

Метод прогнозування на основі усереднення поліноміальної екстраполяційної послідовності

<https://doi.org/10.31713/MCIT.2025.102>

Юрій Турбал

Кафедра комп'ютерних наук та прикладної
математики
Національний університет водного господарства та
природокористування
Рівне, Україна
turbaly@gmail.com

Маріана Турбал

Кафедра комп'ютерних наук та прикладної
математики
Національний університет водного господарства та
природокористування
Рівне, Україна
Turbal.mariana.1@gmail.com

Олександр Кубай

Кафедра комп'ютерних наук та прикладної
математики
Національний університет водного господарства та
природокористування
Рівне, Україна
o.v.kubai@nuwm.edu.ua

Денис Смірнов

Аспірант,
Національний університет водного господарства та
природокористування
Рівне, Україна
d.a.smirnov@nuwm.edu.ua

Максим Мельничук

Аспірант,
Національний університет водного господарства та природокористування
Рівне, Україна
m.o.melnychuk@nuwm.edu.ua

Анотація — У статті пропонується алгоритм, який значно розширить сферу застосування поліноміальних прогнозів. Цей алгоритм базується на процедурі побудови послідовності всіх поліноміальних прогнозів на основі відомих даних та аналізу цих прогнозів. Потім уточнений прогноз будується як середнє арифметичне певної кількості перших елементів послідовності. У статті досліджуються похибки екстраполяції та запропоновано алгоритм знаходження оптимальної кількості елементів послідовності поліноміальних прогнозів, для яких будується усереднення. Числові результати, представлені в статті, демонструють ефективність методики прогнозування на основі усереднення поліноміальних прогнозів. Зокрема, для тестових прикладів відносна похибка становила близько 2%.

Ключові слова — спостереження; прогнозування; екстраполяція; метод; сітка; алгоритм; многочлен; похибка.

I. ВСТУП

Наука про дані сьогодні є надзвичайно популярною. Всі звикли говорити про великі дані, однак проблеми, пов'язані з малими даними, є не менш важливими. Однією з фундаментальних проблем в науці про дані є проблема прогнозування. Задачі прогнозування часових рядів є актуальними і сьогодні, незважаючи на наявність великої кількості

відомих методів та підходів. Ці задачі тісно пов'язані з екстраполяцією та інтерполяцією функцій.

В той час, коли задачі інтерполяції функцій добре вивчені, існує низка ефективних інтерполяційних методів, проблема екстраполяції залишається в багатьох аспектах відкритою. Адже реальні процеси часто мають суттєву стохастичну складову, експериментальні дані можуть бути неповними, обмеженими, спотвореними. Математичні методи, що використовуються для побудови прогнозів, обмежені проблемою адекватності математичної моделі, ступенем стохастичності процесу, що досліджується. Адже стохастичний процес неможливо спрогнозувати при одному експерименті взагалі і тоді говорять про певні усереднені характеристики.

На практиці часто виникають невеликі часові ряди спостережень для яких проблематично визначити адекватну математичну модель, зокрема, через недостатність інформації. Це стосується, наприклад, різноманітних рядів економічної динаміки, коли ступінь стохастичності залежить від інтервалів між спостереженнями: при великих інтервалах процес можна розглядати як стохастичний, при малих – умовно детермінований, якщо враховувати певну інерційність економічних систем. При цьому задача прогнозування таких процесів часто є критично важливою, особливо якщо

це стосується військової галузі. Таким чином, дослідження, присвячені екстраполяційним задачам є актуальними.

II. ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ

У контексті короткострокового прогнозування детермінованих процесів актуальною є задача екстраполяції функції, заданої таблично. Серед численних методів екстраполяції особливе місце посідають підходи, що ґрунтуються на використанні інтерполяційних многочленів. Проблема такого типу задач висвітлена у працях [1] - [5].

Питання наближення неперервних і диференційовних функцій многочленами є одним із класичних у теорії апроксимації. Зокрема, відомі теореми Вейерштрасса, Дзядика та Тімана [1] доводять можливість наближення будь-якої неперервної функції на скінченному інтервалі многочленом із довільно малою похибкою. На основі цих результатів сформульовано підхід до екстраполяції, який часто позначають принципом “добре проінтерполюй — обережно проекстраполюй”.

У межах такого підходу найчастіше застосовуються інтерполяційні многочлени Ньютона другого виду [2] або узагальнені інтерполяційні многочлени, побудовані на базі систем функцій Чебишева [3], [4], зокрема експоненційних і тригонометричних. Окрім цього, значного поширення набули методи екстраполяції, що базуються на аналізі трендів [5], радіальних базисних функціях, сплайнах (кубічних, В-сплайнах) та кривих типу Безьє, які не проходять безпосередньо через точки інтерполяції.

У сучасних дослідженнях активно застосовуються також статистичні та імовірнісні підходи [5], серед яких варто відзначити геометричні методи, що враховують середньоарифметичні співвідношення елементів часових рядів, а також байєсівські процедури прогнозування. Вагоме місце у задачах прогнозування займають методи машинного навчання та штучних нейронних мереж.

Варто зазначити, що більшість відомих методів ґрунтуються на певному припущенні щодо типу функції, яка описує поведінку процесу. Очевидно, що для будь-якої множини точок існує безліч кривих, які можуть точно проходити через них або наближати їх із певною точністю. Тому складно обґрунтовано стверджувати, що саме одна з таких кривих є істинним законом, який описує досліджуване явище та забезпечує достовірне прогнозування його подальшої еволюції.

Практичні результати свідчать, що екстраполяція на основі многочленів часто демонструє низьку ефективність, особливо за межами інтервалу інтерполяції. Це зумовлює необхідність пошуку способів удосконалення поліноміальних методів та їх поєднання з іншими підходами. Метою даного дослідження є спроба інтеграції методів поліноміального прогнозування зі статистичними підходами та технологіями інтелектуального аналізу даних, що дозволяє підвищити точність прогнозів і розширити практичні можливості класичних методів екстраполяції.

III. АЛГОРИТМ ПРОГНОЗУВАННЯ

Розглянемо спочатку декілька необхідних допоміжних тверджень.

Твердження 1. Нехай маємо рівномірну сітку з кроком Δ , відомі значення деякої функції $f_i = f(i\Delta)$, $i = n, n-1, \dots, n-m+1$. Тоді прогнозне значення функції в точці $n+1$, побудоване на основі інтерполяційного многочлена степеня $m-1$, може бути знайдене на основі співвідношення:

$$f_{n+1}^m = \sum_{k=1}^m (-1)^{k-1} C_m^k f_{n-k+1} \quad (1)$$

Це твердження доведено в роботі [1].

Твердження 2. Має місце співвідношення:

$$\sum_{i=1}^{m-1} f_{n+1}^i = \sum_{k=2}^m (-1)^k C_m^k f_{n-k+2} \quad (2)$$

Розглянемо доведення. Маємо:

$$\begin{aligned} \sum_{k=2}^m (-1)^k C_m^k f_{n-k+2} &= |l = k-1| = \\ &= \sum_{l=1}^{m-1} (-1)^{l+1} C_m^{l+1} f_{n-l+1}, \\ \sum_{i=1}^{m-1} f_{n+1}^i &= \sum_{i=1}^{m-1} \sum_{l=1}^i (-1)^{l-1} C_i^l f_{n-l+1} = \\ &= \sum_{l=1}^{m-1} f_{n-l+1} (-1)^{l-1} \sum_{i=l}^{m-1} C_i^l. \end{aligned}$$

Скористаємося відомим комбінаторним співвідношенням:

$$\sum_{k=i}^n C_k^i = C_{n+1}^{i+1}, i = 0, 1, 2, \dots$$

Тоді виконується рівність:

$$\sum_{i=l}^{m-1} C_i^l = C_m^{l+1}.$$

Отже,

$$\begin{aligned} \sum_{l=1}^{m-1} f_{n-l+1} (-1)^{l-1} \sum_{i=l}^{m-1} C_i^l &= \sum_{l=1}^{m-1} f_{n-l+1} (-1)^{l-1} C_m^{l+1} \\ &= \sum_{l=1}^{m-1} f_{n-l+1} (-1)^{l+1} C_m^{l+1} \end{aligned}$$

Отримали співвідношення (2).

Розглянемо детальніше формулу (1) та виконаємо серію перетворень:

$$\begin{aligned} f_{n+1}^m &= \sum_{k=1}^m (-1)^{k-1} C_m^k f_{n-k+1} = \\ &= C_m^1 f_n + \sum_{k=2}^m (-1)^{k-1} C_m^k f_{n-k+1} = \\ &= C_m^1 f_n + \sum_{k=2}^m (-1)^{k-1} C_m^k f_{n-k+1} - \\ &\quad - \sum_{k=2}^m (-1)^k C_m^k f_{n-k+2} + \sum_{k=2}^m (-1)^k C_m^k f_{n-k+2} = \\ &= C_m^1 f_n + \sum_{k=2}^m (-1)^{k-1} C_m^k (f_{n-k+1} - f_{n-k+2}) - \\ &\quad - \sum_{k=2}^m (-1)^k C_m^k f_{n-k+2} = \\ &= C_m^1 f_n + \sum_{k=2}^m (-1)^{k-1} C_m^k (f_{n-k+1} - f_{n-k+2}) - \sum_{i=1}^{m-1} f_{n+1}^i. \end{aligned}$$

Звідси отримуємо:

$$0 = C_m^1 f_n + \sum_{k=2}^m (-1)^{k-1} C_m^k (f_{n-k+1} - f_{n-k+2}) - \sum_{i=1}^m f_{n+1}^i, \\ + \frac{1}{m} \sum_{k=2}^m (-1)^{k-1} C_m^k (f_{n-k+1} - f_{n-k+2}). \quad (3)$$

Будемо шукати невідоме прогнозне значення f_{n+1} у вигляді:

$$f_{n+1} = \frac{1}{m} \sum_{i=1}^m f_{n+1}^i + \varepsilon(m) \quad (4)$$

Тоді з врахуванням (4) з співвідношення (3) отримуємо:

$$f_{n+1} - \varepsilon(m) = f_n + \frac{1}{m} \sum_{k=2}^m (-1)^{k-1} C_m^k (f_{n-k+1} - f_{n-k+2}), \\ f_{n+1} - f_n - \varepsilon(m) = \frac{1}{m} \sum_{k=2}^m (-1)^k C_m^k (f_{n-k+2} - f_{n-k+1}), \\ \varepsilon(m) = \sum_{k=1}^m (-1)^{k-1} C_m^k (f_{n-k+2} - f_{n-k+1}), \\ \varepsilon(m) = \sum_{k=1}^m (-1)^{k-1} C_m^k f_{n-k+2} - \sum_{k=1}^m (-1)^{k-1} C_m^k f_{n-k+1}. \quad (5)$$

З урахуванням (1) умову (5) можемо записати у вигляді:

$$\varepsilon(m) = f_{n+2}^m - f_{n+1}^m. \quad (6)$$

Співвідношення (6) є фактично необхідною умовою виконання (4).

Нехай має місце умова (6). Тоді з умови (5) отримуємо співвідношення:

$$f_{n+1} = \varepsilon(m) + f_n + \frac{1}{m} \sum_{k=2}^m (-1)^{k-1} C_m^k (f_{n-k+1} - f_{n-k+2})$$

З урахуванням тотожності (5) отримуємо умову (6). Отже, виконання рівності (6) є необхідною та достатньою для виконання співвідношення (4).

Бачимо, що похибка при знаходженні прогнозного значення як середнього арифметичного m перших членів послідовності поліноміальних прогнозів визначається співвідношенням (6). Очевидно, нас цікавитиме знаходження такої кількості членів послідовності поліноміальних прогнозів, при якій відповідна похибка є мінімальною.

Тоді, можемо запропонувати наступний алгоритм екстраполяції, який включає визначення оптимального значення кількості елементів ряду поліноміальних прогнозів, в основі якого можемо покласти ідею мінімізації похибки екстраполяції:

1. Вибираємо послідовно значення $m = 1, 2, \dots, n$.
2. Знаходимо значення

$$f_{n+1}^i = \sum_{k=1}^i (-1)^{k-1} C_i^k f_{n-k+1}, \quad i = \overline{1, m}.$$

3. Знаходимо прогнозне значення $\hat{f}_{n+1}^m$ в точці $n+1$ за формулою: $\hat{f}_{n+1}^m = \frac{1}{m} \sum_{i=1}^m f_{n+1}^i$.
4. Знаходимо прогнозне значення $\hat{f}_{n+2}^m$ в точці $n+2$ за формулою: $\hat{f}_{n+2}^m = \sum_{k=2}^m (-1)^{k-1} C_m^k f_{n-k+2} + m \hat{f}_{n+1}^m$.
5. Знаходимо відхилення $\Delta_m = |\hat{f}_{n+2}^m - \hat{f}_{n+1}^m|$.
6. Знаходимо оптимальне значення m^* з умови мінімізації відхилень: $m^* = \operatorname{argmin}_{m \in \{1, 2, \dots, n\}} \Delta_m$.

IV. ЧИСЕЛЬНІ РЕЗУЛЬТАТИ

Розглянемо курс біткоїна в доларах США за період з 9 вересня 2025 р. по 4 жовтня 2025 р.: 111718; 111504; 114409; 115269; 115966; 115717; 115537; 115383; 116474; 117656; 117005; 115543; 115706; 114688; 112387; 111792; 109653; 109648; 109587; 111908; 114443; 114193; 118786; 122226; 122324.

Відповідний графік зображений на Рис.1

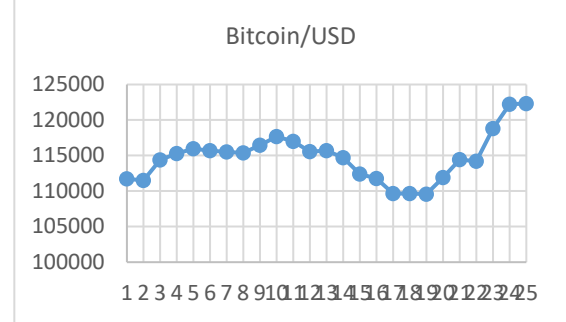


Рис.1. Графік курсу біткоїна

Будемо послідовно виконувати кроки алгоритму. В таблиці 1 наведено числові результати – значення послідовності поліноміальних прогнозів, обрахунок прогнозного значення як середнього арифметичного 6 елементів послідовності, побудову прогнозного значення в точці $n+2$. При цьому використали многочлени до 18 степеня включно.

ТАБЛИЦЯ 1. ВИЗНАЧЕННЯ ПРОГНОЗНИХ ЗНАЧЕНЬ

$\hat{f}_{n+1}$	$\hat{f}_{n+2}$	ступінь	$ \hat{f}_{n+2} - \hat{f}_{n+1} $
122422	124564	2	2142
119080	125586	3	6506
116891	129950	4	13059
120698	136503	5	15805
138129	139249	6	1120
179811	124564	7	-55247
257202	68197	8	-189005
377985	-65561	9	-443546
536834	-320102	10	-856936
695684	-733492	11	-1429176
740709	-1305732	12	-2046441
390045	-1922997	13	-2313042
-996202	-2189598	14	-1193396
-4784337	-1069952	15	3714385

-13728892	3837829	16	17566721
-33057292	17690165	17	50747457
-72059265	50870901	18	1,23E+08

Відоме значення курсу біткоїна 5 жовтня 2025 р., яке рівне 123007 USD. Прогнозне значення 123444 знайдено як середнє арифметичне 5 перших значень послідовності поліноміальних прогнозів. При цьому величина $\Delta_5=1120$.

Відповідні чисельні результати для іншої кількості точок (Таблиці 2-5) наведені нижче. У верхньому лівому куточку кожної таблиці наведено прогнозне значення як відповідне середнє арифметичне.

ТАБЛИЦЯ II.

119464,33			
$\hat{f}_{n+1}$	$\hat{f}_{n+2}$	ступінь	$ \hat{f}_{n+2}-\hat{f}_{n+1} $
122422	116604,7	2	-5817,33
119080	113647	3	-5433
116891	114031,3	4	-2859,67

$\Delta_3=2859$

ТАБЛИЦЯ III.

119772,75			
$\hat{f}_{n+1}$	$\hat{f}_{n+2}$	ступінь	$ \hat{f}_{n+2}-\hat{f}_{n+1} $
122422	117221,5	2	-5200,5
119080	114572,3	3	-4507,75
116891	115265	4	-1626
120698	118146,8	5	-2551,25

$\Delta_4=2551$

ТАБЛИЦЯ IV.

132838,5			
$\hat{f}_{n+1}$	$\hat{f}_{n+2}$	ступінь	$ \hat{f}_{n+2}-\hat{f}_{n+1} $
122422	143353	2	20931
119080	153769,5	3	34689,5
116891	167528	4	50637
120698	183475,5	5	62777,5
138129	195616	6	57487
179811	190325,5	7	10514,5

$\Delta_6=10514$

ТАБЛИЦЯ V.

150604,71			
$\hat{f}_{n+1}$	$\hat{f}_{n+2}$	ступінь	$ \hat{f}_{n+2}-\hat{f}_{n+1} $
122422	178885,4	2	56463,43
119080	207068,1	3	87988,14
116891	238592,9	4	121701,9
120698	272306,6	5	151608,6
138129	302213,3	6	164084,3
179811	314689	7	134878
257202	285482,7	8	28280,71

$\Delta_7=28280$,

Бачимо, що мінімальне значення відхилення, яке визначене у алгоритмі, досягається саме при $m=5$.

V. ВИСНОВКИ

Таким чином, в роботі отримано низку важливих результатів, які становлять теоретичну основу методу прогнозування, який пропонується. Фактично отримано необхідну та достатню умову для представлення прогнозного значення як середнього арифметичного підпослідовності поліноміальних прогнозів. Відповідна умова покладена в основі алгоритму екстраполяції, який містить підпроцедуру для знаходження кількості елементів послідовності поліноміальних прогнозів, для яких необхідно знаходити середнє арифметичне та будувати остаточний прогноз.

Чисельні результати, які наведено у роботі, підтверджують ефективність алгоритму екстраполяції. Зауважимо, що в прикладі використано дані фактично випадкового процесу. Однак, метод дав досить пристойний результат, що підтверджує можливість широкого його застосування.

- [1] Y. Turbal, A. Bomba, M. Turbal, B. Turbal, D. Smirnov «Forecasting Algorithm Based on Intellectual Analysis of Polynomial Extrapolation», Lecture Notes on Data Engineering and Communications Technologies, 2025, 244, pp. 98–113.
- [2] A.S. Kostinsky «On the principles of a spline extrapolation concerning geophysical data. » Reports of the National Academy of Sciences of Ukraine, v.2, 111-117., 2014.
- [3] Z. Fu, Y. Zhan, R.-J. Yang, Z. Xi, L. Shi «A Bayesian Inference based Model Interpolation and Extrapolation.» SAE Int. J. Mater. Manf. 5, 2, 2012,357-364.–350.
- [4] J. I. Monroe, H. W. Hatch, N. A. Mahynski, M. S. Shell, V. K. Shen, «Extrapolation and interpolation strategies for efficiently estimating structural observables as a function of temperature and density. » J. Chem. Phys. 153,2020, 144101.
- [5] N. P. Bakas «Numerical Solution for the Extrapolation of Analytic functions. » Problem of Analytic Functions/Research. 2019.
- [6] S. Makridakis, N. Bakas «Forecasting and uncertainty: a survey.» Risk and Decision Analysis, 6(1), 2016, 37–64.
- [7] N. Demiris, D. Lunn, L.D. Sharples «Survival extrapolation using the poly-Weibull model. » Stat Methods Med Res, 24(2), 2015, 287–301.
- [8] L-Y. Wang, W-C. Lee « One-step extrapolation of the prediction performance of a gene signature derived from a small study.» BMJ Open, 5:e007170, 2014 .
- [9] Y. Turbal, G. Shlikhta, M. Turbal, B.Turbal «The polynomial forecast improvement based on the algorithm of optimal polynomial degree selection», Eastern European Journal of Enterprise Technologies, 2023, 5(4(125)), pp. 34–42
- [10] Y. Turbal, M. Turbal, A. Bomba, A.A.H. Driwi, N. Kunanets «Modification of the “Pyramidal” algorithm of the small time series forecasting. » Ce/ur Workshop Proceedings, 2021, 2853, pp. 223–232.
- [11] Y. Turbal, M. Turbal, A.A. Driwi, S. Al Shukri «On the equivalence of the forecast value construction in the “pyramidal” extrapolation method and cubic forecast. » MCIT , 2020, 67–70.
- [12] R. J. Hyndman, A. V. Kostenko, «Minimum sample size requirements for seasonal forecasting model. » FORESIGHT, 2007, Issue 6.
- [13] S. R. Santanu «Numerical analysis whith algorithms and programming » , CRC Press, 2016, 675 p.