Zadanie egzaminacyjne 1 - Tomasz Woszczyński

1 Polecenie

Standardowy rozkład normalny ma gęstość określoną wzorem:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \text{ dla } x \in \mathbb{R}$$

Dla dystrybuanty otrzymujemy wyrażenie:

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

którego to całka nie ma przedstawienia za pomocą funkcji elementarnych.
Dla ustalonego $t \in \mathbb{R}$ obliczyć wartość całki:

$$G(t) = \int_{-\infty}^{t} \exp\left(-\frac{x^2}{2}\right) dx$$

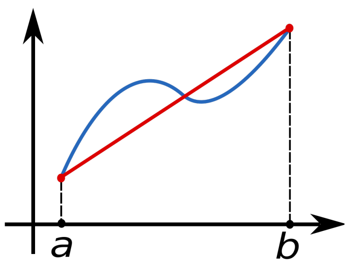
2 Pomysł rozwiązania

Wiemy, że całka dla dystrybuanty $\Phi(t)$ nie ma przedstawienia za pomocą funkcji elementarnych, dlatego skorzystamy z całkowania numerycznego, aby obliczyć przybliżoną wartość całki oznaczonej. Zwykle w celu obliczenia wartości całki nieskończonej korzystamy z przybliżania funkcji całkowanej poprzez inne funkcje, których całki można w łatwy sposób policzyć.

2.1 Wzór trapezów, złożony wzór trapezów

Wzór trapezów jest jednym ze wzorów przydatnych do przybliżonego obliczania całek oznaczonych. Dla funkcji f na przedziale $x \in [a, b]$ możemy oszacować wartość całki $\int_a^b f(x)dx$ za pomocą kwadratury:

$$Q_1(f) = \frac{b-a}{2}\left(f(a) + f(b)\right)$$



Rysunek 1: Zasada działania wzoru trapezów

Niestety wzór trapezów jest dokładny tylko dla funkcji stałych i liniowych, dlatego powinniśmy skorzystać z mocniejszego narzędzia, czyli złożonego wzoru trapezów. Oznacza to, że cały przedział $[a, b]$ dzielimy na równe części i na każdej z części wykorzystujemy wzór trapezów. Otrzymujemy wtedy dla węzłów $t_k := a + h_n k$, $(k = 0, 1, \ldots, n)$, $h_n := \frac{b-a}{n}$ wzór:

$$\int_a^b f(x)dx = T_n(f) + R_n^T$$

gdzie $T_n(f)$ jest złożonym wzorem trapezów, a R_n^T błędem złożonego wzoru trapezów. Złożony wzór trapezów wyraża się wzorem:

$$T_n(f) := h_n \sum_{k=0}^{n}{}'' f(t_k) = h_n(\frac{1}{2}f(t_0) + f(t_1) + \ldots + f(t_{n-1}) + \frac{1}{2}f(t_n))$$

W powyższym wzorze $\sum'' x$ oznacza, że pierwszy i ostatni wyraz mnożymy przez $\frac{1}{2}$.

2.2 Metoda Romberga

Kolejną metodą na numeryczne obliczanie wartości całek oznaczonych jest metoda Romberga. Jest ona rozszerzeniem złożonego wzoru trapezów i daje lepsze przybliżenie całki poprzez zasadniczą redukcję błędu. Niech $n = 2^k$ dla $k \in \mathbb{N}$, $h_k := \frac{b-a}{2^k}$, $x_i^{(k)} := a + ih_k$ dla $i = 0, 1, \ldots, 2^k$. Zdefiniujmy $T_{0,k} := T_{2^k}(f) = h_k \sum_{i=0}^{2^k}{}'' f(x_i^{(k)})$ (złożone wzory trapezów). Kolejne elementy $T_{m,k}$ dla $k = 0, 1, \ldots$ oraz $m = 1, 2, \ldots$ definiujemy rekurencyjnie za pomocą wzoru:

$$T_{m,k} := \frac{4^m T_{m-1,k+1} - T_{m-1,k}}{4^m - 1}$$

Na podstawie tych wartości możemy stworzyć tablicę Romberga:

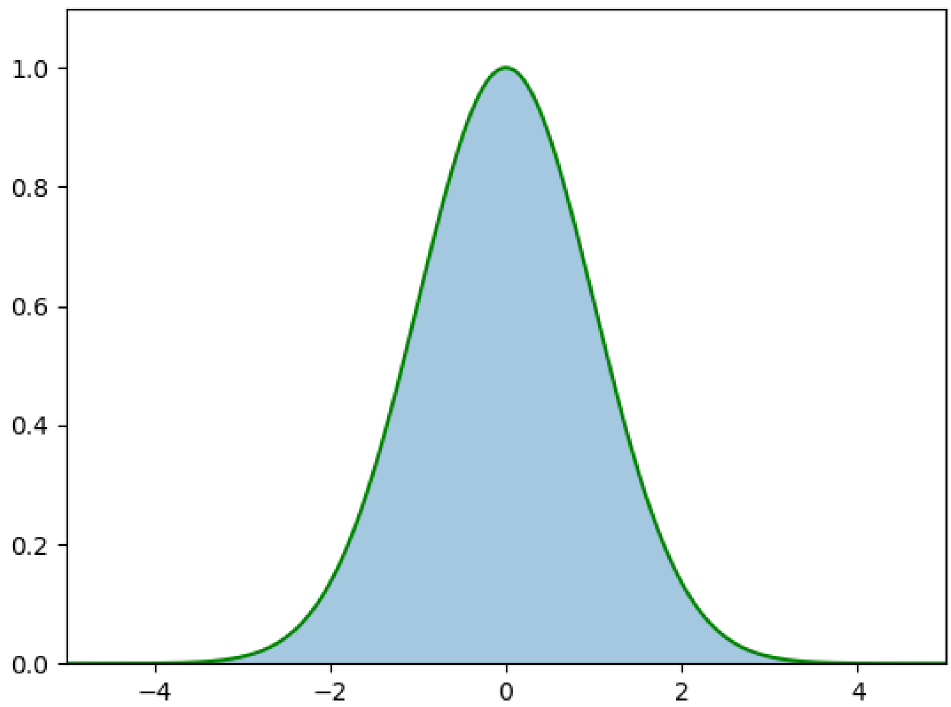
$$\begin{array}{lllllll}
T_{0,0} \\
T_{0,1} & T_{1,0} \\
T_{0,2} & T_{1,1} & T_{2,0} \\
T_{0,3} & T_{1,2} & T_{2,1} & T_{3,0} \\
T_{0,4} & T_{1,3} & T_{2,2} & T_{3,1} & T_{4,0} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\
T_{0,m} & T_{1,m-1} & T_{2,m-2} & \ldots & \ldots & \ldots & T_{m,0}
\end{array}$$

Przeważnie każda kolejna kolumna tablicy Romberga zbiega szybciej do wartości całki, a najszybsza zbieżność występuje na przekątnej tablicy.

3 Rozwiązania zadania

Celem tego zadania jest obliczenie wartości całki

$$G(t) = \int\limits_{-\infty}^{t} \exp\left(-\frac{x^2}{2}\right) dx$$



Rysunek 2: Wykres funkcji podcałkowej $f(x) = \exp\left(-\frac{x^2}{2}\right)$

3.1 Rozwiązanie nr 1

Zauważmy, że $f(x)$ dąży do 0 dla $-\infty$ oraz ∞, a wartości bardzo bliskie zera osiąga dla wszystkich $x \notin [-10, 10]$ (wartość $f(10) \approx 1.92874984 \cdot 10^{-22}$, a dla większych wartości x, ta wartość się jeszcze bardziej zmniejsza). Oznacza to, że możemy całkę z funkcji $f(x)$ przybliżyć w na przykład taki sposób:

$$G(t) = \int\limits_{-\infty}^{t} \exp\left(-\frac{x^2}{2}\right) dx \approx \int\limits_{-50}^{t} \exp\left(-\frac{x^2}{2}\right) dx$$

Wprost ze wzoru pozwalającego obliczyć kolejne wyrazy tablicy Romberga widać, że powstała tablica będzie zawsze różna dla $t \in \mathbb{R}$, ponieważ jeden koniec przedziału (a) mamy w tym momencie ustalony, a więc różnica $b - a$ będzie zawsze różna.

Poszukiwania wartości całki kończymy, gdy dwie ostatnie wartości na przekątnej różnią się o mniej niż oczekiwana dokładność (10^{-8}).
Weźmy więc $t = 1$ i obliczmy wszystkie wartości tablicy:

1	51.000000	15.466532										
2	25.500000	7.733266	5.155511									
4	12.750000	3.866633	2.577755	2.405905								
8	6.375000	1.933320	1.288882	1.202957	1.183863							
16	3.187500	1.257978	1.032863	1.015796	1.012825	1.012154						
32	1.593750	1.966428	2.202578	2.280559	2.300635	2.305685	2.306950					
64	0.796875	2.076131	2.112698	2.106706	2.103947	2.103175	2.102977	2.102927				
128	0.398438	2.100872	2.109119	2.108880	2.108914	2.108934	2.108940	2.108941	2.108941			
256	0.199219	2.106930	2.108949	2.108938	2.108939	2.108939	2.108939	2.108939	2.108939	2.108939		
512	0.099609	2.108437	2.108939	2.108939	2.108939	2.108939	2.108939	2.108939	2.108939	2.108939	2.108939	
1024	0.049805	2.108813	2.108939	2.108939	2.108939	2.108939	2.108939	2.108939	2.108939	2.108939	2.108939	2.108939

Oczekiwaną dokładność otrzymujemy więc po 1025 przybliżeniach funkcji i wartość całki dla $t = 1$ wynosi $G(1) = 2.1089385292093854 \pm 10^{-8}$. Wartością dystrybuanty jest $\Phi(t) = 5.286324946774961 \pm 10^{-8}$.

3.2 Rozwiązanie nr 2

W tym rozwiązaniu skorzystam z faktu z ćwiczeń, a więc tego, że

$$I = \int_{-\infty}^{-\infty} \exp\left(-\frac{-x^2}{2}\right) dx = \sqrt{2\pi}$$

Wiemy też, że funkcja podcałkowa jest symetryczna. Skorzystajmy więc z tego faktu i podzielmy funkcję $G(t)$ na dwa przypadki:

$$G(t) = \int_{-\infty}^{t} \exp\left(-\frac{x^2}{2}\right) dx = \begin{cases} \frac{1}{2}\sqrt{2\pi} + \int\limits_0^t \exp\left(-\frac{x^2}{2}\right) dx & \text{dla } t \geqslant 0 \\ \frac{1}{2}\sqrt{2\pi} - \int\limits_0^t \exp\left(-\frac{x^2}{2}\right) dx & \text{dla } t < 0 \end{cases}$$

Teraz podobnie jak w przypadku rozwiązania 1. możemy wziąć $t = 1$ i obliczyć wartości tablicy Romberga. Ważne jest to, że tablica ta ma całkiem inne wartości niż ta z rozwiązania pierwszego, jako że pracujemy na innym, znacznie mniejszym przedziale.

1	1.000000	0.803265				
2	0.500000	0.842881	0.856086			
4	0.250000	0.852459	0.855651	0.855622		
8	0.125000	0.854834	0.855626	0.855624	0.855624	
16	0.062500	0.855427	0.855624	0.855624	0.855624	0.855624

W tym przypadku oczekiwaną dokładność otrzymujemy już po 17 przybliżeniach funkcji, a więc wykonywane obliczenia są dokładne na tym samym poziomie, jednak wymagają znacznie mniej nakładu czasowego na procesorze. Wynikiem tego obliczenia jest $G(t) = 2.1089385292082303 \pm 10^{-8}$. Wartością dystrybuanty jest $\Phi(t) = 5.286324946772066 \pm 10^{-8}$.

4 Porównanie wyników, podsumowanie

Przedstawione powyżej rozwiązania przybliżają funkcję $G(t)$ z dokładnością do 8 cyfr po przecinku. Wyniki rozwiązań całek $G(t)$ różnią się o $1.1550760348200129 \cdot 10^{-12}$, a różnica obliczonych wartości dystrybuanty wynosi $2.8954616482224083 \cdot 10^{-12}$,a więc są bardzo bliskie sobie, jednak dojście do nich różni się znacząco. Pierwsze podejście jest dość błahe, za $-\infty$ podstawiamy jakąś wartość bliską zeru, jednak to powoduje, że im mniejszą wartość wybierzemy, tym wiecej iteracji będziemy musieli wykonać aby obliczyć wszystkie potrzebne wartości w tablicy Romberga. Drugie podejście jest znacznie lepsze i sprytniejsze. Wykorzystujemy własność znanej nam funkcji oraz operujemy na wartościach łatwych do obliczenia, znalezienie $\sqrt{2\pi}$ nie powinno sprawiać żadnych problemów, jako że dzisiejsze algorytmy są bardzo dokładne. Kolejnym plusem drugiego podejścia jest to, że liczba iteracji znacząco się zmniejsza, dzięki czemu wielokrotne wyliczenie wartości całki $G(t)$ zajęłoby nam znacznie mniej czasu, niż gdybyśmy korzystali z pierwszego rozwiązania.