Paweł Tkocz 332168

$$N(\mu, \sigma^2): \; f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , x \in \mathbb{R}$$

Rozwiązaniem zadania jest funkcja obliczająca wartości dystrybuanty w punkcie x.

Jeśli $X \sim N(\mu, \sigma^2)$ to dystrybuantą zmiennej losowej X nazywamy funkcję

$$\boxed{F_X(t) = P(X \leq t)} \quad \text{czyli} \quad F_X(t) = \int_{-\infty}^{t} f(x)\,dx = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$

Chcąc wyznaczyć wartości tej całki można by spróbować dokonać podstawienia

$$\int \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \left\{\begin{matrix} u = \frac{x-\mu}{\sigma} \\ du = \frac{1}{\sigma}dx \\ dx = \sigma\,du \end{matrix}\right\} = \int \frac{1}{\sqrt{2\pi}\,\cancel{\sigma}} e^{-\frac{1}{2}u^2} \cancel{\sigma}\,du = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}u^2}\,du$$

Nie ma jednak prostego sposobu na policzenie całki $\int e^{-\frac{1}{2}u^2}\,du$

Aby poradzić sobie z tym problemem zastosuję metodę całkowania numerycznego poznaną na analizie numerycznej: zastosuję złożony wzór trapezów.

$$\int_a^b f(x)\,dx \approx \frac{1}{2}h\left[f(a) + 2\sum_{i=1}^{n-1} f(a+ih) + f(b)\right] \quad , \text{gdzie } h = \frac{b-a}{n}$$

n - liczba podprzedziałów całkowania

Z tego wzoru korzysta funkcja integ(mi, sigma, a, b, max_b) odpowiedzialna za obliczanie przybliżonej wartości całki z $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ na przedziale $[a, b]$

Aby uzyskać dobrą dokładność i jednocześnie nie czekać długo na obliczenie wyniku, zdecydowałem się na obranie $n = 10000$

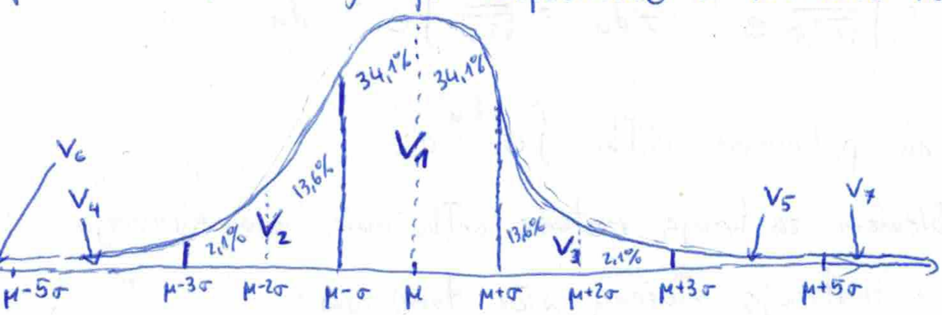
Dodatkowo, ponieważ funkcja ta będzie używana do liczenia dystrybuanty $F_X(t)$, argument max_b wykorzystuję, by upewnić się, że nie obliczam wartości całki dla $x > t$, a w jednym dodatkowym przypadku, by upewnić się że nie liczę całki z pewnego przedziału dwa razy.

Na potrzeby mojego rozwiązania przyjąłem też, że jeśli $a \geq b$ to wartość całki $\int_a^b f(x)\,dx$ wynosi 0.

Paweł Tkocz 332168

Funkcja density(mi, sigma, x) oblicza wartość prawdopodobieństwa w punkcie x, czyli wartość $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma}}$, funkcja ta jest wykorzystywana przez funkcję integ do liczenia wartości całki na przedziale $[a, b]$

Ostatnią funkcją z której korzystam w moim rozwiązaniu jest funkcja dystrybuanta(mi, sigma, x), której zadaniem jest policzenie $F_Z(x)$ dla pewnej zmiennej losowej $Z \sim N(mi, sigma^2)$. W funkcji tej liczę wartości całki na kilku przedziałach, które wynikają bezpośrednio z rozkładu normalnego:



Dla zmiennej losowej w rozkładzie normalnym około 99.7% z wartości $\int_{-\infty}^{\infty} f(x)dx$ to $\int_{\mu-3\sigma}^{\mu+3\sigma} f(x)dx$

Wynika z tego, że licząc przybliżoną wartość całki z zastosowaniem złożonego wzoru trapezów należy zwrócić szczególną uwagę na przedział $[\mu-3\sigma, \mu+3\sigma]$ i jego otoczenie. Dlatego też w funkcji dystrybuanta() liczę kilka całek, na przedziałach:

$V_1 - [\mu-\sigma, \mu+\sigma]$
$V_2 - [\mu-3\sigma, \mu-\sigma]$
$V_3 - [\mu+\sigma, \mu+3\sigma]$
$V_4 - [\mu-5\sigma, \mu-3\sigma]$
$V_5 - [\mu+3\sigma, \mu+5\sigma]$
$V_6 - [\mu-50\sigma, \mu-5\sigma]$
$V_7 - [\mu+5\sigma, \mu+50\sigma]$

obliczając wartości z całek na tych przedziałach przekazuję do funkcji integ(mi, sigma, a, b, max-b) argument max-b = x aby nie policzyć wartości całki dla argumentów większych niż x.

$V_8 - [\mu+50\sigma, x]$
$V_9 - [x-50\sigma, x]$ ← z argumentem max-b = $\mu-50\sigma$, obliczam tę wartość jeśli $x<\mu$ aby zwiększyć dokładność wyniku.

Ostatecznie, przybliżona wartość dystrybuanty $F(x)$ to $v = v_1+v_2+v_3+v_4+v_5+v_6+v_7+v_8+v_9$

W załączonym skrypcie zamieściłem przykładowe wywołanie funkcji dystrybuanta() dla argumentów $\mu=12$ $\sigma=5$ $x=5.5$