal XB? Quel est le rôle de la boucle ?
1. Comment s'est faite la détermination des paramètres du filtre (expliquer en détail chaque ligne)
2. Quels sont les paramètres que la détermination du filtre de Wiener nécessite ?
3. Faire varier la puissance du bruit et l'amplitude de l'écho puis l'ordre du modèle M et commenter.
4. Quel est le but de ce programme

Exercice 4

```
clear all, close all;
I = im2double(imread('peppers.png'));
imshow(I);
title('Image originale');
noise_mean = 0; noise_var = 0.01;
noisy = imnoise(I, 'gaussian',noise_mean, noise_var);
figure, imshow(noisy);title('Image bruitée')
% Application du filtre de Wiener
F = 7;
wnr(:,:,1) = wiener2(noisy(:,:,1),[F F});
wnr(:,:,2) = wiener2(noisy(:,:,2),[F F});
wnr(:,:,3) = wiener2(noisy(:,:,3),[F F});
figure, imshow(wnr);title('Image Restaurée')
```

1. Commenter le programme.
2. Faire varier F, la moyenne, la variance et la nature du bruit et commenter ?