

Stochastyczne i deterministyczne modele epidemii

Marta ZALEWSKA*, Wojciech NIEMIRO**

* Zakład Profilaktyki Zagrożeń Środowiskowych i Alergologii, Warszawski Uniwersytet Medyczny
 ** Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

O procesach Markowa można przeczytać na przykład w $\Delta_{13}^{9,12}$.

Założenie o dużej populacji pozwala zignorować fakt, że przyrost $I(t)$ zmniejsza liczbę osób narażonych na zarażenie. To uproszczenie jest uzasadnione na początku epidemii.

Zjawiska opisywane na poziomie „mikro” przez procesy Markowa zachowują się na poziomie „makro” jak funkcje deterministyczne, będące rozwiązaniami równań różniczkowych. Można to zaobserwować na przykładzie prostego (ale, niestety, bardzo aktualnego w chwili pisania tego artykułu) modelu początkowej, wykładniczej fazy epidemii.

Rozważmy bardzo dużą populację. Niech $I(t)$ oznacza liczbę zarażonych w chwili t . Zakładamy, że w krótkim okresie czasu h każdy z nich zaraża średnio αh innych ludzi, ale przy tym z prawdopodobieństwem βh zdrowieje lub umiera i przestaje zarażać. Deterministyczny model jest następujący:

$$(1) \quad I(t+h) = I(t) + (\alpha - \beta)I(t)h + o(h) \quad \text{dla } h \searrow 0.$$

(Traktujemy $I(t)$ jako wielkość ciągłą. Symbol $o(h)$ oznacza nieskończenie małą rzędu wyższego niż h , czyli funkcję spełniającą równość $\lim_{h \searrow 0} o(h)/h = 0$.) Oczywiście (1) sprowadza się do zwyczajnego równania różniczkowego:

$$\frac{dI(t)}{dt} = (\alpha - \beta)I(t).$$

Popatrzmy na to samo zjawisko z bliska (w skali mikro). Potraktujmy $I(t)$ jako zmienną losową o wartościach całkowitoliczbowych. Ewolucja procesu $I(t)$ jest opisana równaniami

$$(2) \quad \begin{aligned} \mathbb{P}(I(t+h) = i+1 | I(t) = i) &= \alpha i h + o(h) \quad \text{dla } h \searrow 0, \\ \mathbb{P}(I(t+h) = i-1 | I(t) = i) &= \beta i h + o(h) \quad \text{dla } h \searrow 0, \\ \mathbb{P}(I(t+h) = i | I(t) = i) &= 1 - (\alpha + \beta) i h + o(h) \quad \text{dla } h \searrow 0. \end{aligned}$$

Zakładamy, że $\{I(t) : t \geq 0\}$ jest procesem Markowa, to znaczy ewolucja procesu po chwili t zależy tylko od $I(t)$, czyli stanu w chwili t , a nie od wcześniejszej „historii” procesu. Jest to skokowy proces Markowa: w losowych momentach $T_1 < T_2 < \dots < T_n < \dots$ proces rośnie lub maleje o 1. Żeby lepiej zrozumieć strukturę tego procesu i sposób jego symulowania, przyjrzyjmy się czasom pomiędzy skokami, $W_{n+1} = T_{n+1} - T_n$, gdzie

$$T_{n+1} = \inf\{t > T_n : I(t) \neq I(T_n)\} \quad (\text{rys. 1}).$$

Okazuje się (co zostanie uzasadnione na końcu artykułu), że W_{n+1} jest zmienną losową o rozkładzie *wykładniczym z parametrem* $(\alpha + \beta)i_n$. Ponadto

$$\mathbb{P}(I(T_{n+1}) = i_n + 1 | I(T_n) = i_n) = \frac{\alpha}{\alpha + \beta}, \quad \mathbb{P}(I(T_{n+1}) = i_n - 1 | I(T_n) = i_n) = \frac{\beta}{\alpha + \beta}.$$

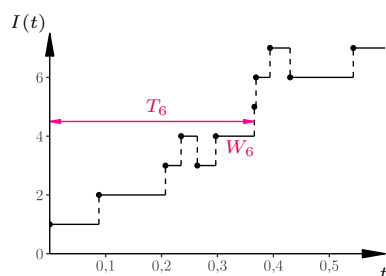
Łatwo w to uwierzyć, patrząc na równania (2).

Przebieg procesu $I(t)$ symulujemy zatem, losując kolejno czasy T_n i stany $I(T_n)$ zgodnie z dwiema powyższymi regułami. Chemicy, biolodzy i epidemiolodzy znają tę metodę pod nazwą algorytmu Gillespie’go. W rzeczywistości taki opis skokowego procesu Markowa podał znakomity amerykański matematyk czeskiego pochodzenia, Joseph Doob. Przykładowy przebieg krótkich symulacji przedstawiamy w poniższej tabelce i na rysunku 1.

krok n	1	2	3	4	5	6	7	8	9	10
czas T_n	0,088	0,207	0,235	0,264	0,297	0,366	0,369	0,394	0,430	0,543
stan $I(T_n)$	2	3	4	3	4	5	6	7	6	7

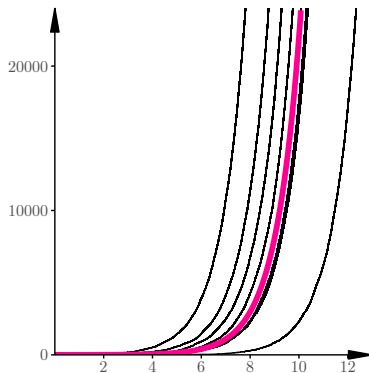
$\mathbb{P}(A|B)$ to prawdopodobieństwo zdarzenia A pod warunkiem zajścia zdarzenia B , czyli $\mathbb{P}(A \cap B)/\mathbb{P}(B)$.

Co to znaczy, że zmienna ma rozkład wykładniczy? Niech λ będzie dodatnią liczbą rzeczywistą. Jeśli wylosujemy liczbę U jednostajnie z odcinka $[0, 1]$ (tzn. szansa na to, że U wpadnie w dowolnie wybrany odcinek $J \subset [0, 1]$, jest równa długości J), to zmienna losowa $W = -(\ln U)/\lambda$ ma rozkład wykładniczy z parametrem λ .

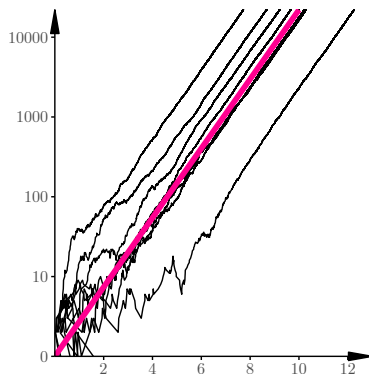


Rys. 1

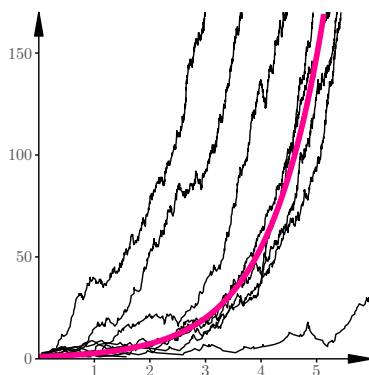
Popatrzmy teraz na dłuższe przebiegi symulacji i porównajmy trajektorie procesu losowego z rozwiązaniem równania różniczkowego, czyli funkcją wykładniczą $I(t) = i_0 \exp\{(\alpha - \beta)t\}$. Przyjęliśmy następujące wartości parametrów: $\alpha = 2$, $\beta = 1$, $i_0 = I(0) = 1$ (śledzimy rozwój epidemii od pierwszego



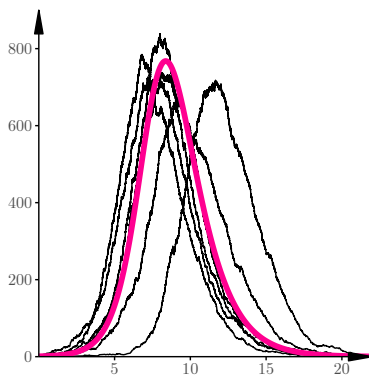
Rys. 2



Rys. 3. Skala logarytmiczna



Rys. 4. Początkowe fragmenty 20 trajektorii

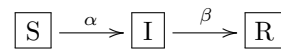


Rys. 5. Trajektorie procesu $I(t)$ w modelu SIR

zarażonego). Na rysunku 2 przedstawionych jest 20 niezależnych trajektorii procesu Markowa. Spośród tych 20 trajektorii 12 wpadło dość wcześnie w stan 0. Jest to stan „pochłaniający”: jeśli $I(T_n) = 0$, to dla każdego $t \geq T_n$ mamy $I(t) = 0$. Pogrubiona, kolorowa linia to wykres funkcji wykładniczej.

Uderzającą cechą procesu losowego opisanego równaniami (2) jest zgodność trajektorii w makroskali z rozwiązaniem równania różniczkowego, czyli z funkcjami wykładniczymi $I(t) = \exp\{(\alpha - \beta)(t - c)\}$. Poszczególne realizacje procesu różnią się jednak wyraźnie przesunięciem w fazie c . To widać na rysunku 3, na którym oś pionowa została przedstawiona w skali logarytmicznej. Rysunek 4 pokazuje te same 20 trajektorii w początkowym okresie rozwoju. Wspomniane wyżej przesunięcia w fazie wynikają z losowych fluktuacji na początku. Rezultatem tych losowych fluktuacji jest to, że 12 spośród przedstawionych trajektorii jest „pochłoniętych” przez zero. Dalszy przebieg procesu jest niemal deterministyczny.

Model SIR. Rozważmy populację złożoną z ℓ osobników. Niech $I(t)$ oznacza liczbę zarażonych w chwili t , zaś $R(t)$ łączną liczbę uodpornionych i zmarłych. Liczba osobników narażonych na zakażenie jest równa $S(t) = \ell - I(t) - R(t)$. Osobnik typu S może przejść do kategorii I , a stąd do kategorii R (stąd nazwa „model SIR”). Schematycznie:



Zakładamy, że w krótkim okresie czasu h każdy osobnik I zaraża średnio $\alpha(S(t)/\ell)h$ innych ludzi, a z prawdopodobieństwem βh zdrowieje lub umiera i przestaje zarażać. Zauważmy, że jeśli $S(t)/\ell \simeq 1$, to dostajemy model fazy wykładniczej, przedstawiony w zadaniu poprzednim.

Klasyczny model deterministyczny jest następujący:

$$(3) \quad \begin{aligned} I(t+h) &= I(t) + \left(\alpha \frac{S(t)}{\ell} - \beta \right) I(t)h + o(h) \quad \text{dla } h \searrow 0, \\ S(t+h) &= S(t) - \alpha \frac{S(t)}{\ell} I(t)h + o(h) \quad \text{dla } h \searrow 0, \end{aligned}$$

oraz $R(t) = \ell - I(t) - S(t)$. Oczywiście (3) jest układem równań różniczkowych zwyczajnych:

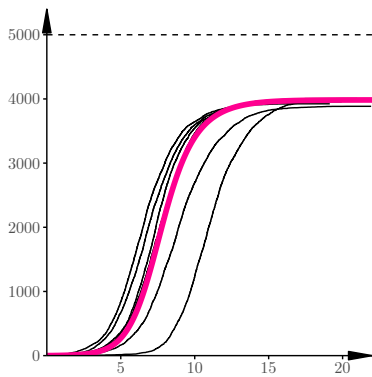
$$(4) \quad \frac{dI(t)}{dt} = \left(\alpha \frac{S(t)}{\ell} - \beta \right) I(t), \quad \frac{dS(t)}{dt} = -\alpha \frac{S(t)}{\ell} I(t).$$

W skali mikro traktujemy $I(t)$ i $S(t)$ jako zmienne losowe o wartościach całkowitoliczbowych. Stan układu jest parą $(I, S)_t := (I(t), S(t)) = (i, s)$. Ewolucja procesu jest opisana równaniami

$$(5) \quad \begin{aligned} \mathbb{P}((I, S)_{t+h} = (i+1, s-1) | (I, S)_t = (i, s)) &= \alpha \frac{s}{\ell} ih + o(h) \quad \text{dla } h \searrow 0, \\ \mathbb{P}((I, S)_{t+h} = (i-1, s) | (I, S)_t = (i, s)) &= \beta ih + o(h) \quad \text{dla } h \searrow 0, \\ \mathbb{P}((I, S)_{t+h} = (i, s) | (I, S)_t = (i, s)) &= 1 - \left(\alpha \frac{s}{\ell} + \beta \right) ih + o(h) \quad \text{dla } h \searrow 0. \end{aligned}$$

Symulacja procesu opisanego powyższymi równaniami również może zostać przeprowadzona przy użyciu algorytmu Gillespie’go. Tym razem jednak musimy monitorować dwie wielkości – liczbę zainfekowanych (I) oraz liczbę osobników podatnych na zakażenie (S). Jeśli liczby te wynoszą i i s , odpowiednio, to czas oczekiwania na kolejne „zdarzenie” (czyli zainfekowanie nowej osoby lub wyzdrowienie/śmierć osoby zarażonej) ma rozkład wykładniczy z parametrem $\frac{\alpha}{\ell} si + \beta i$. Kiedy już dojdzie do zdarzenia, to z prawdopodobieństwem $\frac{\frac{\alpha}{\ell} si}{\frac{\alpha}{\ell} si + \beta i}$ jest to nowe zarażenie (oraz wyzdrowienie/śmierć w przeciwnym przypadku). Nowe stany odpowiadające tym dwóm rodzajom zdarzeń to, odpowiednio, $(i+1, s-1)$ oraz $(i-1, s)$. Oczywiście symulacja kończy się, kiedy $i = 0$.

Na rysunku 5 widać kilka trajektorii procesu opisanego równaniami (5), wraz z zaznaczonym kolorem rozwiązaniem równania różniczkowego (4). Przyjęliśmy następujące wartości parametrów: $\alpha = 2$, $\beta = 1$, $i_0 = I(0) = 1$, $\ell = 5000$.



Rys. 6. Liczba wszystkich dotychczas zainfekowanych. Przerwaną linią oznaczono rozmiar populacji. We wszystkich symulowanych przebiegach po wygaśnięciu epidemii około 1000 osób pozostaje w stanie S (są to osoby, które nie zostały zarażone)

Początkowo trajektorie zachowują się tak, jak w modelu poprzednim: najpierw sporo losowych fluktuacji, później następuje faza wykładniczego wzrostu. Jeszcze później widać zupełnie inne zjawisko: hamowania wzrostu wskutek spadku liczby narażonych $S(t)$, aż do zupełnego wygaśnięcia epidemii.

Dla różnych realizacji tego samego procesu SIR widać duże przesunięcia w fazie, wynikające z losowego charakteru początkowego fragmentu. Losowe fluktuacje (głównie początkowe) mają też pewien (niewielki) wpływ na maksymalną liczbę zarażonych (maksyma poszczególnych krzywych).

W opracowaniach dotyczących rozwoju epidemii często przedstawia się również liczbę wszystkich osób, które zostały zainfekowane do danego momentu (tzn. liczbę zdarzeń polegających na zarażeniu nowej osoby). Liczba ta odpowiada procesowi $\ell - S(t)$, widocznemu na rysunku 6. Warto zwrócić uwagę na niewielkie fluktuacje liczby wszystkich osób, które zostały zakażone do czasu wygaśnięcia epidemii.

Przedstawione w tym artykule modele rozwoju epidemii są skrajnie uproszczone i nie nadają się do ilościowego opisu rzeczywistego zjawiska. Niemniej nawet takie modele pozwalają trochę zrozumieć mechanizm epidemii na poziomie jakościowym.

Dlaczego czasy oczekiwania na skok mają rozkład wykładniczy?

Aby udowodnić, że zmienna W ma rozkład wykładniczy z parametrem λ , wystarczy pokazać, że zmienna $e^{-\lambda W}$ ma rozkład jednostajny na odcinku $[0, 1]$, tzn. że dla dowolnego $t > 0$: $\mathbb{P}(W < t) = \mathbb{P}(e^{-\lambda W} > e^{-\lambda t}) = 1 - e^{-\lambda t}$.

Przyjrzyjmy się czasowi *pierwszego* skoku, $W_1 = T_1 = \inf\{t : I(t) \neq i_0\}$. Nietrudno uwierzyć, że

$$\mathbb{P}(W_1 \leq t + h | W_1 > t) = \mathbb{P}(I(t+h) \neq i_0 | I(t) = i_0) + o(h) = (\alpha + \beta)i_0 h + o(h),$$

gdzie druga równość wynika z (2). Jeśli $F(t) = \mathbb{P}(W_1 \leq t)$, to dostajemy stąd równanie $\frac{F(t+h) - F(t)}{1 - F(t)} = (\alpha + \beta)i_0 h + o(h)$, zatem $\frac{d}{dt} \log(1 - F(t)) = \frac{(1 - F(t))'}{1 - F(t)} = -(\alpha + \beta)i_0 t$, skąd wnioskujemy, że $1 - F(t) = \exp\{-(\alpha + \beta)i_0 t\}$. Identyczne rozumowanie stosuje się do W_{n+1} i prowadzi ono do analogicznego rezultatu, z i_0 zastąpionym przez i_n .

Fibonacci spotyka Banacha

Jarosław GÓRNICKI*

* Politechnika Rzeszowska

Fibonacci (właściwie Leonardo z Pizy, ok. 1170–1240) nauczył się zasad arytmetyki hindusko-arabskiej, gdy razem z ojcem przebywał w Bougie (obecnie algierska Bidżaja). Poszerzał swoją wiedzę podczas podróży do Egiptu, Syrii, Grecji, na Sycylię, do Prowansji. Gdy osiadł w Pizie, w 1202 roku napisał traktat *Liber Abaci* (Księga rachunków), z myślą o rozpowszechnieniu w Europie notacji dziesiętnej opartej na wykorzystaniu cyfr $0, 1, 2, \dots, 9$. Pokazał w nim użyteczność nowych metod na wielu przykładach rachunkowych, szczególnie związanych z przeliczaniem miar i wag, obliczaniem zysków i odsetek, wymianą pieniędzy. W 1225 roku napisał rozprawę *Liber Quadratorum*, która miała być kontynuacją *Arytmetyki* Diofantosa. Problemy rozwiązywane przez Leonarda nie są banalne. Przekonujemy się o tym, rozwiązując zadania zamieszczone w jego tekstach:

1. Znaleźć liczbę wymierną x taką, że liczby $x - 5$, x oraz $x + 5$ są kwadratami.
[Najmniejszym rozwiązaniem jest: $x = \frac{1681}{144} = \left(\frac{41}{12}\right)^2$,
 $x - 5 = \frac{961}{144} = \left(\frac{31}{12}\right)^2$, $x + 5 = \frac{2401}{144} = \left(\frac{49}{12}\right)^2$.]
2. Wybrać pięć odważników tak, aby można było na wadze szalkowej zważyć każdy ładunek o masie $1, 2, 3, \dots, 30$, jeżeli podczas ważenia odważniki można układać tylko na jednej szalce.
[Odpowiedź: $1, 2, 4, 8, 16$.]
3. Wybrać możliwie najmniej odważników tak, aby na wadze szalkowej można było zważyć każdy ładunek o masie $1, 2, 3, \dots, 30$.
[Odpowiedź: $1, 3, 9, 27$.]
4. Ile będzie po roku par królików, które urodzą się jako potomstwo jednej pary, jeśli każda para wydaje na świat co miesiąc nową parę, zdolną z kolei po miesiącu do rozmnażania, i jeśli żadna para w tym czasie nie ginie?
[Odpowiedź: 377 .]