Metody Statystyczne

kacper.topolnicki@uj.edu.pl

2 XII 2017

Symulacja procesu Markova

- 2 użytkowników
- 1 komputer
- Do komputera zalogowanych może być:

$x = 0$ (użytkowników)
$x = 1$ (użytkownik)
$x = 2$ (użytkowników)

niezalogowani

- Prawdopodobieństwo logowania

$$P_{logowania} = 0.2$$

- Prawdopodobieństwo pozostania niezalogowanym

$$1 - P_{logowania} = 0.8$$

zalogowani

- Prawdopodobieństwo wylogowania

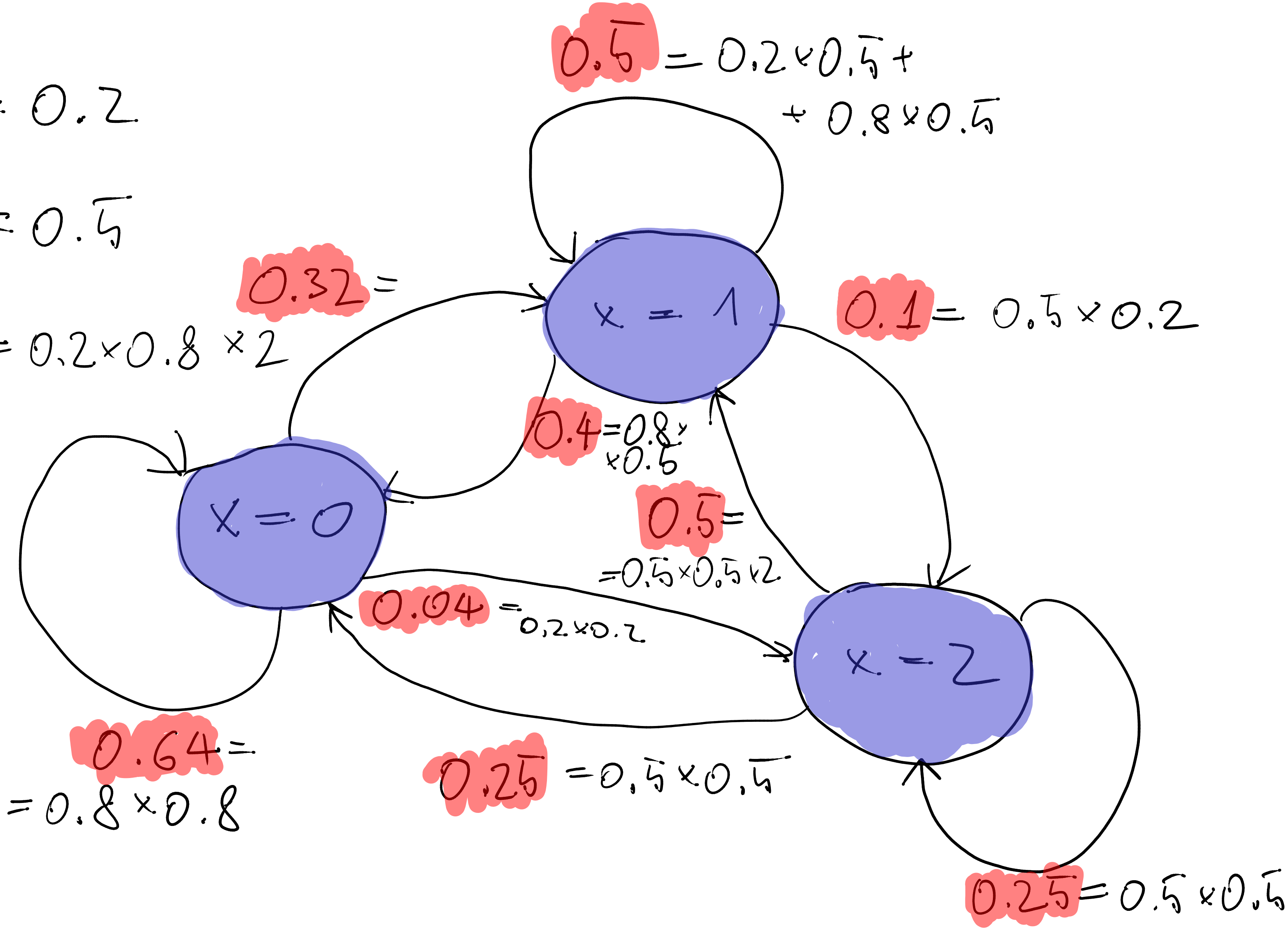
$$P_{wylogowania} = 0.5$$

- Prawdopodobieństwo pozostania zalogowanym

$$1 - P_{wylogowania} = 0.5$$

$P_{logowania} = 0.2$

$P_{wylogowania} = 0.5$



Macierz przejść

$$P = \begin{bmatrix} 0.64 & 0.32 & 0.04 \\ 0.4 & 0.5 & 0.1 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

prawdopodobieństwo ucieczki z $x = 0$

$0.64 + 0.32 + 0.04 = \underline{1}$

$x = 0$

$x = 1$ do $x = 2$

$x = 2$

prawdopodobieństwa przejścia z $x = 1$ do $x = 2$

- Mamy stan w którym układ jest z prawdopodobieństwem

$$p_0 \quad w \quad x=0$$
$$p_1 \quad w \quad x=1$$
$$p_2 \quad w \quad x=2$$

- Jak policzyć stan po iteracji?

kolumna

wiersz (stan)

$$[p_0, p_1, p_2] \begin{bmatrix} 0.64 & 0.32 & 0.04 \\ 0.4 & 0.5 & 0.1 \\ 0.25 & 0.5 & 0.25 \end{bmatrix} = p\,P$$

- Jak policzyć macierz prawdopodobieństw po 2 iteracjach procesu Markova?

$$PP = \begin{bmatrix} 0.64 & 0.32 & 0.04 \\ 0.4 & 0.5 & 0.1 \\ 0.25 & 0.5 & 0.25 \end{bmatrix} \begin{bmatrix} 0.64 & 0.32 & 0.04 \\ 0.4 & 0.5 & 0.1 \\ 0.25 & 0.5 & 0.25 \end{bmatrix} =$$

wiersz × kolumna

$$= \begin{bmatrix} \underbrace{0.64 \times 0.64 + 0.32 \times 0.4 + 0.04 \times 0.25}_{\text{praw. przejścia } x=0 \to x=0 \text{ po 2 iter}} = 0.5476 & 0.3848 & 0.0676 \\ 0.481 & 0.428 & 0.091 \\ 0.4225 & 0.455 & 0.1225 \end{bmatrix}$$

Stan stacjonarny

- Co się stanie po $N \to \infty$ iteracjach?
- Spodziewamy się, że osiągniemy stan stacjonarny?
- Jak w takim przypadku będzie wyglądała macierz

$$[P]^N = \underbrace{[P][P]\dots[P]}_{N}$$

$$[p_0, p_1, p_2]\begin{bmatrix} \pi_0 & \pi_1 & \pi_2 \\ \pi_0 & \pi_1 & \pi_2 \\ \pi_0 & \pi_1 & \pi_2 \end{bmatrix} =$$

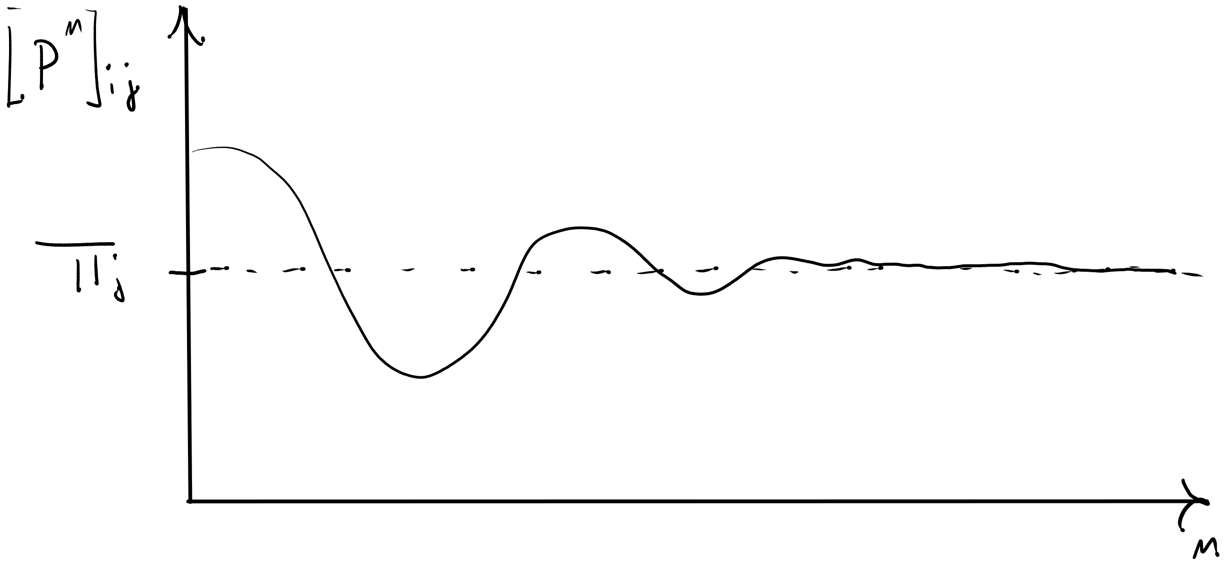
$$N \to \infty \quad [P]^N \to \begin{bmatrix} \pi_0 & \pi_1 & \pi_2 \\ \pi_0 & \pi_1 & \pi_2 \\ \pi_0 & \pi_1 & \pi_2 \end{bmatrix};$$

$$[\pi_0(p_0+p_1+p_2), \pi_1(p_0+p_1+p_2), \pi_2(p_0+p_1+p_2)] =$$

stan stacjonarny $\longrightarrow$ $= [\pi_0, \pi_1, \pi_2]$

A

- Policzyć $[P]^N$ dla dużych N.
- Kryterium zbieżności $\underbrace{\left|P^n - P^{n-1}\right|}_{\text{na przykład } |\max(\ \ )|} < 10^{-5}$
- Można narysować wykres:



B

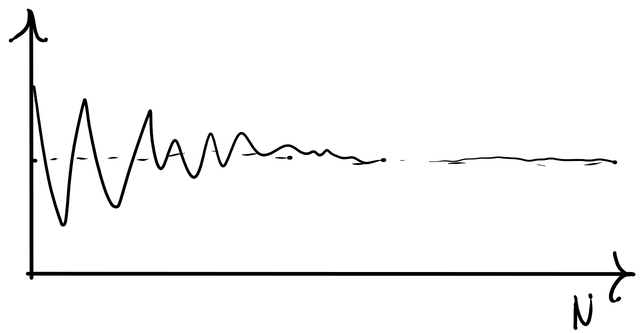
- Start z wybranego węzła $x=0,1,2$
- Losowanie kolejnego węzła zgodnie z P
- Przejście do nowego węzła
-

- Losowanie $\approx N=10^4$ trajektorii
- Policzenie $\pi_i^{exp} = \dfrac{N_i}{N}$ $\quad N_i$ - ile razy odwiedzone $x=i$ $(0,1,2)$
- Porównanie z $[P]^N$, start z $x=0,1,2$, <u>zbieżność</u>

C

- 100 użytkowników
- $x = 0, 1, 2, \ldots, 100$
- $P_{logowanie} = 0.2$, $P_{wylogowanie} = 0.5$ (tak jak poprzednio)
- Trudno jest skonstruować macierz prawdopodobieństw.
- Wykonujemy symulację trajektorii.
- Ile wynosi $\overline{\Pi}_i^{exp}$ dla $i = 0, 1, \ldots, 100$?
- Wykresy zbieżności

$$\overline{\Pi}_i^{exp} = \frac{W_i}{N}$$



D

- Podobna symulacja jak w punkcie C, inne prawdopodobieństwa

dla wylogowanych

$$\overbrace{\left(P_{logowania} = 0.2 \quad P_{pozostania\ niezalogowanym} = 0,8\right)}^{\text{dla wylogowanych}}$$

$$\underbrace{P_{wylogowania} = 1 - (0.008x + 0.1) \quad P_{pozostania\ zalogowanym} = 0.008x + 0.1}_{\text{dla zalogowanych}}$$

dla zalogowanych

Metoda odwrotnej dystrybuanty

Zmiana zmiennych całkowania:

$$y = y(x)$$

stara zmienna losowa

nowa zmienna losowa

Warunek dla FGP:

$$|f_x(x)\,dx| = |f_y(y)\,dy|$$

$$f_y(y) = f_x(x)\left|\frac{dx}{dy}\right|$$

Załóżmy, że $f_X(x)$ to rozkład jednostajny:

$$\begin{cases} f_X(x) = 1 & 0 \leq x \leq 1 \\ f_X(x) = 0 & x < 0 \lor x > 1 \end{cases}$$

można stosunkowo łatwo losować liczby z takiego rozkładu na komputerze

Jak dobrać $y(x)$ tak aby y było losowane z zadanej FGP $f_Y(y)$?

Wystarczy rozwiązać:

$$\frac{dx}{dy} = f_y(y) \quad \left(\text{uwaga} \quad | \quad | \right)$$

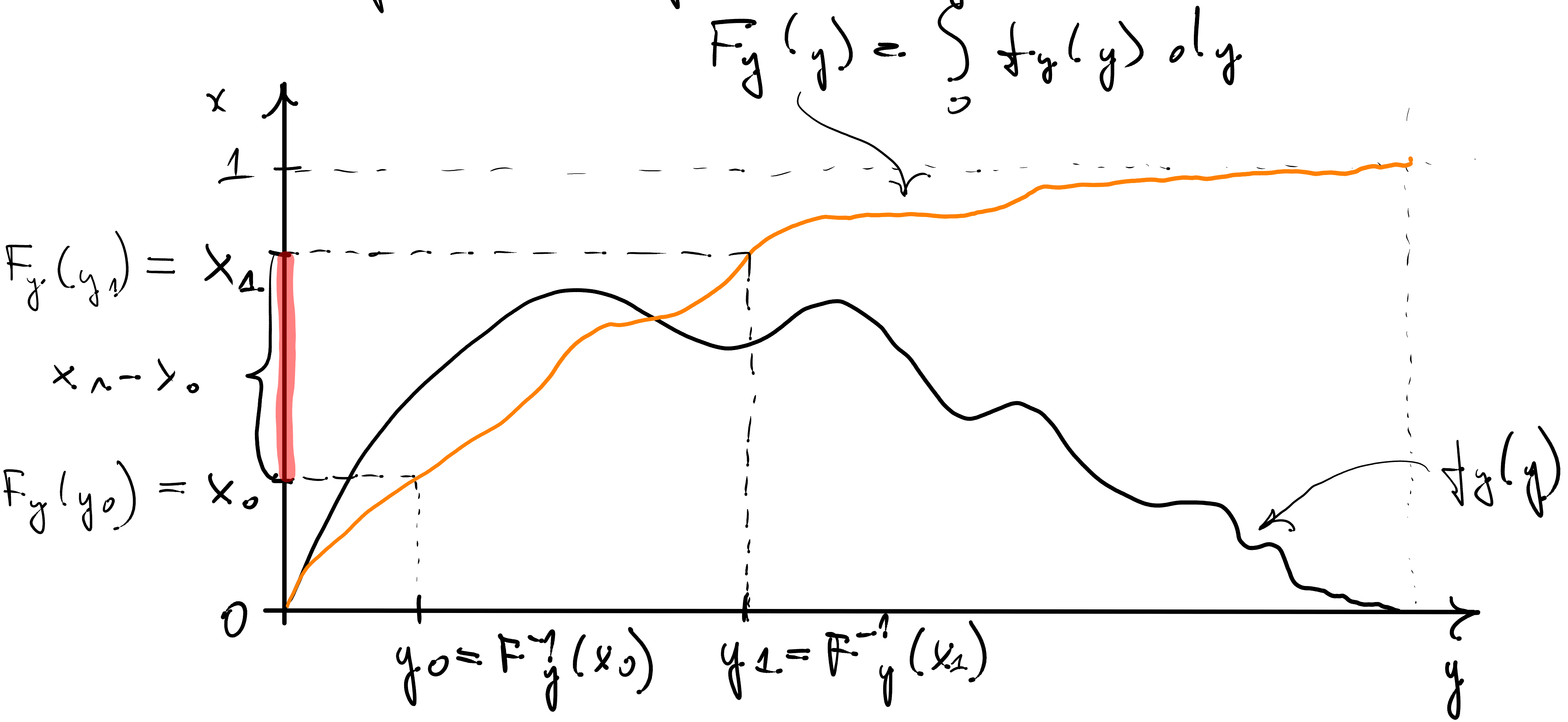
$$x = \int_0^y f_y(v)\,dv = F_y(y)$$

$$y = y(x) = F_y^{-1}(x)$$

znamy FGP

szukamy

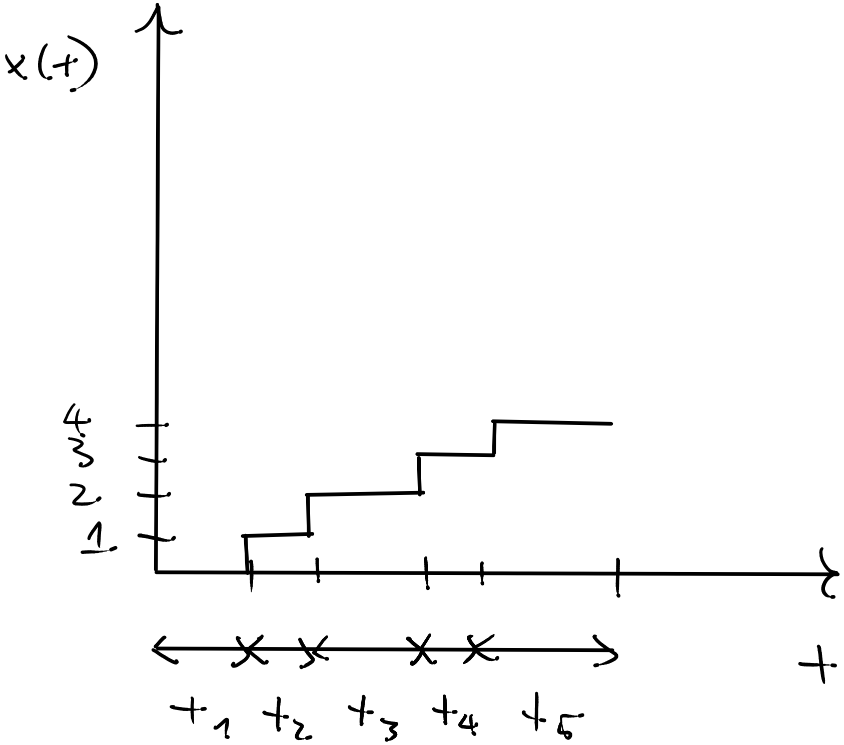
Interpretacja geometryczna:



Proces Poissona

$x(t=0)=0$

skok o 1 co t_i



t_i losowane z rozkładu $f(t)=\lambda e^{-\lambda t}$

$\lambda = 1\,[\mathrm{min}^{-1}]$

metoda odwracania dystrybuanty

$$t_i = \frac{-\ln(u_i)}{\lambda}$$

losowane z rozkładu jednorodnego na przedziale $(0, 1)$

E

- Czy istnieje stan stacjonarny?
- Zebrać $\approx 10^4$ trajektorii, otrzymać rozkład prawdopodobieństwa dla $t = 1, 20, 90$.

- Porównać z rozkładem Poissona $P_t(x=k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$.