ISSN 2078-3744. Вісник Львів. ун-ту. Серія мех.-мат. 2026. Випуск 98. С. 108–116
Visnyk of the Lviv Univ. Series Mech. Math. 2026. Issue 98. P. 108–116
http://publications.lnu.edu.ua/bulletins/index.php/mmf
doi: http://dx.doi.org/10.30970/vmm.2026.98.108-116

УДК 519.217

БАЙЄСІВСЬКЕ ВИСНОВУВАННЯ АГРЕГОВАНОГО ЯДРА ПЕРЕХОДІВ У ПРИХОВАНІЙ СУМІШІ МАРКОВСЬКИХ ЛАНЦЮГІВ

Андрій ДРЕБОТ

Львівський національний університет імені Івана Франка,
вул. Університетська, 1, м. Львів, 79000
e-mail: Andrii.Drebot.AMTS@lnu.edu.ua
ORCID: 0009-0005-2764-3334

У статті розглядається задача байєсівського висновування для агрегованої матриці переходів дискретного Марковського процесу, породженого сумішшю прихованих компонент. На відміну від більшості робіт, де основна увага приділяється відновленню прихованих режимів або компонентних матриць переходів, у даній роботі об'єктом дослідження є агреговане ядро переходів, яке є ідентифікованим функціоналом моделі навіть за неідентифікованості компонент. Для моделі з латентними незалежними індикаторами режимів, Діріхле апріором на ваги суміші та Діріхле апріорами на рядки компонентних матриць, запропоновано алгоритм який базується на методі варіаційного Байєса з середньопольовим наближенням, включно з повним виведенням ELBO (Evidence Lower Bound) та координатними оновленнями. Отримано аналітичні формули для перших двох моментів агрегованого ядра з варіаційним апостеріором та запропоновано компактну Діріхле проєкцію на основі методу моментного узгодження. Запропонований підхід забезпечує інтерпретоване подання апостеріорної невизначеності агрегованої динаміки без необхідності ідентифікації компонентів суміші.

Ключові слова: суміш Марковських процесів, розподіл Діріхле, Байєсівське висновування, середньопольова факторизація, розходження Кульбака-Лейблера, нижня межа правдоподібності.

Нехай $X_0, X_1, \ldots, X_T$, $X_t \in \{1, \ldots, K\}$, є дискретним часовим процесом, для якого спостерігається лише послідовність станів. Для кожного моменту часу $t = 0, \ldots, T-1$ вводиться латентна змінна $Z_t \in \{1, \ldots, M\}$, що визначає компоненту суміші, яка породжує перехід $X_t \to X_{t+1}$.

2020 Mathematics Subject Classification: 60J25

Модель задається такими співвідношеннями:

$$\pi \sim \mathrm{Dirichlet}(\gamma_1, \dots, \gamma_M), \tag{1}$$

$$\mathbf{p}_i^{(m)} \sim \mathrm{Dirichlet}(\alpha_{i1}^{(m)}, \dots, \alpha_{iK}^{(m)}),\ i = 1, \dots, K,\ m = 1, \dots, M, \tag{2}$$

$$Z_t \mid \pi \sim \mathrm{Categorical}(\pi), \quad t = 0, \dots, T-1, \tag{3}$$

$$X_{t+1} \mid X_t = i, Z_t = m \sim \mathrm{Categorical}(\mathbf{p}_i^{(m)}). \tag{4}$$

Тут $\pi = (\pi_1, \dots, \pi_M)$ є вектором ваг суміші, а $\mathbf{p}_i^{(m)} = (P_{i1}^{(m)}, \dots, P_{iK}^{(m)})$ — i-й рядок компонентної матриці переходів $P^{(m)}$.

Агрегованим ядром переходів назвемо матрицю

$$P^- := \sum_{m=1}^{M} \pi_m P^{(m)}.$$

Для фіксованого рядка i маємо $\mathbf{p}_i^- = \sum_{m=1}^{M} \pi_m \mathbf{p}_i^{(m)}$.

Оскільки Z_t є незалежними та однаково розподіленими за умовою π, спостережуваний процес $\{X_t\}$ є однорідним марковським ланцюгом з ядром P^-, тобто

$$P(X_{t+1} = j \mid X_t = i, \pi, P) = P_{ij}^-.$$

Наша ціль полягає в побудові апроксимації апостеріорного розподілу для $\mathbf{p}_i^-$, спираючись лише на спостережені дані $X_{0:T}$.

Позначимо через P усю сукупність параметрів $\{\mathbf{p}_i^{(m)}\}_{i,m}$, а через $Z = (Z_0, \dots, Z_{T-1})$ вектор прихованих індикаторів. Повний спільний розподіл має вигляд

$$p(X, \pi, P, Z) = p(\pi) \prod_{m=1}^{M} \prod_{i=1}^{K} p(\mathbf{p}_i^{(m)}) \prod_{t=0}^{T-1} p(Z_t \mid \pi) \prod_{t=0}^{T-1} p(X_{t+1} \mid X_t, Z_t, P). \tag{5}$$

Відповідно, логарифмічне перетворення розподілу дорівнює:

$$\log p(X, \pi, P, Z) =$$

$$= \log p(\pi) + \sum_{m=1}^{M} \sum_{i=1}^{K} \log p(\mathbf{p}_i^{(m)}) + \sum_{t=0}^{T-1} \log p(Z_t \mid \pi) + \sum_{t=0}^{T-1} \log p(X_{t+1} \mid X_t, Z_t, P). \tag{6}$$

Розпишемо кожен доданок окремо:

Для апріора $\pi \sim \mathrm{Dir}(\gamma)$

$$\log p(\pi) = \log \Gamma(\gamma_0) - \sum_{m=1}^{M} \log \Gamma(\gamma_m) + \sum_{m=1}^{M} (\gamma_m - 1) \log \pi_m, \qquad \gamma_0 = \sum_{m=1}^{M} \gamma_m. \tag{7}$$

Для апріора на рядок $\mathbf{p}_i^{(m)}$

$$\log p(\mathbf{p}_i^{(m)}) = \log \Gamma(\alpha_{i0}^{(m)}) - \sum_{j=1}^{K} \log \Gamma(\alpha_{ij}^{(m)}) + \sum_{j=1}^{K} (\alpha_{ij}^{(m)} - 1) \log P_{ij}^{(m)},\ \alpha_{i0}^{(m)} = \sum_{j=1}^{K} \alpha_{ij}^{(m)}. \tag{8}$$

Якщо ввести індикатор $z_{tm} \in \{0, 1\}$ — в момент часу t була активна компонента m; $\sum_m z_{tm} = 1$, то

$$\log p(Z_t \mid \pi) = \sum_{m=1}^{M} z_{tm} \log \pi_m. \tag{9}$$

Нарешті, якщо $X_t = i$, $X_{t+1} = j$, то

$$\log p(X_{t+1} \mid X_t, Z_t, P) = \sum_{m=1}^{M} z_{tm} \log P_{ij}^{(m)}. \tag{10}$$

Підставляючи (7)–(10) в (6), одержуємо логарифмічне перетворення розподілу у розгорнутій формі:

$$\begin{aligned} \log p(X, \pi, P, Z) = \sum_{m=1}^{M} (\gamma_m - 1) \log \pi_m + \sum_{m=1}^{M} \sum_{i=1}^{K} \sum_{j=1}^{K} (\alpha_{ij}^{(m)} - 1) \log P_{ij}^{(m)} \\ + \sum_{t=0}^{T-1} \sum_{m=1}^{M} z_{tm} \log \pi_m + \sum_{t=0}^{T-1} \sum_{m=1}^{M} z_{tm} \log P_{X_t, X_{t+1}}^{(m)} + \text{const}. \end{aligned} \tag{11}$$

Нехай Y має щільність $p_Y(y)$, а $X = f(Y)$ є вимірною функцією від Y. Тоді розподіл X називається *індукованим* розподілом і має формальний запис

$$p_X(x) = \int \delta(x - f(y))\, p_Y(y)\, dy.$$

У нашому випадку $Y = (\pi, P, Z)$, а функція $f(\pi, P, Z) = \sum_{m=1}^{M} \pi_m P^{(m)}$. Оскільки P^{-} не входить до моделі як первинний параметр, а визначається через (π, P), його апостеріорний розподіл не задається напряму, а індукується повним апостеріором $p(\pi, P, Z \mid X)$.

Апостеріорний розподіл агрегованого ядра має вигляд

$$p(P^{-} \mid X) = \int \delta\left(P^{-} - \sum_{m=1}^{M} \pi_m P^{(m)} \right) p(\pi, P, Z \mid X)\, d\pi\, dP\, dZ.$$

Оскільки $P^{-} = f(\pi, P)$ є вимірною функцією випадкових змінних (π, P), розподіл P^{-} є образом-мірою апостеріора $p(\pi, P, Z \mid X)$ відносно відображення f. Стандартний запис щільності через дельта-функцію дає наведену формулу. Інтегрування по Z еквівалентне маргіналізації по прихованих індикаторах, оскільки P^{-} від Z не залежить.

Загалом індукований апостеріорний розподіл $\mathbf{p}_i^{-} \mid X$ не належить до сімейства Діріхле.

Для фіксованого рядка $\mathbf{p}_i^{-} = \sum_{m=1}^{M} \pi_m \mathbf{p}_i^{(m)}$, навіть якщо кожен $\mathbf{p}_i^{(m)}$ має Діріхле апостеріор та π має Діріхле апостеріор, сімейство Діріхле не замкнене відносно випадкових опуклих комбінацій незалежних Діріхле векторів. Зокрема, коваріаційна структура $\mathbf{p}_i^{-}$ загалом не може бути представлена однією концентраційною константою β_{i0}, як це вимагає Діріхле розподіл. Отже, розподіл $\mathbf{p}_i^{-} \mid X$ у загальному випадку не є Діріхле.

Точний апостеріор $p(\pi, P, Z \mid X) = \frac{p(X,\pi,P,Z)}{p(X)}$ важко обчислити через сумування по всіх можливих комбінаціях Z. Тому вводимо варіаційне наближення $q(\pi, P, Z)$ і апроксимуємо апостеріор через мінімізацію розходження Кульбака-Лейблера(KL-divergence) $\mathrm{KL}\big(q(\pi, P, Z) \,\|\, p(\pi, P, Z \mid X)\big)$.

Відомо, що

$$\log p(X) = \mathcal{L}(q) + \mathrm{KL}\big(q(\pi, P, Z) \,\|\, p(\pi, P, Z \mid X)\big), \tag{12}$$

де

$$\mathcal{L}(q) = E_q[\log p(X, \pi, P, Z)] - E_q[\log q(\pi, P, Z)] \tag{13}$$

і є ELBO. Оскільки розходження Кульбака-Лейблера є невід'ємним, отримуємо

$$\mathcal{L}(q) \le \log p(X),$$

а тому максимізація ELBO еквівалентна мінімізації розходження Кульбака-Лейблера ([1, 3]).

Тепер застосуємо середньопольову факторизацію для Байєсівського висновування. Припускаємо повністю факторизовану форму

$$q(\pi, P, Z) = q(\pi) \prod_{m=1}^{M} \prod_{i=1}^{K} q(\mathbf{p}_i^{(m)}) \prod_{t=0}^{T-1} q(Z_t). \tag{14}$$

Далі позначимо:

$$q(\pi) = \mathrm{Dir}(\delta_1, \ldots, \delta_M),\ q(\mathbf{p}_i^{(m)}) = \mathrm{Dir}(\theta_{i1}^{(m)}, \ldots, \theta_{iK}^{(m)}),\ q(Z_t) = \mathrm{Cat}(r_{t,1}, \ldots, r_{t,M}).$$

Далі, виводимо загальні правила координатного оновлення. Для будь-якого фактора $q(\theta_i)$ оптимальне оновлення має вигляд

$$\log q^*(\theta_i) = E_{-\theta_i}[\log p(X, \pi, P, Z)] + \mathrm{const}, \tag{15}$$

де математичне очікування береться за всіма іншими варіаційними факторами [4, 3].

Опишемо детальне Оновлення $q(Z_t)$. Фіксуємо t та спостерігаємо перехід $X_t = i \to X_{t+1} = j$. Із повного розподілу (11) залежність від Z_t входить лише через два доданки

$$\sum_{m=1}^{M} z_{tm} \log \pi_m + \sum_{m=1}^{M} z_{tm} \log P_{ij}^{(m)}.$$

Тому за правилом (15)

$$\begin{aligned} \log q^*(Z_t = m) &= E_{q(\pi)q(P)}[\log \pi_m + \log P_{ij}^{(m)}] + \mathrm{const} = \\ &= E[\log \pi_m] + E[\log P_{ij}^{(m)}] + \mathrm{const}. \end{aligned} \tag{16}$$

Оскільки $q(Z_t) = \mathrm{Cat}(r_{t,1}, \ldots, r_{t,M})$, маємо

$$\log r_{t,m} \propto E[\log \pi_m] + E[\log P_{ij}^{(m)}].$$

Експоненціюючи та нормалізуючи, отримаємо форму:

$$r_{t,m} = \frac{\exp\{E[\log \pi_m] + E[\log P_{ij}^{(m)}]\}}{\sum\limits_{\ell=1}^{M} \exp\{E[\log \pi_\ell] + E[\log P_{ij}^{(\ell)}]\}}. \tag{17}$$

Залишилося обчислити очікування логарифмів величини яка розподілена за Діріхле. Нагадаємо властивість математичного сподівання таких величин.

Якщо

$$\pi \sim \mathrm{Dir}(\delta_1, \ldots, \delta_M),$$

то стандартна властивість експоненціального сімейства дає

$$E[\log \pi_m] = \psi(\delta_m) - \psi(\delta_0), \qquad \delta_0 = \sum_{m=1}^{M} \delta_m, \tag{18}$$

де $\psi(x) = \frac{d}{dx} \log \Gamma(x)$ — дигама-функція.

Аналогічно, якщо

$$\mathbf{p}_i^{(m)} \sim \mathrm{Dir}(\theta_{i1}^{(m)}, \ldots, \theta_{iK}^{(m)}),$$

то

$$E[\log P_{ij}^{(m)}] = \psi(\theta_{ij}^{(m)}) - \psi(\theta_{i0}^{(m)}), \qquad \theta_{i0}^{(m)} = \sum_{k=1}^{K} \theta_{ik}^{(m)}. \tag{19}$$

Підставляючи (18) та (19) в (17), отримуємо

$$r_{t,m} = \frac{\exp\Big(\psi(\delta_m) - \psi(\delta_0) + \psi(\theta_{ij}^{(m)}) - \psi(\theta_{i0}^{(m)})\Big)}{\sum\limits_{\ell=1}^{M} \exp\Big(\psi(\delta_\ell) - \psi(\delta_0) + \psi(\theta_{ij}^{(\ell)}) - \psi(\theta_{i0}^{(\ell)})\Big)}. \tag{20}$$

Щодо оновлення новлення $q(\pi)$, то залежність повного розподілу від π визначається членами

$$\log p(\pi) + \sum_{t=0}^{T-1} \log p(Z_t \mid \pi).$$

Підставляючи (7) та (9), отримаємо

$$\log p(\pi) + \sum_{t} \log p(Z_t \mid \pi) = \sum_{m=1}^{M} (\gamma_m - 1) \log \pi_m + \sum_{t=0}^{T-1} \sum_{m=1}^{M} z_{tm} \log \pi_m + \text{const}.$$

Беремо очікування за $q(Z)$. Оскільки $E[z_{tm}] = r_{t,m}$, маємо

$$\log q^*(\pi) = \sum_{m=1}^{M} \left(\gamma_m - 1 + \sum_{t=0}^{T-1} r_{t,m} \right) \log \pi_m + \text{const}. \tag{21}$$

Це знову форма Діріхле розподілу, отже

$$q(\pi) = \mathrm{Dir}(\delta_1, \ldots, \delta_M), \qquad \delta_m = \gamma_m + \sum_{t=0}^{T-1} r_{t,m}. \tag{22}$$

Нарешті, можна вивести оновлення $q(\mathbf{p}_i^{(m)})$. Фіксуємо компоненту m і рядок i. Із повного розподілу залежність від $\mathbf{p}_i^{(m)}$ задається членами

$$\log p(\mathbf{p}_i^{(m)}) + \sum_{t : X_t = i} \log p(X_{t+1} \mid X_t = i, Z_t, P).$$

Підставляючи (8) та (10), маємо

$$\log q^*(\mathbf{p}_i^{(m)}) = \sum_{j=1}^{K}(\alpha_{ij}^{(m)} - 1)\log P_{ij}^{(m)} + \sum_{j=1}^{K}\left(\sum_{t:X_t=i,X_{t+1}=j} r_{t,m}\right)\log P_{ij}^{(m)} + \text{const} =$$
$$= \sum_{j=1}^{K}\left(\alpha_{ij}^{(m)} - 1 + \sum_{t:X_t=i,X_{t+1}=j} r_{t,m}\right)\log P_{ij}^{(m)} + \text{const}. \tag{23}$$

Отже,

$$q(\mathbf{p}_i^{(m)}) = \mathrm{Dir}(\theta_{i1}^{(m)}, \dots, \theta_{iK}^{(m)}), \qquad \theta_{ij}^{(m)} = \alpha_{ij}^{(m)} + \sum_{t:X_t=i,X_{t+1}=j} r_{t,m}. \tag{24}$$

Тепер розглянемо аналітичні моменти агрегованого ядра. Після збору достатньої кількості даних та збіжності апостеріорів маємо

$$q(\pi) = \mathrm{Dir}(\delta), \qquad q(\mathbf{p}_i^{(m)}) = \mathrm{Dir}(\theta_i^{(m)}).$$

У рамках середньопольової факторизації (14) ці фактори незалежні, а отже моменти агрегованого ядра можна вивести аналітично.

Спочатку виведемо моменти для π. З властивостей розподілу Діріхле отримуємо

$$E[\pi_m] = \frac{\delta_m}{\delta_0}, \tag{25}$$
$$E[\pi_m^2] = \frac{\delta_m(\delta_m + 1)}{\delta_0(\delta_0 + 1)}, \tag{26}$$
$$E[\pi_m \pi_\ell] = \frac{\delta_m \delta_\ell}{\delta_0(\delta_0 + 1)}, \qquad m \neq \ell. \tag{27}$$

Аналогічно, моменти для компонентних рядків $q(\mathbf{p}_i^{(m)}) = \mathrm{Dir}(\theta_i^{(m)})$

$$E[P_{ij}^{(m)}] = \frac{\theta_{ij}^{(m)}}{\theta_{i0}^{(m)}}, \tag{28}$$
$$E[(P_{ij}^{(m)})^2] = \frac{\theta_{ij}^{(m)}(\theta_{ij}^{(m)} + 1)}{\theta_{i0}^{(m)}(\theta_{i0}^{(m)} + 1)}, \tag{29}$$
$$E[P_{ij}^{(m)} P_{ik}^{(m)}] = \frac{\theta_{ij}^{(m)}\theta_{ik}^{(m)}}{\theta_{i0}^{(m)}(\theta_{i0}^{(m)} + 1)}, \qquad j \neq k. \tag{30}$$

Теорема 1. *Для кожного i, j перший момент елемента агрегованого ядра під варіаційним апостеріором має вигляд*

$$\mu_{ij} := E[P_{ij}^{-}] = \sum_{m=1}^{M} E[\pi_m]E[P_{ij}^{(m)}] = \sum_{m=1}^{M} \frac{\delta_m}{\delta_0} \cdot \frac{\theta_{ij}^{(m)}}{\theta_{i0}^{(m)}}. \tag{31}$$

Доведення. За означенням $P_{ij}^{-} = \sum_{m=1}^{M} \pi_m P_{ij}^{(m)}$. Тоді за лінійністю математичного сподівання

$$E[P_{ij}^{-}] = \sum_{m=1}^{M} E[\pi_m P_{ij}^{(m)}].$$

Оскільки у середньопольовому наближенні π та $P^{(m)}$ незалежні,

$$E[\pi_m P_{ij}^{(m)}] = E[\pi_m] E[P_{ij}^{(m)}],$$

після чого використовуємо (25) та (28). □

Теорема 2. *Другий момент елемента агрегованого ядра має вигляд*

$$E[(P_{ij}^{-})^2] = \sum_{m=1}^{M} E[\pi_m^2] E[(P_{ij}^{(m)})^2] + 2 \sum_{m<\ell} E[\pi_m \pi_\ell] E[P_{ij}^{(m)}] E[P_{ij}^{(\ell)}]. \tag{32}$$

Звідси

$$\mathrm{Var}(P_{ij}^{-}) = E[(P_{ij}^{-})^2] - \mu_{ij}^2. \tag{33}$$

Доведення. Починаємо з $P_{ij}^{-} = \sum_m \pi_m P_{ij}^{(m)}$. Підносимо до квадрату

$$(P_{ij}^{-})^2 = \sum_m \pi_m^2 (P_{ij}^{(m)})^2 + 2 \sum_{m<\ell} \pi_m \pi_\ell P_{ij}^{(m)} P_{ij}^{(\ell)}.$$

Беремо математичне сподівання. За властивістю незалежності середньопольового наближення

$$E[\pi_m^2 (P_{ij}^{(m)})^2] = E[\pi_m^2] E[(P_{ij}^{(m)})^2], \quad E[\pi_m \pi_\ell P_{ij}^{(m)} P_{ij}^{(\ell)}] = E[\pi_m \pi_\ell] E[P_{ij}^{(m)}] E[P_{ij}^{(\ell)}].$$

Після підстановки отримуємо (32); формула (33) є означенням дисперсії. □

Аналогічно для $j \neq k$ маємо

$$\mathrm{Cov}(P_{ij}^{-}, P_{ik}^{-}) = \sum_{m=1}^{M} E[\pi_m^2] E[P_{ij}^{(m)} P_{ik}^{(m)}] + \sum_{m \neq \ell} E[\pi_m \pi_\ell] E[P_{ij}^{(m)}] E[P_{ik}^{(\ell)}] - \mu_{ij} \mu_{ik}. \tag{34}$$

Теорема 3. *Нехай для фіксованого рядка агрегованого ядра і відомі перші два моменти індукованого апостеріора:*

$$\mu_{ij} = E[P_{ij}^{-} \mid X], \qquad v_{ij} = \mathrm{Var}(P_{ij}^{-} \mid X), \quad j = 1, \ldots, K.$$

Тоді можливо побудувати наближення апостеріору:

$$\mathbf{p}_i^{-} \mid X \approx \mathrm{Dir}(\beta_{i1}, \ldots, \beta_{iK}),$$

де параметри β_i визначаються таким чином, щоб узгодити моменти.

Тоді Діріхле розподіл параметризується через середній вектор і концентрацію

$$\mu_{ij}^{\star} = \frac{\beta_{ij}}{\beta_{i0}}, \qquad \beta_{i0} = \sum_{j=1}^{K} \beta_{ij}, \tag{35}$$

$$\mathrm{Var}(p_{ij}) = \frac{\mu_{ij}^{\star}(1 - \mu_{ij}^{\star})}{\beta_{i0} + 1}. \tag{36}$$

Доведення. Середні значення повинні співпадати $\mu^{\star}_{ij} = \mu_{ij}$. З (35) отримуємо

$$\beta_{ij} = \beta_{i0}\mu_{ij}. \tag{37}$$

Отже, після узгодження першого моменту невідомим лишається лише параметр концентрації β_{i0}.

Підставляючи (37) у формулу дисперсії (36), одержимо $v_{ij} = \frac{\mu_{ij}(1-\mu_{ij})}{\beta_{i0}+1}$.

Звідси випливає оцінка концентрації

$$\beta^{(j)}_{i0} = \frac{\mu_{ij}(1-\mu_{ij})}{v_{ij}} - 1. \tag{38}$$

Було з'ясовано, що індукований апостеріорний розподіл $\mathbf{p}^{-}_{i} \mid X$ не є Діріхле, тому, всі оцінки $\beta^{(j)}_{i0}$ не є однаковими. Позначення $\beta^{(j)}_{i0}$ використовується для позначення координатно-специфічної оцінки параметра концентрації, отриманої з дисперсії j-ої компоненти, оскільки індукований апостеріор не є Діріхле і різні координати дають різні значення цієї оцінки. Це означає, що не існує єдиного параметра концентрації, який точно узгоджував би всі координатні дисперсії.

Отже, виникає задача побудови *єдиної узгодженої оцінки* β_{i0}.

Розглянемо задачу пошуку β_{i0}, яка мінімізує відхилення від локальних оцінок

$$L(\beta_{i0}) = \sum_{j\in\mathcal{J}} w_{ij}\big(\beta_{i0} - \beta^{(j)}_{i0}\big)^2, \tag{39}$$

де w_{ij} — ваги, що відображають інформативність координати j, а $\mathcal{J}$ — множина координат, що використовуються в оцінюванні.

Мінімізуючи (39), отримуємо $\frac{dL}{d\beta_{i0}} = 2\sum_{j\in\mathcal{J}} w_{ij}\big(\beta_{i0} - \beta^{(j)}_{i0}\big) = 0$.

Звідси безпосередньо випливає

$$\beta_{i0} = \frac{\sum\limits_{j\in\mathcal{J}} w_{ij}\beta^{(j)}_{i0}}{\sum\limits_{j\in\mathcal{J}} w_{ij}}. \tag{40}$$

Таким чином, агрегована оцінка є розв'язком задачі зваженого найменшоквадратичного узгодження.

Множину $\mathcal{J}$ доцільно обмежувати $\mathcal{J} = \{j : \varepsilon \le \mu_{ij} \le 1-\varepsilon\}$, де $\varepsilon > 0$ є малим числом. Це дозволяє уникнути числової нестабільності при $\mu_{ij} \approx 0$ або $\mu_{ij} \approx 1$.

Запропонована агрегація не відтворює всі координатні дисперсії одночасно, але забезпечує найкраще узгодження в сенсі зваженої квадратичної похибки.

Після визначення β_{i0} параметри Діріхле проєкції задаються як $\beta_{ij} = \beta_{i0}\mu_{ij}$, $j = 1,\dots,K$. □

У статті запропоновано байєсівське висновування агрегованої матриці переходів у латентній суміші марковських ланцюгів. Основна увага приділена не оцінюванню компонентних матриць або прихованих індикаторів, а безпосередньо агрегованому ядру, яке є ідентифікованим функціоналом моделі. Виведено повний варіативний Байєсівський алгоритм, показано, що індукований апостеріор агрегованого ядра загалом не є Діріхле, та отримано аналітичні формули для його перших двох моментів. На цій основі побудовано компактну Діріхле проєкцію, яка дає інтерпретоване представлення невизначеності агрегованої динаміки.

Список використаної літератури

1. Jordan M.I., Ghahramani Z., Jaakkola T.S., Saul L.K. *An introduction to variational methods for graphical models*, Machine Learning, **37** (1999), no.2, 183–233.
2. Wainwright M.J., Jordan M.I., *Graphical models, exponential families, and variational inference*, Foundations and Trends in Machine Learning, **1** (2008), no.1–2, 1–305.
3. Blei D.M., Kucukelbir A., McAuliffe J.D., *Variational inference: A review for statisticians*, Journal of the American Statistical Association, **112** (2017), no.518, 859–877.
4. Bishop C.M., Pattern recognition and machine learning, Springer, 2006.
5. Murphy K.P., Machine learning: a probabilistic perspective, MIT Press, 2012.

Стаття: вперше надійшла 23.07.2025
прийнята до друку 15.04.2026
опублікована 29.04.2026

BAYESIAN INFERENCE OF THE AGGREGATED KERNEL OF TRANSITIONS IN A HIDDEN MIXTURE OF MARKOV CHAINS

Andriy DREBOT

Ivan Franko Lviv National University,
Universytetska str., 1, L'viv, 79000
e-mail: Andrii.Drebot.AMTS@lnu.edu.ua
ORCID: 0009-0005-2764-3334

The paper considers the problem of Bayesian inference for the aggregated transition matrix of a discrete Markov process generated by a mixture of hidden components. Unlike most studies, where the main attention is paid to the recovery of hidden modes or component transition matrices, in this work the object of study is the aggregated transition kernel, which is an identified functional of the model even if the components are not identified. For a model with latent independent mode indicators, Dirichlet priors on the mixture weights and Dirichlet priors on the rows of the component matrices, an algorithm based on the variational Bayes method with a mean-field approximation is proposed, including full ELBO (Evidence Lower Bound) derivation and coordinate updates. Analytical formulas for the first two moments of the aggregated kernel with a variational posterior are obtained and a compact Dirichlet projection based on the moment matching method is proposed. The proposed approach provides an interpretable representation of the a posteriori uncertainty of the aggregated dynamics without the need to identify the mixture components.

Key words: mixture of Markov processes, Dirichlet distribution, Bayesian inference, mean-field factorization, Kullback-Leibler divergence, evidence lower bound.