

УДК 57.087

**Юрчук Анастасія Олександрівна**

бакалавр прикладної математики,  
Національний технічний університет України  
«Київський політехнічний інститут»

**Юрчук Анастасия Александровна**

бакалавр прикладной математики,  
Национальный технический университет Украины  
«Киевский политехнический институт»

**Yurchuk A.**

Bachelor of applied mathematics  
The National Technical University of Ukraine  
«Kyiv Polytechnic Institute»

**ПРОГНОЗУВАННЯ РЕСПІРАТОРНИХ ВІРУСНИХ ЗАХВОРЮВАНЬ З  
ВИКОРИСТАННЯМ ПРИХОВАНИХ МАРКІВСЬКИХ ЛАНЦЮГІВ  
ПРОГНОЗИРОВАНИЕ РЕСПІРАТОРНЫХ ВИРУСНЫХ  
ЗАБОЛЕВАНИЙ С ИСПОЛЬЗОВАНИЕМ СКРЫТЫХ МАРКОВСКИХ  
ЦЕПЕЙ  
PREDICTION RESPIRATORY VIRAL DISEASES USING HIDDEN  
MARKOV CHAINS**

**Анотація:** Досліджено метод прогнозування респіраторних вірусних захворювань з використанням прихованих марківських ланцюгів.

**Ключові слова:** марківський процес, прогнозування, прихована марківська модель, метод Сьорфлінга, теорема Баєса.

**Аннотация:** Исследован метод прогнозирования респираторных вирусных заболеваний с использованием скрытых марковских цепей.

**Ключевые слова:** марковский процесс, прогнозирования, скрытая марковская модель, метод Сёрфлинга, теорема Байеса.

**Abstract:** The method of predicting respiratory viral diseases using Hidden Markov chains.

**Keywords:** Markov process, forecasting, Hidden Markov Model, Serfing's method, Bayes' theorem.

Гострі респіраторні захворювання утворюють найбільш поширену групу патологій, яка в структурі інфекційних захворювань займає 95-97% кількості хворих. Математичне моделювання та статистичний аналіз дозволяють досліджувати сезонні захворювання і дають розуміння характеру циркуляції вірусу на глобальному і регіональному масштабах [1, с. 14].

Для прогнозування респіраторних вірусних захворювань використовують метод Сьорфлінга, оскільки захворюваність на грип та ГРВІ має сезонний характер, а формула Сьорфлінга описує синусоїду з заданими параметрами. Для пошуку параметрів для даного методу було вирішено використати приховані марківські моделі.

Останні роботи в [2, с. 3] запропонували використовувати ПММ для часових рядів показників грипу в епідемічних і неепідемічних фазах. Є дві переваги в цьому підході. Перша перевага полягає в тому, що метод може бути застосований до історичних даних без необхідності проведення різниці між епідемічним і неепідемічним періодом в даних, тим самим дозволяючи розробку автоматизованої системи спостереження за грипом. Друга перевага полягає в тому, що спостереження повинні бути незалежними дані знання про епідемію, в той час як метод Сьорфлінга передбачає граничну незалежність даних.

Для захворювань, що мають епідемічний характер метод Сьорфлінга застосовується в 2 етапи. На першому етапі визначається базисні дані, що описують очікуваний шаблон історичних даних захворюваності. Так як

базисні дані передбачають не епідемічну фазу, періоди, що характеризуються підвищеною кількістю хворих повинні бути виключені, щоб уникнути завищеної оцінки базисних даних.

Основною проблемою є визначення періодів епідемії, для цього в роботі використано ПММ. На другому етапі, знаючи які періоди були епідемічними можна визначити модель часового ряду для прогнозування очікуваної кількості хворих в наступних періодах. В результаті отримаємо 2 рівняння, що описують часовий ряд в епідемічні та неепідемічні періоди:

$$y_{j,t} = \mu_j + \theta_j t + \alpha_j \sin\left(\frac{2\pi t}{r}\right) + \beta_j \cos\left(\frac{2\pi t}{r}\right),$$

де  $j = 0..1$  (0- відповідає неепідемічному періоду, 1 – епідемічному).

Отже, модель для прогнозування даних буде описана, як:

$$y_t = (1 - Z)y_{0t} + Zy_{1t},$$

де  $Z = 0$  – в неепідемічний період, 1 – в епідемічний [3, с. 56-59].

Для застосування методу Сьорфлінга нам необхідно визначити апріорні розподіли його параметрів. В літературі по методу Сьорфлінга пропонується використовувати нормальний розподіл для базового значення кількості хворих без урахування сезонних трендів в обох станах [4, с. 340], але враховуючи особливість респіраторних вірусних захворювань, а саме стрімке збільшення хворих за короткий період в даній роботі пропонується використати для епідемічного періоду експоненціальний розподіл.

Враховавши рекомендації Вінклера, щодо вибору розподілів Для інших параметрів методу Сьорфлінга обрано нормальний розподіл.

$$\mu_0 \sim \text{dnorm}(0.0, 1.0E - 6),$$

$$\mu_1 \sim \text{dexp}(0.0),$$

$$\theta_j \sim \text{dnorm}(0.0, 1.0E - 6),$$

$$\alpha_j \sim \text{dnorm}(0.0, 1.0E - 6),$$

$$\beta_j \sim \text{dnorm}(0.0, 1.0E - 6),$$

де  $dnorm(\mu, \sigma)$  представляє нормальний розподіл з математичним очікуванням  $\mu$  і дисперсією  $\sigma^2$ ;  $dexp(\lambda)$  з математичним очікуванням  $\frac{1}{\lambda}$ .

Для розв'язання поставленої задачі використовується ПММ з двома станами, тобто  $S_t$  може приймати значення 0, якщо  $t$  не період епідемії і 1 –, якщо навпаки. Рис. 1. ілюструє ПММ для поставленої задачі. При баєсовому підході передбачається, що параметри слідує апіорному розподілу, а потім вони оновлюються за допомогою методів Монте-Карло ланцюга Маркова (МКЛМ). Після оцінки параметрів, розшифровуються найбільш імовірна послідовність прихованих станів.

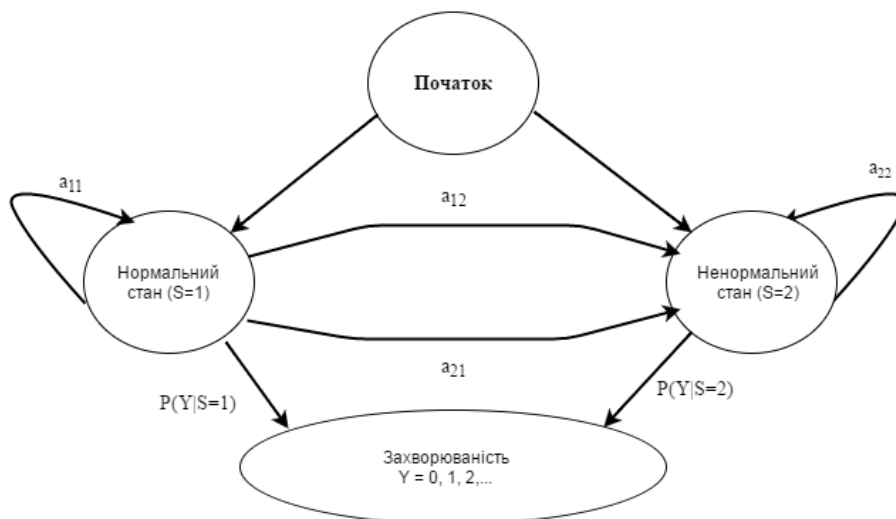


Рис. 1. Опис стану моделі при переході зі стану в стан

Основним методом МКЛМ являється семпсування Гіббса. Визначається модель у вигляді повного спільного розподілу всіх величин, будь то параметри або спостережувані величини, невідомі параметри семплуються по їхнім постеріорних розподілах в кожній вершині.

Алгоритм на кожному кроці бере одну випадкову величину і вибирає її значення за умови фіксованих інших [5, с. 30 ].

Для послідовності прихованих станів  $z_j$  обрано розподіл Бернуллі, оскільки, можливі тільки два значення цього параметру (0 і 1). Бета-розподіл

широко використовується в байєсівській статистиці, так як він є спряженим апіорним розподілом для розподілу Бернуллі [6, с.799].

Отже, для ПММ обрано наступні розподіли:

$$z_j \sim dbern(0.5),$$

$$p_{j,1:K} \sim dbeta(1, 1),$$

де  $z_j$  – послідовність прихованих станів  $dbern(p)$  розподіл Бернуллі з математичним очікуванням  $p$ ;  $p_{j,1:K}$  матриця переходу зі стану  $j$  в стан  $1:K$ ,  $dbeta(\alpha, \beta)$  бета-розподіл з математичним очікуванням  $\frac{\alpha}{\alpha+\beta}$  та дисперсією  $\frac{\alpha\beta}{(\alpha+\beta)^2+(\alpha+\beta+1)}$ .

Отже, для прогнозування кількості захворювань на ГРВІ та грип було обрано метод Сьорфлінга. Оскільки, ГРВІ та грип мають епідемічний характер, то метод Сьорфлінга застосовується в 2 етапа. Для визначення періодів епідемій використовується алгоритм Вітербі для прихованих марківських ланцюгів. Обчислення параметрів формули Сьорфлінга відбувається з використанням баєсового підходу до ПММ. В результаті досліджень було запропоновано математичні моделі для прогнозування захворюваності та оптимізації профілактики респіраторних вірусних інфекцій на основі прихованих марківських моделей з баєсовим підходом.

В якості вихідних даних використано дані спостережень захворюваності населення респіраторними вірусними інфекціями за 1994-2009 роки Національної медичної академії післядипломної освіти ім. П.Л. Шупика.

Було створено баєсову модель, ініціалізовано початкові значення параметрів. За допомогою ПММ визначено періоди епідемій. Епідемічними було названо 154 місяців, а неепідемічними відповідно було визнано – 50 місяці.

Обраховані значення кількості захворювань на ГРВІ згідно обчисленої моделі Сьорфлінга зображені на рис. 3.8. у вигляді графіка.

При цьому було отримано наступні значення параметрів формули Сьорфлінга:

$$y_0 = 434902,2 - 138,1t + 165413,3 \sin\left(\frac{2\pi t}{r}\right) + 81362,8 \cos\left(\frac{2\pi t}{r}\right),$$

$$y_1 = 551608,4 - 6,5t + 346764,8 \sin\left(\frac{2\pi t}{r}\right) + 157977,9 \cos\left(\frac{2\pi t}{r}\right),$$

де перше рівняння описує не епідеміологічний період, друге - епідемічний.

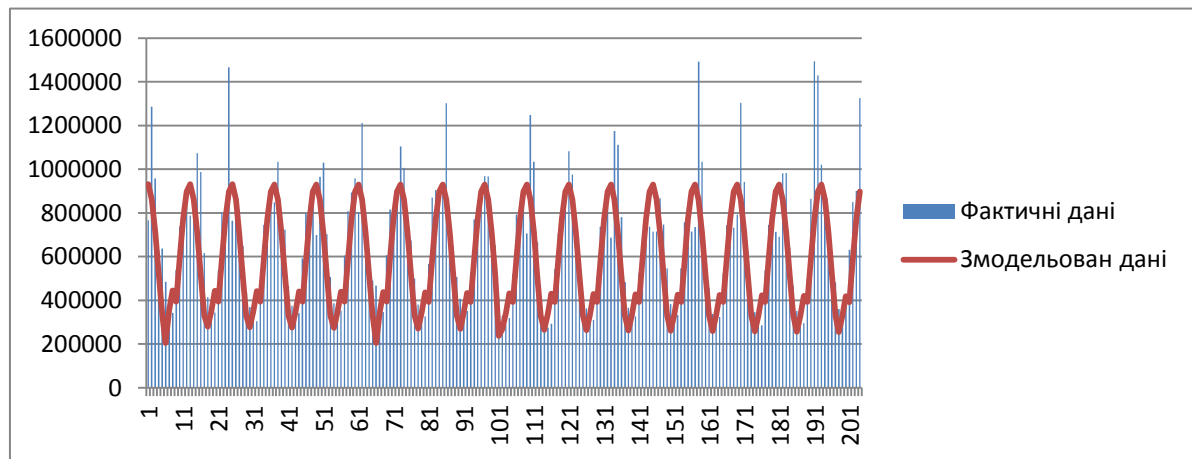


Рис. 2. Графік змодельованих і фактичних даних захворюваності ГРВІ за 1994-2010 роки.

З використанням експериментальних даних захворюваності на грип та ГРВІ було визначено параметри такої моделі з відносною похибкою для ГРВІ - 0,2234, що вказує на адекватність такої моделі. В той же час для грипу модель показала відносну похибку - 0,742652, що пояснюється розбіжністю вихідних експериментальних даних.

#### Література:

1. Simonsen L, Reichert TA, Viboud C, Blackwelder WC, Taylor RJ, Miller MA - Impact of influenza vaccination on seasonal mortality in the US elderly population. Arch Intern Med -2005- P.17 – 234 с.
2. Y. LeStrat and F. Carrat. Monitoring epidemiologic surveillance data using hidden Markov models. Statistics in Medicine, 2004. – 31 с.

3. R. E. Serfling. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Reports*, 1963. - 494–506 c.
4. S. Scott. *Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century*, 2002. - 337-351 c.
5. Juang, B. H., Rabiner, L. R. Hidden Markov models for speech recognition, *Technometrics*, 1991. – 49 c.
6. Winkler, R.L. The Assessment of Prior Distributions in Bayesian Analysis: *Journal of the American Statistical Association* 1967. – 821 c.