Laboratoria zestaw 6

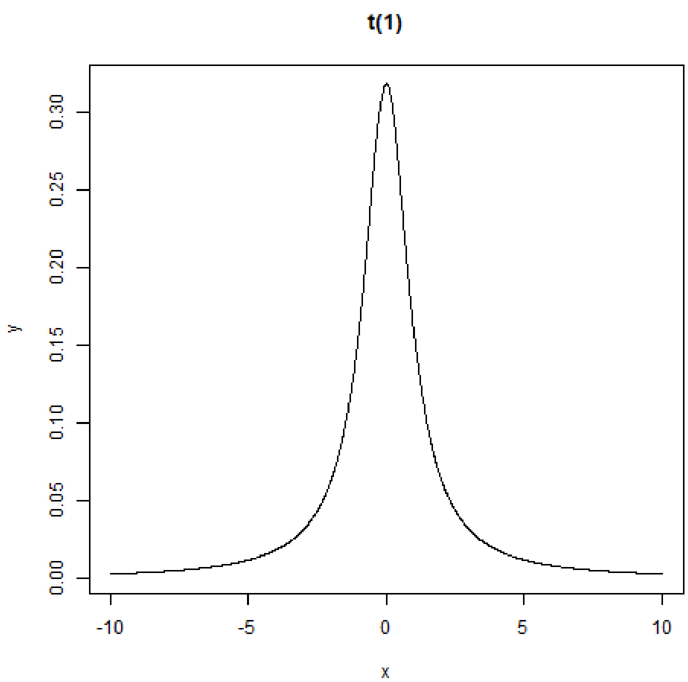
Krystian Baran 145000

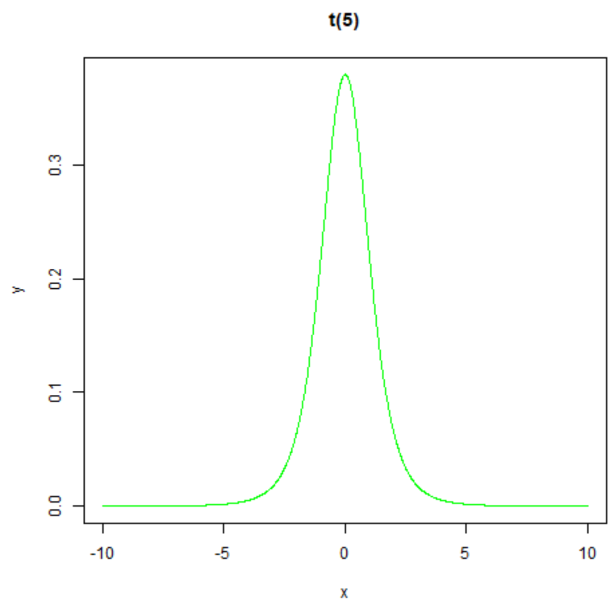
7 kwietnia 2021

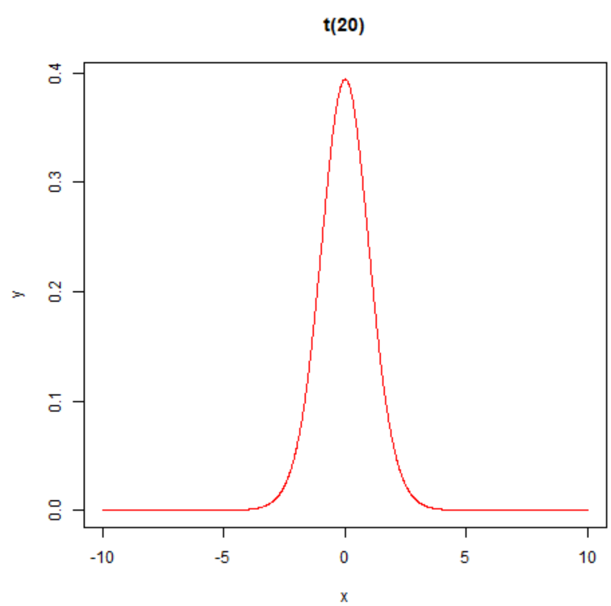
Zadanie 3

Sporządzić krzywe gęstości dla rozkładów t-Studenta $t(1)$,$t(5)$, $t(20)$. Przyjmując, że zmienna losowa X ma podane rozkłady t-Studenta obliczyć prawdopodobieństwa zdarzeń $(X < -2)$ i $(-1 < X < 0)$. Wyznaczyć kwantyle tych rozkładów.

Poniżej przedstawione zostały wykresy funkcji gęstości sporządzone na pomocą R.




t(5)
y
0.0
0.1
0.2
0.3
-10
-5
0
5
10
x


t(20)
y
0.0
0.1
0.2
0.3
0.4
-10
-5
0
5
10
x

Dane prawdopodobieństwa wyliczone zostały także w R za pomocy dystrybuanty funkcji t-Studenta, czyli $P(X < -2) \overset{R}{=} pt(-2, a)$ i $P(-1 < X < 0) \overset{R}{=} pt(0, a) - pt(-1, a)$, gdzie przez a oznaczone zostały podane stopnie swobody ($a = 1, 5, 20$).

Dla $t(1)$:

$$P(X < -2) \overset{R}{=} pt(-2, 1) \approx 0.1475836$$

$$P(-1 < X < 0) \overset{R}{=} pt(0, 1) - pt(-1, 1) \approx 0.25$$

Dla $t(5)$:

$$P(X < -2) \overset{R}{=} pt(-2, 5) \approx 0.05096974$$

$$P(-1 < X < 0) \overset{R}{=} pt(0, 5) - pt(-1, 5) \approx 0.3183913$$

Dla $t(20)$:

$$P(X < -2) \overset{R}{=} pt(-2, 20) \approx 0.02963277$$

$$P(-1 < X < 0) \overset{R}{=} pt(0, 20) - pt(-1, 20) \approx 0.3353717$$

Kwantyle wyliczone zostały za pomocą funkcji kwantylowej $qt(p, n)$ gdzie p to prawdopodobieństwo, a n to stopnie swobody.

Dla $t(1)$:

$$x_{0.25} \overset{R}{=} qt(0.25, 1) \approx -1$$
$$x_{0.5} \overset{R}{=} qt(0.5, 1) \approx 0$$
$$x_{0.75} \overset{R}{=} qt(0.75, 1) \approx 1$$

Dla $t(5)$:

$$x_{0.25} \overset{R}{=} qt(0.25, 5) \approx -0.7266868$$
$$x_{0.5} \overset{R}{=} qt(0.5, 5) \approx 0$$
$$x_{0.75} \overset{R}{=} qt(0.75, 5) \approx 0.7266868$$

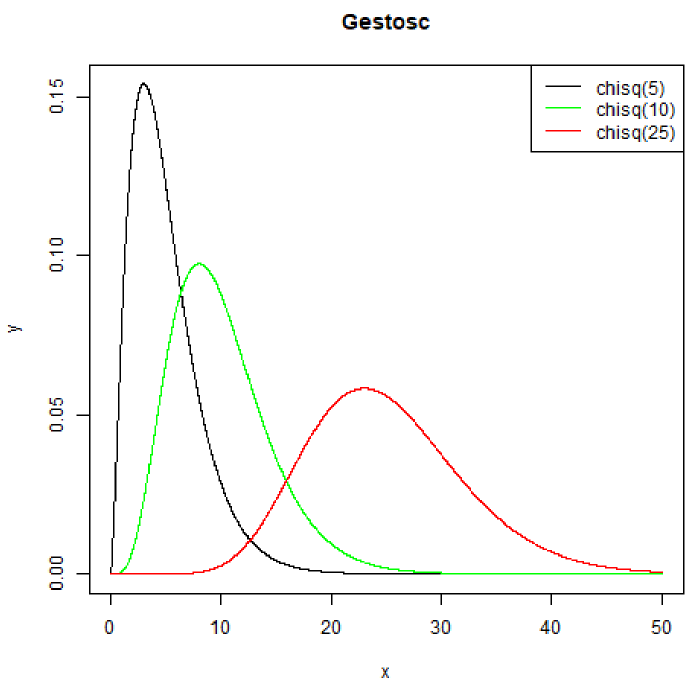
Dla $t(20)$:

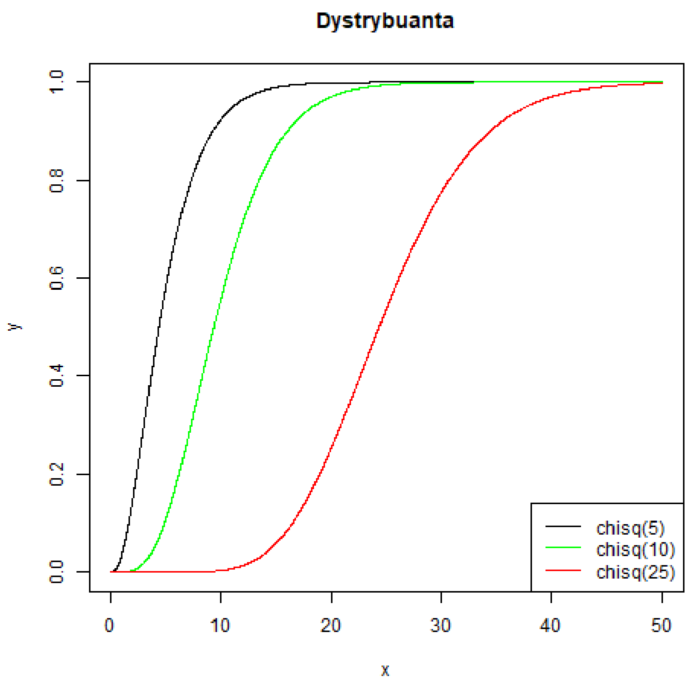
$$x_{0.25} \overset{R}{=} qt(0.25, 20) \approx -0.6869545$$
$$x_{0.5} \overset{R}{=} qt(0.5, 20) \approx 0$$
$$x_{0.75} \overset{R}{=} qt(0.75, 20) \approx 0.6869545$$

Zadanie 4

Sporządzić krzywe gęstości i wykresy dystrybuant zmiennych losowych o rozkładach *chi-kwadrat* z 5, 10 i 25 stopniami swobody. Czy można zauważyć jakąś prawidłowość, analizując kolejne wykresy? Wiedząc, że $X \sim CHIS(25)$, wyznaczyć prawdopodobieństwa zdarzeń $(X < 15)$, $(X > 25)$, $(20 < X < 30)$. Wyznaczyć kwantyle tego rozkładu.

Wykresy sporządzone zostały za pomocą R i wyglądają następująco:





Można zauważyć że gęstość dla $n \to \infty$, gdzie n oznacza stopnie swobody, dąży do rozkładu normalnego.

Prawdopodobieństwa dla $X \sim \chi^2_{25}$ wyliczone zostały w R i wyglądają następująco:

$$P(X < 15) \overset{R}{=} pchisq(15, 25) \approx 0.05861743$$
$$P(X > 25) = 1 - P(X < 25) \overset{R}{=} 1 - pchisq(25, 25) \approx 0.4623737$$
$$P(20 < X < 30) \overset{R}{=} pchisq(30, 25) - pchisq(20, 25) \approx 0.5225363$$

Kwantyle tego rozkładu są następujące:

$$x_{0.25} \overset{R}{=} qchisq(0.25, 25) \approx 19.93934$$
$$x_{0.5} \overset{R}{=} qchisq(0.5, 25) \approx 24.33659$$
$$x_{0.75} \overset{R}{=} qchisq(0.75, 25) \approx 29.33885$$

Zadanie 5

Wiedząc, że $X \sim F(5, 10)$ wyznaczyć prawdopodobieństwo zdarzenia $X > 1.8027$ oraz kwantyle tego rozkładu

Prawdopodobieństwo wyznaczone, jak poprzednio, za pomocą funkcji w R dla rozkładu F.

$$P(X > 1.8027) = 1 - P(X < 1.8027) \stackrel{R}{=} 1 - pf(1.8027, 5, 10) \approx 0.2000048$$

Zatem jest to zbliżone do 0.2.

Kwantyle także wyznaczone w R i wyglądają następująco:

$$x_{0.25} \stackrel{R}{=} qf(0.25, 5, 10) \approx 0.5291417$$
$$x_{0.5} \stackrel{R}{=} qf(0.5, 5, 10) \approx 0.9319332$$
$$x_{0.75} \stackrel{R}{=} qf(0.75, 5, 10) \approx 1.585323$$

Zadanie 6

Rozważmy eksperyment symulacyjny, w którym rozkład populacji istotnie różni się od rozkładu normalnego.

a) Czas zdatności pewnego typu elektronicznego sterownika ma rozkład wykładniczy z wartością oczekiwaną 5000 dni.

b) Czas oczekiwania na autobus ma rozkład jednostajny na przedziale (0, 15)

minut. Wyznaczyć rozkład średniej arytmetycznej dla $n = 5, 10, 30$.

Poniżej przedstawiono skrypt R-owski użyty do wygenerowania rozkładu średnich.

```
means = c()
Means = c()

for(i in 1:10000){
   means = c(means,round(mean(rexp(5,1/5000))))
   Means = c(Means,round(mean(runif(5,0,15))))
   } means = table(means)
Means = table(Means)
png("n5.png")
plot(means,type = "h", col = "red", xlab = "mean", ylab = "n", main = "n = 5, exp(1/5000)")
dev.off()
png("n5_1.png")
plot(Means,type = "h", col = "red", xlab = "mean", ylab = "n", main = "n = 5, unif(0,15)")
dev.off()
write.table(means,file = "n5.csv", sep = ",")
write.table(Means,file = "n5.csv", sep = ",", append = TRUE)

# n = 10 means = c()
Means = c()
for(i in 1:10000){
   means = c(means,round(mean(rexp(10,1/5000))))
   Means = c(Means,round(mean(runif(10,0,15))))
   }
means = table(means)
Means = table(Means)
png("n10.png")
plot(means,type = "h", col = "blue", xlab = "mean", ylab = "n", main = "n = 10, exp(1/5000)")
dev.off()
png("n10_1.png")
plot(Means,type = "h", col = "red", xlab = "mean", ylab = "n", main = "n = 10, unif(0,15)")
dev.off()
write.table(means,file = "n10.csv", sep = ",")
```

```
write.table(Means,file = "n10.csv", sep = ",", append = TRUE)

# n = 30
means = c()
Means = c()
for(i in 1:10000){
   means = c(means,round(mean(rexp(30,1/5000))))
   Means = c(Means,round(mean(runif(30,0,15))))
   }
means = table(means)
Means = table(Means)
png("n30.png")
plot(means,type = "h", col = "green", xlab = "mean", ylab = "n", main = "n = 30, exp(1/5000)")
dev.off()
png("n30_1.png")
plot(Means,type = "h", col = "red", xlab = "mean", ylab = "n", main = "n = 30, unif(0,15)")
dev.off()
write.table(means,file = "n30.csv", sep = ",")
write.table(Means,file = "n30.csv", sep = ",", append = TRUE)
```

a)

Znając wartość oczekiwaną można wyznaczyć parametr λ rozkładu wykładniczego. Wynosi ona $\frac{1}{5000}$. Następnie wygenerowana została próba losowa $n = 5, 10, 30$ elementów i wyliczona z nich średnia. Średnich wygenerowano 10000 dla każdego n. Wartości zostały zaokrąglone do liczb całkowitych.
Następnie wygenerowane wartości wgrano do MS Excel w celu łatwiejszego obliczenia wartości średniej i odchylenia standardowego. Ponieważ dane zapisano w tabeli w formie rozkładu punktowego korzystano ze wzoru na średnia i odchylenie standardowe rozkładu punktowego tj:

$$\overline{X} = \frac{\sum x_i n_i}{N}$$

$$\sigma = \sqrt{\frac{\sum (x_i - \overline{X})^2 n_i}{N}}$$

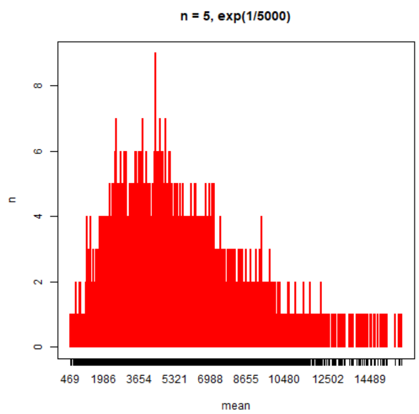
Gdzie N wynosi 10000, x_i to są średnie a n_i ich liczebność. Uzyskano następujące wartości:

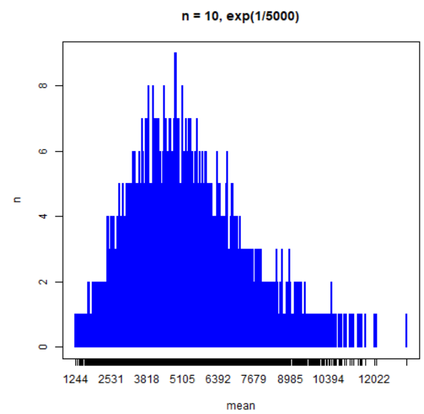
n	5	10	30
$\overline{X}$	5020.5888	5025.0357	5003.5542
σ	2266.355364	1579.013947	922.8496014

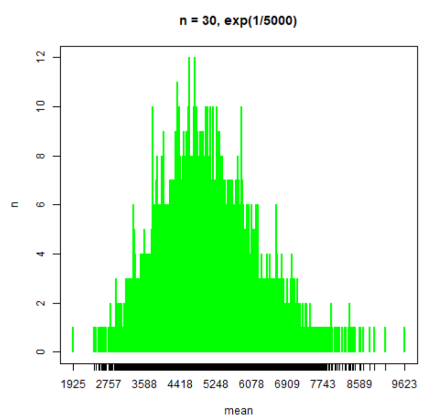
Następnie obliczono wartości teoretyczne korzystając z centralnego twierdzenia granicznego dla średniej.

n	5	10	30
$\mathbb{E}X$	5000	5000	5000
$\mathbb{D}X$	2236.067977	1581.13883	912.8709292

Widzimy zatem że rozkład średniej arytmetycznej dla $n \to \infty$ dąży do rozkładu normalnego z parametrami $\mathbb{E}X$ i $\mathbb{D}X/\sqrt{n}$. Natomiast uzyskaliśmy wartości nie co mniejsze przez zaokrąglenie wartości, co zostało wykonane po to aby wykres liczebności nie wyglądał jak pasek. Poniżej przedstawiono wygenerowane wykresy.







b)

Jak w podpunkcie **a)** wygenerowano 10000 prób średnich dla $n = 5, 10, 30$ i każda średnia została zaokrąglona do liczby całkowitej. Uzyskane dane wprowadzono w MS Excel i jak poprzednio obliczono średnią i odchylenie standardowe.

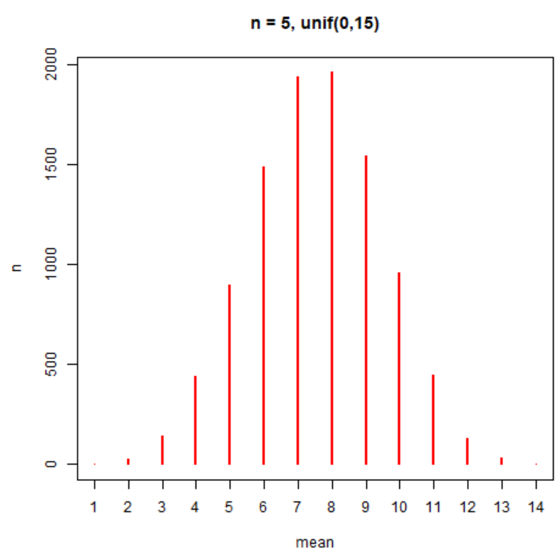
Uzyskano następujące wartości:

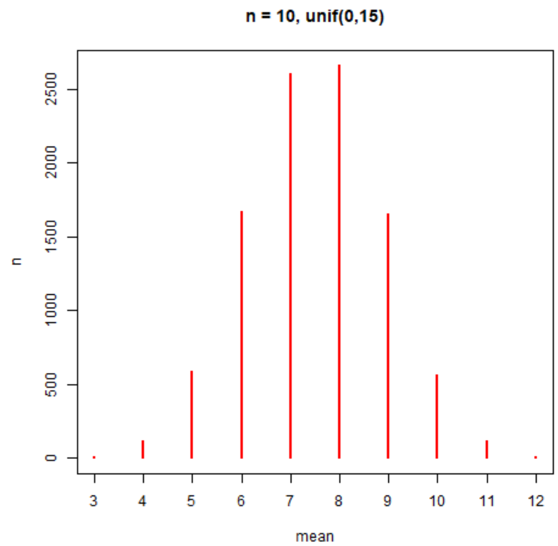
n	5	10	30
$\overline{X}$	7.521	7.4937	7.4909
σ	1.941277672	1.387141056	0.840069753

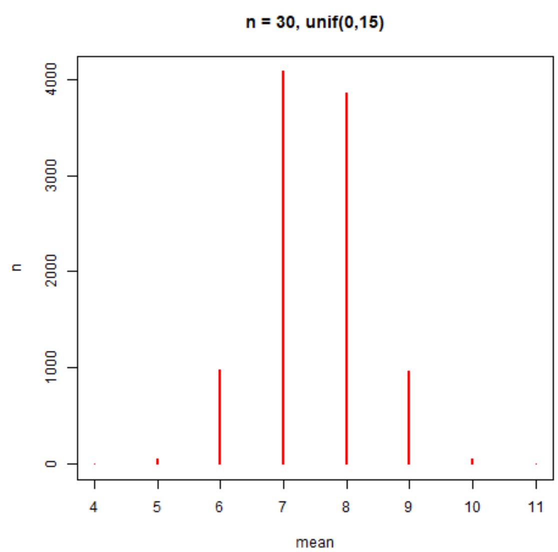
Obliczono także wartości teoretyczne wynikające z centralnego twierdzenia granicznego. Dla rozkładu jednostajnego $\mathbb{E}X = \frac{a+b}{2} = \frac{15}{2} = 7.5$ natomiast $\mathbb{D}^2(X) = \frac{b-a}{12} = \frac{15^2}{12} = 18.75$. Obliczone wartości zapisano w tabeli poniżej:

n	5	10	30
$\mathbb{E}X$	7.5	7.5	7.5
$\mathbb{D}X$	1.936491673	1.369306394	0.790569415

Widzimy zatem dla $n \to \infty$ wartość oczekiwana zmniejsza się bardzo powili, natomiast wariancja bardzo różni się od wartości teoretycznej. Taka różnica może wynikać z tego że średnie zostały zaokrąglane do liczby całkowitej; zatem trzeba uważać gdy dokonuję się próbę i chce się zrobić tak żeby wykres wyglądał bardziej gładko. Poniżej przedstawiono wykresy.




n = 10, unif(0,15)
n
0
500
1000
1500
2000
2500
3
4
5
6
7
8
9
10
11
12
mean


n = 30, unif(0,15)
n
0
1000
2000
3000
4000
4
5
6
7
8
9
10
11
mean

Zadanie 7

Korzystając z twierdzenia o odwracaniu dystrybuanty, wygenerować realizację 5- elementowej próby według rozkładu $BT(2,1)$.

W R dostępna jest funkcja generująca losowe liczby według danego rozkładu. Natomiast wykorzystamy dostępną funkcje odwrotną dystrybuanty czyli funkcja kwantylowa.
Wygenerujemy najpierw 5 losowych liczb z rozkładu jednostajnego na przedziale (0,1) za pomocy funkcji R-owskiej *runif(n, a, b)* która generuje n liczb losowych według rozkładu jednostajnego na przedziale (a,b). Liczb te są na przykład następujące:
[1] 0.8740339 0.1719382 0.7142548 0.4222149 0.2073910

Następnie skorzystamy z tych liczb w funkcji odrotnej dystrybuanty aby wyznaczyć losowe liczb według rozkładu Beta. Czyli, jeżeli liczby powyżej zapisano do zmiennej p:

$$qbeta(p,2,1) =$$

[1] 0.9348978 0.4146543 0.8451359 0.6497807 0.4554020

Liczby tę są losowe liczby według rozkładu $BETA(2,1)$.

Zadanie 8

Wygenerować 5-elementową próbę losową zgodnie z rozkładem o gęstości danej wzorem:

a) $f(x) = 2(x-1)\mathbb{I}_{1;2}(x)$,

b) $f(x) = 2x \cdot e^{-x^2}$

Aby wygenerować losową próbę z podanych rozkładów, skorzystamy z twierdzenia o odwróceniu dystrybuanty. tj. jeżeli X jest zmienną losową typu ciągłego o dystrybuancie F_X, to $Z = F_X(X) \sim U(0,1)$.
Zatem generując losową liczbę rozkładu jednostajnego na przedziale $(0,1)$, i obliczając odwrotność dystrybuanty danego dowolnego rozkładu, otrzymamy próbę z tego rozkładu. Zatem musimy najpierw obliczyć dystrybuanty i odwrócić je.

a)

Jeżeli funkcję gęstości scałkujemy uzyskamy dystrybuantę. Zatem:

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} 2(t-1)\mathbb{I}_{(1,2)}(t)dt \\ &= \int_{1}^{x} 2(t-1)dt = (t-1)^2\Big|_1^x \\ &= \begin{cases} 0 & , x < 1 \\ (x-1)^2 & , 1 \leqslant x \leqslant 2 \\ 1 & , x > 2 \end{cases} \end{aligned}$$

Odwracając teraz dystrybuantę:

$$\begin{aligned} y &= (x-1)^2 \\ \sqrt{y} &= x - 1 \\ x &= 1 + \sqrt{y} = F^{-1}(y) \end{aligned}$$

Gdzie $y \in (0,1)$.

Za pomocą R wygenerowana została próba 5 liczb z rozkładu jednostajnego na przedziale $(0,1)$. Skrypt R-owski jest następujący:

```
x = c()
F1 = function(x) {
    if(x < 0) { F1 = 0 }
    if(x > 1) { F1 = 0 }
    if(0 <= x && x <= 1)  F1 = 1 + sqrt(x)
}
for(i in 1:5) {
   x = c(x, F1(runif(1,0,1)))
}
print(x)
```

Wynik takiego skryptu jest następujący:
[1] 1.808751 1.956776 1.498300 1.565768 1.814085

b)

Podobnie jak poprzednio obliczymy dystrybuantę i ja odwrócimy:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} 2t \cdot e^{-t^2} dt \\ &= \lim_{a \to -\infty} e^{-t^2}\Big|_a^x \\ &= -e^{-x^2} + c \end{aligned}$$

Ponieważ dystrybuanta musi być większa od 0, ale mniejsza od 1, wartość stałej c wynosi $+1$. Można wtedy obrócić dystrybuantę:

$$\begin{aligned} y &= -e^{-x^2} + 1 \\ 1 - y &= e^{-x^2} \\ \ln(1-y) &= -x^2 \\ \sqrt{-\ln(1-y)} &= x = F^{-1}(y) \end{aligned}$$

Wtedy, za pomocą skryptu R-owskiego, można łatwo wygenerować 5 elementową próbe według danego rozkładu:

```
x1 = c()
F2 = function(x) {
    a = 1-x
    if(x < 0)  F2 = 0
    if(x > 1)  F2 = 0
    if(x <= 1 && x >= 0) F2 = sqrt(-log(a, base = exp(1)))
}
for (i in 1:5){
    x1 = c(x1, F2(runif(1,0,1)))
}
print(x1)
```

Wynik takiego działania jest 5 elementowa próba, na przykład:
[1] 0.3928419 0.7199171 0.8786929 0.2694350 0.2226001