Rozkład Weibulla

- Funkcja gęstości prawdopodobieństwa

$$f_X(x) = \alpha\sigma^{-\alpha}x^{\alpha-1}e^{-(x/\sigma)^\alpha} \qquad x\in[0,+\infty), \quad \alpha>0, \sigma>0$$

- Parametr α decyduje o kształcie, σ o położeniu i wysokości (skali).
  dla α=1 - rozkład wykładniczy,
  dla α=2,3 - rozkład zbliżony do normalnego
- Dystrybuanta

$$F_X(x) = 1 - e^{-(x/\sigma)^\alpha}$$

- Wartość oczekiwana i wariancja

$$E(X) = \sigma\,\Gamma(1+\alpha^{-1}) \qquad \mathrm{var}(X) = \sigma^2\left\{\Gamma(1+2\alpha^{-1}) - (\Gamma(1-\alpha^{-1}))^2\right\}$$

- Funkcja Gamma Eulera

$$\Gamma(p) = \int_0^{+\infty} x^{p-1}e^{-x}dx \qquad \Gamma(1)=1 \qquad \Gamma(0.5)=\sqrt{\pi}$$

$$\Gamma(p) = (p-1)\Gamma(p-1) \qquad \Gamma(n) = (n-1)! \quad n\in N$$

Rozkład Weibulla

- $\int_0^{\sigma} f_X(x)\,dx \approx 0.6321$

  $\longrightarrow$ 63,21% populacji umiera do czasu σ (niezależnie od α).
- Zastosowania: czas życia (ubezpieczenia), zużywanie się elementów w technice, czas dostawy produktu (logistyka), rozkład siły wiatru.
- Generowanie:
  1) Jeżeli zmienna losowa X ma rozkład wykładniczy o parametrze σ to zmienna losowa $Y=X^{(1/\alpha)}$ ma rozkład Weibulla $f_Y(y)$ o parametrach α i σ .
  2) Jeżeli zmienna losowa X ma rozkład jednorodny na przedziale (0,1) to zmienna losowa $Y=\sigma(-\ln(X))^{(1/\alpha)}$ ma rozkład Weibulla $f_Y(y)$ o parametrach α i σ .

Rozkład normalny N(μ,σ^2)

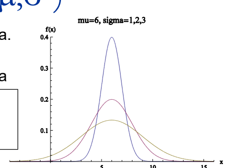
- Zwany jest również rozkładem Gaussa.
- Jest to najważniejszy z rozkładów.
- Funkcja gęstości prawdopodobieństwa

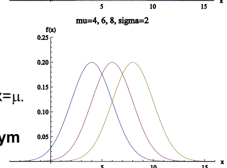
$$\boxed{f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad x\in R}$$

- Rozkład ten jest unormowany:

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

- Rozkład jest symetryczny względem x=μ.
- Dla μ=0 i σ=1 nazywany jest **standardowym rozkładem normalnym** (standardowym rozkładem Gaussa).





Dystrybuanta rozkładu normalnego N(0,1)

- Dystrybuanta rozkładu normalnego
  Nie potrafimy policzyć analitycznie. Można tego uniknąć licząc ją numerycznie, korzystając z tablic dla N(0,1) lub korzystając z przybliżonych wzorów dla N(0,1)

$$F_X^{(1)}(x) = \frac{1}{2} \pm \frac{1}{2}\sqrt{1-e^{\frac{-2x^2}{\pi}}}$$

$$F_X^{(2)}(x) = \frac{1}{2} \pm \frac{1}{2}\sqrt{1-e^{\frac{-2x^2}{\pi}} - \frac{2(\pi-3)}{3\pi^2}x^4e^{\frac{-x^2}{2}}}$$

  gdzie górny znak odpowiada x>0 a dolny x≤0.

- Średnia arytmetyczna $F_X^{(3)}(x) \equiv (F_X^{(1)}(x) + F_X^{(2)}(x))/2$
  przybliża prawdziwą dystrybuantę z dokładnością 0.1%.
- Wzory te mogą posłużyć za punkt wyjścia do metody odwracania dystrybuanty (numerycznego) i napisaniu generatora liczb pseudolosowych o rozkładzie N(0,1)
- Funkcja błędu Gaussa bywa zaimplementowana w niektórych kompilatorach (Java)

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt \qquad \text{wtedy} \qquad F_X(x) = 0.5\left(1 + erf(x/\sqrt{2})\right)$$

Rozkład normalny -własności

- Rolę rozkładu normalnego podkreśla Centralne Twierdzenie Graniczne.
- Zastosowanie: tam gdzie szereg czynników ma wpływ na wielkość X i wpływają na nią w mniej więcej jednakowym stopniu. Jest szeroko stosowany w naukach społecznych, ekonomicznych, biologicznych (dla X lub ln(X)), np. wzrost, ciśnienie krwi, inteligencja, wielkość rocznych opadów deszczu jak również przy szacowaniu niepewności przypadkowych, testach statystycznych. Często jest używany w sytuacjach gdy nie znamy prawdziwego rozkładu. Nie mniej rzadko występuje naprawdę.
- Generowanie, przykładowe metody:
  1) Suma zmiennych o rozkładzie jednorodnym
  2) Transformacja Box-Mullera
  3) Metoda biegunowa (Marsaglia), v_1 i v_2 mają rozkład N(0,1)

$$x,y\in(-1,1) \to z \equiv x^2+y^2$$

$$z<1 \;\Rightarrow\; v_1 \equiv x\sqrt{-2z^{-1}ln(z)} \quad i \quad v_2 \equiv y\sqrt{-2z^{-1}ln(z)}$$

Rozkład normalny -własności

- Wartość oczekiwana i wariancja: $\boxed{E(X)=\mu \qquad var(X)=\sigma^2}$
- **Standaryzacja dowolnego rozkładu** pozwala przechodzić pomiędzy rozkładami normalnymi:
  Jeżeli zmienna losowa X ma rozkład N(μ,σ^2) to zmienna losowa Y=(X-μ)/σ ma rozkład N(0,1) (dowód na ćwiczeniach).
- **„Reguła trzech sigma”** :

k	P(μ–kσ<X<μ+kσ)
1	0.6826895
2	0.9544997
3	0.9973002
4	0.9999367
5	0.9999994

P(μ–kσ<X<μ+kσ)	k
0.50	0.6745
0.80	1.2816
0.90	1.6449
0.95	1.96
0.99	2.5758

k	Surprisal #bits S=$\log_2$(1/P) dla P=0.5 (rzut monetą)
1	1.66
3	8.53
5	20.67

Rozkład normalny -własności

- Szerokość połówkowa FWHM (full width at half maximum)

$$f_X(x_{1/2}) = 0.5 \cdot f_X^{\max} = 0.5 \cdot f_X(\mu) = \frac{1}{2\sigma\sqrt{2\pi}}$$

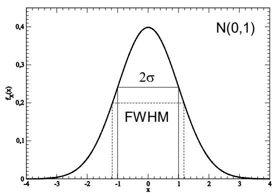
$$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\mu - x_{1/2})^2}{2\sigma^2}} = \frac{1}{2\sigma\sqrt{2\pi}}$$

$$x_{1/2}^{(\pm)} = \mu \pm \sigma\sqrt{2\ln(2)}$$

$$FWHM \equiv \Gamma = x_{1/2}^{(+)} - x_{1/2}^{(-)} =$$

$$= 2\sigma\sqrt{2\ln(2)} \approx 2.355\sigma$$

$$\rightarrow \sigma = \frac{\Gamma}{2.355} = 0.425\Gamma$$



Rozkład Pareto

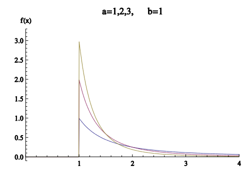
$$f_X(x) = \frac{ab^a}{x^{a+1}} \qquad x > b,\ a > 0, b > 0$$

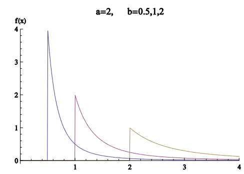
- Parametr b opisuje położenie, a-kształt
- Wartość oczekiwana i wariancja

$$E(X) = \frac{ab}{a-1} \qquad dla\ a > 1$$

$$var(X) = \frac{ab^2}{(a-1)^2(a-2)} \qquad dla\ a > 2$$

- Zastosowanie: rozkład dochodów, rozkład długości plików, wielkość cząstek piasku, czasopisma w bibliotece (rozkład Bradforda)
- Zasada Pareto: np. 20% populacji ma 80% bogactw (wartości zależą od parametrów)
- Pojawia się przy grupowaniu w rangi





Rozkład Gamma

$$f_X(x) = \frac{(\lambda x)^{\alpha-1}\lambda e^{-\lambda x}}{\Gamma(\alpha)} \qquad x \ge 0, \lambda > 0, \alpha > 0 \quad \Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

- Parametr λ opisuje położenie i wielkość, parametr α opisuje kształt.
- Wartość oczekiwana i wariancja

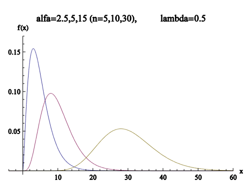
$$E(X) = \frac{\alpha}{\lambda} \qquad var(X) = \frac{\alpha}{\lambda^2}$$

- Z tego rozkładu wyprowadza się inne rozkłady, np:

dla α=1 $\quad f_X(x) = \frac{(\lambda x)^{1-1}\lambda e^{-\lambda x}}{\Gamma(1)} = \lambda e^{-\lambda x}$

dla α=n/2, λ=1/2 $\quad f_X(x) = \frac{\left(\frac{x}{2}\right)^{n/2-1} e^{-\frac{x}{2}}}{2\Gamma(n/2)}$

Jest to rozkład $\chi_n^2(x)$ „Chi-kwadrat o *n* stopniach swobody"



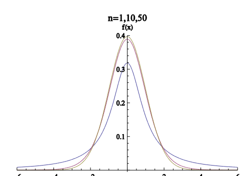
Rozkład t-Studenta

$$f_X(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left((n+1)/2\right)}{\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

- Parametr n=1,2,… opisuje kształt.
- Wartość oczekiwana (dla n>1) i wariancja (dla n>2):

$$E(X) = 0 \qquad var(X) = \frac{n}{n-2}$$

- Dla n=1 jest to rozkład Cauchy
- Dla n>30 dobrym przybliżeniem jest rozkład normalny N(0,n/(n-2))
- Generowanie: Jeżeli zmienna losowa X ma rozkład N(0,1), a zmienna losowa Y ma rozkład $\chi_n^2(y)$ to $Z = X\sqrt{n/Y}$ ma rozkład t-Studenta o n stopniach swobody.



Porównanie rozkładu t-Studenta i rozkładu normalnego