

О.В. БОГДАНОВ

Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна.

П.С. КНОПОВ

Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,

e-mail: *knopov1@yahoo.com*.

СТОХАСТИЧНІ МОДЕЛІ В ЗАДАЧАХ ПРОГНОЗУВАННЯ ЕПІДЕМІОЛОГІЧНОЇ СИТУАЦІЇ¹

Анотація. У роботі досліджуються деякі стохастичні моделі з дискретним та неперервним часом для розв'язання важливих задач прогнозування поширення епідеміологічних захворювань серед населення. Враховуються різні фактори поширення епідемій та основні параметри, що впливають на оцінювання прогнозу. Проведено тестові розрахунки, які спираються на запропоновані методи.

Ключові слова: методи, оптимізація, моделювання, стохастичні рівняння, оцінювання, епідемія, дискретний та неперервний час.

ВСТУП

Пандемія COVID-19 стала глобальним викликом для людства у XXI столітті. Вона потребує адекватних методів та засобів її контролю. За відсутності у населення імунітету та ліків від коронавірусу, а також нерівномірного доступу до вакцин епідемія загрожує життю і здоров'ю людей. Але в той же час введення довготермінового карантину та заходів для обмеження пандемії спричиняє економічні збитки, гальмуючи економічний розвиток країн. Тому рішення стосовно контролю за поширенням хвороби потребують особливої виваженості: адже, з одного боку, йдеться про життя і здоров'я великої кількості людей, з іншого — мають місце значні економічні втрати та потенційне зuboжіння населення.

За цих умов зростає необхідність у засобах моделювання та підтримки прийняття рішень, що базуються на точних обрахунках їхніх наслідків. До таких засобів належать різноманітні моделі прогнозування епідеміологічної ситуації і потреб медичної допомоги населенню, моделі прогнозування економічних наслідків урядових (регіональних) рішень для обмеження епідемії тощо. Також потрібно враховувати різноманітні ризики та невизначеності, що виникають під час моделювання таких складних процесів зі стохастичною (невизначеною) природою їхніх компонент. Це потребує відповідних математичних методів, зокрема використання випадкових процесів і полів, стохастичних диференціальних рівнянь, регресій спеціального вигляду, сучасного апарату мір ризику тощо. У подальшому викладено деякі підходи до розв'язання наведених вище проблем.

Початковими в цьому тексті вибрано SIR (SEIR) та подібні епідеміологічні моделі, що дають змогу прогнозувати вплив на динаміку процесу поширення хвороби, обмежувальних заходів для його контролю. У цих моделях головним фактором є коефіцієнт реплікації вірусу (репродуктивний коефіцієнт), що суттєво залежить від таких заходів. У роботі [1] досліджено вплив на щоденну швидкість поширення захворювання таких факторів як закриття шкіл, робочих місць, відміна публічних заходів, заборона роботи громадського транспорту, обмеження внутрішніх та міжнародних переміщень.

¹Робота виконана за часткової підтримки Національного фонду досліджень України. Грант № 2020.02/0121.

Основна проблема для таких моделей — це труднощі їхнього налаштування (ідентифікації) за реальними даними. Детальніші моделі потребують більш повних даних щодо профілю захворювання та його поширення. У детермінованих моделях більшість їхніх параметрів обчислює середні величин. Вони враховують стохастичність процесів у середньому, успішно апроксимуючи ситуацію щодо великих популяцій в однорідному середовищі. Стохастичні моделі, на відміну від їхніх детермінованих аналогів, адекватніше відображають перебіг процесів особливо в локальних або перехідних процесах.

У цій статті описано деякі моделі для прогнозування епідеміологічної ситуації та спроби їхнього налаштування. Зауважимо, що більш детальні моделі для налаштування потребують доступу до більш повних даних щодо профілю захворювання та його поширення. Також однією з проблем є недостатньо повне тестування населення і прихований перебіг захворювань зі слабо вираженими симптомами у частини населення.

ДЕТЕРМІНОВАНІ ЕПІДЕМІОЛОГІЧНІ МОДЕЛІ

Згідно з [2] за моделлю SIR населення розподіляється на категорії (compartments): S (Susceptible, вразливі), I (Infectious, інфіковані (виявлені випадки хвороби)) та R (Removed, особи, що перехворіли), динаміка яких описується у вигляді відповідних систем диференціальних рівнянь:

$$\begin{aligned}\frac{dS}{dt} &= -\beta \frac{1}{N_0} SI, \\ \frac{dI}{dt} &= \beta \frac{1}{N_0} SI - \gamma I, \\ \frac{dR}{dt} &= \gamma I,\end{aligned}$$

де β та γ — рівень захворюваності та швидкість одужання відповідно, N_0 — кількість населення, що знаходиться під загрозою інфікування.

Ця система обґрунтована таким чином: збільшення кількості інфікованих прямо пропорційно залежить від наявної кількості інфікованих (адже чим більшою є чисельність інфікованих людей, тим більше маємо джерел інфекції) та наявної кількості вразливих осіб. Кількість тих, хто одужав з набуттям імунітету, та кількість померлих прямо пропорційно залежить від кількості виявлених інфікованих. Така модель достатньо чітко описує початок епідемії, коли щільність інфікованих серед вразливих є низькою.

Часто змінними можуть бути не абсолютні величини S , I , R , а їхні пропорції відносно N_0 , тобто $S_* = S / N_0$; $I_* = I / N_0$; $R_* = R / N_0$. Таку систему можна записати у компактнішому вигляді.

Використаємо звичні позначення: $T_{\text{inf}} = 1 / \gamma$, $R_0 = \beta / \gamma$, де T_{inf} — середній період, коли інфікована особа y_j^* поширює хворобу, R_0 — відповідно репродуктивний коефіцієнт, тобто середня кількість заражень однією хворою людиною за період її інфекційності. Тоді $\gamma = 1 / T_{\text{inf}}$, $\beta = R_0 \gamma = R_0 / T_{\text{inf}}$. Отже, маємо рівняння

$$\begin{aligned}\frac{dS}{dt} &= -\frac{R_0}{T_{\text{inf}}} \frac{1}{N_0} SI, \\ \frac{dI}{dt} &= \frac{R_0}{T_{\text{inf}}} \frac{1}{N_0} SI - \frac{1}{T_{\text{inf}}} I, \\ \frac{dR}{dt} &= \frac{1}{T_{\text{inf}}} I.\end{aligned}\tag{1}$$

Зауважимо, що розподіл населення за відповідними категоріями задовольняє загальному балансу $N_0(t) = S(t) + I(t) + R(t)$. Тому на середньо- та довгостроко-

вому періоді для моделювання необхідно враховувати динаміку кількості населення $N_0(t)$ (демографію), зважаючи на кількість народжуваних b за одиницю часу та коефіцієнт смертності μ населення:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{R_0}{T_{\text{inf}}} \frac{1}{N_0} SI + b - \mu S, \\ \frac{dI}{dt} &= \frac{R_0}{T_{\text{inf}}} \frac{1}{N_0} SI - \frac{1}{T_{\text{inf}}} I - \mu I, \\ \frac{dR}{dt} &= \frac{1}{T_{\text{inf}}} I - \mu R.\end{aligned}\quad (2)$$

У моделі SEIR долучають до розгляду додаткову категорію E (Exposed) для осіб, що перебувають в інкубаційному періоді, коли особа тільки інфікувалась, але не заражає інших. До попередніх рівнянь крім динаміки категорії $E(t)$ додають параметр T_{inc} — середній інкубаційний період, а система має вигляд

$$\begin{aligned}\frac{dS}{dt} &= -\frac{R_0}{T_{\text{inf}}} \frac{1}{N_0} SI, \\ \frac{dE}{dt} &= \frac{R_0}{T_{\text{inf}}} \frac{1}{N_0} SI - \frac{1}{T_{\text{inc}}} E, \\ \frac{dI}{dt} &= \frac{1}{T_{\text{inc}}} E - \frac{1}{T_{\text{inf}}} I, \\ \frac{dR}{dt} &= \frac{1}{T_{\text{inf}}} I.\end{aligned}\quad (3)$$

У цю модель можна увести змінну D , що визначає смертність, спричинену інфікуванням [3]. Тоді третє рівняння системи (3) змінюють на таке:

$$\frac{dI}{dt} = \frac{1}{T_{\text{inc}}} E - \frac{1}{T_{\text{inf}}} (1 + \delta)I$$

та додають п'яте рівняння

$$\frac{dD}{dt} = \frac{1}{T_{\text{inf}}} \delta I.$$

Тут параметр $\delta = Cfp / (1 - Cfp)$, де Cfp описує середню пропорцію фатальних (летальних) випадків для інфікованих.

Описані моделі в подальшому уточнюють, додаючи нові категорії, наприклад категорії осіб, що підлягають госпіталізації та потребують застосування апаратів ШВЛ.

Проте для детальних моделей потрібне відповідне налаштування. Навіть проста модель (3) потребує даних щодо розподілу інфікованих осіб на категорії E та I , тобто на тих, хто перебуває в інкубаційному періоді, та на тих, хто вже вийшов з нього.

Оскільки тестування не можна вважати достатнім, у модель (1) для налаштування можна ввести ще один параметр. Затим що офіційні дані відображають лише певну частку P_{inf} інфікованих осіб (кількість виявлених випадків захворювання), а замість реальних величин I відомі лише $I_* = P_{\text{inf}} I$, то, вводячи позначення $I = I_* / P_{\text{inf}}$, отримаємо

$$\begin{aligned}\frac{dS}{dt} &= -\frac{R_0}{T_{\text{inf}}} \frac{1}{N_0} S \frac{1}{P_{\text{inf}}} I_*, \\ \frac{dI_*}{dt} &= \frac{R_0}{T_{\text{inf}}} \frac{1}{N_0} SI_* - \frac{1}{T_{\text{inf}}} I_*, \\ \frac{dR}{dt} &= \frac{1}{T_{\text{inf}} P_{\text{inf}}} I_*.\end{aligned}\quad (4)$$

У цьому разі параметри R_0 , T_{inf} , P_{inf} системи (4) можна оцінити за близькістю траєкторій $I_*(t)$ та $S(t)$ до відповідних даних спостережень. Оскільки їх можна оцінити лише за траєкторією $I_*(t)$, параметр P_{inf} на розв'язок не впливатиме. Отже, система (4) не відрізнятиметься від системи (1).

НАЛАШТУВАННЯ МОДЕЛІ SIR

Повернімося до системи (1), яку налаштуватимемо за даними спостережень активних інфікованих y_t , $t=1, \dots, n$, які визначають як добову «загальну кількість інфікованих» мінус «загальну кількість тих, що перехворіли». Оскільки досліджуваний процес динамічно змінюється, як критерій розглядатимемо підсумовування зваженого модуля відхилення траєкторії від точок спостережень. Величини ваг підбирають залежно від плину часу. Для спостережень y_i , $i=1, \dots, n$, такий критерій має вигляд

$$\text{Crit} = \sum_{t=1}^n w_t |I(t) - y_t|, \quad w_t = 2t / (n(n+1)). \quad (5)$$

Зафіксуємо чисельність населення України на час моделювання на рівні $N_0 = 41$ млн. 858 тис. осіб (за деякими оцінками на 01.03.2020 року). Динаміку чисельності населення можна врахувати згідно з моделлю (2).

Зауважимо, що параметр T_{inf} визначає середній період, під час якого інфікована особа поширює хворобу. Цей параметр визначається дією вірусу та організмом людини, тому не може довільно змінюватись. За оцінками Оксфордської моделі для COVID-19 $T_{\text{inf}} \approx 4.5$. Тому можна обмежувати підбір параметрів умовою $4.4 \leq T_{\text{inf}} \leq 4.6$, а підбір параметрів R_0 , T_{inf} — мінімізацією критерію (5), яку можна виконувати за допомогою стандартних процедур для обмеженої глобальної оптимізації.

На цей час розроблено велику кількість моделей такого «детермінованого» типу; поведінку і властивості багатьох з них досліджено. Принциповим недоліком цих моделей є відсутність стохастичних підходів, хоча процес розвитку епідемії за своєю суттю є випадковим.

МОДЕЛЬ GUY KATRIEL ТА ЇЇ УЗАГАЛЬНЕННЯ

Ця модель базується на результатах роботи [3]. У ній запропоновано стохастичну модель епідемії з дискретним часом, в якій щоденна кількість нових захворювань має біноміальний розподіл, залежний від кількості захворювань у попередні дні. Ця модель має такі переваги.

1. Модель враховує зміни рівня інфекційності у процесі розвитку хвороби, тобто ймовірність передавання захворювання у кожний день хвороби окремим хворим.

2. Модель є стохастичною, що відповідає реальному характеру поширення інфекції серед популяції.

3. Модель є простою для застосування, використовуються відомі формули для обчислення оцінок методом максимальної правдоподібності параметра (базового репродуктивного коефіцієнта), що дорівнює середній кількості людей, які інфікуються одним хворим протягом усього періоду захворювання. Цей параметр визначає швидкість поширення епідемії. Така оцінка дає змогу визначити параметр з використанням попередньої статистики щоденної кількості нових захворювань для подальшого прогнозування розвитку епідемії.

У роботі [3] також розглянуто варіанти моделей, за якими населення поділяють на підгрупи, наприклад за віком, за відсутністю набутого імунітету.

Також розроблено розширений варіант моделі [4–6].

1. Уведено додатковий параметр: ймовірність виявлення захворювання. Оскільки в реальних умовах не всі випадки захворювання виявлено або враховано статистикою, оцінка параметра на підставі попередніх даних не є точною, тому параметр використовують для коригування статистичних даних, враховуючи певний рівень недостовірності.

2. Додано можливість розбиття тривалості епідемії на кілька періодів з різними значеннями параметрів на різних етапах. Оцінки параметрів на окремих етапах не є незалежними, тому потрібно максимізувати наближення статистики всієї епідемії. Розбиття на етапи необхідно у разі введення нових карантинних заходів (при цьому змінюється параметр) або зміни рівня моніторингу населення (зміна параметра DR).

У випадках тривалих епідемій (таких як пандемія COVID-19) динаміка поширення захворювання може бути сезонною (через вплив погоди на рівень інфекційності і/або сезонну зміну кількості контактів серед населення).

Розроблено програму для оцінювання параметрів і подальшої симуляції розвитку епідемії.

ВИПАДКОВІ МОДЕЛІ ПОШИРЕННЯ ЕПІДЕМІЙ

Розглянемо деякі моделі і методи, за допомогою яких визначають поширення епідемій як випадкові, а не детерміновані процеси [7–10].

Нехай n — кількість людей, що захворіли. Щодня кожна хвора людина (незалежно від інших хворих) може одужати з ймовірністю β/n і померти — з ймовірністю γ/n . Також щодня хворі отримують деяку кількість ліків x , які в нашій моделі вважаються абсолютно ефективними. Процес завершується, коли всі хворі або одужають, або помруть. Задача полягає в знаходженні для заданих значень параметрів γ , β та n такого значення x , для якого досягається максимальна ефективність від наданих ліків.

Розглянемо особливості задачі.

1. На відміну від більшості моделей епідеміології до класу хворих під час процесу лікування не додають нових індивідів. Це має місце, наприклад, коли хвороба є генетичною або викликана деякою поодинокую катастрофою.

2. Вважатимемо, що всі параметри задачі (γ , β , n , x) додатні числа.

3. Вважатимемо, що x не змінюється в часі. На відміну від параметрів γ , β , n ми регулюємо кількістю одиниць ліків x , але ця кількість залишається сталою.

Розв'язання задачі. Нехай $N(t)$ — кількість хворих у момент часу t . Розглянемо процес, що задається рівнянням

$$M(t) = n - \sum_{i=0}^t (\xi(i) + x + \mu(i)), \quad M(0) = n, \quad (6)$$

де $\xi(i)$ — кількість людей, що померли в момент i ; $\mu(i)$ — кількість тих, хто одужав самостійно в момент i .

Для будь-якої траєкторії матимемо

$$N(t) = \begin{cases} M(t), & M(t) > 0; \\ 0 & M(t) \leq 0. \end{cases}$$

Дійсно, $N(t)$ перестає задовольняти рівняння (6), лише коли кількість хворих стає меншою за щоденну кількість одиниць ліків x . Надалі будемо використовувати $M(t)$ і покажемо, що отримані результати можна застосувати до $N(t)$.

Для знаходження ефективного методу надання ліків було доведено три леми [10].

Лема 1. Для математичного сподівання має місце твердження

$$E[(M(t))] = \left(1 - \frac{\gamma}{n} - \frac{\beta}{n}\right)^t n \left(1 + \frac{x}{\gamma + \beta}\right) - \frac{nx}{\gamma + \beta}.$$

Лема 2. Для другого моменту $M(t)$ має місце твердження

$$E[M^2(t)] = a_1^t \left(n^2 - \frac{a_4}{1 - a_1}\right) + a_2 a_3 \frac{a_1^t - a_3^t}{a_1 - a_2} + \frac{a_4}{1 - a_1},$$

де

$$a_1 = \frac{n(n - \gamma - \beta) + (\gamma + \beta)^2}{n^2},$$

$$a_2 = \frac{-2xn^2 + n(2x + 1)(\gamma + \beta) - (\gamma + \beta)^2}{n} \left(1 + \frac{x}{\gamma + \beta} \right),$$

$$a_3 = 1 - \frac{\beta}{n} - \frac{\gamma}{n},$$

$$a_4 = \frac{2xn^2 - n(2x + 1)(\gamma + \beta) + (\gamma + \beta)^2}{n} \left(\frac{x}{\gamma + \beta} \right) + x^2.$$

Лема 3. Нехай $a \in \mathbf{R}$. Тоді

$$\frac{D[M(an)]}{(E[M(an)])^2} \rightarrow 0, \quad n \rightarrow \infty.$$

Використовуючи ці леми, одержуємо твердження [10].

Теорема 1. Нехай $T(n)$ — загальна тривалість хвороби для n інфікованих, тобто $T(n) = \min_{t \in N} \{t: N(t) = 0\}$. Тоді

$$\forall \varepsilon > 0 \quad P(|T(n) - a_0 n| > \varepsilon n) \rightarrow 0, \quad n \rightarrow \infty,$$

$$\text{де } a_0 = \frac{\ln \frac{x + \gamma + \beta}{x}}{\gamma + \beta}.$$

Теорема 2. Для математичних сподівань загальної кількості витрачених одиниць ліків та загальної кількості смертей має місце твердження

$$1) \quad \eta(x, n) = xT(n) \Rightarrow E\eta(x, n) = xET(n) \approx a_0 xn;$$

$$2) \quad E[\xi(x, n)] = E \left[\sum_{i=1}^{T(n)} \xi(i) \right] = \frac{\gamma}{n} \sum_{i=1}^{T(n)} EN(t) \approx n \left[(1 - e^{-a_0(\gamma + \beta)}) \frac{(x + \gamma + \beta)\gamma}{(\gamma + \beta)^2} - \frac{a_0 x \gamma}{\gamma + \beta} \right].$$

Результат теореми 2 дає змогу знаходити за допомогою чисельних методів для заданих значень параметрів γ , β та n таке значення x , для якого досягається максимальна ефективність надання ліків хворим.

РЕГРЕСІЙНІ МОДЕЛІ ПОШИРЕННЯ ЕПДЕМІЙ, РЕГРЕСІЇ З ПЕРЕМІКАННЯМИ

Останнім часом все більше уваги приділяють моделюванню, аналізу та прогнозуванню об'єктів зі змінною структурою та (або) змінними у часі параметрами. Такі процеси мають місце в медицині, економіці та техніці.

Існує кілька підходів до побудови моделей таких об'єктів. Одним з найбільш перспективних, на наш погляд, є використання регресій з перемиканнями, в яких точки перемикання невідомі, тому потребують оцінювання.

Суть регресії з перемиканнями полягає у тому, що параметри регресії не постійні на всьому інтервалі спостереження. Вони є постійними на його підінтервалах, які відокремлені один від одного точками перемикання. Оцінюючи точки перемикання, можна визначити періоди часу, в які відбувалися структурні зміни об'єкта. Існує дві форми регресій з перемиканнями: з неперервною лінією регресії і лінією регресії, що має розриви у точках перемикання.

У дослідженнях таких регресій з дискретним часом вагомим є доробок П. Перона (P. Perron) зі співавторами [7–9]. Вони показали можливість застосування регресій з перемиканнями в економіці.

У роботі [11] уведено новий клас регресій з перемиканнями у неперервному часі і запропоновано метод їхньої побудови. На основі цієї роботи в статті [12] запропоновано метод побудови регресій з перемиканнями у дискретному часі.

У роботі [13] наведено попередній статистичний аналіз поширення коронавірусної інфекції в Україні на основі використання регресії з перемиканнями. Описану там процедуру обчислень можна автоматизувати, що дасть змогу обробляти дані у реальному масштабі часу.

Регресію з перемиканнями можна також використовувати для визначення тривалості лікування інфікованих коронавірусом, а також для контролювання перебігу різних епідемій.

ВИСНОВКИ

Розглянуто деякі підходи для створення стохастичних моделей прогнозування епідемій, а також моделі, математичні методи та програмні засоби для їхньої реалізації. Подальший розвиток цих моделей прогнозування та оцінювання поширення епідемій пов'язаний із застосуванням регресійних моделей із неперервним часом та стохастичних дифузійних рівнянь.

СПИСОК ЛІТЕРАТУРИ

1. Бровченко І. Розробка математичної моделі поширення епідемії COVID-19 в Україні. *Світгляд*. 2020. № 2 (82). С. 2–14.
2. Keeling M.J., Rohani P. Modeling infectious diseases in humans and animals. Princeton, NJ: Princeton University Press, 2007. 385 p.
3. Katriel G. Stochastic discrete-time age-of-infection epidemic models. *International Journal of Biomathematics*. 2013. Vol. 6, N 1. P. 999–1005.
4. Chowell G., Hyman J.M., Bettencourt L.M.A., Castillo-Chavez C. Mathematical models in population biology and epidemiology. Dordrecht: Springer, 2009. 363 p.
5. Kermack W., McKendrick A. Contributions to the mathematical theory of epidemics. I. *Bulletin of Mathematical Biology*. 1991. Vol. 53 (1–2). P. 33–55.
6. Кнопов П.С., Богданов О.В. Использование стохастической модели для прогнозирования длительных эпидемий *Проблемы управления и информатики*. 2021. № 3. С. 50–57.
7. Garcia R., Perron P. An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics*. 1996. Vol. 78, N 1. P. 111–125.
8. Bai J., Perron P. Estimating and testing linear models with multiple structural changes. *Econometrica*. 1998. Vol. 66, N 1. P. 47–78.
9. Bai J., Perron P. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*. 2003. Vol. 18, N 1. P. 1–22.
10. Кнопов П.С., Богданов О.В. Моделювання епідемій. *Кибернетика та комп'ютерні технології*. 2020. № 2. С. 30–44.
11. Кнопов П.С., Корхин А.С. Метод построения регрессии с переключениями в непрерывном времени с неизвестными точками переключения. *Кибернетика и системный анализ*. 2020. Т. 56, № 1. С. 82–96.
12. Корхин А.С. Приближенный метод построения регрессии с переключениями с неизвестными точками переключения. *Кибернетика и системный анализ*. 2020. Т. 56, № 3. С. 97–110.
13. Кнопов П.С., Корхин А.С. Статистический анализ динамики инфицирования коронавирусом с помощью пошаговой регрессии с переключениями. *Кибернетика и системный анализ*. 2020. Т. 56, № 6. С. 96–106.

O.V. Bogdanov, P.S. Knopov

STOCHASTIC MODELS IN THE PROBLEMS OF FORECASTING THE EPIDEMIOLOGICAL SITUATION

Abstract. The paper investigates some stochastic models with discrete and continuous time to solve important problems in predicting the spread of epidemiological diseases among the population. Various factors of epidemic spread and the main parameters influencing the assessment of the forecast are taken into account. Some test calculations based on the proposed methods have been performed.

Keywords: methods, optimization, modeling, stochastic equations, estimation, epidemic, discrete and continuous time.

Надійшла до редакції 08.09.2021