Problèmes d'estimation

Problème 1 : Identification de la réponse d'un détecteur SPOT

Un modèle physique donnant le courant I passant dans une diode soumise à une tension V est donné par la loi de Shockley :
$$I = I_s(e^{\frac{eV}{kT}} - 1)$$
où I_s est le courant inverse de saturation, e la charge élémentaire, k la constante de Boltzman et T la température absolue. Les détecteurs CCD des satellites d'observation SPOT permettent de transformer la luminance reçue par l'instrument en un courant. La relation est alors du même type que pour les diodes.
On considère le modèle :
$$y_{\mathcal{M}}(n, \boldsymbol{\theta}) = \theta_0 e^{\theta_1 x[n]} \tag{1}$$
où $x[n]$ correspond à la luminance reçue et $y_{\mathcal{M}}[n]$ à l'intensité en sortie. Alors que l'instrument était encore au sol, des mesures $y[n]$ ont été effectuées pour différentes luminances $x[n]$, pour $n = 1 \dots N$, dans le but d'identifier les paramètres $\underline{\boldsymbol{\theta}} = [\theta_0, \theta_1]^T$ pour chaque détecteur. On considère que les perturbations sur ces mesures sont additives indépendantes, gaussiennes $\mathcal{N}(0, \sigma_\epsilon^2)$:
$$y[n] = y_{\mathcal{M}}(n, \boldsymbol{\theta}) + \epsilon[n].$$

I Modèle exponentiel

I.1) Peut-on écrire la sortie du modèle sous la forme $y_{\mathcal{M}}(n, \boldsymbol{\theta}) = \underline{\boldsymbol{r}}_n^T \underline{\boldsymbol{\theta}}$ avec $\underline{\boldsymbol{r}}_n$ vecteur de régression pour l'échantillon n ? Si oui, donner $\underline{\boldsymbol{r}}_n$ et sa dimension.

I.2) Ecrire le critère quadratique non pondéré $J(\underline{\boldsymbol{\theta}})$ pour ce modèle.

I.3) Ce critère correspond-il au critère à minimiser pour calculer l'estimateur du maximum de vraisemblance ? Justifier rapidement votre réponse.

I.4) Calculer gradient de J (dérivée par rapport à chacun des paramètres).

I.5) Donner les conditions sur θ_0 et θ_1 permettant d'annuler ce gradient. Est-il immédiat de trouver $\underline{\boldsymbol{\theta}}$ satisfaisant ces conditions ? Si l'on a trouvé un tel $\underline{\boldsymbol{\theta}}$, est-on *a priori* assuré qu'il soit solution de notre problème ? Justifier rapidement vos réponses.

II Modèle logarithmique

En prenant le logarithme de l'équation (2) on obtient la relation : $\log(y_{\mathcal{M}}(n, \underline{\boldsymbol{\theta}})) = \log(\theta_0) + \theta_1 x[n]$ que l'on va considérer comme nouveau modèle :
$$\widetilde{y}_{\mathcal{M}}(n, \underline{\boldsymbol{\theta}}) = \widetilde{\theta}_0 + \widetilde{\theta}_1 x[n]$$
avec pour vecteur de paramètres $\underline{\widetilde{\boldsymbol{\theta}}} = [\widetilde{\theta}_0, \widetilde{\theta}_1]^T$.

II.1) Montrer que l'on peut écrire la sortie de ce nouveau modèle sous la forme $\widetilde{y}_{\mathcal{M}}(n, \underline{\widetilde{\boldsymbol{\theta}}}) = \underline{\boldsymbol{r}}_n^T \underline{\widetilde{\boldsymbol{\theta}}}$ avec $\underline{\boldsymbol{r}}_n$ vecteur de régression pour l'échantillon n ? Donner $\underline{\boldsymbol{r}}_n$ et sa dimension.

II.2) En construisant le vecteur $\underline{\widetilde{\boldsymbol{y}}}_{\mathcal{M}}(\underline{\widetilde{\boldsymbol{\theta}}}) = [\widetilde{y}_{\mathcal{M}}(1, \underline{\widetilde{\boldsymbol{\theta}}}), y_{\mathcal{M}}(2, \underline{\widetilde{\boldsymbol{\theta}}}) \dots \widetilde{y}_{\mathcal{M}}(N, \underline{\widetilde{\boldsymbol{\theta}}})]^T$, donner le contenu de la matrice $\mathbf{R}$ vérifiant $\underline{\widetilde{\boldsymbol{y}}}_{\mathcal{M}}(\underline{\widetilde{\boldsymbol{\theta}}}) = \mathbf{R}\underline{\widetilde{\boldsymbol{\theta}}}$ et sa dimension.

II.3) Pour ce nouveau modèle, écrire le critère quadratique non pondéré $\widetilde{J}(\underline{\widetilde{\boldsymbol{\theta}}})$ et donner $\underline{\widetilde{\boldsymbol{\theta}}}$ minimisant ce critère.

II.4) Ce critère correspond-il au critère à minimiser pour calculer l'estimateur du maximum de vraisemblance ? Justifier rapidement votre réponse.

III Synthèse

III.1) Donner les avantages et inconvénients des deux méthodes utilisées en I et II pour identifier les paramètres des détecteurs. Quelle méthode vous semble préférable ?

Problème 2 : Calibration de détecteurs

On souhaite mesurer N quantités x_n ($n = 1 \dots N$) via un détecteur. Malheureusement le gain du détecteur n'est pas connu ce que l'on peut modéliser par le fait que l'on mesure effectivement en sortie du détecteur les quantités $y_n = \theta x_n$ où θ est le gain du détecteur (scalaire). L'objectif est bien évidement de retrouver à partir des mesures y_n les quantités observées x_n. On souhaite pour cela calibrer le détecteur c'est-à-dire estimer la valeur du paramètre θ. On note par la suite $\underline{\boldsymbol{x}}$ et $\underline{\boldsymbol{y}}$ les vecteurs colonnes contenant les quantités x_n et y_n respectivement.

Les partie III, IV.2 et IV.3 peuvent être traitées séparément.

I Calibration en laboratoire

On positionne en entrée du détecteur des quantités x_n connues, on mesure les sorties y_n correspondantes et l'on souhaite estimer le paramètre θ par moindres carrés.

I.a) Exprimer la valeur du paramètre $\overset{o}{\theta}_{MC}$, estimé par moindres carrés, en fonction des x_n et des y_n (Vous pouvez, si vous le souhaitez, exprimer r_n tel que $y_n = r_n\theta$ et $\mathbf{R}$ tel que $\underline{\boldsymbol{y}} = \mathbf{R}\theta$ en précisant leur dimension).

I.b) Quelle valeur obtient-on pour $\overset{o}{\theta}_{MC}$ si l'on ne fait qu'une seule mesure ($N = 1$) ? Pourquoi est il préférable d'effectuer plusieurs mesures ?

II Auto-calibration

On veut s'affranchir de cette étape de calibration en laboratoire et estimer θ à partir de données mesurées pour des x_n inconnus. On suppose néanmoins que les x_n sont indépendants et identiquement distribués suivant une loi gaussienne centrée (moyenne $E\{x_n\} = 0$) de variance $v_x = \sigma_x^2$ (soit $E\{x_n^2\} = v_x$) :
$$\forall n, \quad f_X(x_n) = \frac{1}{\sqrt{2\pi v_x}} e^{-\frac{1}{2}\frac{x_n^2}{v_x}}$$

II.1 Moindres carrés

II.1.a) Peut-on encore estimer θ par moindres carrés ? Commenter votre réponse...

II.2 Estimateur empirique

II.2.a) Calculer la moyenne $E\{y_n\}$ et la variance $E\left\{(y_n - E\{y_n\})^2\right\}$ des mesures y_n en fonction de celles de x_n et de θ (considéré comme déterministe).

II.2.b) Dans le cas où la variance v_x des x_n est connue, déduisez de votre réponse à la question précédente un estimateur empirique $\overset{o}{\theta}_E$ du paramètre θ en fonction des y_n.

II.3 Maximum de vraisemblance

II.3.a) Écrire la fonction de vraisemblance pour les paramètres θ et v_x

$$L(\theta, v_x; \underline{\boldsymbol{y}}) = f_Y(\underline{\boldsymbol{y}}|\theta, v_x)$$

en utilisant la formule de changement de variable, ou en exploitant le fait que $\underline{\boldsymbol{y}}$ est gaussien, de moyenne et matrice de covariance à préciser. En déduire la neg-log-vraisemblance $NLL(\theta, v_x; \underline{\boldsymbol{y}})$, c'est-à-dire l'opposé du logarithme de la vraisemblance : $NLL(\theta, v_x; \underline{\boldsymbol{y}}) = -\ln(L(\theta, v_x; \underline{\boldsymbol{y}}))$.

On considère pour l'instant que la variance v_x des x_n est connue.

II.3.b) Écrire le critère $J_1(\theta)$ à minimiser pour estimer le paramètre θ par maximum de vraisemblance.

II.3.c) Calculer $\frac{\partial J_1(\theta)}{\partial \theta}$ et en déduire $\overset{o}{\theta}_{MV}$, l'estimateur de θ par maximum de vraisemblance, en fonction des y_n et de v_x. Comparer $\overset{o}{\theta}_{MV}$ à votre estimateur empirique $\overset{o}{\theta}_E$.

On considère maintenant que la variance v_x des x_n est inconnue et on va essayer de l'estimer simultanément avec θ.

II.3.d) Écrire le critère $J_2(\theta, v_x)$ à minimiser pour estimer conjointement les paramètres θ et v_x au sens du maximum de vraisemblance.

II.3.e) Calculer $\frac{\partial J_2(\theta, v_x)}{\partial \theta}$ et $\frac{\partial J_2(\theta, v_x)}{\partial v_x}$ et en déduire des conditions nécessaires que doivent vérifier θ et v_x pour être les estimateurs au sens du maximum de vraisemblance.

II.3.f) Ces conditions sont elles suffisantes pour déduire l'expression de θ et v_x à partir des y_n ?

II.3.g) Envisagez vous une autre façon de voir le problème ?

Problème 3 : Mesure d'une table avec des mètres élastiques

On nous demande de mesurer la longueur ℓ d'une table. Malheureusement, les mètres dont on dispose pour effectuer les mesures sont légèrement élastiques. Le fait d'effectuer plusieurs mesures et notre connaissance de la théorie de l'estimation devrait nous permettre d'effectuer au mieux notre besogne.
On a effectué 5 mesures (en mètre, au millimètre près) avec deux mètres différents. Les erreurs peuvent être considérées additives :

$$y[n] = \ell + \epsilon[n],$$

les types d'erreurs $\epsilon[n]$ étant différents pour les deux mètres (les deux types d'erreur peuvent être traités séparément).

I Erreur gaussienne

On peut considérer que le premier mètre qu'on nous a donné commet une erreur blanche (variables aléatoires indépendantes et identiquement distribuées), gaussienne, centrée d'écart-type 5 cm.

I.a) Exprimer la vraisemblance[1] du paramètre ℓ.

I.b) En déduire l'estimateur du maximum de vraisemblance $\overset{o}{\ell}_{\mathrm{MV1}}$. En donner son interprétation pratique.

I.c) Calculer le biais et la variance de cet estimateur. La précision obtenue augmente-t-elle si l'on augmente le nombre de données ?

I.d) Application numérique : calculer $\overset{o}{\ell}_{\mathrm{MV1}}$ ainsi que la précision obtenue sur cette valeur pour les mesures suivantes.

	y_1	y_2	y_3	y_4	y_5
mètre 1	1.242	1.243	1.126	1.160	1.197

II Erreur uniforme

On peut considérer que le second mètre qu'on nous a donné commet une erreur blanche (variables aléatoires indépendantes et identiquement distribuées), uniformément répartie sur l'intervalle [-5 cm ; 5 cm].

II.1 Estimation empirique par intervalle

De façon empirique, on va essayer de déterminer, à partir des données, un intervalle d'appartenance pour le paramètre ℓ : $\ell \in [\ell_- \; ; \; \ell_+]$.

II.1.a) A quel intervalle appartient une donnée $y[n]$? En déduire la distance maximale entre $y[n]$ et ℓ, et donc un intervalle d'appartenance $[\ell_-[n] \; ; \; \ell_+[n]]$ pour ℓ connaissant un tel $y[n]$.

II.1.b) Connaissant N données $y[n]$, $n = 1 \ldots N$, déduire de l'appartenance de l aux intervalles $[\ell_-[n] \; ; \; \ell_+[n]]$, un intervalle d'appartenance $[\ell_- \; ; \; \ell_+]$ pour ℓ (un graphique peut vous aider à construire cet intervalle). La largeur de cet intervalle décroit-elle si l'on augmente le nombre de données ? Proposer un estimateur $\overset{o}{\ell}_{\mathrm{Int2}}$ pour ℓ.

II.1.c) Application numérique : calculer cet estimateur ainsi que l'intervalle correspondant pour les mesures suivantes. Quelle est la précision obtenue ?

	y_1	y_2	y_3	y_4	y_5
mètre 2	1.179	1.206	1.222	1.190	1.224

II.2 Maximum de vraisemblance

II.2.a) Exprimer la densité de probabilité de l'erreur $\epsilon[n]$.

II.2.b) Exprimer la vraisemblance du paramètre ℓ.

1. Rappel : une variable aléatoire x gaussienne $\mathrm{N}(m, \sigma^2)$ a pour densité $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{1}{2}\frac{(x-m)^2}{\sigma^2}$ où m est sa moyenne et σ^2 sa variance.

II.2.c) L'estimation au sens du maximum de vraisemblance nous fournit-elle dans ce cas un unique estimateur ? Proposer un tel estimateur $\overset{o}{\ell}_{\mathrm{MV2}}$.

II.2.d) Application numérique : calculer $\overset{o}{\ell}_{\mathrm{MV2}}$ pour les mesures précédentes.

III Conclusion

III.a) Quel mètre vous semble le plus fiable pour effectuer vos mesures ? Pouvait-on le prévoir avant d'effectuer les mesures ? Justifier votre réponse.

Problème 4 : Mesure avec erreur systématique

On cherche à estimer la réponse impulsionnelle $h[n]$ d'un canal de transmission. On considère que cette réponse est nulle au dela de M échantillons, on a donc pour modèle une relation entrée sortie de la forme :

$$y_{\mathcal{M}}[n] = \sum_{k=0}^{M-1} h[k]x[n-k]$$

Pour estimer cette réponse, on émet un message connu $x[n], n = 0 \ldots N-1$ et l'on mesure le message reçu $y[n]$. On ne peut faire aucune hypothèse sur le passé de x et de y.
Hormis le fait que les mesures sont nécessairement bruitées – on considère ici des perturbations indépendantes identiquement distribuées gaussiennes centrées de variance σ_b^2 – on sait que notre instrument de mesure commet une erreur systématique m_b. On peut donc écrire :

$$y[n] = y_{\mathcal{M}}[n] + m_b + \epsilon[n].$$

On considère dans un premier temps que la valeur m_b de l'erreur systématique est connue. On cherche alors à estimer le vecteur $\underline{\boldsymbol{h}} = [h[0], h[1], \ldots h[M-1]]^T$ à partir des données par moindres carrés.

1. Donner l'expression de $\overset{o}{\underline{\boldsymbol{h}}}$ estimé par moindres carrés, en prenant soin de définir les matrices et vecteurs utilisés et de préciser leur dimensions.
2. Dans ce cadre, $\overset{o}{\underline{\boldsymbol{h}}}$ correspond-il à l'estimateur du maximum de vraisemblance ? Justifier votre réponse...

On considère dans un second temps que la valeur m_b de l'erreur systématique est inconnue et l'on va chercher à l'estimer à partir des données. On construit donc le vecteur $\underline{\boldsymbol{\theta}} = [\underline{\boldsymbol{h}}^T m_b]^T$

3. Donner l'expression du vecteur $\overset{o}{\underline{\boldsymbol{\theta}}}$ estimé par moindres carrés, en prenant soin de définir les matrices et vecteurs utilisés et de préciser leur dimensions.
4. Dans ce cadre, $\overset{o}{\underline{\boldsymbol{\theta}}}$ correspond-il à l'estimateur du maximum de vraisemblance ? Justifier votre réponse...

Problème 5 Identification d'un système

Soit un système caractérisé par la fonction de transfert suivante :

$$H(z) = \frac{z^{-1} - bz^{-2}}{1 - az^{-1}}.$$

On place un signal $x[N]$ en entrée de ce système et on mesure sa sortie $y[n]$.

1. Donner l'équation de récurrence liant $y[n]$ à $x[n]$.

On cherche dans un premier temps à identifier les paramètres a et b de ce système à partir de N échantillons des données non bruitées de l'entrée et de la sortie (pour les indices $n = 1 \ldots N$).

2. Écrire sous forme matricielle la relation liant les paramètres à identifier et les données disponibles. Donner les dimensions des matrices et vecteurs entrant en jeu.
3. Écrire le critère des moindres carrés et la solution au sens des moindres carrés sous forme matricielle.
4. Écrire le système à résoudre pour calculer les paramètres minimisant ce critère sous la forme : $$\begin{cases} \alpha_1 a + \beta_1 b = \gamma_1, \\ \alpha_2 a + \beta_2 b = \gamma_2; \end{cases}$$ en précisant les valeurs de α_1, β_1, γ_1, α_2, β_2 et γ_2.
5. A partir de combien d'observations peut-on espérer identifier a et b ? Commenter votre réponse en vous appuyant sur vos réponses aux questions 3 et 4.
6. D'après votre expérience en traitement du signal, est-on assuré d'identifier correctement a et b à partir de n'importe quel signal d'entrée (bruit blanc, chirp, sinusoïde, signal constant) ou est il indispensable de le choisir judicieusement ? Justifier votre réponse...

Malheureusement, en pratique, les observations sont toujours perturbées. On dispose donc uniquement des entrées $x[n]$ et des sorties bruitées $y[n] + \epsilon[n]$.

7. Dans ces conditions, peut-on encore calculer avec la relation précédente la solution des moindres carrés ? Commenter...

Problème 6 : Étude du trafic sur internet

On observe le trafic dans un noeud du réseau internet. Pour cela on mesure l'intervalle de temps t entre le passage de deux paquets. Un modèle couramment admis pour cet intervalle est une variable aléatoire T dont la densité de probabilité est exponentielle positive :

$$f(t) = \theta e^{-\theta t} \text{ si } t > 0 \text{ et } f(t) = 0 \text{ sinon.}$$

On cherche à estimer la valeur du paramètre $\theta > 0$ de cette distribution, qui est image de la densité du trafic (plus θ est grand, plus la probabilité d'avoir t grand est faible, donc plus le trafic est dense), à partir de N mesures indépendantes $\{t_1, t_2, \ldots t_N\}$.

Question préliminaire : Montrer par récurrence que $\forall \beta > 0 : I_n(\beta) = \int_0^{+\infty} u^n e^{-u\beta}\, du = \frac{n!}{\beta^{n+1}}$.

1. Calculer la moyenne et la variance de la variable aléatoire T.
2. Par une méthode des moments, proposer deux estimateurs possibles pour θ.
3. Écrire la vraisemblance $f(\underline{\boldsymbol{t}}|\theta)$ et en déduire l'expression de l'estimateur du maximum de vraisemblance. Cet estimateur est-il biaisé ?

Des études expérimentales ont permis de montrer que le paramètre λ pouvait être modélisée *a priori* par une densité de probabilité exponentielle (de paramètre λ connu) : $f(\theta) = \lambda e^{-\lambda\theta}$.

4. Écrire la loi *a posteriori* $f(\theta|\underline{\boldsymbol{t}})$ à une constante multiplicative près. Calculer cette constante en exploitant la propriété de normalisation.
5. En déduire l'expression de l'estimateur du maximum *a posteriori*.
6. En déduire l'expression de l'estimateur de la moyenne *a posteriori*.