

Математика больших данных

Модели распространения эпидемий, в частности, COVID-19 как модель стохастической химической кинетики

Бабкин Петр Б05-003
Дорин Даниил Б05-003
Киселев Никита Б05-003
Крейнин Матвей Б05-003
Никитина Мария Б05-003

Модель SIR (без рождения и смерти)

Динамика эпидемии, например гриппа, часто намного быстрее, чем динамика рождаемости и смертности, поэтому в простых моделях рождаемость и смертность часто опускаются. Пусть популяция включает в себя N человек. Введем в рассмотрение следующие функции:

1. $S(t)$ — число людей, еще не переболевших вирусом, но восприимчивых к заболеванию, в момент времени t ;
2. $I(t)$ — число людей, инфицированных вирусом, способных заразить восприимчивых;
3. $R(t)$ — число людей, получивших иммунитет.

В модели SIR вводятся следующие естественные предположения:

1. Суммарное число людей в популяции остается постоянным и равным N ;
2. Число заболеваний пропорционально числу контактов между людьми;
3. Скорости заражения и выздоровления не меняются с течением времени.

Из определения ясно, что на данные функции накладывается естественное условие:

$$S(t) + I(t) + R(t) \equiv N. \quad (0.1)$$

Составим систему уравнений, которая бы на макроскопическом уровне описывала то, что происходит в популяции. Рассмотрим некоторый промежуток времени Δt . Пусть за это время произошло число заражений $\Delta I = -\Delta S$ и число выздоровлений ΔR , тогда доля зараженных людей в популяции изменилась на $\Delta I/N = -\Delta S/N$, а доля выздоровевших на $\Delta R/N$. С другой стороны, заметим, что каждую из этих величин можно выразить следующим образом:

$$\frac{\Delta S}{N} = -\frac{I}{N} \cdot \frac{S}{N} \cdot \beta \cdot \Delta t \quad (0.2)$$

$$\frac{\Delta I}{N} = \frac{I}{N} \cdot \frac{S}{N} \cdot \beta \cdot \Delta t - \frac{I}{N} \cdot \gamma \cdot \Delta t \quad (0.3)$$

$$\frac{\Delta R}{N} = \frac{I}{N} \cdot \gamma \cdot \Delta t \quad (0.4)$$

Здесь I/N и S/N — вероятности встретить зараженного и восприимчивого человека в популяции соответственно, β — скорость заражения, а γ — скорость выздоровления зараженного человека. При промежутке времени $\Delta t \rightarrow 0$ динамика системы может быть выражена систе-

мной обыкновенных дифференциальных уравнений (СОДУ):

$$\frac{dS}{dt} = -\beta \frac{IS}{N} \quad (0.5)$$

$$\frac{dI}{dt} = \beta \frac{IS}{N} - \gamma I \quad (0.6)$$

$$\frac{dR}{dt} = \gamma I \quad (0.7)$$

Из введенной в рассмотрение системы сразу же следует, что

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0, \quad (0.8)$$

откуда

$$S(t) + I(t) + R(t) = \text{const} = N. \quad (0.9)$$

Во-первых, обратим внимание, что вышеприведенное соотношение также означает то, что можно изучать систему только для двух из трех переменных. Во-вторых, отметим, что динамика класса инфицированных зависит от следующего соотношения:

$$R_0 = \frac{\beta}{\gamma}, \quad (0.10)$$

так называемого *базового индекса репродукции*. Эта величина рассчитывается как ожидаемое количество новых инфекций (эти новые инфекции иногда называют вторичными инфекциями) от одной инфекции в популяции, где все субъекты восприимчивы.

Роль как базового индекса репродукции, так и начального числа восприимчивых людей чрезвычайно велика. Фактически, если переписать уравнение для заразных лиц следующим образом:

$$\frac{dI}{dt} = \left(R_0 \frac{S}{N} - 1 \right) \gamma I,$$

получается, что если

$$R_0 \cdot S(0) > N,$$

то

$$\frac{dI}{dt}(0) > 0,$$

т.е. произойдет собственно эпидемическая вспышка с увеличением числа заразных (которые могут охватить значительную часть населения). Напротив, если

$$R_0 \cdot S(0) < N,$$

то

$$\frac{dI}{dt}(0) < 0,$$

т.е. независимо от первоначальной численности восприимчивого населения болезнь никогда не может вызвать настоящую эпидемическую вспышку.

Модели стохастической химической кинетики

Общие модели макросистем с точки зрения математики можно понимать как разнообразные модели стохастической химической кинетики [1, 2]. В общем и целом, подход следующий. Предположим, что некоторая макросистема может находиться в различных состояниях, характеризуемых вектором n с неотрицательными целочисленными компонентами (по сути это просто число частиц/людей в данном состоянии). Будем считать, что в системе происходят случайные превращения (химические реакции).

Пусть $n \rightarrow n - \alpha + \beta$, $(\alpha, \beta) \in J$ — все возможные типы реакций, где α и β — вектора с неотрицательными целочисленными компонентами. Введем интенсивность реакции:

$$\lambda_{(\alpha, \beta)}(n) = \lambda_{(\alpha, \beta)}(n \rightarrow n - \alpha + \beta) = N^{1 - \sum_i \alpha_i} K_\beta^\alpha \prod_{i: \alpha_i > 0} n_i \cdot \dots \cdot (n_i - \alpha_i + 1),$$

где $K_\beta^\alpha \geq 0$ — константа реакции; при этом $\sum_{i=1}^m n_i(t) \equiv N \gg 1$. Другими словами, $\lambda_{(\alpha, \beta)}(n)$ — вероятность осуществления в единицу времени перехода $n \rightarrow n - \alpha + \beta$. Здесь не предполагается, что число состояний $m = \dim n$ и число реакций $|J|$ не зависят от числа агентов N .

Возникающий марковский процесс считается неразложимым (все состояния являются сообщающимися, т.е. каждое достижимо из каждого). Далее приводится теорема Куртца [1, 2], которая говорит о том, что если существует

$$\lim_{N \rightarrow \infty} \frac{n(0)}{N} = c(0),$$

где $K_\beta^\alpha := K_\beta^\alpha(n/N)$, а m и J не зависят от N , то для любого $t > 0$ с вероятностью 1 существует

$$\lim_{N \rightarrow \infty} \frac{n(t)}{N} = c(t) \quad \left[\text{т.е. } \frac{n(t)}{N} \xrightarrow{\text{п.н.}} c(t) \right],$$

где $c(t)$ — не случайная вектор-функция, удовлетворяющая СОДУ Гульдберга-Вааге (закон действующих масс):

$$\frac{dc_i}{dt} = \sum_{(\alpha, \beta) \in J} (\beta_i - \alpha_i) K_\beta^\alpha(c) \prod_j c_j^{\alpha_j}.$$

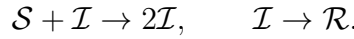
Более того, случайный процесс $n(t)/N$ слабо сходится при $N \rightarrow \infty$ к $c(t)$ на любом конечном промежутке времени. Отметим, что в каком-то смысле жизнь нелинейной динамической системы определяется линейными законами сохранения, унаследованными ею при каноническом скейлинге (по сути заключающемся в замене концентраций их средними значениями).

SIR как модель стохастической химической кинетики

Заметим, что модель SIR, изложенная в первом разделе, в действительности является моделью стохастической кинетики. В наших обозначениях

$$n(t) = \begin{bmatrix} n_1(t) \\ n_2(t) \\ n_3(t) \end{bmatrix} = \begin{bmatrix} S(t) \\ I(t) \\ R(t) \end{bmatrix}, \quad n_i(t) \in \mathbb{N} \cup \{0\}, \quad i = 1, 2, 3.$$

Все возможные переходы описываются как



Тогда всего есть две ситуации:

$$\alpha^1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \beta^1 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \quad \text{и} \quad \alpha^2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \beta^2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Обозначим $\nu(t) = n(t)/N$. Тогда, предполагая, что $\nu(0) \xrightarrow{\text{п.н.}} c(0)$, имеем $\nu(t) \xrightarrow{\text{п.н.}} c(t)$, а СОДУ Гильдберга-Вааге примет вид

$$\begin{aligned} \frac{dc_1}{dt} &= -\beta(t)c_1c_2 \\ \frac{dc_2}{dt} &= \beta(t)c_1c_2 - \gamma(t)c_2 \\ \frac{dc_3}{dt} &= \gamma(t)c_2 \end{aligned}$$

В самом деле, эта система совпадает с СОДУ, полученной в модели SIR, однако там мы не акцентировали внимание на том, что $n(t)$ вообще-то является случайным вектором, $N \gg 1$, а коэффициенты β и γ могут зависеть от времени.

Марковская модель распространения эпидемии

Попробуем получить марковскую модель описания микроскопической динамики эпидемии, которая в макромасштабе описывается моделью SIR.

В популяции из $N \gg 1$ человек распространяется эпидемия. Люди делятся на восприимчивых к вирусу, зараженных и переболевших. Каждый восприимчивый к вирусу человек может сопротивляться ему некоторое время, имеющее показательное распределение с параметром $\alpha(t)$, пропорциональным доле зараженных людей. Кроме того, каждый зараженный человек имеет шанс выздороветь, причем время выздоровления имеет показательное распределение с постоянным параметром γ . Известно, что в начальный момент времени $t = 0$ было заражено I_0 человек.

Поскольку $\alpha(t) \propto I(t)/N$, то можно записать $\alpha(t) = \beta I(t)/N$, где $\beta = \text{const}$.

Аналогично введем величины $S(t)$, $I(t)$ и $R(t)$, причем $S(t) + I(t) + R(t) \equiv N$. Рассмотрим случайный процесс

$$\xi(t) = \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix} = \begin{bmatrix} S(t) \\ I(t) \end{bmatrix}, \quad t \geq 0, \quad \xi(0) = \begin{bmatrix} \xi_1(0) \\ \xi_2(0) \end{bmatrix} = \begin{bmatrix} S(0) \\ I(0) \end{bmatrix} = \begin{bmatrix} N - I_0 \\ I_0 \end{bmatrix}.$$

Поскольку по постановке задачи приращения процесса являются независимыми, а сам процесс принимает значения из конечного множества $\{0, \dots, N\} \times \{0, \dots, N\}$, то мы имеем дело с непрерывной марковской цепью.

Обозначим переходные вероятности (учитывая, что цепь однородная)

$$p_{(s_1, i_1) \rightarrow (s_2, i_2)}(\Delta t) = \mathbb{P}(S(\Delta t) = s_2, I(\Delta t) = i_2 \mid S(0) = s_1, I(0) = i_1).$$

Распределение вероятностей состояний:

$$\pi_{(s, i)}(t) = \mathbb{P}(S(t) = s, I(t) = i).$$

Выведем формулы для вероятностей переходов. Рассмотрим момент времени $t \geq 0$ и случайную величину $\eta(t)$, равную времени до заражения случайного восприимчивого человека, начиная с момента времени t . Введем случайную величину $\varphi(t) = \min\{\eta_i(t)\}_{i=1}^{S(t)}$ — время до заражения первого из $S(t)$ восприимчивых человек, начиная с момента времени t . Поскольку по условию задачи $\eta(t) \sim \text{Exp}(\beta I(t)/N)$, то

$$\mathbb{P}(\varphi(t) < \Delta t) = 1 - \mathbb{P}(\varphi(t) \geq \Delta t) = 1 - \prod_{i=1}^{S(t)} \mathbb{P}(\eta_i(t) \geq \Delta t) = 1 - \exp\left(-\beta \frac{I(t)S(t)}{N} \Delta t\right),$$

откуда имеем, что $\varphi(t) \sim \text{Exp}(\beta I(t)S(t)/N)$.

Аналогично, рассмотрев случайную величину $\psi(t)$, равную времени до выздоровления первого из $I(t)$ зараженных человек, получим, что $\psi(t) \sim \text{Exp}(\gamma I(t))$.

Теперь запишем вероятности переходов [3]:

$$\begin{aligned}\mathbb{P}((S(t + \Delta t), I(t + \Delta t)) - (S(t), I(t)) = (-1, 1)) &= \mathbb{P}(\varphi(t) < \Delta t) \simeq \beta \frac{I(t)S(t)}{N} \Delta t + o(\Delta t), \\ \mathbb{P}((S(t + \Delta t), I(t + \Delta t)) - (S(t), I(t)) = (0, -1)) &= \mathbb{P}(\psi(t) < \Delta t) \simeq \gamma I(t) \Delta t + o(\Delta t), \\ \mathbb{P}((S(t + \Delta t), I(t + \Delta t)) - (S(t), I(t)) = (0, 0)) &= 1 - \left(\beta \frac{I(t)S(t)}{N} + \gamma I(t) \right) \Delta t + o(\Delta t).\end{aligned}$$

Таким образом, имеем

$$p_{(s,i) \rightarrow (s+k,i+j)}(\Delta t) = \begin{cases} \beta \frac{is}{N} \Delta t + o(\Delta t), & (k, j) = (-1, 1), \\ \gamma i \Delta t + o(\Delta t), & (k, j) = (0, -1), \\ 1 - \left(\beta \frac{is}{N} + \gamma i \right) \Delta t + o(\Delta t), & (k, j) = (0, 0), \\ o(\Delta t), & \text{иначе.} \end{cases}$$

Таким образом, элементы инфинитезимальной матрицы (в нашем случае — 4-х мерного тензора) Q есть

$$q_{(s,i) \rightarrow (s+k,i+j)} = \begin{cases} \beta \frac{is}{N}, & (k, j) = (-1, 1), \\ \gamma i, & (k, j) = (0, -1), \\ - \left(\beta \frac{is}{N} + \gamma i \right), & (k, j) = (0, 0), \\ 0, & \text{иначе.} \end{cases}$$

Итак, поведение цепи можно представлять себе следующим образом. В состоянии (s, i) цепь находится некоторое время $t \sim \text{Exp}(q_{(s,i)})$, где $q_{(s,i)} = -q_{(s,i) \rightarrow (s,i)}$. По прошествии этого времени наступает момент скачка цепи в другое состояние. В этот момент скачка цепь может попасть в произвольное состояние $(s + k, i + j)$ с вероятностью $q_{(s,i) \rightarrow (s+k,i+j)}/q_{(s,i)}$. Таким образом,

▷ Вероятность возникновения одного нового заражения есть

$$\frac{\left(\beta \frac{is}{N}\right)}{\left(\beta \frac{is}{N} + \gamma i\right)},$$

▷ Вероятность одного выздоровления есть

$$\frac{\gamma i}{\left(\beta \frac{is}{N} + \gamma i\right)}.$$

Имея построенную марковскую модель, можно моделировать течение эпидемии. Ниже приведен псевдокод, позволяющий осуществить это [4].

Algorithm 1 stochastic SIR model of epidemic spreading

Input: β, γ, N, I_0

Output: arrays t, S, I, R

$j = 0$

$t[j] = 0$

$S[j] = N - I_0$

$I[j] = I_0$

$R[j] = 0$

while $I[j] > 0$ **do**

$a = \beta I[j] S[j] / N$

$b = \gamma I[j]$

$p_1 = a / (a + b)$

▷ probability of a new infection

$p_2 = b / (a + b)$

▷ probability of a recovery from infection

 generate $u_1, u_2 \sim U[0, 1]$

$t[j + 1] = t[j] - \ln(u_1) / (a + b)$

▷ if $u_1 \sim U[0, 1]$ then $-\ln(u_1) / (a + b) \sim \text{Exp}(a + b)$

if $0 < u_2 \leq p_1$ **then**

$S[j + 1] = S[j] - 1$

$I[j + 1] = I[j] + 1$

$R[j + 1] = R[j]$

else if $p_1 < u_2 \leq 1$ **then**

▷ $p_1 + p_2 = 1$

$S[j + 1] = S[j]$

$I[j + 1] = I[j] - 1$

$R[j + 1] = R[j] + 1$

end if

$j = j + 1$

end while

Остается открытым вопрос, почему полученная марковская модель микроскопической динамики эпидемии в макромасштабе описывается моделью SIR?

Список литературы

- [1] N. Buzun, A. Gasnikov, F. Goncharov, O. Gorbachev, S. Guz, E. Krymova, A. Natan, and E. Chernousova. Stochastic analysis in problems, part 1 (in russian), 2017.
- [2] A. Gasnikov, E. Gorbunov, S. Guz, E. Chernousova, M. Shirobokov, and E. Shulgin. Lecture notes on stochastic processes, 2021.
- [3] P. E. Greenwood and L. F. Gordillo. *Stochastic Epidemic Modeling*, pages 31–52. Springer Netherlands, Dordrecht, 2009.
- [4] O. Ige. Markov chain epidemic models and parameter estimation, May 2020. Available at <https://mds.marshall.edu/etd/1307>.