ESTIMATION PAR MAXIMUM DE VRAISEMBLANCE MOINDRES CARRÉS

L'objectif de cette manipulation de Travaux Pratiques est d'illustrer sous MATLAB la synthèse et l'évaluation d'un estimateur par maximum de vraisemblance sur un modèle linéaire Gaussien.

De nombreux systèmes peuvent être modélisés par une relation entrée-sortie de type convolution avec bruit additif. Si cette représentation est à temps discret, et si x désigne le signal d'entrée (supposé déterministe), alors elle peut s'écrire

$$\forall n,\ Y[n] = h[n] * x[n] + B[n]$$

où

- h désigne la réponse impulsionnelle du filtre linéaire (supposé déterministe) mis en jeu ;
- $*$ désigne le produit de convolution ;
- $B[n]$ désigne un bruit Gaussien additif à l'instant n ; on suppose ce bruit blanc, c'est à dire que pour tout couple d'instants distincts (n, n'), $B[n]$ et $B[n']$ sont mutuellement indépendants ;
- $Y[n]$ désigne la sortie du système à l'instant n ; il s'agit donc d'une variable aléatoire.

I Modélisation

Disposant d'une séquence de valeurs du signal d'entrée x et d'une séquence de mesures (réalisations des variables Y) pour un ensemble d'instants consécutifs donnés, on se propose d'estimer la réponse impulsionnelle h. On suppose que le filtre est causal et d'ordre $M-1$, i.e.

$$\forall n,\ \text{si } n < 0 \text{ ou } n > M-1,\ \text{alors } h[n] = 0.$$

Ainsi, on peut écrire

$$\forall n,\ Y[n] = y_m[n] + B[n],\ \text{avec } y_m[n] = \sum_{k=0}^{M-1} h[k]x[n-k],\ \text{où } B[0], B[1], \ldots \text{ i.i.d. selon } \mathcal{G}(0, \sigma^2). \quad (1)$$

La variance du bruit est donnée : $\sigma^2 = 0.4$.

En d'autres termes, on cherche à estimer le vecteur $h = \big(h[0], h[1], \ldots, h[M-1]\big)^T$ des coefficients de la réponse impulsionnelle du filtre sur la base des entrées $x[0], x[1], \ldots, x[N-1]$ (on suppose que $x[n]$ est nul en tout instant n strictement négatif) et du vecteur des observations $y = \big(y[0], y[1], \ldots, y[N-1]\big)^T$, réalisations du vecteur aléatoire de mesure $Y = \big(Y[0], Y[1], \ldots, Y[N-1]\big)^T$ pour l'expérience en cours.

1. Qu'obtient-on en sortie du système si on place en entrée une impulsion discrète (i.e. $x[n] = \delta[n]$, où $\delta[0] = 1$ et $\delta[n] = 0\ \forall n \neq 0$) ?
   Montrer qu'on peut en déduire un estimateur de la réponse impulsionnelle.
   Préciser le biais et la variance de celui-ci.
2. Le modèle $y_m[n]$ exprimé dans (1) est-il linéaire en les paramètres inconnus ?

3. Écrire la relation $Y = Ah + B$ qui unit le vecteur aléatoire de sortie Y, le vecteur h des paramètres à estimer, et le vecteur des bruits. Préciser les dimensions des quantités mises en jeu, ainsi que le contenu de la matrice A.
4. Écrire le critère que doit maximiser l'estimé $\hat{h}$ du maximum de vraisemblance de h, puis donner l'expression de $\hat{h}$.
   Établir le biais et la matrice de covariance de l'estimateur $\hat{H}$ (cet estimateur se réalisant en $\hat{h}$ pour l'expérience en cours). Déduire la variance de l'erreur d'estimation associée à chaque paramètre $h[m]$, $m = 0, \ldots, M-1$.

II Implémentation et analyse sous MATLAB

La sortie du système étudié pour une entrée quelconque `x` est calculée au moyen de la fonction `syst` présentée en Annexe. La réponse impulsionnelle `h` de l'étude est contenue dans le fichier `datamc.mat`, de même que le signal d'entrée `x`.

II.1 Estimateur pour une entrée impulsionnelle

On place ponctuellement une impulsion discrète $\delta[n]$ en entrée du système.

5. Simuler la sortie du système.
   En déduire un estimé $\breve{h}$ du vecteur de paramètres h.
6. Tracer la réponse impulsionnelle ainsi estimée et la réponse impulsionnelle originale. Calculer la distance quadratique normalisée $\frac{1}{M}\|\breve{h} - h\|^2$. Cette estimation paraît-elle satisfaisante ?
7. Pour `Nr = 1000` réalisations du bruit, estimer la réponse impulsionnelle pour une même impulsion placée en entrée. Calculer la moyenne empirique et la variance empirique de l'estimateur $\breve{H}$ sur ces expériences.
   Comparer ces résultats aux valeurs théoriques du biais et de la variance de cet estimateur.

II.2 Estimateur du maximum de vraisemblance - Estimateur par moindres carrés

On place le signal `x` fourni en entrée du système.

8. Écrire le programme MATLAB permettant l'obtention de l'estimé $\hat{h}$ du maximum de vraisemblance du vecteur de paramètres h pour la sortie obtenue. On pourra exploiter la fonction `toeplitz` explicitée en annexe.
9. Tracer la réponse impulsionnelle ainsi estimée et la réponse impulsionnelle originale. Calculer la distance quadratique normalisée $\frac{1}{M}\|\hat{h} - h\|^2$. Cette estimation paraît-elle satisfaisante ?
10. Pour `Nr = 1000` réalisations du bruit, renouveler l'estimation pour le même signal `x` placé en entrée. Calculer la moyenne empirique et la variance empirique de l'estimateur $\hat{H}$ sur ces expériences.
    Comparer ces résultats aux valeurs théoriques du biais et de la variance de cet estimateur.

II.3 Données aberrantes

On se propose d'étudier la robustesse de l'estimateur précédent aux données aberrantes.

11. Remplacer 5 données de sortie par une valeur aberrante, grande par rapport aux observations. Estimer la réponse impulsionnelle à partir de ces nouvelles données sur la base du schéma précédent. Tracer la réponse impulsionnelle estimée et la réponse impulsionnelle originale. Calculer l'erreur d'estimation quadratique normalisée. Comparer à l'estimation sans données aberrantes.

12. Si l'estimation était parfaite, la différence entre les données et la convolution de l'entrée par la réponse impulsionnelle estimée devrait correspondre uniquement au bruit. Tracer cette erreur avec et sans données aberrantes. Peut-on utiliser un tel résultat pour « détecter » les données aberrantes ?
13. Supprimer les « outliers » ainsi détectés des données sur la base du modèle. L'estimation de h est-elle de meilleure qualité ?

III Annexe

III.1 Fonction syst

```
Fonction syst
      y = syst(x,h,sigma2);

Calcule la sortie bruitée du système de réponse impulsionnelle h pour une
entrée donnée x et un bruit blanc Gaussien

Entrées : x, signal d'entrée du système
          h, réponse impulsionnelle de durée finie du système
          sigma2, variance du bruit (par défaut sigma2 = 0.4).

Sortie : y, signal de sortie de même longueur que le signal d'entrée

Attention, l'amplitude de l'entrée doit être inférieure ou égale à 1,
sinon elle est seuillée (non linéarité du modèle).
```

III.2 Fonction Toeplitz

Une matrice de *Toeplitz* $T(\underline{c}, \underline{l})$ est entièrement définie à partir de sa première colonne $\underline{c} = [c_1, c_2, c_3 \dots c_N]^T$ et de sa première ligne $\underline{l} = [l_1, l_2, l_3 \dots l_M]$ et se construit ainsi :

$$T(\underline{c}, \underline{l}) = \begin{bmatrix} c_1 & l_2 & l_3 & & \cdots & l_M \\ c_2 & c_1 & l_2 & l_3 & \cdots & l_{M-1} \\ c_3 & c_2 & c_1 & l_2 & \cdots & l_{M-2} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ c_N & c_{N-1} & c_{N-2} & & \cdots & . \end{bmatrix}$$

C'est ce qu'effectue la fonction Matlab **toeplitz** :

```
>> help toeplitz

  TOEPLITZ Toeplitz matrix.
    TOEPLITZ(C,R) is a non-symmetric Toeplitz matrix having C as its
    first column and R as its first row.

    TOEPLITZ(C) is a symmetric (or Hermitian) Toeplitz matrix.

    See also HANKEL.
```

Remarque : Si $c_1 \neq l_1$, Matlab affichera le message suivant :

```
  Warning: Column wins diagonal conflict.
  > In /usr/local/matlab/toolbox/matlab/elmat/toeplitz.m at line 18
```