Liczby losowe

Wykorzystywana literatura:
- Wykłady: https://kacpertopol.github.io/
- Załączone zdjęcia/notatki (data dostępu 05.12.2024):

* https://henryk-dabrowski.pl/index.php?title=Liczby_losowe_%E2%80%93_metoda_odwracania_dystrybuanty
* https://pl.wikipedia.org/wiki/Rozk%C5%82ad_jednostajny_ci%C4%85g%C5%82y
* https://www.if.pw.edu.pl/~agatka/numeryczne/wyklad_12.pdf
* własne notatki

Gęstość prawdopodobieństwa i dystrybuanta

Rozważmy funkcję $f(x)$ określoną na $\mathbb{R}$, nieujemną i całkowalną. Powiemy, że $f(x)$ jest gęstością rozkładu prawdopodobieństwa P, jeżeli dla dowolnego zbioru $A \subset \mathbb{R}$ jest

$$P(A) = \int_A f(x)dx$$

gdzie $P(A)$ jest prawdopodobieństwem przypisanym zbiorowi A.

Z powyższej definicji wynika natychmiast, że funkcja $f(x)$ musi być unormowana:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Dystrybuantą gęstości prawdopodobieństwa $f(x)$ nazywamy funkcję:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Prawdopodobieństwo, że zmienna losowa przyjmie wartość należącą do przedziału $[a,b]$ wynosi:

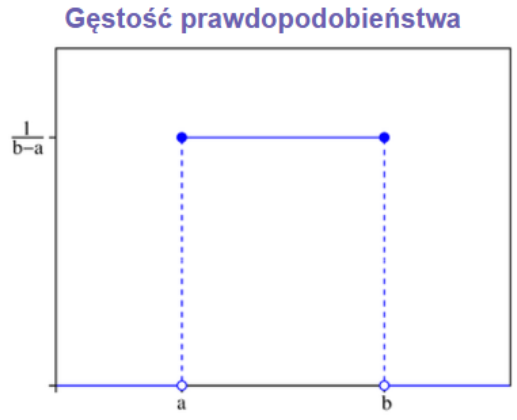
$$P(a \leqslant x \leqslant b) = \int_a^b f(t)dt = F(b) - F(a)$$

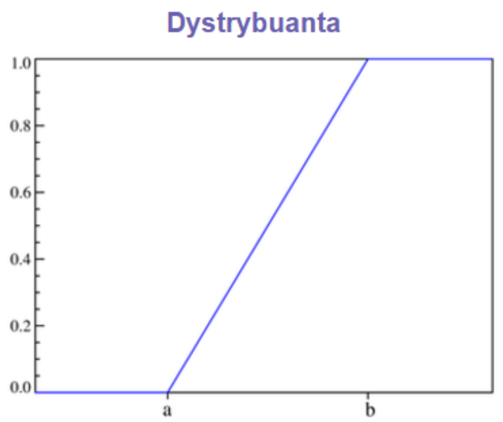
Przykład:

Rozkład jednostajny (jednorodny, równomierny, prostokątny albo płaski)

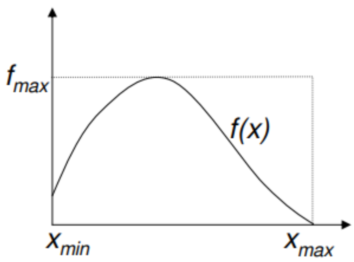
$$f(x) = \begin{cases} 0 & \text{dla} \quad x < a \\ \frac{1}{b-a} & \text{dla} \quad x \in [a,b] \\ 0 & \text{dla} \quad x > b \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{dla} \quad x < a \\ \frac{x-a}{b-a} & \text{dla} \quad x \in [a,b] \\ 1 & \text{dla} \quad x > b \end{cases}$$





Generacja liczb pseudolosowych z zadanego rozkładu: metoda hit-and-miss / metoda akceptacji-odrzucenia / metoda eliminacji



Chcemy wygenerować liczby losowe x w przedziału $[x_{min}, x_{max}]$ zgodnie z rozkładem prawdopodobieństwa $f(x)$.

Niech f_{max} – maksymalna wartość $f(x)$ na przedziale $[x_{min}, x_{max}]$

1. Losujemy liczbę losową x z rozkładu jednorodnego

$$x = x_{\min} + (x_{\max} - x_{\min})U$$

2. Losujemy inną liczbę losową r z rozkładu jednorodnego (między 0 i 1)

$$r = U$$

3. Jeśli

$$r < \frac{f(x)}{f_{\max}}$$

akceptujemy liczbę losową x.

W przeciwnym wypadku, powracamy do kroku 1.

Generacja liczb pseudolosowych z zadanego rozkładu: metoda odwracania

dystrybuanty

Dystrybuanta odwrotna:

Dystrybuanta odwrotna $F^{-1}(u)$ przekształca zmienną losową U(0,1) o rozkładzie równomiernym w zmienną losową X o rozkładzie f(x), któremu odpowiada dystrybuanta F(x):

$X = F^{-1}(U)$

Przykład:

Rozkład jednostajny (jednorodny, równomierny, prostokątny albo płaski)

$F^{-1}(U) = a + (b - a)\,u$

Generacja liczb pseudolosowych:

1. Wyznaczenie dystrybuanty
2. Wyznaczenie dystrybuanty odwrotnej
3. Losowanie liczby pseudolosowej z rozkładu jednorodnego (0,1)
4. Przeliczenie wylosowanej liczby na liczbę z szukanego rozkładu poprzez wstawienie jej do funkcji dystrybuanty odwrotnej
5. Powrót do kroku 3

Problemy:

Na ogół wymagane jest aby dystrybuanta była znana i odwracalna analitycznie → stosunkowo niewielka liczba funkcji!

Generacja liczb pseudolosowych z rozkładu normalnego

Dystrybuanta nieodwracalna - potrzebne inne metody:

* Hit-and-miss: powyżej

* Transformacja Boxa-Mullera:

1. Losowanie dwóch liczb pseudolosowych x_1, x_2 z rozkładu jednorodnego (0,1)
2. Przeliczenie

$$y_1 = \sqrt{-2\ln(x_1)}\,\cos(2\pi x_2)$$

$$y_2 = \sqrt{-2\ln(x_1)}\,\sin(2\pi x_2)$$

gdzie y_1 i y_2 będą należały do standardowego rozkładu normalnego N(0,1)

* Metoda polarna:

1. Losowanie dwóch liczb pseudolosowych x_1, x_2 z rozkładu jednorodnego (0,1)
2. Sprawdzenie warunku $x_1^2 + x_2^2 \leq 1$
3. Jeśli warunek spełniony przejście do punktu 4, jeśli nie to powrót do punktu 1
4. Przeliczenie

$$R^2 = x_1^2 + x_2^2$$

$$y_1 = \sqrt{-2\ln(R^2)}\,x_1/R$$

$$y_2 = \sqrt{-2\ln(R^2)}\,x_2/R$$

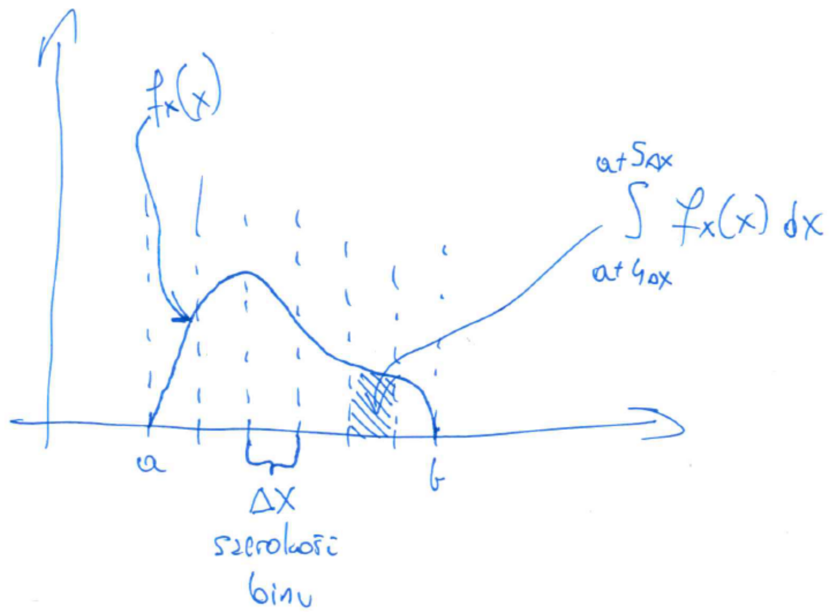
gdzie y_1 i y_2 będą należały do standardowego rozkładu normalnego N(0,1)

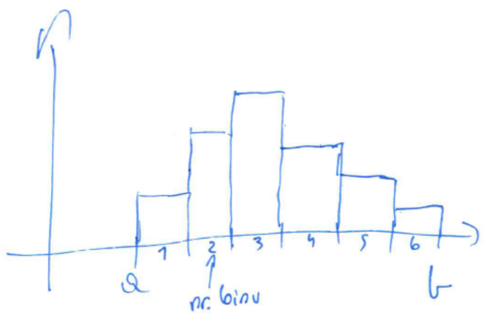
Generacja liczb pseudolosowych z zadanego rozkładu: hybrydowa metoda

eliminacji

W standardowej metodzie hit-and-miss niejednokrotnie odrzucamy wylosowaną liczbę co zmniejsza efektywność algorytmu. W przypadku, gdy pole szukanego rozkładu jest dużo mniejsze od pola “pudełka”, w którym losujemy punkt spadek efektywności może być znaczący.
W hybrydowej metodzie eliminacji zastępujemy losowanie w “pudełku” losowaniem liczby z innego rozkładu g(x), większego od f(x). Tym sposobem, zwiększamy procentowy udział pola f(x), a zatem poprawiamy wydajność. Rozkład g(x) musimy zaś wybrać w taki sposób, aby generacja liczby pseudolosowej z niego mogła być wykonana metodą odwracania dystrybuanty.
Przykład w zadaniach.

Tworzenie histogramów





h_i – wysokość binu "i"

1° $h_i = n_i$ – liczba zliczeń w binie "i"

2° $h_i = \frac{n_i}{N}$, gdzie $N = \sum_i n_i$ względna częstotliwość

3° $h_i = \frac{n_i}{N \cdot w}$, gdzie w – szerokość binu gęstość doświadczalna