ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 519.879

ПУАСОНІВСЬКІ ПРИХОВАНІ МАРКОВСЬКІ МОДЕЛІ

Баклан І.В., Степанкова Г.А.,
Національна академія управління, м. Київ

У статті досліджено властивості Пуасонівських прихованих марковських моделей (ППММ) та показано практичну можливість їх використання для прогнозування часових рядів, які мають марковські властивості.
Ключові слова: Пуасонівські приховані марковські моделі, марковські процеси, прогнозування часових рядів, функція (оцінка) правдоподібності, ймовірність.

Вступ. Приховані марковські моделі (ПММ) широко використовуються не тільки для таких класичних галузей, як розпізнавання образів, розпізнавання генетичних ланцюжків ДНК та інших. Нами розроблена модель використання ПММ для аналізу та прогнозування фінансово-економічних часових рядів [1, 2].

Постановка задачі. Пуасонівські приховані марковські моделі (ППММ) є різновидом ПММ. ППММ є дискретними часовими стохастичними процесами $\{(X_t, Y_t)\}_{t\in N}$, такими що $\{X_t\}_{t\in N}$ є недосліджуваною нескінченою множиною марковських ланцюжків та $\{Y_t\}_{t\in N}$ є досліджуваною послідовністю випадкових змінних, залежних від $\{X_t\}_{t\in N}$. Це змодельоване припущення про те, що умовний розподіл кожного досліджуваного Y_t задається послідовністю $\{X_t\}_{t\in N}$ і залежить виключно від одночасно недосліджуваних X_t. Більш того, задані $\{X_t\}_{t\in N}$ та $\{Y_t\}_{t\in N}$ є послідовностями умовно незалежних спадкових величин. Якщо ми припустимо, що для кожного t, Y_t, який задає множину X_t, є пуасонівською випадковою величиною, то будемо мати так звані ППММ. У цьому випадку X_t визначає паусонівський параметр, який використовується для генерації Y_t.

Нехай існують деякі зауваження та припущення. Припустимо недосліджуваний процес $\{X_t\}_{t\in N}$ є дискретним, однорідним, апереодичним і нескорочуваним марковським ланцюжком на скінченному просторі станів $S_X = \{1,2,...,m\}$ [3, 4]. Ми позначимо через $\lambda_{i,j}$ перехідну ймовірність від стану i на момент $t-1$ в стан j на момент t (для будь-яких станів i, j на будь-який момент t). Тобто $\lambda_{i,j} = P(X_t = j | X_{t-1} = i) = P(X_2 = j | X_1 = i)$.

Нехай $\Omega = \{\lambda_{i,j}\}$ буде $(m \times m)$ транзитивною матрицею ймовірностей з сумою всіх $\sum_{j\in S_X} \lambda_{i,j} = 1$ для будь-яких $i \in S_X$. Граничний розподіл X_1 є початковим розподілом і позначається $\delta = (\delta_1, \delta_2, ..., \delta_m)'$, де $\delta_i = P(X_1 = i)$ для будь-яких $i = 1,2,...,m$ та $\sum_{i \in S_X} \delta_i = 1$. Як безпосередня послідовність припущень на марковському ланцюжку $\{X_t\}_{t \in N}$ δ є стаціонарним розподілом та дорівнює $\delta' = \delta'\Omega$. Тобто δ є лівим власним вектором матриці Ω, пов′язаний з власним значенням 1, яке завжди існує, оскільки Ω є стохастичною матрицею [4].

Тепер розглянемо досліджувану послідовність $\{Y_t\}_{t \in N}$. У ППММ кожна досліджувана величина Y_t, обумовлена на X_t, є пуасонівською для будь-якого t. Якщо X_t знаходиться в стані i ($i \in S_X; t \in N$), тоді умовний розподіл Y_t є пуасонівською випадковою величиною з параметром γ_i для кожного $y \in N$. Залежні від стану ймовірності задаються

$$\pi_{y,i} = P(Y_t = y | X_t = i) = e^{-\gamma_i \frac{\gamma_i^y}{y!}}$$

При цьому $\sum_{y \in N} \pi_{y,i} = 1$ для кожного $i \in S_X$.

Оскільки $\{X_t\}_{t \in N}$ строго стаціонарний процес, то досліджуваний процес $\{Y_t\}_{t \in N}$ є строго стаціонарним. Тим більш Y_t для кожного t має граничний розподіл

$$P(Y_t = y) = \sum_{i \in S_X} P(Y_t = y, X_t = i) = \sum_{i \in S_X} P(Y_t = y | X_t = i) P(X_t = i) = \sum_{i \in S_X} \delta_i \gamma_i ,$$

який є скінченою сумішшю пуасонівських розподілів. Більш того, легко показати, що очікувана величина Y_t для кожного t задається

$$E(Y_t) = \sum_{i \in S_X} \delta_i \gamma_i .$$

Нарешті зауважимо, що величини Y_t є понадрозподіленими. Це вказує на те, що відхилення більше за середнє значення. Фактично, $V(Y_t) = \gamma' D \gamma + \delta' \gamma - (\delta' \gamma)^2 > E(Y_t) = \delta' \gamma$ для будь-якого t при $\gamma = (\gamma_1, ..., \gamma_m)'$ та $D = diag(\delta)$ [5].

Результати досліджень. ППММ була нами введена залежною від множини параметрів: початково-стаціонарного розподілу $\delta = (\delta_1, \delta_2, ..., \delta_m)'$, транзитивних ймовірностей $\lambda_{i,j}$ $(i, j \in S_X)$ та ймовірностей, залежних від стану $\pi_{y,i}$ $(y \in N, i \in S_X)$.

Зараз ми дамо деякі оцінки для цих параметрів. Зокрема, ми шукаємо максимально правдоподібні оцінки $m^2 - m$ перехідних ймовірностей $\lambda_{i,j}$ з $i \neq j$, тобто недіагональних елементів матриці Ω, і максимальних оцінувачів правдоподібності від m пуасонівських параметрів γ_i, що мають у собі станозалежні імовірності $\pi_{y,i}$. Використовуючи оцінену матрицю Ω, ми також задаємо оцінювач початкового розподілу δ від $\delta' = \delta'\Omega$. Позначимо через φ вектор невідомих параметрів, щоб оцінити їх вектором максимальної правдоподібності

$$\varphi = \left(\lambda_{1,2}, \lambda_{1,3}, \ldots, \lambda_{m,m-1}, \lambda_1, \ldots, \lambda_m\right)',$$

а через Φ позначимо простір параметрів.

Нехай $y = \left(y_1, \ldots, y_T\right)'$ вектор досліджуваних даних, тобто послідовність T реалізацій стохастичних процесів $\left\{Y_t\right\}_{t\in N}$, вектор y є неповним, оскільки послідовність станів ланцюжка $\left\{X_t\right\}_{t\in N}$ є відсутньою. Нехай $x = \left(i_1, \ldots, i_T\right)'$ буде вектор недосліджуваних станів ланцюжка $\left\{X_t\right\}_{t\in N}$. Таким чином $\left(i, y_1, \ldots, i_T, y_T\right)'$ є вектором повних даних.

Функція правдоподібності повних даних $L_T^c(\varphi)$ визначається як об’єднана ймовірність T досліджень і T недосліджуваних станів. Використовуючи марковську залежність, умовно-незалежні і одночасно залежні умови, ми легко отримуємо

$$L_T^c(\varphi) = P\left(Y_1 = y_1, \ldots, Y_T = y_T, X_1 = i_1, \ldots, X_T = i_T\right) = \delta_{i1} \pi_{y_1, i_1} \prod_{t=2}^{T} \lambda_{i_{t-1}, i_t} \pi_{y_t, i_t},$$

підсумовуючи $i_1, \ldots, i_T$ з обох боків, ми отримуємо функцію правдоподібності неповних даних:

$$L_T(\varphi) = P\left(Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_T\right) = \sum_{i_1 \in S_X} \sum_{i_2 \in S_X} \ldots \sum_{i_T \in S_X} \delta_{i1} \pi_{y_1, i_1} \prod_{t=2}^{T} \lambda_{i_{t-1}, i_t} \pi_{y_t, i_t},$$

де $\pi_{y,i}$ є станозалежною ймовірністю y_t, зумовленою на стані i_t $\left(t = 1, \ldots, T\right)$:

$$\pi_{y,i} = e^{-\gamma_{i_t}} \frac{\gamma_{i_t}^{y_t}}{y_t!} \tag{1}$$

Для того, щоб знайти оцінувач максимальної правдоподібності φ, ми розв’яжемо систему правдоподібності, але дуже тяжко знайти аналітичне розв’язання, тому ми змушені використовувати чисельний метод. Оскільки ми маємо справу з неповними даними, ми застосуємо EM-алгоритм [6], який базується на ітеративній процедурі з двома кроками на кожній ітерації: *перший крок*, E-крок, забезпечує очікування, *другий крок* забезпечує максимізацію.

Нехай $Q(\varphi; \varphi')$ функція визначена на E-кроці:

$$Q(\varphi; \varphi') = E_{\varphi'}\left(\ln L_T^c(\varphi) \middle| y\right),$$

для кожного заданого вектору φ', що має відношення до простору параметрів Φ.

Демпстер, Леад, Рубін [7] в 1977 році довели, що достатньою умовою для максимізації $L_T(\varphi)$ є максимізація $Q(\varphi;\varphi')$ по відношенню до φ. Особливо не деталізуючи, ітеративна схема ЕМ-алгоритму буде наступною.

Нехай $\varphi^{(k)}$ буде вектором оцінки, що отриманий на k-ій ітерації:

$$\varphi^{(k)} = \left(\lambda_{1,2}^{(k)}, \lambda_{1,3}^{(k)}, \ldots, \lambda_{m,m-1}^{(k)}, \lambda_1^{(k)}, \ldots, \lambda_m^{(k)}\right)',$$

а на $(k+1)$-ій ітерації Е та М кроки задаються наступним чином:

- Е-крок – дано $\varphi^{(k)}$, обчислюємо $Q(\varphi;\varphi^{(k)}) = E_{\varphi^{(k)}}\left(\ln L_T^c(\varphi) \middle| y\right)$;
- М-крок – шукаємо $\varphi^{(k+1)}$, яка максимізує $Q(\varphi;\varphi^{(k)})$.

Тобто,

$$Q\left(\varphi^{(k+1)};\varphi^{(k)}\right) \geq Q\left(\varphi;\varphi^{(k)}\right)$$

для будь-яких $\varphi \in \Phi$.

Кроки Е та М повинні повторюватися, доки послідовність величин правдоподібності $\left\{\ln L_T\left(\varphi^{(k)}\right)\right\}$ не зійдеться, тобто доки різниця

$$\ln L_T\left(\varphi^{(k+1)}\right) - \ln L_T\left(\varphi^{(k)}\right)$$

не стане меншою або рівною достатньо малій величині.

Коли деяка послідовність умов на просторі параметрів Φ та на функціях $L_T(\varphi)$ і $Q(\varphi;\varphi^{(k)})$ задовольняється, ми можемо сказати, що алгоритм сходиться на $(k+1)$-ій ітерації. Коли $\left(\varphi^{(k+1)};\ln L_T\left(\varphi^{(k+1)}\right)\right)$ є стаціонарною точкою, та $\varphi^{(k+1)} = \left(\lambda_{1,2}^{(k+1)}, \lambda_{1,3}^{(k+1)}, \ldots, \lambda_{m,m-1}^{(k+1)}, \lambda_1^{(k+1)}, \ldots, \lambda_m^{(k+1)}\right)'$ є оцінкою максимальної правдоподібності невідомого параметру φ.

Для ПММ поверхня правдоподібності є нерівномірною та характеризується багатьма локальними максимумами або стаціонарними точками. Зрозуміло, що стаціонарна точка, для якої ЕМ-алгоритм сходиться, може і не бути глобальним максимумом. Таким чином, для того щоб визначити глобальний максимум, вибір початкової точки є визначальним.

Застосування алгоритму для пошуку оцінки невідомих параметрів з ЕМ-алгоритмом може бути спрощено за допомогою використання ймовірностей «вперед» та «назад», які були введені Баумом у 1970 році [8]. Імовірність «вперед», яка позначається $\alpha_t(i)$, є об'єднаною імовірністю попереднього та даного дослідження. Та даний стан ланцюжка такий

$$\alpha_t(i) = P(Y_1 = y_1, \ldots, Y_t = y_t, X_t = i).$$

У той час, як зворотна імовірність (імовірність «назад»), яка позначається $\beta_t(i)$, є імовірністю майбутніх досліджень, зумовлених теперішнім станом ланцюжка

$$\beta_t(i) = P(Y_{t+1} = y_{t+1},\ldots,Y_T = y_T \mid X_t = i).$$

Імовірності $\alpha_t(i)$ та $\beta_t(i)$ можуть бути рекурсивно обчислені наступним чином:

$$\alpha_1(i) = \delta_i \pi_{y_1,i}, \text{ де } i = 1,2,\ldots,m \tag{2}$$

$$\alpha_t(j) = \left(\sum_{i \in S_X} \alpha_{t-1}(i)\lambda_{i,j}\right)\pi_{y_t,j}, \text{ де } t = 2,\ldots,T \text{ та } j = 1,2,\ldots,m$$

для імовірностей «вперед», та

$$\beta_T(i) = 1, \text{ де } i = 1,2,\ldots,m \tag{3}$$

$$\beta_t(i) = \sum_{j \in S_X} \pi_{y_{t+1},j}\beta_{t+1}(j)\lambda_{i,j}, \text{ де } t = T-1,\ldots,1 \text{ та } i = 1,2,\ldots,m$$

для ймовірностей «назад» [5].

Потім ми отримаємо наступні вирази для функцій $Q(\varphi;\varphi^{(k)})$ на Е-кроці $(k+1)$-ої ітерації ЕМ-алгоритму [9]

$$Q(\varphi;\varphi^{(k)}) = E_{\varphi^{(k)}}(\ln L_T^c(\varphi)|y) = \sum_{i \in S_X} \frac{\alpha_1^{(k)}(i)\beta_1^{(k)}(i)}{\sum_{l \in S_X}\alpha_t^{(k)}(l)\beta_t^{(k)}(l)} \ln \delta_i +$$

$$+\sum_{i \in S_X}\sum_{j \in S_X} \frac{\sum_{t=1}^{T-1}\alpha_t^{(k)}(i)\lambda_{i,j}^{(k)}\pi_{y_{t+1},j}^{(k)}\beta_{t+1}^{(k)}(j)}{\sum_{l \in S_X}\alpha_t^{(k)}(l)\beta_t^{(k)}(l)} \ln \lambda_{i,j} + \sum_{i \in S_X} \frac{\sum_{t=1}^{T}\alpha_t^{(k)}(i)\beta_t^{(k)}(i)}{\sum_{l \in S_X}\alpha_t^{(k)}(l)\beta_t^{(k)}(l)} \ln \pi_{y_t,i}, \tag{4}$$

де $\pi_{y_t,i}^{(k)}, \alpha_t^{(k)}(i)$ та $\beta_t^{(k)}(i)$ обчислені згідно з формулами (1), (2) та (3) відповідно, використовуючи значення параметру $\varphi^{(k)}$, який обчислений у k-тій ітерації. У той час, як $\delta^{(k)}$ обчислена як $\delta'^{(k)} = \delta'^{(k)}\Omega^{(k)}$.

Слід зауважити, що δ, стаціонарне припущення, містить дані про матрицю транзитивних ймовірностей Ω, оскільки $\delta_j = \sum_{i \in S_X}\delta_i\lambda_{i,j}$ для кожного $j \in S_X$. Тим не менш, для великих T ефект δ є незначним [10]. Тим паче, на M кроці $(k+1)$-ої ітерації для того, щоб отримати $\varphi^{(k+1)}$, ми можемо проігнорувати перший додаток у формулі (4), коли максимізується $Q(\varphi;\varphi^{(k)})$ у відношенні до $m^2 - m$ параметрів $\lambda_{i,j}$.

Вираз для оцінки максимальної правдоподібності $\lambda_{i,j}$ отриманий на $(k+1)$-ій ітерації ЕМ-алгоритму, задається [9]

$$\lambda_{i,j}^{(k+1)} = \frac{\sum_{t=1}^{T-1}\alpha_t^{(k)}(i)\lambda_{i,j}^{(k)}\pi_{y_{t+1},j}^{(k)}\beta_{t+1}^{(k)}(j)}{\sum_{t=1}^{T-1}\alpha_t^{(k)}(i)\beta_t^{(k)}(i)} \tag{5}$$

для кожного стану i та кожного стану j, $j \neq i$ марковського ланцюжка $\{X_t\}$.

Оцінювач максимальної правдоподібності γ_i, отриманий на $(k+1)$-ій ітерації ЕМ-алгоритму, задається співвідношенням

$$\gamma_i^{(k+1)} = \frac{\sum_{t=1}^{T} \alpha_t^{(k)}(i)\beta_t^{(k)}(i)y_t}{\sum_{t=1}^{T} \alpha_t^{(k)}(i)\beta_t^{(k)}(i)} \tag{6}$$

для будь-якого стану i марковського ланцюжка $\{X_t\}$.

Леруа та Бікель, Рітов, Райден [11-15] довели, що оцінювачі в формулах (5) та (6) несуперечливі та асимптотично нормальні.

Висновки. Досліджені нами властивості ППММ показують практичну можливість їх використання для прогнозування часових рядів, які мають марковські властивості.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Баклан І.В., Степанкова Г.А. Імовірнісні моделі для аналізу та прогнозування часових рядів // Искусственный интеллект. – 2008. – № 3. – С. 505-515.

2. Баклан І.В., Степанкова Г.А. Основні проблеми при застосуванні прихованих марковських моделей: матеріали міжнародної наукової конференції [«Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту»]. – Том 2. – Херсон: ХНТУ, 2009. – С. 430-432.

3. Grimmett G.R. and Stirzaker D.R. *Probability and Random Processes.* – 2nd edition. – Oxford : Clarendon Press; New York : Oxford University Press, – Oxford science publications. – 1992. – 541 p.

4. Guttorp P. *Stochastic Modeling for Scientific Data.* – Chapman & Hall, London. – 1995. – 372 p.

5. MacDonald I.L. and Zucchini W. *Hidden Markov and Other Models for Discrete-valued Time Series.* – Chapman & Hall, London, 1997. – P. 60, 70.

6. McLachlan, G.J. and Krishnan T. *The EM algorithm and extensions.* – New York; John Wiley & Sons, Inc., 1997. – 274 p.

7. Dempster A.P., Laird N.M., Rubin D.B. Maximum likelihood from incomplete data via. the EM algorithm (with Discussion) // *Journal* of *the Royal Statistical Society. Series B,* 39. – 1977. – P. 1-38.

8. Baum L.E., Petrie T., Soules G., Weiss N. A maximization technique occurring in the statistical estimation for probabilistic functions of Markov chains // *The Annals of Mathematical Statistics.* – Vol. 41, no. 1. – 1970. – P. 164-171.

9. Spezia L. *Stima dei parametri di un modello markoviano binomiale negative parzialmente osservato.* Tesi di dottorato, Universith degli Studi di Trento. – 1999. – P. 64-66, 70-75.

10. Basawa I.V. and Prakasa Rao B.L.S. *Statistical Inference for Stochastic Processes.* – Academic Press, London, 1980. – P. 53-54.

11. Le N.D., Leroux B.G., Puterman M.L. Reader Reaction: Exact Likelihood Evaluation in a Markov Mixture Model for Time Series of Seizure Counts. – *Biometrics,* 48, 1992. – P. 317-323, 19.

12. Leroux B.G. Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and their Applications,* 40, 1992. – P. 127-143.

13. Leroux B.G. and Puterman M.L. Maximum-Penalized-Likelihood Estimation for Independent, and Markov-Dependent, Mixture Models. *Biometrics,* 48, 1992. – P. 545-558.

14. Ryden, T. (1999). Likelihood inference in ergodic hidden Markov models: a unified approach, implications and future directions. *Abstracts* of *Second European Conference on Highly Structured Stochastic System.*

http:// www.unipv.it/hsss99/abstracts/tobias.ps.

15. Giudici P., Ryden T., Vandekerkhove P. (1998). *Likelihood ratio tests for hidden Markov models.* Technical report, 1998: 19, Lund University, Sweden.

Баклан И.В., Степанкова А.А. ПУАСОНОВСКИЕ СКРЫТЫЕ МАРКОВСКИЕ МОДЕЛИ

В статье исследованы свойства Пуасоновских скрытых марковских моделей (ППММ) и показана практическая возможность их применения для прогнозирования временных рядов, которые имеют марковские свойства.

Ключевые слова: Пуасоновские скрытые марковские модели, марковские процессы, прогнозирование временных рядов, функция (оценка) правдоподобности, вероятность.

Baklan I.V., Stepankova A.A. POISSON’S HIDDEN MARKOV’S MODELS

Properties of Poisson’s hidden Markov’s models (PHMM) were studied in the paper and practical possibility of their application for prediction of time series that have Markov’s properties was shown.

Key words: Poisson’s hidden Markov’s models, Markov’s processes, prediction of time series, function (estimation) of plausibility, probability.