Niektóre rozkłady zmiennych losowych

Zmienna losowa skokowa
- ♦Rozkład dwumianowy
- ♦Rozkład Poissona

Zmienna losowa ciągła
- ♦Rozkład prostokątny (mikrokanoniczny)
- ♦Rozkład normalny (Gaussa)
- ♦Rozkład χ^2 (chi kwadrat)
- ♦Rozkład Studenta

Rozkład dwumianowy: wielokrotna realizacja doświadczenia, w wyniku którego otrzymać można tylko jedno z dwu wykluczających się zdarzeń –zdarzenie A (z prawdopodobieństwem p) lub nie-A (z prawdopodobieństwem 1-p). Jako przykład można podać wielokrotnie powtarzany rzut monetą (zdarzenie A- wyrzucenie np. reszki, p=0.5). Jeżeli wyniki kolejnych doświadczeń oznaczymy przez x_i (0 lub 1 w rzucaniu monetą), to łączny rezultat n doświadczeń charakteryzuje zmienna losowa X zdefiniowana wzorem

$$X = \sum_{i=1}^{n} x_i$$

Rozkład dwumianowy – rozkład zależności prawdopodobieństwa P(X=k) od wartości k w n doświadczeniach

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

Wartość oczekiwana w rozkładzie dwumianowym dla x_i=k

$$E(X) = \sum_{k=1}^{n} kP(X = k) = np$$

Wariancja rozkładu dwumianowego

$$D^2(X) = \sum_{k=1}^{n} \left[k - E(X)\right]^2 P(X = k) = np(1-p)$$

Rozkład Poissona: szczególny przypadek rozkładu dwumianowego zachodzącym wtedy, gdy prawdopodobieństwo p sukcesu jest bardzo małe, a liczba realizacji n na tyle duża, że iloczyn $n\cdot p=\lambda$ jest wielkością stałą, dodatnią i niezbyt dużą.

$$P(X = k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

Wartość oczekiwana zmiennej losowej w rozkładzie Poissona

$$E(X) = \lambda$$

Wariancja

$$D^2(X) = \lambda$$

Zastosowanie rozkładu Poissona – tam, gdzie liczba obserwowanych przypadków n jest bardzo duża, a prawdopodobieństwo sukcesu p bardzo małe. Przykłady

- rozpad promieniotwórczy: liczba jąder n duża, prawdopodobieństwo rozpadu konkretnego jądra bardzo małe;
- zderzenia cząstek elementarnych, duża ilość cząstek, mała szansa na zderzenie;
- statystyczna kontrola jakości produktów, duża ilość sprawdzanych produktów, mała ilość produktów wybrakowanych.

Rozkład prostokątny: Ma zastosowanie przy analizie niepewności systematycznych. Gęstość prawdopodobieństwa f(x) jest stała wewnątrz przedziału (a, b) i równa zero poza nim.

$$f(x) = \begin{cases} \dfrac{1}{b-a} & dla \quad a < x < b \\ 0 \quad dla & x < a \quad i \quad x > b \end{cases}$$

Wartość oczekiwana rozkładu prostokątnego

$$E(X) = \frac{b-a}{2}$$

Wariancja dla rozkładu prostokątnego

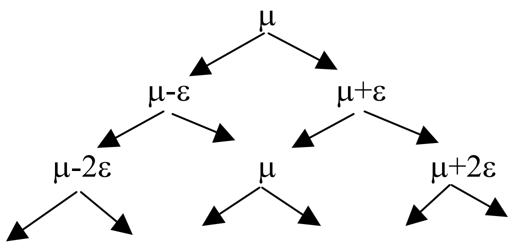
$$D^2(X) = \frac{(b-a)^2}{12}$$

Dla pomiarów obarczonych niepewnością systematyczną Δx, mamy b – a = 2Δx, zatem

$$S_x = \sqrt{D^2(X)} = \frac{\Delta x}{\sqrt{3}}$$

Rozkład normalny:

Mamy do czynienia z rozkładem normalnym wtedy, gdy pomiar pewnej wielkości, mającej wartość μ zakłócany jest bardzo dużą liczbą niezależnych czynników, z których każdy z prawdopodobieństwem ½ powoduje odchylenie o niewielką wartość ±ε.



Gęstość prawdopodobieństwa *rozkładu normalnego standaryzowanego*

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$$

Rozkład ten oznaczany jest także jako N(0, 1). Wartość oczekiwana i wariancja rozkładu normalnego standaryzowanego

$$E(U) = 0, \qquad D^2(U) = 1$$

Dystrybuanta Φ(u) rozkładu standaryzowanego

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} exp\left(-\frac{1}{2}u^2\right) du$$

Wartości dystrybuanty dla u>0 są stabelaryzowane. Wartości dystrybuanty dla u<0 wyznaczyć można z równania: Φ(-u)=1 - Φ(u).

Dokonując podstawienia $u = \frac{x-\mu}{\sigma}$ otrzymamy postać niestandaryzowaną rozkładu Gaussa.

Gęstość prawdopodobieństwa *rozkładu normalnego niestandaryzowanego*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Rozkład ten oznaczany jest także jako N(μ, σ). Wartość oczekiwana i wariancja rozkładu normalnego niestandaryzowanego

$$E(X) = \mu, \qquad D^2(X) = \sigma^2$$

Rozkład χ^2: Gdy X_i są zmiennymi losowymi losowanymi z rozkładu normalnego N(0,1), to $\sum_{i=1}^{k} X_i^2$ ma rozkład chi-kwadrat o *k* stopniach swobody. Gdy losowanie odbywa się z rozkładu normalnego N(μ,σ), to zmienną losową χ^2 definiujemy następująco

$$\chi^2 = \sum_{i=1}^{k} \frac{(X_i - \mu)^2}{\sigma^2}$$

Gęstość prawdopodobieństwa rozkładu chi-kwadrat

$$f(x)=\begin{cases}\dfrac{1}{\Gamma\left(\dfrac{k}{2}\right)2^{\frac{k}{2}}}x^{\frac{k}{2}-1}e^{-\frac{x}{2}} & dla \quad x>0\\ 0 & dla \quad x\le 0\end{cases}$$

gdzie Γ jest funkcją gamma Eulera, a parametr k nazywa się liczbą stopni swobody. Gdy $k<2$, to funkcja f jest malejącą dla $x>0$, natomiast dla $k>2$ funkcja ta ma maksimum przy $x=k-2$. Dla dużych k funkcja f jest zbliżona do krzywej rozkładu normalnego. Wartość oczekiwana zmiennej losowej o rozkładzie chi kwadrat jest równa liczbie stopni swobody k, zaś wariancja jest równa 2k.

Największe znaczenie praktyczne dla rozkładu chi kwadrat mają tablice wartości krytycznych $\chi^2_{\alpha,k}$ zmiennej losowej χ^2, dla których

$$P\left\{\chi^2 \ge \chi^2_{\alpha,k}\right\}=\alpha$$

α nazywa się *poziomem istotności*. Wielkość (1-α) nazywa się *poziomem ufności*.

Rozkład Studenta: Zmienną losową t Studenta definiujemy wzorem

$$t=\frac{Z}{U}\sqrt{k}$$

gdzie Z jest zmienną losową standaryzowaną N(0, 1), a U zmienną losową o rozkładzie chi kwadrat i k stopniach swobody.

Gęstość prawdopodobieństwa rozkładu Studenta

$$f(t)=\frac{\Gamma\left(\dfrac{k+1}{2}\right)}{\Gamma\left(\dfrac{k}{2}\right)\sqrt{k\pi}}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}}$$

gdzie Γ jest funkcją gamma Eulera, a parametr k nazywa się liczbą stopni swobody.

Rozkład Studenta jest identyczny z rozkładem Gaussa N(0, 1) dla $k=\infty$ i staje się coraz bardziej spłaszczony dla malejących k. Wartość oczekiwana rozkładu Studenta jest równa zero, wariancja jest równa $k/(k-2)$.

Tablice Studenta zawierają zazwyczaj tak zwane wartości krytyczne $t_{n,\alpha}$ zmiennej losowej Studenta, zdefiniowane wyrażeniem

$$P\left\{|t|<t_{n,\alpha}\right\}=1-\alpha \qquad \text{lub} \qquad P\left\{|t|>t_{n,\alpha}\right\}=\alpha$$

gdzie α jest ustalonym z góry prawdopodobieństwem, zwanym poziomem istotności.

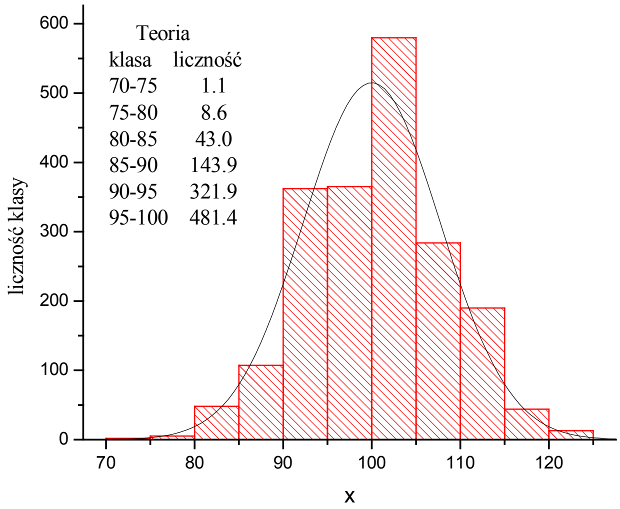
PRZYKŁADY:

Rozkład normalny N[μ,σ]. Wygenerowany zostanie zbiór liczb zgodnie z przedstawionym powyżej schematem powstania rozkładu normalnego. Przyjmuję: μ=100, ε=1, ilość niezależnych czynników zaburzających (poziomów na rysunku): 60, ilość powtórzeń („pomiarów”): 2000.

Wygenerowane liczby (program MATHEMATICA) umieszczam w tablicy gauss i eksportuję do pliku gauss.dat

Parametry zbioru tych liczb: rozstęp: 48, wartość średnia: 100.02, wariancja: 61.49, odchylenie standardowe: 7.84, skośność: -0.0375, eksces: -0.0096.

Plik gauss.dat wczytuję do programu Origin i sporządzam histogram. Na histogram nałożono wykres funkcji Gaussa z następującymi parametrami: μ=100, $\sigma=(60)^{1/2}$, współczynnik liczbowy przed gęstością prawdopodobieństwa =(ilość pomiarów)·(szerokość klasy)=2000·5=10000.



Teoretyczna liczebność danej klasy obliczono korzystając z programu MATHEMATICA. Ponieważ krzywa Gaussa jest symetryczna względem x=100, to liczebności klasy x>100 jest równa liczebności symetrycznie położonej klasy x<100.

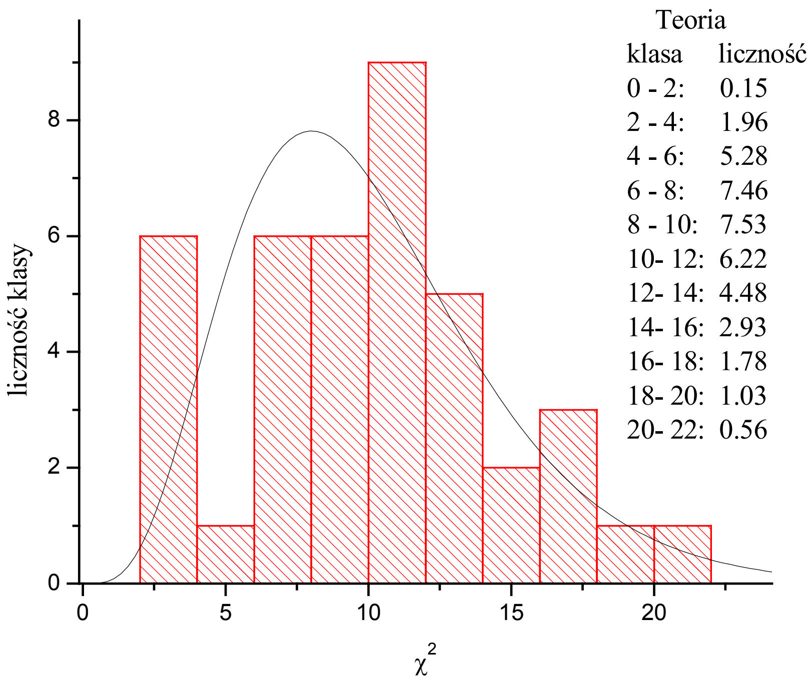
Rozkład χ^2: 1. Z rozkładu normalnego N(0,1) losuję 10 liczb:

-0.5373	1.2410	0.4724	-0.0535	-0.0396
0.9565	2.3512	-0.5431	-1.2842	0.7915

2. Znajduję sumę kwadratów tych liczb: S=11.07
3. Sposób postępowania z p. 1-2 powtarzam 40 razy, otrzymując następujące liczby (są to sumy kwadratów wylosowanych 10 liczb):

11.07	12.6938	9.2834	16.3849	11.2446	7.7831
14.2588	16.6633	12.0763	3.6088	6.77	4.4977
8.8891	7.1067	6.4685	8.9148	13.8252	13.7997
10.0956	11.8455	9.6063	8.2365	7.4864	3.4553
2.4258	8.6349	10.3829	11.1247	7.2504	15.0595
20.7971	3.6683	2.6999	10.714	2.9807	11.0749
11.3654	19.1772	16.514	12.8986		

4. Sporządzam histogram dla otrzymanych czterdziestu liczb



5. Na ten histogram nakładam wykres funkcji rozkładu gęstości prawdopodobieństwa dla rozkładu chi-kwadrat z k=10 stopniami swobody. Ponieważ funkcja gamma Eulera Γ(5)=(5-1)!=24, to ta funkcja ma postać

$$f(x) = \frac{1}{768} x^4 e^{-\frac{x}{2}}$$

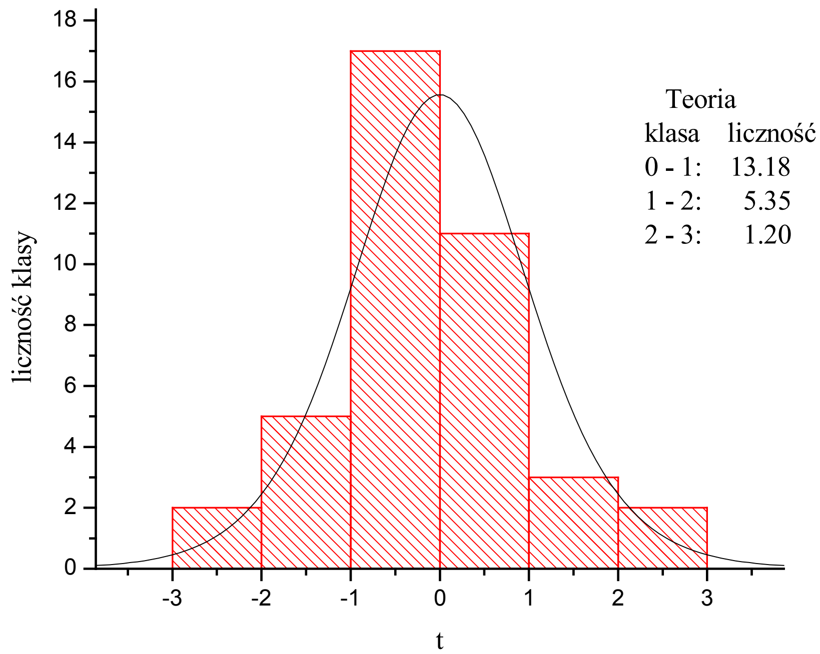
Ponadto na rysunku przedstawiono, korzystając z dystrybuanty rozkładu, obliczone ilości liczb w każdej z klas.

Rozkład Studenta. Niech liczba stopni swobody k=10.

1. Z rozkładu normalnego N(0,1) losuję liczbę, którą oznaczę jako Z, z rozkłady chi-kwadrat o dziesięciu stopniach swobody losuję liczbę, którą oznaczę jako U.
2. Obliczmy wartość parametru t z równania $t = \frac{Z}{U}\sqrt{10}$
3. Czynności z p.1-2 powtarzam 40 razy, otrzymując poniższe 40 liczb

-1.71664	-2.78894	-0.4572	0.42744	0.03692	-0.08404
0.20736	-1.66423	-0.9279	-0.77786	-0.02006	-1.06892
2.66153	-0.73455	1.54518	-0.71934	-1.28696	1.09912
-0.96341	-0.175	-0.87677	-0.30628	0.42025	0.60262
-1.36794	-0.68633	0.21973	2.6117	0.42999	-0.20936
-0.05301	0.39089	-0.2769	0.78756	0.12263	-2.01082
-0.18158	1.65041	0.92888	-0.25252		

4. Sporządzam histogram dla otrzymanych czterdziestu liczb



5. Na ten histogram nakładam wykres funkcji rozkładu gęstości prawdopodobieństwa dla rozkładu Studenta z k=10 stopniami swobody. Ponieważ funkcja gamma Eulera $\Gamma(5)=24$, $\Gamma(11/2)=(945/32)\pi^{1/2}$, to ta funkcja ma postać

$$f(t) = 0.3891\left(1+\frac{t^2}{10}\right)^{-11/2}$$

Ponadto na rysunku przedstawiono, korzystając z dystrybuanty rozkładu, obliczone ilości liczb w każdej z klas.