Analiza danych
przykładowe zadania z rozkładów prawdopodobieństwa

Marek Kubiak

1 Prawdopodobieństwa zdarzeń w standaryzowanym rozkładzie normalnym

1.1 Zadanie 1.

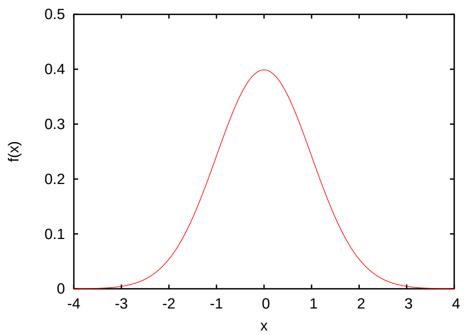
1.1.1 Treść zadania

Zmienna losowa X ma standaryzowany rozkład normalny, czyli $X \sim N(0, 1)$. Obliczyć prawdopodobieństwo: $P(X > 1, 5)$.

1.1.2 Wzorcowe rozwiązanie

Etapy rozwiązania:

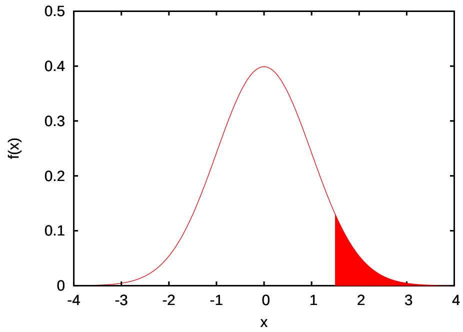
1. Pokazać funkcję gęstości rozkładu $N(0, 1)$ (zob. rysunek 1).



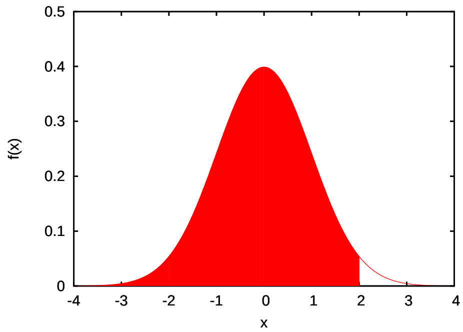
Rysunek 1: Funkcja gęstości standaryzowanego rozkładu normalnego $N(0, 1)$.

2. Pokazać na wykresie funkcji gęstości szukane prawdopodobieństwo (zob. rysunek 2). Jest to po prostu odpowiednie pole pod wykresem gęstości.

3. Stwierdzić, że prawdopodobieństwa dla takiego rozkładu można znaleźć w tablicach, ale tylko dla prawdopodobieństw postaci: $P(X < x) = \Phi(x)$ i dla $x > 0$. Pokazać takie przykładowe prawdopodobieństwo na wykresie na tablicy (zob. rysunek 3).



Rysunek 2: Prawdopodobieństwo zdarzenia $X > 1,5$ na wykresie funkcji gęstości rozkładu $N(0,1)$.



Rysunek 3: Prawdopodobieństwo zdarzenia $X < 2$, które można znaleźć w tablicy jako $\Phi(x)$.

4. Przekształcić zadane prawdopodobieństwo tak, by można było skorzystać z tablic.

   Przekształcenie:

   $$P(X > 1,5) = 1 - P(X \leq 1,5) = 1 - P(X < 1,5).$$

   wykorzystuje dwa fakty:

   - własność zdarzeń odwrotnych,
   - własność rozkładu ciągłego (jeden punkt nie ma znaczenia dla prawdopodobieństwa).

   Teraz można już zamienić prawdopodobieństwo na wartość dystrybuanty:

   $$1 - P(X < 1,5) = 1 - \Phi(1,5).$$

5. Wartość dystrybuanty odczytujemy z tablic:

   $$\Phi(1,5) = 0,9332.$$

   Uzyskujemy wynik:

   $$P(X > 1,5) = 1 - 0,9332 = 0,0668.$$

1.2 Zadanie 2.

1.2.1 Treść zadania

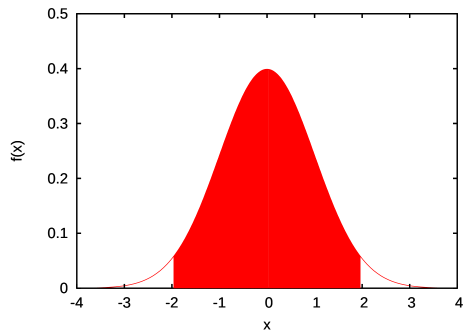
Zmienna losowa X ma standaryzowany rozkład normalny, czyli $X \sim N(0,1)$. Obliczyć prawdopodobieństwo: $P(|X| < 1,96)$.

1.2.2 Wzorcowe rozwiązanie

Etapy rozwiązania:

1. Pokazać funkcję gęstości rozkładu $N(0,1)$ (jak na rysunku 1).

2. Pokazać na wykresie funkcji gęstości szukane prawdopodobieństwo (zob. rysunek 4). Jest to po prostu odpowiednie pole pod wykresem gęstości.



Rysunek 4: Prawdopodobieństwo zdarzenia $|X| < 1,96$ na wykresie funkcji gęstości rozkładu $N(0,1)$.

3. Przekształcić zadane prawdopodobieństwo tak, żeby można było skorzystać z tablic.

   Przekształcamy najpierw wartość bezwzględną:

   $$P(|X| < 1,96) = P(X < 1,96) - P(X < -1,96)$$

   Możliwość tego przekształcenia uzasadniona jest rysunkiem. Korzystamy z tej własności prawdopodobieństwa, że jeżeli jedno zdarzenie $(X < -1,96)$ zawiera się w drugim $(X < 1,96)$, to możemy odejmować prawdopodobieństwa od siebie (zob. rysunek 5).

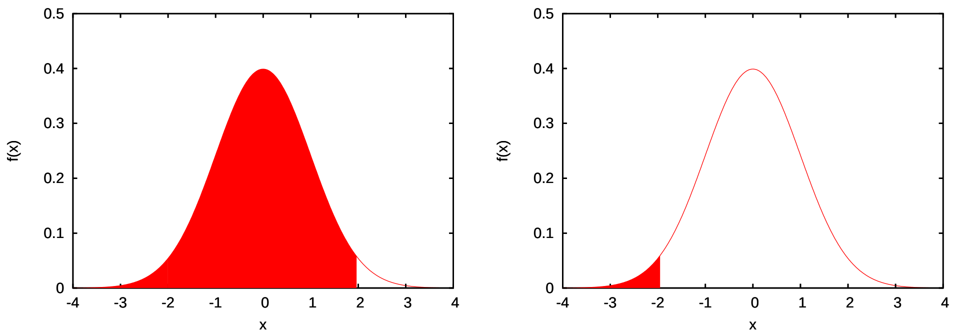
   Pierwsze prawdop. jest już w tablicach – musimy zająć się drugim. Możemy skorzystać z tego, że standaryzowany rozkład normalny jest symetryczny względem 0 i pola $P(X < -1,96)$ i $P(X > 1,96)$ są sobie równe (zob. rysunek 6).

   Otrzymujemy:

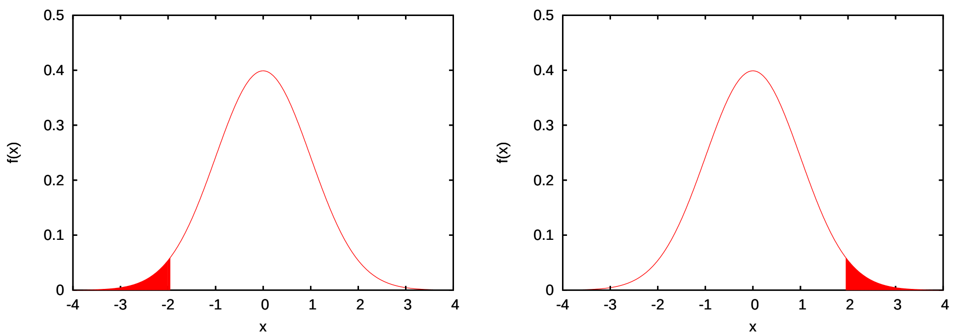
   $$P(X < 1,96) - P(X < -1,96) = P(X < 1,96) - P(X > 1,96)$$

   Następnie korzystamy z własności prawdopodob. zdarzeń odwrotnych i możemy już skorzystać ze stablicowanej dystrybuanty rozkładu normalnego:

   $$P(X < 1,96) - P(X > 1,96) = P(X < 1,96) - [1 - P(X < 1,96)] =$$



Rysunek 5: Zdarzenie $X < 1,96$ (po lewej) zawiera w sobie zdarzenie $X < -1,96$ (po prawej).



Rysunek 6: Zdarzenia $X < -1,96$ (po lewej) i $X > 1,96$ (po prawej) mają takie same prawdopodobieństwa (pola są równe).

$$= \Phi(1,96) - [1 - \Phi(1,96)] = 2 \cdot \Phi(1,96) - 1.$$

4. Wartość dystrybuanty odczytujemy z tablic:

   $$\Phi(1,96) = 0,975$$

   Ostatecznie otrzymujemy wynik:

   $$P(|X| < 1,96) = 2 \cdot 0,975 - 1 = 1,95 - 1 = 0,95.$$

Komentarz do zadania:
Prawdopodobieństwo odchylenia od wartości średniej o *mniej* niż 1,96 sigma wynosi 95%. Czyli odchylenie o *więcej* (w dowolną stronę) jest już raczej mało prawdopodobne (tylko 5%).

2 Prawdopodobieństwa zdarzeń w dowolnym rozkładzie normalnym

2.1 Zadanie 1.

2.1.1 Treść zadania

Zmienna losowa X ma rozkład $N(4,9)$. Obliczyć prawdopodobieństwo: $P(|X-4| < 14)$.

2.1.2 Wzorcowe rozwiązanie

Etapy rozwiązania:

1. Uproszczenie (rozwikłanie) zapisu zdarzenia losowego.

   Upraszczamy wartość bezwzględną w zdarzeniu $|X-4| < 14$:

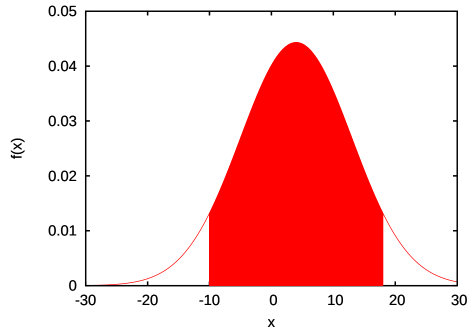
   $$|X-4| = \begin{cases} (X-4) & \text{gdy } X \geq 4 \\ -(X-4) & \text{gdy } X < 4 \end{cases}$$

   Czyli mamy dla $X \geq 4$:

   $$\begin{aligned} |X-4| < 14 \quad \Leftrightarrow \quad & X - 4 < 14 \\ & X < 18 \end{aligned}$$

   A dla $X < 4$:

   $$\begin{aligned} |X-4| < 14 \quad \Leftrightarrow \quad & -(X-4) < 14 \\ & -X + 4 < 14 \\ & -X < 10 \\ & X > -10 \end{aligned}$$



Rysunek 7: Obliczane prawdopodobieństwo na wykresie funkcji gęstości rozkładu $N(4, 9)$.

Ostatecznie otrzymujemy:

$$|X - 4| < 14 \Leftrightarrow (X > -10 \wedge X < 18)$$

i mamy do obliczenia prawdopodobieństwo tego ostatniego, prostszego, równoważnego zdarzenia, które geometrycznie przedstawia rysunek 7

Poszukiwane prawdopodobieństwo można jeszcze uprościć (patrz rysunek 7):

$$P(X > -10 \wedge X < 18) = P(X < 18) - P(X < -10).$$

2. Standaryzacja zmiennej losowej X.

   Jako że zmienna X ma rozkład normalny $N(4, 9)$, to nie można z tablic odczytać wartości jej dystrybuanty – trzeba tę zmienną ustandaryzować, wykonując przekształcenie:

   $$Y = \frac{X - E(X)}{D(X)}$$

   $E(X) = 4$, $D(X) = 9$, więc:

   $$Y = \frac{X - 4}{9} \sim N(0, 1)$$

3. Przetłumaczenie zdarzeń zmiennej X w zdarzenia ustandaryzowanej zmiennej Y.

   Każde z prawdopodobieństw do obliczenia rozważamy oddzielnie.

   Musimy wyznaczyć wartość a zgodną z równaniem:

   $$P(X < 18) = P(Y < a)$$

   W opisie zdarzenia dla zmiennej X kolejno obustronnie odejmujemy $E(X) = 4$ i dzielimy wynik przez $D(X) = 9$:

$$\begin{aligned} P(X < 18) &= P(X - 4 < 18 - 4) = \\ &= P(\frac{X-4}{9} < \frac{18-4}{9}) = \\ &= P(Y < \frac{14}{9}) = \\ &= P(Y < 1,56). \end{aligned}$$

Czyli $a = 1,56$.

Drugie prawdopodobieństwo traktujemy podobnie: musimy wyznaczyć wartość b zgodną z równaniem:

$$P(X < -10) = P(Y < b)$$

W opisie zdarzenia dla zmiennej X kolejno obustronnie odejmujemy $E(X) = 4$ i dzielimy wynik przez $D(X) = 9$:

$$\begin{aligned} P(X < -10) &= P(X - 4 < -10 - 4) = \\ &= P(\frac{X-4}{9} < \frac{-10-4}{9}) = \\ &= P(Y < \frac{-14}{9}) = \\ &= P(Y < -1,56). \end{aligned}$$

Czyli $b = -1,56$.

4. Wykorzystanie ustandaryzowanej zmiennej Y do obliczenia prawdopodobieństwa.

   Dzięki standaryzacji zmiennej X i przetłumaczeniu zdarzeń mamy do obliczenia równoważne prawdopodobieństwo dla zmiennej losowej $Y \sim N(0,1)$:

   $$P(|X - 4| < 14) = P(Y < 1,56) - P(Y < -1,56).$$

   Podobnie jak w zadaniach z punktu 1 przekształcamy prawdopodobieństwa do dystrybuanty $\Phi(y)$:

   $$\begin{aligned} P(|X - 4| < 14) &= P(Y < 1,56) - P(Y < -1,56) = \\ &= P(Y < 1,56) - P(Y > 1,56) = \\ &= P(Y < 1,56) - [1 - P(Y < 1,56)] = \\ &= P(Y < 1,56) - 1 + P(Y < 1,56) = \\ &= 2 \cdot P(Y < 1,56) - 1 = \\ &= 2 \cdot \Phi(1,56) - 1 = \\ \text{(z tablic)} &= 2 \cdot 0,9406 - 1 = 0,8812. \end{aligned}$$

Komentarz do zadania:

- Widać, że: $P(|X-4| < 14) = P(|Y| < 1,56)$. Przedział dla zmiennej Y jest symetryczny względem 0 dlatego, że przedział dla zmiennej X $(-10, 18)$ jest symetryczny względem wartości oczekiwanej $E(X) = 4$.

- Obliczane prawdopodobieństwo można inaczej zapisać: $P(|X - E(X)| < 1,56 \cdot D(X))$. Jak widać prawdopodob. odchylenia wartości zmiennej losowej normalnej od wartości oczekiwanej o mniej niż 1,56 odchylenia standardowego jest duże. Nadal jest jednak zbyt małe, by odchylenie o *więcej* niż 1,56 sigma było pomijalnie małe (por. reguła trzysigmowa).

- W ostatnim etapie tego zadania, pod dojściu już do określenia zdarzenia losowego dla zmiennej ustandaryzowanej, niejako *wywołujemy podprogram* obliczania prawdopodobieństwa w rozkładzie normalnym ustandaryzowanym – dokładnie tak, jak w punkcie 1.

2.2 Zadanie 2.

2.2.1 Treść zadania

Zmienna losowa X ma rozkład $N(105; 0,4)$. Obliczyć prawdopodobieństwo: $P(X < 106 \wedge X > 103,7)$.

2.2.2 Wzorcowe rozwiązanie (wersja skompresowana)

Konieczna jest standaryzacja zmiennej X przez przekształcenie:

$$Y = \frac{X - E(X)}{D(X)} = \frac{X - 105}{0,4}$$

Wykorzystujemy to przekształcenie w określeniu zdarzenia, by przejść na zmienną Y i dla niej obliczyć prawdopodobieństwo – korzystając także z tablic:

$$\begin{aligned}
P(X < 106 \wedge X > 103,7) &= P(\frac{X-105}{0,4} < \frac{106-105}{0,4} \wedge \frac{X-105}{0,4} > \frac{103,7-105}{0,4}) = \\
&= P(Y < 2,5 \wedge Y > -3,25) = \\
&= P(Y < 2,5) - P(Y \leq -3,25) = \\
&= \Phi(2,5) - \Phi(-3,25) = \\
&= \Phi(2,5) - 1 + \Phi(3,25) = \\
&= 0,9938 - 1 + 0,9994 = \\
&= 0,9932 \approx 99,3\%
\end{aligned}$$