Metody statystyczne

Zestaw zadań numer 1

Vitalii Urbanevych

vitalii.urbanevych@doctoral.uj.edu.pl

Literatura

"Numerical recipes"

W.H. Press, S.A. Teukolsky

W.T. Vetterling, B.P. Flannery

Cambrige University Press

Funkcja gęstości prawdowodobieństwa(FGP)

FGP

$f_X(x),\ x \in [a, b]$

X - ciągła zmienna losowa

Prawdopodobieństwo że zmienna x jest w przedziale $x_0 \leq x \leq x_1$:

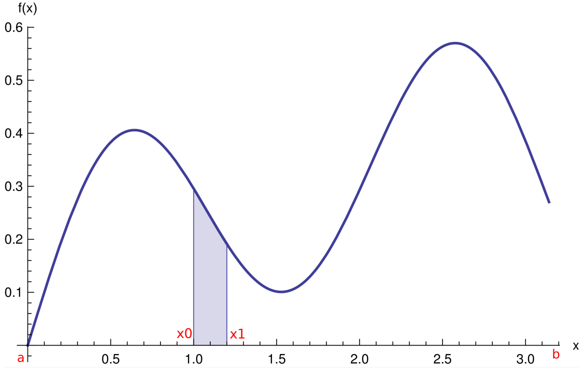
$P(x_0 \leq x \leq x_1) = \int_{x_0}^{x_1} f_X(x)\,\mathrm{d}x$

Normalizacja

$P(a \leq x \leq b) = \int_a^b f_X(x)\,\mathrm{d}x = 1$

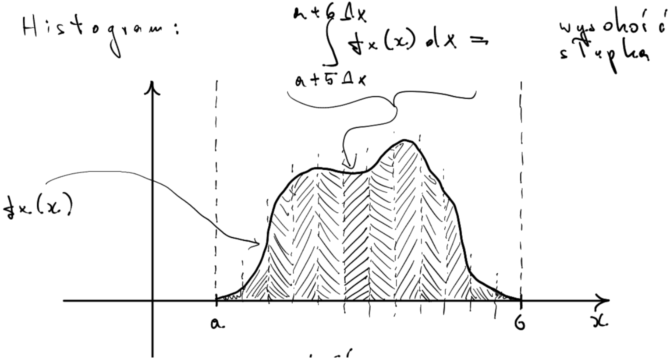
Funkcja gęstości prawdowodobieństwa(FGP)

$P(x = x_0)$ nie ma sensu!



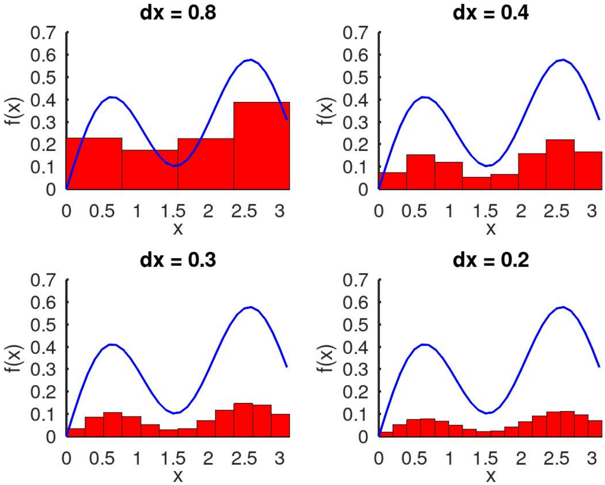
Powierzchnia pod linią:
$P(x_0 \leq x \leq x_1) = \int_{x_0}^{x_1} f_X(x)\,\mathrm{d}x$

Histogram

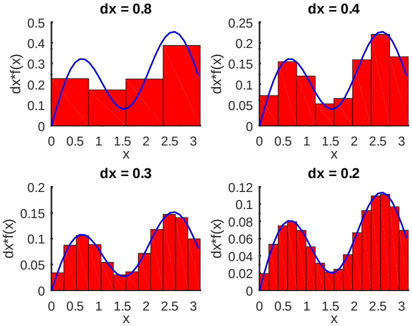


Powierzchnia pod linią = wysokość słupka
Szerokość słupka - Δx

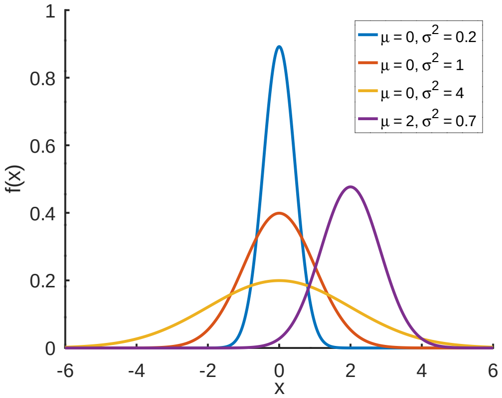
Histogram



Porówniając histogram z FGP należy pamiętać o Δx!



Rozkład normalny $N(\mu, \sigma^2)$



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ - wartość oczekiwana;
σ - odchylenia standardowe
(równoważnie: $var(X) = \sigma^2$)

x_1, x_2 - losowane z rozkładu jednorodnego na przedziale (0,1)

Chcemy zgenerować y_1, y_2 - z rozkładu normalnego

Metoda Boxa-Mullera

x_1, x_2 - losowane z rozkładu jednorodnego na przedziale (0,1)

$$\begin{aligned} y_1 &= \sqrt{-2\ ln(x_1)}cos(2\pi x_2) \\ y_2 &= \sqrt{-2\ ln(x_1)}sin(2\pi x_2) \end{aligned} \quad \Longrightarrow \quad \begin{aligned} x_1 &= exp(-\tfrac{1}{2}(y_1^2 + y_2^2)) \\ x_2 &= \tfrac{1}{2\pi} atan(\tfrac{y_2}{y_1}) \end{aligned}$$

$$f_Y(y_1, y_2)dy_1 dy_2 = f_X(x_1, x_2) \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| dy_1 dy_2,$$

$f_X(x_1, x_2) = 1$ (rozkład jednorodny)

Jakobian

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = (-)\frac{1}{\sqrt{2\pi}} exp\left(-\frac{y_1^2}{2}\right)\frac{1}{\sqrt{2\pi}} exp\left(-\frac{y_2^2}{2}\right)$$

$$f_Y(y_1, y_2)dy_1 dy_2 = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{y_1^2}{2}\right)\frac{1}{\sqrt{2\pi}} exp\left(-\frac{y_2^2}{2}\right) dy_1 dy_2 = f_{Y_1}(y_1) f_{Y_2}(y_2) dy_1 dy_2$$

rozkład normalny dla dwoch niezależnych zmiennych losowych

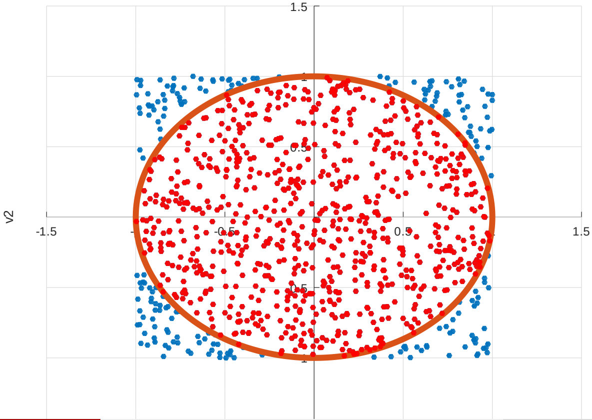
$$f_{Y_1}(y_1) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{y_1^2}{2}\right) = N(0, 1)$$

Algorytm

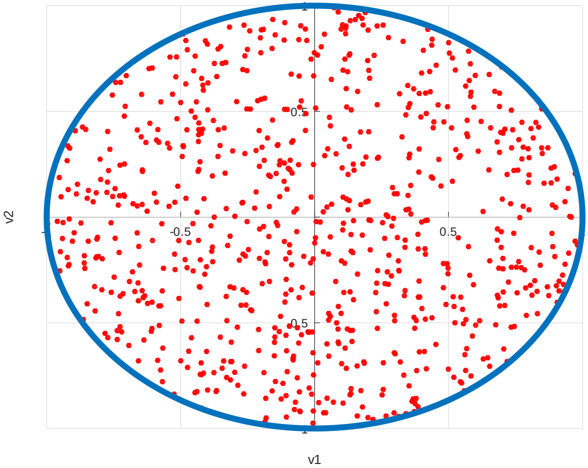
- x_1, x_2 - losowane z rozkładu jednorodnego na przedziale (0,1)
- $y_1 = \sqrt{-2\ ln(x_1)}cos(2\pi x_2)$
  $y_2 = \sqrt{-2\ ln(x_1)}sin(2\pi x_2)$
- $Y_1 \sim N(0,1)$, $Y_2 \sim N(0,1)$
- Jeśli $Y \sim N(0,1)$ i $Z \sim N(\mu, \sigma^2)$, to $z = \sigma y + \mu$

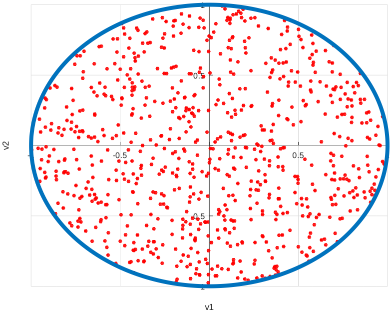
Metoda Polarna

v_1, v_2 z rozkładu jednorodnego (-1,1)
są najpierw losowane w okręgu:



chcemy wziąć tylko tę, które są w środku:





Nowe zmienne losowe:

$$R^2 = v_1^2 + v_2^2$$ rozkład jednorodny (0,1)

$$\theta = arctan\left(\frac{v_1}{v_2}\right)$$ rozkład jednorodny $(0,2\pi)$

Metoda polarna

$$
\begin{aligned}
y_1 &= \sqrt{-2\ ln(x_1)}cos(2\pi x_2) \\
y_2 &= \sqrt{-2\ ln(x_1)}sin(2\pi x_2)
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
y_1 &= \sqrt{-2\ ln(R^2)}\,\overbrace{cos(\theta)}^{v_1/R} \\
y_2 &= \sqrt{-2\ ln(R^2)}\,\underbrace{sin(\theta)}_{v_2/R}
\end{aligned}
\quad \Rightarrow
$$

y_1, y_2 - rozkład normalny N(0, 1)

$$
\boxed{
\begin{aligned}
y_1 &= \sqrt{-2\ ln(R^2)}\frac{v_1}{R} \\
y_2 &= \sqrt{-2\ ln(R^2)}\frac{v_2}{R}
\end{aligned}
}
$$

Metoda odwracania dystrybuanty

Dystrybuanta

$$F_X(x) = P(X \leq x)$$
$$f_X(x) = \frac{d}{dx} F_X(x)$$

Chcemy dobrać taką funkcję (transformację zmiennej losowej) $g(x)$, żeby $y = g(x)$ - nowa zmienna losowa z zadanej $f_Y(y)$

Zakladamy, że X - z rozkładu jednostajnego:

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

$$F_X(x) = x$$

Niechaj

$$X = F_Y(y) \to Y = F_Y^{-1}(X)$$

Wtedy

$$F_X(x) = P(X \leq x) = P(F_Y(y) \leq x) =$$
$$= P(Y \leq F_Y^{-1}(X)) = F_Y(F_Y^{-1}(X)) = X$$

Metoda odwracania dystrybuanty

Czyli X jest z rozkładu jednorodnego.

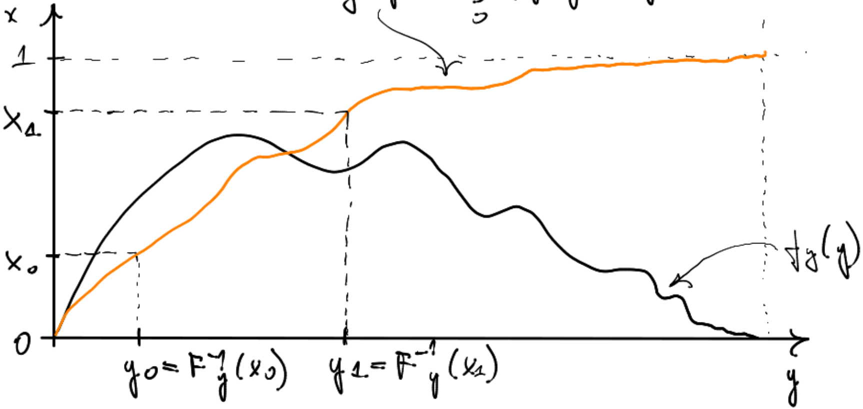
$$F_X(x) = X$$

Znaczy jeśli X jest z rozkładu jednorodnego i Y jest z rozkładu znanej dystrybuanty, to można zrobicz transformację:

$$Y = F_Y^{-1}(X)$$

Interpretacja geometryczna:

$$F_y(y) = \int_0^y f_y(y)\,dy$$



Problem A1

- Implementacja generatoru liczb losowych z rozkładu normalnego $N(\mu, \sigma^2)$ metodą polarną
  $\mu = 0$ - wartość oczekiwania;
  $\sigma^2 = 1$ - wariacja;
- Narysowanie histogramu i porównanie ze wzorem analitycznym

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

- Obliczyć eksperymentalne wartości średniej oraz wariancji

Problem A2

- Implementacja generatoru liczb losowych z rozkładu Cauchy'ego $C(y_0, \gamma)$, metodą odwróconej dystrybuanty
  FGP:
  $$f(y) = \frac{1}{\pi\gamma\left[1 + \left(\frac{y-y_0}{\gamma}\right)^2\right]}, \qquad y \in (-\infty, \infty)$$
- Narysowanie histogramu i porównianie ze wzorem analitycznym dla różnych wartościej y_0 i γ
- Obliczyć eksperymentalne wartości średniej oraz wariancji

Problem ruiny gracza

Gracz A	początkowy kapitał a ($a \in \mathbb{Z}$)
Gracz B	początkowy kapitał b ($b \in \mathbb{Z}$)

$$z = a + b$$

Pod czas każdej rozgrywki gracz A może wygracz jednostkę kapitału u gracza B, lub na odwrót.

A wygrywa 1	prawdopodobieństo p
B wygrywa 1	prawdopodobieństo $q = 1 - p$

Q_i - zdarzenie ruiny A przy kapitale początkowym i

M - zdarzenie wygrana gracza A w pierwszej kolejce

Prawdopodobieństwo całkowite:

$$P(Q_i) = P(Q_i|M)P(M) + P(Q_i|\bar{M})P(\bar{M})$$

$$r_i \equiv P(Q_i)$$

r_i - prawdopodobieństwo ruiny gracza A przy kapitale początkowym i

$$P(Q_i|M) = P(Q_{i+1}) = r_{i+1}$$

$$P(Q_i|\bar{M}) = P(Q_{i-1}) = r_{i-1}$$

$$\boxed{\begin{gathered} r_i = r_{i+1}p + r_{i-1}q \\ r_0 = 1; \qquad r_z = 0 \end{gathered}}$$

$$r_i = r_{i+1}p + r_{i-1}q$$

$$r_i(\overbrace{p+q}^{1}) = r_{i+1}p + r_{i-1}q$$

$$q(r_i - r_{i-1}) = p(r_{i+1} - r_i)$$

$$\frac{q}{p} = \frac{r_{i+1} - r_i}{r_i - r_{i-1}}$$

$$\Delta_i \equiv r_i - r_{i-1}$$

$$\frac{q}{p} = \frac{\Delta_{i+1}}{\Delta_i}$$

$$\Delta_{i+1} = \Delta_i \frac{q}{p}$$

$$\tfrac{q}{p} = const$$

Δ_i - ciąg geometryczny

Ciąg geometryczny

$$\Delta_{i+1} = \Delta_i \frac{q}{p}$$

Suma n czlonków

$$\sum_{i=1}^{n} \Delta_i = S_n = \Delta_1 \frac{1 - \left(\frac{q}{p}\right)^n}{1 - \frac{q}{p}}$$

$$S_n = \underbrace{\Delta_1}_{r_1 - r_0} + \underbrace{\Delta_2}_{r_2 - r_1} + \underbrace{\Delta_3}_{r_3 - r_2} + \ldots + \underbrace{\Delta_n}_{r_n - r_{n-1}} = r_n - r_0 = r_n - 1$$

$$\begin{cases} S_n = \Delta_1 \frac{1-\left(\frac{q}{p}\right)^n}{1-\frac{q}{p}} = r_n - 1 \\ S_z = \Delta_1 \frac{1-\left(\frac{q}{p}\right)^z}{1-\frac{q}{p}} = r_z - 1 = -1 \end{cases}$$

$$-(r_n - 1) = \frac{1-\left(\frac{q}{p}\right)^n}{1-\left(\frac{q}{p}\right)^z}$$

$\mathbf{p \neq q \neq 0.5}$

$$\boxed{r_n = \frac{\left(\frac{q}{p}\right)^n - \left(\frac{q}{p}\right)^z}{1-\left(\frac{q}{p}\right)^z} \qquad (1)}$$

$$p = q = 0.5 \qquad \rightarrow \qquad \frac{q}{p} = 1$$

$$r_n = \frac{\left(\frac{q}{p}\right)^n - \left(\frac{q}{p}\right)^z}{1 - \left(\frac{q}{p}\right)^z} = \frac{1^n - 1^z}{1 - 1^z} = \frac{0}{0}$$

Reguła de l'Hospitala

Jeżeli $\lim\limits_{x \to a} f(x) \to 0$ i $\lim\limits_{x \to a} g(x) \to 0$ oraz istnieją (skończone) pochodne $f'(a)$ i $g'(a)$, przy czym $g'(a) \neq 0$, to

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$$

$$x \equiv \tfrac{q}{p} \to 1$$

$$r_n = \lim_{x\to 1} \frac{x^n - x^z}{1 - x^z} = \lim_{x\to 1} \frac{(x^n - x^z)'}{(1 - x^z)'} = \lim_{x\to 1} \frac{nx^{n-1} - zx^{z-1}}{-zx^{z-1}} =$$
$$\lim_{x\to 1} \frac{x^{z-1}\left(nx^{n-z} - z\right)}{-zx^{z-1}} = \lim_{x\to 1} \frac{nx^{n-z} - z}{-z} = = \frac{n-z}{-z} = 1 - \frac{n}{z}$$

p=q=0.5

$$\boxed{r_n = 1 - \frac{n}{z} \qquad (2)}$$

Jakie prawdopodobieństwo, że gra się zakończy (ktoś przegra)?

$$r_a^A = 1 - \frac{a}{z}$$

prawdopodobieństwo, że gracz A przegra

$$r_b^B \to \left[r_b^A : \begin{matrix} q \to p \\ p \to q \end{matrix} \right] = 1 - \frac{b}{z}$$

p=q=0.5

$$r_a^A + r_b^B = \left(1 - \frac{a}{a+b}\right) + \left(1 - \frac{b}{a+b}\right) = \frac{b+a}{a+b} = 1$$

Jakie prawdopodobieństwo, że gra się zakończy (ktoś przegra)?

p≠q≠0.5

$$r_a^A = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^z}{1 - \left(\frac{q}{p}\right)^z}$$

$$r_b^B = \frac{\left(\frac{p}{q}\right)^b - \left(\frac{p}{q}\right)^z}{1 - \left(\frac{p}{q}\right)^z} = \frac{\left(\frac{q}{p}\right)^{-b} - \left(\frac{q}{p}\right)^{-z}}{1 - \left(\frac{q}{p}\right)^{-z}} = \frac{\left(\frac{q}{p}\right)^{z-b} - 1}{\left(\frac{q}{p}\right)^z - 1} = \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^z}$$

$$r_a^A + r_b^B = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^z}{1 - \left(\frac{q}{p}\right)^z} + \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^z} = 1$$

Gra się zawsze zakończy

Problem:

A ma ∞ kapital, **B** ma kapital **b**, $P(A_{\text{wygrywa}})$ - ?

p=q=0.5

$$P(A_{\text{wygrywa}}) = P(B_{\text{bankrutuje}}) = \lim_{a\to\infty} r_b^B = \lim_{a\to\infty}\left(1 - \frac{b}{a+b}\right) =$$
$$= \lim_{a\to\infty}\frac{a}{a+b} = \lim_{a\to\infty}\frac{1}{1+b/a} = 1$$

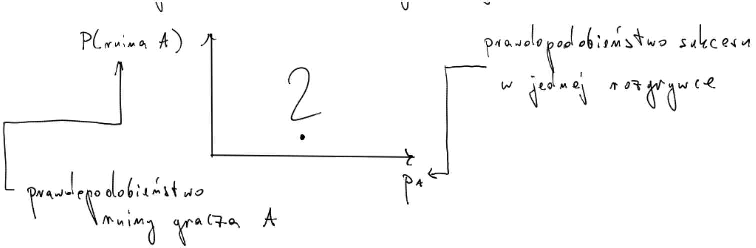
p≠q≠0.5

$$P(A_{\text{wygrywa}}) = \lim_{a\to\infty} r_b^B = \lim_{a\to\infty}\frac{1-\left(\frac{q}{p}\right)^a}{1-\left(\frac{q}{p}\right)^{a+b}} = \begin{cases} 1; & q < p \\ \left(\frac{p}{q}\right)^b; & q > p \end{cases}$$

Problem B

Symulacja N gier z różnymi wartościami p_A. Dla każdej wartości p_A obliczyć prawdopodobieństwo ruiny gracza A.

- Ruina gracza dla 2 graczy A,B



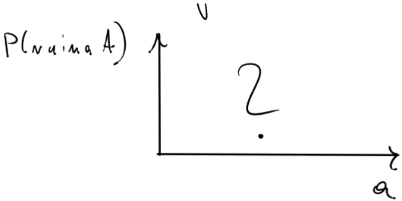
  Kapitały początkowe A,B:
  a = 50; b=50

- Porównanie z wynikiem analitycznym dla różnych a,b

Problem C

Symulacja N gier z różnymi wartościami a. Dla każdej wartości obliczyć prawdopodobieństwo ruiny gracza A.

- Ruina gracza dla 2 graczy A,B



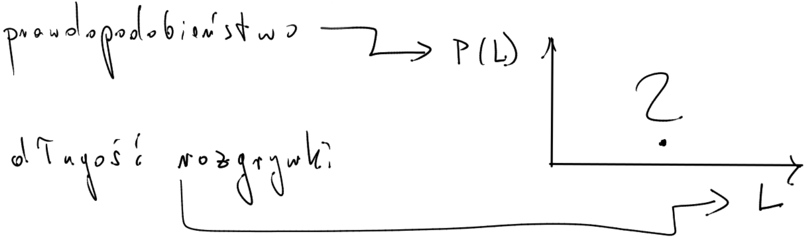
a+b=100;
$p_A = \frac{1}{2}, \dots$

- Porównanie wyniku z teorią

Problem D

Symulacja N gier, dla każdej gry obliczyć iłość rozgrywek.

- Liczba rozgrywek do ukończenia gry - **L**



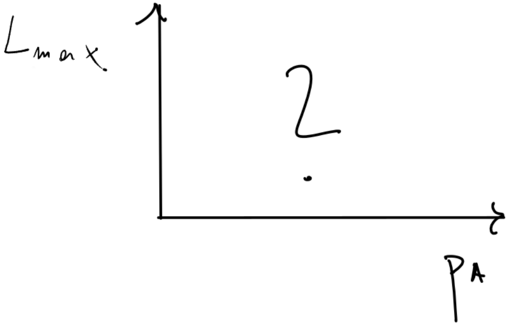
$$p_A = \frac{1}{2}, \frac{1}{5}, \frac{4}{5}; \qquad a = b = 50; \qquad \text{całkowita liczba gier} = 20000$$

- Wyliczyć średnią długość rozgrywki

Problem E

Symulacja N gier z różnymi wartościami p_A. Dla każdej wartości p_A obliczyć maksymalną iłość rozgrywek.

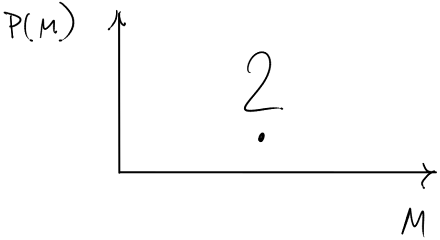
- Histogram maksymalną długości rozgrywek przy 1000 rozgrywkach - L_{max}



p_A - prawdopodobieństwo wygrania kolejki przez A

Problem F

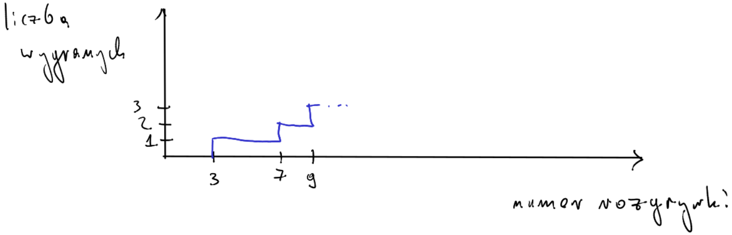
Prawdopodobieństwo że gracz **A** ma kapitał **M** po **n** rozgrywkach



$$n = 2,\ 10, 20, \ldots 100;$$
$$a=b=50;$$
$$p_A = \frac{1}{2}, \frac{1}{5}$$

Problem G

- Trajektoria liczby wygranych dla 1 z 2 graczy



  dla kilku gier(do 10)
  dla różnych wartościej p_A: $\frac{1}{2}, \frac{1}{5}, \frac{4}{5}$
- Trajektoria kapitału dla 1 z 2 graczy

Problem H

- B, C, D, G dla kilku, np. pięciu, graczy
(w G teraz trajektorie dla wszystkich graczy)

różne kombinacje p_i(prawdopodobieństwo wygrania gracza "i" w kolejce)
$a_i = 20$ - kapitały początkowe graczej (lub spróbować inne wartości).