

Anna Bartkowiakowa (Wrocław)

PROCESY STOCHASTYCZNE W BIOLOGII I MEDYCYNIE

Wstęp

W niniejszym artykule chcę dać przegląd rozmaitych zagadnień biologicznych, które mogą być sformułowane w terminach procesów stochastycznych. Zagadnienia te wynikły z rozważań konkretnych sytuacji przyrodniczych, które stały się bodźcem do tworzenia różnych modeli matematycznych, opisujących w sposób abstrakcyjny pewną rzeczywistość. Niektóre zagadnienia (np. rozkład liczby gatunków w obrębie rodzaju, zagadnienia epidemiologiczne) zostały postawione kilkadziesiąt lat temu, kiedy nie istniała jeszcze teoria procesów stochastycznych, inne (np. różne teorie wiążące się ze zjawiskiem mutagenezy) powstały na tle nowszych osiągnięć nauk przyrodniczych.

Jest rzeczą bardzo trudną wskazać na rolę i istotne znaczenie procesów stochastycznych w ogólnym rozwoju wiedzy przyrodniczej. Mówiąc ogólnie, metody te pozwalają na ilościowy opis zjawisk i sprawdzanie pewnych hipotez co do modelu opisującego dane zjawisko.

Modele, konstruowane przez matematyka, są i pozostaną tylko pewnym przybliżeniem rzeczywistości, która jest zawsze bardziej różnorodna i złożona. Stąd też bardzo ważną rzeczą jest konfrontacja założeń danego modelu z rzeczywistością, jak również umiejętność wydobycia najważniejszych faktów, które charakteryzują dane zjawisko.

W artykule niniejszym chcę pokazać pewne zagadnienia i sposoby ich opisu. Z konieczności nie są to kompletne dyskusje, ale

jak gdyby pierwowzory, które można z uwzględnieniem konkretnej sytuacji dalej rozwijać i dyskutować.

Spis literatury jest również bardzo skromny. Spośród monografii, związanych tematycznie z tytułem niniejszego artykułu można wymienić zbiory referatów [30] i [33] wygłoszonych na sympozjach poświęconych tym zagadnieniom. Dobrym przeglądem rozmaitych prac z tej dziedziny z przeszło 300 referencjami do innych prac jest monografia rumuńska [16]. Wiele tych zagadnień znaleźć można też w książkach Bailey'a [3] i Bartletta [4].

1. Podstawowe metody procesów stochastycznych

Zjawiska przyrodnicze - takie jak wzrost i rozwój hodowli komórek, zwalczanie bakterii przez antybiotyki, rozprzestrzenianie się zaraźliwych chorób - są zjawiskami zachodzącymi w czasie. Najbardziej precyzyjny opis takich zjawisk to procesy stochastyczne. Weźmy pod uwagę ilość komórek danej hodowli. Ilość ta jest zmienną losową, której wielkość jest funkcją czasu. Interesującą dla nas rzeczą jest wiedzieć, jakie jest prawdopodobieństwo, że w chwili t zaobserwowana liczba komórek, którą oznaczать będziemy przez $x(t)$, będzie wynosić dokładnie n . Prawdopodobieństwo to będziemy oznaczać następująco:

$$\Pr\{x(t) = n\} = P_n(t), \quad \text{gdzie } n = 0, 1, 2, \dots$$

Mówiąc ogólnie, będziemy rozpatrywać pewien układ (populację), dla którego w każdej chwili t jest określona pewna zmienna losowa $x(t)$, przyjmująca wartości całkowite nieujemne $0, 1, 2, \dots$

Zmienną losową $x(t)$ możemy interpretować jako ilość zachorowań do chwili t , ilość rozmów telefonicznych przeprowadzonych w danej instytucji do chwili t , ilość pacjentów zgłoszonych w rejestracji przychodni do chwili t , ilość porad udzielonych przez lekarza od rozpoczęcia jego dyżuru, itp.

W rozdziałach 2-7 będziemy opisywać takie zdarzenia, w których $x(t)$ jest niemalejącą funkcją czasu, tj. $x(t_1) \leq x(t_2)$, gdy $t_1 \leq t_2$. Są to ogólnie mówiąc, tzw. procesy urodzin.

W dalszych rozważaniach zajmiemy się również tzw. procesami urodzin i śmierci, w których $x(t)$ rozpatrywane jako funkcja czasu może nie tylko wzrastać, ale również maleć. Bardzo często interesuje nas szczególnie $P_0(t)$ czyli prawdopodobieństwo wymarcia (wy-

ginięcia) hodowli. Jeśli np. rozważaną hodowlą są bakterie chorobotwórcze, na które działamy antybiotykiem, to jest dla nas mniej interesujące, czy po pewnym czasie t , w którym stosowaliśmy pewną terapię będzie jeszcze 5 czy też 15 żywych bakterii, ale ważną jest rzeczą, czy zwalczane przez nas bakterie jeszcze w ogóle istnieją - bo jeśli tak, to w szybkim tempie się rozmnożą i właściwie nasze działanie nie zostało uwieńczone pozytywnym rezultatem.

Często interesuje nas tzw. wartość oczekiwana rozpatrywanej zmiennej losowej, inaczej wartość średnia, zdefiniowana wzorem

$$(1.1) \quad M(t) = Ex(t) = \sum_{k=0}^{\infty} kP_k(t),$$

oraz wariancja zmiennej losowej $x(t)$ czyli wielkość

$$(1.2) \quad D^2x(t) = \sum_{k=0}^{\infty} k^2P_k(t) - [Ex(t)]^2.$$

Dla wyjaśnienia tych pojęć posłużymy się następującym przykładem. Przypuśćmy, że obserwujemy równoległe dużą ilość próbek z taką samą ilością komórek w początkowym momencie obserwacji. Hodowlę tę kultywujemy w identycznych warunkach. Ilość żywych komórek w jednej próbce rozpatrywana jako funkcja czasu, to jak gdyby jedna realizacja procesu urodzin i śmierci (patrz rozdz. 8). Gdybyśmy takich próbek mieli dużą ilość, powiedzmy N , wtedy rachując ilość żywych komórek w chwili t , powinniśmy w przybliżeniu otrzymać w $NP_1(t)$ próbkach po jednej żywej komórce, w $NP_2(t)$ próbkach po dwie żywe komórki itp., a w $NP_0(t)$ próbkach nie powinno w ogóle być żadnej żywej komórki. Frakcje próbek z k żywymi komórkami powinny być tym bliższe prawdopodobieństwa $P_k(t)$, im większe N . Za oszacowanie oczekiwanej liczby komórek w chwili t , czyli wielkości $Ex(t) = M(t)$ może służyć średnia arytmetyczna zaobserwowanych liczb komórek w poszczególnych próbkach.

Metodą, za pomocą której w wielu wypadkach dochodzimy do wzorów na $P_n(t)$ będących explicite funkcją czasu, jest metoda równań różniczkowych opartych na badaniu prawdopodobieństwa pojawienia się nowego elementu w bardzo małym odcinku czasu $[t, t + \Delta t)$. Zmierzając z Δt do zera otrzymujemy równania różniczkowe, których rozwiązania są poszukiwanymi funkcjami $P_n(t)$. Metodę tę stosujemy przy tzw. procesach urodzin, w których nowe komórki (cząsteczki) mogą powstać, ale nie mogą ginąć. Modele takie są rozważane w rozdziałach 2-7 niniejszej pracy. Równania, jakie tam otrzymujemy, są

rekurencyjne, to znaczy znając $P_{n-1}(t)$ potrafimy wyznaczyć $P_n(t)$. Jeśli rozpatrywać procesy, w których cząsteczki nie tylko powstają, ale i giną lub ulegają mutacji (tzw. procesy urodzin i śmierci), to na ogół dość łatwo napisać równania różniczkowe, które powinny być spełnione przez poszukiwane funkcje $P_n(t)$. Jednak nie są to już równania rekurencyjne i dla wyznaczenia $P_n(t)$ trzeba znać nie tylko $P_{n-1}(t)$, ale i $P_{n+1}(t)$; inaczej mówiąc, trzeba rozwiązywać cały układ równań jednocześnie. Rozwiązywanie takiego układu równań, na ogół nieskończonego, nie jest bynajmniej rzeczą łatwą. Często w tych warunkach pewne rezultaty może dać metoda funkcji tworzących (por. [10], str. 209 i [12], str. 115). Są to funkcje pomocniczego parametru z oznaczane najczęściej dużymi greckimi literami, np. $\Phi(z, t)$, $\pi(z, t)$, $\psi(z, t)$. Funkcje te definiuje się przez szereg, w którym szukane prawdopodobieństwa $P_n(t)$ są współczynnikami przy kolejnych potęgach z :

$$(1.3) \quad \Phi(z, t) = P_0(t) + P_1(t)z + P_2(t)z^2 + P_3(t)z^3 + \dots + P_n(t)z^n + \dots$$

Jeśli we wzorze powyższym podstawimy $z = 0$, to otrzymamy

$$\Phi(0, t) = P_0(t) + P_1(t) \cdot 0 + P_2(t) \cdot 0 + \dots = P_0(t).$$

Widzimy więc, że wartość funkcji tworzącej w punkcie $z = 0$ daje nam prawdopodobieństwo $P_0(t)$, co możemy zapisać inaczej tak:

$$\Phi(z, t) \Big|_{z=0} = P_0(t).$$

Obliczymy teraz kolejne pochodne funkcji $\Phi(z, t)$. Otrzymujemy kolejno

$$\frac{\partial \Phi}{\partial z} = 0 + P_1(t) + 2 \cdot P_2(t)z + 3 \cdot P_3(t)z^2 + \dots + nP_n(t)z^{n-1} + \dots$$

$$\frac{\partial^2 \Phi}{\partial z^2} = 0 + 0 + 2 \cdot P_2(t) + 3 \cdot 2 \cdot P_3(t)z + \dots + n(n-1)P_n(t)z^{n-2} + \dots$$

.....

$$\frac{\partial^n \Phi}{\partial z^n} = 0 + 0 + 0 + \dots \quad + 0 + n(n-1) \cdot \dots \cdot 2 \cdot 1 \cdot P_n(t) + \dots$$

.....

Jeśli weźmiemy teraz wartości tych pochodnych w punkcie $z = 0$ otrzymamy

$$\begin{aligned} \left. \frac{\partial \Phi}{\partial z} \right|_{z=0} &= P_1(t) \\ \left. \frac{\partial^2 \Phi}{\partial z^2} \right|_{z=0} &= 2P_2(t) \\ &\dots\dots\dots \\ \left. \frac{\partial^n \Phi}{\partial z^n} \right|_{z=0} &= n! \cdot P_n(t) \\ &\dots\dots\dots \end{aligned}$$

Tak więc, aby otrzymać wyrażenie na $P_n(t)$, należy zróżniczkować $\Phi(z, t)$ n razy względem argumentu z , podstawić w otrzymanym w ten sposób wyrażeniu $z = 0$, a otrzymany wynik podzielić przez $n!$:

$$(1.4) \quad P_n(t) = \frac{1}{n!} \left. \frac{\partial^n \Phi(z, t)}{\partial z^n} \right|_{z=0}$$

W analogiczny sposób przy znanej funkcji tworzącej można otrzymać wyrażenie na oczekiwaną wartość $Ex(t)$ oraz wariancję $D^2x(t)$. A mianowicie

$$(1.5) \quad Ex(t) = M(t) = \left. \frac{\partial \Phi}{\partial z} \right|_{z=1}$$

$$(1.6) \quad D^2x(t) = \left. \left\{ \frac{\partial^2 \Phi(z, t)}{\partial z^2} + \frac{\partial \Phi(z, t)}{\partial z} - \left[\frac{\partial \Phi(z, t)}{\partial z} \right]^2 \right\} \right|_{z=1}$$

lub

$$(1.6a) \quad D^2x(t) = \left. \frac{\partial}{\partial z} \left[z \cdot \frac{\partial \Phi(z, t)}{\partial z} \right] \right|_{z=1} - [Ex(t)]^2$$

Z układu równań różniczkowych dla danego procesu jest dość łatwo otrzymać jedno równanie o pochodnych cząstkowych względem funkcji tworzącej. Jednak rozwiązywanie równań różniczkowych o pochodnych cząstkowych jest na ogół rzeczą trudną. Jeśli chodzi o wartość średnią $M(t)$ i wariancję $D^2x(t)$ to często można na podstawie układu równań różniczkowych dla danego procesu napisać jedno równanie różniczkowe zwyczajne na funkcję $M(t)$ i rozwiązując to równanie otrzymać wyrażenie na $M(t)$ z pominięciem funkcji tworzącej danego procesu.

2. Proces Poissona

Proces ten opisuje takie zjawiska jak: rozpad promieniotwórczy, nadejścia sygnałów do centrali telefonicznej, zmiany w chromosomach wywołane naświetlaniem promieniami X (por. [10]). Ogólnie mówiąc, obserwujemy liczbę zajęć pewnego zdarzenia w pewnym określonym odcinku czasu. Liczba ta może przyjmować wartości ze zbioru liczb całkowitych nieujemnych. Dla ustalenia uwagi będziemy mówić w dalszym ciągu o zachorowaniach na jakąś chorobę A w pewnej populacji. Oznaczmy przez $P_n(t)$ prawdopodobieństwo tego, że w czasie $[0, t)$ wystąpi dokładnie n zachorowań. Robimy następujące założenia o obserwowanym procesie:

I. Ilości zachorowań w rozłącznych prawostronnie otwartych przedziałach czasowych $[t_1, t_2)$ i $[t_2, t_3)$ są niezależnymi zmiennymi losowymi. Warunek ten charakteryzuje tzw. proces o przyrostach niezależnych.

II. Prawdopodobieństwo pojawienia się określonej liczby zachorowań w różnych przedziałach czasowych o jednakowej długości jest takie samo i zależy jedynie od długości tych przedziałów. Warunek ten charakteryzuje tzw. proces o przyrostach jednorodnych.

III. Zachodzą następujące relacje:

$$(a) \quad \lim_{t \rightarrow 0} \frac{P_1(t)}{t} = \lambda \quad (\lambda > 0)$$

$$(b) \quad \lim_{t \rightarrow 0} \frac{1 - P_0(t) - P_1(t)}{t} = 0$$

Jaki jest sens tych założeń w odniesieniu do naszego przykładu zachorowań na chorobę A?

Warunek I mówi, że pojawienie się zachorowań w okresie $[t_1, t_2)$ nie wpływa na prawdopodobieństwo zachorowań w okresie $[t_2, t_3)$. Jeśli zachorowania w dwóch rozłącznych przedziałach czasowych są niezależne, to prawdopodobieństwo zdarzenia polegającego np. na wystąpieniu w pierwszym okresie k zachorowań, a w drugim okresie m zachorowań jest równe iloczynowi prawdopodobieństwa, że w pierwszym okresie wystąpi k zachorowań, przez prawdopodobieństwo, że w drugim okresie wystąpi m zachorowań. Warunek ten nie byłby speł-

niony, gdyby choroba A była chorobą zaraźliwą; wtedy bowiem pojawienie się jakiegoś zachorowania zwiększa prawdopodobieństwo pojawienia się dalszych zachorowań.

Warunek II mówi, że prawdopodobieństwo pojawienia się pewnej ilości zachorowań jest takie samo np. w pierwszym i szóstym miesiącu obserwacji. Znaczy to, że badana przez nas populacja pozostaje jak gdyby przez cały czas w tych samych warunkach powodujących powstawanie tej choroby. Wyklucza się tu więc wszelki wpływ sezonowy. Warunek ten nie byłby spełniony, gdyby rozważana przez nas populacja pozostawała wprawdzie w tych samych warunkach, sprzyjających tej chorobie, ale za to licznosc populacji zwiększała się dość znacznie z czasem; wtedy w miarę upływu czasu należy się spodziewać coraz większej ilości zachorowań.

Warunek III jest jak gdyby pewną liczbową charakterystyką omawianego przez nas procesu. Wprowadza się tu parametr λ , czyli tzw. intensywność zachorowań. W naszym przykładzie możemy interpretować λ jako średnią liczbę zachorowań w jednostce czasu (por. wzór (2.3)). Warunek III(a) mówi, że prawdopodobieństwo wystąpienia jednego zachorowania w małym przedziale czasowym o długości Δt jest równe $\lambda \cdot \Delta t + o(\Delta t)$, natomiast z warunku III (b) wynika, że prawdopodobieństwo wystąpienia dwu lub więcej zachorowań jest rzędu $o(\Delta t)$; przy czym $o(\Delta t)$ oznacza wielkość, która przy $\Delta t \rightarrow 0$ dąży do zera szybciej niż Δt , czyli $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$. Z warunków III(a) i III(b) będziemy korzystali przy obliczaniu prawdopodobieństwa wystąpienia jednego i więcej oraz niewystąpienia żadnego zachorowania w bardzo małym odcinku czasu Δt , przy $\Delta t \rightarrow 0$. Opierając się na tych warunkach będziemy przyjmowali, że w przybliżeniu prawdopodobieństwo wystąpienia jednego zachorowania w czasie Δt jest wprost proporcjonalne do długości rozpatrywanego przedziału Δt . Stała λ jest współczynnikiem proporcjonalności. Wskazuje ona na intensywność procesu. Prawdopodobieństwo wystąpienia więcej niż jednego zachorowania w czasie Δt jest rzędu $o(\Delta t)$, czyli praktycznie równe zero, wobec tego tej możliwości nie będziemy uwzględniali.

Przy powyższych założeniach możemy napisać równania różniczkowe, rozwiązania których dadzą nam explicite szukane prawdopodobieństwa jako funkcje czasu. Weźmy pod uwagę dwa sąsiednie momenty czasowe t i $t + \Delta t$. Jeśli w chwili $t + \Delta t$ jest 0 zachorowań, to w chwili t też musiało być 0 zachorowań, a to znaczy, że od chwili t do chwili $t + \Delta t$ nie wystąpiła żadna zmiana. Wobec tego,

oznaczając przez $P_0(t)$ prawdopodobieństwo zdarzenia, że od chwili 0 do chwili t było 0 zachorowań, otrzymujemy następujący związek między obu prawdopodobieństwami:

$$P_0(t + \Delta t) = P_0(t)(1 - \lambda \cdot \Delta t).$$

Zakładając, że $\Delta t \neq 0$, dzielimy tę równość obustronnie przez Δt i otrzymujemy:

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = - \frac{\lambda \cdot \Delta t \cdot P_0(t)}{\Delta t}$$

Przy Δt dążącym do zera, lewa strona powyższej równości dąży do pochodnej funkcji $P_0(t)$ względem czasu, czyli

$$\lim_{\Delta t \rightarrow 0} \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = \frac{d P_0(t)}{dt},$$

natomiast

$$\lim_{\Delta t \rightarrow 0} \frac{\lambda \cdot \Delta t \cdot P_0(t)}{\Delta t} = \lambda \cdot P_0(t).$$

Tak więc po przejściu z Δt do zera otrzymujemy następujące równanie różniczkowe na nieznanne prawdopodobieństwo $P_0(t)$:

$$(2.1) \quad \frac{d P_0(t)}{dt} = -\lambda P_0(t)$$

Rozwiązując to równanie przez rozdzielenie zmiennych otrzymujemy kolejno

$$\frac{d P_0}{P_0} = -\lambda dt,$$

$$\ln P_0(t) - \ln P_0(0) = -\lambda t,$$

$$P_0(t) = P_0(0) \cdot e^{-\lambda t}$$

Jeśli wiadomo, że w chwili $t = 0$ było 0 zachorowań, czyli $P_0(0) = 1$ (warunek początkowy), to otrzymujemy ostatecznie:

$$P_0(t) = e^{-\lambda t}$$

W podobny sposób można wyznaczyć $P_n(t)$, czyli prawdopodobieństwo zdarzenia, że od chwili 0 do chwili t wystąpi n zachorowań. Rozważmy znowu dwa sąsiednie momenty czasowe t i $t + \Delta t$. Jeśli do momentu $t + \Delta t$ wystąpiło n zmian, to w chwili t mogła być aktualna jedna z następujących możliwości: 1^o w chwili t było już n zachorowań i żadne dalsze nie wystąpiły; 2^o w chwili t było $n - 1$

zachorowań, a jedno wystąpiło w czasie między t a $t + \Delta t$. Pozostałe możliwości mają prawdopodobieństwa dążące wraz z Δt do zera. Nie uwzględniamy ich tu, gdyż i tak by znikły przy obliczaniu granicy. Wobec powyższego $P_n(t)$ i $P_{n-1}(t)$ są związane następującymi zależnościami:

$$P_n(t + \Delta t) = P_n(t)(1 - \lambda \cdot \Delta t) + P_{n-1}(t) \cdot \lambda \cdot \Delta t. \quad (2.1)$$

Przenosząc $P_n(t)$ na lewą stronę równania, oraz dzieląc otrzymaną równość obustronnie przez Δt ($\Delta t \neq 0$), otrzymujemy:

$$\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = -\lambda P_n(t) + \lambda P_{n-1}(t).$$

Przechodząc z Δt do zera otrzymujemy w granicy

$$(2.1a) \quad \frac{d P_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t) \quad \text{dla } n \geq 1.$$

Jest to rekurencyjny układ równań różniczkowych. Ponieważ znamy $P_0(t)$, możemy wyznaczyć $P_1(t)$ podstawiając w powyższym równaniu $n = 1$:

$$\frac{d P_1(t)}{dt} = -\lambda P_1(t) + \lambda P_0(t).$$

Jest to równanie różniczkowe liniowe. Sposób rozwiązywania takich równań jest podany np. w [32]. Jako rozwiązanie powyższego równania (podstawiając $P_0(t) = e^{-\lambda t}$) otrzymujemy: $P_1(t) = \lambda t e^{-\lambda t}$ (przy warunku początkowym $P_1(0) = 0$). Mając $P_1(t)$ możemy wyznaczyć $P_2(t)$ itd. Ogólnie przy warunkach początkowych

$$P_k(0) = \begin{cases} 1 & \text{gdy } k = 0 \\ 0 & \text{gdy } k \neq 0 \end{cases}$$

otrzymujemy następujące rozwiązanie układu równań (2.1) i (2.1a):

$$(2.2) \quad P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad \text{dla } n = 0, 1, 2, \dots$$

Oznaczmy przez $M(t)$ oczekiwaną liczbę zachorowań od chwili 0 do chwili t . Z definicji $M(t) = \sum_{n=0}^{\infty} n \cdot P_n(t)$. Podstawiając na $P_n(t)$ odpowiednie wyrażenia według wzoru (2.2) otrzymujemy:

$$(2.3) \quad M(t) = \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \lambda t.$$

Wynika stąd, że λ możemy interpretować jako oczekiwaną ilość zachorowań występujących w jednostce czasu.

Znając $P_n(t)$ możemy również obliczyć funkcję tworzącą procesu. Mamy

$$(2.4) \quad \Phi(z, t) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \cdot z^n = e^{-\lambda t} \cdot \sum_{n=0}^{\infty} \frac{(\lambda t z)^n}{n!} = \\ = e^{-\lambda t} \cdot e^{\lambda t z} = e^{\lambda t(z-1)}.$$

Gdybyśmy nie znali wyrażenia na $P_n(t)$, moglibyśmy otrzymać wyrażenie na funkcję tworzącą bezpośrednio z układu równań (2.1) i (2.1a). Mianowicie, mnożąc każde z tych równań odpowiednio przez z^n , otrzymujemy:

$$\frac{d P_0(t)}{dt} \cdot z^0 = -\lambda P_0(t) \cdot z^0$$

$$\frac{d P_n(t)}{dt} \cdot z^n = -\lambda P_n(t) \cdot z^n + \lambda P_{n-1}(t) \cdot z^n \quad \text{dla } n = 1, 2, \dots$$

Sumując powyższe równości stronami dla $n = 0, 1, 2, \dots$ otrzymujemy:

$$\sum_{n=0}^{\infty} \frac{d P_n(t)}{dt} \cdot z^n = -\lambda \sum_{n=0}^{\infty} P_n(t) \cdot z^n + \lambda \sum_{n=0}^{\infty} P_n(t) \cdot z^{n+1} = \\ = (\lambda z - \lambda) \cdot \sum_{n=0}^{\infty} P_n(t) \cdot z^n.$$

Alte

$$\sum_{n=0}^{\infty} \frac{d P_n(t)}{dt} \cdot z^n = \frac{d}{dt} [P_0(t) \cdot z^0 + P_1(t) z^1 + \dots] = \\ = \frac{\partial}{\partial t} \Phi(z, t).$$

Wobec tego otrzymujemy następujące równanie różniczkowe na $\Phi(z, t)$:

$$(2.5) \quad \frac{\partial \Phi(z, t)}{\partial t} = \lambda(z - 1) \cdot \Phi(z, t).$$

Jeśli założymy warunki początkowe $\Phi(z, 0) = 1$, to rozwiązaniem równania (2.5) jest funkcja

$$\Phi(z, t) = e^{\lambda t(z-1)}$$

Otrzymaliśmy więc wyrażenie (2.4) w inny sposób.

W jaki sposób na podstawie próbki (obserwowanej liczby zachorowań) wyestymować nieznaną wartość parametru λ ?

Pokażemy tu poszukiwanie estymatora parametru λ metodą największej wiarygodności. Przypuśćmy, że obserwowaliśmy nasz proces w kilku przedziałach czasowych:

w przedziale $[0, t_1)$ zaobserwowaliśmy n_1 zachorowań;
 w przedziale $[t_1, t_2)$ zaobserwowaliśmy n_2 zachorowań;

 w przedziale $[t_{i-1}, t_i)$ zaobserwowaliśmy n_i zachorowań.

Ponieważ nasz proces jest jednorodny w czasie, prawdopodobieństwo, że w przedziale czasowym $[t_{i-1}, t_i)$ wystąpi n_i zachorowań jest takie samo, jak prawdopodobieństwo, że wystąpi n_i zachorowań w przedziale czasowym $[0, T_i)$ dla

$$T_i = t_i - t_{i-1}.$$

Prawdopodobieństwo zaobserwowania wskazanych liczb zachorowań przy intensywności λ równa się

$$L = \frac{(\lambda T_1)^{n_1}}{n_1!} e^{-\lambda T_1} \cdot \frac{(\lambda T_2)^{n_2}}{n_2!} e^{-\lambda T_2} \cdot \dots \cdot \frac{(\lambda T_i)^{n_i}}{n_i!} e^{-\lambda T_i} =$$

$$= \frac{\lambda^{n_1+n_2+\dots+n_i} \cdot T_1^{n_1} \cdot T_2^{n_2} \cdot \dots \cdot T_i^{n_i} \cdot e^{-\lambda(T_1+T_2+\dots+T_i)}}{n_1! \cdot n_2! \cdot \dots \cdot n_i!}.$$

Prawdopodobieństwo to, rozważane jako funkcja zmiennej λ przy ustalonych $n_1, n_2, \dots, n_i, T_1, T_2, \dots, T_i$, bywa nazywane wiarygodnością. Znajdziemy teraz taką wartość λ , dla której wiarygodność L , a zarazem i logarytm naturalny $\ln L$ przyjmuje wartość największą. Jest to wartość, dla której zeruje się pierwsza pochodna $\ln L$. Obliczając logarytm naturalny $\ln L$ i różniczkując następnie względem λ , otrzymujemy:

$$\ln L = (n_1 + n_2 + \dots + n_i) \cdot \ln \lambda - (T_1 + T_2 + \dots + T_i) \cdot \lambda +$$

$$+ \ln(T_1^{n_1} \cdot T_2^{n_2} \cdot \dots \cdot T_i^{n_i}) - \ln(n_1! \cdot n_2! \cdot \dots \cdot n_i!),$$

$$\frac{d \ln L}{d \lambda} = (n_1 + n_2 + \dots + n_i) \frac{1}{\lambda} - (T_1 + T_2 + \dots + T_i).$$

Pochodna ta przyjmuje wartość zero, gdy

$$(2.6) \quad \lambda = \frac{n_1 + n_2 + \dots + n_i}{T_1 + T_2 + \dots + T_i}$$

Zatem estymatorem parametru λ , czyli intensywności naszego procesu jest stosunek ogólnej liczby zaobserwowanych zachorowań do łącznego czasu obserwacji.

3. Czysty proces urodzin

Przedstawimy teraz uogólnienie omawianego w poprzednim rozdziale procesu Poissona. Zmienną losową $x(t)$ będzie liczba zajęć pewnego zdarzenia w przedziale czasowym od zera do t . Zmienna ta podobnie jak poprzednio będzie niemalejącą funkcją czasu, a jej możliwymi wartościami będą liczby całkowite nieujemne.

Wzory na prawdopodobieństwa $P_n(t) = \Pr\{x(t) = n\}$ wyprowadzimy przy niezmiennionych założeniach I i II poprzedniego rozdziału, to znaczy, że rozważany proces będzie procesem o przyrostach niezależnych i jednorodnych w czasie. Zmienimy natomiast warunek III, czyli liczbową charakterystykę procesu. W procesie Poissona intensywność nie zależała od tego, ile zajęć zdarzenia nastąpiło do czasu t . Teraz o intensywności procesu uczynimy następujące założenia:

III'. Jeśli w chwili t zmienna losowa $x(t)$ wynosiła n ($n = 0, 1, 2, \dots$), wówczas prawdopodobieństwo, że w ciągu czasu $[t, t + \Delta t)$ nastąpi zmiana w jej wartości o 1 (czyli $x(t + \Delta t)$ przybierze wartość $n + 1$) wynosi $\lambda_n \cdot \Delta t + o(\Delta t)$; prawdopodobieństwo więcej niż jednej zmiany w czasie Δt zmierza do zera szybciej niż Δt .

Wobec tych założeń prawdopodobieństwo, że nie będzie żadnej zmiany w czasie $(t, t + \Delta t)$, wynosi $1 - \lambda_n \cdot \Delta t + o(\Delta t)$.

Jeśli zmienną losową $x(t)$ będziemy interpretować jako ilość bakterii w danej hodowli, to warunek III' oznacza, że prawdopodobieństwo "urodzin" nowej bakterii (czyli podziału jakiejś bakterii) w czasie $[t, t + dt)$ zależy jedynie od ilości bakterii w chwili t , nie zależy natomiast od tego, co działo się z bakteriami tej hodowli przed czasem t . Założenie to (markowskość procesu) nie jest słuszne w odniesieniu do jednej bakterii; jeśli bowiem bakteria ta w pewnym momencie podzieliła się, to nie może się od razu dzielić po raz wtóry. Jeśli jednak rozpatrujemy dużą populację bakterii, to przyjmujemy, że takich bakterii, które podzieliły się bezpośrednio przed czasem t jest mało w stosunku do ogólnej liczby bakterii i wobec tego ich wpływ na badane prawdopodobieństwo podziału jednej z n bakterii jest jedynie rzędu $o(dt)$.

Przy powyższych założeniach, postępując podobnie jak w poprzednim rozdziale, otrzymujemy następujące równania różniczkowe, w których należy wyznaczyć $P_n(t)$ jako funkcję czasu:

$$(3.1) \quad \begin{cases} \frac{dP_0(t)}{dt} = -\lambda_0 P_0(t) \\ \frac{dP_n(t)}{dt} = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t) \quad \text{dla } n \geq 1. \end{cases}$$

Jest to rekurencyjny układ równań różniczkowych liniowych. Równanie na $P_0(t)$ jest tzw. równaniem różniczkowym o rozdzielonych zmiennych. Równania te rozwiązujemy sukcesywnie: najpierw znajdziemy $P_0(t)$, następnie znając $P_0(t)$ możemy wyznaczyć $P_1(t)$ itd.

Przy warunkach początkowych $P_0(0) = 1$, $P_n(t) = 0$ dla $n \neq 0$, oraz zakładając, że żadne λ_n nie są sobie równe, otrzymujemy rozwiązanie w następującej postaci (por. [8], str. 406):

$$(3.2) \quad \begin{cases} P_0(t) = e^{-\lambda_0 t} \\ P_n(t) = (-1)^n \cdot \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_{n-1} \sum_{k=1}^n e^{-\lambda_k t} \{(\lambda_k - \lambda_1) \dots \\ \dots (\lambda_k - \lambda_{k-1})(\lambda_k - \lambda_{k+1}) \dots (\lambda_k - \lambda_n)\}^{-1} \quad \text{dla } n \geq 1 \end{cases}$$

Przy pomocy metod, które omawialiśmy dotychczas, nie można uzyskać explicite wyrażenia na funkcję tworzącą tego procesu oraz wartość oczekiwaną i wariancję zmiennej $x(t)$ przy dowolnym układzie (ciągu) $\{\lambda_n\}$. Udaje się to niekiedy, gdy układowi $\{\lambda_n\}$ nadamy pewną szczególną postać np. $\lambda_n = \lambda \cdot n$ (proces Yule'a), $\lambda_n = -\frac{1+nd}{1+td}$, $d > 0$ (proces Polya'i), $\lambda_n = \lambda N(t)$ (proces zachorowań na

rzadko spotykaną chorobę przy zmieniającej się liczbie ludności $N(t)$ podlegającej ryzyku zachorowania), $\lambda_n = \beta n(N+1-n)$ (proces czystej epidemii w populacji liczącej N osób).

W dalszych rozdziałach omówimy dokładniej te szczególne przypadki czystego procesu urodzin.

4. Proces Furry'ego-Yule'a

Przypuśćmy, że rozpatrujemy rozmnażającą się populację komórek w warunkach nieograniczonego rozwoju. Oznaczmy przez $x(t)$ ilość komórek w chwili t . Zatrzymując założenia I i II z poprzed-

nich rozdziałów, precyzujemy warunek III¹ następująco. Wszystkie komórki są jednakowe. W czasie od t do $t + \Delta t$ każda z istniejących komórek może się podzielić na dwie nowe z prawdopodobieństwem $\lambda \cdot \Delta t$. Jeśli w chwili t było już n komórek, to prawdopodobieństwo, że w czasie Δt ich liczba wzrośnie o jeden jest wprost proporcjonalne do liczby komórek już istniejących i wynosi $n \cdot \lambda \cdot \Delta t$. Oznaczmy przez $P_n(t)$ prawdopodobieństwo zdarzenia, że $x(t)$, czyli liczba komórek w chwili t wyniesie dokładnie n . Podstawiając $\lambda_n = \lambda \cdot n$ do równań (3.1) otrzymujemy następujący układ równań:

$$(4.1) \quad \begin{cases} \frac{dP_1(t)}{dt} = -\lambda P_1(t) \\ \frac{dP_n(t)}{dt} = -\lambda n \cdot P_n(t) + \lambda(n-1)P_{n-1}(t) \quad \text{dla } n > 1 \end{cases}$$

Jeśli założymy następujące warunki początkowe:

$$P_1(0) = 1, \quad P_n(0) = 0 \quad \text{dla } n \neq 1,$$

otrzymamy następujące rozwiązanie układu równań (4.1):

$$(4.2) \quad P_n(t) = e^{-\lambda t} (1 - e^{-\lambda t})^{n-1} \quad \text{dla } n \geq 1$$

Weźmy teraz pod uwagę jedną komórkę i oznaczmy przez T czas, po którym komórka ta podzieli się. Z (4.2) wynika, iż prawdopodobieństwo podzielenia się komórki przed czasem t wynosi

$$\Pr \{T \leq t\} = 1 - e^{-\lambda t}$$

Gęstość tego rozkładu prawdopodobieństwa wynosi

$$f_T(t) = \frac{d}{dt} \Pr \{T \leq t\} = \lambda e^{-\lambda t}$$

Oczekiwana wartość zmiennej losowej T wynosi wobec tego

$$E(T) = \int_0^{\infty} t \cdot f_T(t) dt = - \int_0^{\infty} t \lambda (-e^{-\lambda t}) dt = \frac{1}{\lambda}$$

Tak więc stałą λ możemy interpretować jako odwrotność średniego czasu upływającego między kolejnymi podziałami komórki. Inaczej jeszcze można powiedzieć, że jest to oczekiwana liczba podziałów komórki w jednostce czasu.

Jeśli obserwujemy daną hodowlę w ciągu pewnego czasu t , to może nas interesować oczekiwana liczba komórek w chwili t , czyli wielkość $M(t) = \sum_{n=0}^{\infty} n P_n(t)$. Możemy uzyskać ją w następujący spo-

sób. Każde z równań różniczkowych (3.1) mnożymy przez odpowiadającą mu n i sumujemy od $n = 1$ do $n = \infty$. Otrzymamy wtedy następujące równanie różniczkowe na szukaną funkcję $M(t)$:

$$\frac{dM(t)}{dt} = \lambda \cdot M(t).$$

Stąd przy warunku początkowym $M(0) = 1$ otrzymujemy:

$$(4.3) \quad M(t) = e^{\lambda t}.$$

W podobny sposób można znaleźć wyrażenie na wariancję liczby komórek w chwili t . Wynosi ona

$$(4.4) \quad D^2x(t) = e^{2\lambda t} - e^{\lambda t}.$$

Furry stosował omawiany proces przy opisie promieniowania kosmicznego, a Yule [37] zastosował go do wyprowadzenia rozkładu ilości gatunków (species) w obrębie rodzaju (genus) dla różnych rodzin zwierząt i roślin.

5. Proces Polya'i

Rozpatrzmy zachorowania na jakąś zaraźliwą chorobę. Jest rzeczą naturalną spodziewać się, że prawdopodobieństwo nowego zachorowania w chwili t , do której wystąpiło już n zachorowań, jest tym większe im większa jest aktualna liczba zachorowań (bo wtedy jest więcej szans na zarażenie się), ale jest tym mniejsze, im dłuższy czas upłynął od początku epidemii (bo przedsięwzięto odpowiednie środki ostrożności).

Założmy, iż prawdopodobieństwo, że w przedziale czasowym $[t, t + \Delta t)$ wystąpi nowe zachorowanie pod warunkiem, że do chwili t wystąpiło n zachorowań, wynosi

$$(5.1) \quad \lambda_n \cdot \Delta t = \frac{1 + n \cdot d}{1 + td}, \quad \text{gdzie } d > 0.$$

Oznaczmy przez $P_n(t)$ prawdopodobieństwo, że w czasie od 0 do t zaobserwujemy n zachorowań. Prawdopodobieństwa te powinny spełniać następujący układ równań różniczkowych:

$$(5.2) \quad \left\{ \begin{array}{l} \frac{dP_0(t)}{dt} = - \frac{1}{1 + td} \cdot P_0(t) \\ \frac{dP_n(t)}{dt} = - \frac{1 + nd}{1 + td} \cdot P_n(t) + \frac{1 + (n-1)d}{1 + td} \cdot P_{n-1}(t) \quad \text{dla } n \geq 1 \end{array} \right.$$

Przy warunku początkowym, że w chwili $t = 0$ nie było zachorowań, rozwiązaniami układu (5.2) są następujące funkcje:

$$(5.3) \quad \begin{cases} P_0(t) = (1 + td)^{-1/d} \\ P_n(t) = \left(\frac{t}{1 + td} \right)^n \cdot \frac{1(1+d) \dots (1+(n-1)d)}{n!} \cdot P_0(t) \end{cases}$$

dla $n \geq 1$

Wartość oczekiwana i wariancja zmiennej losowej $x(t)$, czyli liczby zachorowań do czasu t , wyrażają się następującymi wzorami:

$$(5.4) \quad \begin{cases} M(t) = \sum_{k=0}^{\infty} k \cdot P_k(t) = t \\ D^2 x(t) = \sum_{k=0}^{\infty} [k - M(t)]^2 \cdot P_k(t) = t(1 + td) \end{cases}$$

Rozkład (5.3) otrzymali po raz pierwszy dla szczególnego przypadku $t = 1$ Eggenberger i Polya [7]. Ogólna postać (5.1) pochodzi od Lundberga [24]. Eggenberger zebrał dane mówiące o ilości śmiertelnych zachorowań na ospę w poszczególnych miesiącach lat 1877-1900 w Szwajcarii. Rozważając pewien schemat urnowy, w którym wyciągnięcie kuli czarnej (zachorowanie) zwiększa prawdopodobieństwo wyciągnięcia kuli czarnej w następnych losowaniach, natomiast wyciągnięcie kuli białej (brak zachorowania) prawdopodobieństwo to zmniejsza, Eggenberger i Polya dopasowali swoje dane do rozkładu (5.3), znajdując parametry tego rozkładu przez przyrównanie momentów teoretycznych z wzorów (5.4) do momentów empirycznych obliczonych na podstawie danych z 288 miesięcy lat 1877-1900.

Eggenberger i Polya uważali swój rozkład za typowy rozkład opisujący zjawiska zaraźliwe, przy których pojawienie się zjawiska u jednego osobnika, zwiększa prawdopodobieństwo pojawienia się tego zjawiska u innych osobników.

Jest rzeczą ciekawą, że ten sam rozkład, jak to pokazał Lundberg [24], można otrzymać przy innych założeniach. Poprzednio zakładaliśmy, że intensywność procesu jest funkcją stanu i czasu trwania procesu. Obecnie pokażemy, że takie same prawdopodobieństwa można otrzymać, nie zakładając zależności intensywności od stanu procesu, a tylko przyjmując, że proces jest mieszaniną jednorodnych procesów Poissona o różnych intensywnościach.

Niech rozważaną przez nas populacją będzie populacja ludzi, którzy ubezpieczyli się od nieszczęśliwego wypadku, zaś zmienną losową $x(t)$ niech będzie ilość nieszczęśliwych wypadków, jakim uległ ubezpieczony w przedziale czasowym $[0, t)$. Danymi takimi dysponuje towarzystwo ubezpieczeniowe. Lundberg porachował frekwencje polis ubezpieczeniowych, na które w określonym czasie wystąpiło n reklamacji i otrzymał w ten sposób rozkład ilości nieszczęśliwych wypadków, jakim uległy osoby ubezpieczone w pewnym określonym czasie. Następnie Lundberg powtórzył pewne rozumowanie Greenwooda i Yule'a w odniesieniu do zwykłych rozkładów prawdopodobieństwa; o każdym człowieku z osobna można założyć, że ulega on wypadkowi spontanicznie, z pewną intensywnością λ , podobnie jak to się dzieje przy procesie Poissona, opisanym w rozdziale 2. Tak więc prawdopodobieństwo, że dany człowiek w ciągu danego okresu ulegnie n razy nieszczęśliwemu wypadkowi, wyraża się wzorem

$$(5.5) \quad P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t},$$

gdzie λ można interpretować jako średnią ilość nieszczęśliwych wypadków w ciągu jednostki czasu. Rzecz cała w tym, że wartość λ nie jest ta sama dla wszystkich ludzi. Jedni są pechowcami i średnia ilość wypadków jest u nich duża; innym na odwrót - udaje się tych nieszczęśliwych wypadków uniknąć i ich λ jest małe. Innymi słowy osoby badanej populacji dzielą się na warstwy ze względu na λ , a więc λ nie jest wielkością stałą dla danej populacji, ale też pewną zmienną losową. Jeśli tak, to ma też swój rozkład prawdopodobieństwa. Można przyjąć, że rozkład λ ma gęstość

$$(5.6) \quad f(\lambda) = \begin{cases} \frac{a^v}{\Gamma(v)} \lambda^{v-1} e^{-a\lambda} & \text{dla } \lambda > 0 \\ 0 & \text{dla } \lambda = 0 \end{cases}$$

przy czym $v > 0$, $a > 0$. Jest to tzw. rozkład gamma lub rozkład Pearsona typu III.

Aby otrzymać prawdopodobieństwo, że losowo z danej populacji wybrany człowiek ulegnie n nieszczęśliwym wypadkom, należy obliczyć prawdopodobieństwo warunkowe, że człowiek ten ulegnie n nieszczęśliwym wypadkom, jeśli pochodzi z warstwy o intensywności λ (wzór (5.5)), pomnożyć to prawdopodobieństwo przez prawdopodobieństwo, że człowiek ten pochodzi z warstwy λ (wzór (5.6)) i następnie zsumować ze względu na wszystkie możliwe λ . W ten sposób ot-

rzymujemy następujące wyrażenia na prawdopodobieństwa złożonego rozkładu Poissona

$$(5.7) \quad P_n(t) = \int_0^{\infty} \frac{(\lambda t)^n}{n!} = e^{-\lambda t} f(\lambda) d\lambda$$

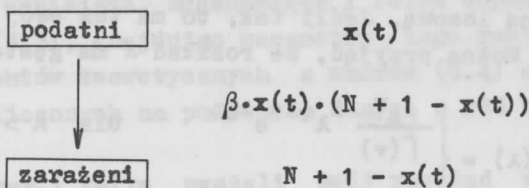
Podstawiając do (5.7) funkcję $f(\lambda)$ określoną wzorem (5.6) i wykonując całkowanie, otrzymujemy

$$(5.8) \quad \begin{cases} P_0(t) = \left(\frac{a}{a+t} \right)^v \\ P_n(t) = \frac{(v)(v+1) \dots \{v+(n-1)\}}{n!} \left(\frac{t}{a+t} \right)^n P_0(t) \quad \text{dla } n \geq 1 \end{cases}$$

Wyrażenia (5.8) są równoważne wyrażeniom (5.3), co łatwo sprawdzić podstawiając $a = v = 1/d$.

6. Proces opisujący prostą epidemię

Przypuśćmy, że w chwili $t = 0$ mamy N osobników, podatnych na zarażenie się określoną chorobą i że do tej grupy wprowadziliśmy jednego osobnika chorego. Choroba przyjmuje się od razu i osobnik nią dotknięty natychmiast staje się zaraźliwy dla otoczenia. Osobnik chory nie jest izolowany, ani nie zdrowieje (nie przestaje być zaraźliwy).



Rys. 1. Schemat prostej epidemii

Założenie, że choroba przeniesiona od razu czyni osobnika zaraźliwym jest oczywiście wyidealizowane, gdyż w rzeczywistości potrzeba pewnego czasu, aby osoba zarażona stała się zarażającą swoje otoczenie. W modelu, który mam zamiar opisać, zakłada się ponadto, że osoba zarażona (zarażająca) pozostaje w normalnych kontaktach z resztą osób jeszcze nie zarażonych.

Założmy, że prawdopodobieństwo zachorowania jednego osobnika

w przedziale czasowym $[t, t + \Delta t)$, gdy jest już $N - n + 1$ zarażonych jest wprost proporcjonalne do Δt , do ilości osób już zarażonych i do ilości osób jeszcze nie zarażonych, czyli równa się $\beta n \cdot (N + 1 - n) \Delta t + o(\Delta t)$. Załóżmy ponadto, że rozpatrywany przez nas proces czyni zadość warunkom I i II z rozdziału 2 (jednorodność w czasie i niezależność przyrostów w rozłącznych przedziałach czasowych). Niech $P_n(t)$ oznacza prawdopodobieństwo, że w chwili t będzie jeszcze n osobników niezarażonych. Prawdopodobieństwa $P_n(t)$ ($n = 0, 1, 2, \dots, N$) spełniają następujący układ równań różniczkowych:

$$(6.1) \quad \begin{cases} \frac{d P_n(t)}{dt} = \beta(n+1)(N-n)P_{n+1} - \beta n(N-n+1)P_n(t) \\ \qquad \qquad \qquad \text{dla } n = 0, 1, 2, \dots, N-1, \\ \frac{d P_N(t)}{dt} = -\beta N P_N(t) \end{cases}$$

Warunek początkowy, wyznaczający jednoznacznie rozwiązanie układu (6.1) jest następujący: $P_0(0) = 1$, $P_n(0) = 0$ dla $n \neq 0$.

Równania (6.1) wyglądają bardzo prosto, ale ich rozwiązanie nie jest wcale proste. Bailey [2], używając transformacji fourierowskich uzyskał explicite rozwiązania, ale są one dość skomplikowane, i dlatego ich tutaj nie przytaczam.

Można pokazać, że gdy $t \rightarrow \infty$, $P_0(t) \rightarrow 1$. Znaczy to, że epidemia z prawdopodobieństwem 1 obejmie populację podatnych.

Bailey podał również metody estymowania parametru β , gdy zaobserwowaliśmy momenty t_1, t_2, \dots, t_N kolejnych zachorowań. Wykazał on, że za estymator wielkości β^{-1} można przyjąć wyrażenie

$$(6.2) \quad \hat{\beta}^{-1} = \sum_{k=1}^N g_k t_k,$$

gdzie $g_k = (1 + k - 1)(N - k + 1)$, a t_k jest to moment, w którym nastąpiło k -te z kolei zachorowanie. Wariancja tego estymatora wynosi $\beta^{-2} N^{-1}$.

Jeśli obserwujemy przebieg epidemii na dwóch grupach indywidualów i gdy dysponujemy danymi podającymi momenty kolejnych zachorowań, to istotność różnic między współczynnikami β , czyli intensywnościami zachorowań dla tych dwu grup, jednej o liczności n_1 i drugiej o liczności n_2 , możemy badać, tworząc iloraz

$$(6.3) \quad F = \frac{\frac{1}{n_1} \sum_{k=1}^{n_1} g_k t_k}{\frac{1}{n_2} \sum_{k=1}^{n_2} g_k t_k}$$

Można wykazać, że licznik i mianownik mają rozkłady χ^2 z $2n_1$ ($i = 1$ dla licznika; $i = 2$ dla mianownika) stopniami swobody. Wobec tego iloraz (6.3) ma rozkład F Snedecora o $2n_1$ i $2n_2$ stopniach swobody.

Jeśli chodzi o średnią ilość niezarażonych osobników w chwili t , to po żmudnych obliczeniach Haskey uzyskał następujący wynik (cytowany za Baileyem [2]):

$$(6.4) \quad M(t) = \sum_{k=0}^N k P_k(t) =$$

$$= \sum_{r=1}^N \frac{N!}{(N-r)!(r-1)!} \left\{ (N - 2r + 1)^2 \beta t + 2 - \right.$$

$$\left. - (N - 2r + 1) \cdot \sum_{u=r}^{N-r} u^{-1} \right\} \cdot e^{-r(N-r+1) \cdot \beta t}$$

przy czym sumowanie po r przebiega od 1 do $\frac{1}{2}N$, gdy N jest parzyste, i od 1 do $\frac{1}{2}(N+1)$, gdy N jest nieparzyste. W tym ostatnim przypadku za współczynnik przy $e^{-\frac{1}{2}(N+1)^2 \beta t}$, tj. dla $r = \frac{1}{2}(N+1)$, należy wziąć $\frac{1}{2}$.

Wzór (6.4) podaje nam co prawda wartość oczekiwaną w dowolnej chwili t , ale w postaci pewnej sumy, która przy dużej liczności populacji jest kłopotliwa w rachunkach. Jeśli jednak liczność populacji jest duża, można rozważany proces opisać w sposób przybliżony za pomocą tzw. modelu deterministycznego, w którym zaniebuje się wpływy losowe i liczbę zachorowań $x(t)$ w pewnej chwili t traktuje się nie jako zmienną losową, ale tylko i wyłącznie jako funkcję rzeczywistą czasu. Funkcja $x(t)$ wynikająca z modelu deterministycznego jest pewnym analogonem wartości oczekiwanej w procesie stochastycznym.

Założmy, że liczba osób podatnych na zachorowanie, którą oznaczać będziemy przez $x(t)$, jest liczbą dużą, tak że w przybliżeniu

możemy ją traktować jako ciągłą funkcję czasu. Do populacji wprowadziliśmy jedną osobę chorą (zarażającą). Podobnie jak w modelu stochastycznym zakładamy, że osoba zarażona staje się od razu zaraźliwą. Wszystkie osoby danej populacji ulegają pewnemu jednorodnemu mieszanemu tak, że każda osoba jeszcze niezarażona ma takie same szanse spotkać się z dowolną z osób zarażających i być zarażoną. Zakładamy, że w chwili $t = 0$ było N osób podatnych, a więc $x(0) = N$. Następnie, w miarę zarażania się coraz to nowych osób, $x(t)$ maleje. Przy tych założeniach można przyjąć, że $\Delta x = x(t+\Delta t) - x(t)$ jest proporcjonalne 1^o do $x(t)$ czyli liczby osób podatnych w chwili t , 2^o do $(N + 1 - x(t))$, czyli liczby osób zarażających w chwili t , 3^o do Δt , czyli długości czasu, dla którego obliczamy ubytek Δx . Oznaczając współczynnik proporcjonalności przez β i zapatrując Δx w znak "-" otrzymujemy następującą przybliżoną równość:

$$(6.5) \quad -\Delta x(t) = \beta x(t)(N - x(t) + 1)\Delta t + o(\Delta t).$$

Dzieląc powyższą równość przez Δt ($\Delta t \neq 0$) i zmierzając z Δt do zera otrzymujemy w granicy równanie różniczkowe

$$(6.6) \quad \frac{d x(t)}{dt} = -\beta x(t) \cdot (N - x(t) + 1).$$

Rozwiązaniem tego równania różniczkowego przy uwzględnieniu warunku początkowego $x(0) = N$ jest funkcja

$$(6.7) \quad x(t) = \frac{N(N + 1)}{N + e^{(N+1)\beta t}}$$

Jest to krzywa, podająca średnią ilość osób podatnych na zachorowanie w chwili t , gdy nie uwzględniamy wahań losowych. Jak wykazał Bailey [2], różni się ona tylko nieznacznie od odpowiadającej jej stochastycznej średniej (6.4), obliczonej przez Haskey'a i posiada punkt przegięcia w tym samym w przybliżeniu miejscu.

Epidemiolodzy bardziej interesują się inną krzywą, ukazującą nasilenie epidemii. Jest to krzywa wskazująca szybkość przyrostów liczby zachorowań. Ilość nowych zachorowań Δy w przedziale czasowym $[t, t + \Delta t)$ równa się ubytkowi funkcji $x(t)$ w tym czasie wziętemu ze znakiem "-".

Mamy więc

$$\Delta y = -\Delta x(t) = \beta x(t)(N + 1 - x(t)).$$

Stąd otrzymujemy równanie różniczkowe na $y(t)$, czyli liczbę osób zarażających w chwili t :

$$(6.8) \quad \frac{dy}{dt} = \beta x(t) \cdot (N + 1 - x(t))$$

Podstawiając za $x(t)$ wyrażenie z wzoru (6.7), otrzymujemy:

$$(6.9) \quad \frac{dy}{dt} = \frac{\beta N(N+1)^2 e^{(N+1)\beta t}}{[N + e^{(N+1)\beta t}]^2}$$

Pochodna dy/dt wyraża szybkość zmiany funkcji $y(t)$. Największa szybkość zmiany występuje tam, gdzie zeruje się pochodna dy/dt .

Łatwo sprawdzić, że następuje to dla $t = \frac{\ln N}{(N+1)}$. W tym momencie występuje największe nasilenie epidemii, gdyż ilości nowych zachorowań są największe.

7. Proces zachorowań na rzadko występującą chorobę przy zmieniającej się liczbie ludności

Przypuśćmy, że badamy wśród populacji polskiej zachorowalność na jakąś rzadką chorobę, np. białaczkę. Przypuśćmy dalej, że dysponujemy rocznymi liczbami zachorowań w poszczególnych powiatach (są one rzędu kilkunastu zachorowań), oraz liczbą ludności tych powiatów w badanych latach.

Oznaczmy przez $y(t)$ liczbę ludności w chwili t . O zachorowaniach zakładamy, że opisują się one procesem o przyrostach niezależnych i jednorodnych w czasie (warunki I i II z rozdziału 2). Prawdopodobieństwo nowego zachorowania w przedziale czasowym o długości Δt jest wprost proporcjonalne do długości tego przedziału czasowego, czyli do Δt oraz do liczby osób $y(t)$ podatnych na tę chorobę. Współczynnik proporcjonalności będziemy oznaczać przez λ . Jest to intensywność naszego procesu. Oznaczając przez $P_1[t, t + \Delta t)$ prawdopodobieństwo wystąpienia jednego nowego zachorowania w przedziale czasowym $[t, t + \Delta t)$, mamy:

$$P_1[t, t + \Delta t) = \lambda \cdot y(t) \cdot \Delta t + o(\Delta t).$$

Oznaczmy przez $P_n(t)$ prawdopodobieństwo, że w przedziale $[0, t)$ wystąpiło n nowych zachorowań. Przy powyższych założeniach prawdopodobieństwo $P_n(t + \Delta t)$ równa się sumie prawdopodobieństw następujących zdarzeń: 1° że w chwili t było już n zachorowań i w następnym przedziale czasowym o długości Δt nie wystąpiło żadne nowe zachorowanie; 2° że w chwili t było $n - 1$ zachorowań i w na-

stępnym przedziale czasowym o długości Δt wystąpiło jedno zachorowanie; 3^o uważamy, że przyjęte przez nas Δt jest tak małe, iż prawdopodobieństwo, że w czasie tym wystąpi więcej niż jedno zachorowanie, jest wielkością rzędu $o(\Delta t)$.

Tak więc otrzymujemy następujący układ równań:

$$(7.1) \quad \begin{cases} P_0(t + \Delta t) = P_0(t)(1 - \lambda y(t)\Delta t) \\ P_n(t + \Delta t) = P_n(t)(1 - \lambda y(t)\Delta t) + \\ \quad + P_{n-1}(t) \cdot \lambda y(t)\Delta t + o(\Delta t) \quad \text{dla } n > 0 \end{cases}$$

Dzieląc te równania obustronnie przez Δt i przechodząc do granicy, gdy $\Delta t \rightarrow 0$, otrzymujemy następujący układ równań różniczkowych:

$$(7.2) \quad \begin{cases} \frac{d P_0(t)}{dt} = -\lambda y(t) \cdot P_0(t) \\ \frac{d P_n(t)}{dt} = -\lambda y(t) P_n(t) + \lambda y(t) P_{n-1}(t) \quad \text{dla } n > 0 \end{cases}$$

W chwili $t = 0$ było 0 zachorowań, a więc warunki początkowe są tu następujące:

$$(7.3) \quad \begin{cases} P_0(0) = 1 \\ P_n(0) = 0 \quad \text{dla } n > 0 \end{cases}$$

Układ (7.2) jest układem tzw. równań różniczkowych liniowych rekurencyjnych. Przy znanym $P_{n-1}(t)$ potrafimy napisać wyrażenie na $P_n(t)$. Najpierw znajdziemy $P_0(t)$. Po scałkowaniu pierwszego z równań (7.2) otrzymujemy

$$(7.4) \quad P_0(t) = e^{-\lambda \int_0^t y(t) dt}$$

Następnie znajdujemy $P_1(t)$ z równania

$$\frac{d P_1(t)}{dt} = -\lambda y(t) \cdot P_1(t) + \lambda y(t) \cdot P_0(t)$$

i tak dalej.

Te same prawdopodobieństwa można w bardzo łatwy sposób otrzymać z funkcji tworzącej. Pomnóżmy mianowicie każde z równań (7.2) odpowiednio przez z^n i zsumujmy dla $n = 0, 1, 2, \dots$. Otrzymujemy

$$(7.5) \quad \sum_{n=0}^{\infty} \frac{d P_n(t)}{dt} \cdot z^n = -\sum_{n=0}^{\infty} \lambda \cdot y(t) \cdot P_n(t) \cdot z^n + \\ + \lambda \sum_{n=0}^{\infty} y(t) \cdot P_n(t) \cdot z^{n+1}$$

Oznaczmy przez $\Phi(z, t)$ funkcję tworzącą naszego procesu. Z definicji

$$\Phi(z, t) = P_0(t) \cdot z^0 + P_1(t) \cdot z^1 + P_2(t) \cdot z^2 + \dots$$

Pochodną cząstkową funkcji $\Phi(z, t)$ względem zmiennej t jest następujące wyrażenie:

$$\frac{\partial \Phi(z, t)}{\partial t} = \frac{d P_0(t)}{dt} \cdot z^0 + \frac{d P_1(t)}{dt} \cdot z^1 + \frac{d P_2(t)}{dt} \cdot z^2 + \dots,$$

czyli akurat to, które znajduje się po lewej stronie równania (7.5). Mamy więc:

$$(7.6) \quad \frac{\partial \Phi(z, t)}{\partial t} = \Phi(z, t) \cdot \lambda \cdot (z - 1)$$

Rozwiązaniem tego równania różniczkowego o pochodnych cząstkowych jest funkcja

$$\Phi(z, t) = e^{\lambda(z-1) \int_0^t y(t) dt} + C(z)$$

przy czym stałą $C(z)$ należy tak dobrać, aby $\Phi(z, 0) = 1$. Wynika stąd, że $e^{C(z)} = 1$. Tak więc funkcją tworzącą naszego procesu jest funkcja

$$(7.7) \quad \Phi(z, t) = e^{\lambda(z-1) \int_0^t y(t) dt}$$

Stąd już bardzo łatwo znajdujemy kolejne prawdopodobieństwa $P_n(t)$, stosując wzór (1.4):

$$(7.8) \quad \left\{ \begin{aligned} P_0(t) &= \Phi(z, t) \Big|_{z=0} = e^{-\lambda \int_0^t y(t) dt} \\ P_n(t) &= \frac{1}{n!} \frac{\partial^n \Phi(z, t)}{\partial z^n} \Big|_{z=0} = \\ &= \frac{1}{n!} \left[\lambda \int_0^t y(t) dt \right]^n \cdot e^{-\lambda \int_0^t y(t) dt} \\ &\text{dla } n = 1, 2, \dots \end{aligned} \right.$$

Tak więc prawdopodobieństwo, że w przedziale czasowym $[0, t]$ wystąpi n zachorowań opisuje się rozkładem Poissona, przy czym rolę

wyrażenia λt ze wzoru (2.2) odgrywa tu wielkość $\lambda \int_0^t y(t) dt$.

Oczekiwana ilość zachorowań do chwili t czyli $M(t)$ oraz wariancję tej liczby zachorowań $D^2(t)$ obliczamy, stosując wzory (1.5) i (1.6):

$$(7.9) \quad M(t) = \frac{\partial \Phi(z, t)}{\partial z} \Big|_{z=1} = \lambda \int_0^t y(t) dt$$

$$(7.10) \quad D^2(t) = \frac{\partial}{\partial z} \left[z \cdot \frac{\partial \Phi(z, t)}{\partial z} \right] \Big|_{z=1} - [M(t)]^2 = \lambda \int_0^t y(t) dt$$

Często interesują nas nie same prawdopodobieństwa, że w danym okresie wystąpi n zachorowań na białaczkę, ale wielkość współczynnika λ , mierzącego intensywność zachorowań. Gdybyśmy brali pod uwagę tylko jednego człowieka, to prawdopodobieństwo, że nie choruje on do chwili t wynosi $e^{-\lambda t}$. Wobec tego prawdopodobieństwo, że dowolny człowiek z rozważanej populacji zachoruje do chwili t wynosi $1 - e^{-\lambda t}$. Przyjmijmy $t =$ jeden rok. Ponieważ λ jest małe, $e^{-\lambda}$ można rozwinąć w szereg i przyjąć z dobrym przybliżeniem, że $e^{-\lambda} \cong 1 - \lambda$. Ostatecznie więc prawdopodobieństwo, że dowolny człowiek zachoruje w ciągu roku na białaczkę wynosi $1 - e^{-\lambda} \cong \lambda$. Współczynnik λ możemy więc w tym modelu interpretować jako szanse jednego człowieka na zachorowanie w ciągu jednego roku na tę chorobę. Przy badaniu rozwoju i rozprzestrzeniania się danej choroby interesuje nas wielkość współczynnika λ czyli intensywność zachorowań: zapytujemy się, czy λ zmienia się z upływem czasu, lub czy różne powiaty wykazują różne intensywności.

Przykład A. Przypuśćmy, że dysponujemy danymi, mówiącymi o liczbie zachorowań w ciągu kilku okresów rocznych i że dane te są następujące:

okres	liczba zachorowań n_k
1955-1956	3
1956-1957	0
1957-1958	1
1958-1959	2
1959-1960	8

Przypuśćmy dalej, że liczba ludności w tych latach wynosiła odpowiednio (stan na 1 I odpowiedniego roku):

rok	1955	1956	1957	1958	1959	1960
	$t = t_0$	$t = t_1$	$t = t_2$	$t = t_3$	$t = t_4$	$t = t_5$
liczba ludności $y(t)$	14 000	14 500	15 000	15 500	16 000	16 500

Widzimy, że liczba ludności $y(t)$ jest wielkością zmieniającą się liniowo w czasie. Jeśli rok 1955 przyjąć za $t = 0$, a za jednostkę czasu przyjąć rok, to $y(t)$ będzie następującą funkcją czasu:

$$(7.11) \quad y(t) = 500 \cdot t + 14\,000$$

stąd

$$(7.12) \quad \int_{t_0}^{t_1} y(t) dt = 500 \frac{t_1^2}{2} + 14\,000 \cdot t_1 - 500 \frac{t_0^2}{2} - 14\,000 \cdot t_0$$

Prawdopodobieństwo, że w okresie od $t = t_{i-1}$ do $t = t_i$ wystąpi n_i zachorowań, wynosi według wzoru (7.8)

$$P_{n_i}(t_{i-1}, t_i) = \frac{1}{n_i!} \left[\lambda \int_{t_{i-1}}^{t_i} y(t) dt \right]^{n_i} \cdot e^{-\lambda \int_{t_{i-1}}^{t_i} y(t) dt}$$

Oznaczmy

$$\int_{t_{i-1}}^{t_i} y(t) dt = C(t_i, t_{i-1}).$$

Wiarogodność L zdarzenia, że w rozpatrywanych pięciu okresach zaobserwujemy odpowiednio n_1, n_2, \dots, n_5 zachorowań, wynosi:

$$\begin{aligned} L &= \prod_{i=1}^5 P_{n_i} = \lambda^{n_1+n_2+\dots+n_5} \cdot \prod_{i=1}^5 \frac{C(t_i, t_{i-1})^{n_i}}{n_i!} \cdot e^{-\lambda[C(t_1, t_0)+\dots+C(t_5, t_4)]} = \\ &= \lambda^{n_1+n_2+\dots+n_5} \cdot e^{-\lambda C(t_5, t_0)} \cdot \prod_{i=1}^5 \frac{C(t_i, t_{i-1})^{n_i}}{n_i!} \end{aligned}$$

Szukamy takżej wartości λ , dla której wiarogodność L osiąga maksimum. Zadanie to jest równoważne znalezieniu miejsca zerowego pierwszej pochodnej funkcji $\ln L$. Wobec tego, że

$$\frac{d \ln L}{d \lambda} = 0 \quad \text{gdy} \quad \frac{n_1 + n_2 + \dots + n_5}{\lambda} = C(t_5, t_0)$$

otrzymujemy

$$(7.13) \quad \hat{\lambda} = \frac{n_1 + n_2 + \dots + n_5}{C(\tau_5, \tau_0)} = \frac{n_1 + n_2 + \dots + n_5}{\int_{\tau_0}^{\tau_5} y(t) dt}$$

Jest to estymator współczynnika λ , otrzymany metodą największej wiarygodności.

Estymator ten jest funkcją liniową liczebności n_1 , które są zmiennymi losowymi o rozkładzie Poissona, wobec czego ich wartości oczekiwane i wariancje równają się $\lambda \cdot C(\tau_1, \tau_{1-1})$. Korzystając ze znajomości wartości oczekiwanych i wariancji poszczególnych n_1 otrzymujemy:

$$(7.14) \quad E(\hat{\lambda}) = E\left(\frac{n_1 + \dots + n_5}{C(\tau_5, \tau_0)}\right) = \frac{En_1 + \dots + En_5}{C(\tau_5, \tau_0)} = \\ = \frac{\lambda \cdot C(\tau_1, \tau_0) + \dots + \lambda C(\tau_5, \tau_4)}{C(\tau_5, \tau_0)} = \lambda.$$

Obliczymy teraz wariancję tego estymatora

$$D^2(\hat{\lambda}) = D^2\left(\frac{n_1 + \dots + n_5}{C(\tau_5, \tau_0)}\right) = \frac{D^2(n_1) + \dots + D^2(n_5)}{[C(\tau_5, \tau_0)]^2} = \\ = \frac{\lambda C(\tau_1, \tau_0) + \dots + \lambda C(\tau_5, \tau_4)}{[C(\tau_5, \tau_0)]^2} = \lambda \cdot \frac{C(\tau_5, \tau_0)}{[C(\tau_5, \tau_0)]^2} = \frac{\lambda}{C(\tau_5, \tau_0)}.$$

Ostatecznie

$$(7.15) \quad D^2(\hat{\lambda}) = \frac{\lambda}{C(\tau_5, \tau_0)}.$$

Podstawiając na miejsce λ jego estymator $\hat{\lambda}$ według wzoru (7.13) otrzymujemy

$$(7.15a) \quad D^2(\hat{\lambda}) \approx \frac{\sum_{i=1}^5 n_i}{\left(\int_{\tau_0}^{\tau_5} y(t) dt\right)^2}$$

Dla naszego przykładu

$$\sum_{i=1}^5 n_i = 3 + 0 + 1 + 2 + 8 = 14,$$

$$\int_0^5 y(t) dt = 500 \cdot \frac{5^2}{2} + 14000 \cdot 5 = 76\ 250$$

Stąd otrzymujemy:

$$\hat{\lambda} = \frac{14}{76\ 250} = 0,000\ 184$$

$$D^2(\hat{\lambda}) = \frac{14}{(76\ 250)^2} = 0,000\ 000\ 0024$$

$$\sigma(\hat{\lambda}) = \sqrt{D^2(\hat{\lambda})} = 0,000\ 05$$

Gdybyśmy mieli dane dotyczące zachorowań w dwóch powiatach, to moglibyśmy na podstawie wzorów (7.13) i (7.14) testować hipotezę H_0 , że intensywności zachorowań są w obu powiatach takie same. Oznaczmy intensywność zachorowań w pierwszym powiecie przez λ_1 , a w drugim przez λ_2 , ogólnie przez λ_p , gdzie $p = 1, 2$.

Przypuśćmy, że w pierwszym powiecie obserwowaliśmy zachorowania w ciągu k_1 , a w drugim w ciągu k_2 okresów rocznych. Wtedy estymatorami intensywności λ_p będą wyrażenia analogiczne do (7.13):

$$(7.16) \quad \hat{\lambda}_p = \frac{n_1^{(p)} + \dots + n_{k_p}^{(p)}}{C^{(p)}(t_{k_p}, t_0)}$$

Jeśli założymy, że intensywności zachorowań w obu powiatach są takie same, czyli $\lambda_1 = \lambda_2 = \lambda$, to różnica $d = \lambda_1 - \lambda_2$ powinna mieć w przybliżeniu rozkład normalny o wartości oczekiwanej $E(d) = 0$ i dyspersji

$$(7.17) \quad \sigma_d = \sqrt{\lambda \left(\frac{1}{C(t_{k_1}, t_0)} + \frac{1}{C(t_{k_2}, t_0)} \right)}$$

Za λ możemy tu podstawić estymator $\tilde{\lambda}$, uzyskany na podstawie dwu próbek metodą największej wiarygodności:

$$(7.18) \quad \tilde{\lambda} = \frac{n_1^{(1)} + \dots + n_{k_1}^{(1)} + n_1^{(2)} + \dots + n_{k_2}^{(2)}}{C^{(1)}(t_{k_1}, t_0) + C^{(2)}(t_{k_2}, t_0)}$$

Ponieważ rozkład różnicy jest jedynie w przybliżeniu normalny, przyjmujemy następującą zasadę: 1^o jeśli $t = \left| \frac{d}{\sigma_d} \right| < 2$, to różnice między obu intensywnościami uważamy za statystycznie nieis-

totne; 2^o jeśli $2 \leq t = \left| \frac{d}{\sigma_d} \right| \leq 3$, to będziemy mówić, że zaobserwowana różnica leży na granicy istotności; 3^o jeśli $t = \left| \frac{d}{\sigma_d} \right| > 3$, to różnice między obu intensywnościami będziemy uważać za statystycznie istotne.

Przykład B. Jako dane dla pierwszego powiatu przyjmijmy dane z przykładu A czyli: $k_1 = 5$, $(n_1, \dots, n_5) = (3, 0, 1, 2, 8)$,

$$C^{(1)}(\tau_5, \tau_0) = \int_{\tau_0}^{\tau_5} y^{(1)}(t) dt = 76\,250, \text{ skąd } \hat{\lambda}_1 = 0,000\,184. \text{ Przypuśćmy,}$$

że dla drugiego powiatu, dla którego również $k_2 = 5$ i $\int_{\tau_0}^{\tau_5} y^{(2)}(t) dt = 76\,250$ zaobserwowaliśmy następujące liczby zachorowań: $n_1^{(2)} = 4$, $n_2^{(2)} = 8$, $n_3^{(2)} = 4$, $n_4^{(2)} = 5$, $n_5^{(2)} = 7$, skąd otrzymaliśmy następującą wartość estymatora $\hat{\lambda}_2$:

$$\hat{\lambda}_2 = \frac{4 + 8 + 4 + 5 + 7}{76\,250} = \frac{28}{76\,250} = 0,000\,367.$$

Wobec tego $d = \hat{\lambda}_1 - \hat{\lambda}_2 = -0,000\,183$.

Estymator $\hat{\lambda}$ obliczony dla całego materiału przy założeniu, że intensywności zachorowań są w obu powiatach takie same, wynosi według wzoru (7.18)

$$\hat{\lambda} = \frac{42}{2 \cdot 76\,250} = 0,000\,275$$

Dyspersja różnicy $d = \hat{\lambda}_1 - \hat{\lambda}_2$ wynosi

$$\begin{aligned} \sigma_d &= \sqrt{0,000\,275 \cdot \frac{2}{76\,250}} = \sqrt{\frac{0,000\,55}{76\,250}} = \sqrt{0,000\,000\,007\,223\,87} = \\ &= 0,000\,085 \end{aligned}$$

Stąd

$$t = \left| \frac{d}{\sigma_d} \right| = \left| - \frac{0,000\,183}{0,000\,085} \right| = 2,15$$

Otrzymaliśmy tu wartość $2 < t = 2,15 < 3$, a więc zaobserwowana różnica leży na granicy istotności. Różnica $d = 0,000\,184$ przy założeniu H_0 jest dość mało prawdopodobna, chociaż nie występuje ona aż tak rzadko, byśmy stanowczo mogli twierdzić, że nie można jej otrzymać, jeśli intensywności zachorowań w obu powiatach są takie same.

8. Proces urodzin i śmierci

Dotychczas mówiliśmy o takich procesach, dla których obserwowana zmienna losowa $x(t)$, interpretowana jako liczba zajść pewnego zdarzenia w przedziale czasowym $[0, t)$, była niemalejącą funkcją czasu. Obecnie omówimy takie zdarzenia, które mogą nie tylko pojawiać się, ale i zanikać. Liczba zdarzeń trwających do chwili t jest zmienną losową i będziemy ją w dalszym ciągu oznaczać przez $x(t)$.

Takim "zdarzeniem" może być np. istnienie żywej komórki (bakterii, wirusa) w jakiejś hodowli, w której komórki mogą rozmnażać się lub obumierać. Wtedy $x(t)$ będzie oznaczać liczbę żywych komórek w chwili t . Innym przykładem zdarzeń pojawiających się i zanikających jest zaraźliwość jakiejś osoby w modelu epidemii, w którym zakłada się, że osoba zarażająca może w pewnym momencie stracić swą zaraźliwość. W tym przypadku $x(t)$ można interpretować jako liczbę osób zarażających w chwili t .

Proces, w którym zdarzenia mogą pojawiać się i zanikać, jest nazywany procesem urodzin i śmierci.

Podobnie jak poprzednio, będziemy starali się znaleźć prawdopodobieństwa $P_n(t)$, tj. prawdopodobieństwa, że w chwili t zmienna losowa $x(t)$ przyjmie wartość n ($n = 0, 1, \dots$). Interesującą rzeczą będzie również oczekiwana liczba zdarzeń, trwających w chwili t czyli $E x(t)$, jak również wariancja $D^2 x(t)$.

Aby znaleźć równania różniczkowe na prawdopodobieństwa $P_n(t)$, uczynimy następujące założenia (por. warunki I-III z rozdziału 2):

I. Przyrosty zmiennej losowej $x(t)$ w rozłącznych przedziałach czasowych są zmiennymi losowymi niezależnymi.

II. Przyrosty zmiennej losowej $x(t)$ są jednorodne w czasie.

III. Dla dowolnego $n > 0$

$$a) \Pr \{x(t + \Delta t) = n + 1 \mid x(t) = n\} = \lambda_n \cdot \Delta t + o(\Delta t) \quad \lambda_n \geq 0.$$

$$b) \Pr \{x(t + \Delta t) = n - 1 \mid x(t) = n\} = \mu_n \cdot \Delta t + o(\Delta t) \quad \mu_n \geq 0.$$

$$c) \Pr \{x(t + \Delta t) = r \mid x(t) = n\} = o(\Delta t) \quad \text{dla } 0 \leq r < n-1 \text{ i } r > n+1$$

Wzór a) mówi, że warunkowe prawdopodobieństwo pojawienia się w przedziale czasowym $[t, t + \Delta t)$ nowego zdarzenia - pod warunkiem, że w chwili t trwało n zdarzeń - jest w przybliżeniu proporcjo-

nalne do długości tego przedziału. Współczynnik proporcjonalności λ_n , czyli intensywność pojawienia się nowego zdarzenia jest funkcją n , czyli liczby zdarzeń trwających w chwili t .

Wzór b) mówi, że prawdopodobieństwo zaniku (zginięcia, zniknięcia) jednego zdarzenia w przedziale czasowym $[t, t + \Delta t)$ jest proporcjonalne do długości tego przedziału ze współczynnikiem proporcjonalności μ_n , który w ogólnym przypadku może zależeć od tego, ile zdarzeń trwało w chwili t .

Wzór c) mówi, że prawdopodobieństwo, że wystąpi więcej niż jedna zmiana w ciągu przedziału czasowego o długości Δt jest w przybliżeniu równe zero.

Przy tych założeniach można wyprowadzić równania różniczkowe na nieznanne prawdopodobieństwa $P_n(t)$ oraz można pokazać, że rozwiązaniem tych równań rzeczywiście są funkcje, które są poszukiwanymi prawdopodobieństwami. Są to jednak zagadnienia dość skomplikowane (por. [10], str. 375).

W dalszym ciągu opiszemy proces urodzin i śmierci przy pewnej dalszej specyfikacji warunku III, opisujący rozwój populacji bakterii, na które działamy antybiotykiem. Zmienną losową $x(t)$ będzie liczba żywych bakterii w chwili t . Bakterie te mogą rodzić nowe bakterie oraz mogą ginąć (obumierać). Jeśli weźmiemy pod uwagę krótki przedział czasowy o długości Δt , to wydaje się naturalnym przyjąć, że prawdopodobieństwo obumarcia jednej spośród n żywych bakterii powinno być: 1° proporcjonalne do liczby żywych bakterii (im większa liczba bakterii, tym większe prawdopodobieństwo, że któraś z nich umrze); 2° proporcjonalne do Δt (im dłużej obserwujemy naszą populację, tym większe prawdopodobieństwo, że któraś z obserwowanych w tym czasie bakterii umrze). Wobec tego możemy napisać, iż prawdopodobieństwo zdarzenia, że w przedziale czasowym $[t, t + \Delta t)$ zaobserwujemy śmierć jednej bakterii, wynosi w przybliżeniu $\mu \cdot n \cdot \Delta t$. Współczynnik μ jest tu współczynnikiem proporcjonalności. Im większe μ , tym większe prawdopodobieństwo śmierci dla każdej z obserwowanych bakterii. Mówimy, że μ charakteryzuje intensywność umierania bakterii. W szczególnym przypadku μ jest wielkością stałą dla całego rozpatrywanego okresu rozwoju obserwowanej populacji bakterii. W ogólnym wypadku μ może być funkcją czasu: $\mu = \mu(t)$. Np. w omawianym przykładzie śmiertelność wśród bakterii na pewno zależy od stężenia antybiotyku, to zaś z kolei może być funkcją czasu, gdy podawany antybiotyk jest stopniowo rozkładany i wydalany z organizmu.

Podobne rozważania można przeprowadzić co do intensywności rozmnażania się bakterii; w ogólnym przypadku należy przyjmować, że intensywność podziału jest również funkcją czasu.

Wyprowadzimy teraz równania na prawdopodobieństwa $P_n(t)$, że w czasie t zaobserwujemy n żywych bakterii. Przypuśćmy więc, że rozpatrujemy hodowlę komórek, w której komórki nie tylko rozmnażają się (powstają nowe), ale i giną. Niech prawdopodobieństwo, że jedna komórka w małym przedziale czasowym Δt rozmnoży się (podzieli na dwie), wynosi w przybliżeniu $\lambda \Delta t$. Wtedy jeśli w chwili t było n komórek, prawdopodobieństwo zdarzenia, że w przedziale czasowym $[t, t + \Delta t)$ przybędzie jedna nowa komórka, wynosi $\lambda n \Delta t + o(\Delta t)$. Analogicznie, niech prawdopodobieństwo śmierci dla jednej komórki w czasie Δt będzie równe $\mu \Delta t + o(\Delta t)$. Jeśli w chwili t liczba żywych komórek wynosi n , to od chwili t do chwili $t + \Delta t$ mogły zajść następujące zdarzenia: 1^o w chwili t było już n żywych komórek i w przedziale czasowym $[t, t + \Delta t)$ żadna nowa się nie urodziła ani żadna nie obumarła; 2^o w chwili t było $n - 1$ żywych komórek i w przedziale czasowym $[t, t + \Delta t)$ przybyła jedna żywa komórka; 3^o w chwili t było $n + 1$ żywych komórek, ale w przedziale czasowym $[t, t + \Delta t)$ jedna komórka obumarła; 4^o w przedziale czasowym $[t, t + \Delta t)$ wystąpiła więcej niż jedna zmiana. Prawdopodobieństwa zdarzenia 1^o wynosi

$$P_n(t) \cdot [1 - \lambda n \Delta t + o(\Delta t)] \cdot [1 - \mu n \Delta t + o(\Delta t)] \cong \\ \cong P_n(t) \cdot [1 - (\lambda + \mu) n \Delta t + o(\Delta t)] ,$$

prawdopodobieństwo zdarzenia 2^o wynosi

$$P_{n-1}(t) \cdot [\lambda(n-1)\Delta t + o(\Delta t)] ,$$

prawdopodobieństwo zdarzenia 3^o wynosi

$$P_{n+1}(t) \cdot [\mu(n+1)\Delta t + o(\Delta t)] ,$$

wreszcie prawdopodobieństwo zdarzenia 4^o przyjmujemy za wielkość rzędu $o(\Delta t)$.

Wobec powyższego otrzymujemy następujące równania na $P_n(t)$, czyli prawdopodobieństwo, że w chwili t będzie n żywych komórek:

$$P_n(t + \Delta t) = P_n(t) \cdot [1 - (\lambda + \mu) \Delta t - o(\Delta t)] + \\ + P_{n-1}(t) \cdot [\lambda(n-1)\Delta t + o(\Delta t)] + \\ + P_{n+1}(t) \cdot [\mu(n+1)\Delta t + o(\Delta t)] + o(\Delta t).$$

Przenosząc na lewą stronę składnik $P_n(t)$ i dzieląc obie strony równania przez $\Delta t \neq 0$ otrzymujemy po lewej stronie iloraz różnicowy

wy prawdopodobieństwa $P_n(t)$. Dla $\Delta t \rightarrow 0$ otrzymujemy:

$$(8.1) \quad \begin{cases} \frac{d P_0(t)}{dt} = \mu P_1(t) \\ \frac{d P_n(t)}{dt} = -(\lambda + \mu)nP_n(t) + \lambda(n-1)P_{n-1}(t) + \mu(n+1)P_{n+1}(t) \end{cases} \quad \text{dla } n > 0$$

Widzimy, że otrzymany układ równań różniczkowych nie jest już typu rekurencyjnego: aby znaleźć $P_0(t)$ trzeba znać $P_1(t)$, ogólnie, aby znaleźć $P_n(t)$ należy znać zarówno $P_{n-1}(t)$ jak i $P_{n+1}(t)$. Układy takie rozwiązuje się metodą funkcji tworzących. Wypiszmy mianowicie kolejno dla $n = 0, 1, 2, \dots$ równania (8.1), przemnożmy każde z tych równań obustronnie odpowiednio przez z^0, z^1, z^2, \dots i zsumujmy. Otrzymamy wówczas:

$$(8.2) \quad \begin{aligned} \sum_{n=0}^{\infty} \frac{d P_n(t)}{dt} \cdot z^n &= \sum_{n=0}^{\infty} \left\{ \lambda \cdot z(z-1)nP_n(t)z^{n-1} + \right. \\ &\quad \left. + \mu n(1-z)P_n(t)z^{n-1} \right\} = \\ &= \sum_{n=0}^{\infty} [\lambda z(z-1) - \mu(z-1)] nP_n(t) \cdot z^{n-1} = \\ &= \sum_{n=0}^{\infty} (z-1)(\lambda z - \mu) nP_n(t) z^{n-1} \end{aligned}$$

Ale

$$\sum_{n=0}^{\infty} \frac{d P_n(t)}{dt} \cdot z^n = \frac{\partial \Phi(z, t)}{\partial t} \quad \text{ i } \quad \sum_{n=0}^{\infty} nP_n(t) z^{n-1} = \frac{\partial \Phi(z, t)}{\partial z}$$

Wobec tego otrzymujemy z (8.2) następujące równanie różniczkowe o pochodnych cząstkowych:

$$(8.3) \quad \frac{\partial \Phi(z, t)}{\partial t} - (z-1)(\lambda z - \mu) \frac{\partial \Phi(z, t)}{\partial z} = 0.$$

Zakładając, że w chwili $t = 0$ była jedna komórka, mamy następujący warunek początkowy:

$$(8.4) \quad \Phi(z, 0) = z.$$

Metody rozwiązywania takich równań są podane np. w [31] lub [32].

Jeśli intensywności λ i μ są stałymi, niezależnymi od czasu, to rozwiązaniem równania (8.3) jest funkcja

$$(8.5) \quad \Phi(z, t) = \begin{cases} \frac{\mu(1 - e^{(\lambda-\mu)t}) - z(\lambda - \mu e^{(\lambda-\mu)t})}{\mu - \lambda e^{(\lambda-\mu)t} - \lambda z(1 - e^{(\lambda-\mu)t})} & \text{dla } \lambda \neq \mu \\ 1 - \frac{z-1}{\lambda t(z-1) - 1} & \text{dla } \lambda = \mu \end{cases}$$

Gdyby w chwili $t = 0$ było s komórek, to funkcja tworząca takiego procesu byłaby równa funkcji $\Phi(z, t)$ z wzoru (8.5) podniesionej do s -tej potęgi:

$$\Phi^{(s)}(z, t) = [\Phi(z, t)]^s$$

Mając funkcję tworzącą, możemy znaleźć prawdopodobieństwa $P_n(t)$, korzystając z wzorów (1.4).

Palm podał wzory na $P_n(t)$ w następującej postaci (cytowane za Fellerem [8]): gdy $\lambda \neq \mu$

$$(8.6) \quad \begin{cases} P_0(t) = \mu A \\ P_n(t) = (1 - \mu A)(1 - \lambda A)(\lambda A)^{n-1} \end{cases} \quad \text{dla } n \geq 1,$$

gdzie

$$A = (1 - e^{(\lambda-\mu)t})(\mu - \lambda e^{(\lambda-\mu)t})^{-1};$$

gdą $\lambda = \mu$

$$(8.7) \quad \begin{cases} P_0(t) = \lambda t (1 + \lambda t)^{-1} \\ P_n(t) = (\lambda t)^{n-1} (1 + \lambda t)^{-(n+1)} \end{cases}$$

Prawdopodobieństwo ostatecznego wymarcia hodowli wynosi:

$$\lim_{t \rightarrow \infty} P_0(t) = \begin{cases} 1 & \text{dla } \mu \geq \lambda \\ \frac{\mu}{\lambda} & \text{dla } \mu \leq \lambda \end{cases}$$

Wartość średnią liczby komórek w czasie t możemy wyznaczyć jako wartość pierwszej pochodnej funkcji $\Phi(z, t)$, określonej wzorem (8.5), po podstawieniu do tego wzoru wartości $z = 1$. Otrzymujemy w ten sposób:

$$(8.8) \quad M(t) = \left. \frac{\partial \Phi(z, t)}{\partial z} \right|_{z=1} = e^{(\lambda-\mu)t}$$

W podobny sposób można wyznaczyć wariancję liczby komórek w chwili t . Wynosi ona:

$$(8.9) \quad D^2(t) = \begin{cases} \left\{ [M(t)]^2 - M(t) \right\} \cdot \frac{\lambda + \mu}{\lambda - \mu} & \text{dla } \lambda \neq \mu \\ 2\lambda t & \text{dla } \lambda = \mu \end{cases}$$

Widzimy, że wariancja liczby żywych komórek w chwili t przy procesie urodzin i śmierci różni się dla $\lambda \neq \mu$ czynnikiem $\frac{\lambda + \mu}{\lambda - \mu}$ w porównaniu z wariancją czystego procesu urodzin (porównaj wzór (4.4)).

Gdy $\lambda = \lambda(t)$ i $\mu = \mu(t)$, wówczas rozwiązanie równania (8.3) i tym samym wzory na prawdopodobieństwa $P_n(t)$ są bardziej skomplikowane. Można je znaleźć w pracy Kendalla [18].

Niech $x(t)$ oznacza licznosc populacji bakterii w chwili t . Urbanik [35] znalazł rozkład prawdopodobieństwa dla maksymalnej ilości bakterii, czyli zmiennej losowej $S(x) = \max_{0 \leq t \leq \infty} x(t)$ w pro-

cesie urodzin i śmierci o stałych intensywnościach. Znalazł on również oczekiwaną długość czasu, potrzebnego do osiągnięcia maksymalnej licznosci $S(x)$ po raz pierwszy. Urbanik rozważał również rozstęp między momentami czasowymi, w których rozwijająca się według procesu urodzin i śmierci o stałych intensywnościach populacja bakterii osiąga swą wartość maksymalną. W pracy [34] podany jest wzór na wartość oczekiwaną tego rozstępu pod warunkiem, że maksymalna liczba bakterii w rozważanym czasie wynosi n .

Znacznie trudniejsze jest znalezienie wyrażeń na $P_{n,m}(t)$, czyli prawdopodobieństwo, że w chwili t będzie dokładnie n żywych i m martwych komórek. Zagadnieniem tym zajmował się Kendall [18]. Pewien sposób postępowania prowadzący do otrzymania wyrażeń na $P_{n,m}(t)$ w przypadku, gdy $\lambda(t)$ i $\mu(t)$ są niezależne od czasu, oraz gdy $\varrho = \frac{\mu(t)}{\lambda(t)}$ jest niezależne od czasu, podał Klonecki [23].

9. Zwalczanie komórek pasożytniczych przez makrofagi

Rozważmy teraz jako przykład następujący proces urodzin i śmierci. Przypuśćmy, że mamy dwie kategorie wzajemnie zwalczających się komórek tzw. makrofagi oraz komórki pasożytnicze tzw. drożdżaki. Jeśli makrofag pochłonie komórkę drożdżaka, to mogą

zajść dwa zdarzenia: 1° drożdżak rozwija się (rozmnaża) wewnątrz makrofaga, powodując zahamowanie jego procesów witalnych i w ostatecznym stadium jego zniszczenie, po czym uwolnione komórki rozmnażają się swobodnie według prawa logarytmicznego wzrostu; 2° drożdżak wewnątrz makrofaga ulega lizie, czyli następuje śmierć drożdżaka.

Przypuśćmy, że możemy wyznaczyć w ustalonych momentach czasu t ilość żywych komórek drożdżaka w danej hodowli. Jeśli za czas zerowy przyjmiemy czas, w którym nastąpiło wzajemne spotkanie drożdżaków i makrofagów, przy czym drożdżaków było względnie mało, a makrofagów dużo, tak że każdy makrofag pochłoniął najwyżej jednego drożdżaka, to można się spodziewać, że w początkowym okresie średnia ilość żywych komórek drożdżaków będzie zmniejszać się (bo drożdżaki ulegają lizie), po czym po pewnym czasie zacznie znowu wzrastać, bo komórki, które uwolniły się z makrofagów rozpoczynają fazę swobodnego rozwoju.

Oznaczmy przez $x(t)$ liczbę żywych komórek drożdżaków w chwili t . Prawdopodobieństwo podziału jednej komórki na dwie nowe w ciągu czasu Δt niech wynosi $\lambda \cdot \Delta t$, zaś prawdopodobieństwo śmierci jednej komórki w czasie t niech wynosi $\mu \cdot \Delta t$. Intensywność śmierci na pewno jest funkcją czasu. Przyjmijmy, że jest to funkcja kształtu

$$(9.1) \quad \mu(t) = \frac{\mu}{t + A}$$

przy czym parametry μ oraz A należy wyznaczyć z danych doświadczalnych. O intensywności podziału wiadomo, że jest ona bardzo mała, gdy komórki znajdują się wewnątrz makrofagów; po uwolnieniu się rozwijają się już według praw logarytmicznego wzrostu. Tak więc intensywność ta jest funkcją czasu. Załóżmy, że jest to funkcja kształtu

$$(9.2) \quad \lambda(t) = \lambda \left(1 - \frac{A}{t + A}\right)$$

Z (9.2) wynika, że $\lambda(0) = 0$, a dla dużych t intensywność $\lambda(t)$ zbliża się asymptotycznie do λ jako wartości granicznej.

Niech $P_n(t)$ oznacza prawdopodobieństwo, że w chwili t będzie n żywych komórek. Rozważając, co mogło być w chwili t jeśli w chwili $t + \Delta t$ było n żywych komórek, otrzymujemy następujące równania różniczkowe na prawdopodobieństwa $P_n(t)$:

$$(9.3) \left\{ \begin{aligned} \frac{d P_0(t)}{dt} &= \mu(t) \cdot P_1(t) \\ \frac{d P_1(t)}{dt} &= -[\mu(t) + \lambda(t)] \cdot P_1(t) + \mu(t) \cdot P_2(t) \\ \frac{d P_n(t)}{dt} &= -[\lambda(t) + \mu(t)] \cdot n \cdot P_n(t) + \lambda(n-1) P_{n-1}(t) + \\ &\quad + \mu(t) \cdot (n+1) \cdot P_{n+1}(t) \end{aligned} \right. \quad \text{dla } n \geq 2$$

Przypuśćmy, że w chwili $t = 0$ zaczęliśmy obserwować N komórek drożdżaków. Losy jednej komórki możemy uważać za jedną realizację procesu urodzin i śmierci o intensywnościach $\lambda(t)$ i $\mu(t)$. Przypuśćmy, że w chwili t zaobserwowaliśmy $N(t)$ żywych komórek. Są to bądź komórki, które istniały w chwili $t = 0$, bądź też ich potomstwo. Komórki pochodzące od jednej z N komórek macierzystych nazywać będziemy rodziną komórki. Dla każdej chwili t możemy zbadać rozkład rodzin o danej liczności n . Ponieważ N jest duże, spodziewamy się, że będzie $N \cdot P_0(t)$ takich rodzin, które już nie istnieją, bo wszystkie komórki w nich wymarły; będzie $N \cdot P_1(t)$ takich rodzin, które będą liczyć po jednej komórce; i ogólnie będzie $N \cdot P_n(t)$ rodzin takich, które będą liczyć po n komórek. Zaobserwowaną łączną ilość żywych komórek w chwili t możemy więc w przybliżeniu przedstawić jako następującą sumę:

$$N(t) \cong 0 \cdot N \cdot P_0(t) + 1 \cdot N \cdot P_1(t) + 2 \cdot N \cdot P_2(t) + \dots$$

Jeśli równość powyższą podzielimy obustronnie przez N , otrzymamy:

$$(9.4) \quad \frac{N(t)}{N} \cong \sum_{k=0}^{\infty} k P_k(t) = E x(t).$$

Tak dzieląc zaobserwowaną ilość komórek w chwili t przez ilość komórek w chwili $t = 0$ otrzymujemy w przybliżeniu wartość oczekiwaną liczby komórek w chwili t przy założeniu, że w chwili $t = 0$ była jedna komórka.

Wyrażenie przedstawiające oczekiwaną wartość naszego procesu przy założonych intensywnościach λ i μ można uzyskać w łatwy sposób z układu równań (9.3). Mnożąc każde z tych równań odpowiednio przez $n = 0, 1, 2, \dots$ i sumując stronami otrzymujemy:

$$\sum_{n=0}^{\infty} n \cdot \frac{d P_n(t)}{dt} = -[\lambda(t) + \mu(t)] \cdot \sum_{n=0}^{\infty} n^2 \cdot P_n(t) + \lambda(t) \sum_{n=0}^{\infty} (n+1) n \cdot P_n(t) +$$

$$+ \mu(t) \sum_{n=0}^{\infty} (n-1) \cdot n \cdot P_n(t) = \lambda(t) \cdot \sum_{n=0}^{\infty} n \cdot P_n(t) - \mu(t) \cdot \sum_{n=0}^{\infty} n \cdot P_n(t) =$$

$$= [\lambda(t) - \mu(t)] \cdot M(t).$$

Otrzymujemy więc następujące równania różniczkowe zwyczajne na oczekiwaną wartość liczby komórek w chwili t :

$$(9.5) \quad \frac{dM(t)}{dt} = \left[\lambda \left(1 - \frac{1}{t+A} \right) - \mu \frac{1}{t+A} \right] M(t).$$

Rozwiązaniem tego równania, przy warunku początkowym, że w chwili $t = 0$ była jedna komórka, jest funkcja

$$(9.6) \quad M(t) = e^{\lambda t} \left(\frac{A}{t+A} \right)^{\mu + \lambda}$$

Równanie (9.6) przedstawia krzywą trójparametrową: parametrami są λ , μ oraz A . λ jest to intensywność podziału rozpatrywanych komórek w warunkach swobodnego rozwoju. Prowadząc doświadczenie, znamy na ogół ten parametr. I tak, jeśli znamy np. okres T , w którym populacja komórek podwaja się w warunkach swobodnego rozwoju, to $\lambda = \frac{\ln 2}{T}$. Pozostaje sprawa dobrania dwóch pozostałych paramet-
trów tj. μ oraz A tak, aby krzywa $M(t; \lambda, \mu, A)$ pasowała dobrze do

zaobserwowanej krzywej empirycznej $\frac{N(t)}{N(0)}$. Jeśli krzywa empiryczna ma dobrze zaznaczone minimum, to można odczytać wartość t_{\min} z empirycznego wykresu i następnie przyrównać tę wartość do wartości t_{\min} obliczonej ze wzoru (9.6). Otrzymujemy w ten sposób prosty związek między nieznanymi wartościami μ i A , z którego wyznaczamy jeden parametr jako funkcję drugiego. Pozostała więc do znalezienia wartość jednego parametru, który wyznaczamy np. z warunku, że-
by krzywa $M(t)$ przechodziła przez jakiś odczytany z wykresu punkt.

Gani [14] rozważał sytuację odwrotną - duża ilość komórek pasożytniczych i mała ilość makrofagów (które u Ganiego nazywają się bakteriami). W tej sytuacji na każdą bakterię przypada pewna ilość komórek pasożytniczych, które przyklejają się do bakterii i przedostają się do jej wnętrza. Przypuśćmy, że do jednej bakterii może dokleić się maksymalnie r komórek pasożytniczych. Oznaczając przez $P(n_0, n_1, \dots, n_r; v_0, t)$ prawdopodobieństwo, że w chwili t będzie odpowiednio n_0 wolnych bakterii, n_1 bakterii z jedną komórką pasożytniczą, n_2 bakterii z dwiema komórkami pasożytniczymi itd. aż do n_r bakterii z r komórkami pasożytniczymi, podczas gdy

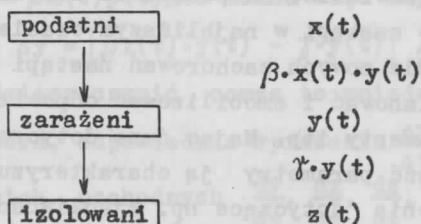
v_0 komórek pasożytniczych jest wolnych, Gani znalazł wyrażenie na funkcję tworzącą prawdopodobieństwa $P(n_0, n_1, \dots, n_r; v_0, t)$ przy upraszczającym założeniu, że v_0 czyli ilość wolnych komórek pasożytniczych przyjmuje nie dokładne wartości wynikające z losowego ubytku wskutek doklejania się do różnych bakterii, ale jest pewną funkcją czasu wynikłą z rozważania tzw. modelu deterministycznego.

10. Model epidemii z uwzględnieniem izolowania chorych osobników

Rozważmy następujący model epidemii: w populacji, liczącej ogółem n osobników w chwili $t = 0$ mamy $x(0)$ osobników zdrowych, podatnych na zarażenie się określoną chorobą oraz $y(0)$ osobników chorych, którzy przy kontaktach z osobnikami zdrowymi mogą je z pewnym prawdopodobieństwem zarażać. Osoby zarażane natychmiast stają się chore (widoczne są objawy choroby) i zarażają z kolei inne zdrowe osoby. Osoby chore można izolować od reszty populacji w ten sposób, że nie mają już one więcej kontaktu z osobami zdrowymi i nie mogą być źródłem nowych zarażeń. Izoluje się jednak nie wszystkie osoby chore, ale tylko pewną ich część, co można wyrazić inaczej, że dla każdej osoby istnieje pewne prawdopodobieństwo, że zostanie ona odizolowana od populacji osób zdrowych. W ten sposób cała populacja została podzielona na trzy klasy: podatnych na zarażenie, chorych i izolowanych. Liczności poszczególnych klas w chwili t będziemy oznaczać przez $x(t)$, $y(t)$, $z(t)$, przy czym oczywiście

$$x(t) + y(t) + z(t) = n.$$

Możliwe są przejścia z klasy zdrowych do chorych oraz z klasy chorych do klasy izolowanych. Inne przejścia są niemożliwe. Warunki powyższe przedstawiono schematycznie na rys. 2.



Rys. 2. Schemat epidemii z izolowaniem zarażonych osobników

Taki podział, zastosowany przez Mc Kendricka w latach dwudziestych przy badaniu zachorowań na cholera w Indiach jest oczywiście bardzo uproszczony. W rzeczywistości przy konkretnej chorobie kategorii takich możemy wyróżnić więcej, np. osoby odporne na daną chorobę, osoby będące nosicielami danej choroby, ale u których choroba się nie manifestuje, osoby z nietypowymi objawami danej choroby, itp.

Taką ogólną sytuację, w której badana populacja dzieli się na r kategorii badali Firescu i Tautu [11]. Przyjmowali oni, że przejścia z jednej kategorii do innych mogą się odbywać jedynie w pewnych określonych momentach czasu, które możemy ponumerować: $k = 0, 1, 2, \dots$. Niektóre z tych kategorii są tzw. stanami pochłaniającymi, to znaczy, że osoba, która znalazła się w tej kategorii, nie może jej już zmienić. Tak jest np. gdy chory umrze, lub gdy ma trwałą odporność na daną chorobę. Inne kategorie tworzą tzw. stany przejściowe, to znaczy możliwe jest opuszczanie i ewentualne późniejsze wracanie do tych stanów.

Prawdopodobieństwa przejścia z jednego stanu do drugiego można charakteryzować macierzą przejścia, która w szczególnym przypadku może nie zależeć od czasu:

$$P = \begin{pmatrix} p_{11} & \dots & p_{1r} \\ \dots & \dots & \dots \\ p_{r1} & \dots & p_{rr} \end{pmatrix} \quad \text{gdzie } p_{ij} \geq 0 \quad \text{i} \quad \sum_{j=1}^r p_{ij} = 1 \quad \text{dla każdego } i$$

W pracy [11] podane są różne wzory, dotyczące zachowania się tych prawdopodobieństw po długim czasie trwania epidemii (np. ile razy średnio osobnik może przechodzić ze stanu i do stanu j , jak długo osobnik średnio pozostaje w danym stanie przejściowym j , jaka część osób z kategorii i przejdzie po n odstępach czasu do kategorii j).

Matematyczny opis epidemii jest bardzo ważny dla służby zdrowia. Pozwala on przewidzieć na podstawie dotychczasowego przebiegu epidemii to, co nastąpi w najbliższym czasie. Jeśli zaś potrafimy przewidzieć, ile nowych zachorowań nastąpi w najbliższym czasie, to można zaplanować i zmobilizować odpowiednią opiekę lekarską, łóżka, medykamenty itp. Mając dane dotyczące pewnej epidemii staramy się znaleźć parametry ją charakteryzujące, aby później wysuwać spostrzeżenia dotyczące np. zjadliwości zarazków, wpływu profilaktyki lekarskiej itp. Ciekawą rzeczą jest porównywanie różnych epidemii (z różnych okresów, z różnych krajów) nie na podsta-

wie subiektywnych wrażeń, ale na podstawie obliczonych parametrów danej epidemii. Aby jednak to uczynić, trzeba umieć opisać daną epidemię za pomocą pewnego modelu matematycznego. Niestety, matematyczny opis epidemii nawet przy najprostszych modelach nie jest rzeczą łatwą.

Podaję tu opis epidemii za pomocą tzw. modelu deterministycznego dla schematu, rozpatrywanego po raz pierwszy przez Kermacka i Mc Kendricka (rys. 2). Rozpatrywali oni zachorowania na cholereę. Okres czasu, jaki upływa od momentu zarażenia się tą chorobą do stania się zaraźliwym jest stosunkowo krótki i dlatego był przez nich zaniedbywany.

Zakładamy, że licznosc populacji jest duża, tak że zmiany zachodzące w jej licznosci możemy w przybliżeniu uważać za zmiany ciągłe, a nie skokowe. Przyjmujemy, że ilość osób zarażonych w małym przedziale czasowym $[t, t + \Delta t)$ jest wprost proporcjonalna do liczby osób zdrowych podatnych na zarażenie, czyli do $x(t)$, do aktualnej liczby chorych, czyli $y(t)$ oraz do długości rozpatrywanego przedziału czasowego czyli Δt . Można więc napisać

$$\Delta y = \beta \cdot x(t) \cdot y(t) \cdot \Delta t,$$

gdzie β jest współczynnikiem proporcjonalności, mierzącym intensywność zachorowań. Jednocześnie, ile osób przybyło do klasy chorych, tyle osób ubyło w klasie zdrowych. A więc

$$(10.1a) \quad \Delta x = -\Delta y = -\beta \cdot x(t) \cdot y(t) \cdot \Delta t.$$

Jeśli chodzi o ilość izolowanych w przedziale czasowym $[t, t + \Delta t)$, to przyjmujemy, że ilość ta jest wprost proporcjonalna do liczby chorych, czyli $y(t)$ oraz do Δt . Można więc napisać

$$(10.1b) \quad \Delta z = \gamma y(t) \Delta t.$$

Tyle samo osób ubywa z klasy chorych. Ostatecznie $\Delta y = y(t + \Delta t) - y(t)$ składa się z dwóch wielkości: ubytku Δy równego $-\beta y(t) \Delta t$ oraz nowo zarażonych $\Delta y = \beta x(t) y(t) \Delta t$. Można więc napisać:

$$(10.1c) \quad \Delta y = [\beta x(t) \cdot y(t) - \gamma \cdot y(t)] \Delta t.$$

Jeśli Δt będziemy czynić coraz to mniejszym, czyli z Δt będziemy dążyć do zera, odpowiednie wyrażenia $\frac{\Delta y}{\Delta t}$, $\frac{\Delta x}{\Delta t}$ i $\frac{\Delta z}{\Delta t}$ będą dążyć do odpowiednich pochodnych $\frac{dy}{dt}$, $\frac{dx}{dt}$, $\frac{dz}{dt}$, i z równań (10.1a)

-(10.1c) otrzymamy następujący układ równań różniczkowych:

$$(10.2) \quad \begin{cases} \frac{dx}{dt} = -\beta xy \\ \frac{dy}{dt} = \beta xy - \gamma y \\ \frac{dz}{dt} = \gamma y \end{cases}$$

Zmienne $x(t)$, $y(t)$, $z(t)$ są oczywiście powiązane zależnością $x(t) + y(t) + z(t) = N$.

Eliminując y z pierwszego i trzeciego równania i oznaczając $\frac{z}{\beta} = \xi$ oraz $x(0) = x_0$ otrzymujemy

$$(10.3) \quad x = x_0 e^{-z/\xi}$$

$$(10.4) \quad \frac{dz}{dt} = \gamma(N - z - x_0 e^{-z/\xi})$$

Dokładne rozwiązanie równania (10.4) nie jest znane, ale można rozwinąć występującą w nim funkcję wykładniczą w szereg; biorąc trzy pierwsze składniki tego rozwinięcia można równanie (10.4) zastąpić innym, które już łatwo rozwiązać:

$$(10.4a) \quad \frac{dz}{dt} = \gamma \left\{ N - x_0 + \left(\frac{x_0}{\xi} - 1 \right) z - \frac{x_0}{2\xi^2} z^2 \right\}.$$

Rozwiązaniem tego równania jest funkcja

$$(10.5) \quad z = \frac{\xi^2}{x_0} \left\{ \frac{x_0}{\xi} - 1 + \alpha \operatorname{tgh} \left(\frac{1}{2} \alpha \gamma t - \Phi \right) \right\},$$

gdzie

$$\alpha = \left\{ \left(\frac{x_0}{\xi} - 1 \right)^2 + \frac{2x_0 y_0}{2} \right\}^{\frac{1}{2}},$$

$$\Phi = \operatorname{arc} \operatorname{tgh} \frac{1}{\alpha} \left(\frac{x_0}{\xi} - 1 \right)$$

Interesującą rzeczą jest wiedzieć, jaki będzie zasięg epidemii, czyli ile osób obejmie ona po nieskończeniu długim czasie. Ostatecznie wszystkie osoby, które zachorowały, powinny zostać wcześniej lub później izolowane, tak że za miarę zasięgu epidemii może służyć wielkość $z_\infty = \lim_{t \rightarrow \infty} z(t)$. Z (10.5) otrzymujemy

$$(10.6) \quad z_\infty = \frac{\xi^2}{x_0} \left(\frac{x_0}{\xi} - 1 + \alpha \right)$$

Jeśli w wyrażeniu na α ze wzoru (10.5) składnik $2x_0y_0$ jest mały w porównaniu z $(\frac{x_0}{\varrho} - 1)^2$ (co jest słuszne, np. gdy x_0 jest rzędu kilka tysięcy, a ϱ rzędu kilka jednostek), wtedy w przybliżeniu można przyjąć, że $\alpha = \frac{x_0}{\varrho} - 1$. Po podstawieniu tej wartości na α do wzoru (10.6) otrzymujemy

$$(10.7) \quad z_{\infty} \cong 2\varrho(1 - \frac{\varrho}{x_0})$$

Z równości (10.7) widzimy, że warunkiem koniecznym na to, aby epidemia mogła się rozwinąć, jest, aby ilość podatnych w chwili rozpoczęcia epidemii, czyli x_0 było większe od wielkości ϱ , gdyż tylko wtedy wartości na z są dodatnie. Szczegółową dyskusję równania (10.4a) przeprowadził Kendall w pracy [19].

Ten sam model epidemii możemy rozpatrywać w ujęciu stochastycznym. Zamiast pytać się o funkcje $x(t)$, $y(t)$, $z(t)$ możemy pytać się o prawdopodobieństwa $P_{r,s}(t)$ zdarzenia, że w chwili t będzie r zdrowych, podatnych na zarażenie oraz s zarażających chorych. Zakładając, że prawdopodobieństwo nowego zarażenia w przedziale czasowym o długości Δt jest rzędu $\beta r s \Delta t$, a prawdopodobieństwo odizolowania $\gamma s \Delta t$, można napisać równania różniczkowe na szukane prawdopodobieństwa:

$$(10.8) \quad \begin{cases} \frac{dP_{r,s}(t)}{dt} = \beta(r+1)(s-1)P_{r+1,s-1}(t) \\ \quad - \beta s(r + \frac{\gamma}{\beta})P_{r,s}(t) + \gamma(s+1)P_{r,s+1}(t) \\ \frac{dP_{N,a}(t)}{dt} = -\beta a(N + \frac{\gamma}{\beta}) \cdot P_{N,a}(t), \end{cases}$$

gdzie $0 \leq r + s \leq N + a$, $0 \leq r \leq N$, $0 \leq s \leq N + a$,

z warunkiem początkowym, że w chwili $t = 0$ było N osób zdrowych i a osób chorych:

$$(10.8a) \quad P_{N,a}(0) = 1.$$

Równania te można rozwiązać przy użyciu transformat fourierowskich, jednak otrzymane rozwiązania są bardzo złożone. Po pewnych operacjach matematycznych można z nich uzyskać wyrażenia na prawdopodobieństwa, że po nieskończenie długim czasie epidemia obejmie w osób: $P_w = \lim_{t \rightarrow \infty} P_{N-w,0}(t)$, $0 \leq w \leq N$. Mają one znaczenie przy obserwowaniu wielokrotnej realizacji tego samego procesu sto-

chastycznego, np. jeśli obserwowane populacje to rodziny, a dane, którymi dysponujemy, to ilości członków tych rodzin, które zachorowały na badaną chorobę. Bailey [2] podaje dla populacji o liczności $N = 1, 2, 3, 4$ i 5 oraz $a = 1$ wyrażenia na p_w oraz równania na nieznaną wartość parametru $g = \frac{\lambda}{\beta}$ wyprowadzone metodą największej wiarygodności, ponadto zamieszcza kilka diagramów obrazujących rozkłady wartości p_w dla wybranych wartości N . Kendall [19] wykazał, że rozkłady p_w dla $N \leq g$ są typu "L", a dla $N > g$ są one bimodalne. Dla rozkładów typu "L" najbardziej prawdopodobnymi wartościami są wartości małe, czyli epidemia ma znikome szanse na rozwinięcie się. Otrzymane rozkłady bimodalne są tego typu, że wartości modalne są położone na krańcach możliwych wartości w . Oznacza to, że najbardziej prawdopodobnymi są epidemie albo o dużym, albo o małym zasięgu.

Whittle (cyt. za [2]) badał prawdopodobieństwo π_1 , że epidemia obejmie nie więcej niż i -tą część populacji. Otrzymał on następujące nierówności w zależności od wielkości N , czyli liczności rozpatrywanej populacji:

$$(10.9) \quad \left\{ \begin{array}{l} \left(\frac{g}{N}\right)^a \leq \pi_1 \leq \left\{\frac{g}{N(1-i)}\right\}^a, \text{ gdy } g < N < (1-i); \\ \left(\frac{g}{N}\right)^a \leq \pi_1 \leq 1, \text{ gdy } N(1-i) < g < N; \\ \pi_1 = 1 \text{ gdy } N < g. \end{array} \right.$$

Widzimy, że prawdopodobieństwo rozszerzenia się epidemii na więcej niż i -tą część populacji wynosi 0 dla $N \leq g$, podczas gdy przy małym i oraz $g < N$ prawdopodobieństwo to w przybliżeniu równa się $1 - \left(\frac{g}{N}\right)^a$.

Kendall przybliżał rozważany proces dwoma innymi, będącymi znanymi procesami urodzin i śmierci. Jeśli mianowicie $g \geq N$, a więc prawdopodobieństwo wygaśnięcia epidemii przed objęciem całej populacji jest pewne, przyjmował on, że rozważany proces epidemii można przybliżyć procesem urodzin i śmierci o intensywnościach urodzin $\beta \cdot N$ i śmierci λ . Wartość średnia liczby osób objętych taką epidemią (łącznie z a osobami zarażającymi w chwili $t = 0$) wynosi

$$(10.10) \quad \bar{w} = \frac{a \cdot N}{g - N}, \text{ gdy } g \geq N.$$

Gdy $g < N$, wtedy prawdopodobieństwo wygaśnięcia epidemii w skończonym czasie wynosi $\left(\frac{g}{N}\right)^a$ i wtedy możliwe są dwa sposoby rozwinęcia się epidemii, nazwijmy je sposobami A i B odpowiednio z prawdopodobieństwami

$$(10.11) \quad P(A) = \left(\frac{g}{N}\right)^a \quad \text{i} \quad P(B) = 1 - \left(\frac{g}{N}\right)^a$$

W przypadku A epidemia wygasa w skończonym czasie, a więc rozważany proces możemy przybliżyć procesem urodzin i śmierci o intensywnościach $\beta \cdot N$ i γ . W przypadku B epidemia nie wygasa w skończonym czasie, a więc ilości zarażonych i zarażających nie są równe zero dla żadnego skończonego t . Możemy wtedy użyć modelu deterministycznego, opisanego równaniami (10.2), przy czym $(x(t), y(t), z(t))$ dla $t = 0$ wynosi $(N, a, 0)$, a dla $t = \infty$ równa się $(N - \zeta a, 0, \zeta)$, gdzie ζ jest całkowitą ilością osób objętych epidemią po upływie nieskończonego czasu (łącznie z początkowymi a osobami). Ponieważ między ilością osób izolowanych z a ilością osób podatnych x musi zachodzić związek (10.4), otrzymujemy następujące równanie na nieznaną wartość ζ :

$$(10.12) \quad a - \zeta + N(1 - e^{-\zeta/g}) = 0$$

Ilość nowo powstałych zachorowań podczas rozważanej epidemii wynosi $\zeta - a$. Ostatecznie więc na średnią ilość nowych zachorowań w przypadku epidemii, dla której $g < N$, otrzymujemy wyrażenie

$$(10.13) \quad \bar{w} = \left(\frac{g}{N}\right)^a \cdot \frac{a \cdot g}{N - g} + \left\{1 - \left(\frac{g}{N}\right)^a\right\} (\zeta - a), \quad \text{gd}y \quad g < N$$

Wartości przybliżone na \bar{w} ze wzorów (10.10) i (10.13) były porównywane z wartościami dokładnymi, uzyskanymi z wyrażeń na prawdopodobieństwa p_w . Zgodność obu wartości jest duża i prowadzi do praktycznego wniosku, że przybliżenia Kendalla są wystarczająco dobre, aby móc je stosować w praktyce.

W książce Bailey'a [2] można znaleźć przykłady zastosowań powyższej teorii do opisu zachorowań na szkarlatynę, jak również referencje do dalszej literatury na ten temat.

11. Model epidemii z uwzględnieniem okresu inkubacji

Omawiając w poprzednich rozdziałach modele epidemiczne zakładaliśmy, że indywidua zarażone od razu stają się zarażającymi

w stosunku do reszty populacji podatnych. Dla wielu chorób jest to nieprawda. Zarazek, przedostawszy się do organizmu, przeżywa w nim okres utajonego rozwoju, po czym dopiero po upływie pewnego czasu zaczynają się manifestować objawy choroby i osobnik staje się zaraźliwy dla innych. Tak więc w rozwoju choroby powinniśmy uwzględnić 3 okresy: 1^o okres utajonego rozwoju (the latent period), czyli okres od momentu zarażenia się aż do momentu stania się zaraźliwym dla otoczenia; 2^o okres zaraźliwości (the infectious period), czyli okres w czasie którego chory osobnik jest źródłem zarażeń dla innych podatnych osobników ze swego otoczenia; 3^o okres inkubacji (the incubation period), czyli okres od momentu zarażenia się do momentu wystąpienia objawów danej choroby. Najczęściej przy poważnych zaraźliwych chorobach osobnik, u którego wystąpią objawy danej choroby bywa izolowany od reszty podatnych, tak że nie ma możliwości zarażenia innych, a więc praktycznie okres inkubacji jest sumą okresu utajonego rozwoju i okresu zaraźliwości.

Dla ospy czarnej (variola vera) okres inkubacji wynosi około 2 tygodni. Może on być różny dla różnych osobników (np. we wrocławskiej epidemii z roku 1963 zanotowano okresy inkubacji i 7-dniowe i 21-dniowe). Naturalną rzeczą jest przyjąć, że okres inkubacji jest zmienną losową o jakiejś dystrybucji. Podobnie można zakładać, że i okresy utajonego rozwoju i okresy zaraźliwości są zmiennymi losowymi z odpowiednimi dystrybucjami (rozkładami prawdopodobieństwa).

Bharucha-Reid [6] rozpatrywał model epidemii z uwzględnieniem czasu utajonego rozwoju w sensie przytoczonej przez nas za Bailey'em definicji. Choroba jest tego rodzaju, że u osobnika zarażonego następuje okres utajonego rozwoju choroby o długości τ , a następnie każdy z osobników może zarazić $n = 0, 1, 2, \dots$ innych osobników odpowiednio z prawdopodobieństwami q_n , sam zaś jest izolowany lub traci zaraźliwość. Czas τ jest zmienną losową o gęstości prawdopodobieństwa $g(\tau)$. U każdego z nowo zarażonych osobników następuje okres utajonego rozwoju, który znowu może trwać τ jednostek czasu z prawdopodobieństwem $\int_0^{\tau} g(\tau) d\tau$, po czym każdy z osobników może zarazić $n = 0, 1, 2, \dots$ innych osobników z tymi samymi prawdopodobieństwami q_0, q_1, q_2, \dots

Niech $x(t)$ będzie liczbą zarażonych osobników w chwili t . Oznaczmy prawdopodobieństwo zdarzenia, że w chwili t będzie do-

kładnie x zarażonych osobników symbolem $p(x, t)$. A więc

$$p(x, t) = \Pr \{x(t) = x\} \quad \text{dla } x \geq 0$$

przy warunku początkowym, że w chwili $t=0$ został zarażony 1 osobnik. Aby otrzymać szukane prawdopodobieństwa $p(x, t)$ należy dla każdego $t < t$ uwzględnić możliwość wystąpienia zaraźliwości w momencie τ (gęstość prawdopodobieństwa tego zdarzenia wynosi $g(\tau)$) i pomnożyć ją przez warunkowe prawdopodobieństwo zdarzenia polegającego na tym, że jeśli w chwili τ wystąpiła zaraźliwość u pierwszego osobnika, to dalszy przebieg procesu był taki, że w chwili t czyli po czasie $t - \tau$ było x zarażonych. Otrzymujemy w ten sposób

$$(11.1) \quad p(x, t) = \Pr \{x(t) = x\} = \int_0^t \Pr \{x(t) = x \mid \tau\} g(\tau) d\tau$$

Zastanówmy się, jak można inaczej przedstawić $\Pr \{x(t) = x \mid \tau\}$. Wiadomo tu, że w chwili τ nastąpił moment zaraźliwości, w wyniku którego została zarażona pewna ilość osób, powiedzmy n osób, które to osoby z kolei zapoczątkowały nowy cykl epidemii, tak że ostatecznie w chwili t było x osób zarażonych. Gdyby n wynosiło 1, to mielibyśmy sytuację podobną jak w chwili $t = 0$, i wtedy prawdopodobieństwo warunkowe, że od momentu τ do momentu t zostanie zarażonych x osób wynosi $p(x, t - \tau)$. Jeśli $n \neq 1$, to jak gdyby zmienia się warunek początkowy; można wykazać, że jeśli $\pi(z, t)$ jest funkcją tworzącą dla prawdopodobieństwa $p(x, t)$ przy warunku początkowym $x(0) = 1$, to $\pi^n(z, t)$ jest funkcją tworzącą dla prawdopodobieństw $p(x, t)$ przy warunku początkowym $x(0) = n$. Odpowiednie prawdopodobieństwa, czyli współczynniki rozwinięcia funkcji $\pi^n(z, t)$ przy potęgach z mają wtedy postać

$$(11.2) \quad p^{(n)}(x, t) = \sum_{i_1+i_2+\dots+i_n=x} \prod_{k=1}^n p(i_k, t).$$

W naszym zadaniu od osobnika inicjującego epidemię mogło się zarazić w chwili wystąpienia jego zaraźliwości 0, 1, 2, ..., n , ... osobników z prawdopodobieństwem $q_0, q_1, q_2, \dots, q_n, \dots$ Przypuśćmy, że zarażyło się n osób. Jeśli tak, to prawdopodobieństwo zdarzenia polegającego na tym, że tych n osób spowoduje, że po czasie $t - \tau$ będzie x zarażonych, wynosi $p^{(n)}(x, t - \tau)$ i w myśl (11.2) jest równe

$$(11.3) \quad p^{(n)}(x, t - \tau) = \sum_{i_1+i_2+\dots+i_n=x} \prod_{k=1}^n p(i_k, t - \tau)$$

Podany wzór zachodzi dla $n = 0, 1, 2, \dots$. Jeśli w chwili t ma być x zarażonych, to z prawdopodobieństwem $q_1 \cdot p^{(1)}(x, t - \tau)$ zaistniała taka sytuacja, że w chwili τ został zarażony jeden osobnik i od niego bezpośrednio lub pośrednio zaraził się inni osobnicy, tak że w chwili t było x zarażonych; z prawdopodobieństwem $q_2 \cdot p^{(2)}(x, t - \tau)$ zaistniała taka sytuacja, że w chwili τ zostało zarażonych dwóch osobników i od nich bezpośrednio lub pośrednio zaraził się inni osobnicy tak, że w chwili t było x zarażonych i ogólnie mówiąc, z prawdopodobieństwem $q_n \cdot p^{(n)}(x, t - \tau)$ w chwili τ zostało zarażonych n osobników, którzy spowodowali, że w chwili t było x zarażonych. Przyjmując, że dla każdego $0 < \tau < t$

$$p^{(0)}(x, t - \tau) = \begin{cases} 1 & \text{dla } x = 0 \\ 0 & \text{dla } x > 0 \end{cases}$$

możemy przedstawić warunkowe prawdopodobieństwo $\Pr \{x(t) = x | \tau\}$ jako następującą sumę:

$$(11.4) \quad \Pr \{x(t) = x | \tau\} = \sum_{n=0}^{\infty} q_n \cdot p^{(n)}(x, t - \tau)$$

Podstawiając wyrażenie (11.4) do wzoru (11.1) otrzymujemy wyrażenie na $p(x, t)$:

$$(11.5) \quad p(x, t) = \int_0^t \sum_{n=0}^{\infty} q_n \cdot p^{(n)}(x, t - \tau) g(\tau) d\tau \quad \text{dla } x > 0 \text{ i } x \neq 1$$

Aby otrzymać wyrażenie $p(1, t)$, do prawdopodobieństwa określonego wzorem (11.5), które wynikało z sytuacji, że na pewno do czasu t wystąpił okres zaraźliwości u pierwszego osobnika, należy dodać prawdopodobieństwo zdarzenia, że okres taki do czasu t jeszcze nie wystąpił. Prawdopodobieństwo tego ostatniego zdarzenia wynosi

$$1 - G(t) = 1 - \int_0^t g(\tau) d\tau. \text{ Ostatecznie więc:}$$

$$(11.5a) \quad p(1, t) = \int_0^t \sum_{n=0}^{\infty} q_n p^{(n)}(x, t - \tau) g(\tau) d\tau + 1 - G(t)$$

Wprowadzimy teraz wzór na funkcję tworzącą prawdopodobieństw $p(x, t)$ czyli funkcję

$$(11.6) \quad \pi(z, t) = p(0, t) + p(1, t)z + p(2, t)z^2 + \dots + p(x, t)z^x + \dots$$

Mnożąc wyrażenie (11.5a) przez z , a wyrażenie $p(x, t)$ z wzoru (11.5)

przez z^x i sumując dla $x = 0, 1, 2, \dots$ otrzymujemy:

$$\sum_{x=0}^{\infty} p(x, t) z^x = \int_0^t \sum_{n=0}^{\infty} q_n \sum_{x=0}^{\infty} p^{(n)}(x, t - \tau) z^x g(\tau) d\tau + z [1 - G(t)].$$

Uwzględniając zależność (11.6) i (11.2), otrzymujemy:

$$(11.7) \quad \pi(z, t) = \int_0^t \sum_{n=0}^{\infty} q_n \pi^n(z, t - \tau) g(\tau) d\tau + z [1 - G(t)].$$

Jeśli wprowadzimy funkcję tworzącą prawdopodobieństwa q_n

$$(11.8) \quad h(z) = q_0 + q_1 z + \dots + q_n z^n + \dots$$

to zależność (11.7) możemy zapisać w nieco prostszej postaci:

$$(11.9) \quad \pi(z, t) = \int_0^t h[\pi(z, t - \tau)] g(\tau) d\tau + z [1 - G(t)]$$

Otrzymaliśmy w ten sposób równanie na nieznaną funkcję $\pi(z, t)$ w postaci równania całkowego. W praktyce znamy kształt funkcji $g(t)$ i $h(z)$, a szukamy $\pi(z, t)$.

Z wzoru (11.9) możemy otrzymać łatwo równanie na oczekiwaną liczbę zarażonych osób w chwili t :

$$M(t) = \sum_{k=1}^{\infty} k \cdot p(k, t) = \left. \frac{\partial \pi(z, t)}{\partial z} \right|_{z=1}$$

Różniczkując równanie (11.9) ze względu na parametr z otrzymujemy

$$M(t) = \int_0^t \left. \frac{\partial h}{\partial \pi} \cdot \frac{\partial \pi}{\partial z} \right|_{z=1} \cdot g(\tau) d\tau + 1 - G(t)$$

skąd po wykonaniu działań

$$(11.10) \quad M(t) = 1 - G(t) + K \int_0^t M(t - \tau) g(\tau) d\tau, \quad \text{gdzie } K = \sum_{n=0}^{\infty} n q_n$$

Przy znanych $G(t)$ oraz stałej K można próbować znaleźć wartość oczekiwaną rozważanego procesu spełniającą powyższe równanie całkowe.

Dotychczas w tym rozdziale mówiliśmy o "zaraźliwości", o "zarażeniu", o liczbie osób "zarażonych". Należy jednak zwrócić uwagę, że proces tutaj opisany jest procesem ogólniejszym. Można on opisywać również proces urodzin i śmierci. Gdybyśmy, np. jako fun-

kcję $h(z)$, przyjęli funkcję

$$h(z) = q_0 + q_2 z^2,$$

to mielibyśmy do czynienia z procesem, w którym nowo powstała cząsteczka po czasie t dzieli się na dwie części z prawdopodobieństwem q_2 lub ginie z prawdopodobieństwem q_0 .

Proces Furry'ego-Yule'a, opisany w rozdziale 4 jest wyznaczony przez funkcję $h(z) = z^2$ oraz $g(t) = e^{-\lambda t}$. Równanie (11.9) na funkcję tworzącą prawdopodobieństwa tego procesu ma postać:

$$(11.11) \quad \pi(z, t) = z e^{-\lambda t} + \int_0^t \pi^2(z, t - \tau) \lambda e^{-\lambda \tau} d\tau.$$

Równanie to możemy zamienić na równanie różniczkowe o pochodnych cząstkowych. Różniczkując mianowicie równość (11.11) obustronnie ze względu na t , otrzymujemy:

$$\begin{aligned} \frac{d\pi}{dt} &= -\lambda z e^{-\lambda t} + \int_0^t \frac{d\pi^2(z, t - \tau)}{d\tau} \cdot \lambda e^{-\lambda \tau} d\tau + \pi^2(z, t - \tau) \cdot \lambda e^{-\lambda \tau} \Big|_{\tau=t} = \\ &= -\lambda z e^{-\lambda t} - \int_0^t \frac{d\pi^2(z, t - \tau)}{d\tau} \cdot \lambda e^{-\lambda \tau} d\tau + \pi^2(z, 0) \cdot \lambda e^{-\lambda t} = \\ &= -\lambda z e^{-\lambda t} + z^2 \lambda e^{-\lambda t} - \left[\pi^2(z, t - \tau) \cdot \lambda e^{-\lambda \tau} \Big|_0^t + \lambda \int_0^t \pi^2(z, t - \tau) \cdot \lambda e^{-\lambda \tau} d\tau \right] = \\ &= -\lambda z e^{-\lambda t} + \pi^2(z, t) \cdot \lambda - \lambda \int_0^t \pi^2(z, t - \tau) \lambda e^{-\lambda \tau} d\tau = \\ &= \pi^2(z, t) \cdot \lambda - \lambda \left[z e^{-\lambda t} + \int_0^t \pi^2(z, t - \tau) \cdot \lambda e^{-\lambda \tau} d\tau \right] = \lambda \pi(\pi - 1). \end{aligned}$$

Ostatecznie więc otrzymujemy następujące równanie różniczkowe:

$$(11.12) \quad \frac{d\pi}{dt} = \lambda \pi(\pi - 1)$$

Łatwo sprawdzić, że rozwiązaniem tego równania jest funkcja

$$(11.13) \quad \pi(z, t) = \frac{z \cdot e^{-\lambda t}}{1 - z(1 - e^{-\lambda t})}$$

Rozwijając tę funkcję w szereg otrzymujemy

$$(11.14) \quad \pi(z, t) = \sum_{n=1}^{\infty} e^{-\lambda t} (1 - e^{-\lambda t})^{n-1} \cdot z^n$$

Widzimy, że współczynniki przy z^n są identyczne z prawdopodobieństwami procesu Furry'ego - Yule'a otrzymanymi wcześniej na innej drodze jako wzory (4.2).

12. Model epidemii na płaszczyźnie

Neyman i Scott opisali [28] następujący model epidemii: Rozpatrujemy taką chorobę, dla której okres utajonego rozwoju jest wielkością stałą i przypuścmy, że wynosi on T . Po upływie tego okresu osobnik zarażony staje się zaraźliwy dla otoczenia i może zarażić inne osoby, z którymi będzie akurat w kontakcie. Weźmy pod uwagę płaszczyznę dwuwymiarową, na której żyje i porusza się rozważana populacja. Przypuścmy, że w chwili $t=0$ w pewnym punkcie u tej płaszczyzny znalazł się osobnik A , u którego akurat wystąpił okres zaraźliwości. Wobec tego może on zarażić pewną ilość osób, powiedzmy $v(u)$. Ilość ta jest zmienną losową, zależną od miejsca, w którym osobnik zarażający właśnie się znajduje. Weźmy pod uwagę s rozłącznych obszarów R_1, R_2, \dots, R_s znajdujących się na rozważanej płaszczyźnie. Po upływie czasu nT w każdym z tych obszarów zostanie zarażona pewna liczba osób, które będą stanowiły kontynuację epidemii zapoczątkowanej przez osobę A . Oznaczmy ilość nowych zachorowań w obszarze o numerze i przez $N_{ni}(u)$. Tak więc po czasie nT łączna ilość nowych zachorowań w obszarze H wywodzących się od osoby A jest scharakteryzowana przez wektor

$$N_n(u) = (N_{n1}(u), N_{n2}(u), \dots, N_{ns}(u))$$

Neyman i Scott wyprowadzili wzory na funkcję tworzącą wektora $N_n(u)$ przy założeniu, że znane są funkcje tworzące prawdopodobieństwa dla zmiennej losowej $v(u)$ oraz funkcja $f(x,y)$ wyrażająca prawdopodobieństwo przemieszczenia się osobnika w czasie T z punktu x do punktu y . W pracy ich można również znaleźć rozważania, dotyczące łącznej liczby osób objętych epidemią, wpływu akcji immunizacyjnej na tę liczbę, oraz opis epidemii inicjowanej przez losowo pojawiających się osobników zarażających na obszarze H (np. wskutek mutacji). Autorzy nie podają żadnych przykładów numerycznych, ilustrujących opisane twierdzenia. Bartoszyński, Łoś i Wycech - Łoś [5] badali, jak zachowuje się funkcja tworząca wektora $N_n(u)$ po nieskończeniu długim czasie trwania epidemii.

13. Zagadnienia mutagenезy

Istnieje duża ilość prac, poświęconych zagadnieniom tzw. mutagenезy. Weźmy pod uwagę bakterię wrażliwą na działanie antybiotyku. W potomstwie tej bakterii po pewnym czasie można wykryć bakterie, odporne na działanie tego antybiotyku. Powstały one wskutek tzw. spontanicznej mutacji odpowiednich genów. Podobnie jak przy procesie urodzin i śmierci można zdefiniować pewien proces urodzin, śmierci i mutacji z odpowiednimi intensywnościami λ , μ , ν , przy czym zdarza się, że zmutowana bakteria ma inną intensywność podziału. Zagadnienie to zostało bardziej szczegółowo omówione w książce Bailey'a [3]. Zagadnieniem tym zajmował się m. in. Nissen-Mayer [29], którego interesowało prawdopodobieństwo $P_0(t)$, że po czasie t będzie 0 zmutowanych bakterii. Chciał on wyznaczyć schematy terapeutyczne, to znaczy taki sposób podawania antybiotyku, żeby zmaksymizować prawdopodobieństwo $P_0(t)$.

Również w teorii karcinogenезy mówi się o mutacji komórek w komórki rakowe. Zagadnienie to posiada bardzo obszerną literaturę, a z polskich prac wymienić tu należy prace Kloneckiego [21], [23]. Przegląd tych zagadnień można znaleźć w [30].

Doświadczenia wykazały, że jeśli daną hodowlę poddamy pewnym zabiegom (np. naświetlaniu promieniami X), to ilość mutantów w potomstwie wzrasta. Często okazuje się jednak, że odpowiednio zmieniiony gen nie manifestuje się natychmiast (to znaczy nie można stwierdzić zmiany od razu), ale dopiero po kilku pokoleniach. Modele, wyjaśniające to zjawisko tak zwanej indukowanej mutacji dla bakterii zbudował m. in. Armitage [1].

Badanie zjawiska mutacji dla wirusów jest znacznie bardziej utrudnione z powodu trudności eksperymentalnych, gdyż możemy rachoować wirusy tylko w określonym stadium rozwoju. Green i Krieg [15] rozważając różne modele powstawania mutantów, zaproponowali szereg doświadczeń, które miały wyjaśnić zagadnienie mutacji tzw. fagów T4 z formy r do formy r+. Fagi te po przedostaniu się do komórki bakteryjnej rozpoczynają okres wegetatywnego rozwoju, rozmnażając się przez podział. Po kilku godzinach niektóre fagi osiągną tzw. stadium dojrzałości i już więcej nie rozmnażają się. Inne dzielą się dalej tak długo, aż komórka bakterii ulegnie znisz-

czeniu i cała rodzina znajdujących się w niej fagów uwolni się. Eksperymentalnie potrafimy wykryć jedynie fagi, które osiągnęły stadium dojrzałości.

Kimball [20] znalazł rozkład ilości dojrzałych fagów w rodzinie wyhodowanej wewnątrz jednej bakterii przyjmując, że fagi osiągają dojrzałość począwszy od k -tej generacji, przy czym prawdopodobieństwo osiągnięcia dojrzałości przez faga, należącego do k -tej lub wyższej generacji wynosi $\frac{1}{2}$, natomiast cały cykl rozwoju (aż do zniszczenia bakterii) trwa $k+m$ generacji.

Jeśli chodzi o zmutowane formy fagów, to zakłada się następujący mechanizm: gen, w którym nastąpiła zmiana i który ulegnie później mutacji, w początkowym stadium rozwoju fagów jest przekazywany w populacji tylko przez jednego faga. Po pewnym czasie w genie znajdującym się w fagu - nosicielu N nastąpi taka zmiana jakościowa, że gen ten należy uznać za zmutowany. Począwszy od tego momentu gen ten ulega podziałowi, to znaczy obydwa fagi pochodzące od faga N i ich potomstwo należą już do formy zmutowanej. Zmiany w genie mogą się nie ujawnić, jeśli fag - nosiciel osiągnie stadium dojrzałości, zanim w genie dokona się zmiana jakościowa. Kimball [20] podał wzory na oczekiwane liczby zmutowanych fagów, które osiągnęły stadium dojrzałości przy założeniach wyżej opisanego modelu. W pracy jego można znaleźć również pewne konfrontacje wyprowadzonych przez niego wzorów z danymi doświadczalnymi Greena i Kriega.

14. O procesach stochastycznych w genetyce

Przypuśćmy, że mamy populację składającą się z N osobników. Każdy z nich jest nosicielem dwóch genów. Możliwe są dwa rodzaje genów; nazwijmy je a i A . Wobec tego każdy osobnik może być nosicielem genów aa , aA lub AA . Po pewnym okresie czasu cała ta populacja umiera, a na jej miejsce pojawia się potomstwo, liczące również N osobników. Każdy z nich posiada również po dwa geny dobране na zasadzie losowania dwóch spośród $2N$ genów pierwszego pokolenia. Taka sytuacja istnieje, np. przy uprawie roślin jednoletnich, gdy w następnym roku wysiewa się nasiona zebrane poprzedniego roku.

Przypuśćmy, że w pierwszym pokoleniu było j genów typu A . Wtedy prawdopodobieństwo wylosowania genu typu A jest $\frac{j}{2N}$, a prawo-

podobieństwo wylosowania genu typu a odpowiednio $1 - \frac{j}{2N}$. Wobec tego prawdopodobieństwo, że w następnym pokoleniu otrzymamy k genów A wynosi:

$$(14.1) \quad p_{jk} = \binom{2N}{k} \cdot \left(\frac{j}{2N}\right)^k \cdot \left(1 - \frac{j}{2N}\right)^{2N-k}$$

W ten sposób dla dowolnego j oraz k ($j = 0, 1, 2, \dots, 2N$; $k = 0, 1, 2, \dots, 2N$) możemy określić całą macierz prawdopodobieństw p_{jk} , że jeśli w pierwszym pokoleniu było j genów typu A to w następnym pokoleniu genów tych będzie k . Macierz tę nazywamy macierzą prawdopodobieństw przejścia. Każdą z możliwości $j = 0, 1, \dots, 2N$ będziemy nazywać stanem rozważanego układu. W naszym przykładzie genetycznym mamy $2N + 1$ możliwych stanów. Dwa spośród nich, mianowicie $j = 0$ oraz $j = 2N$ wyróżniają się jako tzw. stany pochłaniające. Jeśli układ znajdzie się w jednym z tych stanów, to nie może już żadnego z nich opuścić. Praktyczny sens (np. dla hodowcy) mają następujące pytania: jakie są prawdopodobieństwa przejścia układu do jakiegoś stanu v i po jak długim czasie to nastąpi?

W zagadnieniu tym mamy do czynienia ze specjalnym przypadkiem procesów stochastycznych: czas nie jest tu parametrem ciągłym. Zmiany w układzie występują tylko co pewien czas, i wobec tego wystarczy znajomość układu tylko w pewnych określonych momentach czasu. Ponadto w przykładzie genetycznym omawianym powyżej można było wyznaczyć prawdopodobieństwa poszczególnych stanów w pewnej generacji n na podstawie znajomości stanu w poprzedniej generacji $n - 1$. Proces, rozpatrywany w dyskretnych momentach czasu, w którym prawdopodobieństwa poszczególnych stanów w chwili $t = t_n$ zależą tylko od stanu w chwili $t = t_{n-1}$, nie zależą natomiast od stanów w chwilach wcześniejszych t_{n-2}, t_{n-3}, \dots , nazywamy łańcuchem Markowa. Metody badania prawdopodobieństwa dla łańcuchów Markowa omówione są, np. w książce Feller'a [10].

Zasadniczym pojęciem w teorii łańcuchów Markowa jest tzw. macierz prawdopodobieństw przejścia $p_{ij}^{(n)}$, czyli prawdopodobieństw warunkowych, że jeśli układ w momencie początkowym (zerowym) znajdował się w stanie i , to po n przejściach (krokach) znajdzie się w stanie j , przy czym wskaźniki i, j dotyczą wszystkich możliwych stanów, których może być w ogólnym przypadku nieskończona ilość.

Przypuśćmy, że mamy w urnie dwa geny, przy czym mogą to być geny aa, aA lub AA . Dwa geny a nazwijmy stanem E_0 , dwa różne geny - stanem E_1 i dwa geny AA stanem E_2 . Stany E_0 i E_2 są to tzw. sta-

ny homozygotyczne, stan E_1 - heterozygotyczny. Jeśli z urny tej wylosujemy dwa geny (losowanie ze zwracaniem), to otrzymany wynik zależy od tego, jaki był skład urny. I tak, jeśli w urnie były geny aa lub AA, to w wyniku losowania otrzymamy z prawdopodobieństwem 1 również dwa jednakowe geny. Jeśli jednak w urnie były dwa różne geny, to z prawdopodobieństwem $\frac{1}{4}$ otrzymamy dwa geny a, z prawdopodobieństwem $\frac{1}{2}$ dwa geny różne i z prawdopodobieństwem $\frac{1}{4}$ dwa geny A. Oznaczając przez p_{ik} ($i, k = 0, 1, 2$) prawdopodobieństwo przejścia ze stanu i do stanu k możemy wszystkie te prawdopodobieństwa zapisać w postaci macierzy:

$$P^{(1)} = \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix}$$

Znajdziemy teraz prawdopodobieństwa otrzymania poszczególnych stanów po dwóch losowaniach. Prawdopodobieństwa te zależą od stanu urny przed pierwszym losowaniem oraz od wyniku pierwszego losowania. I tak:

$$\begin{aligned} (2) \quad P_{00} &= P_{00} \cdot P_{00} + P_{01} \cdot P_{10} + P_{02} \cdot P_{20} = 1 \cdot 1 + 0 \cdot \frac{1}{4} + 0 \cdot 0 = 1 \\ (2) \quad P_{01} &= P_{00} \cdot P_{01} + P_{01} \cdot P_{11} + P_{02} \cdot P_{21} = 1 \cdot 0 + 0 \cdot \frac{1}{2} + 0 \cdot 0 = 0 \\ (2) \quad P_{02} &= P_{00} \cdot P_{02} + P_{01} \cdot P_{12} + P_{02} \cdot P_{22} = 1 \cdot 0 + 0 \cdot \frac{1}{4} + 0 \cdot 1 = 0 \\ (2) \quad P_{10} &= P_{10} \cdot P_{00} + P_{11} \cdot P_{10} + P_{12} \cdot P_{20} = \frac{1}{4} \cdot 1 + \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{4} \cdot 0 = \frac{3}{8} \\ (2) \quad P_{11} &= P_{10} \cdot P_{01} + P_{11} \cdot P_{11} + P_{12} \cdot P_{21} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 0 = \frac{1}{4} \\ (2) \quad P_{12} &= P_{10} \cdot P_{02} + P_{11} \cdot P_{12} + P_{12} \cdot P_{22} = \frac{3}{4} \cdot 0 + \frac{1}{2} \cdot \frac{1}{4} + \frac{3}{4} \cdot 0 = \frac{3}{8} \\ (2) \quad P_{20} &= P_{20} \cdot P_{00} + P_{21} \cdot P_{10} + P_{22} \cdot P_{20} = 0 \cdot 1 + 0 \cdot \frac{1}{4} + 1 \cdot 0 = 0 \\ (2) \quad P_{21} &= P_{20} \cdot P_{01} + P_{21} \cdot P_{11} + P_{22} \cdot P_{21} = 0 \cdot 1 + 0 \cdot \frac{1}{4} + 1 \cdot 0 = 0 \\ (2) \quad P_{22} &= P_{20} \cdot P_{02} + P_{21} \cdot P_{12} + P_{22} \cdot P_{22} = 0 \cdot 0 + 0 \cdot \frac{1}{4} + 1 \cdot 1 = 1 \end{aligned}$$

Łatwo zauważyć, że macierz $P^{(2)}$ można otrzymać jako iloczyn macierzy $P^{(1)}$ przez siebie samą, czyli $P^{(2)} = P^{(1)} \cdot P^{(1)}$. Ogólniej

słuszny jest następujący wzór:

$$(14.2) \quad P^{(n)} = P^{(1)} \cdot P^{(n-1)}, \quad \text{gdzie } n > 1$$

Dla naszego przykładu

$$P^{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ 0 & 0 & 1 \end{pmatrix}$$

$$P^{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{7}{16} & \frac{1}{8} & \frac{7}{16} \\ 0 & 0 & 1 \end{pmatrix}$$

i tak dalej.

Widzimy, że w rozważanej przez nas macierzy po dowolnej liczbie kroków pierwszy i trzeci wiersz pozostają takie same, czyli nie zależą od n . Prawdopodobieństwa drugiego wiersza zmieniają się w ten sposób, że ze wzrostem n zwiększa się prawdopodobieństwo stanu E_0 i E_2 , a zmniejsza prawdopodobieństwo stanu E_1 . Oznacza to, że jeśli w momencie zerowym były różne geny, to z biegiem czasu zmniejsza się heterogeniczność rozważanego układu. W naszym przykładzie prawdopodobieństwo, że układ pozostanie heterogeniczny zmniejsza się z każdym krokiem dwa razy.

Istnieją metody algebraiczne, które pozwalają znaleźć prawdopodobieństwa przejścia $p_{jk}^{(n)}$ ze stanu j do stanu k po n krokach bez n -krotnego mnożenia macierzy P . Można wykazać [10], że jeśli liczba możliwych stanów równa się N , to prawdopodobieństwa $p_{jk}^{(n)}$ dają się przedstawić w postaci sumy

$$(14.3) \quad p_{jk}^{(n)} = \frac{c_{jk}^{(1)}}{s_1^n} + \frac{c_{jk}^{(2)}}{s_2^n} + \dots + \frac{c_{jk}^{(a)}}{s_a^n}$$

gdzie s_1, s_2, \dots, s_a są pierwiastkami następującego równania:

$$(14.4) \quad \begin{vmatrix} p_{11} - s & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} - s & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{vmatrix} = 0$$

a $c_{jk}^{(r)}$ ($r = 1, \dots, a$) są to pewne stałe.

Można wykazać, że w naszym przykładzie

$$p_{11}^{(n)} = \frac{1}{2^n}$$

$$p_{10}^{(n)} = p_{12}^{(n)} = \frac{1}{2} \left[1 - \frac{1}{2^n} \right]$$

Ostatecznie, po nieskończonej ilości kroków układ nasz znajdzie się z prawdopodobieństwem 1 w stanie E_0 lub E_2 , bo

$$\lim_{n \rightarrow \infty} p_{1k}^{(n)} = \begin{cases} \frac{1}{2} & \text{dla } k = 0 \\ 0 & \text{dla } k = 1 \\ \frac{1}{2} & \text{dla } k = 2 \end{cases}$$

Stany E_0 i E_2 są tzw. stanami pochłaniającymi, a stan E_1 tzw. stanem chwilowym. Po nieskończonej więc liczbie generacji otrzymamy populację, w której z prawdopodobieństwem $1 - j/2N$ będą tylko geny a i z prawdopodobieństwem $j/2N$ będą tylko geny A (por. [10], str. 338). Jest to zgodne z intuicją, że jeśli w pewnym momencie geny A i a znajdują się w proporcji $j:(2N - j)$, to ich szanse przeżycia (utrzymania się w tej populacji) są w tym samym stosunku.

Nieco inną sytuację spotykamy w modelach genetycznych, w których uwzględniamy również możliwość mutacji poszczególnych genów. I tak Wright [36] zbudował model genetyczny, w którym każdy z N osobników danej populacji podobnie jak poprzednio jest nosicielem dwóch genów, przy czym po czasie T na miejsce tej populacji pojawia się inna, licząca również N osobników, z których każdy ma dwa geny, wybrane losowo ze zbioru $2N$ genów poprzedniej generacji. Dodatkowo w modelu tym zakłada się możliwość mutacji. Gen a może więc z prawdopodobieństwem α_1 zmienić się na gen A oraz przeciwnie - gen A może z prawdopodobieństwem α_2 zmienić się w gen a . Uwzględniając to zjawisko mutacji, przy założeniu, że pierwotnie było j genów typu a , otrzymujemy następujące prawdopodobieństwo wylosowania genu typu a :

$$(14.5) \quad p_j = \frac{j}{2N}(1 - \alpha_1) + (1 - \frac{j}{2N}) \cdot \alpha_2$$

i odpowiednio prawdopodobieństwo wylosowania genu typu A :

$$(14.6) \quad q_j = \frac{j}{2N} \cdot \alpha_1 + (1 - \frac{j}{2N}) \cdot (1 - \alpha_2)$$

Macierz prawdopodobieństw przejścia określona jest za pomocą prawdopodobieństw

$$(14.7) \quad p_{jk} = \binom{2N}{k} \cdot p_j^k \cdot q_j^{2N-k}$$

gdzie p_j i q_j określone zostały wzorami (14.5) i (14.6).

Można się zapytać o rozkład frakcji genów typu a po pewnym czasie t liczonym w ilościach generacji jako jednostkach. Jeśli frakcję tę oznaczymy przez X_t , to dla dużych N możemy frakcję tę przybliżyć pewną zmienną losową ciągłą $x(t)$, gdzie $0 \leq x \leq 1$. Oznaczmy gęstość prawdopodobieństwa zmiennej losowej $x(t)$ przez $u(x, t)$. Feller [9] podał równanie różniczkowe, jakie powinna spełniać funkcja $u(x, t)$, jak również zbadał różne własności rozwiązania tego równania.

Rozpatrując poprzednio granicę $\lim_{n \rightarrow \infty} p_{jk}^{(n)}$ stwierdziliśmy, że granica tego wyrażenia zależy od numeru wiersza, z którego to wyrażenie pochodzi, czyli od początkowego stanu układu. Może się jednak zdarzyć, że po pewnym czasie prawdopodobieństwa $p_{jk}^{(n)}$ będą dążyć do pewnej granicy p_k niezależnej od stanu początkowego E_j . Jeśli rozważany układ jest układem o skończonej ilości stanów, to zbieżność taka jest zagwarantowana, gdy istnieje takie r , że wszystkie wyrazy macierzy $P^{(r)}$ są dodatnie (twierdzenie Markowa), lub gdy wszystkie wyrazy przynajmniej jednej kolumny macierzy $P^{(r)}$ są dodatnie [12].

Prawdopodobieństwa graniczne p_k noszą nazwę prawdopodobieństw ergodycznych. Określają one pewien stan równowagi rozważanego układu. Mówimy wtedy, że układ jest w stanie równowagi (stanie stabilnym, equilibrium). Oznacza to, że niezależnie od stanu wyjściowego z biegiem czasu ustala się pewien graniczny rozkład prawdopodobieństw możliwych stanów systemu.

Dla dopiero co rozpatrywanego modelu Wright'a przy dużej liczności populacji, czyli przy dużym N i przy małym α_1 i α_2 rozkład frakcji genów a dla stanu równowagi można przybliżyć przez funkcję (cytowane za [25], str. 61)

$$(14.8) \quad u(x) = \frac{\Gamma(2\beta_1 + 2\beta_2)}{\Gamma(2\beta_1)\Gamma(2\beta_2)} x^{2\beta_2-1} (1-x)^{2\beta_1-1}$$

gdzie x ($0 \leq x \leq 1$) jest frakcją genów rozważanego typu, $\beta_1 = \alpha_1 \cdot 2N$, $\beta_2 = \alpha_2 \cdot 2N$. Funkcja ta jest gęstością prawdopodobieństwa, przy czym nie wykazuje ona koncentracji prawdopodobieństwa dla $x = 0$

if $x = 1$. Znaczy to, że prawdopodobieństwo zniknięcia genu a lub A w tak trwającej populacji wynosi 0.

Moran [25], [26] badał model genetyczny nieco bardziej złożony, w którym pokolenia mogą na siebie zachodzić: nie wszystkie osobniki giną po upływie czasu T i przekazują swe geny nowej generacji, ale w pewnym momencie czasu t jakiś z osobników umiera, zaś na jego miejsce pojawia się inny o genach, które zostały wybrane spośród wszystkich genów znajdujących się w populacji w chwili t -dt. Model ten może mieć zastosowanie do pewnych populacji ekologicznych lub hodowli komórkowych w warunkach ograniczonego rozwoju (liczność populacji jest w przybliżeniu stała dzięki np. ograniczonej ilości pożywienia; natomiast przez cały czas występuje wymieranie i narodziny osobników tej hodowli). Jeśli rozpatrywać nasz układ w momentach śmierci poszczególnych osobników, to będziemy tu znowu mieli pewien łańcuch Markowa, ale czas nie będzie już liczony w jednostkach T długości życia danej generacji, ale w odstępach odpowiadających odstępom między czasem śmierci poszczególnych osobników. Moran zbadał również, co się dzieje z prawdopodobieństwem $p_{jk}^{(n)}$ po nieskończenie długim czasie ($n \rightarrow \infty$).

Te same zagadnienia można rozważać przy rozmaitych innych założeniach co do warunków rozwoju i rozmnażania się badanej populacji (por. [26], [27]), jak również przy bogatszym garniturze genów ([22]).

Literatura cytowana

- [1] Armitage, P., The statistical theory of bacterial populations subject to mutation, *Journal of the Royal Statistical Society, Series B* 14 (1952), str. 1-33.
- [2] Bailey, N.T.J., *The mathematical theory of epidemics*, London 1957.
- [3] Bailey, N.T.J., *The elements of stochastic processes with application to the natural sciences*, New York 1964.
- [4] Bartlett, M. S., *Stochastic population models in ecology and epidemiology*, London 1960.
- [5] Bartoszyński, R., Łoś, J., Wycech-Łoś, M., Contribution to the theory of epidemics, Bernoulli-Bayes-Laplace Anniversary Volume, Berlin-Heidelberg-New York 1965, str. 1-8.
- [6] Bharucha-Reid, A. T., On the stochastic theory of epidemics, *Proceedings of the Third Berkeley Symp. on Math. Stat. and Prob.*, Vol. IV, Berkeley and Los Angeles 1956, str. 111-119.
- [7] Eggenberger, F., Polya, G., Über die Statistik verketteter Vorgänge, *Zeitschrift für angewandte Mathematik und Mechanik* 3 (1923), str. 279-289.
- [8] Feller, W., On the theory of stochastic processes with particular reference to applications, *Proceedings of the Berkeley Symp. on Math. Stat. and Prob.*, Berkeley and Los Angeles 1949, str. 403-432.
- [9] Feller, W., Diffusion processes in genetics, *Proceedings of the Second Berkeley Symp. on Math. Stat. and Prob.*, Berkeley and Los Angeles 1951, str. 227-246.
- [10] Feller, W., *Wstęp do rachunku prawdopodobieństwa*, Warszawa 1966.
- [11] Firescu, D., Tautu, P., A stochastic model of focal epidemic, *Revue Roumaine de Mathématiques Pures et Appliquées* 12 (1967), str. 653-664.
- [12] Fisz, M., *Rachunek prawdopodobieństwa i statystyka matematyczna*, Warszawa 1958.
- [13] Gani, J., An approximate stochastic model for phage reproduction in a bacterium, *Journal of the Australian Mathematical Society* 2 (1962), str. 478-483.

- [14] Gani, J., Stochastic phage attachment to bacteria, *Biometrics* 21 (1965), str. 134-139.
- [15] Green, D., Krieg, D., The delayed origin of mutants induced by exposure of extracellular phage T4 to ethyl methane sulfonate, *Proceedings of the National Academy of Sciences* 47 (1961), str. 64-72 (cytowane za [20]).
- [16] Iosifescu, M., Tautu, P., *Procese stohastice și aplicații în biologie și medicină*, Bukarest 1968 (w jęz. rumuńskim).
- [17] Irwin, J.O., The contributions of G.U. Yule and A.G. Mc Kendrick to stochastic process methods in biology and medicine, *Stochastic models in medicine and biology*, *Proceedings of a symposium at the University of Wisconsin, Madison 1964*, str. 147-164.
- [18] Kendall, D.G., On generalised birth and death processes, *Annals of Mathematical Statistics* 19 (1948), str. 1-15.
- [19] Kendall, D.G., Deterministic and stochastic epidemics in closed populations, *Proceedings of the Third Berkeley Symp. on Math. Stat. and Prob., Vol. IV, Berkeley and Los Angeles 1956*, str. 149-165.
- [20] Kimball, A. W., A model for chemical mutagenesis in bacteriophage, *Biometrics* 21 (1965), str. 875-889.
- [21] Klonecki, W., A method for derivation of probabilities in a stochastic model of population growth for carcinogenesis, *Colloquium Mathematicum* 13 (1965), str. 273-288.
- [22] Klonecki, W., On the partition of the genotypic function, *Zastosowania Matematyki* 9 (1966), str. 21-30.
- [23] Klonecki, W., On the distribution of the number of survivors and deaths in a birth and death process, *Coll. Math.* 16 (1967), str. 269-279.
- [24] Lundberg, O., *On random processes and their application to sickness and accident statistics*, Uppsala 1940.
- [25] Moran, P. A.P., Random processes in genetics, *Proceedings of the Cambridge Philosophical Society* 54 (1958), str. 60-71.
- [26] Moran, P. A.P., The effect of selection in a haploid genetic population, *Proceedings of the Cambridge Philosophical Soc.* 54 (1958), str. 463-467.
- [27] Moran, P.A.P., The distribution of gene frequency in a bisexual diploid population, *Proceedings of the Cambridge Philosophical Soc.* 54 (1958), str. 468-474.

- [28] Neyman, J., Scott, E., A stochastic model of epidemics, Stochastic models in medicine and biology, Proceedings of a symposium at the University of Wisconsin, Madison 1964, str. 45-83.
- [29] Nissen-Mayer, S., Analysis of effects of antibiotics on bacteria by means of stochastic models, Biometrics 22 (1966), str. 761-780.
- [30] Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume IV, Biology and Problems of Health, Berkeley and Los Angeles 1961.
- [31] Sneddon, I., Równania różniczkowe cząstkowe, Warszawa 1962.
- [32] Stepanow, W. W., Równania różniczkowe, Warszawa 1964.
- [33] Stochastic models in medicine and biology, Proceedings of a symposium at the University of Wisconsin, Madison 1964.
- [34] Urbanik, K., On a problem concerning birth and death processes, Acta Mathematica Scientiarum Hungaricae 7 (1956), str. 99-106.
- [35] Urbanik, K., Uwagi o maksymalnej ilości bakterii w populacji, Zast. Mat. 2 (1956), str. 341-348.
- [36] Wright, S., Statistical genetics in relation to evolution, Annals of Eugenics, London 15 (1950), str. 323-354 (cytowane za [24]).
- [37] Yule, G.U., A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, F.R.S., Philosophical Transactions, B 213, str. 21-87 (cytowane za [17]).