Metoda odwrotnej dystrybuanty.

Piotr Fonferek

12 stycznia 2024

Rozkład jednostajny

Załóżmy, że istnieje ciąg niezależnych zmiennych losowych $U_1, U_2, \ldots,$ takich że:

$$P(U_i \leqslant u) = \begin{cases} 0, & u < 0 \\ u, & 0 \leqslant u \leqslant 1 \\ 1, & u > 1 \end{cases}$$

Czyli te zmienne mają rozkład $U[0,1]$.

Metoda odwrotnej dystrybuanty

Twierdzenie

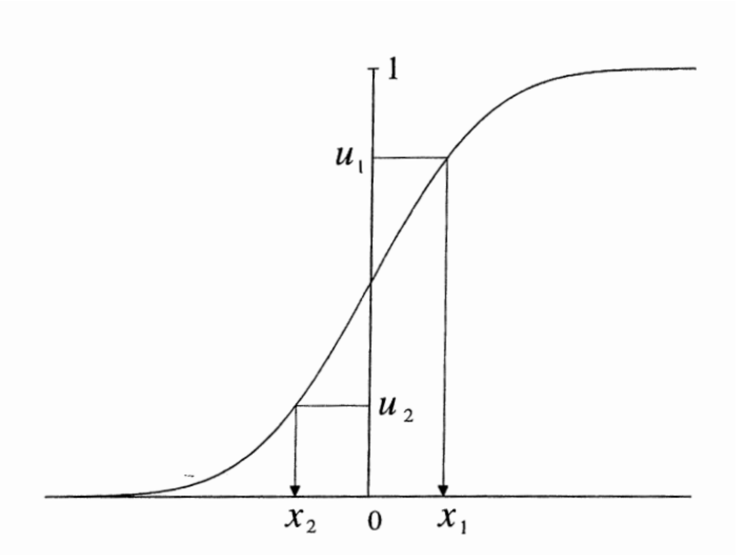
Załóżmy, że mamy rozkład Φ, jego dystrybuantę F i zmienną losową U o rozkładzie $U[0,1]$. Wtedy dla zmiennej losowej

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

zachodzi $P(X \leqslant x) = F(x)$ dla każdego x.

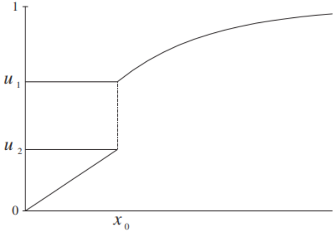
Czyli możemy uzyskać próbkę o dowolnym rozkładzie, znając tylko wzór na jego odwróconą dystrybuantę i mając generator liczb losowych $U[0,1]$.

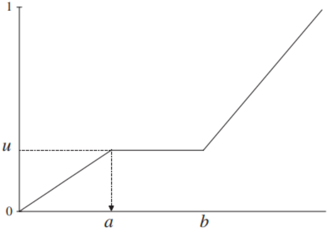
Metoda odwrotnej dystrybuanty



Metoda odwrotnej dystrybuanty

1. Jeśli w dystrybuancie F występują skoki, to $u_1 = F(x-) \leqslant u \leqslant u_2 = F(x+)$ są odwzorywane przez F^{-1} na x_0;
2. Jeśli w dystrybuancie F występują spłaszczenia, to przyjmujemy $F^{-1}(u) = \inf\{x : F(x) \geqslant u\}$.





Metoda odwrotnej dystrybuanty

Sprawdzenie, że ta metoda rzeczywiście zwraca próbkę o rozkładzie Φ.

$$\begin{aligned} P(X \leqslant x) &= P(F^{-1}(U) \leqslant x) \\ &= P(U \leqslant F(x)) \\ &= F(x). \end{aligned}$$

Percentyle

Definicja

Percentyle są podziałem liczb rzeczywistych na rozłączne zbiory $A_1, \ldots, A_{100}$, takie że $\bigcup_{i=1}^{100} A_i = \mathbb{R}$, dla których zachodzi

$$\Phi(A_k) = \Phi(A_i) \tag{1}$$

dla dowolnych $k, i \in \{1, \ldots, 100\}$.

Uwaga

Równość (1) zachodzi również, jeśli zbiory A_k i A_i składają się z sumy n różnych zbiorów $A_1, \ldots, A_{100}$.

Percentyle

Przykład

Uzyskanie wyniku losowania z rozkładu Φ pomiędzy dwudziestym a trzydziestym percentylem jest równie prawdopodobne, jak uzyskanie wyniku pomiędzy osiemdziesiątym, a dziewięćdziesiątym percentylem.

- Metoda odwrotnej dystrybuanty pozwala na szybkie znalezienie percentyli.
- Ze wzoru (1) wynika, że percentyle mają rozkład jednostajny.

Przykład - rozkład wykładniczy

Dystrybuanta F rozkładu wykładniczego o średniej θ wynosi

$$F(x) = 1 - e^{-x/\theta}, \ x \geqslant 0.$$

Teraz obliczamy F^{-1}:

$$\begin{aligned} U &= 1 - e^{-\frac{x}{\theta}} \\ e^{-\frac{x}{\theta}} &= 1 - U \\ X &= -\theta \ln(1 - U) = -\theta \ln(U) \end{aligned}$$

Przykład - rozkład wykładniczy

```
#Metoda odwrotnej dystrybuanty dla rozkładu wykładniczego
x = seq(from=0.01, to=20, by=0.01)
theta <- 0.5 #średnia
lambda <- 1/theta
probka_wykladniczego <- c()
probka_wykladniczego <- -theta*log(runif(100))
hist(probka_wykladniczego, breaks = seq(0, 20, 0.5),
     main="Histogram próbki rozkładu wykładniczego", prob=TRUE)
lines(x, lambda*exp(-lambda*x), col="red", lwd=2)
```

Przykład - rozkład wykładniczy

Rozkład arcus sinusa

Rozkład arcus sinusa

Rozkład o dystrybuancie $F(x) = \frac{2}{\pi}\arcsin(\sqrt{x}),\ \ 0 \leqslant x \leqslant 1$ jest nazywany rozkładem arcus sinusa.
Jego gęstość wynosi $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$.

Prawa arcus sinusa

Niech zmienna losowa M określa czas, w którym proces Wienera osiągnął maksimum w czasie $[0,1]$:

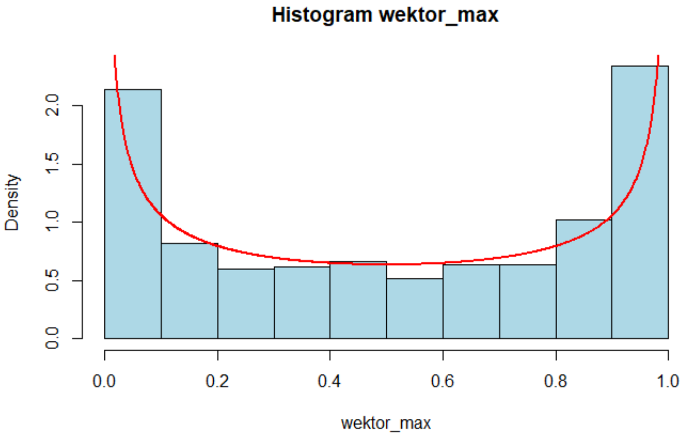
$$W_M = sup\{W_s :\ \ s \in [0,1]\}.$$

Wtedy M ma rozkład arcus sinusa.

Prawo arcus sinusa

```r
sigma <- 0.2
p <- 500 # the number of samples
delta <- 0.001 # delta*n=T in years
n <- 999 # the number of steps
wektor_max <- c() #wektor maksimów


for (k in 1:p) {
  S <- 0
  t <- 0
  wektorS <- c()
  for (j in 1:n) {
    y <- rnorm(1) # Generate a random number from a standard normal distribution
    #t <- t + delta
    S <- S + sigma * y * sqrt(delta)
    wektorS[j] <- S
    #output <- rbind(output, data.frame(t = t, x = x))
  }
  wektor_max[k] <- which.max(wektorS)/1000
}

hist(wektor_max, col = "lightblue", main = "Histogram", prob=TRUE)

# Definiowanie funkcji gęstości rozkładu beta
x <- seq(from = 0.001, to = 0.999, by = 0.001)
y <- 1 / (pi * sqrt(x * (1 - x)))

# Rysowanie linii na tym samym wykresie
lines(x, y, col = "red", lwd = 2)
```

Prawo arcus sinusa



Bez obliczania F^{-1}

Metoda odwrotnej dystrybuanty jest stosowana, nawet jeśli nie można analitycznie wyznaczyć odwzorowania F^{-1}. Wtedy obliczenie $F^{-1}(u)$ jest równoważne znalezieniu miejsca zerowego równania $F(x) - u = 0$. Jeżeli $f(x)$ jest gęstością naszego rozkładu, to korzystając z metody Newtona dostajemy:

$$x_{n+1} = x_n - \frac{F(x_n) - u}{f(x_n)},$$

gdzie, przy rozpoczęciu iteracji w punkcie x_0, x_n zbiega do szukanego pierwiastka.

Przykład - rozkład dyskretny

Rozważmy zmienną losową X o wartościach dyskretnych $c_1 < c_2 < \cdots < c_n$. Niech p_i będzie prawdopodobieństwem osiągnięcia c_i. Następnie ustalmy $q_0 = 0$ oraz

$$q_i = \sum_{j=1}^{i} p_j, \ \ i = 1, \ldots, n.$$

Zauważmy, że $q_i = F(c_i), \ i = 1, \ldots, n$. Teraz stosujemy metodę odwrotnej dystrybuanty:

1. losujemu liczbę $U \sim U[0,1]$;
2. szukamy $K \in \{1, \ldots, n\}$, takie że $q_{K-1} \leqslant U \leqslant q_K$;
3. przypisujemy $X = c_k$.

Rozkład warunkowy

Rozważmy zmienną losową X i losowanie warunkowane $a < X \leqslant b$, gdzie $a, b \in \mathbb{R}$ są, takie że $F(a) < F(b)$. Zdefiniujmy zmienną losową V:

$$V = F(a) + (F(b) - F(a))U.$$

Zauważmy, że V ma rozkład $U[F(a), F(b)]$ i rozkład:

$$\begin{aligned} P(F^{-1}(V) \leqslant x) &= P(F(a) + (F(b) - F(a))U \leqslant F(x)) \\ &= P(U \leqslant [F(x) - F(a)]/[F(b) - F(a)] \\ &= [F(x) - F(a)]/[F(b) - F(a)]. \end{aligned}$$

Dostaliśmy rozkład warunkowy dla $a < X \leqslant b$.

Zadanie

Zadanie

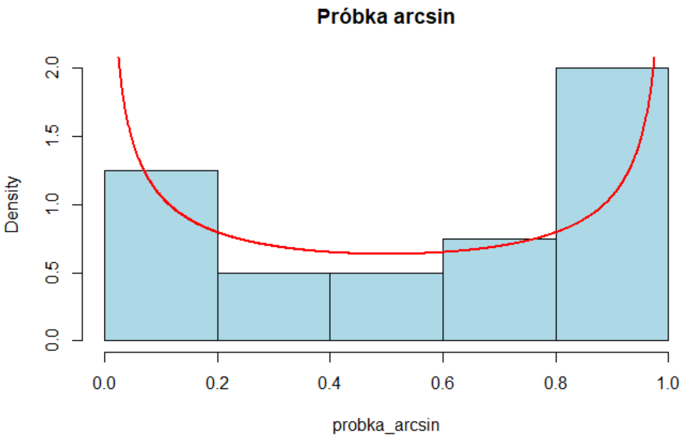
Napisz program, który, przy pomocy metody odwrotnej dystrybuanty, wygeneruje próbkę 20 liczb z rozkładu arcus sinusa.

Rozwiązanie:

1. znaleźć funkcję $F^{-1}(u) = \sin^2(u\frac{\pi}{2}) = \frac{1}{2} - \frac{1}{2}\cos(u\pi)$;
2. napisać kod.

```
#Metoda odwrotnej dystrybuanty dla rozkładu arcsin

probka_arcsin <- c()
probka_arcsin <- 0.5-0.5*cos(runif(20)*pi)

hist(probka_arcsin, col = "lightblue", main = "Próbka arcsin", prob=TRUE)
x <- seq(from = 0.001, to = 0.999, by = 0.001)
y <- 1 / (pi * sqrt(x * (1 - x))) #gęstość arcsin
lines(x, y, col = "red", lwd = 2)

```

Zadanie




```
> print(probka_arcsin)
 [1] 0.006881820 0.827185241 0.901494517 0.933918682 0.867600915 0.467945633 0.230592614 0.581440301 0.978017077
[10] 0.491967789 0.062035470 0.755816095 0.111089756 0.167381695 0.895387184 0.838456811 0.997979273 0.272097322
[19] 0.002545446 0.742030710
```

Źródła

1. Paul Glasserman: "Monte Carlo Methods in Financial Engineering";
2. https://en.wikipedia.org/wiki/Arcsine_distribution
3. https://en.wikipedia.org/wiki/Arcsine_laws_(Wiener_process)