

УДК 519.21

*П.С. Кнопов, А.В. Богданов*

## ИСПОЛЬЗОВАНИЕ СТОХАСТИЧЕСКОЙ МОДЕЛИ ДЛЯ ПРОГНОЗИРОВАНИЯ ДЛИТЕЛЬНЫХ ЭПИДЕМИЙ \*

**Ключевые слова:** стохастическая дискретная модель эпидемии, оценка параметров, симуляция эпидемии.

### Введение

В нынешнее время проблема распространения эпидемий в связи с всемирной пандемией COVID-19 является актуальной. Из-за этого возникает необходимость разработки методов прогнозирования числа больных. Такая информация дает возможность оценить будущее влияние введения карантинных мер, а также эффективность вакцинации ограниченного количества населения.

Одним из наиболее исследованных типов моделей эпидемий являются полигамные модели [1], такие как модели SI и SIR, в которых население разбивается на несколько групп (уязвимые и инфицированные в случае SI; уязвимые, инфицированные и выздоровевшие в случае SIR). Динамика перехода людей из одной группы в другую обычно описывается дифференциальными уравнениями в случае непрерывного времени или разностными уравнениями в случае дискретного времени. Модели также могут включать зависимость уровня инфекционности от времени болезни отдельного больного (ageoinfection).

Другим типом моделей являются сетевые модели, базирующиеся на изображении схемы контактов между инфицированными и уязвимыми лицами с помощью графов [2–4]. Также моделирование может базироваться на ветвящихся процессах [5]. Такие модели особенно полезны для описания начала эпидемий, когда количество инфицированных недостаточно большое для применения полигамных моделей, в которых группы считаются равномерно смешанными между собой.

Модели эпидемий могут быть детерминированными или стохастическими. Обычно стохастические модели являются более точными, ведь распространение болезни — процесс случайный в связи с произвольностью контактов между больными и уязвимыми людьми и случайностью передачи инфекции при контакте. Тем не менее, обычно анализ стохастических моделей гораздо сложнее или невозможен по сравнению с детерминированными моделями [6].

Рассматриваемая модель основана на результатах работы [7]. В ней предложена стохастическая модель эпидемии с дискретным временем, в которой ежедневное количество новых заболеваний имеет биномиальное распределение, зависящее от количества заболеваний в предыдущие дни. Эта модель имеет ряд следующих преимуществ.

---

\* Работа выполнена при частичной поддержке Национального фонда исследований Украины. Грант № 2020.02/0121.

© П.С. КНОПОВ, А.В. БОГДАНОВ, 2021

1. Модель учитывает изменения уровня инфекционности заболевания в процессе развития болезни, т.е. вероятность передачи заболевания каждый день болезни отдельного больного.

2. Модель является стохастической, что соответствует реальному характеру передачи инфекции среди популяции.

3. Модель проста в работе, имеет известные формулы для оценки методом максимального правдоподобия параметра  $R_0$  (базовое репродуктивное число), равного среднему числу людей, которые инфицируются одним больным в течение всего заразного периода болезни. Этот параметр указывает на скорость распространения эпидемии. Такая оценка дает возможность определить параметр  $R_0$ , используя предварительную статистику ежедневного числа новых заболеваний для дальнейшего прогнозирования развития эпидемии.

Также в [7] рассмотрены варианты модели с разбивкой населения на подгруппы (например, по возрасту) и с отсутствием приобретенного иммунитета, а также приближенная и детерминированная версии модели.

Был разработан расширенный вариант модели.

1. Введен дополнительный параметр: вероятность выявления заболевания  $DR$ . Поскольку в реальных условиях не все случаи заболевания выявлены или учтены в статистике, оценка параметра  $R_0$  на основании предварительных данных не является точной, поэтому параметр  $DR$  используется для корректировки статистических данных, учитывая определенный уровень недоверности.

2. Добавлена возможность разбиения длительности эпидемии на несколько периодов с различными значениями параметров на разных этапах. При этом оценки параметров на отдельных этапах не являются независимыми, а стремятся максимизировать приближение статистики всей эпидемии. Разбиение на этапы необходимо в случаях введения новых карантинных мер (при этом изменяется параметр  $R_0$ ) или корректировки уровня мониторинга населения (изменение параметра  $DR$ ). Также в случаях длительных эпидемий, таких как пандемия COVID-19, динамика распространения заболевания может быть сезонной (из-за влияния погоды на уровень инфекционности и/или сезонное изменение количества контактов среди населения).

Также была разработана программа для оценки параметров  $R_0$  и  $DR$  и последующей симуляции развития эпидемии.

### Описание модели и программы для симуляции эпидемии

Для сравнения использована модель, предложенная в [7].

В основе модели лежит предположение, что вероятность передачи болезни при контакте зависит от длительности заболевания и она известна, т.е. имеется вектор  $(P_1, \dots, P_d)$ , где  $P_r$  — вероятность больного заразить восприимчивого человека при контакте на  $r$ -й день болезни,  $d$  — длительность периода заразности. Тогда число новых заболевших  $i(t)$  в день  $t$  является случайной величиной. Также предполагается, что число восприимчивых  $S(t)$  в день эпидемии  $t$  уменьшается только в связи с новыми случаями заболеваний, т.е.

$$S(t) = S(t-1) - i(t). \quad (1)$$

Предположим, что в среднем один человек совершает  $\beta$  контактов с другими людьми в день. В этом случае в [5] представлено распределение числа новых заболевших в день  $t$ , которое является биномиальным со следующими параметрами:

$$i(t) \cong \text{Binomial} \left( s(t-1), 1 - \exp \left( -\frac{\beta}{N-1} \sum_{r=1}^d P_r i(t-r) \right) \right), \quad (2)$$

где  $S(t)$  — число восприимчивых к болезни людей в день  $t$ ,  $\beta$  — среднее число контактов в день,  $N$  — размер популяции,  $d$  — длительность периода заразности,  $P_r$  — вероятность больного заразить уязвимого на  $r$ -й день болезни.

Видим, что распределение числа новых заболевших  $i(t)$  в день  $t$  зависит от числа восприимчивых  $S(t-1)$  в предыдущий день эпидемии и числа новых инфицированных  $i(t-r)$ ,  $1 \leq r \leq d$  в предыдущие  $d$  дней. Благодаря формуле (1) это позволяет симулировать всю эпидемию, если известно число новых инфицированных в первые  $d$  дней эпидемии.

Также в [7] представлена формула для базового репродуктивного числа  $R_0$  :

$$R_0 = \beta \sum_{r=1}^d P_r. \quad (3)$$

Исходя из формулы (2) модель не меняется при умножении  $P_r$  на константу и делении  $\beta$  на ту же константу. Таким образом, можно ввести ограничение

$$\sum_{r=1}^d P_r = 1. \quad (4)$$

Тогда  $\beta$ ,  $P_r$  теряют изначальную интерпретацию, однако из формулы (3) получаем

$$R_0 = \beta. \quad (5)$$

В состав модели добавлен еще один параметр:  $DR$  — вероятность выявления заболевания, показывающий, какая часть реально инфицированных людей была и будет обследована, выявлена как заболевшая и задокументирована. Это позволяет более точно симулировать реальные данные.

Была разработана программа, симулирующая число больных в каждый день эпидемии согласно представленной модели, а также производящая оценку параметров  $\beta$  и  $DR$ . Для демонстрации эффективности модели было проведено два типа симуляции. В первом случае параметры  $\beta$  и  $DR$  подбирались вручную, во втором — автоматически с помощью программы.

### Ручной подбор параметров модели

Для оценки эффективности модели и программы симуляции использовано число заболевших вирусом COVID-19 в разных областях в течение первого месяца распространения эпидемии в Украине. Это позволяет проверить эффективность работы модели на наиболее проблематичном в плане описания отрезке времени, где число заболевших и качество статистических данных являются низкими.

Сначала рассмотрим вариант с достаточно короткой длительностью эпидемии с постоянными значениями параметров  $\beta$  и  $DR$ . В случае более длительных промежутков времени значения параметров могут существенно меняться в течение эпидемии в связи с введением карантинных мер, изменением качества обследования и погодными условиями, и этот случай рассмотрен отдельно ниже.

Полагаем, что длительность периода инфекционности равна 10 дням. Инфекционность заболевания считалась постоянной в течение болезни:  $P_1 =$

$= P_2 = \dots = P_{10} = \frac{1}{10}$ . Размер популяции равен населению рассматриваемой области.

Начальное число инфицированных:  $1/10000$  населения рассматриваемой области.

В случае ручного подбора параметров рассматривался график суммарного числа заболевших, и значения параметров  $R_0$  и  $\beta$  варьировались для достижения максимального приближения результатов симуляции к реальным данным.

Для примера результата работы модели представлен город Киев. Его население равно 2 967 360. На рис. 1 на оси  $Ox$  показан день эпидемии, а на оси  $Oy$  — суммарное число заболевших до этого дня.

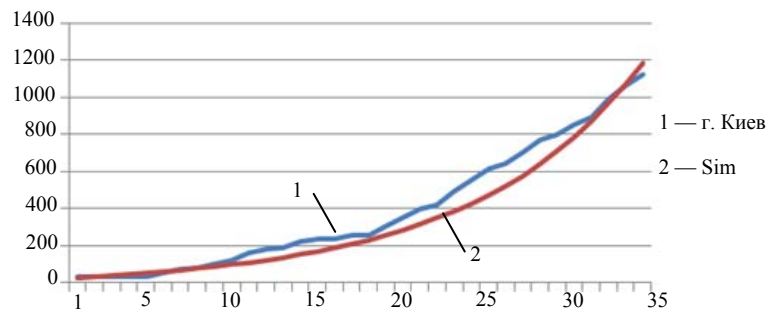


Рис. 1

Значения наилучшей пары параметров для Киева равны

$$R_0 = \beta = 1,5, \quad (6)$$

$$DR = 0,2. \quad (7)$$

Видим, что с помощью подбора пары параметров  $(\beta, DR)$  можно достигнуть высокой точности приближения реального числа заболевших.

Для остальных областей Украины подбор подходящих параметров также позволяет достаточно точно приблизить кривую суммарного числа заболевших.

#### Автоматический подбор параметров

Разработана возможность автоматического подбора параметров  $\beta$  и  $DR$ . В этом случае алгоритм выглядит так.

1. Значение  $\hat{DR}$  выбирается 0,01 и меняется до единицы с интервалом 0,05.
2. Для каждого значения  $\hat{DR}$  производится оценка  $\beta$  методом максимального правдоподобия. В [5] приведена формула этой оценки:

$$\hat{\beta} = \frac{(N-1) \sum_{t=d+1}^T i(t)}{\sum_{t=d+1}^T \left[ N - \sum_s i(s) \right] \sum_{\tau=1}^d P_{\tau} i(s-\tau)}. \quad (8)$$

3. Для каждой пары  $(\hat{\beta}, \hat{DR})$  генерируется эпидемия.
4. Каждая сгенерированная эпидемия сравнивается с исходными данными посредством подсчета суммы квадратов отклонений количества новых заболевших каждый день.

5. Сгенерированная эпидемия с наименьшим значением суммы квадратов отклонений принимается за конечный результат.

Кроме отсутствия необходимости ручного подбора параметров, автоматический подбор позволяет избавиться от влияния выбросов (дней со значительно более высоким или низким числом новых инфицированных, чем в предыдущие или следующие дни). Такие выбросы не связаны с качественным изменением поведения эпидемии, однако заметно влияют на статистику суммарного числа заболевших за период пандемии, и поэтому использование суммарного числа заболевших для описания поведения эпидемии может привести к неточностям.

Для примера результата работы модели представлен город Киев. На первом графике представлено ежедневное число новых заболевших, на втором — суммарное число заболевших, как в случае ручного подбора параметров (рис. 2, а, б).

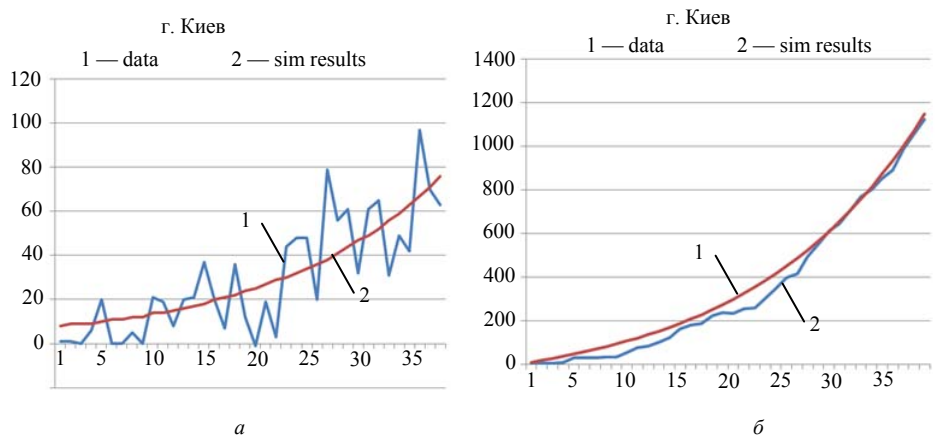


Рис. 2

Значения наилучшей пары параметров для Киева равны:

$$R_0 = \beta = 1,39, \tag{9}$$

$$DR = 0,2. \tag{10}$$

Рассматривая результаты подбора параметров для всех областей Украины, заметим, что модель хорошо описывает реальные данные, особенно число новых инфицированных. В случаях суммарного числа заболевших и больших выбросов результаты симуляции менее точно описывают статистические данные, однако это, скорее, является недостатком самой выборки. В основном приближение достаточно близкое, и с увеличением размера выборки эта проблема становится менее значительной, как показано в следующем разделе.

### Моделирование с разными этапами эпидемии

Разработана возможность подбора разных значений  $\beta$  и  $DR$  на разных этапах эпидемии и последующая симуляция эпидемии с этими значениями.

Производился подбор параметров на каждом этапе, начиная с первого. Последние  $d$  значений числа инфицированных и числа восприимчивых предыдущего этапа принимались как начальные условия следующего промежутка времени, пока не была сгенерирована вся эпидемия. Результаты симуляции сравнивались

с исходными данными методом наименьших квадратов, и набор пар  $(\beta, DR_s)$ ,  $1 \leq s \leq n$  с наименьшим отклонением и симуляция эпидемии с этим набором параметров считались результатом оценки параметров.

Исследование проводилось с данными по Киеву (стандартный вариант) и Чехии (экстремальный вариант), при этом эпидемия рассматривалась на значительно более длительном промежутке времени, чем в случае предыдущих симуляций.

Сначала рассмотрим результаты симуляции с единственной парой параметров  $(\beta, DR)$  на всем промежутке времени  $[1, 231]$  (рис. 3, а). Как и прежде, на графике на оси  $Ox$  показан день эпидемии, а на оси  $Oy$  — суммарное число заболевших до этого дня.

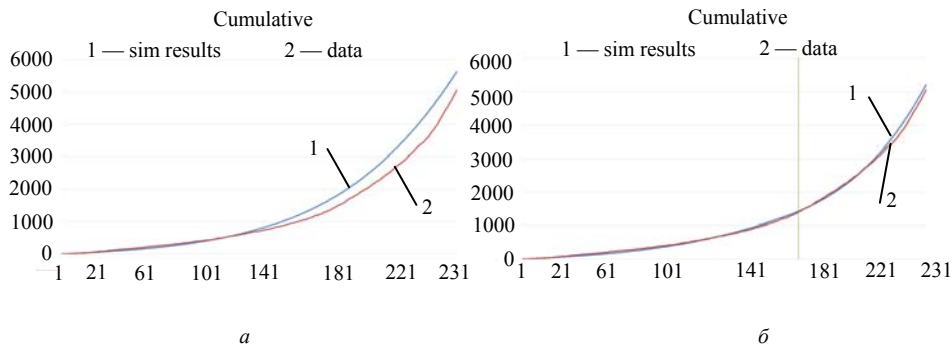


Рис. 3

Мы видим заметное отклонение результатов симуляции от реальных данных, т.е. единственная пара параметров  $(\beta, DR)$  не может адекватно описать поведение эпидемии на всем рассмотренном промежутке времени.

Рассмотрим результат симуляции с двумя промежутками:  $[1, 181]$  и  $[181, 231]$  (рис. 3, б).

Заметно значительное улучшение качества приближения, так как разные значения параметров позволяют описать этапы с низкой и высокой скоростью распространения заболевания.

Рассмотрим случай Чехии. В этих данных через 6 месяцев после начала эпидемии произошел резкий всплеск числа заболевших. В случае использования единственного значения параметров на всем промежутке времени  $[1, 256]$  получен следующий результат (рис. 4).

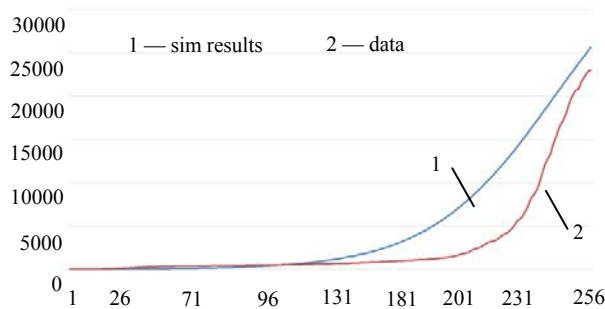


Рис. 4

Существенное отклонение говорит о невозможности объяснения динамики эпидемии при неизменных значениях параметров.

Рассмотрим результат симуляции с двумя промежутками: [1, 181] и [181,256] (рис. 5).

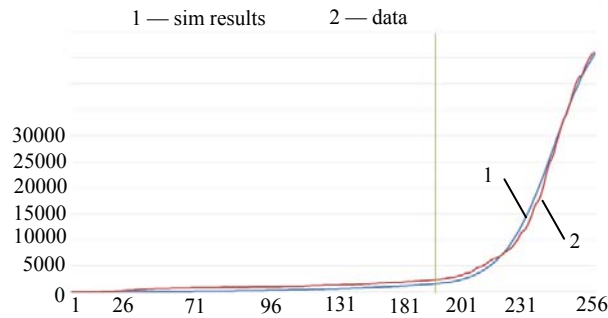


Рис. 5

Заметно значительное улучшение качества приближения. Таким образом, даже в случае резкого изменения динамики эпидемии, разбиение на промежутки с разными значениями параметров позволяет использовать модель для точного описания числа инфицированных. Однако не стоит использовать слишком большое число промежутков, так как при этом уменьшается размер выборки на каждом из них, что может привести к ухудшению оценки.

### Заключение

Резюмируя полученные результаты, отметим, что представленная модель дает возможность оценки базового репродуктивного числа и вероятности выявления заболевания, базируясь на существующей статистике эпидемии, и впоследствии эта модель с полученными параметрами может быть использована для прогнозирования будущего развития эпидемии. Также представлена возможность разбиения эпидемии на несколько отрезков с разными значениями параметров и оценка этих параметров для описания длительных эпидемий с изменениями динамики.

*П.С. Кнопов, О.В. Богданов*

## ВИКОРИСТАННЯ СТОХАСТИЧНОЇ МОДЕЛІ ДЛЯ ПРОГНОЗУВАННЯ ТРИВАЛИХ ЕПІДЕМІЙ

Розглядається стохастична дискретна модель епідемії, в якій рівень інфекційності залежить від тривалості хвороби, з оцінкою максимальної вірогідності для параметра, що відповідає за швидкість розповсюдження інфекції. Для більш ефективного використання реальної статистики кількості хворих введено додатковий параметр моделі — ймовірність виявлення захворювання. Розроблено програму для автоматичної оцінки параметрів моделі з використанням минулої статистики з подальшою симуляцією розвитку епідемії. Наведено порівняння результатів симуляції з існуючими даними кількості хворих COVID-19 у місті Києві з ручним та автоматичним підбором параметрів. Розглянуто можливість та доцільність розбиття епідемії на декілька проміжків з різними значеннями параметрів для описання тривалих епідемії зі значними змінами динаміки. Виконано порівняння результатів симуляції з різною кількістю проміжків для описання довгострокового розвитку епідемії COVID-19 у місті Києві та Чехії, що мають різну динаміку розвитку епідемії.

**Ключові слова:** стохастична дискретна модель епідемії, оцінка параметрів, симуляція епідемії.

P.S. Knopov, O.V. Bogdanov

## APPLICATION OF STOCHASTIC MODEL FOR LENGTHY EPIDEMIC FORECASTING

In this paper we consider a stochastic discrete-time epidemic model, with the infectivity depending on the age of infection and existing formula for the maximum likelihood estimation of the parameter responsible for the rate of the infection spread. In order to utilize the real number of infection cases statistics, a detection rate parameter is introduced. A program for automatic parameter estimation using past data with future epidemic simulation is developed. We present the comparison between the simulation of COVID-19 cases in Kyiv and real data using manual and automatic parameter estimation. We consider the possibility of the epidemic partition into several intervals with different parameters in order to simulate lengthy epidemics with significant changes in dynamics. We present the comparison between different numbers of partitions for long-term COVID-19 simulation in Kyiv (Ukraine) and Czech Republic, which have different dynamics of the epidemic development.

**Keywords:** stochastic discrete-time epidemic model, parameter estimation, epidemic simulation.

1. Kermack W., McKendrick A. Contributions to the mathematical theory of epidemics. *I. Bulletin of Mathematical Biology*. 1991. **53** (1–2). P. 33–55.
2. Daduna G., Knopov P.S., Chornej R.K. Local control of Markovian processes of interaction on graph with compact set of states. *Cybernetics and Systems Analysis*. 2001. № 3. P. 62–78.
3. Knopov P.S., Samosonok A.S. On Markov stochastic processes with local interaction for solving some applied problems. *Cybernetics and Systems Analysis*. 2011. **47**(3). P. 346–359.
4. Brauer F., Castillo-Chavez C. *Mathematical models in population biology and epidemiology*. New York : Springer, 2001. 501 p.
5. Ball F., Neal P. Network epidemic models with two levels of mixing. *Mathematical Biosciences*. 2008. **212**(1). P. 69–87.
6. Allen L.J.S. *Stochastic population and epidemic models: persistence and extinction*. Cham : Springer, 2015.
7. Katriel G. Stochastic discrete-time age-of-infection epidemic models. *International Journal of Biomathematics*. 2013. **6**, N 1. 1250066. 20 p.

Получено 03.03.2021