

# Rekordy długowieczności i procesy Poissona

Wojciech NIEMIRO\*

## Część I: Jak często umiera najstarszy człowiek na Ziemi?

Jaka jest szansa, że będę kiedyś najstarszym człowiekiem na świecie? W Polsce?

Oczywiście, na tak postawione pytanie matematyka nie udzieli odpowiedzi. Zbyt wiele czynników ma na to wpływ. Są wśród nich czynniki trudne do zmierzenia: w jakich jednostkach wyrazilibyśmy, powiedzmy, swój stan zdrowia lub łaskę bogów (przecież wybrańcy bogów umierają młodo)? Można jednak rozważyć bardzo uproszczony model, w którym postawione na wstępie pytanie nabierze matematycznego sensu. Wyobraźmy sobie świat sprawiedliwy, w którym każdy człowiek w momencie urodzenia ma jednakowe prawdopodobieństwo  $S(t)$  przeżycia ponad  $t$  lat. Nazwiemy  $S$  funkcją przeżycia. Z tego, co powiedzieliśmy, wynika, że  $S(0) = 1$  i  $\lim_{t \rightarrow \infty} S(t) = 0$ . Nie przesadzajmy, czy maksymalny możliwy czas życia jest skończony, czyli czy istnieje takie  $t_{\max}$ , że  $S(t_{\max}) = 0$ . To się okaże nieistotne w naszych rozważaniach. Nasze wyjściowe założenia możemy sformułować w następujący sposób:



**T1.** Każdy noworodek ma jednakową funkcję przeżycia  $S$ .

**T2.** Długości życia różnych noworodków są statystycznie niezależne.

Co znaczy założenie T2? Jeśli rozważymy dwóch osobników, to prawdopodobieństwo tego, że pierwszy przeżyje ponad  $t_1$  lat i drugi  $t_2$  lat, jest równe  $S(t_1)S(t_2)$  dla dowolnej pary liczb nieujemnych  $t_1$  i  $t_2$ . Podobnie prawdopodobieństwa się „przemnażają” dla trzech i więcej osobników.

Potrzebny nam będzie jeszcze opis funkcji przeżycia  $S(t)$  za pomocą tak zwanej gęstości prawdopodobieństwa. Zdarzenie polegające na tym, że długość życia pojedynczego osobnika należy do „krótkiego odcinka czasu”  $(t, t + h]$  ma prawdopodobieństwo  $S(t) - S(t + h)$ . Niech

$$\sigma(t) = \lim_{h \rightarrow 0} \frac{1}{h} (S(t) - S(t + h)).$$

Nieformalnie znaczy to, że  $S(t) - S(t + h) \approx h\sigma(t)$ . Żeby wyrazić funkcję  $S$  poprzez funkcję  $\sigma$ , podzielmy przedział  $(t, \infty)$  na krótkie odcinki długości  $h$ . Mamy

$$S(t) = \sum_{i=0}^{\infty} (S(t + ih) - S(t + (i + 1)h)) \approx \sum_{i=0}^{\infty} h\sigma(t + ih) \approx \int_t^{\infty} \sigma(u) du.$$

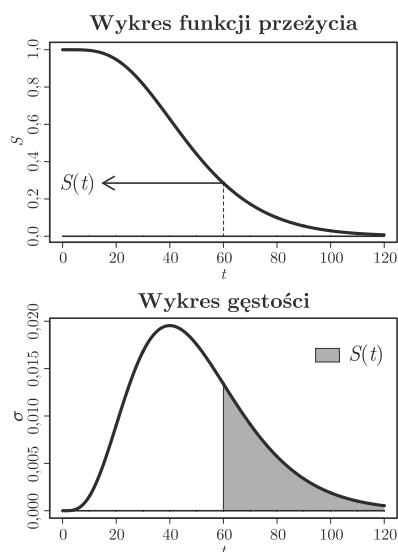
Przechodząc do granicy z  $h \rightarrow 0$ , otrzymamy dokładną równość (patrz rysunek 1).

Druga grupa założeń mówi z grubsza tyle, że dzieci przychodzą na świat całkowicie losowo i z jednakową intensywnością. W języku rachunku prawdopodobieństwa „strumień narodzin” stanowi *jednorodny proces Poissona*. Ten niezwykle ciekawy obiekt matematyczny opisujemy poprzez następujące założenia:

**N1.** Dla dowolnego momentu  $x$  prawdopodobieństwo urodzenia się dziecka w „krótkim” odcinku czasu  $(x, x + h]$  jest w przybliżeniu równe  $\lambda h$ , prawdopodobieństwo zaś urodzenia się więcej niż jednego dziecka jest tak małe, że możemy je zaniedbać.

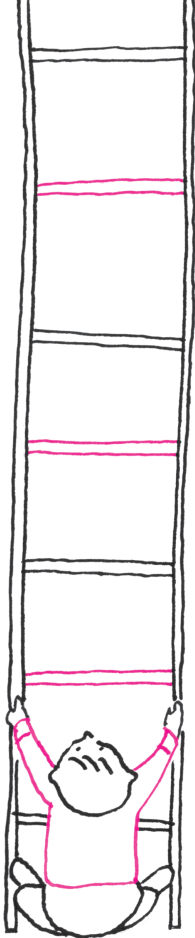
**N2.** Liczby noworodków pojawiających się w rozłącznych odcinkach czasu są statystycznie niezależne.

Zwróćmy uwagę, że mówimy teraz o „czasie kalendarzowym”, liczonym od pewnego umownego momentu 0. Wyjaśnimy te założenia nieco dokładniej.



Rys. 1. Całka oznacza pole obszaru pod wykresem funkcji  $\sigma$ , zaznaczonego na dolnej części rysunku. Górna część rysunku jest wykresem funkcji  $S$ .

\*Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika, Toruń; Instytut Matematyki Stosowanej i Mechaniki, Uniwersytet Warszawski



Niech  $p_n(x)$  oznacza prawdopodobieństwo tego, że w przedziale czasowym długości  $x$  urodzi się dokładnie  $n$  dzieci. Założenie N1 mówi, że

$$\lim_{h \rightarrow 0} \frac{1}{h} p_1(h) = \lambda, \quad \lim_{h \rightarrow 0} \frac{1}{h} \sum_{n=2}^{\infty} p_n(h) = 0.$$

Oczywiście, z tego wynika, że  $\lim_{h \rightarrow 0} p_0(h) = 1$ . Ukryte w warunku N1 jest założenie, że liczba narodzin w określonym odcinku czasu zależy tylko od długości tego odcinka. W naszym szybko zmieniającym się rzeczywistym świecie oczywiście tak nie jest, ale rozważamy uproszczony model stacjonarny.

Zacznijmy od wyprowadzenia wzoru wyrażającego  $p_0(u)$  dla dowolnego  $u > 0$ . Podzielmy odcinek  $(t, t + u]$  na sumę  $n$  krótkich odcinków  $(t + ih, t + (i + 1)h]$  o długości  $h = u/n$ . Dla każdego z tych odcinków prawdopodobieństwo nienarodzenia się dziecka jest w przybliżeniu równe  $1 - \lambda h = 1 - \lambda u/n$ , z założenia N1. Z założenia N2 wynika, że

$$p_0(u) \approx \left(1 - \frac{\lambda u}{n}\right)^n \approx \left(\exp\left(-\frac{\lambda u}{n}\right)\right)^n = \exp(-\lambda u).$$

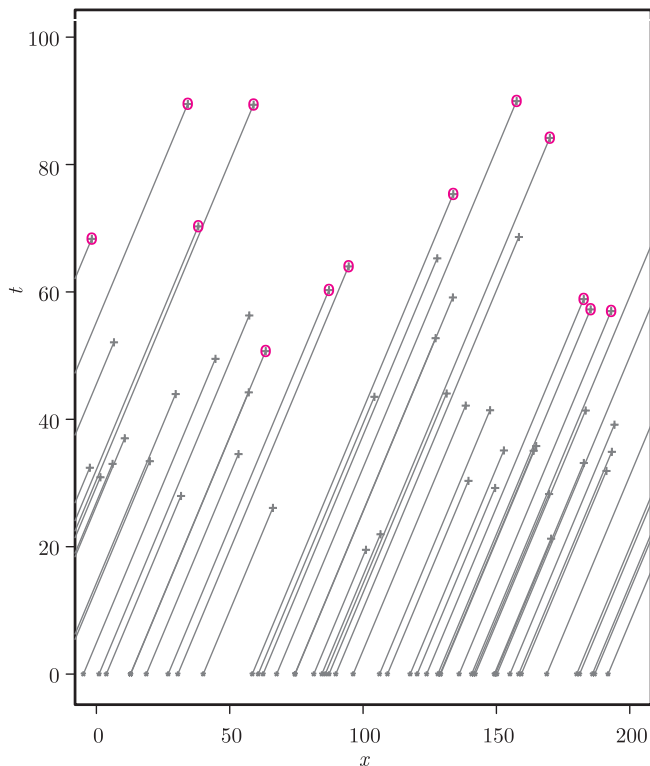
Funkcja „exp” w tym wzorze jest to funkcja wykładnicza,  $\exp(x) = e^x$ , gdzie podstawa potęgi  $e = 2,71828 \dots$  jest tak wybrana, żeby  $\lim_{h \rightarrow 0} (e^h - 1)/h = 1$ . Nieformalnie mówiąc,  $\exp(h) \approx 1 + h$  dla  $h$  „bliskiego zeru”. Tę własność wykorzystaliśmy w wyprowadzeniu wzoru na  $p_0$ . Wykorzystaliśmy także dobrze znaną własność przysługującą każdej funkcji wykładniczej, mianowicie  $e^{u+w} = e^u e^w$ .

**Stwierdzenie 1.** Prawdopodobieństwo tego, że w odcinku czasu  $(t, t + u]$  nie urodzi się ani jedno dziecko, jest dane wzorem:  $p_0(u) = \exp(-\lambda u)$ .

Sam ten wynik nie będzie bezpośrednio używany, ale dalsze rozumowania (w nieco bardziej skomplikowanej sytuacji) będą podobne.

Przyjmijmy jeszcze jedno założenie.

**TN.** Długości życia wszystkich osobników są statystycznie niezależne od procesu narodzin.



Rys. 2. Symulowany przebieg procesu; kolorowe kółka oznaczają rekordy długowieczności.

Spostrzeżenie, które pozwoli nam na rozwiązanie postawionego zadania, jest niezwykle proste.

**Stwierdzenie 2.** Zdarzenie polegające na tym, że w „krótkim” odcinku czasu  $(x, x + h]$  narodzi się osobnik, który przeżyje ponad  $t$  lat, jest w przybliżeniu równe  $\lambda h S(t)$ .

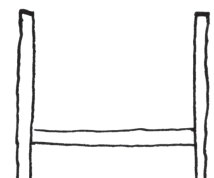
Nasz główny rezultat możemy sformułować w następującej postaci.

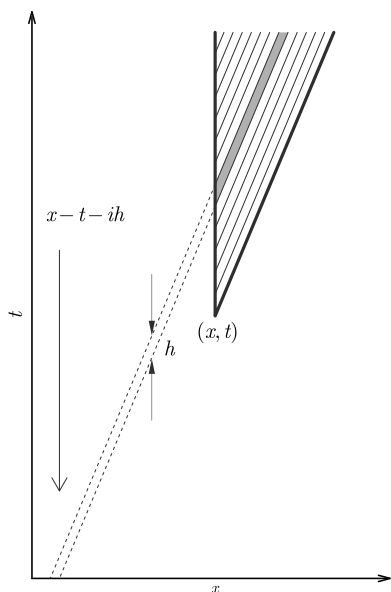
**Twierdzenie.** Prawdopodobieństwo zdarzenia polegającego na tym, że pojedynczy osobnik w momencie śmierci będzie starszy od wszystkich aktualnie żyjących, jest równe

$$\int_0^{\infty} \sigma(t) \exp\left(-\lambda \int_t^{\infty} S(u) du\right) dt.$$

Zanim podamy dowód, zaproponujemy pewną geometryczną interpretację badanego procesu. Rozważmy układ współrzędnych na płaszczyźnie. Oś poziomą oznaczmy literką  $x$  i będziemy na niej zaznaczali czas „kalendarzowy”. Oś pionowa, oznaczona  $t$ , „mierzy” czas życia. Życie osobnika, który urodził się w momencie  $x$  i przeżył  $t$  lat, przedstawimy w postaci odcinka o końcach  $(x, 0)$  i  $(x + t, t)$ .

Na rysunku 2 widzimy przykładową realizację opisywanego przez nas procesu.





Rys. 3. Obszar  $\mathcal{D}$  podzielony na paski.

Punkty na osi  $x$  (czyli momenty narodzin) stanowią, jako się rzekło, jednorodny proces Poissona. Proces na naszym rysunku ma intensywność  $\lambda = 0,2$ , co oznacza, że średnio na jednostkę czasu (powiedzmy, rok) przypada 0,2 narodzin. Współrzędne pionowe punktów stanowią, w języku statystyki matematycznej, próbkę z rozkładu prawdopodobieństwa długości życia, opisanego funkcją  $S(t)$ . W naszym przykładzie jest to funkcja przedstawiona na rysunku 1.

Sformułujemy następujący wynik pomocniczy.

**Lemat.** Jeśli pewien osobnik umrze w wieku  $t$  lat, to prawdopodobieństwo tego, że w momencie śmierci będzie starszy od wszystkich aktualnie żyjących, jest równe

$$\exp\left(-\lambda \int_t^{\infty} S(u) du\right).$$

**Dowód.** Najpierw opiszmy interesujące nas zdarzenie losowe geometrycznie.

Śmierć osobnika, o którym mowa, jest reprezentowana przez punkt  $(x, t)$ .

W chwili  $x$  ten osobnik jest najstarszy ze wszystkich wtedy i tylko wtedy, gdy nie zdarzy się śmierć opisana takim punktem  $(x', t')$ , że  $x' > x$  i  $t' > t + (x' - x)$ .

Innymi słowy, mamy obliczyć prawdopodobieństwo tego, że w obszarze  $\mathcal{D}$ , zaznaczonym na rysunku 3, nie ma „punktów śmierci”.

Podzielmy obszar  $\mathcal{D}$  na „wąskie paski” wysokości  $h$ , tak jak pokazano na rysunku 3. Powiemy, że podstawy pasków są to (pionowe) odcinki pomiędzy punktami  $(x, t), (x, t + h), \dots, (x, t + ih), (x, t + (i + 1)h), \dots$ . Zajmiemy się bliżej „typowym paskiem” numer  $i$  o podstawie  $(x, t + ih), (x, t + (i + 1)h)$ . Zgodnie ze Stwierdzeniem 2 w tym pasku z prawdopodobieństwem bliskim  $\lambda h S(t + ih)$  leży jeden „punkt śmierci”. Prawdopodobieństwo tego, że w tym pasku leżą dwa punkty lub więcej jest tak małe, że możemy je zaniedbać. Stąd wynika, że prawdopodobieństwo braku punktów w pasku rozsądnie przybliża liczba  $1 - \lambda h S(t + ih)$ . Rzecz jasna, brakuje punktów w całym obszarze  $\mathcal{D}$ , gdy w każdym pasku brak punktów. Z Założeń T2, N2 i TN wnioskujemy, że obliczane przez nas prawdopodobieństwo jest iloczynem odpowiednich prawdopodobieństw dla pasków, a więc w przybliżeniu

$$\prod_{i=0}^{\infty} (1 - \lambda h S(t + ih)) \approx \prod_{i=0}^{\infty} \exp(-\lambda h S(t + ih)) = \exp\left(-\lambda \sum_{i=0}^{\infty} h S(t + ih)\right) \approx \exp\left(-\lambda \int_0^{\infty} S(t + u) du\right).$$

To jest wyrażenie, które chcieliśmy otrzymać. Wykorzystaliśmy pewne przybliżone równości, które stają się coraz dokładniejsze, jeśli  $h$  maleje do zera. Przyjmijmy, z przymrużeniem oka, że „wykazaliśmy słuszność lematu”.

W Lemacie pojawiła się funkcja, którą odtąd będziemy oznaczać

$$R(t) = \int_t^{\infty} S(u) du.$$

**Dowód Twierdzenia.** Najtrudniejsze już mamy za sobą. Wystarczy teraz zsumować prawdopodobieństwa zdarzeń polegających na tym, że rozpatrywana osoba umarła w przedziale wieku  $(ih, (i + 1)h]$  i była w momencie śmierci starsza od wszystkich innych. Korzystając z Lematu, otrzymujemy

$$\sum_{i=0}^{\infty} (S(ih) - S((i + 1)h)) \exp(-\lambda R(ih)) \approx \sum_{i=0}^{\infty} h \sigma(ih) \exp(-\lambda R(ih)) \approx \int_0^{\infty} \sigma(t) \exp(-\lambda R(t)) dt.$$

Przejście do granicy z  $h \rightarrow 0$  zmienia przybliżenia w dokładne równości i kończy dowód twierdzenia.

Dygresje i komentarze za miesiąc.



#### Rozwiązanie zadania M 1419.

Odp. Tak!

Dowód przeprowadzimy nie wprost. Ustalmy trzy wektory jednostkowe  $v_1, v_2, v_3$  i założmy, że dla każdego zestawu znaków  $\epsilon_1, \epsilon_2, \epsilon_3$  zachodzi

$$|\epsilon_1 v_1 + \epsilon_2 v_2 + \epsilon_3 v_3|^2 < 3.$$

Ponieważ  $|x|^2 = x \cdot x$ , gdzie  $\cdot$  oznacza iloczyn skalarny, dostajemy

$$\sum_{i,j} \epsilon_i \epsilon_j v_i v_j < 3,$$

co przy założeniu o tym, że wektory  $v_i$  mają długość jeden, daje

$$\sum_{i \neq j} \epsilon_i \epsilon_j v_i v_j < 0,$$

Sumując otrzymane nierówności stronami po wszystkich zestawach znaków  $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3)$ , dostajemy

$$0 > \sum_{\epsilon} \sum_{i \neq j} \epsilon_i \epsilon_j v_i v_j = \sum_{i \neq j} (v_i \cdot v_j) \left( \sum_{\epsilon} \epsilon_i \epsilon_j \right).$$

Ponieważ  $\sum_{\epsilon} \epsilon_i \epsilon_j = 0$  dla  $i \neq j$ , więc otrzymujemy sprzeczność.