

Медведєв М.Г.

Таврійський національний університет імені В.І. Вернадського

Юсупів Т.В.

Таврійський національний університет імені В.І. Вернадського

ВИКОРИСТАННЯ ФУНКЦІЇ ЛОГНОРМАЛЬНОГО РОЗПОДІЛУ ДЛЯ МОДЕЛЮВАННЯ СМЕРТНОСТІ ВІД COVID-19

У статті запропоновано підхід до апроксимації статистичних даних щодо загальної кількості померлих від COVID-19 для таких країн, як Австралія та Італія. Ці країни вибрані для моделювання, оскільки мають суттєво різні дані щодо померлих, а це дає змогу перевірити модель у різних ситуаціях.

Основним критерієм вибору моделі було найкраще наближення статистичних даних, а саме:

- характер зміни значень загальної кількості померлих від COVID-19 (швидке зростання та повільне падіння, висота точки перегину);
- максимальне значення.

Обрана функція отримала назву «логнормальна модель», оскільки за її основу було взяти функцію логнормального розподілу. Вона показала гарне наближення реальної статистики за певних значень своїх параметрів. Вдалося вказати вузькі межі для двох величин: σ^2 – параметр моделі та y_0 – ордината точки перегину графіка моделі.

Похибка оцінювання максимального значення померлих не перевищувала 0,76%, при тому, що порядок параметра, який відповідає цій кількості, був зоразу різний (102, 905 та 35 473).

Застосування моделі чітко показало вплив на поведінку графіка параметрів, які можуть бути інтерпретовані в таких термінах: A – граничне значення графіка, t_0 – абсциса точки перегину, σ^2 – величина, що впливає на висоту точки перегину графіка, тобто характеризує якість і швидкість медичних заходів країни щодо виявлення та лікування від конкретного штаму вірусу.

Також цікавим моментом у процесі моделювання виявилася необхідність введення додаткового параметра d для часового лагу між тими даними, які вважаються першим летальним випадком від COVID-19, і стартовим моментом моделі. Для Австралії цей параметр двічі виявився додатним, тоді як для Італії він дорівнював нулю. Однією з можливих інтерпретацій тут може бути те, що не завжди вдається вчасно виявити першу жертву вірусу.

Ця модель має перспективи бути використаною для подальшого моделювання та, на його основі, прогнозування смертності від COVID-19 і для інших країн світу (хвиль), де наявна досить структурована статистика смертності.

Ключові слова: модель, COVID-19, соціальна статистика, наближення, логнормальний розподіл.

Постановка проблеми. Весь 2020 рік пройшов під гаслом світової боротьби з COVID-19. 11 березня 2020 р. ВООЗ характеризувала ситуацію з вірусом як пандемію. Станом на кінець листопада 2020 р. у світі налічується понад 58 млн зафіксованих випадків зараження, з яких більше 1,3 млн виявилися летальними.

Багато вчених по всьому світу намагаються дослідити статистику, яка стосується COVID-19, і використати її для моделювання, зокрема смертності. Такі дослідження можуть допомогти у процесі подальшого прогнозування або вказати на певні загальні закономірності.

Постановка завдання. Метою статті є опис, обґрунтування вибору та наближення до реальних даних моделі, що відображає тенденції зміни

загальної кількості померлих від COVID-19 у різних країнах. Для цього були використані дані смертності по Австралії (дві хвили) та Італії (перша хвиля), які були в числі перших у боротьбі із захворюваністю та веденні статистики, що, на нашу думку, є достовірною.

Виклад основного матеріалу дослідження. Серед багатьох функцій для моделювання була обрана така:

$$y(t) = \frac{A}{2} \left(1 + \operatorname{erf} \left(\frac{\ln t - \mu}{\sqrt{2\sigma^2}} \right) \right), \mu \in \mathbb{R}, t, \sigma^2 > 0,$$

де $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ – функція помилок.

Ця функція розподілу ймовірностей для логнормального розподілу, помножена на число A .

Сам логнормальний розподіл широко застосовується у статистиці для ймовірнісної оцінки вартості активу в майбутньому, а для опису загальної кількості померлих від COVID-19 цікавими є дві його характеристики:

– графік цього розподілу має правобічну скошеність (асиметричність), притаманну графіку реальної статистики померлих;

– за необмеженого зростання параметру часу t цей графік асимптотично наближається до деякого числа (в моделі – числа A), що характеризує загальну кількість померлих.

Друга похідна вибраної функції $y(t)$ має такий вигляд:

$$y''(t) = -\frac{Ae^{-\left(\frac{\ln(t)-\mu}{\sqrt{2\sigma^2}}\right)^2} (\ln(t) - (\mu - \sigma^2))}{\sqrt{2\pi\sigma^2} \sigma^2 t^2}, \mu \in \mathbb{R}, t, \sigma^2 > 0.$$

Прирівнюючи її до нуля, отримуємо абсцису точки перегину t_0 (до точки перегину кількість померлих за добу збільшується, після – зменшується):

$$\ln(t_0) - (\mu - \sigma^2) = 0, \text{ тобто } t_0 = e^{\mu - \sigma^2}.$$

Ордината цієї точки:

$$y_0 = y(t_0) = \frac{A}{2} \left(1 + \operatorname{erf} \left(\frac{\mu - \sigma^2 - \mu}{\sqrt{2\sigma^2}} \right) \right) = \frac{A}{2} \left(1 + \operatorname{erf} \left(\frac{-\sigma^2}{\sqrt{2\sigma^2}} \right) \right) = \frac{A}{2} \left(1 + \operatorname{erf} \left(-\sqrt{\frac{\sigma^2}{2}} \right) \right).$$

Таким чином, точка перегину розташована нижче прямої $y = 0,5A$ (оскільки функція помилок від від'ємного аргументу є від'ємною) та цілком визначається параметром σ^2 . На практиці перший факт добре узгоджується зі статистикою.

Для потреб моделі найкраще оперувати такими трьома параметрами: A , t_0 та σ^2 . Зауваживши, що $\mu - \sigma^2 = \ln(t_0)$ (а отже, $\mu = \ln(t_0) + \sigma^2$), отримуємо:

$$y(t) = \frac{A}{2} \left(1 + \operatorname{erf} \left(\frac{\ln t - \ln t_0 - \sigma^2}{\sqrt{2\sigma^2}} \right) \right) = \frac{A}{2} \left(1 + \operatorname{erf} \left(\frac{\ln \left(\frac{t}{t_0} \right) - \sigma^2}{\sqrt{2\sigma^2}} \right) \right), t_0, t, \sigma^2 > 0$$

Проте графік обраної функції розпочинається з дуже малих значень, тому введення додаткової змінної – $d \geq 0$ (стартового часу) допомагає «змістити» вліво його початок і, як результат, збільшити його початкові значення. Пояснити це можна тим, що важко точно визначити час смерті першої людини від COVID-19, тоді як модель починає наближати значення з точки $t = 1$, де її значення ще дуже малі.

Отже, остаточно було використано таку модель:

$$y(t) = \frac{A}{2} \left(1 + \operatorname{erf} \left(\frac{\ln \left(\frac{t+d}{t_0+d} \right) - \sigma^2}{\sqrt{2\sigma^2}} \right) \right), d, t_0, t, \sigma^2 > 0.$$

Той факт, що точка перегину має абсцису t_0 , наочно підтверджується для тих наборів даних, де можна було явно виділити дні з піковою смертністю.

Надалі модель має робочу назву «логнормальна модель».

Австралія. В Австралії, починаючи з 3 листопада 2020 р., перестали фіксувати нові випадки

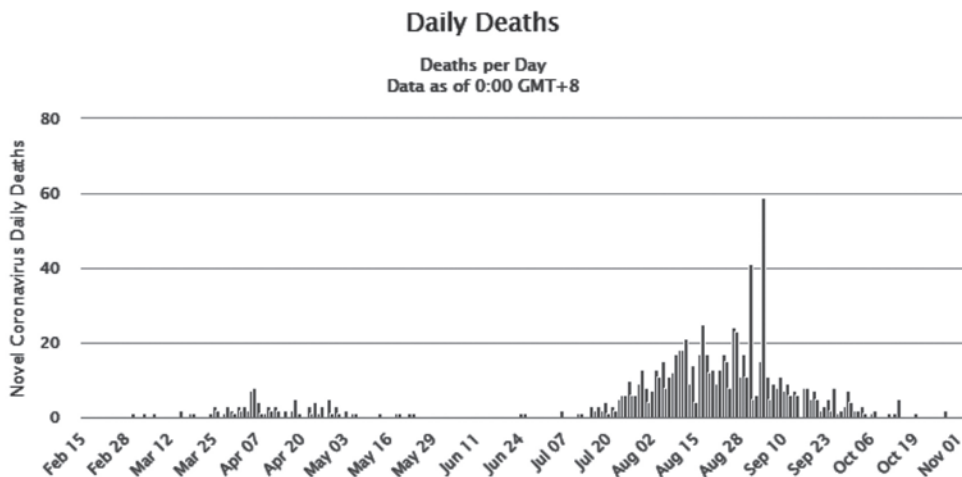


Рис. 1. Графік щоденної смертності за весь період, Австралія

зараження COVID-19. Країна пережила дві хвилі розповсюдження вірусу та, відповідно, дві хвилі смертності, відображені на графіку.

Загалом (із 15 лютого до 2 листопада 2020 р.) в Австралії зафіксовано 27 602 випадки захворювань, із яких 907 виявилися летальними. Між двома хвилями смертності був проміжок в один місяць (із 24 травня до 23 червня), коли жодна людина не померла, хоча випадки захворюваності фіксувалися.

Проаналізувавши статистику, виділяємо два періоди:

- перша хвиля смертності в Австралії – 84 дні (з 1.03 до 23.05);
- друга хвиля смертності в Австралії – 127 днів (з 24.06 до 28.10).

Перша хвиля забрала життя 102 людей, а друга – 805. Друга хвиля характеризувалася не тільки довшим у 1,5 раза періодом, але й більшою максимальною кількістю померлих за добу – 59 проти 8.

Водночас ці графіки можна добре наблизити обраною функцією.

Використовуючи «логнормальну модель», ми отримали такий результат для першої хвилі смертності.

Параметри моделі мають такі значення:

- $A = 103,22$;
- $\sigma^2 = 0,069$;
- $t_0 = 37,5$;
- $d = 17,8$.

Перегин відбувається в точці з ординатою $y_0 = 0,397A$. Максимальне значення відповідно до моделі становило 101,22, тоді як реальне значення (що відповідає 23.05.2020) становило 102 померлі (відносна похибка – 0,76%).

Параметри моделі для другої хвилі виявилися такими:

- $A = 809,81$;
- $\sigma^2 = 0,018$;
- $t_0 = 57,58$;
- $d = 81,14$.

Перегин відбувається в точці з ординатою $y_0 = 0,447A$. Максимальне значення відповідно до моделі становило 808,27, тоді як реальне значення кількості померлих дру-

гої хвилі (що відповідає 28.10) – 805 (відносна похибка – 0,41%).

Італія. Цікавим стало застосування «логнормальної моделі» для моделювання загальної кількості летальних випадків від COVID-19 в Італії. Тут перша хвиля смертності плавно перейшла у другу, досягши свого мінімуму 29.08.2020, коли померла одна людина. Таким чином, перша хвиля смертності для Італії тривала 191 день (із 21.02 до 29.08) і забрала життя 35 473 людей.

Як видно із графіка, модель дуже точно відображає реальний стан речей. Параметри моделі в цьому разі будуть такими:

- $A = 35402,81$;
- $\sigma^2 = 0,204$;
- $t_0 = 39,32$;
- $d = 0$.

Ордината точки перегину $y_0 = 0,326A$. Максимальне значення відповідно до моделі становило



Рис. 2. Моделювання першої хвилі смертності в Австралії



Рис. 3. Моделювання другої хвилі смертності в Австралії



Рис. 4. Моделювання першої хвилі смертності в Італії

35 361,73, тоді як реальне значення, що відповідає 29.08, – 35 473. Відносна похибка оцінки максимального значення в цьому разі становить 0,31%. $d = 0$, що свідчить про відсутність лагу між реальним початком смертності та модельним.

Висновки. Отже, «логнормальна модель» добре описує статистику летальних випадків

двох хвиль в Австралії та першої хвилі в Італії. Ця модель включає чотири параметри: A , σ^2 , t_0 та d . Функція, використана в моделі – це функція розподілу ймовірностей для логнормального розподілу, помножена на коефіцієнт A , який показує максимально можливу кількість померлих до закінчення поточної хвилі.

У процесі наближення було встановлено, що параметр $\sigma^2 \in [0,018; 0,204]$, ордината точки перегину $y_0 \in [0,326A; 0,447A]$. Відносно вузькі межі цих параметрів можуть свідчити про влучний вибір моделі, яка описує процеси смертності від COVID-19, для різних країн.

Порядок величини кількості померлих, що задається параметром A , може залежати від багатьох факторів, таких як мутація вірусу, недотримання правил санітарної безпеки, прориви у лікуванні чи настання холодного сезону.

Список літератури:

1. Химмельблау Д. Прикладное нелинейное программирование. Москва : Мир, 1975. 536 с.
2. Хвищун І.О. Програмування і математичне моделювання : підручник / Львівський національний ун-т ім. Івана Франка. Київ : Ін Юре : Видавничий центр Львівського національного університету ім. Івана Франка, 2007. 544 с.
3. Статистика смертності від COVID-19 у світі. 20 листопада 2020 р. URL: <https://www.worldometers.info/coronavirus>.

Medvedev M.H., Yusyiv T.V. USE OF THE LOGNORMAL DISTRIBUTION FUNCTION FOR SIMULATION OF MORTALITY FROM COVID-19

The article proposes an approach to approximate statistics on the total number of deaths from COVID-19 for countries such as Australia and Italy. These countries are chosen for modeling because they have significantly different data on deaths, which allows the model to be tested in different situations.

The main criterion for choosing the model was the best approximation of statistics, namely:

- the nature of changes in the values of the total number of deaths from COVID-19 (rapid growth and slow decline, the height of the inflection point);
- maximum value.

The chosen function was called the “lognormal model” because it was based on the lognormal distribution function. It showed a good approximation of real statistics at certain values of its parameters. It was possible to specify narrow limits for two quantities: σ^2 – model parameter and y_0 – ordinate of the inflection point of the model graph.

The error in estimating the maximum value of deaths did not exceed 0.76%, although the order of the parameter corresponding to this number was different each time (102, 905 and 35473).

The application of the model clearly showed the influence on the behavior of the graph of parameters that can be interpreted in the following terms: A – limit value of the graph, t_0 – abscissa of the inflection point, σ^2 – the value that affects the height of the inflection point of the graph, ie for the detection and treatment of a particular strain of the virus.

Another interesting point in the simulation process was the need to introduce an additional parameter d for the time lag between the data that are considered the first fatality from COVID-19, and the starting point of the model. For Australia, this parameter was twice positive, while for Italy it was zero. One possible interpretation here may be that it is not always possible to detect the first victim of the virus in time.

This model has the potential to be used for further modeling and, based on it, the prediction of mortality from COVID-19 and for other countries (waves), where there are fairly structured mortality statistics.

Key words: model, COVID-19, social statistics, simulation, lognormal distribution.