

SdI30 W03: *ROZKŁAD NORMALNY I JEGO ZASTOSOWANIA*

1. Rozkład normalny i jego własności

Przykład 1

Przykład 2

2. Standaryzacja rozkładu normalnego

Przykład 3

3. Funkcja kwantylowa rozkładu normalnego

Przykład 4

4. Wsparcie excelowskie

5. Studium przypadku

6. Zestaw zadań W03

1. Rozkład normalny i jego własności

Zm. 1. X typu ciągłego ma *rozkład normalny* (*normal distribution*) z parametrami μ i σ ($\mu \in \mathbb{R}, \sigma > 0$), co oznaczamy $X \sim \mathcal{N}(\mu, \sigma)$, jeśli jej gęstość wyraża się wzorem:

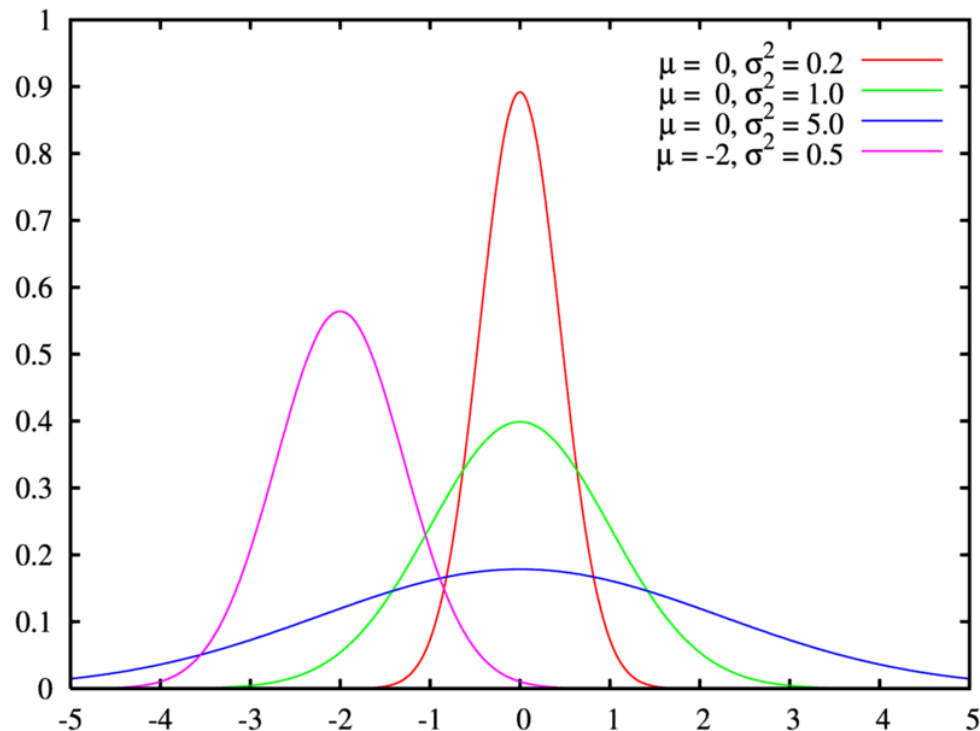
$$\text{PDF: } f_{\mathcal{N}}(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), x \in \mathbb{R}$$

Gęstość rozkładu normalnego zaproponował **Gauss**¹, jako model rozkładu częstości błędów pomiarowych.



¹ Carl Friedrich Gauss (1777-1855). – matematyk niemiecki. Jeden z najwybitniejszych matematyków wszystkich czasów, zwany przez współczesnych księżem matematyków. Profesor uniwersytetu w Getyndze.

Na jego cześć krzywą gęstości rozkładu normalnego nazywamy **krzywą Gaussa**. Gęstość osiąga maksimum w punkcie $x = \mu$. Dla $x = \mu \pm \sigma$ krzywa Gaussa ma punkty przegięcia.



Rys. 3. Krzywe Gaussa

Własność: $(X \sim \mathcal{N}(\mu, \sigma)) \Rightarrow (\mathbb{E}X = \mu, \mathbb{D}X = \sigma).$

Twierdzenie (o kombinacji liniowej). Jeżeli X_1 i X_2 są niezależnymi zm. losowymi oraz

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1), \quad X_2 \sim \mathcal{N}(\mu_2, \sigma_2),$$

to dla dowolnych stałych rzeczywistych c_1, c_2 , z których co najmniej jedna jest różna od zera, kombinacja liniowa

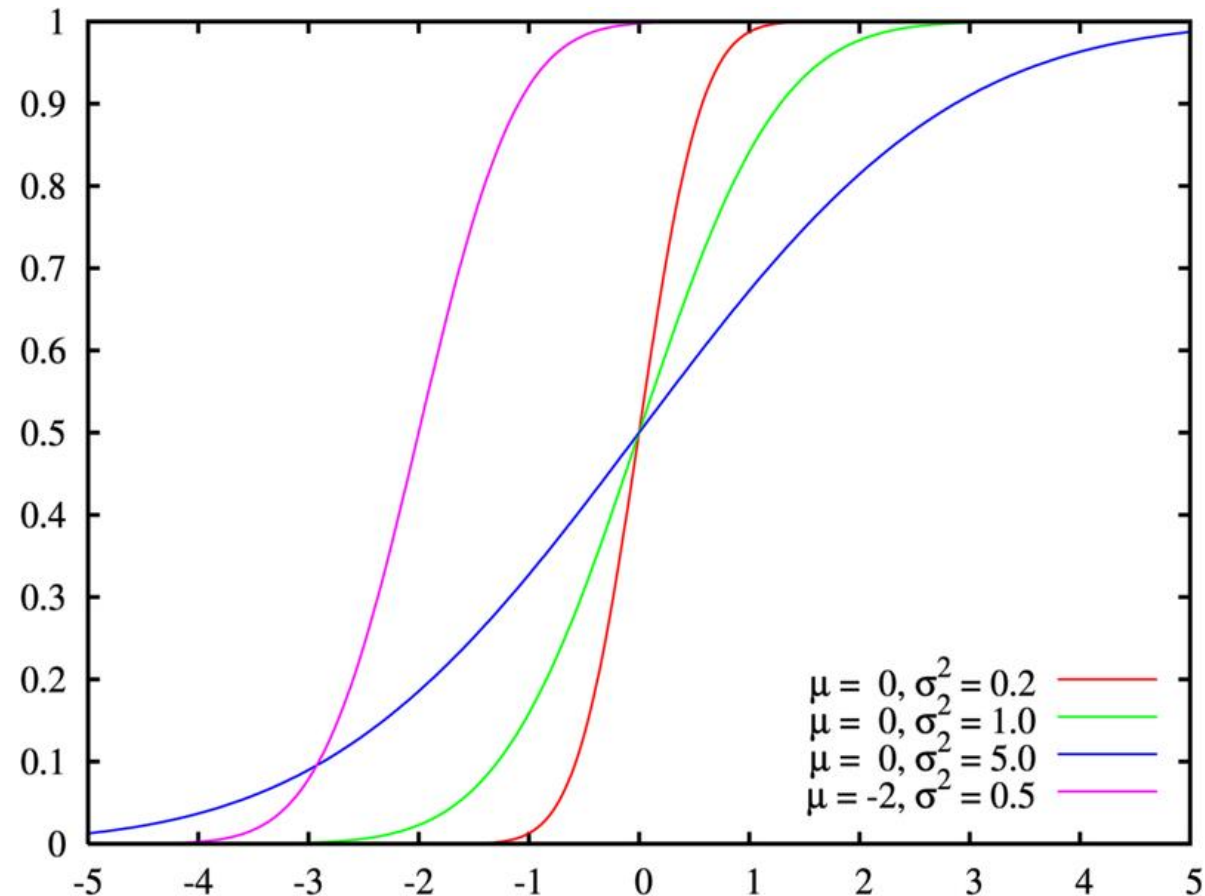
$$U = c_1 X_1 + c_2 X_2$$

ma również rozkład normalny z podanymi parametrami


$$U \sim \mathcal{N}\left(c_1 \mu_1 + c_2 \mu_2, \sqrt{c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2}\right)$$

Dowód. Parametry wynikają z własności operatorów \mathbb{E} i \mathbb{D}^2 .

Twierdzenie można uogólnić na kombinację liniową n zmiennych losowych.



Rys. 4. Wykresy dystrybuant rozkładów normalnych

Przykład 1.  Czas T_1 przejazdu tramwaju między przystankami A i B ma rozkład $\mathcal{N}(180; 12)[sek.]$, a między przystankami B i C czas T_2 ma rozkład $\mathcal{N}(140; 9)[sek.]$.

- a) Ile wynosi oczekiwany czas i wariancja czasu przejazdu tramwaju między przystankami A i C ?
- b) Obliczyć $P(T_1 + T_2 > 6[minut])$.

Przykład 2. Wytrzymałość lin stalowych (wyrażona w [MPa]), pochodzących z masowej produkcji, jest zm. 1. W o gęstości:

$$f(w|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(w - 100)^2}{50}\right), w \in \mathbb{R}$$

Ile wynoszą wartość oczekiwana i wariancja wytrzymałości?

Odp.: $\mathbb{E}W = 100[\text{MPa}]$, $\mathbb{D}^2W = 25[\text{MPa}]^2$.

Zastosowanie. Rozkład normalny jest najważniejszym i najczęściej stosowanym rozkładem w teoretycznych rozważaniach probabilistyki i statystyki matematycznej oraz najczęściej stosowanym rozkładem w zastosowaniach inżynierskich i ekonomicznych.

2. Standaryzacja rozkładu normalnego

Standaryzacja należy do podstawowych przekształceń liniowych zm. l.

Stosując ją dla rozkładu normalnego otrzymamy twierdzenie: jeśli $X \sim \mathcal{N}(\mu, \sigma)$, to zm. l. standaryzowana $Z = \frac{X - \mu}{\sigma}$ ma rozkład $\mathcal{N}(0, 1)$, zwany *standardowym rozkładem normalnym*, tj.:

$$X \sim \mathcal{N}(\mu, \sigma) \Rightarrow Z \sim \mathcal{N}(0, 1)$$

Wikipedia. Standard scores are also called **z-values**, **z-scores**, **normal scores**, and **standardized variables**; the use of "Z" is because the [normal distribution](#) is also known as the "Z distribution".

Dystrybuantę stand. rozkładu normalnego oznaczamy grecką literą Φ .

$$\text{CDF dla } \mathcal{N}(0, 1): \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{x^2}{2}\right) dx, \quad z \in \mathbb{R}$$

Z symetrii gęstości stand. rozkładu normalnego względem osi Oy wynika zależność:

$$\forall_{z \in \mathbb{R}} \quad \Phi(-z) = 1 - \Phi(z)$$

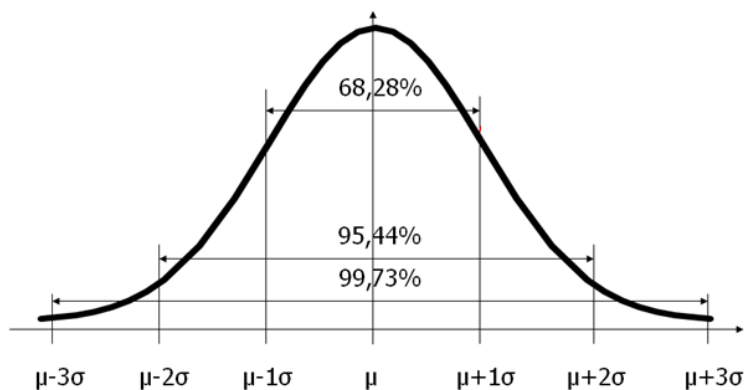
Wartości funkcji Φ są tablicowane.

Korzystając z wartości tablicowych można obliczać prawd. dla dowolnej zm. l. o rozkładzie normalnym dokonując jej standaryzacji (zamiast $=$ piszemy wówczas $\stackrel{STD}{=}$).

Jeżeli $X \sim \mathcal{N}(\mu, \sigma)$, to $Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$. Stąd

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$

$$P(X \leq a) = \Phi\left(\frac{a - \mu}{\sigma}\right), \quad P(X \geq b) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right)$$



Rys. Graficzne przedstawienie reguły sigma

Przykład 3. Wytrzymałość W (w [MPa]) pewnej partii lin stalowych, ma rozkład normalny jak w przykładzie 2.

- a) Obliczyć prawd. zdarzenia, że losowo wybrana lina z tej partii będzie miała wytrzymałość ponad 105[MPa].
- b) Ile przeciętnie lin spośród 1000 ma wytrzymałość mniejszą niż 90 [MPa]?



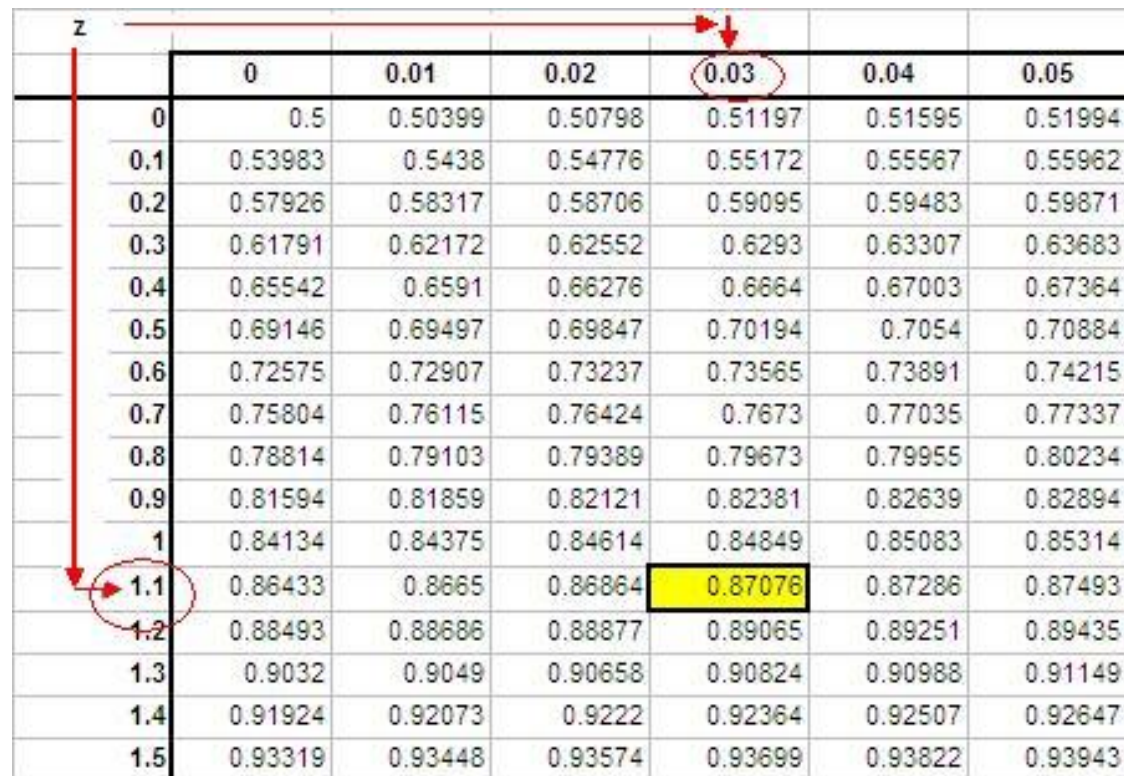
Z praw wielkich liczb możemy przyjąć, że częstość przyjmowania wartości z przedziału $(-\infty, w)$ jest równa prawd. przyjmowania wartości z tego przedziału.

- a) Obliczamy prawd. zdarzenia $W > 105[\text{MPa}]$

$$P(W > 105) = 1 - P(W \leq 105)$$

$$\stackrel{STD}{=} 1 - P\left(\frac{W-100}{5} \leq \frac{105-100}{5}\right) = 1 - P(Z \leq 1) = 1 - \Phi(1)$$

Wartość $\Phi(1)$ odczytana z tablicy statystycznej lub z komputera, wynosi w przybliżeniu 0,8413. Stąd prawd., że losowo wybrana lina ma wytrzymałość większą niż 105 [MPa] wynosi 0,1587.



z	0	0.01	0.02	0.03	0.04	0.05
0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943

Rys. Tablica statystyczna

b) Obliczamy prawd. zdarzenia $W < 90$ [MPa]

$$\begin{aligned} P(W < 90) &\stackrel{STD}{=} P\left(\frac{W - 100}{5} < \frac{90 - 100}{5}\right) \\ &= P(Z < -2) = \Phi(-2) = 1 - \Phi(2) \\ &\stackrel{TABL}{=} 1 - 0,9772 \approx 0,0228 \end{aligned}$$

czyli przeciętnie 23 liny na tysiąc mają wytrzymałość mniejszą niż 90[MPa].

3. Funkcja kwantylowa rozkładu normalnego

Niech $F_X(x|\mu, \sigma)$ będzie dystrybuantą zm. l. $X \sim \mathcal{N}(\mu, \sigma)$. Kwantyle zm. l. X wyznaczamy za pomocą funkcji kwantylowej $F_X^{-1}(p|\mu, \sigma)$, określonej dla $p \in (0,1)$ wzorem:

$$F_X^{-1}(p|\mu, \sigma) = \mu + \sigma\Phi^{-1}(p)$$

gdzie $\Phi^{-1}(p)$ jest funkcją kwantylową rozkładu $\mathcal{N}(0, 1)$.

Ponieważ kwantyl rzędu p stand. rozkładu normalnego przedstawia się wzorem $z_p = \Phi^{-1}(p)$, więc pomiędzy kwantylami zachodzi związek:


$$\Phi^{-1}(p) = -\Phi^{-1}(1 - p), \text{ tj. } z_p = -z_{1-p}$$

Stąd wystarczy znać jedynie wartości kwantyli dla rzędów $p \in [0,5; 1)$.

Wartości funkcji odwrotnej Φ^{-1} podobnie jak samej dystrybuanty są zestawiane w tablicach statystycznych. Można je również wyznaczać za pomocą programów komputerowych.

p	0,75	0,90	0,95	0,975	0,99	0,995	0,999
z_p	0,6745	1,2816	1,6449	1,9600	2,3263	2,5758	3,0902

Tablica. Skrócona tablica kwantyli $z_p = \Phi^{-1}(p)$

Przykład 4.  Zużycie paliwa niezbędnego do przebycia przez odrzutowiec odległości między dwoma bazami lotniczymi jest zm. l. X o rozkładzie $\mathcal{N}(5,7; 0,5)$ [tony].

Ustalić ilość tankowanego paliwa tak, aby prawdopodobieństwo dolotu do drugiej bazy wyniosło ponad 0,999.

Dane: $X \sim \mathcal{N}(5,7; 0,5)[\text{tony}]$.

Szukane: wartość x dla której $P(X \leq x) = 0,999$, czyli kwantyl rzędu 0,999, tj. $x_{0,999}$.

Obliczenia: Korzystamy z zależności

$$x_{0,999} = \mathbb{E}X + \mathbb{D}X \cdot z_{0,999}.$$

Ponieważ $\mathbb{E}X = 5,7$; $\mathbb{D}X = 0,5$; $z_{0,999} = \Phi^{-1}(0,999) = 3,0902$, więc $x_{0,999} = 7,2451$ ton.

Zatankowanie 7,2451 ton paliwa daje pewność rzędu 99,9%, że wystarczy paliwa na ten przelot.

Decyzja: Zatankowanie 7,2451 ton paliwa daje pewność rzędu 99,9%, że wystarczy paliwa na ten przelot.

W internecie:

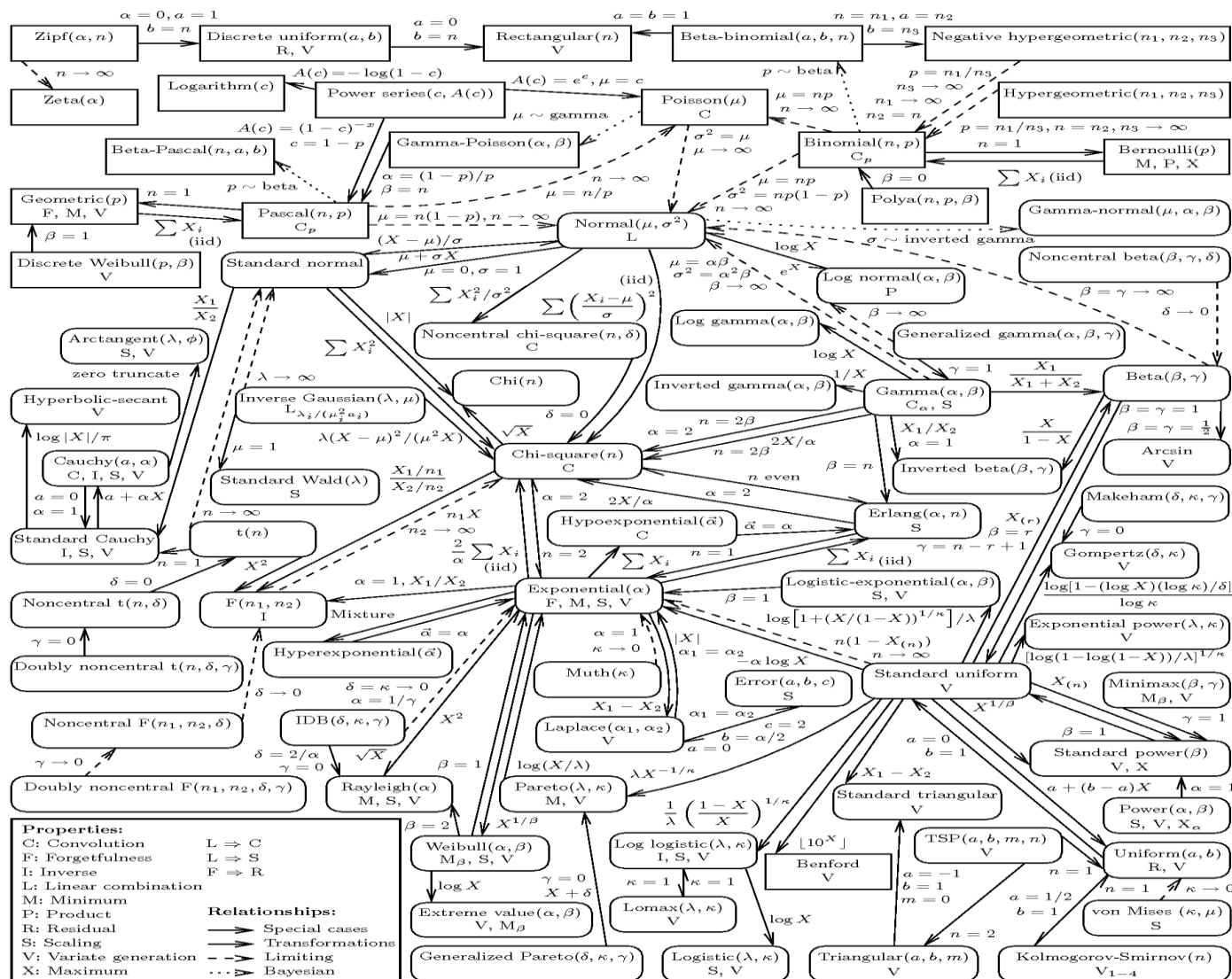
- 1) [Interactivate: Normal Distribution](#) 2) [Normal Distribution Calculator](#)

4. Wsparcie excelowskie

nazwa	definicja	funkcja excela
Rozkład dwumianowy	$f(x n, p) = \binom{n}{x} p^x (1 - p)^{n-x},$ $x = 0, 1, 2, \dots, n$	ROZKŁ.DWUM(x;n;p; 0)
Rozkład Poissona	$f(x \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \dots$ $\lambda > 0$	ROZKŁ.POISSON(x; λ; 0)
Rozkład wykładniczy	$F(x \lambda) = 1 - \exp(-\lambda x), x \geq 0$ $\lambda > 0$	ROZKŁ.EXP(x; λ; 1)
Rozkład normalny	$F(x \mu, \sigma)$ $= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) dt$ $\mu \in \mathbb{R}, \sigma > 0$	ROZKŁ.NORMALNY(x; μ, σ, 1)

Tabela. Wybrane rozkłady dostępne w Excelu.

Zobacz związki między rozkładami <http://www.math.wm.edu/~leemis/chart/UDR/UDR.html>



5. Studium przypadku .

Czas X (w [h]) sprawności pewnych przyrządów pomiarowych jest zmienną losową o gęstości prawdopodobieństwa

$$f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2 - 4000x + 4000000}{80000}\right), x \in \mathbb{R}.$$

- Rozpoznać rozkład czasu sprawności przyrządu i podać jego parametry.
- Obliczyć prawd. zdarzenia $|X - 1800| < 100$.
- Wyznaczyć czas $x_{0,05}$, w którym utraci sprawność 5% przyrządów z dużej ich partii.
- Norma czasu sprawności przyrządów jest określona na co najmniej 1600[h]. Obliczyć prawd., że losowo wybra-

ne dwa przyrządy spełnią normę sprawności. Obliczyć prawd., że co najmniej jeden z nich spełni normę.

e) Obliczyć prawd., że przyrząd będzie sprawny przez 2200[h], jeśli nie utraci sprawności przez 2000[h].

Rozwiązanie. a) $X \sim \mathcal{N}(2000; 200)[h]$.

$$b) P(|X - 1800| < 100) = F_{\mathcal{N}(2000;200)}(1900) - F_{\mathcal{N}(2000;200)}(1700)$$

$$\stackrel{\text{EXCEL 2013}}{=} \text{ROZKŁ. NORMALNY}(1900; 2000; 200; 1) - \text{ROZKŁ. NORMALNY}(1700; 2000; 200; 1) = (0,308538 - 0,066807) = 0,24173.$$

$$c) x_{0,05} = F_X^{-1}(0,05)$$

$$\stackrel{\text{EXCEL 2013}}{=} \text{ROZKŁ. NORMALNY. ODWR}(0,05; 2000; 200) = 1671,029[h]$$

$$d) P(X > 1600) = 1 - P(X \leq 1600) = 1 - F_{\mathcal{N}(2000;200)}(1600)$$

$$\stackrel{\text{EXCEL 2013}}{=} 1 - \text{ROZKŁ. NORMALNY}(1600; 2000; 200; 1) = 1 - 0,02275 = 0,97725$$

$$P(X_1 > 1600, X_2 > 1600) \stackrel{\text{z niezależności}}{=} 0,97725^2 \approx 0,955$$

$$\begin{aligned} P(X_1 > 1600 \text{ lub } X_2 > 1600) \\ = 0,97725 + 0,97725 - 0,97725^2 \approx 0,9995. \end{aligned}$$

6. Zestaw zadań W03

1. Przykład projektu zaliczeniowego cz. 1



Uwaga. Należy przytaczać wzory i składnie funkcji wykorzystywanych w rozwiązaniach. Udzielać pełnych odpowiedzi. Sporządzić tabelę ocen według wzoru.

Etap	1	2	3	4	5	6	7	Łącznie
do uzyskania	4	4	4	2	2	3	9	28
uzyskano								

Długość X (w [mm]) detalu produkowanego na pewnym automacie jest zmienną losową o gęstości prawdopodobieństwa

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x^2 + 40x - 400}{0,08}\right), x \in \mathbb{R}$$

1. Rozpoznać rozkład długości detalu i jego parametry, wyznaczyć drugi moment zwykły długości detalu, sporządzić krzywą gęstości i dystrybuantę.
2. Obliczyć prawd. zdarzeń: $|X - 19,98| \geq 0,02$, $|X - \mathbb{E}X| < \mathbb{D}X$.
3. Dla jakiej wartości stałej b zachodzi równość $P(x_{0,05} < X < b) = 0,90$?
4. Wyznaczyć kwartyle długości detalu oraz obliczyć gęstości dla nich.
5. Wyznaczyć przedział, w którym mieści się 95% produkowanych detali po złomowaniu 5% detali o największej odchyłce długości od wymiaru przeciętnego.
6. Co wynika z faktu, że łączna długość 180 wyprodukowanych detali będzie mniejsza od 358[cm]?

7. Detal spełnia normę długości, jeśli jego długość mieści się w dopuszczalnym przedziale $(19,6; 20,4)$ [mm]. W celu sprawdzenia dokładności produkcji zmierzona zostanie długość 180 losowo wybranych detali.

- a) Wprowadzić zmienną losową opisującą wynik sprawdzania normy długości badanej partii detali. Podać jej rozkład i sporządzić wykresy PMF i CDF.
- b) Obliczyć prawd. zdarzenia, że w badanej partii detali, co najmniej 175 z nich spełni normę długości.
- c) Wyznaczyć wartość oczekiwaną, odchylenie standardowe oraz modę liczby detali, które spełnią normę długości i prawdopodobieństwo dla mody.

2. Z partii produkowanych wyrobów, wśród których jest 10% extra, pobierzemy próbę o liczności $n = 12$, w celu sprawdzenia frakcji wyrobów extra w próbie.

a) Jaki jest rozkład liczby wyrobów extra w próbie, tj. zm.

1. T_n ?

b) Czy rozsądne jest aproksymowanie zm. 1. T_n rozkładem normalnym?

c) Obliczyć prawd. zdarzenia $T_n \geq 2$.

d) Obliczyć wartości oczekiwane i wariancje zm. losowych T_n i \bar{P}_n .



© Can Stock Photo - csp19186097

3. Przypuśćmy, że w pewnej populacji ludzi wysokość kobiety: $X \sim \mathcal{N}(168; 5)$ [cm], a mężczyzny: $Y \sim \mathcal{N}(187; 7)$ [cm]. Z populacji tej wylosowani zostaną jedna kobieta i jeden mężczyzna. Obliczyć prawd., że

- a) mężczyzna będzie wyższy od kobiety o ponad 10[cm];
- b) kobieta będzie wyższa od mężczyzny.
- c) średnia arytmetyczna ich wysokości będzie w przedziale (170; 175) [cm].
- d) niższa z wylosowanych osób będzie niższa niż 160[cm].
- e) wyższa z wylosowanych osób będzie wyższa niż 180[cm].



4. Opór R pewnego typu oporników elektrycznych można opisać rozkładem normalnym $\mathcal{N}(\mu, \sigma)$. Koszt produkcji jednego opornika wynosi k , jego cena rynkowa zaś równa jest $5k$, gdy $R \in (\mu - \sigma, \mu + \sigma)$ i $2k$, jeżeli $R \in (\mu - 2\sigma, \mu - \sigma)$ lub $R \in (\mu + \sigma, \mu + 2\sigma)$. Oporniki, które nie spełnią podanych kryteriów, nie mogą być sprzedawane. Oblicz dochód na jeden opornik.



5. Waga netto X [tony] towarów wysyłanych w kontenerach określonych wymiarów jest normalną zm. l. o nieznanych parametrach. Wiadomo, że 65% kontenerów wykazuje wagę netto ponad 4,9 ton, a 25% kontenerów wagę netto mniejszą niż 4,2 tony.

- a) Wyznacz nieznane parametry rozkładu wagi netto towarów wysyłanych w tych kontenerach.
- b) Oblicz procent kontenerów, które mają wagę w przedziale od 4 do 5 ton?

Odp.: a) $\mathbb{E}X = 5,83$ tony; $\mathbb{D}X = 2,41$ tony.

