Laboratoria zestaw 4

Krystian Baran 145000

30 marca 2021

1 Zadanie 1

Zużycie wody (w hektolitrach) w pewnym osiedlu w ciągu dnia ma rozkład $N(m=?, \sigma = 11)$. Obliczyć prawd. zdarzenia, że empiryczna wariancja zużycia wody w losowo wybranych 90 dniach

a) nie będzie większa niż 100[hl],

b) będzie większa niż 200[hl].

Niech X_i będzie zużycie wody w jednym dniu i niech m będzie parametr m rozkładu normalnego. Oznaczmy $\overline{X}$, jako rozkład średniej zużycia wody w 90 dniach; wtedy $\overline{X}$ będzie także rozkładem normalny następującym:

$$\overline{X} \sim N\Big(m, \frac{11}{\sqrt{90}}\Big) = N\Big(m, \frac{11}{3\sqrt{10}}\Big)$$

Oznaczmy Y, jako $X_i - \overline{X}$; wtedy Y ma rozkład normalny:

$$Y \sim N\Big(m+m, \sqrt{\frac{121}{90} + 121}\Big) = N\Big(2m, \sqrt{\frac{91}{90}}11\Big)$$

Korzystając z twierdzenia mówiącego że, jeżeli $X_i, i = 1, 2, \ldots n$ są zmiennymi losowymi z rozkładem normalnym z parametrami $\mu = 0$ i $\sigma = 1$, to ich suma ma rozkład χ^2 z n stopniami swobody.

$$X_1 + X_2 + \cdots + X_n = \chi^2_n$$

Wtedy wariancja będzie miała następujący wzór:

$$\begin{aligned} S_n^2 &= \frac{1}{89} \sum_{i=1}^{90} (Y_i)^2 \\ &= \frac{1}{89} \sum_{i=1}^{90} \Big(\frac{Y_i \cdot \sqrt{90}}{\sqrt{91} \cdot 11}\Big)^2 \cdot \frac{91 \cdot 11^2}{90} \\ &= \frac{91 \cdot 121}{90 \cdot 89} \chi^2_{90} \end{aligned}$$

a)

Można wtedy obliczyć prawdopodobieństwo, że wariancja będzie mniejsza niż 100 [hl], jako:

$$\begin{aligned} P(S_n^2 < 100) &= P\Big(\frac{91 \cdot 121}{90 \cdot 89} \chi^2_{90} < 100\Big) \\ &= P\Big(\chi^2_{90} < \frac{9000 \cdot 89}{91 \cdot 121}\Big) \\ &\overset{R}{=} pchisq(9000 * 121/(91 * 121), 90) \approx 0.7555559 \end{aligned}$$

Zatem prawdopodobieństwo, że wariancja będzie mniejsza niż 100 [hl] wynosi 0.7556.

b)

Prawdopodobieństwo, że wariancja będzie większa niż 200[hl] wyraża się następująco:

$$
\begin{aligned}
P(S_n^2 > 200) &= 1 - P\Big(\frac{91 \cdot 121}{90 \cdot 89}\chi^2_{90} < 200\Big) \\
&= 1 - P\Big(\chi^2_{90} < \frac{18000 \cdot 89}{91 \cdot 121}\Big) \\
&\overset{R}{=} 1 - pchisq(18000 * 121/(91 * 121), 90) \approx 4.589235e - 10
\end{aligned}
$$

Jest to liczba bardzo bliska zeru, zatem jest mało prawdopodobne, że wariancja będzie większa niż 200 [hl].

2 Zadanie 2

Wiadomo, że błąd pomiaru pewnego przyrządu ma rozkład normalny $N(0, \sigma)$ i z prawd. 0,95 nie wychodzi poza prze-dział (-1,1). Dokonanych zostanie i) 10, ii) 100 niezależ-nych pomiarów tym przyrządem. Oblicz prawd. zdarzenia, że wariancja pomiarów

a) przyjmie wartość między 0,2 a 0,3,

b) będzie większa od 0,28.

Niech $X_i \sim N(0, \sigma)$ będzie zmienną losową opisującą błąd jednego pomiaru. Wiedząc, że $P(-1 < X_i < 1) = 0.95$ błąd pomiaru mieści się w przedziale (-1,1) z prawdopodobieństwem 0.95, można wyznaczyć parametr σ.

$$\begin{aligned} P(-1 < X_i < 1) &= P\Big(-\frac{1}{\sigma} < \frac{X_i}{\sigma} < \frac{1}{\sigma}\Big) \\ &= \Phi\Big(\frac{1}{\sigma}\Big) - \Phi\Big(-\frac{1}{\sigma}\Big) \\ &= 2 \cdot \Phi\Big(\frac{1}{\sigma}\Big) - 1 = 0.95 \end{aligned}$$

$$\begin{aligned} 2 \cdot \Phi\Big(\frac{1}{\sigma}\Big) &= 1.95 \\ \Phi\Big(\frac{1}{\sigma}\Big) &= 0.975 \\ \frac{1}{\sigma} &= \Phi^{-1}(0.975) \\ \sigma &= \frac{1}{\Phi^{-1}(0.975)} \overset{R}{=} \frac{1}{qnorm(0.975, 0, 1)} \\ &\approx 0.5102135 \end{aligned}$$

a)

Wariancja w próby wyznacza się następującym wzorem:

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

Zatem skorzystamy ze wzoru na rozkład średniej:

$$\overline{X}_n \sim N\Big(\mu, \frac{\sigma}{\sqrt{n}}\Big)$$

Wprowadzamy nową zmienna $Y = X_i - \overline{X}_n$, taka zmienna będzie miała także rozkład normalny

$$\sim N\Big(0 + 0, \sqrt{\frac{\sigma^2}{n} + \sigma^2}\Big) = N\Big(0, \sqrt{\frac{n+1}{n}}\sigma\Big)$$

Korzystając z twierdzenia mówiącego że, jeżeli $X_i, i = 1, 2, \dots n$ są zmiennymi losowymi z rozkładem normalnym z parametrami $\mu = 0$ i $\sigma = 1$, to ich suma ma rozkład χ^2 z n stopniami swobody.

$$X_1 + X_2 + \cdots + X_n = \chi^2_n$$

Wtedy wariancja z próby będzie miała następujący wzór:

$$\begin{aligned}
S_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\Big(\frac{Y_i \cdot \sqrt{n}}{\sqrt{n+1}\sigma}\Big)^2 \cdot \frac{(n+1)\sigma^2}{n} \\
&= \frac{(n+1)\sigma^2}{n(n-1)}\chi^2_n
\end{aligned}$$

Następnie można obliczyć szukane prawdopodobieństwo

$$\begin{aligned}
P(0.2 < S_n^2 < 0.3) &= P\Big(0.2 < \frac{(n+1)\sigma^2}{n(n-1)}\chi^2_n < 0.3\Big) \\
&= P\Big(0.2\frac{n(n-1)}{(n+1)\sigma^2} < \chi^2_n < 0.3\frac{n(n-1)}{(n+1)\sigma^2}\Big) \\
&\overset{R}{=} pchisq\Big(0.3\frac{n(n-1)}{(n+1)\sigma^2}, n\Big) - pchisq\Big(0.2\frac{n(n-1)}{(n+1)\sigma^2}, n\Big)
\end{aligned}$$

Dla $n = 10$:

$$\begin{aligned}
&= pchisq\Big(0.3\frac{90}{11\sigma^2}, 10\Big) - pchisq\Big(0.2\frac{90}{11\sigma^2}, 10\Big) \\
&= pchisq(9.429035, 10) - pchisq(6.286024, 10) \\
&\approx 0.2987605
\end{aligned}$$

Dla $n = 100$:

$$\begin{aligned}
&= pchisq\Big(0.3\frac{9900}{101\sigma^2}, 100\Big) - pchisq\Big(0.2\frac{9900}{101\sigma^2}, 100\Big) \\
&= pchisq(112.9617, 100) - pchisq(75.30781, 100) \\
&\approx 0.7918889
\end{aligned}$$

Zatem prawdopodobieństwo, że wariancja będzie w przedziale $(0.2, 0.3)$ wynosi:

i) 0.2988

ii) 0.7919

b)

Podobnie jak w podpunkcie **a** obliczymy prawdopodobieństwo:

$$
\begin{aligned}
P(S_n^2 > 0.28) &= 1 - P(S_n^2 < 0.28) \\
&= 1 - P\Big(\frac{(n+1)\sigma^2}{n(n-1)}\chi_n^2 < 0.28\Big) \\
&= 1 - P\Big(\chi_n^2 < 0.28\frac{n(n-1)}{(n+1)\sigma^2}\Big) \\
&\overset{R}{=} 1 - pchisq\Big(0.28\frac{n(n-1)}{(n+1)\sigma^2}, n\Big)
\end{aligned}
$$

Dla $n = 10$:

$$
\begin{aligned}
&= 1 - pchisq\Big(0.28\frac{90}{11\sigma^2}, 10\Big) \\
&= 1 - pchisq(8.800433, 10) \\
&\approx 0.5511423
\end{aligned}
$$

Dla $n = 100$:

$$
\begin{aligned}
&= 1 - pchisq\Big(0.28\frac{9900}{101\sigma^2}, 100\Big) \\
&= 1 - pchisq(105.4309, 100) \\
&\approx 0.3356967
\end{aligned}
$$

Zatem prawdopodobieństwo, że wariancja będzie większa niż 0.28 wynosi:

i) 0.5511

ii) 0.3357

3 Zadanie 3

Losujemy 100 liczb według rozkładu jednostajnego na przedziale (0, 1).

a) Ustalić rozkład sumy tych liczb.

b) Obliczyć prawd. zdarzenia, że suma wylosowanych liczb nie będzie należała do przedziału (45, 55).

c) Wyznaczyć dystrybuantę największej z wylosowanych liczb i oblicz prawd., że liczba ta będzie mniejsza od 0,95.

d) Jaki wniosek należy wyciągnąć, jeśli że suma wylosowanych liczb będzie mniejsza niż 40?

a)

Niech X_i będzie wylosowana i-ta liczba. $X_i \sim U(0,1)$ jest rozkładem jednostajnym na przedziale $(0,1)$ z następującą funkcją gęstości:

$$f_{X_i}(x) = \begin{cases} 0 & , x \leqslant 0 \vee x \geqslant 1 \\ 1 & , 0 < x < 1 \end{cases}$$

Wtedy dystrybuanta przyjmuję następujący wzór:

$$F_{X_i}(t) = \begin{cases} 0 & , t \leqslant 0 \\ t & , 0 < t < 1 \\ 1 & , t \geqslant 1 \end{cases}$$

Korzystając z centralnego twierdzenia granicznego dla sumy, przyjmując, że 100 jest wystarczającą dużą liczbą, rozkład sumy przyjmuje rozkład normalny z następującymi parametrami.

$$X = \sum_{i=1}^{100} X_i \sim N(n \cdot \mathbb{E}X, \sqrt{n} \cdot \mathbb{D}X)$$

Gdzie $\mathbb{E}X$ jest wartością oczekiwaną, czyli $\frac{b-a}{2} = \frac{1}{2}$,
a $\mathbb{D}X$ jest odchyleniem standardowym czyli $\sqrt{\mathbb{D}^2(X)} = \sqrt{\frac{(b-a)^2}{12}} = \frac{1}{\sqrt{12}}$.
Zatem $X \sim N\left(\frac{100}{2}, \sqrt{\frac{100}{12}}\right) = N\left(50, \frac{5}{\sqrt{3}}\right)$.

b)

Znając rozkład sumy można obliczyć prawdopodobieństwo że suma nie będzie w przedziale $(45, 55)$, $P(X < 45 \vee X > 55)$, przechodząc na prawdopodobieństwo przeciwne:

$$
\begin{aligned}
P(X < 45 \vee X > 55) &= 1 - P(45 < X < 55) \\
&\overset{R}{=} 1 - (pnorm(55, 50, 5/sqrt(3)) - pnorm(45, 50, 5/sqrt(3))) \\
&\approx 0.08326452
\end{aligned}
$$

Zatem prawdopodobieństwo, że suma nie będzie w danym przedziale wynosi 0.0834.

c)

Aby wyznaczyć dystrybuantę liczby maksymalnej skorzystamy z definicji dystrybuanty; niech $Y = \sum_{i=1} 100X_i$.

$$F_Y(t) = P(Y \leqslant t) = \prod_{i=1}^{100} P(X_i \leqslant t)$$

Ponieważ losowanie liczb jest niezależne to $X_i \cap X_j = 0, i \neq j$, wtedy:

$$\prod_{i=1}^{100} P(X_i \leqslant t) = P(X_i \leqslant t)^{100} = \begin{cases} 0 & , t \leqslant 0 \\ t^{100} & , 0 < t < 1 \\ 1 & , t \geqslant 1 \end{cases}$$

Mając już dystrybuantę można wyliczyć $P(Y < 0.95)$:

$$P(Y < 0.95) = F_Y(0.95) = 0.95^{100} \approx 0.0059205292$$

Zatem prawdopodobieństwo, że maksymalna liczba będzie mniejsza niż 0.95 wynosi 0.006.

d)

Jeżeli suma z wylosowanych liczby będzie z dużym prawdopodobieństwem mniejsza od 40, oznacza to, że wartość oczekiwana jest podwyższona lub że stosowane przybliżenie jest błędne.

$$P(X < 40) \overset{R}{=} pnorm(40, 50, 5/sqrt(3)) \approx 0.0002660028$$

Zatem nie jest mocno prawdopodobne, że wartość oczekiwana jest błędna o ± 10.

4 Zadanie 4

Zobacz plik z wykładów.

5 Zadanie 5

Ze zbioru $\{1, 2, 3, 4\}$ wylosowano dwie liczby

a) ze zwracaniem,

b) bez zwracania,

Wyznaczyć rozkład oraz wartość oczekiwaną i wariancję rozstępu

a)

Niech X_1, X_2 będą wylosowane liczby ze zbioru. Ponieważ losowane są ze zwracaniem możliwości będzie 4^2, czyli 16.
Oznaczmy $A = (X_1, X_2)$, $\Omega = 16$, R rozstęp, i niech wylosowane liczby posortujemy oznaczając je posortowane, jako $X_{(1)}$ i $X_{(2)}$ wtedy można sporządzić tabele w następujący sposób:

A	$P(X_1, X_2)$	$X_{(1)}$	$X_{(2)}$	R
(1,1)	1/16	1	1	0
(1,2)	1/16	1	2	1
(1,3)	1/16	1	3	2
(1,4)	1/16	1	4	3
(2,1)	1/16	1	2	1
(2,2)	1/16	2	2	0
(2,3)	1/16	2	3	1
(2,4)	1/16	2	4	2
(3,1)	1/16	1	3	2
(3,2)	1/16	2	3	1
(3,3)	1/16	3	3	0
(3,4)	1/16	3	4	1
(4,1)	1/16	1	4	3
(4,2)	1/16	2	4	2
(4,3)	1/16	3	4	1
(4,4)	1/16	4	4	0

Za pomocą tej tabeli można wyznaczyć rozkład łączy i rozkłady brzegowe (widoczne w ostatniej kolumnie i w ostatnim wierszu):

$x_1 \backslash x_2$	1	2	3	4	$f_{X_{(1)}}(x_1)$
1	1/16	2/16	2/16	2/16	7/16
2	0	1/16	2/16	2/16	5/16
3	0	0	1/16	2/16	3/16
4	0	0	0	1/16	1/16
$f_{X_{(2)}}(x_2)$	1/16	3/16	5/16	7/16	16/16

Wyznaczymy teraz wartość oczekiwaną i wariancje wraz z kowariancją.

$$\begin{aligned}\mathbb{E}(X_{(1)}) &= \sum_{i=1}^{4} x_i \cdot f_{X_{(1)}}(x_i) \\ &= 1\frac{7}{16} + 2\frac{5}{16} + 3\frac{3}{16} + 4\frac{1}{16} \\ &= \frac{30}{16} = \frac{15}{8}\end{aligned}$$

$$\begin{aligned}\mathbb{E}(X_{(1)}^2) &= \sum_{i=1}^{4} x_i^2 \cdot f_{X_{(1)}}(x_i) \\ &= 1\frac{7}{16} + 4\frac{5}{16} + 9\frac{3}{16} + 16\frac{1}{16} \\ &= \frac{70}{16} = \frac{35}{8}\end{aligned}$$

$$\begin{aligned}\mathbb{D}^2(X_{(1)}) &= \mathbb{E}(X_{(1)}^2) - \mathbb{E}(X_{(1)})^2 \\ &= \frac{35}{8} - \frac{225}{64} = \frac{55}{64}\end{aligned}$$

$$\begin{aligned}\mathbb{E}(X_{(2)}) &= \sum_{i=1}^{4} x_i \cdot f_{X_{(2)}}(x_i) \\ &= 1\frac{1}{16} + 2\frac{3}{16} + 3\frac{5}{16} + 4\frac{7}{16} \\ &= \frac{50}{16} = \frac{25}{8}\end{aligned}$$

$$\begin{aligned}\mathbb{E}(X_{(2)}^2) &= \sum_{i=1}^{4} x_i^2 \cdot f_{X_{(2)}}(x_i) \\ &= 1\frac{1}{16} + 4\frac{3}{16} + 9\frac{5}{16} + 16\frac{7}{16} \\ &= \frac{170}{16} = \frac{85}{8}\end{aligned}$$

$$\begin{aligned}\mathbb{D}^2(X_{(2)}) &= \mathbb{E}(X_{(2)}^2) - \mathbb{E}(X_{(2)})^2 \\ &= \frac{85}{8} - \frac{625}{64} = \frac{55}{64}\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}(X_{(1)}X_{(2)}) &= \sum_{i=1}^{4}\sum_{j=1}^{4} x_i \cdot x_j \cdot f_X(x_i, x_j) \\
&= \Big(1\frac{1}{16} + 2\frac{2}{16} + 3\frac{2}{16} + 4\frac{2}{16}\Big) + 2\Big(2\frac{1}{16} + 3\frac{2}{16} + 4\frac{2}{16}\Big) + \\
&+ 3\Big(3\frac{1}{16} + 4\frac{2}{16}\Big) + \Big(4\frac{1}{16}\Big) \\
&= \frac{19}{16} + \frac{32}{16} + \frac{33}{16} + \frac{4}{16} \\
&= \frac{88}{16} = \frac{11}{2}
\end{aligned}
$$

$$
\begin{aligned}
Cov(X_{(1)}, X_{(2)}) &= \mathbb{E}(X_{(1)}X_{(2)}) - \mathbb{E}(X_{(1)}) \cdot \mathbb{E}(X_{(2)}) \\
&= \frac{11}{2} - \frac{15}{8} \cdot \frac{25}{8} \\
&= \frac{352 - 375}{64} = -\frac{33}{64}
\end{aligned}
$$

Rozstęp pomiędzy wylosowanymi liczbami posiada rozkład następujący, jeżeli w pierwszej tabeli odejmiemy wartość $X_{(2)}$ od wartości $X_{(1)}$:

R_i	0	1	2	3
$P(R_i)$	4/16	6/16	4/16	2/16

Wtedy wartość oczekiwana i wariancja rozstępu wynoszą:

$$
\begin{aligned}
\mathbb{E}R &= \sum_{i=0}^{3} R_i \cdot P(R_i) \\
&= 0 + \frac{6}{16} + 2\frac{4}{16} + 3\frac{2}{16} \\
&= \frac{20}{16} = \frac{5}{4}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(R^2) &= \sum_{i=0}^{3} R_i^2 \cdot P(R_i) \\
&= 0 + \frac{6}{16} + 4\frac{4}{16} + 9\frac{2}{16} \\
&= \frac{40}{16} = \frac{5}{2}
\end{aligned}
$$

$$\mathbb{D}^2(R) = \mathbb{E}(R^2) - \mathbb{E}R^2$$
$$= \frac{5}{2} - \frac{25}{16} = \frac{40-25}{16}$$
$$= \frac{15}{16}$$

b)

Podobnie jak poprzednio oznaczymy X_1, X_2, jako wylosowane liczby ze zbioru, $A = (X_1, X_2)$, $\Omega = 12$, R rozstęp, i niech wylosowane liczby posortujemy oznaczając je posortowane, jako $X_{(1)}$ i $X_{(2)}$ wtedy można sporządzić tabele w następujący sposób:

A	$P(X_1, X_2)$	$X_{(1)}$	$X_{(2)}$	R
(1,2)	1/12	1	2	1
(1,3)	1/12	1	3	2
(1,4)	1/12	1	4	3
(2,1)	1/12	1	2	1
(2,3)	1/12	2	3	1
(2,4)	1/12	2	4	2
(3,1)	1/12	1	3	2
(3,2)	1/12	2	3	1
(3,4)	1/12	3	4	1
(4,1)	1/12	1	4	3
(4,2)	1/12	2	4	2
(4,3)	1/12	3	4	1

Wtedy rozkład będzie następujący:

$x_1 \backslash x_2$	2	3	4	$f_{X_{(1)}}(x_1)$
1	2/12	2/12	2/12	6/12
2	0	2/16	2/16	4/16
3	0	0	2/16	2/16
$f_{X_{(2)}}(x_2)$	1/12	4/12	6/12	12/12

Rozstęp pomiędzy wylosowanymi liczbami posiada rozkład następujący, jeżeli w pierwszej tabeli odejmiemy wartość $X_{(2)}$ od wartości $X_{(1)}$:

R_i	1	2	3
$P(R_i)$	6/12	4/12	2/12

Wtedy wartość oczekiwana i wariancja rozstępu wynoszą:

$$
\begin{aligned}
\mathbb{E}R &= \sum_{i=1}^{3} R_i \cdot P(R_i) \\
&= \frac{6}{12} + 2\frac{4}{12} + 3\frac{2}{12} \\
&= \frac{20}{12} = \frac{5}{3}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(R^2) &= \sum_{i=1}^{3} R_i^2 \cdot P(R_i) \\
&= \frac{6}{12} + 4\frac{4}{12} + 9\frac{2}{12} \\
&= \frac{40}{12} = \frac{10}{3}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{D}^2(R) &= \mathbb{E}(R^2) - \mathbb{E}R^2 \\
&= \frac{10}{3} - \frac{25}{9} = \frac{30-25}{9} \\
&= \frac{5}{9}
\end{aligned}
$$