## 1 Wprowadzenie

Często stoimy przed koniecznością generacji liczb losowych o zadanym rozkładzie prawdopodobieństwa. Jeśli dysponujemy liczbami losowymi o rozkładzie równomiernym, to problem sprowadza się do przekształcenia rozkładu równomiernego w rozkład pożądany. Ponieważ w generatory liczb o rozkładzie równomiernym wyposażone są ogólnie dostępne kompilatory popularnych języków programowania np. C++, jako punkt wyjścia przyjmujemy zatem, że dysponujemy takowym.

Jako przykład takiego generatora liczb na odcinku $(0,1)$ może służyć procedura rekurencyjna:

$$x_{n+1} = (ax_n + b)(\operatorname{mod} c),$$

w której $a$, $b$, $c$ są odpowiednio ustalonymi liczbami całkowitymi, natomiast $x_1$ jest liczbą początkową, przy czym $\alpha(\operatorname{mod}\beta)$ jest resztą z dzielenia $\alpha$ przez $\beta$. Jest oczywiste, że kolejne liczby mają w istocie charakter deterministyczny i jedynie symulują losowość. Z tego też względu nazywane są pseudolosowymi. Generator taki uznawany jest jako dobry, gdy pomyślnie przejdzie testy dotyczące rozkładu, niezależności itp.

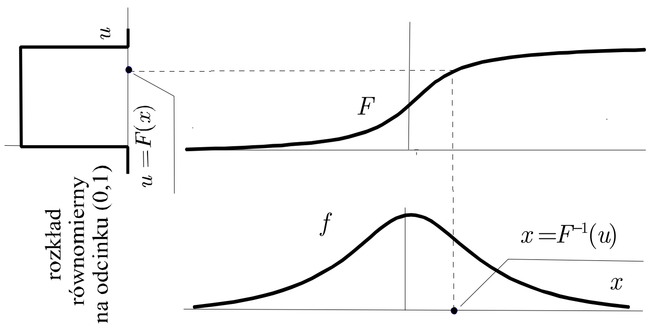
Przedstawimy metody, przy pomocy których, wychodząc ze zmiennych losowych o rozkładzie równomiernym, można otrzymywać dowolne rozkłady. Sporo uwagi poświęcimy rozkładowi normalnemu. Dla każdej przedstawionej metody podamy uzasadnienie teoretyczne, a następnie konkretne generatory podamy w czytelnej postaci algorytmicznej.

Pisząc o rozkładzie równomiernym, mamy zawsze na myśli rozkład równomierny na odcinku $(0,1)$.

## 2 Metoda inwersyjna

### 2.1 Omówienie metody

W metodzie inwersyjnej pożądany rozkład o dystrybuancie $F$ otrzymuje się przez odpowiednie nieliniowe przekształcenie zmiennej losowej o rozkładzie równomiernym. Przekształceniem tym jest $F^{-1}$, patrz rys. 2.1. Uzasadnieniem metody jest poniższe twierdzenie:




Rys. 2.1: Ilustracja metody inwersyjnej.

**Twierdzenie 2.1** *Jeśli zmienna losowa $U$ ma rozkład równomierny na odcinku $(0,1)$, to $F^{-1}(U)$ ma dystrybuantę $F$. I na odwrót. Jeżeli $X$ ma dystrybuantę $F$, to $F(X)$ ma rozkład równomierny.*

**Dowód.** Zauważmy, że

$$P\{F^{-1}(U) < x\} = P\{U < F(x)\} = F(x),$$

skąd wynika pierwsza część twierdzenia. Druga jest konsekwencją tego, że

$$\begin{aligned} P\{F(X) < u\} &= P\{X < F^{-1}(u)\} \\ &= F(F^{-1}(u)) = u \end{aligned}$$

dla każdego $u \in (0,1)$. □

Twierdzenie 2.1 pozwala skonstruować następujący algorytm generacji zmiennej losowej o dystrybuancie $F$:

| Zmienna losowa o dystrybuancie $F$<br>Metoda inwersyjna |
|---|
| ► Generuj $U$ o rozkładzie równomiernym na odcinku $(0,1)$ |
| ► RETURN $X \leftarrow F^{-1}(U)$ |

### 2.2 Wybrane rozkłady

Zasadniczą sprawą dla metody jest to czy można, a jeżeli tak, to czy można łatwo wykonać operację $F^{-1}$. Nadaje się ona zatem do generacji zmiennych losowych, których dystrybuanta znana jest w postaci jawnej, i którą można łatwo odwrócić. Są to rozkłady takie jak np.: wykładniczy, Laplace'a, Cauchy'ego. Nie można tej metody wykorzystać jednak do generacji rozkładu normalnego, gdyż jego dystrybuanta nie ma jawnej postaci.

Dla niektórych ważniejszych rozkładów odwrotności $F^{-1}$ podane są w tabeli 1. Dla pewnych rozkładów, wzory określające $F^{-1}$ podano także w prostszych postaciach alternatywnych. Wynikają one ze spostrzeżenia, że $V = 1-U$ oraz $U$ mają jednakowe rozkłady. Dzięki tabeli tej możemy podać generatory rozkładu wykładniczego, Cauchy'ego, Laplace'a, Pareto oraz logistycznego.

Tabela 1: Gęstości, dystrybuanty i ich odwrotności

| Rozkład | trójkątny<br>$0 \le x \le 1$ | wykładniczy<br>$0 \le x < \infty$ |
|---|---|---|
| $f(x)$ | $2(1-x)$ | $e^{-x}$ |
| $F(x)$ | $1-(x-1)^2$ | $1-e^{-x}$ |
| $F^{-1}(u)$ | $1-\sqrt{1-u}$ | $-\ln(1-u)$ |
| $F^{-1}(v)\big\vert_{v=1-u}$ | $1-\sqrt{v}$ | $-\ln v$ |

| Rozkład | Pareto<br>$0 < b \le x$ | logistyczny |
|---|---|---|
| $f(x)$ | $\dfrac{ab^a}{x^{a+1}}$ | $\dfrac{1}{2+e^x+e^{-x}}$ |
| $F(x)$ | $1-\left(\dfrac{b}{x}\right)^a$ | $\dfrac{1}{1+e^{-x}}$ |
| $F^{-1}(u)$ | $\dfrac{b}{(1-u)^{1/a}}$ | $\ln\dfrac{u}{1-u}$ |
| $F^{-1}(v)\big\vert_{v=1-u}$ | $\dfrac{b}{v^{1/a}}$ | |

| Rozkład | Cauchy'ego |
|---|---|
| $f(x)$ | $\dfrac{1}{\pi(x^2+1)}$ |
| $F(x)$ | $\dfrac{1}{2}+\dfrac{1}{\pi}\arctan x$ |
| $F^{-1}(u)$ | $\tan\left[\pi\left(u-\dfrac{1}{2}\right)\right]$ |

### 2.2.1 Rozkład trókątny

| Rozkład trójkątny<br>Metoda inwersyjna |
|---|
| ▶ Generuj $U$ o rozkładzie równomiernym |
| ▶ RETURN $X \leftarrow 1-\sqrt{U}$ |

### 2.2.2 Rozkład wykładniczy

| Rozkład wykładniczy<br>Metoda inwersyjna |
|---|
| ▶ Generuj $U$ o rozkładzie równomiernym |
| ▶ RETURN $X \leftarrow -\ln U$ |

### 2.2.3 Rozkład Laplace'a

| Rozkład Laplace'a<br>Metoda inwersyjna |
|---|
| ▶ Generuj $U$ o rozkładzie równomiernym |
| ▶ $X = -\ln U$ |
| *$X$ ma rozkład wykładniczy* |
| ▶ Generuj $V$ o rozkładzie równomiernym |
| ▶ IF $V < 1/2$ THEN $X \leftarrow -X$ |

### 2.2.4 Rozkład Cauchy'ego

| Rozkład Cauchy'ego<br>Metoda inwersyjna |
|---|
| ▶ Generuj $U$ o rozkładzie równomiernym |
| ▶ RETURN $X \leftarrow \tan\left(\pi\left(U-\dfrac{1}{2}\right)\right)$ |

### 2.2.5 Rozkład Pareto

| Rozkład Pareto $(a,b)$<br>Metoda inwersyjna |
|---|
| ▶ Generuj $U$ o rozkładzie równomiernym |
| ▶ RETURN $X \leftarrow \dfrac{b}{U^{1/a}}$ |

### 2.2.6 Rozkład logistyczny

| Rozkład logistyczny<br>Metoda inwersyjna |
|---|
| ▶ Generuj $U$ o rozkładzie równomiernym |
| ▶ RETURN $X \leftarrow \ln\dfrac{U}{1-U}$ |

# 3 Metoda odrzucania

## 3.1 Omówienie metody

Istota metody odrzucania polega na tym, że generuje się realizacje zmiennej losowej o pewnym, łatwym do uzyskania rozkładzie, a następnie – według zasady wynikającej z Twierdzenia 3.1 – odrzuca się niektóre z nich. Pozostałe posiadają pożądany rozkład.

Szczególną uwagę zwrócimy na rozkład normalny. Otrzymać go można wychodząc np. z rozkładów Laplace'a, wykładniczego i Cauchy'ego. Wygenerowanie zmiennych losowych o tych rozkładach nie jest trudne, jeśli korzysta się z metody inwersyjnej.

Podstawą metody odrzucania jest poniższe twierdzenie:

**Twierdzenie 3.1** *Dla funkcji $f$, niech*

$$A = \{(x,y) : -\infty < x < \infty, 0 \le y \le cf(x)\},$$

*gdzie $c > 0$. Jeśli $f$ jest gęstością prawdopodobieństwa zmienej losowej $X$ i $U$ ma rozkład równomierny na odcinku $(0,1)$, to para $(X,Y)$, gdzie $Y = cUf(X)$, ma rozkład równomierny na zbiorze $A$. Na odwrót, jeśli $(X,Y)$ ma rozkład równomierny na zbiorze $A$, to $X$ ma gęstość $f$.*

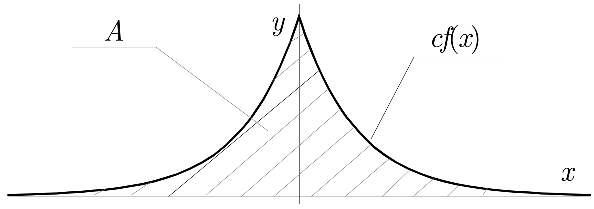
**Dowód.** Ilustracją dowodu jest rys 3.1. Załóżmy, że $X$ ma gęstość $f(x)$, natomiast $U$ ma rozkład równomierny na odcinku $[0,1]$. Zatem

$$f(y|x) = \begin{cases} \dfrac{1}{cf(x)} & \text{dla } (x,y) \in A \\ 0, & \text{dla } (x,y) \notin A, \end{cases}$$

skąd wynika, że

$$f(x,y) = f(y|x)f(x) = \begin{cases} \dfrac{1}{c}, & \text{dla } (x,y) \in A \\ 0, & \text{dla } (x,y) \notin A \end{cases}$$

co oznacza, że para $(X,Y)$ ma rozkład równomierny na zbiorze $A$.




Rys. 3.1: Para $(X,Y)$, gdzie $Y = cUf(X)$, ma rozkład równomierny na zakreskowanym zbiorze $A$.

Teraz zakładamy że, para $(X,Y)$ ma rozkład równomierny na zbiorze $A$, tzn. że

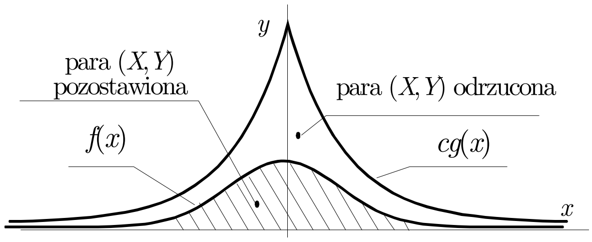
$$f(x,y) = \begin{cases} \dfrac{1}{c}, & \text{dla } (x,y) \in A \\ 0, & \text{dla } (x,y) \notin A, \end{cases}$$

bowiem $\int\int f(x,y)dxdy = c$. Zatem gęstością zmiennej $X$ jest

$$\int f(x,y)dy = \int_0^{cf(x)} \frac{1}{c}dy = f(x),$$

co kończy dowód. □

Idea algorytmu jest następująca. Gęstość $f$ jest oczywiście zadana. Załóżmy, że istnieje łatwa do generacji gęstość $g$ oraz stała $c$ takie, że $f(x) \le cg(x)$ dla wszystkich $x \in (-\infty,\infty)$. Schemat postępowania jest następujący, patrz rys. 3.2:




Rys. 3.2: Ilustracja generacji rozkładu $f$ przy pomocy odrzucania z rozkładu $g$.

- ► Najpierw generuje się zmienną losową o rozkładzie równomiernym na zbiorze zawartym pomiędzy osią $x$ a krzywą $y = cg(x)$. Korzystając z Twierdzenia 3.1 można zrobić to zgodnie z poniższym algorytmem:

  - ▷ Generuj zmienną losową $X$ o gęstości $g$
  - ▷ Generuj zmienną losową $U$ o rozkładzie równomiernym na odcinku $(0, 1)$
  - ▷ $Y \leftarrow cUg(X)$
  - ▷ RETURN $(X, Y)$

- ► Nastęnie odrzuca się wszystkie pary $(X, Y)$, które leżą ponad krzywą $y = f(x)$. Pozostałe pary, zgodnie z podanym twierdzeniem, mają rozkład równomierny na zbiorze zawartym pomiędzy osią $x$ i krzywą $y = f(x)$. Zatem, zgodnie z Twierdzeniem 3.1, $X$ w pozostałych parach ma pożądaną gęstość $f$. Schemat postępowania na tym etapie jest jak poniżej:

  - ▷ IF $Y \le f(X)$ THEN $X \leftarrow X$

    ELSE powtórz wszystkie operacje od początku
  - ▷ RETURN $X$

Łącząc to w całość i dokonując oczywistych uproszczeń otrzymujemy gotową procedurę.

| Zmienna losowa o gęstości $f$<br>Metoda odrzucania |
|---|
| ♦ REPEAT |
| ♢ Generuj $X$ o gęstości $g$ |
| ♢ Generuj $U$ o rozkładzie równomienym |
| ♦ UNTIL $Uc\dfrac{g(X)}{f(X)} \le 1$ |
| ♦ RETURN $X$ |

Aby metoda była efektywna musi posiadać trzy podstawowe cechy:

1) musi istnieć gęstość $g$ i stała $c$ takie, że $f(x) \le cg(x)$ dla wszystkich $x \in (-\infty, \infty)$,

2) rozkład o gęstości $g$ powinien być łatwy do generacji,

3) liczba iteracji procedury powinna być możliwie mała, tzn. prawdopodobieństwo przyjęcia pary $(X, cUg(X))$ powinno być duże.

Odnośnie do punktu 3), oznaczając przez $p$ prawdopodobieństwo przyjęcia wylosowanej pary $(X, Y)$, zauważamy, że

$$p = P\{f(X) \ge cUg(X)\} = \frac{1}{c}.$$

Pożądane jest zatem, aby $c$ było możliwie małe. Oznaczając przez $N$ liczbę iteracji algorytmu, która zapewnia przyjęcie pary, zauważmy ponadto, że $P\{N = i\} = p(1 - p)^{i-1}$. Zatem $N$ ma rozkład geomeryczny i $E\{N\} = 1/p = c$. Zatem wyprowadzony przed chwilą wniosek odnoszący się do $c$ potwierdza się.

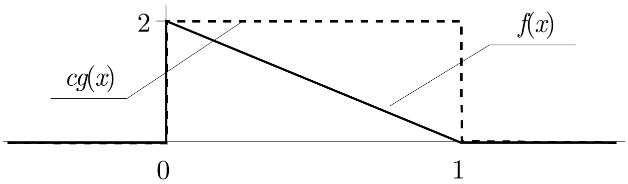
Jest więc oczywiste, że stała $c$ powinna być mała, gdyż zarówno prawdopodobieństwo odrzucenia wylosowanej pary jak i średnia liczba iteracji algorytmu są wtedy małe.

### 3.2 Rozkład trójkątny

Zauważmy, że dla gęstości $f(x)$ rozkładu trójkątnego i gęstości $g(x)$ rozkładu równomiernego, $f(x) \le cg(x)$, gdzie $c = 2$, bowiem

$$\frac{g(x)}{f(x)} = \frac{1}{2(1-x)} \ge \frac{1}{2},$$

dla $0 \le x \le 1$, patrz rys. 3.3.




Rys. 3.3: Generacja rozkładu trójkątnego $f(x)$ (linia ciągła) przez odrzucanie z rozkładu równomiernego (linia przerywana) $g(x)$; $c = 2$.

Pozwala to zbudować algorytm generacji rozkładu trójkątnego jak poniżej:

| Rozkład trójkątny<br>Odrzucanie z rozkładu równomiernego |
|---|
| ♦ REPEAT |
| ♢ Generuj $X$ o rozkładzie równomiernym |
| ♢ Generuj $U$ o rozkładzie równomienym |
| ♦ UNTIL $U \le 1 - X$ |
| ♦ RETURN $X$ |

Nie wykonuje on operacji pierwiastkowania, patrz § 2.2.1, ale odrzuca połowę wygenerowanych liczb.

### 3.3 Rozkład logistyczny

Gęstość rozkładu logistycznego

$$f(x) = \frac{1}{2 + e^x + e^{-x}}$$

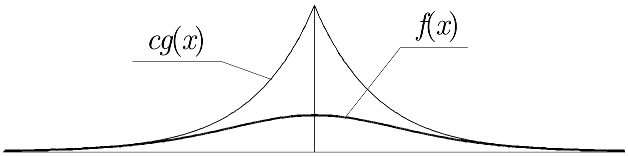
spełnia nierówność $f(x) \le cg(x)$, gdzie $c = 2$ oraz

$$g(x) = \frac{1}{2}e^{-|x|}$$

jest gęstością rozkładu Laplace'a. Zauważmy bowiem, że

$$\begin{aligned}\frac{g(x)}{f(x)} &= \frac{1}{2}(2 + e^{|x|} + e^{-|x|})e^{-|x|} \\ &= \frac{1}{2}(1 + e^{|x|})^2 \geq \frac{1}{2},\end{aligned}$$

patrz rys. 3.4. Krzywa $cg(x)$ leży ponad $f(x)$ na całej prostej $x \in (-\infty, \infty)$.




Rys. 3.4: Generacja rozkładu logistycznego $f(x)$ z rozkładu Laplace'a $g(x)$, $c = 2$.

Wynika stąd podany poniżej algorytm, który zapewnia prawdopodobieństwo przyjęcia na poziomie $1/c = 1/2$:

| Rozkład logistyczny<br>Odrzucanie z rozkładu Laplace'a |
|---|
| ♦ REPEAT |
| ◇ Generuj $X$ o rozkładzie Laplace'a |
| ◇ Generuj $U$ o rozkładzie równomienym |
| ♦ UNTIL $U\left(1 + e^{-\|X\|}\right)^2 \leq 1$ |
| ♦ RETURN $X$ |

**Ćwiczenie 3.1** *Zauważając, że* $f(x) \leq 1/\left(4 + x^2\right)$*, zbudować generator rozkładu logistycznego, w którym* $c = \pi/2 \approx 1,57$.

## 3.4 Rozkład normalny

Podamy teraz generatory rozkładu normalnego $N(0,1)$, czyli o gęstości

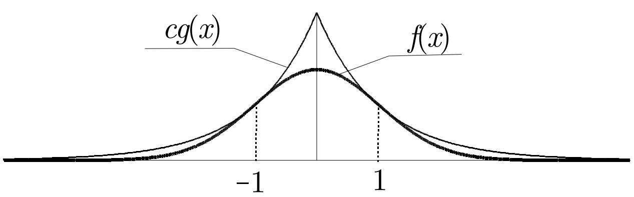
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

### 3.4.1 Odrzucane z rozkładu Laplace'a

Jeśli odrzuca się z rozkładu Laplace'a o gęstości $g(x) = (1/2)e^{-|x|}$, to $c = \sqrt{2e/\pi} \approx 1,315$, bowiem

$$\frac{g(x)}{f(x)} = \sqrt{\frac{\pi}{2e}} e^{(|x|-1)^2/2} \geq \sqrt{\frac{\pi}{2e}}.$$

Dla wyznaczonego $c$, $cg(x) = f(x)$ dla $|x| = 1$. Wzajemne położenie funkcji $f(x)$ i $cg(x)$ pokazuje rys. 3.5. Są one styczne w punktach $x = \pm 1$. Prawdopodobieństwo przyjęcia $X$ jest równe $1/c = 1/\sqrt{2e/\pi} \approx 0,760$.




Rys. 3.5: Generacja rozkładu normalnego $f(x)$ przez odrzucanie z rozkładu Laplace'a $g(x)$, $c = \sqrt{2e/\pi}$.

| Rozkład normalny<br>Odrzucanie z rozkładu Laplace'a |
|---|
| ■ REPEAT |
| □ Generuj $X$ o Laplace'a |
| □ Generuj $U$ o rozkładzie równomiernym |
| ■ UNTIL $Ue^{(\|x\|-1)^2/2} \leq 1$ |
| ■ RETURN $X$ |

### 3.4.2 Odrzucanie z rozkładu Cauchy'ego

W podanym teraz generatorze odrzuca się z rozkładu Cauchy'ego o gęstości

$$g(x) = \frac{1}{\pi(x^2+1)}.$$

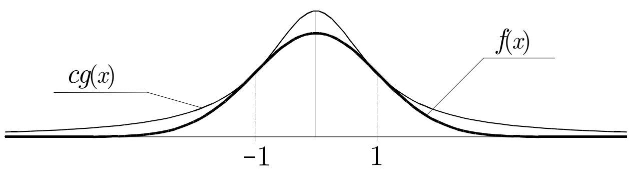
Jak można sprawdzić

$$\begin{aligned}\frac{g(x)}{f(x)} &= \frac{\sqrt{2\pi}e^{x^2/2}}{\pi(x^2+1)} = \sqrt{\frac{2}{\pi}}\frac{e^{x^2/2}}{(x^2+1)} \\ &\geq \sqrt{\frac{e}{2\pi}},\end{aligned}$$

ponieważ

$$\frac{e^{x^2/2}}{x^2+1} \geq \left.\frac{e^{x^2/2}}{x^2+1}\right|_{x=\pm 1} = \frac{\sqrt{e}}{2}.$$

Zatem $c = \sqrt{2\pi/e} \approx 1,520$. Krzywe $f(x)$ i $cg(x)$ pokazane




Rys. 3.6: Generacja rozkładu normalnego $f(x)$ przez odrzucanie z rozkładu Cauchy'ego $g(x)$, $c - \sqrt{2\pi/e}$ .

na rys. 3.6 są styczne w punktach $x = \pm 1$. W procedurze przedstawionej poniżej prawdopodobieństwo przyjęcia $X$ jest równe $1/c = 1/\sqrt{2\pi/e} \approx 0.657$:

| Rozkład normalny<br>Odrzucanie z rozkładu Cauchy'ego |
|---|
| ▼ REPEAT |
| ▽ Generuj $X$ o rozkładzie Cauchy'ego |
| ▽ Generuj $U$ o rozkładzie równomiernym |
| ▽ $S = X^2$ |
| ▼ UNTIL $U \leq \frac{\sqrt{e}}{2}(1+S)e^{\frac{1}{2}S}$ |