

## Stochastyczne procesy narodzin i śmierci

Wiele procesów w fizyce, biologii, ekonomii itp. charakteryzuje się pewnym elementem losowości: na przykład wynikiem zdarzenia kwantowego lub trajektorii cząstki, będącej w ruchu Browna. Są to tak zwane procesy stochastyczne. Zamiast próbować modelować szczegóły dynamiki, często łatwiej jest oddalić się do bardziej grubego (zgrubnego?) modelu, opartego na liczbach losowych.

Przyjrzymy się prostemu modelowi dynamiki populacji, zwanemu procesami urodzenia-śmierci. W tym modelu fluktuacje populacji—w szczególności wskaźniki urodzeń i zgonów, urodzeń i zgonów  $\beta_{\text{birth}}$  i  $\beta_{\text{death}}$ —zależą tylko od wielkości populacji  $l$ . Klasycznym przykładem fizyki jest rozpad jądrowy, w którym szybkość rozpadu zależy tylko od liczby pozostałych jąder.

Rozważmy jednak inną sytuację: syntezę białek. W naszym prostym modelu rozważymy tylko dwa możliwe zdarzenia: narodziny (lub syntezę) i zgony (lub rozpady). Wyobraźmy sobie dalej, że ten system ma tempo syntezy – to znaczy tempo wzrostu populacji, które zależy tylko od obecności składników, a zatem jest stałe  $*$  oraz tempo klirensu (zgony), które jest proporcjonalne do całkowitej populacji  $l$ :

$$\begin{aligned}\beta_{\text{birth}} &= (\text{stała}) \\ \beta_{\text{death}} &= (k_0) * l\end{aligned}$$

$\beta_{\text{birth}}$  i  $k_0$  to parametry modelu, które w zasadzie byłyby określane przez pomiary eksperymentalne. Oba współczynniki  $\beta$  mają jednostki [zdarzenia/czas].

Aby ten model był stochastyczny, niech zarówno czas, jak i kolejność poszczególnych narodzin i zgonów będą probabilistyczne. Odzwierciedla to wiele fizycznych i biologicznych procesów, takich jak rozpad jądrowy, synteza białek i drapieżnictwo, w których efekty kwantowe, termiczne lub społeczne uniemożliwiają nam precyzyjne przewidywanie poszczególnych zdarzeń.

Celem jest zbudowanie modelu obliczeniowego tego prostego procesu urodzeń-śmierci i przeanalizowanie wyników.

Kiedy stanie zrozumiałe, jak działa ten algorytm, można rozszerzyć go na bardziej skomplikowane systemy — na przykład systemy, które zawierają wiele oddziałujących ze sobą populacji, wiele ścieżek rozpadu i/lub różnego rodzaju zdarzenia (np. sytuacje, takie jak przełączanie enzymów, w których tempo parametry są zmieniane.)

\*) Uwaga. Prawdę mówiąc, wskaźniki urodzeń nigdy nie są naprawdę stałe, ale są pewne okoliczności, w których są w przybliżeniu stałe. Na przykład synteza białek, ponieważ „wskaźnik urodzeń” (a raczej „wskaźnik syntezy”) zależy tylko od składników aminokwasów i enzymów – a nie od obecnej populacji białek – to tak długo, jak składniki są utrzymywane przy ustalonej ilości tempo syntezy naszego białka pozostanie stałe.

## Algorytm Gillespiego

Podjęcia obliczeniowe do modelowania systemów stochastycznych zazwyczaj wykorzystują algorytmy Monte-Carlo, które obejmują generowanie liczb losowych i manipulowanie nimi. Aby zmodelować proces od urodzenia do śmierci, zwrócimy się do konkretnego algorytmu Monte-Carlo, zwanego algorytmem Gillespiego, który dzieli procesy stochastyczne na serię kroków, które można modelować numerycznie.

Podstawowy proces wygląda następująco:

1. Obliczyć średnią częstotliwość występowania zdarzeń (dowolnych zdarzeń)
2. Obliczyć czas oczekiwania do następnego zdarzenia (proces losowy)
3. Określić, które zdarzenie faktycznie występuje (drugi losowy proces)
4. Dostosować parametry populacji i czasu na podstawie wyników zdarzenia
5. Powtarzać

Zaczynamy od załadowania dowolnej biblioteki, wymaganej do generowania liczb losowych, kreślenia i manipulowania tablicami; rysowania wykresów; oraz definiujemy parametry modelu:

**beta\_birth = 0.2**

**k0= 0.025**

**N0 = 20.0      # ( początkowa populacja )**

### KROK 1 – LOSOWE CZASY OCZEKIWANIA

Podobnie jak kliknięcia licznika Geigera, zdarzenia w procesach narodzin i śmierci nie występują w regularnych odstępach czasu; zamiast tego możemy tylko powiedzieć, że prawdopodobieństwo wystąpienia zdarzenia w dowolnym małym przedziale czasu  $\Delta t$  jest równe  $\beta \Delta t$ , gdzie  $\beta$  to średnia częstość występowania zdarzeń (a  $1/\beta$  to średni czas między zdarzeniami). Sekwencja losowych czasów oczekiwania jest przykładem tak zwanego „procesu Poissona”.

Chociaż czasy oczekiwania między kolejnymi zdarzeniami zmieniają się losowo, nadal mają rozkład wykładniczy, w którym prawdopodobieństwo wystąpienia następnego zdarzenia po czasie trwania  $t_w$  (mniej-więcej  $\Delta t$ ) jest określone wzorem:

$$P(t_w) = \beta e^{-\beta t_w}$$

Aby komputer generował losowy czas oczekiwania zgodny z tym rozkładem, należy przekonwertować równomiernie rozłożone liczby losowe, generowane przez komputer, na rozkład wykładniczy:

**wait\_time = -(1/beta)\*log(random())**

Napisać funkcję (lub znaleźć odpowiednią wstępnie napisaną funkcję), która zwróci losowy, wykładniczy rozkład czasu oczekiwania dla danej intensywności.

## KROK 2 – OBLICZYĆ ŚREDNIĄ CZĘSTOTLIWOŚĆ ZDARZEŃ

Istnieją dwa możliwe zdarzenia, które mogą wystąpić w procesie od urodzenia do śmierci, a każde z nich może mieć różną częstotliwość:  $\beta_{\text{birth}}$  i  $\beta_{\text{death}}$ .

Określić ogólną zależność, aby obliczyć łączną szybkość występowania zdarzeń. Napisać to w kategoriach parametrów, które zostały już zdefiniowane

(Wskazówka: załóżmy, że porody zdarzają się 2 razy na minutę, a zgony 3 razy na minutę; ile zdarzeń dzieje się na minutę? )

## KROK 3 – WYBIERANIE MIĘDZY WYDARZENIAMI

Ponieważ konkretnym wydarzeniem może być albo narodziny, albo śmierć, musimy być w stanie określić, które wydarzenie faktycznie miało miejsce. Próba Bernoulliego to nazwa procesu, który losowo decyduje między dwoma zdarzeniami o szansach odpowiednio  $\xi_1$  i  $\xi_2 = 1 - \xi_1$ . Klasycznym przykładem jest rzut monetą, dla którego  $\xi_1 = \xi_2 = 1/2$ .

Napisać funkcję, która będzie losowo wybierać między dwoma zdarzeniami z parametrem uczciwości  $\xi$  (wynikiem może być **prawda/falsz**, **1/0** itd.)

## KROK 4 – JAKIE SĄ SZANSE?

Biorąc pod uwagę częstość występowania urodzeń i zgonów ( $\beta_{\text{birth}}$  i  $\beta_{\text{death}}$ ), a także łączną częstość występowania dowolnego zdarzenia  $\beta$ , jakie jest prawdopodobieństwo,  $\xi$ , że dane zdarzenie będzie narodzinami? Czy można użyć tej samej procedury do określenia prawdopodobieństwa, że dane zdarzenie będzie zgonem? Sprawdzamy, czy te szanse są odpowiednio znormalizowane, czyli czy  $\xi_{\text{birth}} + \xi_{\text{death}} = 1$ ?

## KROK 5 – BUDOWANIE DYNAMIKI

Teraz możemy wszystko połączyć. Wewnątrz pętli iteracyjnej dołączamy kod, który:

- Oblicza średnią częstość występowania zdarzeń
- Oblicza czas oczekiwania do następnego wydarzenia
- Określa, które zdarzenie (narodziny lub śmierć) ma miejsce
- Aktualizuje odpowiednio czas i populację
- Rysuje wykres populacji

Uruchamiamy program i krótko opisujemy wyniki. Czy są one losowe?

## KROK 6 – CIĄGŁE ROZWIĄZANIE DETERMINISTYCZNE

Dla rozwiązania analitycznego załóżmy, że populacja  $I$  jest deterministyczna, ciągle zmienna (tzn.

zdarzenia są ciągłe) i bezstopniowo zmienna (tzn. populacja  $I$  nie jest ograniczona do liczb całkowitych).

Napisać równanie różniczkowe, które opisuje tempo zmian  $I$   
(Wskazówka: w jaki sposób mogą się zmieni? Jak możemy to uwzględnić?)

Użyć tego równania różniczkowego, aby określić teoretyczną populację w stanie stacjonarnym.

Rozwiązać równanie różniczkowe, aby znaleźć pełną teoretyczną dynamikę populacji.  
(Wskazówka: jedną z opcji jest przyjrzenie się ogólnemu trendowi krzywej stochastycznej i „zgadnięcie”, jakie powinno być gładkie rozwiązanie analityczne; można następnie przeprowadzić je z powrotem przez równanie różniczkowe i „dostroić” je, aby pasowały. Przy odpowiedniej znajomości z równaniami różniczkowymi, można też od razu „zgadnąć” jego rozwiązanie)

Wykreślić rozwiązanie analityczne obok modelu stochastycznego.

## KROK 7 – ZROZUMIEĆ MODEL

Jak wypada rozwiązanie analityczne w porównaniu z modelem stochastycznym? Jakie są podobieństwa i różnice? Spróbować uruchomić model wielokrotnie bez zmiany jakichkolwiek parametrów; czym różnią się symulacje stochastyczne?

Zmieniamy parametry związane z tym modelem (**beta\_birth**, **k0** i **N0**); jaki to ma wpływ na symulację? Jaki wpływ powinny one mieć na symulację według rozwiązania analitycznego? Czy są one spójne?

Rozwiązanie analityczne dla  $I(t)$  przypomina dynamikę prędkości dla cząstki podlegającej działaniu siły liniowej; dlaczego tak powinno być?

Jakościowo, dlaczego wykładnicza składowa rozwiązania analitycznego jest  $e^{-k_0 t}$ , a nie  $e^{-\beta t}$ ?  
(Innymi słowy, dlaczego parametr  $k_0$  miałby określać szybkość relaksacji do stanu ustalonego, a nie szybkość występowania zdarzeń,  $\beta$ ?)

## KROK 8 – ROZKŁADY HAŁASU I PRAWDOPODOBIENSTWA

Hałas jest nieodłączną częścią systemu stochastycznego; nawet po osiągnięciu stanu ustalonego populacja nadal waha się w zakresie przewidywania analitycznego. Z tego powodu musimy mówić raczej o rozkładzie prawdopodobieństwa, a niż o deterministycznej dynamice populacji. Ale jak wygląda ta dystrybucja? Jak duże są wahania? Czy są zależne od parametrów modelu (tj.  $\beta$  urodzenia lub  $k_0$ )?

Istnieją dwa sposoby na zbadanie szumu stacjonarnego: możemy przeanalizować wyniki naszego modelu lub możemy wymyślić teoretyczną prognozę. Na razie najprościej jest wykorzystać nasz

model do generowania danych, które możemy analizować.

Upewnimy się, że załadowaliśmy odpowiednie biblioteki do kreślenia i analizy statystycznej.

```
import matplotlib.pyplot
```

```
import numpy
```

Aby zbadać stan stacjonarny modelu, zmienimy kod, aby:

- Rozpocząć populację od wartości stanu ustalonego ( $N_0 = N$ ).
- Pamiętać listę wszystkich wartości osiąganych przez populację (po jednej na każdą iterację!)

a następnie napiszemy nowy kod dla:

- Obliczenia średniej, wariancji i odchylenia standardowego tej listy wartości populacji.

- Obliczyć również zupełną liczbę próbek populacji (tj. ile punktów danych posiada rozkład?)

- Narysować histogram rozkładu populacji. Sprawić, aby to wyglądało ładnie.

Chcemy, aby zakres koszy histogramu odpowiadał zakresowi wartości populacji. Aby zwiększyć precyzję miar statystycznych, można wydłużyć czas trwania modelu (i zwiększyć szybkość symulacji, aby nie trwała 10 razy dłużej!), aby zwiększyć liczbę punktów danych (lub uruchomić go wielokrotnie bez nadpisywania danych z poprzednich przebiegów).

Omówimy wyniki. Na przykład:

- Czy nasza próbka dokładnie zgadza się z teoretyczną populacją w stanie stacjonarnym?

Czy powinniśmy się tego spodziewać? Jak różne musiałyby być, żeby zdziwić się?

- Czy próbkowe średnią, wariancja itp. dokładnie pokrywają się z tymi obliczonymi przy ponownych uruchomieniach modelu? Jaka różnica byłaby zaskakująca?

- Otrzymany rozkład prawdopodobieństwa powinien wyglądać dość „normalnie”; dlaczego taki regularny wzór wyłania się z modelu stochastycznego? A dokładniej: jak „wygląda” losowość?

- Porównać, w jaki sposób  $I$  jest używane w modelu stochastycznym/deterministycznym i jak jest używane przy omawianiu rozkładów prawdopodobieństwa (istnieją pewne subtelne, ale ważne różnice.)

## KROK 9 – TEORETYCZNY ROZKŁAD POPULACJI

Gdy populacja osiągnie stan stacjonarny, rozkład wartości populacji zbliża się do tzw. „rozkładu Poissona” (nie mylić z „procesami Poissona”). To znaczy, biorąc pod uwagę rozkład skoncentrowany na  $I_{\text{średnie}} = \mu = \frac{\beta_{\text{birth}}}{k_0}$ . Szansa na to, że populacja faktycznie ma  $I$  wszystkich członków w danym momencie, jest podane przez wzór

$$P_{\text{Poisson}}(I) = \frac{\mu^I}{I!} e^{-\mu}$$

Dla tego konkretnego rozkładu zarówno oczekiwanie (średnia), jak i wariancja (kwadrat odchylenia standardowego) są równe  $\mu$ :

$$E(l) = \mu \qquad \text{Var}(l) = \mu$$

Wykreślić ten rozkład jako funkcję  $l$  i porównać go z histogramem wygenerowanym z modelu. Czy średnia próbek i wariancja są zgodne ze średnią teoretyczną i wariancją? Nie zapominamy, że histogram pokazuje liczbę zliczeń, a nie prawdopodobieństwo, więc trzeba odpowiednio dostosować swoje wykresy. Nie zapominamy też, że  $l$  jest liczbą naturalną (tzn. 0, 1, 2, 3 itd.).

Logistyczne równanie opisuje zachowanie populacji, gdy jak śmierci, tak i narodziny, zależą od rozmiaru populacji, a ilości dostępnych zasobów wystarcza dla populacji co najwyżej rozmiaru  $N$ :

$$\frac{dP}{dt} = k P (N - P), \quad P(0) = N_0$$

(zapis alternatywny  $P'(t) = k P(t) (N - P(t))$ ,  $P(0) = N_0$ )

rozwiązanie

$$P(t) = \frac{N}{\left(\frac{N-N_0}{N_0}\right) e^{-kNt} + 1}$$

Krok 6, analityczne rozwiązanie.

W modelowanym nami przypadku, gdy częstość narodzin **nie** zależy od rozmiaru populacji, a częstość śmierci jest jej proporcjonalna (oraz przy założeniu, że populacja nie jest ograniczona do liczb całkowitych - inaczej nie będzie istnieć pochodna), równanie ma postać:

$$\frac{dP}{dt} = -k_0 P + \beta_{\text{birth}},$$

rozwiązanie równania:  $P[t] = e^{-t k_0} c_1 + \frac{\beta_{\text{birth}}}{k_0}$  z dowolną stałą  $c_1$ .

a rozwiązanie zagadnienia Cauchiego :

$$\frac{dP}{dt} = -k_0 P + \beta_{\text{birth}}, \quad P(0) = N_0$$

(i.e. równania razem z warunkiem początkowym) ma postać :

$$P(t) = \frac{e^{-t k_0} (k_0 N_0 - \beta_{\text{birth}}) + \beta_{\text{birth}}}{k_0}$$

$$\frac{dP}{dt} == -k_0 P + \beta_{\text{birth}}$$

$$\frac{dP}{dt} == -k_{\theta} P \rightarrow$$

$$dP == dt (-k_{\theta} P) \rightarrow$$

$$\frac{dP}{P} == -k_{\theta} dt \rightarrow$$

$$\int \frac{dP}{P} == \int -k_{\theta} dt \rightarrow$$

$$\ln(P) == -k_{\theta} t + C \rightarrow$$

$$P == \underbrace{C(t)}_{\text{stała}} e^{-k_{\theta} t} \dots \underbrace{C(t)}_{\text{stała}} = c_1 + \frac{\beta}{k_{\theta}} e^{tk_{\theta}}$$

$$\frac{dP}{dt} = -k_{\theta} P^2 + \beta_{\text{birth}} P$$

In[215]:=  $N_0 = 20; N_1 = 8; k = 0.05;$

$$f1[t_] = \frac{N_1}{\left(\frac{N_1 - N_0}{N_0}\right) e^{-k N_1 t} + 1}$$

$\beta_{\text{birth}} = 0.2; k_{\theta} = 0.025; N_0 = 20;$

$$f2[t_] = \frac{e^{-tk_{\theta}} (k_{\theta} N_0 - \beta_{\text{birth}} + e^{tk_{\theta}} \beta_{\text{birth}})}{k_{\theta}}$$

Plot[ {f1[t], f2[t]}, {t, 0, 300}, PlotRange -> {All, {0, 20}}]

[wyk...](#) [zakres wykresu](#) [wszystko](#)

Out[216]= 
$$\frac{8}{1 - \frac{3 e^{-0.4 t}}{5}}$$

Out[218]= 
$$40. e^{-0.025 t} (0.3 + 0.2 e^{0.025 t})$$

