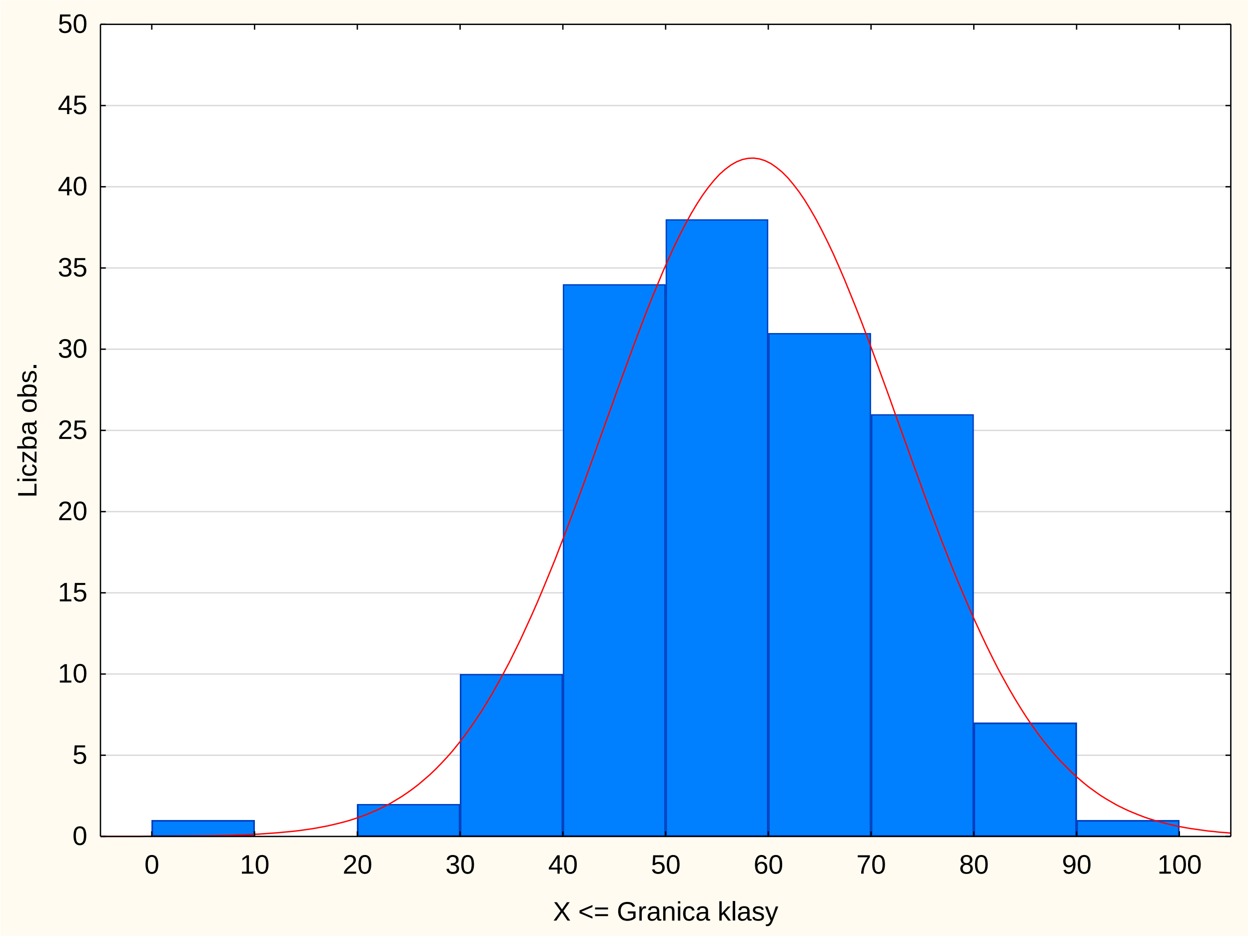
Statystyka i modelowanie w naukach o środowisku

Wykład 2

Rozkład normalny

Histogram vs funkcja gęstości prawdopodobieństwa



Metoda największej wiarygodności

W procesie estymacji na podstawie próbki $x_1, x_2, \ldots, x_n$ wyznaczamy parametry opisujące domniemany rozkład prawdopodobieństwa.

Na podstawie tego rozkładu możemy z kolei określić *a posteriori* (po fakcie) prawdopodobieństwo próbki $x_1, x_2, \ldots, x_n$.

Metoda największej wiarygodności

Parametry należy dobierać tak, aby zmaksymalizować prawdopodobieństwo *a posteriori* próbki, z których je wyznaczamy.

Funkcją wiarygodności nazywać możemy iloczyn prawdopodobieństwa *a posteriori* dla n dostępnych próbek.

$$L = \prod_{i=1}^{n} f(x_i, \lambda) \qquad l = \ln(L) = \sum_{i=1}^{n} \ln f(x_i, \lambda)$$

Metoda największej wiarygodności

Rozważmy próbkę $x_1, x_2, \ldots, x_n$ z rozkładu normalnego $N(\mu, \sigma)$.

Wyznacz estymatory największej wiarygodności parametrów μ i σ.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, -\infty < x < \infty, \sigma > 0$$

Metoda największej wiarygodności

Funkcja największej wiarygodności ma postać:

$$L(\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x_1-m)^2}{2\sigma^2}}\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x_2-m)^2}{2\sigma^2}}\cdot\ldots\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x_n-m)^2}{2\sigma^2}}$$

$$L(\mu,\sigma)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\cdot\sum_{i=1}^{n}(x_i-m)^2}$$

Metoda największej wiarygodności

Szukamy maksimum funkcji największej wiarygodności, czyli (zwykle) zera pochodnej.

Funkcja największej wiarygodności $L(\mu,\sigma)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\cdot\sum_{i=1}^{n}(x_i-m)^2}$

przyjmuje maksimum gdy wykładnik $\sum_{i=1}^{n}(x_i-m)^2$

jest minimalny.

Metoda największej wiarygodności

Estymator parametru μ

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Metoda największej wiarygodności

Funkcja największej wiarygodności ma postać:

$$L(\mu,\sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\left(\sqrt{2\pi}\right)^n \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Metoda największej wiarygodności

Logarytm funkcji największej wiarygodności ma postać:

$$\ln\left(\left(\sqrt{2\pi}\right)^{n}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n} e^{-\frac{1}{2\sigma^{2}}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^{2}}\right)$$

$$\ln\left(\left(\frac{1}{\sigma}\right)^{n} e^{-\frac{1}{2\sigma^{2}}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^{2}}\right)$$

Metoda największej wiarygodności

Pochodną logarytmu funkcji największej wiarygodności przyrównujemy do zera i znajdujemy σ:

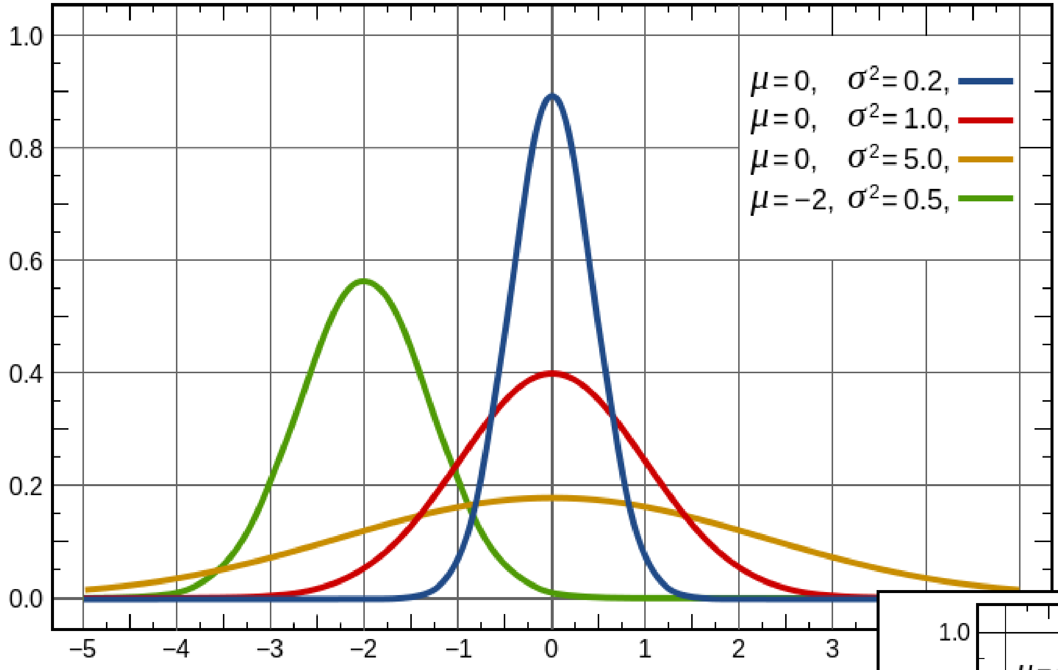
$$-\frac{n}{\sigma}+\frac{1}{\sigma^{3}}\cdot\sum_{i=1}^{n}(x_i-\overline{x})^2=0$$

Metoda największej wiarygodności

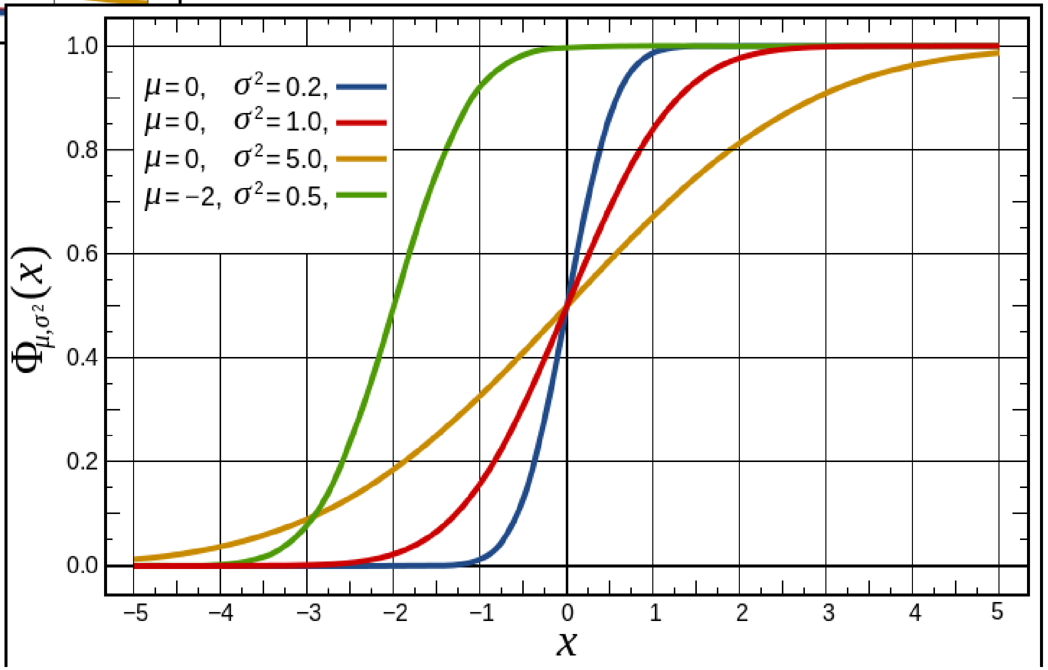
Estymator parametru σ^2:

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$$

Rozkład normalny

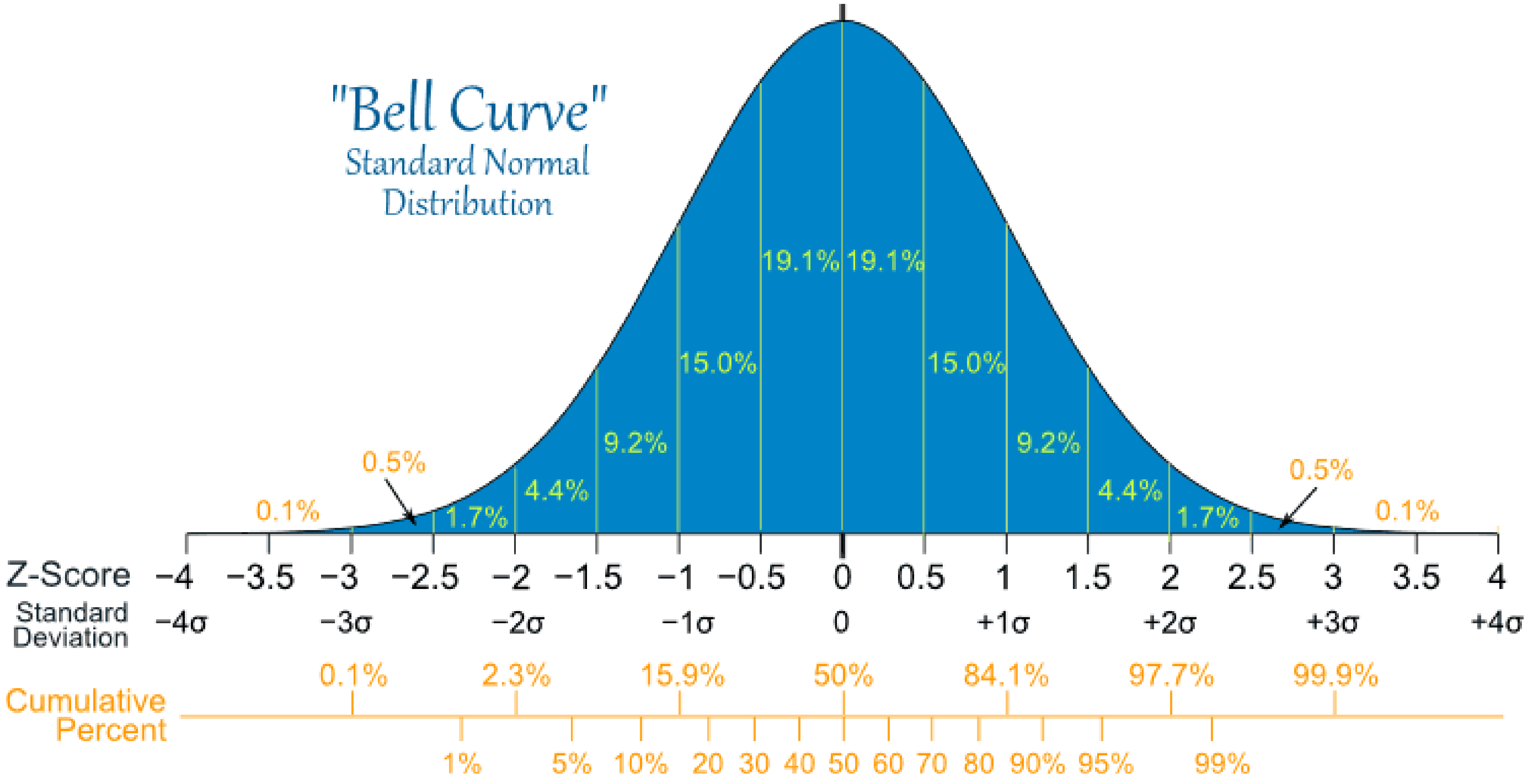


$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



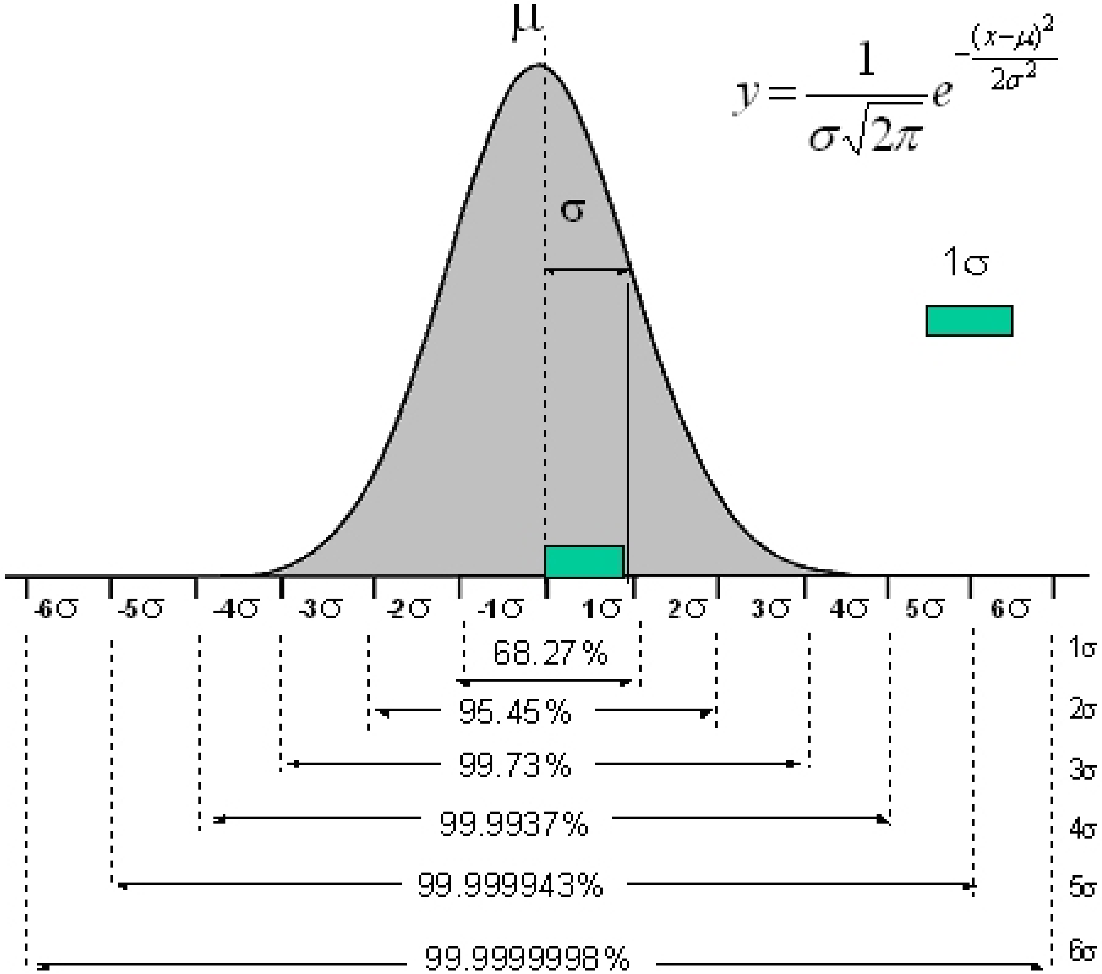
http://en.wikipedia.org/wiki/Normal_distribution

Standardowy rozkład normalny



https://www.mathsisfun.com/data/standard-normal-distribution.html

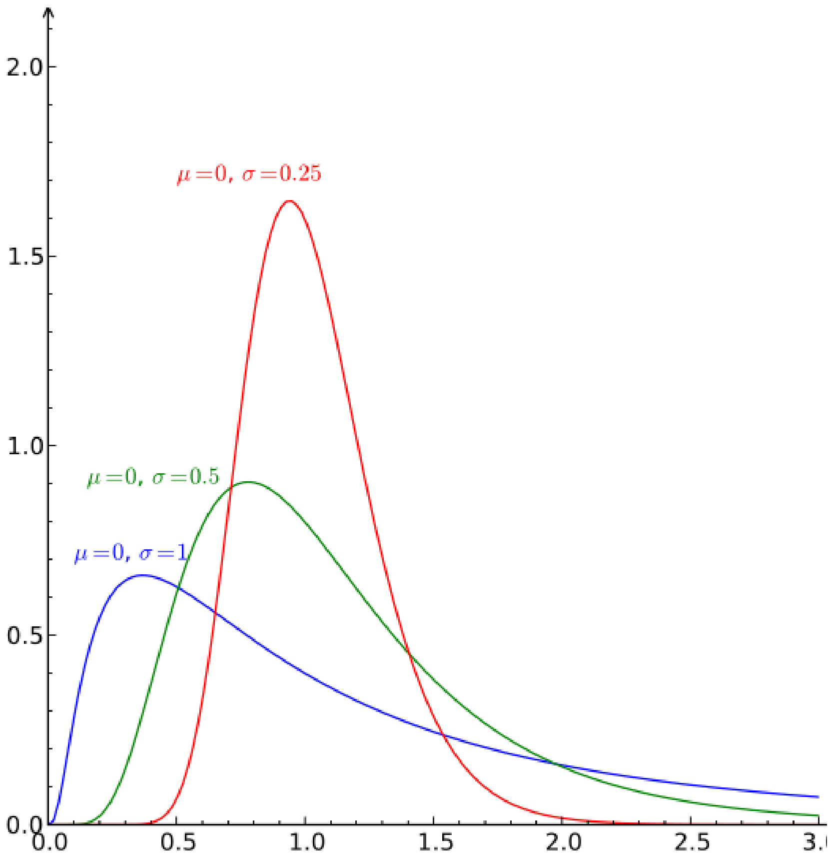
Reguła 3 sigm



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

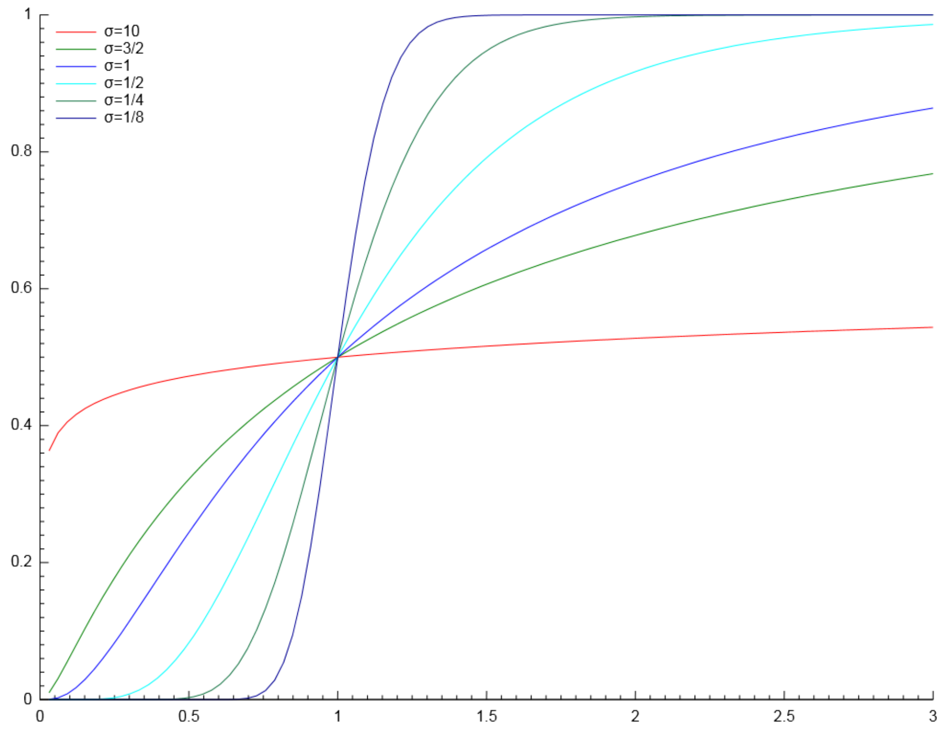
http://icpartnership.com/sixsigma.html

Rozkład log-normaly



$$x \in (0, +\infty)$$

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



http://en.wikipedia.org/wiki/Log-normal_distribution

Miary położenia rozkładu

Średnia arytmetyczna

Mediana rozkładu empirycznego

Wartość modalna (moda, dominanta)

Rozkład normalny

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} \, dt$$

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

$$D^2(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} \, dx = \sigma^2$$