Wybrane rozkłady ciągłe cech statystycznych

rozkład Normalny, t-Studenta oraz Chi-kwadrat

(literatura: Ostasiewicz, Rusnak, Siedlecka. Statystyka elementy teorii i zadania. Wydawnictwo AE we Wrocławiu, Wrocław 1999 (2006 – wydanie 6) – rozdział 5

Przykład 1. Dla jakiej wartości C następująca funkcja jest gęstością prawdopodobieństwa zmiennej losowej X.

$$f(x) = \begin{cases} 0 & dla \quad x < 0 \\ C * (1 - x) & dla \quad 0 \leq x \leq 1 \\ 0 & dla \quad x > 1 \end{cases}$$

a) Po wyznaczeniu C naszkicować wykres funkcji $f(x)$.
b) Wyznaczyć dystrybuantę zmiennej losowej X oraz naszkicować jej wykres.
c) Obliczyć następujące prawdopodobieństawa:
   - $P(X < 0{,}5)$
   - $P(0 < X < 0{,}75)$
   - $P(0{,}5 < X < 0{,}75)$
   - $P(X > 1)$
d) Obliczyć wartość oczekiwaną oraz wariancję dla zmiennej losowej X.

Rozwiązanie:

Ad a)

Aby następujaca funkcja była funkcją gęstości rozkładu prawdopodobieństwa pewnej zmiennej losowej (statystyki) X musi spełniać warunek: $\int_{-\infty}^{+\infty} f(x)dx = 1$.

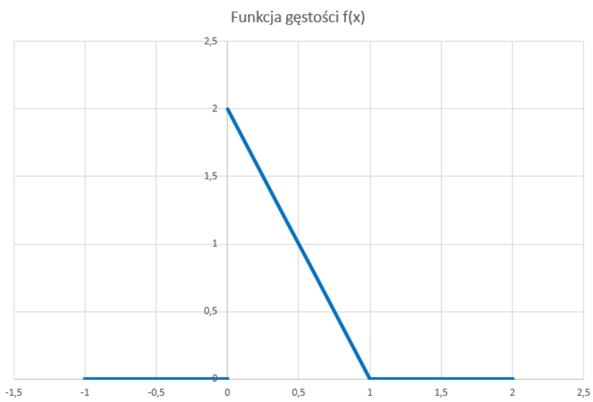
$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{1} C * (1 - x)dx + \int_{1}^{+\infty} 0dx = \left[C * \left(x - \frac{1}{2} * x^2\right)\right]_0^1 = C * \left(1 - \frac{1}{2}\right) - 0 = \frac{1}{2} * C$$

Ponieważ $\frac{1}{2} * C = 1$, to $C = 2$.

Funkcja gęstości zmiennej losowej X jest postaci:

$$f(x) = \begin{cases} 0 & dla \quad x < 0 \\ 2 * (1 - x) & dla \quad 0 \leq x \leq 1 \\ 0 & dla \quad x > 1 \end{cases}$$

Wykres funkcji gęstości przedstawia rysunek 1.



Rys.1. Funkcja gęstości zmiennej losowej X

Ad b)

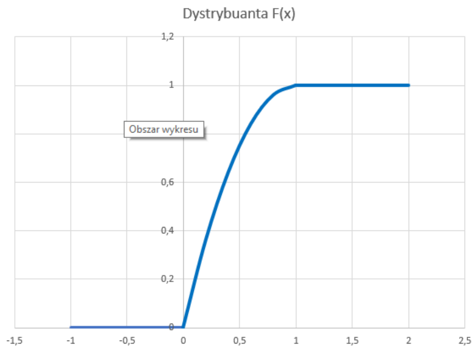
Dystrybuantę zmiennej losowej X wyznaczamy jako funkcję górnej granicy całkowania ze wzoru: $F_d(x) = \int_{-\infty}^{x} f(t)dt$

Dla $x < 0$ wartość dystrybuanty jest oczywiście równa $F_d(x) = 0$

Dla $0 \leq x \leq 1$ wartość dystrybuanty jest równa $F_d(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{x} 2*(1-t)dt = \left[2*\left(t - \frac{1}{2}*t^2\right)\right]_0^x = 2*\left(x-\frac{1}{2}*x^2\right) - 0 = 2*x - x^2$.

Dla $x > 1$ wartość dystrybuanty jest równa $F_d(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{1} 2*(1-t)dt + \int_{1}^{x} 0dt = 0 + 1 + 0 = 1$.

Wykres dystrybuanty zmiennej losowej X przedstawia rysunek 2.



Rys.2. Dystrybuanta zmiennej losowej X

Ad c)

Szukane prawdopodobieństwa możemy wyznaczyć w oparciu o wartości funkcji gęstości lub odczytać z wartości funkcji dystrybuanty.

$P(X < 0{,}5) = \int_{-\infty}^{0{,}5} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{0{,}5} 2*(1-x)dx = 0 + \left[2*\left(x-\frac{1}{2}*x^2\right)\right]_0^{0{,}5} = 0 + 2*\left(0{,}5 - \frac{1}{2}*0{,}5^2\right) - 0 = 1 - 0{,}5^2 = 1 - 0{,}25 = 0{,}75$ (75%).

Od razu z wartości dystrybuanty otrzymujemy: $P(X < 0{,}5) = F_d(0{,}5) = 2*0{,}5 - 0{,}5^2 = 0{,}75$.

$P(0 < X < 0{,}75) = \int_{0}^{0{,}75} f(x)dx = \int_{0}^{0{,}75} 2*(1-x)dx = \left[2*\left(x-\frac{1}{2}*x^2\right)\right]_0^{0{,}75} = 2*\left(0{,}75 - \frac{1}{2}*0{,}75^2\right) - 0 = 0{,}9375$ (93,7 %).

Z wartości dystrybuanty otrzymujemy: $P(0 < X < 0{,}75) = F_d(0{,}75) - F_d(0) = 2*0{,}75 - 0{,}75^2 - 0 = 0{,}9375$.

$P(0{,}5 < X < 0{,}75) = \int_{0{,}5}^{0{,}75} f(x)dx = \int_{0{,}5}^{0{,}75} 2*(1-x)dx = \left[2*\left(x-\frac{1}{2}*x^2\right)\right]_{0{,}5}^{0{,}75} = 2*\left(0{,}75 - \frac{1}{2}0{,}75^2\right) - 2*\left(0{,}5 - \frac{1}{2}*0{,}5^2\right) = 0{,}9375 - 0{,}75 = 0{,}1875$ (18,7 %).

Z wartości dystrybuanty otrzymujemy: $P(0{,}5 < X < 0{,}75) = F_d(0{,}75) - F_d(0{,}5) = (2*0{,}75 - 0{,}75^2) - (2*0{,}5 - 0{,}5^2) = 0{,}1875$.

$P(X > 1) = 1 - P(X \leq 1) = 1 - 1 = 0$ Z dystrybuanty $P(X > 1) = 1 - P(X \leq 1) = 1 - F_d(1) = 1 - 1 = 0$.

Ad d)

Wartość oczekiwaną zmiennej losowej X obliczamy ze wzoru: $E[X] = \int_{-\infty}^{+\infty} x * f(x)dx = \int_0^1 x * 2 * (1 - x)dx = \int_0^1 (2 * x - 2 * x^2)dx = \left[2 * \frac{1}{2} * x^2 - 2 * \frac{1}{3} * x^3\right]_0^1 = \left(1 - \frac{2}{3}\right) - 0 = \frac{1}{3}$

Wartość oczekiwana $E[X] = \frac{1}{3}$.

Wariancję zmiennej losowej X obliczamy ze wzoru: $\mathrm{V}[X] = \sigma^2[X] = E[X^2] - E[X]^2$

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 * f(x)dx = \int_0^1 x^2 * 2 * (1 - x)dx = \int_0^1 (2 * x^2 - 2 * x^3)dx = \left[2 * \frac{1}{3} * x^3 - 2 * \frac{1}{4} * x^4\right]_0^1 = \left(\frac{2}{3} - \frac{1}{2}\right) - 0 = \frac{1}{6}$$

$\mathrm{V}[X] = \frac{1}{6} - \frac{1}{3}^2 = \frac{1}{6} - \frac{1}{9} = \frac{3-2}{18} = \frac{1}{18}$, zaś odchylenie standardowe $\sigma[X] = \sqrt{\frac{1}{18}}$

Przykład 2. Prawdopodobieństwo wykrycia awarii przyrzadów pomiarówych w magazynach pewnej firmy logistycznej w przeciągu czasu nie dłuższego niż t (minut) wynosi: $F(t) = 1 - e^{-2*t}$.

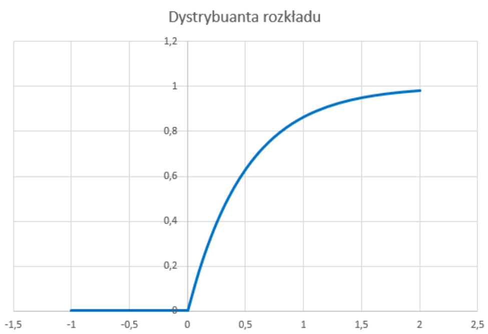
a) Podać funkcję dystrybuanty dla zmiennej losowej T – czas [w min] jaki upłynął do momemntu wykrycia awarii przyrządów i narysować jej wykres.
b) Określić funkcję gęstości rozkładu prawdopodobieństwa zmiennej losowej T i narysować jej wykres.
c) Obliczyć prawdopodobieństwa:
   - na wykrycie awarii pracownik potrzebuje nie więcej niż 5 [min]
   - na wykrycie awarii pracownik potrzebuje więcej niż 3 [min]
d) Obliczyć wartość oczekiwaną oraz wariancję dla czasu potrzebnego na wykrycie awarii.

Ad a)

Podana funkcja dla prawdopodobieństwa wykrycia awarii w czasie nie dłuższym niż t (minut) jest funkcją dystrybuanty badanej zmiennej losowej T: $P(T \leq t) = F(t) = 1 - e^{-2*t}$.

Równanie dystrybuanty będzie zatem następujące: $F_d(t) = \begin{cases} 0 & dla \quad t < 0 \\ 1 - e^{-2*t} & dla \quad t \geq 0 \end{cases}$

Rysunek przedstawia jej wykres:



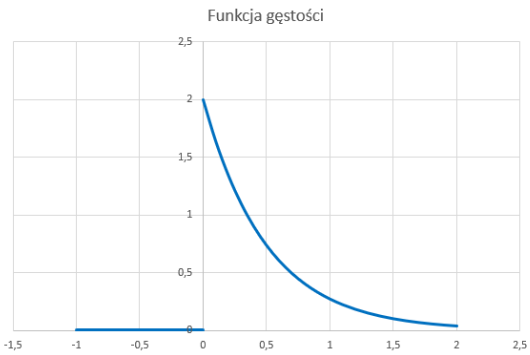
Ad b)

Funkcja gęstości jest pochodna z dystrybuanty rozkładu: $\mathrm{f}(t) = \frac{d}{dt}[1 - e^{-2*t}] = 0 - (-2) * e^{-2*t} = 2 * e^{-2*t}$

Równanie funkcji gęstości rozkładu prawdopodobieństwa jest zatem następujace:

$$f(t) = \begin{cases} 0 & dla \quad t < 0 \\ 2 * e^{-2*t} & dla \quad t \geq 0 \end{cases}$$

Rysunek przedstawia jej wykres:



Ad c)

Szukane prawdopodobieństwa wynoszą:

$P(T \leq 5[min]) = 1 - e^{-2*5} = 1 - 2{,}72^{-10} = 0{,}9999$ (99,9 %)

$P(T > 3[min]) = 1 - P(T \leq 3[min]) = 1 - F_d(3) = 1 - (1 - e^{-2*3}) = 2{,}72^{-6} = 0{,}00247$ (0,2 %)

Ad d)

W rozważanym zadaniu rozkład prawdopodobieństwa jewst tzw. rozkładem wykładniczym. Dla rozkładu wykładniczego z dystrybuantą: $F(t) = 1 - e^{-\lambda*t}$ (w naszym zadaniu parametr $\lambda = 2$) wartość oczekiwana wynosi: $E[X] = \frac{1}{\lambda}$, zaś wariancja $\mathrm{V}[X] = \frac{1}{\lambda^2}$. Zatem $E[X] = \frac{1}{2}$, zaś $\mathrm{V}[X] = \frac{1}{2^2} = \frac{1}{4}$.

Przykład 3. Ustalono, że średni roczny błąd prognozy dochodów pewnej firmy logistycznej ma rozkład normalny ze średnią 31,3 % i odchyleniem standardowym 10 %. Załóżmy, że otrzymaliśmy roczną prognozę dochodów firmy.

a) Jakie jest prawdopodobieństwo, że błąd prognozy będzie zawarty między 20 % i 25 % ?
b) Jakie jest prawdopodobieństwo, że błąd prognozy będzie większy niż 50 % ?

Rozwiazanie:

Zakładamy, że rozkład X - średni roczny błąd prognozy dochodów pewnej firmy logistycznej ma rozkład $N(\mu, \sigma)$=N(31,3;10).

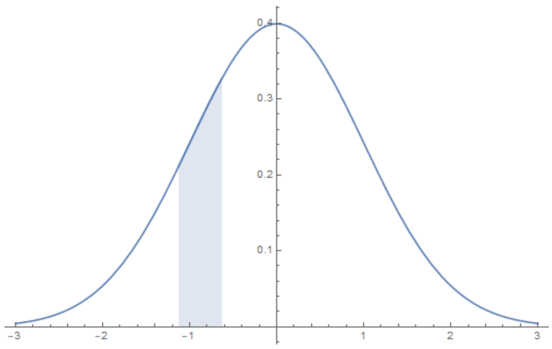
Dokonujemy standartyzacji tedo rozkładu sprowadzając go do rozkładu Normalnego N(0,1).

Jeżeli zmienna losowa X: $N(\mu, \sigma)$, to zmienna $Z = \frac{X-\mu}{\sigma} : N(0,1)$.

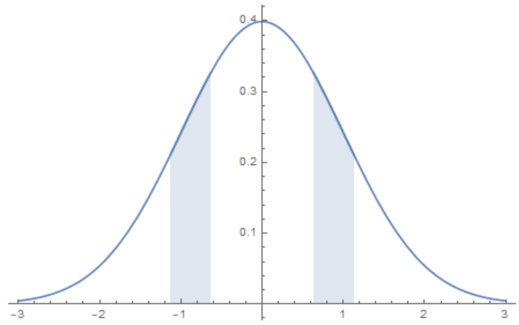
Ad a)

Szukamy prawdopodobieństwa $P(20 \leq X \leq 25) = P\left(\frac{20-31{,}3}{10} \leq \frac{X-31{,}3}{10} \leq \frac{25-31{,}3}{10}\right) = P(-1{,}13 \leq Z \leq -0{,}63)$

Ilustracja graficzna szukanego prawdopodobieństwa jako pole pod krzywą funkcji gęstości:



Z racji symetrii rozkładu normalnego (zob. rys.) mamy $P(20 \leq X \leq 25) = P(-1{,}13 \leq Z \leq -0{,}63) = P(0{,}63 \leq Z \leq 1{,}13) = \varphi(1{,}13) - \varphi(0{,}63)$, gdzie: $\varphi(x) = \int_0^x f(t)dt$, $f(t) = \frac{1}{\sqrt{2*\pi}} * e^{\frac{-t^2}{2}}$ – funkcja gestości rozkładu Normalnego (Gaussa) z parametrami $\mu = 0$, $\sigma = 1$ (standaryzowanego).

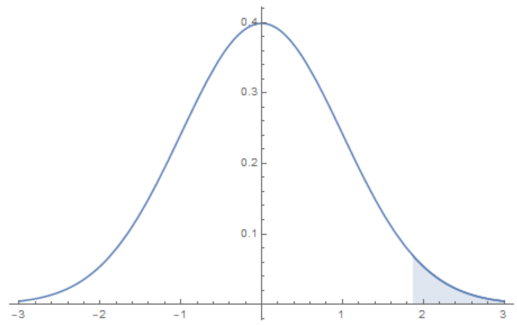


Wartości (całki) funkcji $\varphi(x)$, są stablicowane (Ostasiewicz (red.), Statystyka elementy teorii i zadania, tablica I, str. 379)

Zatem z tablic (tablica 1) $P(20 \leq X \leq 25) = \varphi(1{,}13) - \varphi(0{,}63) = 0{,}3749 - 0{,}2422 = 0{,}13$ (13%)

Ad b)

Szukamy prawdopodobieństwa: $P(X > 50) = P\left(\frac{X-31{,}3}{10} > \frac{50-31{,}3}{10}\right) = P(Z > 1{,}87)$



$P(X > 50) = P(Z > 1{,}87) = \int_{1{,}87}^{+\infty} f(t)dt = \int_0^{+\infty} f(t)dt - \int_0^{1{,}87} f(t)dt = 0{,}5 - \varphi(1{,}87) = 0{,}5 - 0{,}4678 = 0{,}03$ (3%).

Przykład 4. Zmienna losowa $X = T_{df=9}$ ma rozkład t-Studenta o $(df = ss = n = 9)$ stopniach swobody. Obliczyć następujące prawdopodobieństwa:

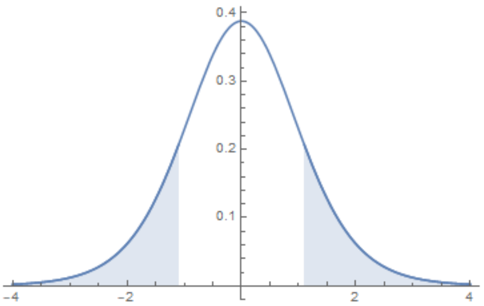
a) $P(|X| > 1{,}1)$
b) $P(|X| \leq 0{,}26)$
c) $P(X \geq 1{,}83)$
d) $P(X < -2{,}8)$

Rozkład t-studenta jest także stablicowany. W książce: (Ostasiewicz (red.), Statystyka elementy teorii i zadania, tablica III, str. 381) mamy prawdopodobieństwa rozkładu postaci: $P\left(\left|T_{df=n=9}\right| \geq t_\alpha\right) = P = \alpha$.

Ad a)

$P(|X = T_9| > 1{,}1)$ odczytujemy bezpośrednio z tablic (dla n=ss=df=9) stopni swobody w wierszu szukamy $t_\alpha = 1{,}1$ i odczytujemy w tabeli w kolumnie odpowiadającą wartośc prawdopodobieństwa, a więc: $P = 0{,}3$.

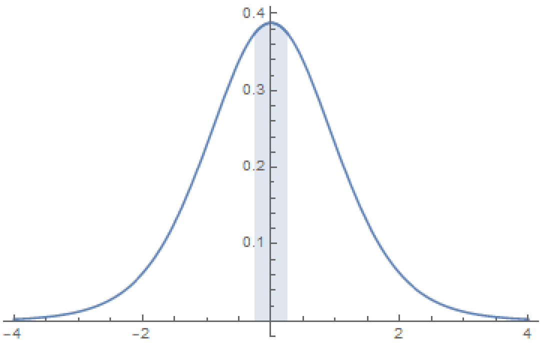
Zatem $P(|X = T_9| > 1{,}1) = 0{,}3$ (30%).



Ad b)

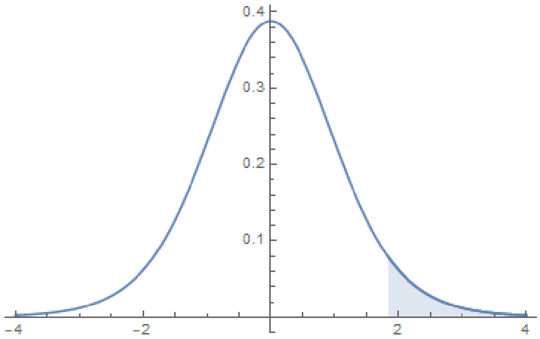
Skorzystamy z prawdopodobieństwa zdarzenia przeciwnego:

$P(|X = T_9| \leq 0{,}26) = 1 - P(|X = T_9| \geq 0{,}26) = 1 - 0{,}8 = 0{,}2$ (20%)



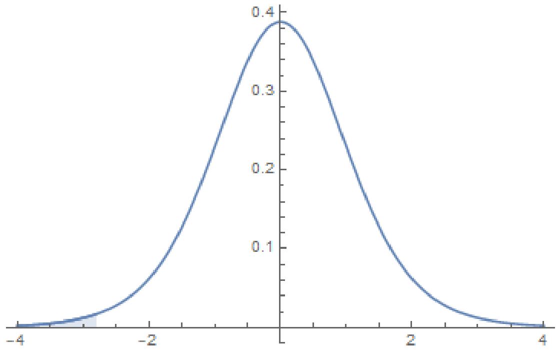
Ad c)

Z racji symetrii rozkładu t-Studenta $P(X \geq 1{,}83) = \frac{P(|X| \geq 1{,}83)}{2} = \frac{0{,}1}{2} = 0{,}05$ (5%)



Ad d)

Z racji symetrii rozkładu t-Studenta $P(X < -2{,}8) = \frac{P(|X| \geq 2{,}8)}{2} = \frac{0{,}02}{2} = 0{,}01$ (1%)



Przykład 5. Zmienna losowa $X = \chi^2_{df=4}$ ma rozkład χ^2- Chi kwadrat z $(df = ss = n = 4)$ stopniami swobody. Obliczyć następujące prawdopodobieństwa:

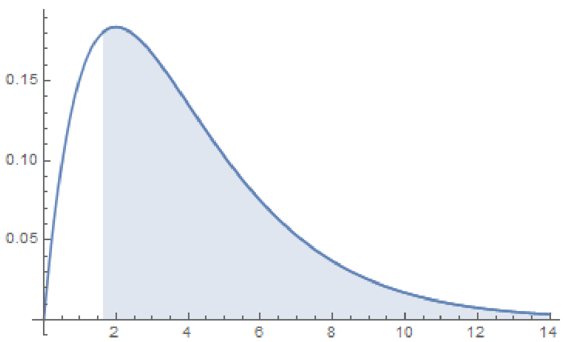
a) $P\left(\chi^2_{df=4} > 1{,}65\right)$
b) $P\left(\chi^2_{df=4} \leq 4{,}88\right)$
c) $P\left(0{,}71 \leq \chi^2_{df=4} \leq 1{,}65\right)$
d) Odczytać taką wartość χ^2_0 dla której $P\left(\chi^2_{df=4} \leq \chi^2_0\right) = 0{,}5$.

W tablicach statystycznych (w książce: Ostasiewicz (red.), Statystyka elementy teorii i zadania, tablica II, str. 380) mamy prawdopodobieństwa rozkładu postaci: $P\left(\chi^2_{df=n} \geq \chi^2_\alpha\right) = P = \alpha$.

Ad a)

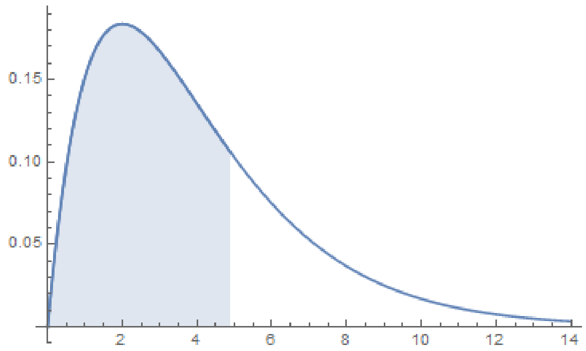
Odczytujemy bezpośrednio z tablic:

$P\left(\chi^2_{df=4} > 1{,}65\right) = 0{,}8$ (80%)



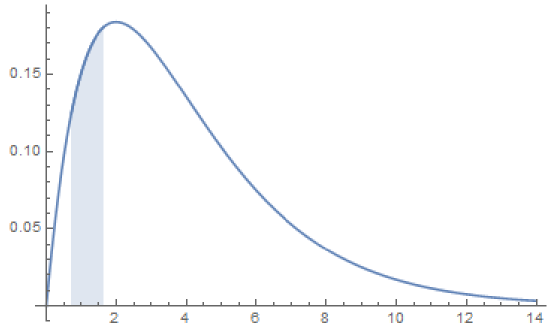
Ad b)

$P\left(\chi^2_{df=4} \leq 4{,}88\right) = 1 - P\left(\chi^2_{df=4} > 4{,}88\right) = 1 - 0{,}3 = 0{,}7$ (70%)



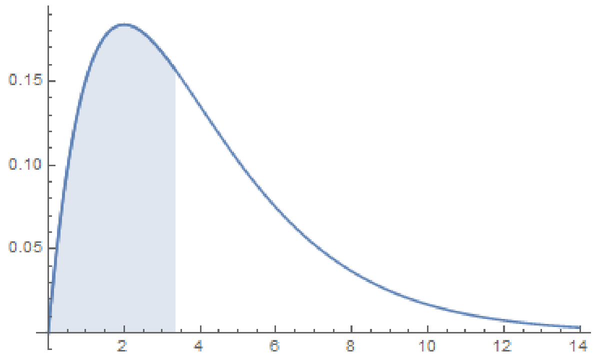
Ad c)

$P\left(0{,}71 \leq \chi^2_{df=4} \leq 1{,}65\right) = P\left(\chi^2_{df=4} > 0{,}71\right) - P\left(\chi^2_{df=4} > 1{,}65\right) = 0{,}95 - 0{,}8 = 0{,}15$ (15%)



Ad d)

$P\left(\chi^2_{df=4} \leq \chi^2_0\right) = 0{,}5 = 1 - P\left(\chi^2_{df=4} > \chi^2_0\right)$, zatem $P\left(\chi^2_{df=4} > \chi^2_0\right) = 0{,}5$, stąd $\chi^2_0 = 3{,}357$.



Zadania do samodzielnego rozwiązania (Ostasiewicz S. [red.], Statystyka elementy teorii i zadania, str. 185-193): Zad. 5.24, 5.25, 5.29, 5.46