

**École Nationale de la Statistique
et de l'Analyse de l'Information
examen du cours**

“Filtrage linéaire et non-linéaire”

lundi 25 janvier 2021, 14:00 à 16:00

L'objectif de ce problème est d'étudier le *lisseur* bayésien dans le cas simple d'un modèle de Markov caché, et d'établir plusieurs formes d'équations récurrentes décrivant son évolution. Ici, l'horizon temporel est fixé, c'est-à-dire que toutes les observations recueillies entre l'instant initial 0 et l'instant final n sont disponibles et peuvent être utilisées pour estimer l'état caché à un instant k intermédiaire entre 0 et n (le cas où $k = n$ correspond bien sûr au cas du *filtre* bayésien vu en cours).

Pour fixer les idées, on considère un modèle de Markov caché, où les états cachés $\{X_k\}$ forment une chaîne de Markov à valeurs dans E , et où *conditionnellement aux états cachés* les observations $\{Y_k\}$ forment une suite de variables aléatoires indépendantes (hypothèse de canal sans mémoire). Ce modèle est caractérisé par

- la distribution de probabilité initiale $\eta_0(dx)$ à l'instant 0,
- le noyau de probabilités de transition $Q_k(x, dx')$, c'est-à-dire que

$$Q_k(x, dx') = \mathbb{P}[X_k \in dx' \mid X_{k-1} = x] ,$$

pour tout $k = 1, \dots, n$,

- la fonction de vraisemblance $g_k(x')$ pour tout $k = 0, 1, \dots, n$.

Ici l'instant final n est fixé, on a accès à toutes les observations $Y_{0:n} = (Y_0, Y_1, \dots, Y_n)$, et on cherche à estimer l'état caché à un instant k intermédiaire entre l'instant initial 0 et l'instant final n . Il s'agit donc de calculer la distribution de probabilité conditionnelle de l'état caché X_k sachant toutes les observations $Y_{0:n}$, ou *lisseur* bayésien, définie par

$$\mu_{k|n}(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:n}] .$$

Si on a seulement accès aux observations $Y_{0:k} = (Y_0, Y_1, \dots, Y_k)$ pour estimer l'état caché à l'instant k , alors on peut juste calculer la distribution de probabilité conditionnelle de l'état caché X_k sachant les observations $Y_{0:k}$, ou *filtre* bayésien, définie par

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k}] .$$

Clairement

$$\mu_{n|n}(dx) = \mathbb{P}[X_n \in dx \mid Y_{0:n}] = \mu_n(dx) ,$$

c'est-à-dire que le lisseur bayésien est égal au filtre bayésien pour $k = n$.

NOTATIONS

L'intégrale d'une fonction intégrable ϕ par rapport à une distribution de probabilité μ est notée

$$\langle \mu, \phi \rangle = \int_E \phi(x) \mu(dx) .$$

Un noyau de probabilités de transition, ou noyau de transition, ou encore noyau markovien, est une collection $K(x, dx')$ indexée par $x \in E$ de distributions de probabilité.

Un noyau de transition agit (à droite) sur les fonctions, c'est-à-dire transforme une fonction ϕ en une autre fonction $K \phi$ définie par

$$K \phi(x) = \int_E K(x, dx') \phi(x') ,$$

pour tout $x \in E$, où l'intégrale porte sur la variable x' . Il agit aussi (à gauche) sur les distributions de probabilité, c'est-à-dire transforme une distribution de probabilité μ en une autre distribution de probabilité μK , vue comme mélange des distributions de probabilité $K(x, dx')$ par rapport à la distribution de mélange $\mu(dx)$, et définie par

$$\mu K(dx') = \int_E \mu(dx) K(x, dx') ,$$

où l'intégrale porte sur la variable x . On note l'identité

$$\langle \mu K, \phi \rangle = \langle \mu, K \phi \rangle ,$$

qui résulte du théorème de Fubini.

On rappelle [voir la planche #42 du troisième et dernier cours] la représentation probabiliste de la loi conditionnelle des états cachés $X_{0:k}$ sachant les observations $Y_{0:k}$

$$\mathbb{P}[X_{0:k} \in dx_{0:k} \mid Y_{0:k}] \propto \eta_0(dx_0) \prod_{p=1}^k Q_p(x_{p-1}, dx_p) \prod_{p=0}^k g_p(x_p) ,$$

à une constante multiplicative près, pour tout $k = 0, 1, \dots, n$. On rappelle aussi qu'en intégrant par rapport aux variables $x_{0:k-1} = (x_0, \dots, x_{k-1})$, on obtient

$$\mu_k(dx_k) = \mathbb{P}[X_k \in dx_k \mid Y_{0:k}] \propto g_k(x_k) \int_E \mu_{k-1}(dx_{k-1}) Q_k(x_{k-1}, dx_k) ,$$

de sorte que (en changeant le nom des variables, x' pour x_k et x pour x_{k-1}) on obtient

$$\mu_k(dx') \propto g_k(x') \int_E \mu_{k-1}(dx) Q_k(x, dx'),$$

c'est-à-dire que la suite $\{\mu_k\}$ vérifie l'équation récurrente dans le sens direct

$$\mu_k \propto g_k \eta_k \quad \text{avec} \quad \eta_k = \mu_{k-1} Q_k,$$

à une constante multiplicative près, pour tout $k = 1, \dots, n$, avec la condition initiale $\mu_0 \propto g_0 \eta_0$ pour $k = 0$. Concrètement, en explicitant la constante de normalisation, on obtient

$$\mu_k = \frac{g_k \eta_k}{\langle \eta_k, g_k \rangle} \quad \text{c'est-à-dire que} \quad \langle \mu_k, \phi \rangle = \frac{\langle \eta_k, g_k \phi \rangle}{\langle \eta_k, g_k \rangle}, \quad (1)$$

pour tout $k = 0, 1, \dots, n$ et pour toute fonction mesurable bornée ϕ .

ÉQUATIONS FORWARD-BACKWARD

On rappelle [voir la planche #42 du troisième et dernier cours] la représentation probabiliste de la loi conditionnelle des états cachés $X_{0:n}$ sachant les observations $Y_{0:n}$

$$\mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n}] \propto \eta_0(dx_0) \prod_{p=1}^n Q_p(x_{p-1}, dx_p) \prod_{p=0}^n g_p(x_p),$$

à une constante multiplicative près.

(i) **Montrer que**

$$\mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n}] \propto \mathbb{P}[X_{0:k} \in dx_{0:k} \mid Y_{0:k}] \prod_{p=k+1}^n Q_p(x_{p-1}, dx_p) \prod_{p=k+1}^n g_p(x_p),$$

à une constante multiplicative près, pour tout $k = 0, 1, \dots, n-1$.

Pour tout $k = 0, 1, \dots, n-1$, on pose

$$v_k(x_k) = \int_E \cdots \int_E \prod_{p=k+1}^n Q_p(x_{p-1}, dx_p) \prod_{p=k+1}^n g_p(x_p),$$

pour tout $x_k \in E$, où l'intégration porte sur les variables $x_{k+1:n} = (x_{k+1}, \dots, x_n)$. Pour $k = n$, on pose par convention $v_n(x_n) = 1$ pour tout $x_n \in E$.

- (ii) **En intégrant par rapport aux variables $(x_0, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$ (sauf la variable x_k) l'identité établie à la question (i), montrer que**

$$\mathbb{P}[X_k \in dx_k \mid Y_{0:n}] \propto \mathbb{P}[X_k \in dx_k \mid Y_{0:k}] v_k(x_k) ,$$

de sorte que (en changeant le nom de la variable, x pour x_k)

$$\mu_{k|n}(dx) \propto v_k(x) \mu_k(dx) ,$$

soit

$$\mu_{k|n} \propto v_k \mu_k ,$$

pour tout $k = 0, 1, \dots, n$, à une constante multiplicative près.

[Indice : Dans le membre de droite de l'identité établie à la question (i), intégrer d'abord par rapport aux variables $x_{k+1:n} = (x_{k+1}, \dots, x_n)$ puis par rapport aux variables $x_{0:k-1} = (x_0, \dots, x_{k-1})$.]

Concrètement, en explicitant la constante de normalisation, on obtient la relation suivante entre le lisseur bayésien et le filtre bayésien

$$\mu_{k|n} = \frac{v_k \mu_k}{\langle \mu_k, v_k \rangle} \quad \text{c'est-à-dire que} \quad \langle \mu_{k|n}, \phi \rangle = \frac{\langle \mu_k, v_k \phi \rangle}{\langle \mu_k, v_k \rangle} \quad (2)$$

pour tout $k = 0, 1, \dots, n$ et pour toute fonction mesurable bornée ϕ .

- (iii) **Montrer que la suite de fonctions $\{v_k\}$ vérifie l'équation récurrente dans le sens rétrograde**

$$v_{k-1} = Q_k(g_k v_k) , \quad (3)$$

pour tout $k = n, n-1, \dots, 1$, avec la condition initiale $v_n \equiv 1$ pour $k = n$.

- (iv) **Montrer que les constantes de normalisation dans (2) vérifient l'équation récurrente dans le sens rétrograde**

$$\langle \mu_{k-1}, v_{k-1} \rangle = \langle \eta_k, g_k \rangle \langle \mu_k, v_k \rangle , \quad (4)$$

pour tout $k = n, n-1, \dots, 1$, avec la condition initiale $\langle \mu_n, v_n \rangle = 1$ pour $k = n$.

Dans cette première approche, on résoud séparément

- l'équation récurrente (1) dans le sens direct pour le filtre bayésien μ_k ,
- et l'équation récurrente (3) dans le sens rétrograde pour la fonction v_k ,

et on obtient le lisseur bayésien $\mu_{k|n}$ grace à la relation (2).

ÉQUATION BACKWARD POUR LE LISSEUR BAYÉSIEN

L'objectif ici est de montrer qu'il est possible de combiner les équations (2) et (1) pour éliminer la fonction v_k et obtenir une équation récurrente dans le sens rétrograde pour le lisseur bayésien $\mu_{k|n}$.

Pour toute distribution de probabilité μ définie sur E , on considère le noyau de probabilités de transition $Q_k^B(\mu, x', dx)$ agissant dans le sens rétrograde, paramétré par la distribution de probabilité μ , et défini implicitement par

$$\mu(dx) Q_k(x, dx') = \mu Q_k(dx') Q_k^B(\mu, x', dx) ,$$

soit, en terme de densités de transition telles que $Q_k(x, dx') = q_k(x' | x) dx'$ si celles-ci existent

$$\mu(dx) q_k(x' | x) = \left[\int_E q_k(x' | x) \mu(dx) \right] Q_k^B(\mu, x', dx) ,$$

ce qui donne l'expression explicite

$$Q_k^B(\mu, x', dx) = \frac{q_k(x' | x) \mu(dx)}{\int_E q_k(x' | x) \mu(dx)} ,$$

pour tout $k = 1, \dots, n$. En particulier pour le filtre bayésien μ_{k-1} , on a

$$\mu_{k-1}(dx) Q_k(x, dx') = \mu_{k-1} Q_k(dx') Q_k^B(\mu_{k-1}, x', dx) ,$$

pour tout $k = 1, \dots, n$.

(v) **Montrer que**

$$\langle \mu_{k-1|n}, \phi \rangle = \langle \mu_{k|n}, Q_k^B(\mu_{k-1}) \phi \rangle ,$$

pour toute fonction mesurable bornée ϕ , de sorte que le lisseur bayésien vérifie l'équation récurrente dans le sens rétrograde

$$\mu_{k-1|n} = \mu_{k|n} Q_k^B(\mu_{k-1}) , \tag{5}$$

pour tout $k = n, n-1, \dots, 1$, avec la condition initiale $\mu_{n|n} = \mu_n$ pour $k = n$.

[Indice : Utiliser successivement la relation (2) à l'instant $(k-1)$, l'équation récurrente (3) dans le sens rétrograde, la définition du noyau de probabilités de transition $Q_k^B(\mu_{k-1})$, l'équation récurrente (1) dans le sens direct, la relation (4) pour les constantes de normalisation, et à nouveau la relation (2) à l'instant k .]

- (vi) **Au vu de l'équation (5), montrer que le lisseur bayésien $\mu_{k|n}$ peut s'interpréter comme la distribution de probabilité inconditionnelle de l'état X_k^B d'une chaîne de Markov rétrograde $\{X_k^B\}$, c'est-à-dire que**

$$\mu_{k|n}(dx) = \mathbb{P}[X_k^B \in dx] ,$$

pour tout $k = n, n-1, \dots, 0$.

Donner les caractéristiques de la chaîne de Markov $\{X_k^B\}$.

Dans cette deuxième approche, on résoud successivement

- d'abord l'équation récurrente (1) dans le sens direct pour le filtre bayésien μ_k ,
- puis l'équation récurrente (5) dans le sens rétrograde pour le lisseur bayésien $\mu_{k|n}$.

ÉQUATION FORWARD POUR LE LISSEUR BAYÉSIEN

L'objectif ici est de montrer qu'il est également possible de combiner les équations (2) et (3) pour éliminer la distribution de probabilité μ_k et obtenir une équation récurrente dans le sens direct pour le lisseur bayésien $\mu_{k|n}$.

Pour toute fonction positive v définie sur E , on considère le noyau de probabilités de transition $Q_k^F(v, x, dx')$ agissant dans le sens direct, paramétré par la fonction v , et défini explicitement par

$$Q_k^F(v, x, dx') = \frac{Q_k(x, dx') g_k(x') v(x')}{Q_k(g_k v)(x)} ,$$

pour tout $k = 1, \dots, n$.

- (vii) **En particulier pour la fonction v_k définie en (2), montrer que**

$$Q_k^F(v_k) \phi = \frac{Q_k(g_k v_k \phi)}{v_{k-1}} .$$

pour tout $k = 1, \dots, n$ et pour toute fonction mesurable bornée ϕ .

- (viii) **Montrer que**

$$\langle \mu_{k|n}, \phi \rangle = \langle \mu_{k-1|n}, Q_k^F(v_k) \phi \rangle ,$$

pour toute fonction mesurable bornée ϕ , de sorte que le lisseur bayésien vérifie l'équation récurrente dans le sens direct

$$\mu_{k|n} = \mu_{k-1|n} Q_k^F(v_k) , \tag{6}$$

pour tout $k = 1, \dots, n$, avec la condition initiale $\mu_{0|n} \propto (v_0 g_0) \eta_0$ pour $k = 0$.

[Indice : Utiliser successivement la relation (2) à l'instant k , l'équation récurrente (1) dans le sens direct, la définition du noyau de probabilités de transition $Q_k^F(v_k)$, à nouveau la relation (2) à l'instant $(k - 1)$, et la relation (4) pour les constantes de normalisation.]

(ix) **Au vu de l'équation (6), montrer que le lisseur bayésien $\mu_{k|n}$ peut s'interpréter comme la distribution de probabilité inconditionnelle de l'état X_k^F d'une chaîne de Markov $\{X_k^F\}$, c'est-à-dire que**

$$\mu_{k|n}(dx) = \mathbb{P}[X_k^F \in dx] ,$$

pour tout $k = 0, 1, \dots, n$.

Donner les caractéristiques de la chaîne de Markov $\{X_k^F\}$.

Dans cette troisième approche, on résoud successivement

- d'abord l'équation récurrente (3) dans le sens rétrograde pour la fonction v_k ,
- puis l'équation récurrente (6) dans le sens direct pour le lisseur bayésien $\mu_{k|n}$.