Elementy rachunku prawdopodobieństwa

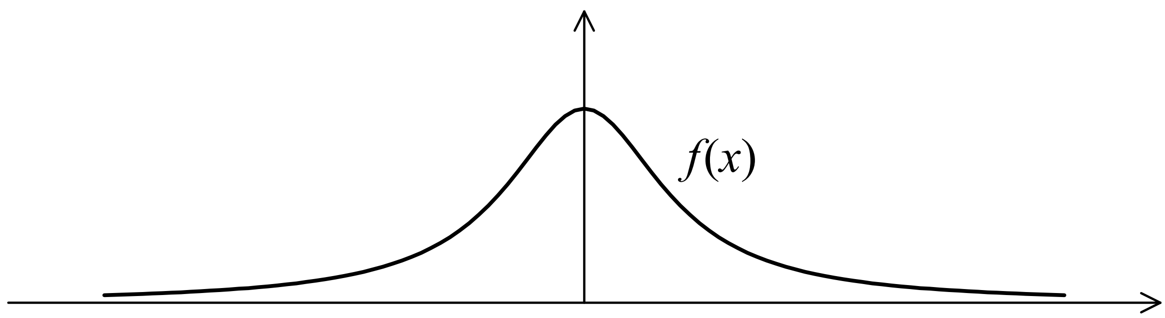
$f(x)$ - gęstość rozkładu prawdopodobieństwa $\quad X \to \begin{cases} f(x) \geq 0 \\ \int_{-\infty}^{\infty} f(x)dx = 1 \end{cases}$

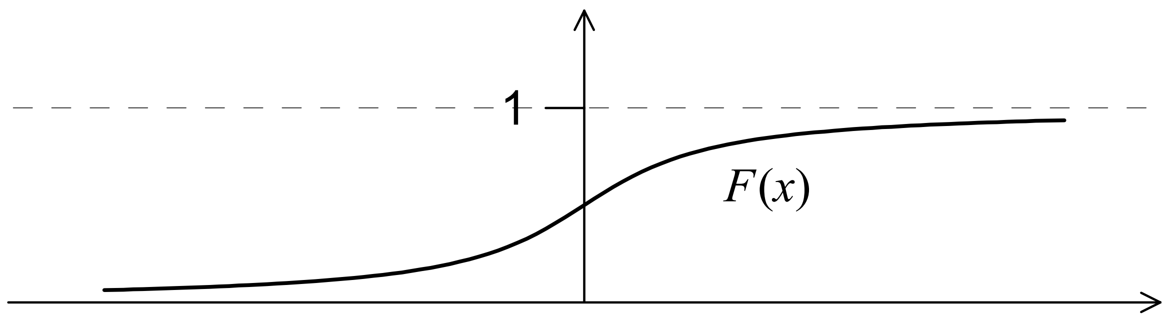
$$P(a < x < b) = \int_a^b f(x)dx$$

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

$E(X) = \int_{-\infty}^{\infty} tf(t)dt$ - wartość średnia

$D^2(X) = E(X^2) - E(X)^2$ - wariancja

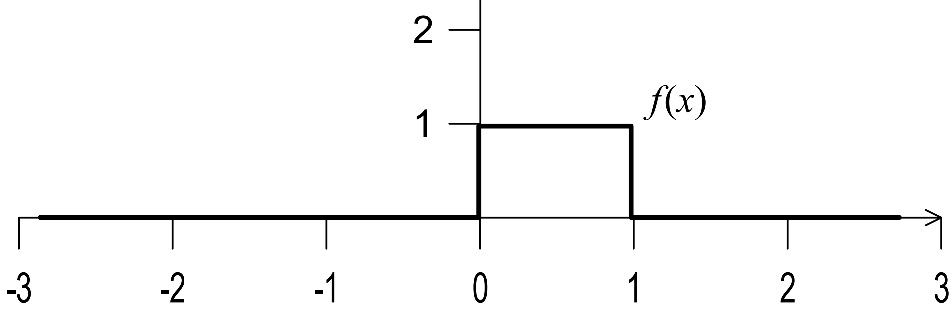




Rozkład równomierny
w przedziale $(0,1)$

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

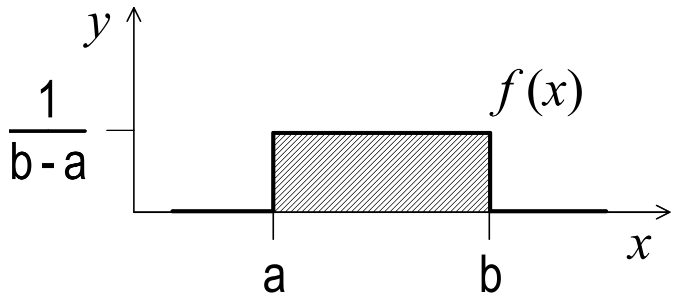
$$E(X) = \frac{1}{2}; \quad D^2(X) = \frac{1}{12}$$



Rozkład równomierny na przedziale [a,b]

$$x_i = \alpha_i (b-a) + a$$

bo: $$\int_a^{x_i} \frac{1}{b-a} dx = \alpha_i$$



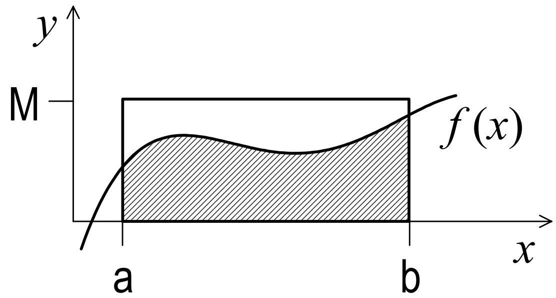
Obliczanie całek metodą Monte Carlo

metoda próbkowania prostego

$$\int_a^b f(x)dx = \frac{n}{N} M(b-a);$$

$0 \le f(x) \le M$

dla $x \in (a,b)$



$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} \ldots \int_{a_K}^{b_K} f(\mathbf{X}) d\mathbf{X} = \frac{n}{N} V_K$$

Przykładowy algorytm

```
randomize;
N:=100000;
nc:=0;
for i:=1 to N do
begin
   xi:=random*(b-a) + a;
   yi:=random*M;
   if yi<f(xi) then Inc(nc);
end;
calka:=nc*M*(b-a)/N;
```

<u>metoda próbkowania średniej</u> (*sample mean method*)

$$\int_a^b f(x)dx = (b-a)\left\langle f \right\rangle \simeq \frac{(b-a)}{N}\sum_{i=1}^{N} f(x_i)$$

<u>Przykładowy algorytm</u>

```
randomize;
N:=100000;
suma:=0;
for i:=1 to N do
begin
   xi:=random*(b-a) + a;
   suma:=suma + f(xi);
end;
calka:=suma*(b-a)/N;
```

<u>ogólny wzór na całkowanie wielowymiarowe</u>

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} f(\mathbf{X})d\Phi(\mathbf{X}) = \langle f \rangle = \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{X}_i)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} f(x^1,x^2,\ldots,x^K)d\Phi(x^1,x^2,\ldots,x^K) = \langle f \rangle = \frac{1}{N}\sum_{i=1}^{N} f(x_i^1,x_i^2,\ldots,x_i^K)$$

gdzie: $\Phi(\mathbf{X})$ - dystrybuanta zmiennej $\mathbf{X}$: $d\Phi(\mathbf{X}) = \varphi(\mathbf{X})d\mathbf{X}$

np: $d\Phi(x) = \dfrac{1}{b-a}dx$ (1-wym. rozkład równomierny)

Metody generowania liczb pseudolosowych o rozkładzie równomiernym

Generatory liniowe

$$X_{n+1} = \left(a_1 X_n + a_2 X_{n-2} + \ldots + a_k X_{n-k+1} + c\right) \bmod m$$

$$\text{k=1} \Rightarrow X_{n+1} = \left(a_1 X_n + c\right) \bmod m$$

przykłady

a	c	m	*autor*
$2^2 \cdot 23^7 + 1$	0	2^{35}	Zieliński (1966)
69069	1	2^{32}	Marsaglia (1972)
40692	0	2^{31}-249	L'Ecuyer (1988)
68909602460261	0	2^{48}	Fishman (1990)

inne ogólne generatory liniowe (Marsaglia, 1995):

$$X_n = \left(1176 X_{n-1} + 1476 X_{n-2} + 1776 X_{n-3}\right) \bmod (2^{32} - 5)$$
$$X_n = 2^{19}\left(X_{n-1} + X_{n-2} + X_{n-3} + c\right) \bmod (2^{35} - 1629)$$

Generatory nieliniowe

$$X_{n+1} = \left(a X_n^{-1} + b\right) \bmod m \qquad \text{Eichenauer-Lehn (1986)}$$

$$X_n = \left(a(n + n_0) + b\right)^{-1} \bmod m \qquad \text{Eichenauer-Hermann (1993)}$$

Inne generatory

- generatory Fibonacciego
- generatory na rejestrach przesuwnych
- generatory oparte na mnożeniu z przeniesieniem

....

- generatory mieszane

<u>Testy zgodności</u>

- $E(X)$; $D^2(X)$; momenty; χ^2; testy wielowymiarowe, testy par...
- Rozkłady punktów w kostce wielowymiarowej $[1,0]^n$:

1. $\left(X_1, X_2, \ldots, X_d\right), \left(X_2, X_3, \ldots, X_{d+1}\right), \ldots$
2. $\left(X_1, X_2, \ldots, X_d\right), \left(X_{d+1}, X_{d+2}, \ldots, X_{2d}\right), \ldots$

Wybrane testy generatorów liczb losowych

Test chi-kwadrat

Niech: $F(a) = 0, \ \ F(b) = 1,$

$a = a_0 < a_1 < a_2 < \ldots < a_k = b$ - rozbicie zbioru wartości X

$p_i = P\{a_{i-1} < X < a_i\}, \qquad i = 1, 2, \ldots$

n_i - liczba takich elementów X ciągu $X_1, \ldots, X_n$ spełniających $a_{i-1} < X < a_i$

Dla dużego n statystyka

$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

ma rozkład bliski rozkładowi chi-kwadrat o k-1 stopniach swobody.

Dla $p_i = 1/k$ mamy:

$$\chi^2_{k-1} = \frac{k}{n} \sum_{i=1}^{k} n_i^2 - n$$

Podobnie z rozkładami wielowymiarowymi $(X_1, \ldots, X_m)$

Test najmniejszej odległości w parach

Generujemy n punktów z kostki m wymiarowej $(0,1)^m$.

Obliczamy odległości w metryce euklidesowej dla wszystkich $\binom{n}{2}$ par pkt.

Niech D oznacza najmniejszą z tych odległości.

$\Rightarrow$ zmienna losowa $T = n^2 D^m / 2$ ma rozkład wykładniczy ze średnią $1/V_m$

gdzie V_m oznacza objętość m-wymiarowej kuli jednostkowej.

Uwaga. Generatory liniowe zwykle nie spełniają tego testu!

Generowanie liczb pseudolosowych o rozkładzie dowolnym

Metoda odwracania dystrybuanty

Niech α ma rozkład równomierny na przedziale (0,1)

Def. $X = F^{-1}(\alpha)$

bo: $P\{X \le x\} = P\{F^{-1}(\alpha) \le x\} = P\{\alpha \le F(x)\} = F(x)$

$$\Rightarrow \quad F(x_i) = \alpha \qquad \Rightarrow \quad x_i = F^{-1}(\alpha)$$

$$\boxed{\int_{-\infty}^{x_i} f(x)dx = \alpha_i \;\Rightarrow\; x_i = ?}$$

Przykład 1 - funkcja trójkątna

$$f(x) = \begin{cases} Ax, & x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

Przykład 2 - funkcja potęgowa

$$f(x) = \begin{cases} Ax^n, & x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

Przykład 3 - funkcja wykładnicza; $\lambda > 0$

$$f(x) = \begin{cases} Ae^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

Przykład 4 - niejednorodny rozkład wykładniczy:

$$f(x) = \begin{cases} \lambda(x)\exp\left[-\int_0^x \lambda(t)dt\right], & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$\int_{-\infty}^{x} f(t)dt = \alpha \quad \Rightarrow \quad \int_0^x \lambda(t)dt = -\ln(1-\alpha)$$

$$\int_0^x \lambda(t)dt = -\ln\alpha$$ - równanie na x

Metoda eliminacji (von Neumann, 1951)

Wariant szczególny

Niech $f(x)$ będzie gęstością interesującego nas rozkładu.
$\exists_d \forall_{x\in(a,b)}\ f(x) < d, \quad \forall_{x\notin(a,b)}\ f(x) = 0$

1. Generujemy dwie zmienne losowe U_1 i U_2 o rozkładach równomiernych $U_1(a,b)$ i $U_2(0,d)$.
2. Jeżeli $U_2 < f(U_1)$, to przyjąć $X = U_1$; w przeciwnym przypadku parę (U_1, U_2) wyeliminować i powtórzyć obliczenia od p.1.

$\Rightarrow$ Zmienna X ma rozkład $f(x)$.

Wariant ogólny

Wybieramy taką gęstość prawdopodobieństwa $g(x)$, aby generowanie liczb losowych o tej gęstości było łatwe i szybkie oraz wyznaczamy stałą $c > 0$, taką, że

$$\forall_{x\in(a,b)}\ f(x) < cg(x)$$

$g(x)$ - gęstość dominująca

1. Generujemy punkt losowy X o rozkładzie $g(x)$ oraz liczbę losową U o rozkładzie równomiernym $U(0,1)$.
2. Powtarzamy generowanie wg. p.1 dopóki nie zostanie spełniony warunek akceptacji

$$cUg(X) \le f(X)$$

$\Rightarrow$ Zmienna X ma rozkład $f(x)$.

Wybór optymalnej stałej:

$$c = \sup_x \left\{ \frac{f(x)}{g(x)} \right\}$$

Przykład.
Generowanie rozkładu normalnego przy użyciu rozkładu wykładniczego N(0,1).
Ogólnie:

$$N(\mu,\sigma): \quad f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \qquad -\infty < x < \infty$$

Metoda:
1. Początkowo generujemy zmienną losową X o gęstości:

$$f(x) = \begin{cases} \sqrt{\frac{2}{\pi}} \exp\left(\frac{-x^2}{2}\right), & x \ge 0 \\ 0, & x < 0 \end{cases}$$

(dodatnia połówka rozkładu normalnego), przy użyciu gęstości dominującej:
$g(x) = e^{-x}$ dla $x > 0$ - oznaczamy go przez *E(0,1)*.

Uwaga: dla $U(0,1)$, zmienna $(-\ln(U))$ ma rozkład wykładniczy o gęstości *g*.
Stała *c* wynosi

$$c = \sup_x \left\{\frac{f(x)}{g(x)}\right\} = \sqrt{\frac{2e}{\pi}}$$

2. Wyposażamy zmienną X w znak (+) lub (-) z prawdopobieństwem 1/2.

Algorytm:

repeat
  Generuj X o rozkładzie wykładniczym E(0,1)
  Generuj U o rozkładzie równomiernym U(0,1)
until

$$\sqrt{\frac{2e}{\pi}} U e^{-X} \le \sqrt{\frac{2}{\pi}} e^{-X^2/2}$$

Generuj U o rozkładzie równomiernym U(0,1)
if $U \le 0.5$ *then* $X \leftarrow (-X)$
return X

Inne metody generowania rozkładu normalnego

I. Metoda odwaracania dystrybuanty (Odeh i Evans, 1974)

Przybliżenie:

$$F^{-1}(u) = \begin{cases} g(u) & \text{dla } 10^{-20} < u < 0.5 \\ -g(1-u) & \text{dla } 0.5 \le u < 10^{-20} \end{cases}$$

gdzie:

$$g(u) = t - \frac{L(t)}{M(t)}, \qquad t = \sqrt{-2\ln u}$$

$$L(t) = 0.322232431088 + t + 0.342242088547t^2 +$$
$$+0.0204231210245t^3 + 0.0000453642210148t^4$$

$$M(t) = 0.0993484626060 + 0.588581570495t + 0.531103462366t^2$$
$$+0.103537752850t^3 + 0.0038560700634t^4$$

II. Metoda dekompozycji (Marsaglia i Bray 1964)

Symulacja Monte Carlo

- Wybór reprezentacji układu
- Obliczenia dla procesu elementarnego

Wybór stanu początkowego
Repeat
    Określenie prawdopodobieństw zmiany stanu
    Generacja zmiennych losowych odpowiadającym
        wszystkim możliwym przejściom
    Realizacja
until koniec
Zapis wyników

Układ o odpowiedzi liniowej

- powtarzamy realizacje procesu dla N cząstek
- wyniki sumujemy

Układ o odpowiedzi nieliniowej

- znajdujemy minimalne N które dobrze opisuje układ makroskopowy
- obliczenia wykonujemy jednocześnie dla N cząstek

1. Symulacja Monte Carlo z czasem dyskretnym

Wybór stanu początkowego
$t \leftarrow t_0 - \Delta t$
Repeat
    $t \leftarrow t + \Delta t$
    Określenie prawdopodobieństw zmiany stanu
    Generacja zmiennych losowych odpowiadającym
        wszystkim możliwym przejściom w czasie Δt
    Realizacja
until koniec or $t > t_k$
Zapis wyników

2. Symulacja Monte Carlo z czasem ciągłym

Wybór stanu początkowego
$t \leftarrow t_0$
Repeat
    Określenie prawdopodobieństw zmiany stanu
    Generacja czasów przejścia Δt_i dla każdego
        procesu elementarnego P_i
    Wybór najkrótszego czasu $\Delta t_{\min}$ odpowiadającego P_r
    Realizacja pierwszego przejścia P_r
    $t \leftarrow t + \Delta t_{\min}$
until koniec or $t > t_k$
Zapis wyników

Typy procesów ze względu na zależność rozkładu prawdopodobieństwa od historii układu
(wybór chwili początkowej)

Procesy niezależne od historii (procesy Markowa)	Procesy z pamięcią

Przykłady symulacji Monte Carlo z czasem dyskretnym (liniowe)

1. Rozpad promieniotwórczy

2. Ruchy Browna (błądzenie przypadkowe - *random walk*)

Przykłady symulacji Monte Carlo z czasem ciągłym (liniowe)

1. Rozpad promieniotwórczy

2. Transport nośników ładunku w materiale dielektrycznym

Przykład symulacji Monte Carlo z czasem ciągłym (nieliniowy)

Rekombinacja promienista w dielektrykach