Introduction aux processus stochastiques

Examen — 2007

Durée : 4h. Pas de gsm ni de calculatrice.

1 Chaînes de Markov

Acides aminés distincts présents dans une séquence.

On considère des séquences de longueur N d'acides aminés, avec M acides aminés différents possibles. On dispose d'un modèle chaîne de Markov (π, Π) qui reflète les règles de succession entre deux acides aminés de la séquence. Pour rappel, π est un vecteur ligne dont l'élément j vaut la probabilité que le premier acide aminé soit j, et Π une matrice dont l'élément ij vaut la probabilité que le $(n+1)$-ème acide aminé soit j si le n-ème acide aminé est i.
On veut connaître l'espérance du nombre d'acides aminés distincts qu'on peut trouver dans une séquence de longueur N.

1. Quelle est la probabilité $p_{\sim j}$ que l'acide aminé j n'apparaisse pas dans une séquence de longueur N ?
   En déduire la probabilité p_j que l'acide aminé j apparaisse dans une séquence de longueur N.

2. Que vaut l'espérance de la fonction indicatrice $\mathcal{X}_A$ d'un événement A ?
   ($\mathcal{X}_A$ est une variable aléatoire qui vaut 1 si l'événement survient, 0 sinon.)

3. Soit $\mathcal{Y} = \{1, 2, \ldots, M\}$ la variable aléatoire qui donne le nombre d'acides aminés distincts rencontrées sur une séquence de longueur N. Quelle est son espérance $\mathbf{E}\{\mathcal{Y}\}$?

A présent on veut connaître la variance du nombre d'acides aminés distincts dans une séquence de longueur N.

4. Quelle est la probabilité $p_{\sim jk}$ que ni l'acide aminé j ni l'acide aminé $k \neq j$ n'apparaissent dans une séquence de longueur N ?

5. Soit A et B deux événements non nécessairement disjoints définis sur Ω. On note $\overline{A}$ le complémentaire de A, et $\overline{B}$ celui de B. Vérifier que
   $$\text{Prob}(A \cap B) = 1 - \text{Prob}(\overline{A}) - \text{Prob}(\overline{B}) + \text{Prob}(\overline{A} \cap \overline{B}),$$
   puis calculer la probabilité p_{jk} que les acides aminés j et $k \neq j$ apparaissent dans une séquence de longueur N.
6. Que vaut l'espérance du carré de la fonction indicatrice $\mathcal{X}_A$ d'un événement A ?
   Que vaut l'espérance du produit des fonctions indicatrices $\mathcal{X}_A$ et $\mathcal{X}_B$ des événements A et B ?

7. Quelle est l'espérance du carré du nombre d'acides aminés distincts dans une séquence de longueur N, i.e. $\mathbf{E}\{\mathcal{Y}^2\}$? Répondre en fonction des $p_{\sim j}$ et des $p_{\sim jk}$.
   En déduire la variance $\text{var}\{\mathcal{Y}\}$.

2 Chaînes de Markov cachées

Inférence à partir d'observations d'état bruitées.

Soit une chaîne de Markov cachée invariante dans le temps (π, Π, Σ) à 3 états S_1, S_2, S_3 et 3 observations possibles E_1, E_2, E_3.
On note $\mathcal{X}_t$ la séquence des états cachés et $\mathcal{Y}_t$ la séquence des observations, pour t = $1, 2, \ldots, N$.
Pour rappel,

$$\pi_i = P(\mathcal{X}_1 = S_i) \qquad \Pi_{ij} = P(\mathcal{X}_{t+1} = S_j | \mathcal{X}_t = S_i) \qquad \Sigma_{ij} = P(\mathcal{Y}_t = E_j | \mathcal{X}_t = S_i).$$

On considère le cas particulier

$$\pi = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}/3 \qquad \Pi = \begin{bmatrix} .7 & .3 & 0 \\ 0 & .7 & .3 \\ .3 & 0 & .7 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} .8 & .1 & .1 \\ .1 & .8 & .1 \\ .1 & .1 & .8 \end{bmatrix}.$$

On dispose d'une séquence $Y_1\ Y_2 \ldots Y_N$ d'observations, où chaque $Y_i \in \{E_1, E_2, E_3\}$.
On considère le cas particulier où $N = 3$ et la séquence observée est

$$E_3\ E_2\ E_1.$$

Autrement dit, on a $Y_1 = E_3$, $Y_2 = E_2$ et $Y_3 = E_1$.

Dans les questions qui viennent, la résolution numérique n'est pas demandée.

1. Dessiner le diagramme de transition d'états de la chaîne de Markov (π, Π).
2. Dessiner le réseau bayesien de la chaîne de Markov cachée (π, Π, Σ) pour les 3 premiers instants.
3. *Propagation α/Estimation d'état.*
   Calculer de façon efficace les probabilités

   $$P(\mathcal{X}_3 = S_i | \mathcal{Y}_1 = Y_1, \mathcal{Y}_2 = Y_2, \mathcal{Y}_3 = Y_3),$$

   pour $i \in \{1, 2, 3\}$ et la séquence particulière observée.
   En notation abrégée, la question revient à demander $P(\mathcal{X}_3 | Y_1 Y_2 Y_N)$ avec $Y_1 = E_3$, $Y_2 = E_2$, $Y_3 = E_1$.
   Suggestion : se ramener au calcul de $P(\mathcal{X}_t = S_i, \mathcal{Y}_1 = Y_1, \ldots, \mathcal{Y}_t = Y_t)$, c'est-à-dire en notation abrégée $P(\mathcal{X}_t, Y_1 \ldots Y_N)$, pour t valant successivement 1,2,3. Particulariser ensuite les expressions obtenues aux observations de l'énoncé, i.e. $Y_1 = E_3$, $Y_2 = E_2$ et $Y_3 = E_1$.
4. A présent, la première observation n'est plus disponible, la matrice de transition devient Π' pour la seconde transition d'état, et la matrice d'observation à l'instant $t = 3$ devient Σ'. Comment calculez-vous

   $$P(\mathcal{X}_3 = S_i | \mathcal{Y}_2 = Y_2, \mathcal{Y}_3 = Y_3),$$

   pour $i \in \{1, 2, 3\}$ dans ces nouvelles conditions et la séquence particulière observée, i.e.

   $$Y_2 = E_2,\ Y_3 = E_1?$$

3 Estimation statique

Le Bon, la Brute et le Truand.

On utilise trois senseurs y_1, y_2, y_3 de qualités différentes pour effectuer simultanément la mesure d'un signal source x :

- Le Bon mesure $y_1 = x + v_1$, où v_1 est un bruit de moyenne nulle et de variance 0.1,
- La Brute mesure $y_2 = x + v_2$, où v_2 est un bruit de moyenne nulle et de variance 0.4,
- Le Truand mesure $y_3 = 0.9x + v_3$, où v_3 est un bruit de moyenne 0.1 et de variance 0.2.

La source suit une distribution normale de moyenne 1 et de variance 2.
Les bruits v_1, v_2, v_3 et la source x sont mutuellement indépendants.

Dans les questions qui viennent, vous pouvez arrêter la résolution numérique à des expressions du type $\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}^{-2} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^{3}$.

1. Calculer l'estimateur $\hat{x} = \mathrm{E}\{x|y_1, y_2, y_3\}$.
   Conseil : Etablissez proprement les expressions de Σ_y et Σ_{xy}, pour disposer d'un calcul de référence lorsque vous effectuerez un calcul similaire au point 3.

2. Calculer l'espérance de l'erreur quadratique d'estimation, c'est-à-dire $\mathrm{E}\{(x - \hat{x})^2\}$.

3. Supposer à présent que le bruit v_3 est corrélé avec la source x.
   Les autres relations d'indépendances sont conservées.
   Le facteur de corrélation est $\rho = 1/\sqrt{10}$.
   Etablir les nouvelles expressions de $\mathrm{E}\{x|y_1, y_2, y_3\}$ et $\mathrm{E}\{(x - \hat{x})^2\}$.
   Rappel : $\rho = \rho_{x,v_3} = \mathrm{covar}\{x, v_3\}/\sqrt{\mathrm{var}\{x\}\mathrm{var}\{v_3\}}$.

$\star$. Bonus. (Ce point peut être résolu indépendamment des réponses aux points précédents.)
Supposer (comme au point précédent) que le bruit v_3 est corrélé avec la source (facteur de corrélation ρ).
A votre avis, que vaut l'espérance de l'erreur quadratique d'estimation $\mathrm{E}\{(x - \hat{x})^2\}$ dans le cas $\rho = -1$? Dans le cas $\rho = 1$?
Esquisser grossièrement le graphe de $\mathrm{E}\{(x - \hat{x})^2\}$ en fonction de $\rho \in [-1, 1]$.
Vous pouvez aussi essayer de calculer l'estimateur $\hat{x} = \mathrm{E}\{x|y_1, y_2, y_3\}$ dans les cas $\rho = -1$, $\rho = 1$, mais ceci demande de l'imagination dans la méthode.

4 Filtre de Kalman

Pour quelques dollars de plus.

Reprenez l'énoncé *Le Bon, la Brute et le Truand* et les sous-questions 1 et 2.
Cette fois, utilisez la technique du filtre de Kalman pour calculer progressivement $\hat{x}(1|1) = \mathrm{E}\{x|y_1\}$, $\hat{x}(2|2) = \mathrm{E}\{x|y_1, y_2\}$, $\hat{x}(3|3) = \mathrm{E}\{x|y_1, y_2, y_3\}$ et les espérances de l'erreur quadratique associées.

Annexe : Filtre de Kalman

- On considère dans la théorie
  - un système dont l'état suit la dynamique $x(t+1) = Ax(t) + w(t)$,
  - l'équation de mesure $y(t) = Cx(t) + v(t)$,
  - $x(0), w(0), w(1), \ldots, v(0), v(1), \ldots$ indépendants et conjointement gaussiens,
  - $w(t)$ bruit de moyenne nulle et de variance W,
  - $v(t)$ bruit de moyenne nulle et de variance V,
  - $x(0)$ condition initiale de moyenne $\bar{x}(0)$ et de variance Σ_0.
- On dispose de la *récursion de Riccati*

$$\hat{\Sigma}_{t+1} = A\hat{\Sigma}_t A^T + W - A\hat{\Sigma} C^T (C\hat{\Sigma}_t C^T + V)^{-1} C\hat{\Sigma}_t A^T \tag{1}$$

qui permet de calculer itérativement

$$\hat{\Sigma}_{t+1} \triangleq \Sigma_{t+1|t} \triangleq \mathbf{E}\{(x(t+1) - \hat{x}(t+1|t))(x(t+1) - \hat{x}(t+1|t))^T\} \tag{2}$$

$$\text{avec} \quad \hat{x}(t+1|t) \triangleq \mathbf{E}\{x(t+1)|y(0), \ldots, y(t)\}. \tag{3}$$

i.e. la matrice de covariance $\hat{\Sigma}_{t+1}$ de l'erreur d'estimation de l'état à l'instant $t+1$ connaissant les observations du temps 0 au temps t.

- Pour le filtre de Kalman, on dispose tout d'abord du gain de l'observateur à l'instant t

$$L_t = A\hat{\Sigma}_t C^T (C\hat{\Sigma}_t C^T + V)^{-1}. \tag{4}$$

On définit $\hat{y}(t)$ comme la prédiction de l'observation au temps t connaissant les observations précédentes :

$$\hat{y}(t) \triangleq \hat{y}(t|t-1) \triangleq \mathbf{E}\{y(t)|y(0), \ldots, y(t-1)\} = C\hat{x}(t). \tag{5}$$

La formule de mise à jour de la prédiction de l'état au temps $t+1$ en fonction d'une nouvelle observation recueillie au temps t est le filtre de Kalman :

$$\hat{x}(t+1) \triangleq \hat{x}(t+1|t) = A\hat{x}(t) + L_t(y(t) - \hat{y}(t)). \tag{6}$$

Solutions

Acides aminés distincts présents dans une séquence.

1. Soit $\pi_{(j)}$ le vecteur ligne obtenu en supprimant l'élément j de la distribution initiale π.
   Soit $\Pi_{(j)}$ la matrice obtenue en supprimant la ligne et la colonne j de la matrice de transition Π.
   Si N vaut un, évidemment $p_{\sim j} = \sum_k (\pi_{(j)})_k = \pi_{(j)}\mathbf{1}$ avec $\mathbf{1}$ un vecteur colonne de 1.
   Plus généralement, pour tout N, on a $p_{\sim j} = \pi_{(j)}\Pi_{(j)}^{N-1}\mathbf{1}$.
   La probabilité p_j que l'acide aminé j apparaisse dans la séquence de longueur N vaut donc $p_j = 1 - p_{\sim j} = 1 - \pi_{(j)}\Pi_{(j)}^{N-1}\mathbf{1}$.
2. Soit p_A la probabilité que l'événement A survienne. Alors,
   $\mathbf{E}\{X_A\} = p_A \cdot 1 + (1 - p_A) \cdot 0 = p_A$.
3. Soit $\mathcal{X}_j$ la variable indicatrice de l'événement "L'acide aminé j est présent dans la séquence de longueur N". Soit $\mathcal{Y}$ le nombre d'acides aminés distincts dans la séquence de longueur N.
   On a $\mathcal{Y} = \sum_{j=1}^{M} \mathcal{X}_j$.
   L'espérance est un opérateur linéaire. Donc $\mathbf{E}\{\mathcal{Y}\} = \sum_{j=1}^{M} \mathbf{E}\{\mathcal{X}_j\} = \sum_{j=1}^{M} p_j$.
   Ainsi, l'espérance du nombre d'acides aminés distincts dans une séquence de longueur N vaut $\mathbf{E}\{\mathcal{Y}\} = M - \sum_{j=1}^{M} \pi_{(j)}\Pi_{(j)}^{N-1}\mathbf{1}$.
4. Soit $\pi_{(jk)}$ le vecteur ligne obtenu en supprimant les éléments j et k de la distribution initiale π.
   Soit $\Pi_{(jk)}$ la matrice obtenue en supprimant les lignes et colonnes j et k de la matrice de transition Π.
   La probabilité que les acides aminés j et $k \neq j$ n'apparaissent pas dans la séquence de longueur N vaut $p_{\sim jk} = \pi_{(jk)}\Pi_{(jk)}^{N-1}\mathbf{1}$.
5. On a $\text{Prob}(A \cap B) = 1 - \text{Prob}(\overline{A \cap B}) = 1 - \text{Prob}(\overline{A} \cup \overline{B}) = 1 - [\text{Prob}(\overline{A}) + \text{Prob}(\overline{B}) - \text{Prob}(\overline{A} \cap \overline{B})]$.
   La probabilité que les acides aminés j et $k \neq j$ apparaissent dans une séquence de longueur N vaut

   $$p_{jk} = 1 - p_{\sim j} - p_{\sim k} + p_{\sim jk} = 1 - \pi_{(j)}\Pi_{(j)}^{N-1}\mathbf{1}_{M-1} - \pi_{(k)}\Pi_{(k)}^{N-1}\mathbf{1}_{M-1} + \pi_{(jk)}\Pi_{(jk)}^{N-1}\mathbf{1}_{M-2}.$$

   On a ici précisé la dimension des vecteurs formés de 1.
6. Soit p_A la probabilité que l'événement A survienne. Comme $\mathcal{X}_A^2$ vaut 1^2 si A est réalisé et 0^2 sinon, $\mathbf{E}\{\mathcal{X}_A^2\} = p_A$.
   Comme $\mathcal{X}_A\mathcal{X}_B$ vaut 1 si A et B sont réalisés et 0 sinon, $\mathbf{E}\{\mathcal{X}_A\mathcal{X}_B\} = p_{A\cap B}$.

7. On a

$$\begin{aligned}
\mathbf{E}\{\mathcal{Y}^2\} &= \mathbf{E}\{(\sum_{j=1}^{M}\mathcal{X}_j)(\sum_{k=1}^{M}\mathcal{X}_k)\} = \mathbf{E}\{\sum_{j=1}^{M}\mathcal{X}_j^2 + \sum_{j=1}^{M}\sum_{\substack{k=1\\k\neq j}}^{M}\mathcal{X}_j\mathcal{X}_k\} \\
&= \sum_{j=1}^{M}\mathbf{E}\{\mathcal{X}_j^2\} + \sum_{j=1}^{M}\sum_{\substack{k=1\\k\neq j}}^{M}\mathbf{E}\{\mathcal{X}_j\mathcal{X}_k\} = \sum_{j=1}^{M}p_j + \sum_{j=1}^{M}\sum_{\substack{k=1\\k\neq j}}^{M}p_{jk} \\
&= M - \sum_{j=1}^{M}p_{\sim j} + (M^2 - M) - 2(M-1)\sum_{j=1}^{M}p_{\sim j} + \sum_{j=1}^{M}\sum_{\substack{k=1\\k\neq j}}^{M}p_{\sim jk} \\
&= M^2 - (2M-1)\sum_{j=1}^{M}p_{\sim j} + \sum_{j=1}^{M}\sum_{\substack{k=1\\k\neq j}}^{M}p_{\sim jk}.
\end{aligned}$$

La variance du nombre d'acides aminés distincts dans une séquence de longueur N vaut donc

$$\begin{aligned}
\text{var}\{\mathcal{Y}\} = \mathbf{E}\{\mathcal{Y}^2\} - \mathbf{E}\{\mathcal{Y}\}^2 &= M^2 - (2M-1)\sum_{j=1}^{M}p_{\sim j} + \sum_{j=1}^{M}\sum_{\substack{k=1\\k\neq j}}^{M}p_{\sim jk} - (M - \sum_{j=1}^{M}p_{\sim j})^2 \\
&= \sum_{j=1}^{M}p_{\sim j} + \sum_{j=1}^{M}\sum_{\substack{k=1\\k\neq j}}^{M}p_{\sim jk} - (\sum_{j=1}^{M}p_{\sim j})^2 \\
&= \sum_{j=1}^{M}\pi_{(j)}\Pi_{(j)}^{N-1}\mathbf{1} + \sum_{j=1}^{M}\sum_{\substack{k=1\\k\neq j}}^{M}\pi_{(jk)}\Pi_{(jk)}^{N-1}\mathbf{1} - (\sum_{j=1}^{M}\pi_{(j)}\Pi_{(j)}^{N-1}\mathbf{1})^2.
\end{aligned}$$

8. Remarque [non demandé].
Notons $\mathcal{Y}_\infty$ la variable aléatoire $\mathcal{Y}$ dans le cas où la longueur de la séquence tend vers $+\infty$.
Une condition suffisante pour que var$\{\mathcal{Y}_\infty\} = 0$ est que la chaîne de Markov soit irréductible.
On a alors

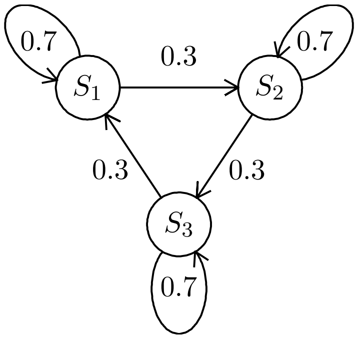
$$\mathbf{E}\{\mathcal{Y}_\infty\} = M.$$

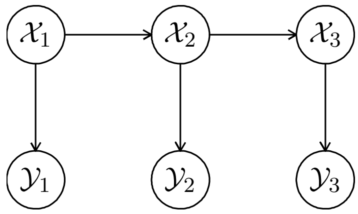
Une condition nécessaire pour que var$\{\mathcal{Y}_\infty\} > 0$ est que la chaîne de Markov ne soit pas irréductible, et que la décomposition de la matrice de transition $\Pi \in \mathbf{R}^{M\times M}$ en blocs irréductibles donne des blocs de dimensions distinctes. On a alors

$$0 < L \leq \mathbf{E}\{\mathcal{Y}_\infty\} \leq U < M$$

où L est la dimension du plus petit bloc irréductible, et U la dimension du plus grand bloc irréductible.
Les valeurs exactes de $\mathbf{E}\{\mathcal{Y}_\infty\}$ et var$\{\mathcal{Y}_\infty\}$ dépendent de la distribution initiale π.

Inférence à partir d'observations d'état bruitées.

1. 

2. 

3. Comme

$$P(\mathcal{X}_t = S_i | Y_1 \dots Y_N) = \frac{P(\mathcal{X}_t = S_i, Y_1 \dots Y_N)}{\sum_{k=1}^{3} P(\mathcal{X}_t = S_k, Y_1 \dots Y_N)},$$

tout se ramène bien à calculer efficacement $P(\mathcal{X}_N = S_i, Y_1 \dots Y_N)$.
Pour y parvenir, on va partir de l'état 1 et ajouter les observations séquentiellement.
Sans observation,

$$\hat{\alpha}_1 = \begin{bmatrix} P(\mathcal{X}_1 = S_1) & P(\mathcal{X}_1 = S_2) & P(\mathcal{X}_1 = S_3) \end{bmatrix} = \pi.$$

Avec Y_1,

$$\begin{aligned} \alpha_1 &= \begin{bmatrix} P(\mathcal{X}_1 = S_1, Y_1) & P(\mathcal{X}_1 = S_2, Y_1) & P(\mathcal{X}_1 = S_3, Y_1) \end{bmatrix} \\ &= \begin{bmatrix} P(\mathcal{X}_1 = S_1) P(Y_1 | \mathcal{X}_1 = S_1) & \dots & P(\mathcal{X}_1 = S_3) P(Y_1 | \mathcal{X}_1 = S_3) \end{bmatrix} \\ &= \hat{\alpha}_1 * (\Sigma \mathbf{e}_{Y_1})^T = \hat{\alpha}_1 * (\Sigma \mathbf{e}_3)^T \end{aligned}$$

où $*$ est le produit de Hadamard (produit élément par élément) et $\mathbf{e}_3 = [0\ 0\ 1]^T$.
Etat suivant sans observation supplémentaire :

$$\begin{aligned} \hat{\alpha}_2 &= \begin{bmatrix} P(\mathcal{X}_2 = S_1, Y_1) & P(\mathcal{X}_2 = S_2, Y_1) & P(\mathcal{X}_2 = S_3, Y_1) \end{bmatrix} \\ &= \begin{bmatrix} \sum_{k=1}^{3} P(\mathcal{X}_1 = S_k, \mathcal{X}_2 = S_1, Y_1) & \dots & \sum_{k=1}^{3} P(\mathcal{X}_1 = S_k, \mathcal{X}_2 = S_3, Y_1) \end{bmatrix} \\ &= \begin{bmatrix} \sum_{k=1}^{3} P(\mathcal{X}_1 = S_k, Y_1) P(\mathcal{X}_2 = S_1 | \mathcal{X}_1 = S_k) & \dots & \sum_{k=1}^{3} P(\mathcal{X}_1 = S_k, Y_1) P(\mathcal{X}_2 = S_3 | \mathcal{X}_1 = S_k) \end{bmatrix} \\ &= \alpha_1 \Pi. \end{aligned}$$

Ajout de l'observation Y_2 :

$$\begin{aligned} \alpha_2 &= \begin{bmatrix} P(\mathcal{X}_2 = S_1, Y_1, Y_2) & P(\mathcal{X}_2 = S_2, Y_1, Y_2) & P(\mathcal{X}_2 = S_3, Y_1, Y_2) \end{bmatrix} \\ &= \begin{bmatrix} P(\mathcal{X}_2 = S_1, Y_1) P(Y_2 | \mathcal{X}_2 = S_1) & \dots & P(\mathcal{X}_2 = S_3, Y_1) P(Y_2 | \mathcal{X}_2 = S_3) \end{bmatrix} \\ &= \hat{\alpha}_2 * (\Sigma \mathbf{e}_{Y_2})^T = \hat{\alpha}_2 * (\Sigma \mathbf{e}_2)^T \end{aligned}$$

où $\mathbf{e}_2 = [0\ 1\ 0]^T$.
Etat suivant sans observation supplémentaire :

$$\hat{\alpha}_3 = \begin{bmatrix} P(\mathcal{X}_3 = S_1, Y_1, Y_2) & \dots & P(\mathcal{X}_3 = S_3, Y_1, Y_2) \end{bmatrix} = \alpha_2 \Pi.$$

Ajout de l'observation Y_3 :

$$\alpha_3 = \left[\begin{array}{ccc} P(\mathcal{X}_3 = S_1, Y_1, Y_2, Y_3) & \ldots & P(\mathcal{X}_3 = S_3, Y_1, Y_2, Y_3) \end{array} \right] = \hat{\alpha}_3 * (\Sigma \mathbf{e}_{Y_3})^T = \hat{\alpha}_3 * (\Sigma \mathbf{e}_1)^T$$

où $\mathbf{e}_1 = [1\ 0\ 0]^T$.
Finalement,

$$p_3 = \left[\begin{array}{ccc} P(\mathcal{X}_3 = S_1 | Y_1, Y_2, Y_3) & \ldots & P(\mathcal{X}_3 = S_3 | Y_1, Y_2, Y_3) \end{array} \right] = \alpha_3 / (\alpha_3 \mathbf{1}),$$

où $\mathbf{1} = [\begin{array}{ccc} 1 & 1 & 1 \end{array}]^T$.

4. L'algorithme d'estimation d'état devient

$$\begin{aligned} \hat{\alpha}_1 &= \pi \\ \alpha_1 &= \hat{\alpha}_1 \\ \hat{\alpha}_2 &= \alpha_1 \Pi \\ \alpha_2 &= \hat{\alpha}_2 * (\Sigma \mathbf{e}_2)^T \\ \hat{\alpha}_3 &= \alpha_2 \Pi' \\ \alpha_3 &= \hat{\alpha}_3 * (\Sigma' \mathbf{e}_1)^T, \end{aligned}$$

et finalement, dans les nouvelles conditions de l'énoncé,

$$p_3' = \left[\begin{array}{ccc} P(\mathcal{X}_3 = S_1 | Y_2, Y_3) & \ldots & P(\mathcal{X}_3 = S_3 | Y_2, Y_3) \end{array} \right] = \frac{\alpha_3}{\alpha_3 \mathbf{1}}.$$

Le Bon, la Brute et le Truand.

On pose $y = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^T$.
L'équation de mesure s'écrit

$$y = Cx + v \qquad \text{avec } C = \begin{bmatrix} 1 \\ 1 \\ 0.9 \end{bmatrix}, \quad v \sim \mathcal{N}(\bar{v}, \Sigma_v)$$

$$\text{où } \bar{v} = \begin{bmatrix} 0 \\ 0 \\ 0.1 \end{bmatrix}, \quad \Sigma_v = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}.$$

Les espérances de x et de y sont respectivement

$$\bar{x} = 1 \qquad \bar{y} = C\bar{x} + \bar{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{1}.$$

Les variables x et y sont conjointement gaussiennes :

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix}).$$

Pour établir les expressions pour Σ_{xy} et Σ_y, on observe que

$$\begin{bmatrix} x \\ v \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} \bar{x} \\ \bar{v} \end{bmatrix}, \begin{bmatrix} \Sigma_x & 0 \\ 0 & \Sigma_v \end{bmatrix}), \quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix}, \quad \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{v} \end{bmatrix},$$

puis on évalue

$$\begin{aligned} \mathrm{E}\begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix} \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}^T &= \mathrm{E}\begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \begin{bmatrix} x - \bar{x} \\ v - \bar{v} \end{bmatrix} \begin{bmatrix} x - \bar{x} \\ v - \bar{v} \end{bmatrix}^T \begin{bmatrix} I & 0 \\ C & I \end{bmatrix}^T \\ &= \begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \begin{bmatrix} \Sigma_x & 0 \\ 0 & \Sigma_v \end{bmatrix} \begin{bmatrix} I & 0 \\ C & I \end{bmatrix}^T = \begin{bmatrix} \Sigma_x & \Sigma_x C^T \\ C\Sigma_x & C\Sigma_x C^T + \Sigma_v \end{bmatrix} \end{aligned}$$

qu'on identifie avec $\begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_y \end{bmatrix}$.
Dès lors,

$$\begin{aligned} \Sigma_x &= 2 \\ \Sigma_{xy} &= \Sigma_x C^T = 2 \cdot \begin{bmatrix} 1 & 1 & 0.9 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 1.8 \end{bmatrix} \\ \Sigma_y &= C\Sigma_x C^T + \Sigma_v \\ \text{[non demandé :]} \quad &= 2 \cdot \begin{bmatrix} 1 & 1 & 0.9 \\ 1 & 1 & 0.9 \\ 0.9 & 0.9 & 0.81 \end{bmatrix} + \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} = \begin{bmatrix} 2.1 & 2 & 1.8 \\ 2 & 2.4 & 1.8 \\ 1.8 & 1.8 & 1.82 \end{bmatrix}. \end{aligned}$$

1. L'espérance de x sachant y_1, y_2, y_3 est

$$\hat{x} = \mathrm{E}\{x|y\} = \bar{x} + \Sigma_{xy}\Sigma_y^{-1}(y - \bar{y})$$

$$\text{[non demandé :]} = 1 + \begin{bmatrix} 0.5865 & 0.1466 & 0.2639 \end{bmatrix} \begin{bmatrix} y_1 - 1 \\ y_2 - 1 \\ y_3 - 1 \end{bmatrix}.$$

2. L'espérance de l'erreur quadratique d'estimation est

$$\mathrm{E}\{(x - \hat{x})^2\} = \mathrm{tr}\ \Sigma_{\text{est}} = \mathrm{tr}(\Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T)$$

$$\text{[non demandé :]} = \mathrm{tr}\left(2 - \begin{bmatrix} 0.5865 & 0.1466 & 0.2639 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 1.8 \end{bmatrix}\right) = 0.0587.$$

3. Cette fois

$$\begin{bmatrix} x \\ v \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{x} \\ \bar{v} \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{xv} \\ \Sigma_{xv}^T & \Sigma_v \end{bmatrix}\right)$$

avec

$$\Sigma_{xv} = \begin{bmatrix} 0 & 0 & \rho\sqrt{\sigma_x^2\sigma_{v_3}^2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0.2 \end{bmatrix},$$

de sorte que

$$\mathrm{E}\begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}\begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}^T = \begin{bmatrix} I & 0 \\ C & I \end{bmatrix}\begin{bmatrix} \Sigma_x & \Sigma_{xv} \\ \Sigma_{xv}^T & \Sigma_v \end{bmatrix}\begin{bmatrix} I & 0 \\ C & I \end{bmatrix}^T$$

$$= \begin{bmatrix} \Sigma_x & \Sigma_x C^T + \Sigma_{xv} \\ C\Sigma_x + \Sigma_{xv}^T & C\Sigma_x C^T + C\Sigma_{xv} + \Sigma_{xv}^T C^T + \Sigma_v \end{bmatrix},$$

et par identification (valeurs numériques non demandées),

$$\Sigma_{xy} = \Sigma_x C^T + \Sigma_{xv} = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix},$$

$$\Sigma_y = C\Sigma_x C^T + \Sigma_v + C\Sigma_{xv} + \Sigma_{xv}^T C^T = \begin{bmatrix} 2.1 & 2 & 2 \\ 2 & 2.4 & 2 \\ 2 & 2 & 2.18 \end{bmatrix}.$$

Les expressions de $\mathrm{E}\{x|y_1, y_2, y_3\}$ et $\mathrm{E}\{(x - \hat{x})^2\}$ restent alors inchangées si on tient compte de la redéfinition de Σ_{xy} et Σ_y.
Non demandé :

$$\mathrm{E}\{x|y_1, y_2, y_3\} = 1 + \begin{bmatrix} 0.5389 & 0.1347 & 0.2994 \end{bmatrix}\begin{bmatrix} y_1 - 1 \\ y_2 - 1 \\ y_3 - 1 \end{bmatrix}$$

$$\mathrm{E}\{(x - \hat{x})^2\} = 0.0539.$$

4. Si $\rho = 1$ ou $\rho = -1$, intuitivement il n'y pas plus d'incertitude sur la source une fois que y_3 est mesuré, et l'erreur d'estimation devrait être nulle. L'espérance de celle-ci serait donc nulle.
Il vaut toutefois mieux s'assurer explicitement qu'on peut bien estimer x avec y_3.
Le bruit v_3 dans l'équation de mesure $y_3 = 0.9x + v_3$ est nécessairement de la forme

$$v_3 = ax + b, \tag{7}$$

pour avoir x, v_3 conjointement gaussiens.
Les coefficients a et b sont obtenus en exprimant les conditions $\overline{v_3} = 0.1$ et $\rho_{xv_3} = \pm 1$.
D'une part,

$$0.1 = \overline{v_3} = a\overline{x} + b = a \cdot 1 + b \quad \Rightarrow \quad b = 0.1 - a. \tag{8}$$

D'autre part,

$$\rho_{xv_3} = \pm 1 \quad \Rightarrow \quad \text{covar}(x, v_3) = \pm\sqrt{\text{var}(x)}\sqrt{\text{var}(v_3)} = \pm\sqrt{2 \cdot 0.2}$$

avec

$$\begin{aligned}\text{covar}(x, v_3) &= \mathbf{E}\{xv_3\} - \mathbf{E}\{x\}\mathbf{E}\{v_3\} = \mathbf{E}\{x(ax+b)\} - 1 \cdot 0.1 \\ &= a\mathbf{E}\{x^2\} + 1 \cdot b - 0.1 = a(\text{var}(x) + \mathbf{E}\{x\}^2) + b - 0.1 \\ &= a(2 + 1^2) + b - 0.1 = 3a - a = 2a,\end{aligned}$$

en tenant compte à la dernière ligne de (8).
Ainsi,

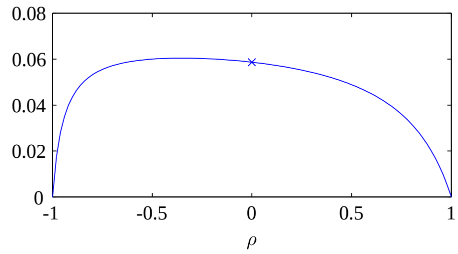
$$2a = \text{covar}(x, v_3) = \pm 2\sqrt{0.1} \quad \Rightarrow \quad a = \pm\sqrt{0.1}. \tag{9}$$

De (7), (8) et (9), on tire $y_3 = 0.9x + v_3 = 0.9x \pm \sqrt{0.1}x + 0.1 \mp \sqrt{0.1}$, d'où

$$x = \frac{y_3 - 0.1 \pm \sqrt{0.1}}{0.9 \pm \sqrt{0.1}} = \mathbf{E}\{x|y_3\}.$$

Dans cette expression, les signes $+$ des symboles $\pm$ sont relatifs au cas $\rho_{xv_3} = 1$, et les signes $-$ au cas $\rho_{xv_3} = -1$.

Pour les autres valeurs de ρ, l'espérance de l'erreur quadratique est positive.
Le graphe exact de la variation de $\mathbf{E}\{(x - \mathbf{E}\{x|y_1, y_2, y_3\})^2\}$ en fonction de $\rho = \rho_{xv_3}$ est donné ici pour comparaison qualitative.



Pour quelques dollars de plus.

Les équations dynamiques sont (trivialement)

$$x_1 = x \quad \text{(condition initiale)}$$
$$x_{t+1} = x_t \quad \text{pour } t = 1, 2,$$

avec

$$x \sim \mathcal{N}\{\overline{x}, \Sigma_1\}, \qquad \overline{x} = 1, \qquad \Sigma_1 = \sigma_x^2 = 2.$$

Par rapport au modèle théorique $x_{t+1} = Ax_t + w_t$, avec $w_t \sim \mathcal{N}\{0, W\}$, on a ici

$$A = 1, \qquad W = 0.$$

Les équations de mesure sont

$$y_t = C_t x_t + v_t \quad \text{pour } t = 1, 2, 3$$

avec

$$\begin{array}{lll} C_1 = 1, & C_2 = 1, & C_3 = 0.9, \\ v_1 \sim \mathcal{N}(\overline{v_1}, V_1), & v_2 \sim \mathcal{N}(\overline{v_2}, V_2), & v_3 \sim \mathcal{N}(\overline{v_3}, V_3), \\ \overline{v_1} = 0, & \overline{v_2} = 0, & \overline{v_3} = 0.1, \\ V_1 = 0.1, & V_2 = 0.4, & V_3 = 0.2. \end{array}$$

On part des conditions initiales

$$t = 1, \qquad \hat{x}(1|0) = \bar{x} = 1, \qquad \hat{\Sigma}_1 = \Sigma(1|0) = \Sigma_1 = 2.$$

Les équations de mise à jour par les mesures (phase de correction) sont

$$\hat{x}(t|t) = \hat{x}(t|t-1) + \Sigma_{t|t-1} C_t^T (C_t \Sigma_{t|t-1} C_t^T + V_t)^{-1} (y(t) - C_t \hat{x}(t|t-1) - \overline{v_t}) \tag{10}$$
$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} C_t^T (C_t \Sigma_{t|t-1} C_t^T + V_t)^{-1} C_t \Sigma_{t|t-1}. \tag{11}$$

Ensuite on doit aux temps $t = 1, 2$ appliquer les équations de mise à jour temporelle (phase de prédiction)

$$\hat{x}(t+1|t) = A\hat{x}(t|t) = \hat{x}(t|t) \tag{12}$$
$$\Sigma_{t+1|t} = A\Sigma_{t|t} A^T + W = \Sigma_{t|t}. \tag{13}$$

L'algorithme de résolution se résume donc à appliquer successivement (10) et (11) pour les temps $t = 1, 2, 3$, en ayant posé $\hat{x}(t|t-1) = \hat{x}(t-1|t-1)$ aux temps $t = 2, 3$.
On a bien sûr

$$\mathbf{E}\{x | y_1, \ldots, y_t\} = \hat{x}(t|t), \qquad \mathbf{E}\{(x - \hat{x}(t|t))^2\} = \text{tr}\{\Sigma_{t|t}\} = \Sigma_{t|t}.$$

L'algorithme explicite (non demandé) est le suivant.

$$
\begin{aligned}
\hat{x}(1|1) &= \overline{x} + \Sigma_x C_1^T[C_1\Sigma_x C_1^T + V_1]^{-1}(y_1 - C_1\overline{x} - \overline{v}_1) \\
&= 1 + 2\cdot 1[1\cdot 2\cdot 1 + 0.1]^{-1}(y_1 - 1\cdot 1 - 0) \\
&= 1 + \frac{20}{21}(y_1 - 1) \\
&= 1 + 0.9524(y_1 - 1) \\
\Sigma_{1|1} &= \Sigma_x - \Sigma_x C_1^T[C_1\Sigma_x C_1^T + V_1]^{-1}C_1\Sigma_x \\
&= 2 - 2\cdot 1[1\cdot 2\cdot 1 + 0.1]^{-1}1\cdot 2 = 2 - \frac{4}{2.1} = \frac{2}{21} = 0.0952 \\
\hat{x}(2|2) &= \hat{x}(1|1) + \Sigma_{1|1}C_2^T[C_2\Sigma_{1|1}C_2^T + V_2]^{-1}(y_2 - C_2\hat{x}(1|1) - \overline{v}_2) \\
&= \hat{x}(1|1) + \Sigma_{1|1}\cdot 1[1\cdot\Sigma_{1|1}\cdot 1 + 0.4]^{-1}(y_2 - 1\cdot\hat{x}(1|1) - 0) \\
&= \hat{x}(1|1) + \frac{5}{26}(y_2 - \hat{x}(1|1)) \\
&= \hat{x}(1|1) + 0.1923(y_2 - \hat{x}(1|1)) \\
\Sigma_{2|2} &= \Sigma_{1|1} - \Sigma_{1|1}C_2^T[C_2\Sigma_{1|1}C_2^T + V_2]^{-1}C_2\Sigma_{1|1} \\
&= \Sigma_{1|1} - \Sigma_{1|1}\cdot 1[1\cdot\Sigma_{1|1}\cdot 1 + 0.4]^{-1}1\cdot\Sigma_{1|1} = \frac{1}{13} = 0.0769 \\
\hat{x}(3|3) &= \hat{x}(2|2) + \Sigma_{2|2}C_3^T[C_3\Sigma_{2|2}C_3^T + V_3]^{-1}(y_3 - C_3\hat{x}(2|2) - \overline{v}_3) \\
&= \hat{x}(2|2) + \Sigma_{2|2}\cdot 0.9[0.9\cdot\Sigma_{2|2}\cdot 0.9 + 0.2]^{-1}(y_3 - 0.9\cdot\hat{x}(2|2) - 0.1) \\
&= \hat{x}(2|2) + \frac{90}{341}(y_3 - \hat{x}(2|2) - 0.1) \\
&= \hat{x}(2|2) + 0.2639(y_3 - \hat{x}(2|2) - 0.1) \\
\Sigma_{3|3} &= \Sigma_{2|2} - \Sigma_{2|2}C_3^T[C_3\Sigma_{2|2}C_3^T + V_3]^{-1}C_3\Sigma_{2|2} \\
&= \Sigma_{2|2} - \Sigma_{2|2}\cdot 0.9[0.9\cdot\Sigma_{2|2}\cdot 0.9 + 0.2]^{-1}0.9\cdot\Sigma_{2|2} = \frac{20}{341} = 0.0587.
\end{aligned}
$$

Le calcul de $\Sigma_{1|1}$, $\Sigma_{2|2}$, $\Sigma_{3|3}$ aurait pu être placé avant celui de $\hat{x}(1|1)$, $\hat{x}(2|2)$, $\hat{x}(3|3)$.