École Nationale Supérieure de Techniques Avancées
examen du cours SOD333
“Filtrage bayésien et approximation particulaire”

vendredi 23 octobre 2020, 13:30 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 $k = 1, \cdots, n$,
- la fonction de vraisemblance $g_k(x')$ pour tout $k = 0, 1, \cdots, n$.

Ici l'instant final n est fixé, 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 $Y_{0:n} = (Y_0, Y_1, \cdots, Y_n)$, définie par
$$\mu_{k|n}(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:n}] \ .$$

On rappelle que le filtre bayésien μ_k peut s'exprimer comme
$$\langle \mu_k, \phi \rangle = \frac{\langle \gamma_k, \phi \rangle}{\langle \gamma_k, 1 \rangle} \ ,$$
où la distribution non–normalisée γ_k est définie par
$$\langle \gamma_k, \phi \rangle = \mathbb{E}[\, \phi(X_k) \, \prod_{p=0}^{k} g_p(X_p) \,] \ ,$$
pour toute fonction mesurable bornée ϕ.

(i) **En s'inspirant de la preuve vue en cours pour le filtre bayésien, montrer que le lisseur bayésien $\mu_{k|n}$ peut s'exprimer comme**

$$\langle \mu_{k|n}, \phi \rangle = \frac{\langle \gamma_{k|n}, \phi \rangle}{\langle \gamma_{k|n}, 1 \rangle} \ ,$$

où la distribution non–normalisée $\gamma_{k|n}$ est définie par

$$\langle \gamma_{k|n}, \phi \rangle = \mathbb{E}[\, \phi(X_k) \prod_{p=0}^{n} g_p(X_p) \,] \ ,$$

pour toute fonction mesurable bornée ϕ.

Équations forward–backward

On introduit la fonction

$$v_k(x) = \mathbb{E}[\prod_{q=k+1}^{n} g_q(X_q) \mid X_k = x \,] \ , \tag{1}$$

pour tout $k = 0, 1, \cdots, n-1$, avec la convention $v_n \equiv 1$ pour $k = n$.

(ii) **En séparant explicitement les termes qui dépendent des états passés $X_{0:k}$ et les termes qui dépendent des états futurs $X_{k+1:n}$, montrer que**

$$\langle \gamma_{k|n}, \phi \rangle = \langle \gamma_k, v_k \, \phi \rangle \ ,$$

pour toute fonction mesurable bornée ϕ, et en déduire que

$$\gamma_{k|n} = v_k \, \gamma_k \qquad \textbf{et} \qquad \mu_{k|n} = \frac{v_k \, \mu_k}{\langle \mu_k, v_k \rangle} = v_k \cdot \mu_k \ , \tag{2}$$

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

Plus généralement, on introduit la fonction

$$f_k(x) = \mathbb{E}[\, \phi(X_n) \prod_{q=k+1}^{n} g_q(X_q) \mid X_k = x \,] \ , \tag{3}$$

pour tout $k = 0, 1, \cdots, n-1$, avec la convention $f_n = \phi$ pour $k = n$. Pour $x \in E$ fixé, cette expression peut s'interpréter comme une espérance pour le modèle de Markov caché caractérisé par

- la distribution de probabilité initiale δ_x à l'instant k,

- le noyau de probabilités de transition $Q_p(x,dx')$, pour tout $p=k+1,\cdots,n$,
- la fonction de vraisemblance $g_p(x')$ pour tout $p=k+1,\cdots,n$.

(iii) **Montrer que la suite de fonctions $\{f_k\}$ définie en (3) vérifie l'équation récurrente dans le sens rétrograde**

$$f_{k-1} = Q_k(g_k\, f_k)\ , \tag{4}$$

pour tout $k=n,n-1,\cdots,1$, avec la condition initiale $f_n=\phi$ pour $k=n$.

Dans le cas particulier $\phi\equiv 1$, en déduire que la suite de fonctions $\{v_k\}$ définie en (1) vérifie l'équation récurrente dans le sens rétrograde

$$v_{k-1} = Q_k(g_k\, v_k)\ , \tag{5}$$

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

On rappelle que la distribution de filtrage non–normalisée γ_k vérifie une équation récurrente dans le sens direct

$$\gamma_k = g_k\,(\gamma_{k-1}\, Q_k)\ , \tag{6}$$

pour tout $k=1,\cdots,n$, avec la condition initiale $\gamma_0 = g_0\,\eta_0$ pour $k=0$.

(iv) **Montrer que les équations (6) et (4) sont des équations duales, dans le sens où l'expression $\langle \gamma_k, f_k\rangle$, et en particulier l'expression $\langle \gamma_k, v_k\rangle$ obtenue pour $\phi\equiv 1$, ne dépend pas de $k=0,1,\cdots,n$.**

En déduire que la constante de normalisation $\langle \gamma_{k|n}, 1\rangle$ de la distribution non–normalisée $\gamma_{k|n}$ ne dépend pas de $k=0,1,\cdots,n$.

Dans cette première approche, on résoud séparément

- une équation récurrente dans le sens direct pour le filtre bayésien μ_k,
- une équation récurrente 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 (6) 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 $Q_k^{\mathrm{B}}(\mu, x', dx)$ agissant dans le sens rétrograde, paramétré par la distribution μ, et défini implicitement par

$$\mu(dx)\; Q_k(x, dx') = \mu\, Q_k(dx')\; Q_k^{\mathrm{B}}(\mu, x', dx) \;,$$

soit, en terme de densités de transition telles que $Q_k(x, dx') = q_k(x' \mid x)\; dx'$ si celles–ci existent

$$\mu(dx)\; q_k(x' \mid x) = [\int_E q_k(x' \mid x)\; \mu(dx)\,]\; Q_k^{\mathrm{B}}(\mu, x', dx) \;,$$

ce qui donne l'expression explicite

$$Q_k^{\mathrm{B}}(\mu, x', dx) = \frac{q_k(x' \mid x)\; \mu(dx)}{\displaystyle\int_E q_k(x' \mid x)\; \mu(dx)} \;,$$

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

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

et pour la distribution non–normalisée γ_{k-1}

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

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

(v) **Montrer que**

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

pour toute fonction mesurable bornée ϕ, de sorte que

$$\gamma_{k-1|n} = \gamma_{k|n}\; Q_k^{\mathrm{B}}(\mu_{k-1}) \;,$$

pour tout $k = n, n-1, \cdots, 1$, avec la condition initiale $\gamma_{n|n} = \gamma_n$ pour $k = n$.

En déduire que le lisseur bayésien vérifie l'équation de récurrence dans le sens rétrograde

$$\mu_{k-1|n} = \mu_{k|n}\; Q_k^{\mathrm{B}}(\mu_{k-1}) \;, \tag{7}$$

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

(vi) **Au vu de l'équation (7), montrer que le lisseur bayésien $\mu_{k|n}$ peut s'interpréter comme la distribution de probabilité inconditionnelle de l'état $X_k^{\rm B}$ d'une chaîne de Markov rétrograde $\{X_k^{\rm B}\}$, c'est–à–dire que**

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

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

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

Dans cette deuxième approche, on résoud

- d'abord une équation récurrente dans le sens direct pour le filtre bayésien μ_k,
- puis une équation récurrente 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 (5) 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 $Q_k^{\rm F}(v, x, dx')$ agissant dans le sens direct, paramétré par la fonction v, et défini explicitement par

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

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

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

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

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

(viii) **Montrer que**

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

pour toute fonction mesurable bornée ϕ, de sorte que

$$\mu_{k|n} = \mu_{k-1|n}\, Q_k^{\rm F}(v_k) \ , \tag{8}$$

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

(ix) **Au vu de l'équation (8), 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 $\{X_k^{\mathrm{F}}\}$, c'est–à–dire que**

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

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

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

Dans cette troisième approche, on résoud

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