Data: 02.04.2025

Laboratorium 5

Autorzy: Kosman Mateusz, Ludwin Bartosz

Temat laboratorium:

Celem niniejszego zadania jest analiza numeryczna problemu aproksymacji danych historycznych dotyczących populacji Stanów Zjednoczonych w latach 1900–1980, z zastosowaniem metody najmniejszych kwadratów oraz ekstrapolacji wyników do roku 1990. W sprawozdaniu podejmujemy próbę wyznaczenia optymalnego stopnia wielomianu interpolacyjnego dla badanych danych przy użyciu zarówno empirycznych miar błędu (błąd względny), jak i kryterium informacyjnego Akaikego (AICc). Dodatkowo analizie podlega ciągła aproksymacja funkcji $f(x) = \sqrt{x}$ w przedziale [0, 2] przy wykorzystaniu wielomianów Czebyszewa, co pozwala na porównanie efektywności metody analitycznej oraz numerycznych technik aproksymacyjnych.

Treść:

Zadanie 1. Wykonaj aproksymację średniokwadratową punktową populacji Stanów Zjednoczonych w przedziale [1900,1980] wielomianami stopnia m dla 0 ≤ m ≤ 6.

a. Dla każdego m dokonaj ekstrapolacji wielomianu do roku 1990. Porównaj otrzymaną wartość z prawdziwą wartością dla roku 1990, wynoszącą 248 709 873.
b. Czy stopień m odpowiadający najmniejszej wartości AIC_c pokrywa się z wartością z poprzedniego podpunktu?

Zadanie 2. Wykonaj aproksymację średniokwadratową ciągłą funkcji $f(x) = \sqrt{x}$ w przedziale $[0, 2]$ wielomianem drugiego stopnia, używając wielomianów Czebyszewa.

Rozwiązanie:

Zadanie 1. a):

Wykonamy aproksymację populacji USA dla różnych stopni wielomianu (od 0 do 6 włącznie). W tym celu użyjemy równania:

$$A^T A c = A^T y \qquad (1)$$

gdzie A - macierz Vandermonde'a o $m + 1$ rzędach, c - szukane współczynniki wielomianu, y - wartości populacji w USA w danym roku.
Po obliczeniu współczynników tworzymy wielomian ze wzoru:

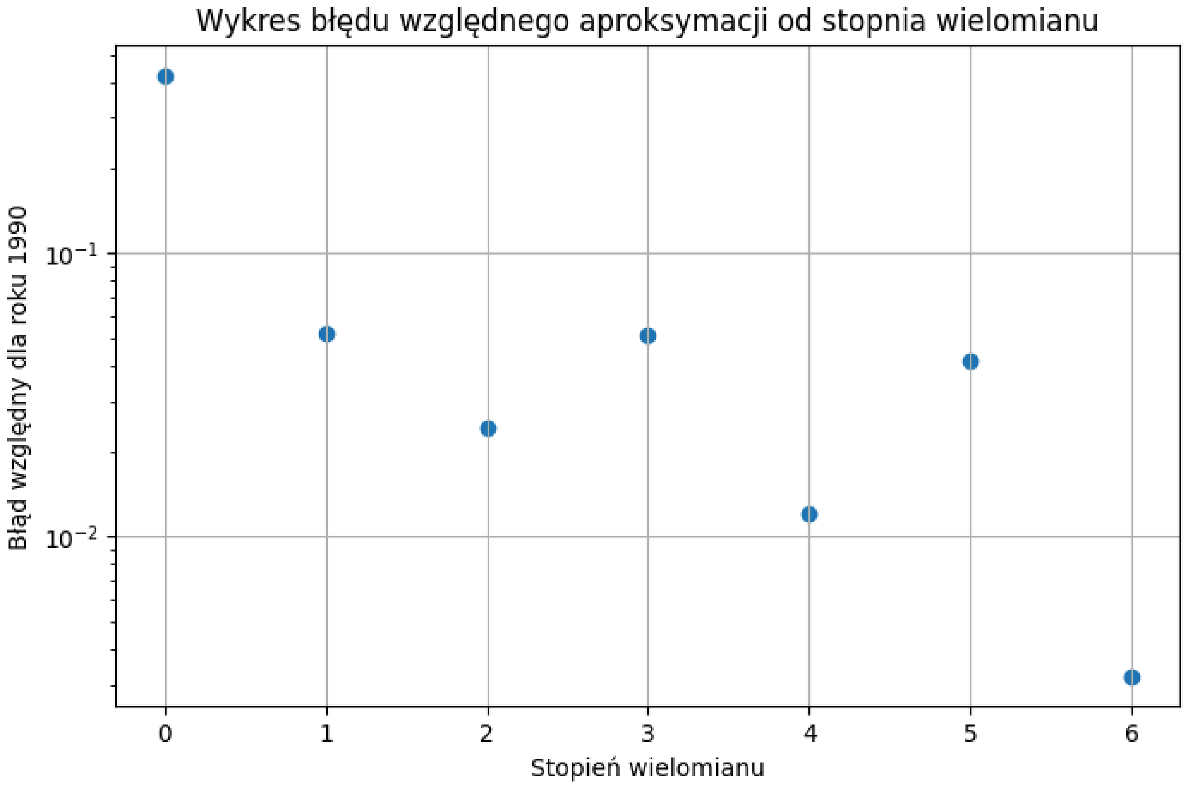
$$p(x) = \sum_{j=0}^{n} c_j x^j \qquad (2)$$

gdzie c_j oznacza współczynnik przy j-tym stopniu x, a następnie podstawiamy argument do wielomianu aby ekstrapolować rok 1990. W kodzie uzyskamy ten rezultat za pomocą funkcji $np.polyval$, oraz wykonamy powyższe operacje dla wszystkich wielomianów (ze stopniami od 0 do 6 włącznie).
Następnie dla każdej otrzymanej w ten sposób wartości liczymy błąd względny, otrzymując wartości:

Stopień wielomianu m	Ekstrapolowana wartość dla roku 1990	Błąd względny
0	143369177	0.42355
1	235808109	0.05187
2	254712945	0.02414
3	261439111	0.05118
4	251719359	0.0121
5	259115342	0.04184
6	249510782	0.00322

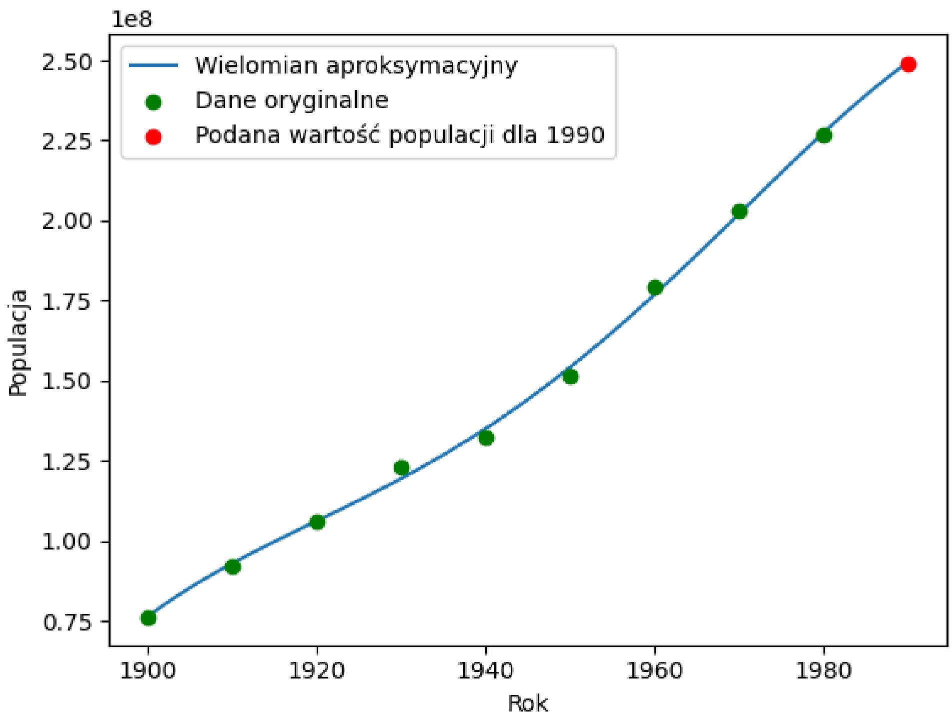
Tabela 1

Sporządzając wykres stopnia wielomianu od błędu względnego otrzymujemy:



Wykres 1

Wyraźnie widać że wartość z najmniejszym błędem względnym została otrzymana dla stopnia wielomianu równego 6. Potwierdziła to również funkcja $find\_smallest\_error$. Rysując wykres tego wielomianu otrzymujemy:



Wykres 2

Jak można zauważyć na wykresie, wielomian bardzo dobrze wpasował się zarazem w oryginalne dane, jak i w podaną w zadaniu prawdziwą wartość dla roku 1990.

Zadanie 1. b):

Do wyboru optymalnego stopnia wielomianu można posłużyć się kryterium informacyjnym Akaikego. Jest ono dane wzorem:

$$AIC = 2k + n\,ln\left(\frac{\sum_{i=1}^{n}[y_i - \hat{y}(x_i)]^2}{n}\right) \tag{3}$$

gdzie $k = m + 1$, $y_i = (1, 2, ..., n)$ oznacza prawdziwą liczbę osób w roku x_i, natomiast $\hat{y}(x_i)$ liczbę osób przewidywaną przez model. Zważywszy jednak na mały rozmiar próbki ($\frac{n}{k} < 40$), należy użyć wzoru z czynnikiem korygującym:

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1} \tag{4}$$

Korzystając ze wzoru (4) otrzymujemy:

Stopień wielomianu m	0	1	2	3	4	5	6
Wartość kryterium Akaike'a	321.01098	289.05648	279.45337	284.8804	292.67239	319.82799	387.93123

Tabela 2

Najmniejszą wartość kryterium otrzymaliśmy dla $m = 2$, co nie pokrywa się z empirycznie uzyskaną wartością ($m = 6$) w podpunkcie a).

Zadanie 2.

W tym zadaniu do aproksymacji wartości funkcji f wykorzystujemy wielomiany Czebyszewa drugiego stopnia.
Wielomian Czebyszewa dla $x \in [-1, 1]$ definiujemy następująco:

$$T_0(x) = 1,$$

$$T_1(x) = x,$$

$$T_k(x) = 2xT_{k-1}(x) - T_{k-1}(x).$$

Rozważamy jedynie $k \leq 2$, stąd nie korzystamy ze wzoru rekurencyjnego w implementacji.
Iloczyn skalarny dwóch funkcji z wagą definiujemy następująco:

$$< f, g >= \int_{-1}^{1} w(x)f(x)g(x)dx$$

Wielomiany Czebyszewa są **ortogonalne**, stąd wiemy, że $T_i(x)T_j(x) = 0$ dla $i \neq j$. Dla $i = j$ mamy:

$$i = j = 0: < T_i, T_j >= \pi$$

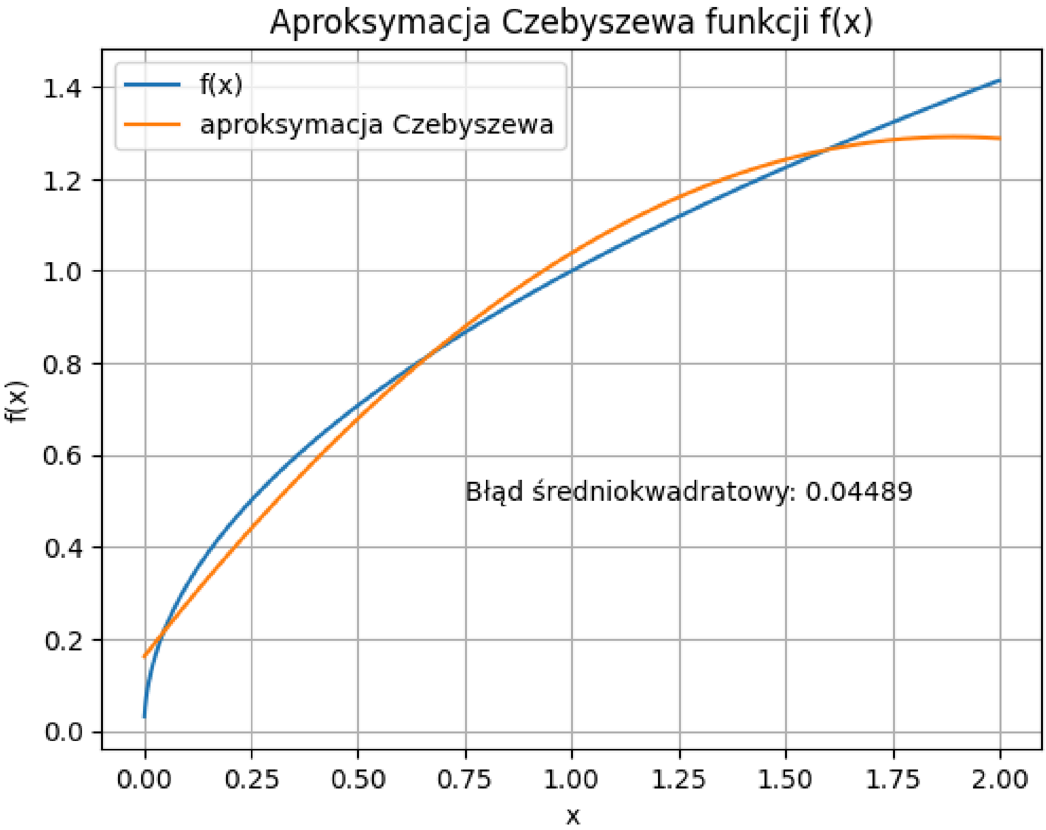
$$i = j \neq 0: < T_i, T_j >= \frac{\pi}{2}$$

Dla $< f, T >$ mamy $w(x) = \frac{1}{\sqrt{1-x^2}}$.

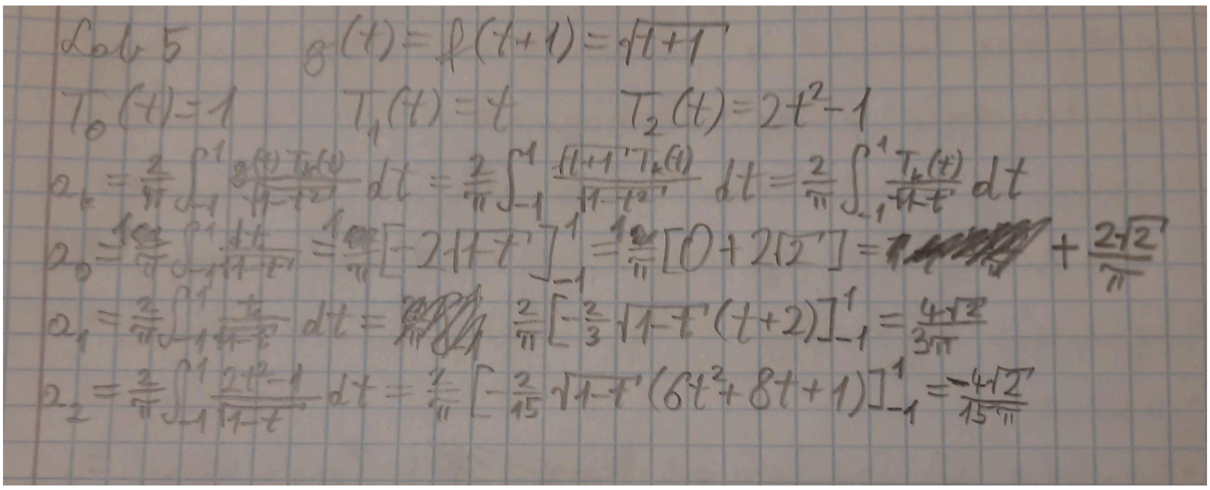
Współczynniki Czebyszewa wynoszą: $c_k = \frac{<f,T_k>}{<T_k,T_k>}$

Ostatecznie nasza aproksymacja p wynosi: $p(x) = \sum_{k=0}^{2} c_k T_k(x)$

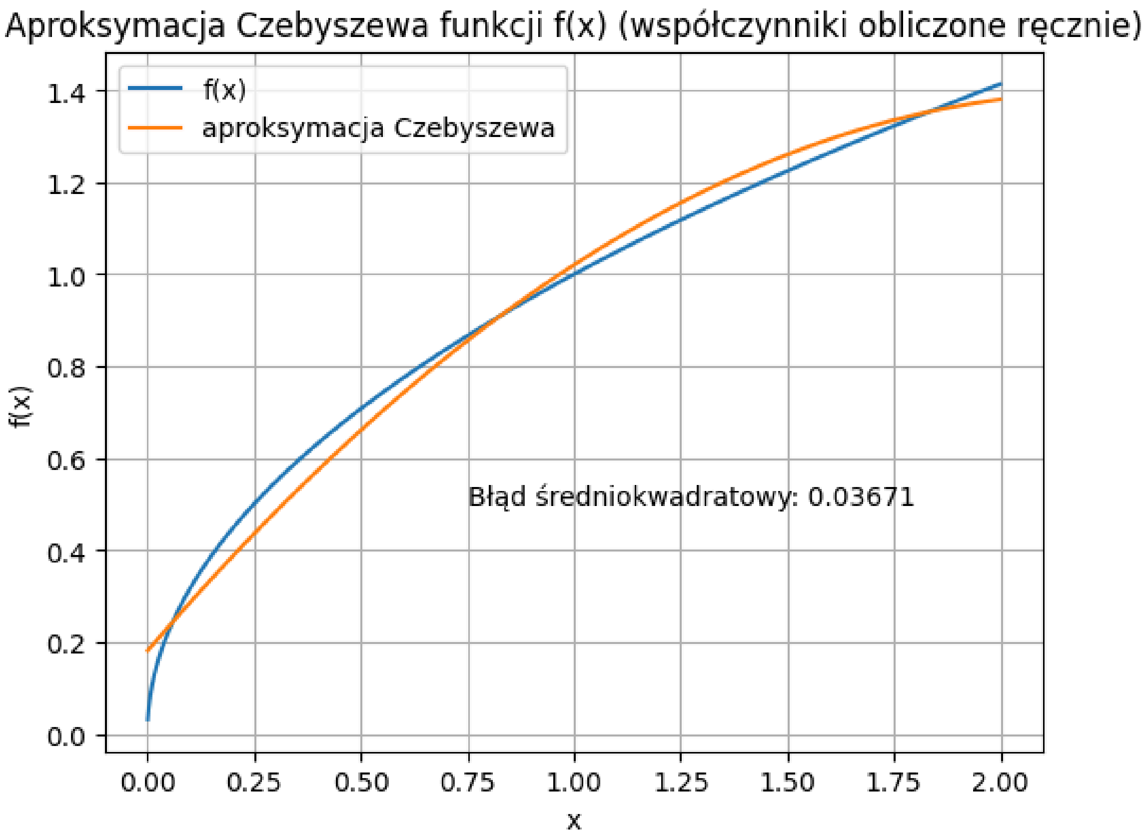
Aproksymacja Czebyszewa prezentuje się następująco względem funkcji f:



Wykres 3 - Aproksymacja Czebyszewa funkcji f(x)



Rysunek 1 - obliczenia współczynników Czebyszewa



Wykres 4 - Aproksymacja Czebyszewa funkcji f(x) (współczynniki obliczone ręcznie)

Widać, że błąd średniokwadratowy dla współczynników obliczonych numerycznie jest wyższy niż dla tych obliczonych analitycznie. Źródłem większego błędu dla współczynników obliczonych numerycznie może być przycięcie dziedziny o 2^{-10} przy jej końcach, co było konieczne ze względu na zerowanie się mianownika funkcji $w(x)$ dla $x \in \{-1, 1\}$.

Wnioski:

- Aproksymacja jest skuteczniejszą techniką niż interpolacja do przewidywania nieznanych wartości. W szczególności jest przydatna, gdy nie zależy nam, aby funkcja aproksymująca przecinała dokładnie zadane punkty, lecz znajdowała się w ich pobliżu.
- Kryterium informacyjne Akaikego daje optymalny, jednak nie zawsze najlepszy stopień wielomianu.
- Do obliczeń wykonywanych na bardzo dużych liczbach lepiej używać typu Double niż Int, aby uniknąć zjawiska przepełnienia (overflow).
- Aproksymacja Czebyszewa jest wydajnym algorytmem (złożoność czasowa $O(nk)$, gdzie n jest liczbą próbek, a k najwyższym stopniem wielomianu Czebyszewa) wyznaczającym aproksymację ciągłą.

- Obliczanie numeryczne całek funkcji dążących do nieskończoności przy krańcach dziedziny wymaga przycięcia tej dziedziny. Skutkuje to jednak mniejszą dokładnością obliczeń względem wyprowadzania całek analitycznie.

Źródła:

- Prezentacje z MS Teams (zespół MOwNiT 2025)
- pythonnumericalmethods.studentorg.berkeley.edu
- https://en.wikipedia.org/wiki/Chebyshev_polynomials