

Zadania z Analizy danych

zestaw 3

Zadanie 1

Pokaż, że rozkład Lorentza $g(x) = \frac{2}{\pi\Gamma} \frac{\Gamma^2}{4(x-a)^2 + \Gamma^2}$ jest zawsze unormowany. Sprawdź, że szerokość połówkowa tego rozkładu jest równa parametrowi Γ , a wartość oczekiwana, wartość modalna i mediana są równe parametrowi a .

Rozwiązanie

Sprawdzamy warunek normalizacyjny.

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{2}{\pi\Gamma} \frac{\Gamma^2}{4(x-a)^2 + \Gamma^2} dx &= \frac{2\Gamma}{\pi} \int_{-\infty}^{\infty} \frac{1}{\Gamma^2 \left(\frac{4}{\Gamma^2} (x-a)^2 + 1 \right)} dx = \left(\begin{array}{l} u = \frac{2}{\Gamma} (x-a) \\ dx = (\Gamma/2) du \end{array} \right) \\ &= \frac{2}{\pi\Gamma} \frac{\Gamma}{2} \int_{-\infty}^{\infty} \frac{du}{u^2 + 1} = \frac{1}{\pi} \operatorname{arctg}(x) \Big|_{-\infty}^{\infty} = \frac{1}{\pi} \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1. \end{aligned}$$

Czyli rozkład jest unormowany zawsze niezależnie od wartości Γ, a .

Liczymy dominantę.

$$\frac{dg(x)}{dx} = \frac{2\Gamma^2}{\pi\Gamma} \frac{-8(x-a)}{(4(x-a)^2 + \Gamma^2)^2} = 0 \rightarrow x = a.$$

$$g(a) = \frac{2}{\pi\Gamma}.$$

Liczymy wartość średnią

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} \frac{2}{\pi\Gamma} \frac{x\Gamma^2}{4(x-a)^2 + \Gamma^2} dx = \frac{2\Gamma}{\pi} \int_{-\infty}^{\infty} \frac{(x-a) + a}{\Gamma^2 \left(\frac{4}{\Gamma^2} (x-a)^2 + 1 \right)} dx = \left(\begin{array}{l} u = \frac{2}{\Gamma} (x-a) \\ dx = (\Gamma/2) du \end{array} \right) = \\ &= \frac{2}{\pi\Gamma} \frac{\Gamma}{2} \int_{-\infty}^{\infty} \frac{u}{u^2 + 1} du + \frac{2}{\pi\Gamma} \frac{\Gamma}{2} \int_{-\infty}^{\infty} \frac{a}{u^2 + 1} du = 0 + \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{du}{u^2 + 1} = \frac{a}{\pi} \operatorname{arctg}(x) \Big|_{-\infty}^{\infty} = \frac{a}{\pi} \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = a. \end{aligned}$$

Liczymy szerokość połówkową

$$\frac{2}{\pi\Gamma} \frac{\Gamma^2}{4(x-a)^2 + \Gamma^2} = \frac{1}{2} g(a) = \frac{1}{\pi\Gamma}.$$

Stąd

$$4(x-a)^2 + \Gamma^2 = 2\Gamma^2.$$

$$x_1 = -\frac{\Gamma}{2} + a, \quad x_2 = \frac{\Gamma}{2} + a$$

Ostatecznie szerokość połówkowa jest równa

$$x_2 - x_1 = \Gamma$$

Zadanie 2

Transport bananów jest odrzucany przez sklep, jeśli w 10 losowo wybranych skrzynkach znajdą się co najmniej dwie zepsute banany. Załóżmy, że w transporcie 2% skrzynek bananów ulega zepsuciu. Jakie jest prawdopodobieństwo, że transport zostanie odrzucony?

Wskazówka:

Zastosuj rozkład dwumienny z $n = 10$ i $p = 0,02$. Oblicz prawdopodobieństwo zdarzenia, że w wylosowanej próbie jest mniej niż 2 skrzynki z zepsutymi bananami $P_{10}(k < 2) = P_{10}(0) + P_{10}(1)$, a następnie prawdopodobieństwo zdarzenia przeciwnego. Do wartości wyliczenia symbolu Newtona $\binom{n}{k}$ skorzystaj z funkcji KOMBINACJE lub wprost skorzystaj z funkcji ROZKŁ.DWUM.

Rozwiązanie

$$\begin{aligned}P_n(k) &= \binom{n}{k} p^k (1-p)^{n-k} \\P_{10}(0) &= \binom{10}{0} 0,02^0 (1-0,02)^{10} \approx 0,817, \\P_{10}(1) &= \binom{10}{1} 0,02^1 (1-0,02)^9 \approx 0,167, \\P_{10}(k \geq 2) &= 1 - (P_{10}(0) + P_{10}(1)) \approx 1 - (0,817 + 0,167) = 0,016\end{aligned}$$

Zadanie 3

Jak duża powinna być próba losowa w problemie z zadania 2, aby z prawdopodobieństwem $P(K \geq 1) = 0,1$ znalazła się w niej co najmniej jedna skrzynka z zepsutymi bananami? Powtórz zadanie dla $P(K \geq 1) = 0,99$.

Wskazówka:

Jak w zadaniu 2 stosujemy rozkład dwumienny z $p = 0,02$, ale nieznanym n . Rozważ zdarzenie przeciwne w wylosowanej n -elementowej próbie wszystkie skrzynki zawierają niezepsute banany z prawdopodobieństwem $P_n(K = 0) = 1 - 0,1 = 0,9$ i wylicz dla jakiego n będzie to miało miejsce (wynik trzeba zaokrąglić do najmniejszej liczby naturalnej większej lub równej otrzymanej).

Rozwiązanie

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Dla $P(K \geq 1) = 0,1$

$$\begin{aligned}P_n(0) &= \binom{n}{0} 0,02^0 (1-0,02)^{n-0} = 0,98^n \\P_n(k > 0) &= 1 - P_n(0) = 1 - 0,98^n = 0,1 \\0,98^n &= 0,9 \\n \ln 0,98 &= \ln 0,9 \\n &= \frac{\ln 0,9}{\ln 0,98} \approx 5,216 \rightarrow n = 6\end{aligned}$$

Dla $P(K \geq 1) = 0,99$

$$\begin{aligned}P_n(0) &= \binom{n}{0} 0,02^0 (1-0,02)^{n-0} = 0,98^n \\P_n(k > 0) &= 1 - P_n(0) = 1 - 0,98^n = 0,99 \\0,98^n &= 0,01 \\n \ln 0,98 &= \ln 0,01 \\n &= \frac{\ln 0,01}{\ln 0,98} \approx 227,948 \rightarrow n = 228.\end{aligned}$$

Zadanie 4

W teście jednokrotnego wyboru znajduje się 20 pytań i po cztery odpowiedzi do każdego z nich. Test jest zaliczony, jeśli student odpowie na co najmniej 12 pytań. Jakie jest prawdopodobieństwo zaliczenia testu przy wyborze odpowiedzi na chybił trafił? Jaka jest najbardziej prawdopodobna liczba poprawnych odpowiedzi? Jaka byłaby średnia liczba poprawnych odpowiedzi?

Wskazówka:

Należy zastosować rozkład dwumienny z $n = 20$ i $p = 1/4$, a następnie policzyć prawdopodobieństwo poprawnej odpowiedzi na 12, 13, ... 20 pytań $P_{20}(k \geq 12) = \sum_{k=12}^{20} P_{20}(k)$.

Rozwiązanie

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

k	$P_{20}(k)$
0	0,003171
1	0,021141
2	0,066948
3	0,133896
4	0,189685
5	0,202331
6	0,168609
7	0,112406
8	0,060887
9	0,027061
10	0,009922
11	0,003007
12	0,000752
13	0,000154
14	2,57E-05
15	3,43E-06
16	3,57E-07
17	2,8E-08
18	1,56E-09
19	5,46E-11
20	9,09E-13

$$P_{20}(k \geq 12) = \sum_{k=12}^{20} P_{20}(k) \approx 0,000935.$$
$$\bar{k} = n \cdot p = 20 \cdot \frac{1}{4} = 5.$$

Zadanie 5

Podczas wielkiej kumulacji gry Lotto w dniu 07-05-2016 liczba wygranych poszczególnych stopni była następująca

Stopień	Liczba wygranych
szóstka	3
piątka	644
czwórka	31 128
trójka	571 008

Na podstawie liczby wygranych każdego stopnia oszacuj liczbę zawartych zakładów, zakładając, że wszystkie zakłady były typu prostego (nie było zakładów systemowych typu 7 i więcej skreśleń). Następnie wylicz średnią ważoną n otrzymanych liczb zakładów i jej odchylenie standardowe. Na podstawie wyliczonej wartości n wylicz wartości oczekiwane liczby wygranych poszczególnego stopnia i ich odchylenia standardowe. Porównaj te wartości z liczbami wygranych. Dla oszacowanej liczby zakładów wylicz prawdopodobieństwa $P_n(k)$, że liczba wygranych szóstego stopnia wyniesie $k = 0, 1, 2, 3, 4, 5$. Jakie jest prawdopodobieństwo, że padnie przynajmniej jedna główna wygrana?

Wskazówka

Do każdego ze stopnia wygranych zastosuj rozkład dwumienny z prawdopodobieństwami sukcesu wynoszącymi

Stopień	Prawdopodobieństwo
szóstka	1/13 983 816
piątka	1/54 201
czwórka	1/1032
trójka	1/57

Do wyliczenia prawdopodobieństw $P_n(k)$ skorzystaj z przybliżenia rozkładu dwumianowego rozkładem Poissona.

Rozwiązanie

Niech n oznacza liczbę zawartych zakładów. Wartość oczekiwana rozkładu dwumiennego wynosi

$$E(k) = np.$$

Stąd dostajemy

$$n = \frac{E(k)}{p}.$$

Jako estymatorów wartości oczekiwanych liczby wygranych każdego stopnia bierzemy liczbę wygranych, stąd dostajemy

$$n = \frac{E(k_i)}{p_i} \approx \frac{\bar{k}_i}{p_i} \approx \frac{k_i}{p_i}.$$

Stopień	Liczba wygranych k_i	Oszacowanie liczby zakładów
6	3	41 951 448
5	644	34 905 444
4	31 128	32 124 096
3	571 008	32 547 456

Wariancja liczby trafień danego stopnia wynosi

$$\sigma^2(k) = np(1 - p).$$

Ponieważ

$$n \approx \frac{k_i}{p_i},$$

to

$$\sigma^2(n) \approx \left(\frac{\partial n}{\partial k_i} \sigma(k_i) \right)^2 = \frac{\sigma^2(k_i)}{p_i^2}$$

Stąd wariancja liczby zakładów jest równa

$$\sigma^2(n) = \frac{\sigma^2(k)}{p} = n(1 - p).$$

Wagi poszczególnych oszacowań liczby n liczymy na podstawie wariancji $\sigma^2(n)$

$$w_i = \frac{\frac{1}{\sigma^2(n_i)}}{\sum_{i=3}^6 \frac{1}{\sigma^2(n_i)}} = \frac{\frac{1}{n_i(1 - p_i)}}{\sum_{i=3}^6 \frac{1}{n_i(1 - p_i)}}.$$

Oczywiście

$$\sum_{i=3}^6 w_i = 1.$$

Stopień	$\sigma^2(k)$	$\sigma^2(n)$	w_i
6	3,0	586 641 287 814 120	0,0000029
5	644,0	1 891 875 064 800	0,0009123
4	31097,8	33 119 942 976	0,0521137
3	560990,3	1 822 657 536	0,9469710

Jak widzimy największy wkład do średniej ważonej będzie pochodził od informacji nt. liczby wygranych stopnia 3. Średnia ważona wynosi

$$\tilde{n} = \sum_{i=3}^6 w_i n_i \cong 32\,527\,572.$$

Wariancji średniej ważonej wynosi

$$\sigma^2(\tilde{n}) = \sum_{i=3}^6 \frac{1}{\sigma^2(n_i)} = \sum_{i=3}^6 \frac{1}{n_i(1 - p_i)} \approx 1\,726\,003\,832.$$

Stąd odchylenie standardowe wynosi

$$\sigma(\tilde{n}) \approx 41\,545.$$

Czyli

$$\tilde{n} = 32\,527(42) \cdot 10^3.$$

Policzmy na podstawie tej liczby wartości oczekiwane liczb wygranych poszczególnych stopni i ich odchylenia standardowe. Porównajmy wartości oczekiwane z faktycznymi liczbami wygranych. Dla rozkładu dwumianowego mamy

$$E(k) = np, \quad \sigma(k) = \sqrt{np(1-p)}$$

Stopień	Liczba wygranych k_i	$E(k)$	$k_i - E(k_i)$	$\sigma(k)$
6	3	2,3	0,7	1,5
5	644	600	44	24
4	31 128	31 519	-391	177
3	571 008	570 659	349	749

Policzmy prawdopodobieństwa $P_n(k)$ dla wyliczonej wartości n i $k = 0, 1, 2, 3, 4, 5$ trafień stopnia 6. Rozkład dwumianowy możemy przybliżyć rozkładem Poissona z wartością parametru $\lambda = np_6 \approx 2,326$.

k	$P_n(k)$
0	0,098
1	0,227
2	0,264
3	0,205
4	0,119
5	0,055

Prawdopodobieństwo, że padnie chociaż jedna wygrana szóstego stopnia wynosi

$$P_n(k > 0) = 1 - P_n(0) \approx 1 - 0,098 = 0,902.$$

Zadanie 6

Powtórz obliczenia z poprzedniego zadania dla wyników z losowania z dnia 10-05-2016 (pierwszego losowania po kumulacji), w którym liczba wygranych była następująca:

Stopień	Liczba wygranych
szóstka	0
piątka	42
czwórka	2 312
trójka	46 767

Rozwiązanie

Dla danych z tego zadania tabele wyników są następujące

Stopień	Liczba wygranych	Oszacowanie liczby zakładów
6	0	-
5	42	2 276 442
4	2 312	2 385 984
3	46 767	2 665 719

Stopień	$\sigma^2(k)$	$\sigma^2(n)$	w_i
6	-	-	-
5	42,0	123 383 156 400	0,00113
4	2 309,8	2 459 949 504	0,0571
3	45 946,5	149 280 264	0,941

$$\tilde{n} = \sum_{i=3}^{i=6} w_i n_i \cong 2\,649\,289.$$

Wariancji średniej ważonej wynosi

$$\sigma^2(\tilde{n}) = \sum_{i=3}^6 \frac{1}{\sigma^2(n_i)} \approx 140\,579\,229.$$

Stąd odchylenie standardowe wynosi

$$\sigma(\tilde{n}) \approx 11\,857.$$

Czyli

$$\tilde{n} = 2\,647(12) \cdot 10^3.$$

Stopień	Liczba wygranych k_i	$E(k)$	$k_i - E(k_i)$	$\sigma(k)$
6	0	0,19	0,19	0,44
5	42	49	-7	7,0
4	2 312	2 567	-255	51
3	46 767	46 479	288	214

k	$P_n(k)$
0	0,827
1	0,157
2	0,0149
3	0,00094
4	0,00004
5	0,0000017

Prawdopodobieństwo, że padnie chociaż jedna wygrana szóstego stopnia wynosi

$$P_n(k > 0) = 1 - P_n(0) \approx 1 - 0,827 = 0,173.$$

Zadanie 7

Pokaż, że parametr μ jest równy wartości średniej, a parametr σ^2 jest równy wariancji rozkładu

$$\text{Gaussa } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Rozwiązanie

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$u = \frac{x-\mu}{\sqrt{2}\sigma}, \quad x = \sqrt{2}\sigma u + \mu, \quad dx = \sqrt{2}\sigma du$$

$$E(X) = \int_{-\infty}^{\infty} (u\sqrt{2}\sigma + \mu) \frac{1}{\sigma\sqrt{2\pi}} e^{-u^2} \sqrt{2}\sigma du =$$

$$\frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{-\infty}^{\infty} u e^{-u^2} du + \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2} du = \frac{\mu}{\sqrt{\pi}} \cdot \sqrt{\pi} = \mu$$

$$\text{var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$u = \frac{x - \mu}{\sqrt{2}\sigma}, \quad (x - \mu)^2 = 2\sigma^2 u^2, \quad x - \mu = \sqrt{2}\sigma u, \quad dx = \sqrt{2}\sigma du$$

$$\sigma^2(X) = \int_{-\infty}^{\infty} 2\sigma^2 u^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-u^2} \sqrt{2}\sigma du = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} u^2 e^{-u^2} du = \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2} = \sigma^2$$

Zadanie 8

Pokaż, że dla $x = \mu \pm \sigma$ rozkład Gaussa ma tzw. punkty przegięcia.

Wskazówka:

W punkcie przegięcia różniczkowalnej funkcji, pochodna tej funkcji ma lokalne ekstremum. Oblicz pochodną rozkładu Gaussa i znajdź jej punkty ekstremalne.

Rozwiązanie

$$\frac{d}{dx} \left[e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right] = -\frac{(x-\mu)}{\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\begin{aligned} \frac{d}{dx} \left[(x - \mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right] &= e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left(1 - (x - \mu) \cdot \frac{(x - \mu)}{\sigma^2} \right) = \\ &= e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left(1 - \frac{(x - \mu)^2}{\sigma^2} \right). \\ \left(1 - \frac{(x - \mu)^2}{\sigma^2} \right) &= 0, \\ (x - \mu)^2 &= \sigma^2, \\ x - \mu &= \pm \sigma \\ x &= \mu \pm \sigma \end{aligned}$$

Zadanie 9

Korzystając z funkcji Excela Rozkład.Normalny sporządź wykresy gęstości prawdopodobieństwa, ich dystrybuanty i kwantyle dla kilku wybranych wartości parametrów μ i σ rozkładu Gaussa.

Rozwiązanie

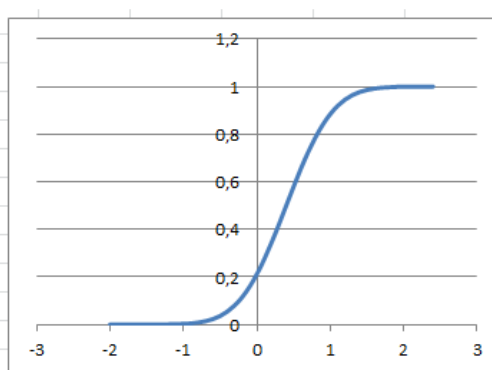
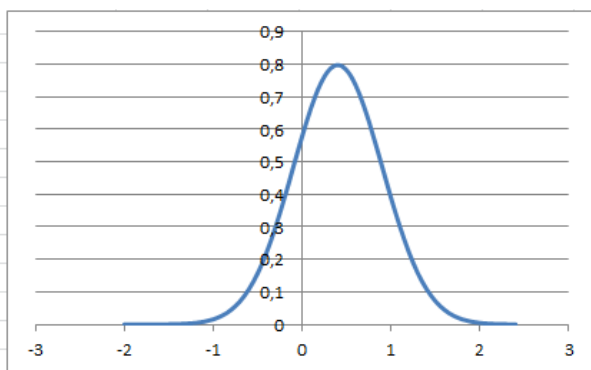
Dystrybuanta

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{2} \left(1 + \text{erf} \left(\frac{t}{\sqrt{2}} \right) \right),$$

gdzie

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^x e^{-t^2} dt$$

jest tzw. funkcją błędu



gęstość	ROZKŁ.NORMALNY(x,mi,sigma,0)
dystybuanta	ROZKŁ.NORMALNY(x,mi,sigma,1)
kwantyle	ROZKŁ.NORMALNY.ODWR(prawdopodobieństwo,mi,sigma)

