DEVOIR II POUR LE 13 NOVEMBRE

Exercice 1 (FILTRAGE DE KALMAN)

Un avion se déplace entre Paris et Londres, en tentant de suivre une trajectoire théorique définie par le plan de vol. L'avion est surveillé au sol par des contrôleurs aériens grâce à un radar qui reçoit un écho de l'avion à intervalles réguliers. La trajectoire effective de l'avion s'écarte de la trajectoire théorique pour de multiples raisons (météorologie, imprécision du pilote automatique, turbulences,...). On cherche donc à localiser l'avion au cours de son vol à partir des observations radars successives. On note X_n l'écart (inconnu) entre la trajectoire théorique et la position de l'avion au temps n. De plus, on note Y_n la mesure donnée par le radar au temps n. Cette mesure est entachée d'erreurs à cause de l'imprécision du radar. Le problème qui se pose à l'aiguilleur est d'estimer au mieux la position de l'avion au temps n au vu des observations $Y_0, \ldots, Y_n$. Pour simplifier l'étude, on supposera que l'objet observé évolue dans un espace de dimension 1.

On modélise les suites aléatoires $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ en posant

$$\begin{cases} X_0 = W_0, & \\ X_n = aX_{n-1} + W_n & n \geq 1 \\ Y_n = X_n + V_n & n \geq 0 \end{cases}$$

où a est un nombre réel déterministe modélisant l'action du pilote, $(V_n)_{n>0}$ et $(W_n)_{n>0}$ sont des suites aléatoires indépendantes, avec $(V_n)_{n>0} \sim_{i.i.d.} \mathcal{N}(0, \tau^2)$ et $(W_n)_{n>0} \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$. Les variables aléatoires $(W_n)_{n>0}$ représentent les fluctuations instantanées de l'écart entre position théorique et position réelle. Les variables aléatoires $(V_n)_{n>0}$ modélisent les erreurs de mesure du radar. Le paramètre a modélise l'action du pilote. En théorie du signal, on dit que $(X_n)_{n>0}$ est un processus autorégressif d'ordre 1, noté AR(1), à bruit gaussien, et on a

$$X_n = \sum_{k=0}^{n} a^k W_{n-k} = \sum_{k=0}^{n} a^{n-k} W_k, n \in \mathbb{N}.$$

À présent, le problème est d'estimer X_n sachant les observations $Y_0, \ldots, Y_n$. Le candidat naturel est l'espérance conditionnelle $\mathbb{E}(X_n | Y_0, \ldots, Y_n)$.

1. Que peut-on dire du vecteur aléatoire $(X_0, \ldots, X_n, Y_0, \ldots, Y_n)$?

   Solution 1 . *C'est un vecteur gaussien car image du vecteur trivialement gaussien*

   $$\begin{pmatrix} W_0 & \ldots & W_n & \vdots & V_0 & \ldots & V_n \end{pmatrix}^T$$

   par la transformation linéaire de matrice

   $$\begin{pmatrix} \mathbf{T} & \vdots & \mathbf{0}_n \\ \cdots & & \cdots \\ \mathbf{T} & \vdots & \mathbf{Id}_n \end{pmatrix}$$

   avec $\mathbf{T}$ *matrice de Toeplitz triangulaire inférieure dont la première colonne est donnée par* $1, a, \ldots, a^n$.

2. Vérifier la formule de Bayes pour le calcul des densités conditionnelles. Si la loi P du vecteur aléatoire $(X, Y_0, \ldots, Y_{n-1})$ admet une densité jointe p, alors presque partout :

   $$p(x_n \mid y_0, \ldots, y_{n-1}, y_n) = \frac{p(x_n, y_n \mid y_0, \ldots, y_{n-1})}{p(y_n \mid y_0, \ldots, y_{n-1})}$$

   Solution 2 . *On suppose dans la suite que les versions des densités conditionnelles considérées sont calculées à partir d'une meme version de la densité jointe. Idem pour les densités marginales.*

On considère un tuple où la densité jointe

$$\begin{aligned} p(x_n \mid y_0, \dots, y_n) &= \frac{p(x_n, y_0, \dots, y_n)}{p(y_0, \dots, y_n)} \\ &= \frac{p(x_n, y_0, \dots, y_n)}{p(y_n \mid y_0, \dots, y_{n-1}) p(y_0, \dots, y_{n-1})} \\ &= \frac{p(x_n, y_0, \dots, y_n)}{p(y_0, \dots, y_{n-1})} \times \frac{1}{p(y_n \mid y_0, \dots, y_{n-1})} \\ &= \frac{p(x_n, y_n \mid y_0, \dots, y_{n-1})}{p(y_n \mid y_0, \dots, y_{n-1})}. \end{aligned}$$

3. Vérifier que si $(X, Y_0, \dots, Y_{n-1})$ est un vecteur gaussien dans $\mathbb{R}^{n+1}$ tel que

$$\mathrm{Loi}(X|Y_0, \dots, Y_{n-1}) = \mathcal{N}(\mu, \gamma^2) \qquad \mathrm{Loi}(Y|Y_0, \dots, Y_{n-1}, X) = \mathcal{N}(X, \delta^2),$$

alors

$$\mathrm{Loi}(X|Y_0, \dots, Y_{n-1}, Y) = \mathcal{N}\left(\rho^2\left(\frac{\mu}{\gamma^2} + \frac{Y}{\delta^2}\right), \rho^2\right), \qquad \frac{1}{\rho^2} := \frac{1}{\gamma^2} + \frac{1}{\delta^2}.$$

Solution 3 . *Conditionnellement à $Y_0, \dots, Y_{n-1}$, la loi de $(X, Y)^T$ est*

$$\mathcal{N}\left(\begin{pmatrix}\mu \\ \mu\end{pmatrix}, \begin{pmatrix}\gamma^2 & \gamma^2 \\ \gamma^2 & \gamma^2 + \delta^2\end{pmatrix}\right).$$

L'énoncé nous indique que X est indépendant de $Y_0, \dots, Y_{n-1}$ et que conditionnellement à X, Y est indépendant de $Y_0, \dots, Y_{n-1}$.
La loi conditionnelle de X sachant Y est gaussienne d'espérance

$$\mu + \frac{\gamma^2}{\gamma^2 + \delta^2}(Y - \mu) = \rho^2\left(\frac{\mu}{\gamma^2} + \frac{Y}{\delta^2}\right)$$

et de variance

$$\gamma^2 - \gamma^2 \frac{1}{\gamma^2 + \delta^2}\gamma^2 = \frac{\gamma^2\delta^2}{\gamma^2 + \delta^2} = \rho^2$$

voir section 2.5 dans les notes du cours.
On retrouve les expressions classiques en inférence bayésienne.

4. On souhaite montrer que la loi conditionnelle de X_n sachant $Y_0, \dots, Y_n$ est une loi normale $\mathcal{N}(\widetilde{X}_n, P_n)$ où

$$\widetilde{X}_n = a\widetilde{X}_{n-1} + \frac{P_n}{\tau^2}(Y_n - a\widetilde{X}_{n-1}), \qquad P_n = \frac{a^2\tau^2 P_{n-1} + \sigma^2\tau^2}{a^2 P_{n-1} + \sigma^2 + \tau^2}.$$

On propose de le faire par récurrence en deux étapes.

a) Initialisation : caractériser la loi de X_0 conditionnellement à Y_0.
Solution 4 . *Nous sommes dans le contexte de la question précédente avec*

$$\gamma^2 = \sigma^2 \quad \textit{et} \quad \delta^2 = \tau^2.$$

La loi jointe de $(X_0, Y_0)^T$ est

$$\mathcal{N}\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 + \tau^2\end{pmatrix}\right).$$

La loi de X_0 sachant Y_0 est donc

$$\mathcal{N}\left(\frac{\sigma^2}{\tau^2 + \sigma^2}Y_0, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$$

On peut convenir de $\widetilde{X}_{-1} = X_{-1} = 0$, $P_{-1} = 0$ et on a alors

$$\begin{aligned} P_0 &= \frac{a^2 P_{-1}\sigma^2 + \sigma^2\tau^2}{a^2 P_{-1} + \sigma^2 + \tau^2} \\ \widetilde{X}_0 &= a \times \widetilde{X}_{-1} + \frac{P_0}{\tau^2}\left(Y_0 - a\widetilde{X}_{-1}\right) \end{aligned}$$

b) Prédiction : à partir de la loi de X_{n-1} sachant $Y_0, \ldots, Y_{n-1}$, en déduire la loi de X_n sachant $Y_0, \ldots, Y_{n-1}$.

Solution 5 . *(Prédiction) On procède par récurrence sur n. On veut établir que la loi conditionnelle de X_{n+1} sachant $Y_0, \ldots, Y_n$ est gaussienne d'espérance*

$$a\mathbb{E}[X_n \mid Y_0, \ldots, Y_n]$$

et de variance

$$\sigma^2 + a^2 \operatorname{var}(X_n \mid Y_0, \ldots, Y_n) .$$

Il suffit de noter

$$\begin{aligned}
\mathbb{E}[X_{n+1} \mid Y_0, \ldots, Y_n] &= \mathbb{E}[aX_n + W_n \mid Y_0, \ldots, Y_n] \\
&= a\mathbb{E}[X_n \mid Y_0, \ldots, Y_n] + \mathbb{E}[W_n \mid Y_0, \ldots, Y_n] \\
&= a\mathbb{E}[X_n \mid Y_0, \ldots, Y_n] + \mathbb{E}W_n \\
&= a\mathbb{E}[X_n \mid Y_0, \ldots, Y_n] .
\end{aligned}$$

$$\begin{aligned}
\operatorname{var}(X_{n+1} \mid Y_0, \ldots, Y_n) &= \operatorname{var}(aX_n + W_n \mid Y_0, \ldots, Y_n) \\
&= \operatorname{var}(aX_n \mid Y_0, \ldots, Y_n) + \operatorname{var}(W_n \mid Y_0, \ldots, Y_n) \\
&= a^2 \operatorname{var}(X_n \mid Y_0, \ldots, Y_n) + \operatorname{var}(W_n) \\
&= a^2 \operatorname{var}(X_n \mid Y_0, \ldots, Y_n) + \sigma^2 .
\end{aligned}$$

c) Filtrage : déterminer la loi de X_n sachant $Y_0, \ldots, Y_n$.

Solution 6 . *On procède par récurrence sur n. Le cas de base $n = 0$ est traité en réponse au a). On suppose que l'hypothèse de récurrence est vérifiée jusqu'au rang n. On invoque la réponse à la question 3 avec*

$$\begin{aligned}
\mu &= \mathbb{E}[X_{n+1} \mid Y_0, \ldots, Y_n] \\
&= a\mathbb{E}[aX_n \mid Y_0, \ldots, Y_n] \\
&= a\widetilde{X}_n
\end{aligned}$$

et $\gamma^2 = a^2 \operatorname{var}(\widetilde{X}_n \mid Y_0, \ldots Y_n) + \sigma^2 = a^2 P_n + \sigma^2$. Et la variance conditionnelle de Y_{n+1} sachant $X_{n+1}, Y_0, \ldots, Y_n$ est τ^2. On peut donc convenir de $\delta^2 = \tau^2$.

Conditionnellement à $Y_0, \ldots, Y_n, Y_{n+1}$, la variable X_n est distribuée selon une gaussienne de variance $P_{n+1} = \rho^2$ avec

$$\begin{aligned}
\frac{1}{\rho^2} &= \frac{1}{\tau^2} + \frac{1}{a^2 P_n + \sigma^2} \\
&= \frac{a^2 P_n + \sigma^2 + \tau^2}{\tau^2 (a^2 P_n + \sigma^2)}
\end{aligned}$$

et d'espérance

$$\begin{aligned}
\widetilde{X}_{n+1} &= P_{n+1} \left(\frac{a\widetilde{X}_n}{a^2 P_n + \sigma^2} + \frac{Y_{n+1}}{\tau^2} \right) \\
&= P_{n+1} \left(\left(\frac{1}{P_{n+1}} - \frac{1}{\tau^2} \right) a\widetilde{X}_n + \frac{Y_{n+1}}{\tau^2} \right) \\
&= a\widetilde{X}_n + \frac{P_{n+1}}{\tau^2} \left(Y_{n+1} - a\widetilde{X}_n \right) .
\end{aligned}$$

Ceci termine l'argument de récurrence.

5. Comparer $\mathbb{E}\left[\left(\widetilde{X}_n - X_n\right)^2\right]$ et $\mathbb{E}\left[(\mathbb{E}[X_n \mid Y_n] - X_n)^2\right]$.

Solution 7 . *L'inclusion des tribus conditionnantes implique*

$$\mathbb{E}\left[\left(\widetilde{X}_n - X_n\right)^2\right] \leq \mathbb{E}\left[(\mathbb{E}[X_n \mid Y_n] - X_n)^2\right] .$$

Le membre gauche vaut P_n, le membre droit vaut $\tau^2/(\tau^2 + \sigma^2(1 - a^{2(n+1)})/(1 - a^2))$

(*aller plus loin : conditionnement et filtre de Kalman*) *cf. Chafaï-Malrieu, section 7.2*

Dans l'exercice suivant, il est question de la borne de Cramer-Rao. Dans sa forme la plus simple, cette borne stipule que la variance d'un estimateur sans biais est au moins aussi grande que l'inverse de l'information de Fisher. Cette borne est vérifiée dans les modèles expontiels de dimension 1 (entre autres).

Exercice 2 Dans cet exercice, $\begin{pmatrix} X \\ Y \end{pmatrix}$ est un vecteur Gaussien d'espérance $\mu = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$ et de covariance connue $C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. L'objectif est d'estimer θ_1. On dispose de n tirages indépendants distribués selon $\mathcal{N}(\mu, C)$: $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$.

Dans le scenario 1, θ_2 est connu du statisticien, dans le scenario 2, θ_2 n'est pas connu (c'est un *paramètre de nuisance* par opposition à θ_1 qui est un *paramètre d'intérêt*).

i) Proposer un estimateur de θ_1 au maximum de vraisemblance dans le scenario 1. Est il biaisé ? Quel est son risque quadratique ?

Solution 8 . *Dans ce scenario, comme θ_2 est connue, la loi de Y est connue. On peut considérer la loi de X conditionnellement à Y. C'est une loi Gaussienne dont la densité est explicitement connue (voir cours), c'est la loi des résidus lorsqu'on cherche à prédire X à partir de Y, c'est à dire $\mathcal{N}(\theta_1 + \rho\sigma_1/\sigma_2(y - \theta_2), \sigma_1^2 - \rho^2\sigma_1^2\sigma_2^2/\sigma_2^2) = \mathcal{N}(\theta_1 + \rho\sigma_1/\sigma_2(y - \theta_2), \sigma_1^2(1 - \rho^2))$.*
La variance de X sachant Y n'est autre que le complément de Schur de σ_2^2 dans C, c'est la variance de X (σ_1^2) diminuée de la variance dite expliquée $\rho^2\sigma_1^2$.
Si on écrit la log-vraisemblance, dans le scenario 1, pour une observation, on obtient

$$\ell(\theta_1) = -\frac{1}{2}\log\left(2\pi\sigma_1^2(1-\rho^2)\right) - \frac{(x - \theta_1 - \rho\sigma_1/\sigma_2(y-\theta_2))^2}{2\sigma_1^2(1-\rho^2)}\,.$$

Ceci conduit à

$$\widehat{\theta}_1 = x - \rho\sigma_1/\sigma_2(y - \theta_2)$$

et pour un n -échantillon à

$$\widehat{\theta}_1 = \overline{X}_n - \rho\sigma_1/\sigma_2(\overline{Y}_n - \theta_2)\,.$$

Cet estimateur linéaire est sans biais, gaussien de variance $\frac{\sigma_1^2(1-\rho^2)}{n}$.

ii) Proposer un estimateur de θ_1 au maximum de vraisemblance dans le scenario 2. Est il biaisé ? Quel est son risque quadratique ?

Solution 9 . *Si on écrit la log-vraisemblance, dans le scenario 2, pour un n-échantillon, on obtient*

$$\ell(\theta_1, \theta_2) = -\frac{1}{2}\log\left((2\pi)^2\sigma_1^2\sigma^2(1-\rho^2)\right) - \frac{1}{2}\begin{pmatrix} \overline{X}_n - \theta_1 & \overline{Y}_n - \theta_2 \end{pmatrix} C^{-1} \begin{pmatrix} \overline{X}_n - \theta_1 \\ \overline{Y}_n - \theta_2 \end{pmatrix}\,.$$

L'estimateur au maximum de vraisemblance est

$$\begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{pmatrix} = \begin{pmatrix} \overline{X}_n \\ \overline{Y}_n \end{pmatrix}$$

Cet estimateur est lui aussi linéaire, gaussien, sans biais de covariance, $\frac{1}{n}C$. L'estimateur de θ_1 est donc sans biais, gaussien, de covariance σ_1^2/n. On n'utilise pas $Y_1, \ldots, Y_n$ pour estimer θ_1.

iii) Calculer l'information de Fisher dans les deux scenarios. Les deux estimateurs réalisent-ils la borne de Cramer-Rao ?

Solution 10 . *Dans les deux modèles, l'information de Fischer est l'inverse de la covariance, les deux estimateurs réalisent la borne de Cramer-Rao.*

iv) Comparer les risques des deux estimateurs. Le paramètre θ_2 est il vraiment une nuisance ?

Solution 11 . *Les risques quadratiques sont $(1-\rho^2)\sigma_1^2$ et σ_1^2. Le paramètre θ_2 est bien un paramètre de nuisance (sauf bien sûr si $\rho = 0$, c'est-à-dire si X et Y sont indépendantes).*