Pracownia 1 - zad 4

Artur Jankowski, indeks: 317928

16 kwietnia 2021

Streszczenie

W zadaniu dla ustalonego **t** mamy policzyć dystrybuantę rozkładu **t-Studenta** z **k** stopniami swobody. Wykorzystamy metodę **złożonych trapezów** i **metodę Romberga** do obliczenia całki z gęstości oraz algorytmu Lanczosa do obliczenia wartości funkcji gamma.
Mamy gęstość:

$$f(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{k}{2})}\left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

Liczymy dla ustalonego $t \in \mathbb{R}$:

$$G(t) = \int_{-\infty}^{t} f(x)dx$$

Spis treści

1 Liczenie wartości funkcji gamma

Dla ustalonego stopnia swobody **k** możemy jednorazowo obliczyć wartości $\Gamma(\frac{k+1}{2})$ oraz $\Gamma(\frac{k}{2})$. Użyjemy do tego przybliżenia Lanczosa [1]. Można by też użyć przybliżenia Stirlinga (używanego w domyślnym module *Scipy* obliczającym wartości Γ), jednak ze względu na prostotę implementacji wybrałem algorytm Lanczosa.

Porównanie algorytmów					
Gamma	Lanczos	Błąd względny	Biblioteczny	Błąd względny	Wartość
$\Gamma(1)$	0.99999999	$\approx 2.22 * 10^{-16}$	1	0	1
$\Gamma(5)$	23.999999	$\approx 1.48 * 10^{-16}$	24	0	24
$\Gamma(10)$	362880.0000000015	$\approx 4.17 * 10^{-15}$	362880	0	3628800
$\Gamma(\frac{7}{2})$	3.3233509	$\approx 6.49 * 10^{-13}$	3.3233509	$\approx 6.49 * 10^{-13}$	3.32335097045

Wartości pochodzą ze wzoru: $\Gamma(n) = (n-1)!$ oraz ze strony Wikipedii [2] dotyczącej funkcji gamma
Jak widać, algorytm dostatecznie dobrze przybliża wartości funkcji gamma.

2 Pomocne własności

$\mathbf{G(t)}$ to całka niewłaściwa co nie odpowiada nam, gdy chcemy jej wartości liczyć numerycznie. Skorzystamy więc z parzystości funkcji $\mathbf{f(x)}$ tak, by zmienić $\mathbf{G(t)}$ i przybliżać całkę znanymi metodami.

Lemat 1. *$f(x)$ jest funkcją parzystą czyli $f(x) = f(-x)$*

Lemat 2. $G(t) + G(-t) = 1$

Dowód.

$$G(t) + G(-t) = \int_{-\infty}^{t} f(x)dx + \int_{-\infty}^{-t} f(x)dx$$

Korzystamy z parzystości f:

$$\int_{-\infty}^{t} f(x)dx + \int_{-\infty}^{-t} f(-x)dx$$

Dokonujemy podstawienia $x = -u$, $dx = -du$:

$$\int_{-\infty}^{t} f(x)dx + \int_{t}^{\infty} f(u)du$$

$$\int_{R} f(x)dx = 1$$

$\square$

Lemat 3. *Dla $t \geqslant 0$:*

$$G(t) = \frac{1}{2} + \int_0^t f(x)dx$$

Dowód. Z parzystości f(x) wiemy, że:

$$\int_0^\infty f(x)dx = \frac{1}{2}\int_{-\infty}^\infty f(x)dx = \frac{1}{2}$$

Z lematu 2:

$$G(t) = 1 - G(-t)$$

$$G(t) = \int_{-\infty}^\infty f(x)dx - \int_{-\infty}^{-t} f(x)dx$$

Rozbijamy pierwszy czynnik ($t > 0$):

$$G(t) = \int_{-\infty}^{-t} f(x)dx + \int_{-t}^{0} f(x)dx + \int_0^\infty f(x)dx - \int_{-\infty}^{-t} f(x)dx$$

$$G(t) = \int_{-t}^{0} f(x)dx + \frac{1}{2}$$

Ponownie z parzystości:

$$G(t) = \frac{1}{2} + \int_0^t f(x)dx$$

□

Lemat 4. *Dla $t < 0$:*

$$G(t) = \frac{1}{2} - \int_0^{-t} f(x)dx$$

Dowód. Korzystamy z poprzedniego lematu:

$$G(t) = 1 - G(-t) = 1 - \left(\frac{1}{2} + \int_0^{-t} f(x)dx\right) = \frac{1}{2} - \int_0^{-t} f(x)dx$$

□

Zatem policzenie $\mathbf{G(t)}$ sprowadza się do policzenia całki $\int_0^t f(x)dx$ oraz wykonania paru operacji arytmetycznych.

3 Metoda Romberga

Do przybliżenia wartości całki oznaczonej wykorzystamy metodę Romberga [3].
Budujemy pierwszą kolumnę przy pomocy złożonego wzoru trapezów:

$$T_{0,k} = T_{2^k} = \frac{b-a}{2^k} \sum_{i=0}^{2^k}{}'' f(x_i)$$

Następne wyliczamy kolejne na podstawie wzoru rekurencyjnego:

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

```
def romberg(function, k, a, b):
    T = [[0.0]]  # tablica Romberga

    T[0][0] = (b-a)/2.0 * (function(b) + function(a))

    n = 1
    for i in range(1, k+1):  # budujemy pierwsza kolumne
        hn = (b-a)/n
        sum_f = 0
        for j in range(1, n + 1):
            sum_f += hn * function(a + 0.5 * (2 * j - 1) * hn)
        rec = 0.5 * (T[0][i-1] + sum_f)
        n *= 2
        T[0].append(rec)

    for i in range(1, k + 1):   # reszta kolumn liczona rekurencyjnie
        col = []
        for j in range(0, k - i + 1):
            col.append((4**i * T[i-1][j+1] - T[i-1][j])
                           / (4**i - 1))
        T.append(col)

    return T[k][0]
```

Listing 1: Funkcja wyliczająca tablicę Romberga rozmiaru k.

Jako wynik otrzymujemy ostatni wynik tablicy (po przekątnej), czyli przybliżenie z najmniejszym błędem (wyniki w tablicy Romberga zbiegają najszybciej do dokładnej wartości po przekątnej).

4 Liczenie wyniku

Aby policzyć dla danego **t** i **k**:

$$G(t) = \int_{-\infty}^{t} f(x)dx = \int_{-\infty}^{t} \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

Liczymy dla ustalonego **k** wartość:

$$\frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{k}{2})}$$

Przybliżamy **metodą Romberga** całkę:

$$\int_0^{\pm t} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} dx$$

Następnie korzystamy z **Lematu 3** lub **Lematu 4** do policzenia $\mathbf{G(t)}$.

$$t \geqslant 0 \implies \phi(t) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{k}{2})} \int_0^{t} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} dx + \frac{1}{2}$$

$$t < 0 \implies \phi(t) = -\frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{k}{2})} - \int_0^{-t} \left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}} dx + \frac{1}{2}$$

```
import numpy as np
def t_student(k, t, precision):
    const = my_gamma((k+1)/2) / (sqrt(k * np.pi)*my_gamma(k/2))

    def func(x):
        return np.power((1 + x**2/k), -((k+1)/2))

    if t >= 0:
        return const * romberg(func, precision, 0, t) + 1.0/2
    else:
        return -(const * romberg(func, precision, 0, -t)) + 1.0/2
```

Listing 2: Liczy *G(t)* dla k stopni swobody przy pomocy tablicy Romberga rozmiaru *precision*

5 Porównanie metody złożonych trapezów i metody Romberga

Tablica błędu względnego dla $k = 1$ oraz $t = 0.5$, oraz precyzji metody Romberga równej 6 (czyli w metodzie trapezów mamy $2^6 = 64$ węzły).
Wtedy dla całki:

$$\int_0^{0.5} \left(1 + x^2\right)^{-1} = arctan\left(\frac{1}{2}\right) dx \approx 0.4636476090008061$$

2.94×10^{-2}						
7.23×10^{-3}	1.67×10^{-4}					
1.80×10^{-3}	1.08×10^{-5}	4.17×10^{-7}				
4.49×10^{-4}	6.75×10^{-7}	5.14×10^{-9}	1.18×10^{-8}			
1.12×10^{-5}	4.21×10^{-8}	8.94×10^{-11}	9.18×10^{-12}	3.72×10^{-11}		
2.80×10^{-5}	2.63×10^{-9}	1.42×10^{-12}	3.04×10^{-14}	5.38×10^{-15}	3.10×10^{-14}	
7.02×10^{-6}	1.64×10^{-10}	2.27×10^{-14}	4.78×10^{-16}	3.59×10^{-16}	3.59×10^{-16}	3.59×10^{-16}

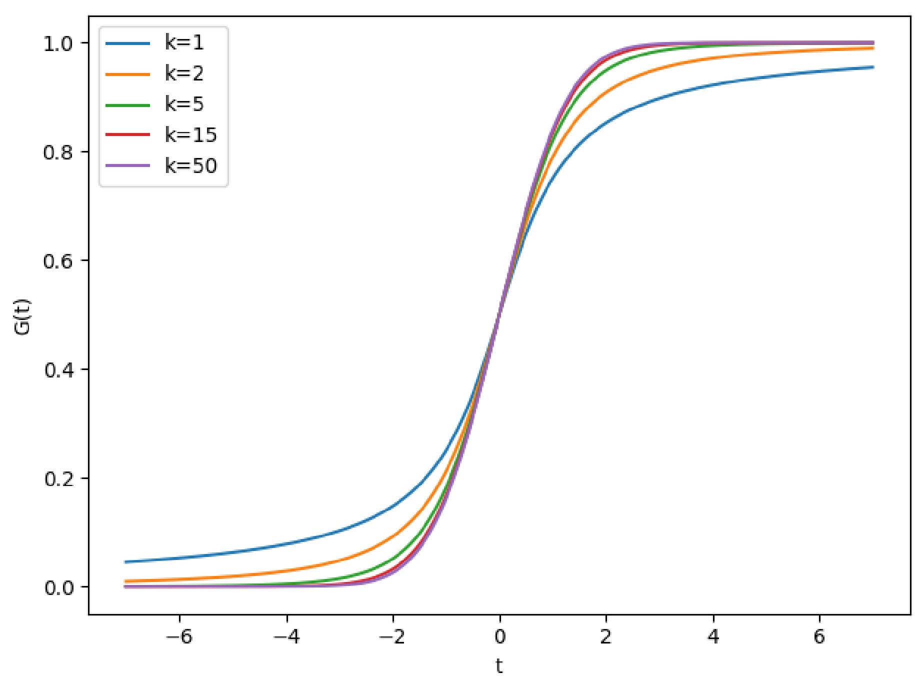
Widzimy, że metoda Romberga zbiega o wiele szybciej od metody trapezów. *(Wyniki w tabeli zostały obcięte do 3 cyfr znaczących w celu lepszej prezentacji)*

6 Wyniki

Parę wyników (precision = 10), porównanych z wynikami z "kalkulatora" [4].

t	k	Wynik algorytmu	Wartość G(t)	Błąd Względny
1	1	0.74999999...	0.75	$\approx 1.48 * 10^{-16}$
-1	1	0.2500000000000001	0.25	$\approx 4.44 * 10^{-16}$
0.5	1	0.6475836176504332	0.64758362	$\approx 3.63 * 10^{-9}$
13	1	0.9755627480278023	$\approx$ 0.97556275	$\approx 2.02 * 10^{-9}$
0	2	0.5	0.5	0
1	2	0.7886751345948	$\approx$ 0.78867513	$\approx 5.83 * 10^{-9}$
-5	2	0.01887477567531	$\approx$ 0.01887478	$\approx 2.30 * 10^{-7}$
1	10	0.8295534338489704	$\approx$ 0.82955343	$\approx 4.64 * 10^{-9}$
6	10	0.9999339455698227	$\approx$ 0.99993395	$\approx 4.43 * 10^{-9}$

Jako że wyniki, z którymi porównujemy, są dokładne, do 8 cyfr znaczących pojawiają się pewne błędy zaokrągleń. Mimo to widać, że nasza funkcja dobrze przybliża wartości G(t).



Rysunek 1: Rozkłady t-Studenta dla wybranych k - stopni swobody.

Literatura

[1] *Algorytm Lanczosa do liczenia wartości funkcji gamma*, Wikipedia.

[2] *Funkcja gamma (oraz wybrane wartości)*, Wikipedia.

[3] *Metoda Romberga*, Numerical Analysis Mathematics of Scientific Computing, David R. Kincaid, E. Ward Cheney, rodział 7.4, str. 502-507.

[4] *Soper, D.S. (2021). Cumulative Distribution Function (CDF) Calculator for the t-Distribution. Dostępny na*, https://www.danielsoper.com/statcalc.