

ДОСЛІДЖЕННЯ ВПЛИВУ НАЯВНОСТІ МЕЖ ПРИ СИМУЛЯЦІЇ ФРАКТАЛЬНОГО БРОУНІВСЬКОГО РУХУ НА ТОЧНІСТЬ ОЦІНКИ ПОКАЗНИКА ХЕРСТА

Лавриненко С. Р., Лавриненко Р. М.

Науковий керівник – к.т.н., доцент Рябова Н. В.

Харківський національний університет радіоелектроніки, каф. ШІ
м. Харків, Україна

e-mail: sofiia.lavrynenko@nure.ua, roman.lavrynenko.cpe@nure.ua

This work is devoted to the influence of boundaries on the estimation of the Hurst exponent during Fractional Brownian Motion simulations in the second Anomalous Diffusion (AnDi) challenge. Ten key methods for estimating the Hurst exponent were applied to experiments with varying series lengths and diffusion coefficients for persistent time series with a Hurst exponent above 0.75. The findings reveal that all examined methods, except for the periodogram method, are susceptible to boundary effects even at short series lengths. The periodogram method proved to be robust against boundary effects when working with persistent time series across all tested series lengths and diffusion coefficient values.

В інтелектуальному аналізі даних одним із важливих напрямів є аналіз часових рядів різної природи та дослідження впливу різних факторів на показник Херста, як важливої характеристики часових рядів. Друге змагання з Аномальної дифузії (AnDi) ставить нові виклики, один з яких оцінка показника аномальності для групи частинок, що може бути виконана також методами машинного навчання. В змаганні моделюються різні аспекти аномальної дифузії, яка спостерігається в живих клітинах. Це може включати дослідження руху молекул, білкових комплексів, везикул та органел у клітинах, їх взаємодію та поведінку відповідно до різних моделей дифузії. Очікується, що однакові молекули демонструють подібні характеристики руху (дифузії). Відповідно спостереження за групою частинок дозволяє більш точно визначити характеристики їх руху. За модель аномальної дифузії прийнято Фрактальний броунівський рух.

Показник аномальності прямо пов'язаний з показником Херста: показник аномальності дорівнює подвоєному показнику Херста, і відповідно може приймати значення від 0 до 1,99. Відповідно оцінка показника аномальності в змаганні зводиться до оцінки показника Херста часових рядів за законами Фрактального броунівського руху.

Організатори змагання AnDi вважають, що траєкторії реєструються пристроєм з полем зору 128×128 пікселів, при цьому розмір пікселя становить 100 нм. Траєкторії досить короткі, від десятку до двохсот кадрів. Щоб уникнути ефектів меж, вони зазвичай симулюють дифузію в коробці розміром в 1.5 рази більшим за розмір поля зору (Рисунок 1).

Аналізуючи симуляцію фрактального броунівського руху з такими параметрами ми встановили, що ефект меж все ж таки може значно впливати на експеримент. За великих значень коефіцієнту дифузії D (визначеного як пропорційного до дисперсії зміщень уздовж одного виміру на найкоротшому часовому інтервалі) та показником аномальності вищим за 1.7 рух є досить направленим, і з часом частинка може досягнути межі коробки. Метою цієї роботи є визначення параметрів руху (довжина траєкторії та коефіцієнт дифузії), для яких наявність межі буде впливати на точність оцінки показника аномальності, а також пошук методів оцінки показника аномальності, на які найменше впливає ефект меж.

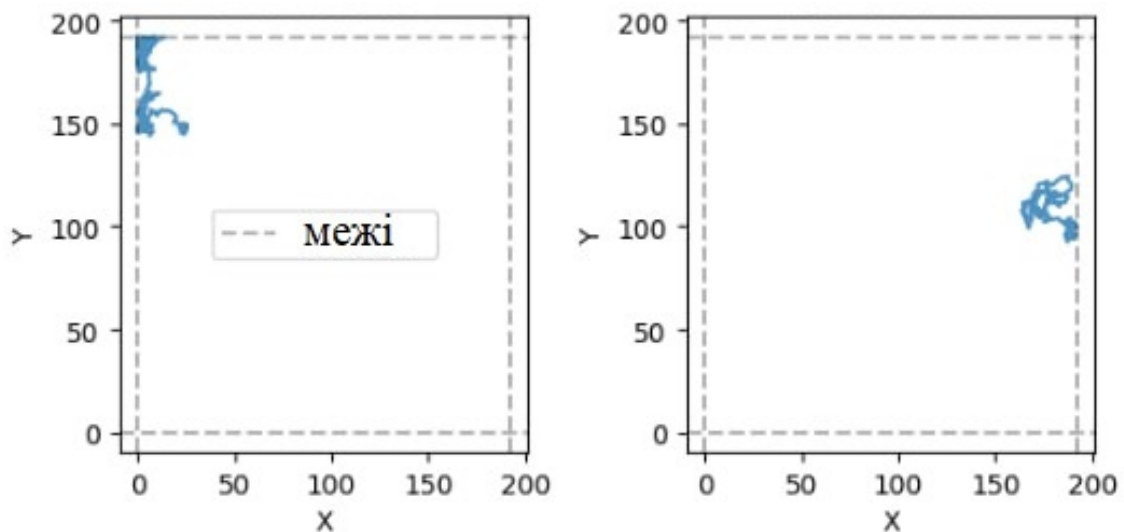


Рисунок 1 – Приклад траєкторій руху частинок біля меж коробки

Дослідження методів оцінки показника Херста було проведене зокрема в [2]. В 2023 році в [3] автори представили огляд та порівняння методів оцінки показника Херста саме для фрактального броунівського руху, зробивши публічними програмний код методів, хід та результати експериментів.

Для досліджень за допомогою інструментів для генерації траєкторій ми створили набори даних з запропонованими організаторами змагання AnDi [1] параметрами (з наявністю межі навколо коробки 192 x 192 пікселя), та такого ж розміру набір даних без меж для порівняння результату і виявлення ефекту межі.

Далі всі траєкторії перетворили в прирости (тобто отримали фрактальний гаусівський шум) і прирости нормували (поділили на середньоквадратичне відхилення приростів).

Дослідили дев'ять методів із запропонованих в [3], а також метод Віттла. Для кожної групи з 200 часових рядів отримано середнє значення показника Херста. Результатом кожного методу прийнято корінь

сереньоквадратичної похибки (RMSE) від дійсного значення показника Херста.

Отримана залежність точності оцінки показника Херста від коефіцієнта дифузії наведена в таблиці 1.

Таблиця 1 – Точність методів для довжини ряду в 25 кроків

Метод	GHE	HM	DFA	Whittle	TTA	PM	AWC	VVL	LW	LSV
Межі відсутні	0.130	0.126	0.030	0.040	0.049	0.027	0.028	0.016	0.025	0.022
Межі, $D = 0.1$	0.136	0.131	0.023	0.051	0.073	0.028	0.029	0.066	0.022	0.044
Межі, $D = 1$	0.148	0.144	0.068	0.101	0.129	0.036	0.134	0.171	0.069	0.090
Межі, $D = 3$	0.166	0.162	0.114	0.146	0.181	0.034	0.225	0.249	0.118	0.138
Межі, $D = 5$	0.180	0.178	0.139	0.171	0.209	0.033	0.271	0.288	0.146	0.169

За умов запропонованих в Другому змаганні з Аномальної дифузії (AnDi) ефект меж проявляється навіть за умов, запропонованих організаторами змагання при довжині траєкторії 25. Всі досліджені методи окрім методу періодограм суттєво знижують точність від ефекту межі.

Зауважимо, що така висока точність методу періодограм працює тільки для оцінки показника Херста групи часових рядів з однаковим показником Херста, де за результат береться середнє значення. Для оцінки показника Херста одного короткого ряду метод періодограм значно поступається точністю іншим методам.

Отримані результати показують, що в Другому змаганні з Аномальної дифузії (AnDi) треба уважно ставитись до генерації навчальних даних. В разі наявності процесів з показником Херста вищим за 0.75, що відповідає показнику аномальності вище 1.5, ефективність певних методів на отриманих траєкторіях може бути заниженою по зрівнянню з їх дісною ефективністю в реальних умовах, де межі для руху частинок відсутні.

Список використаних джерел:

1. Muñoz G., Bachimanchi H., Pineda J., Midtvedt B., Lewenstein M., Metzler R., Manzo C. Quantitative evaluation of methods to analyze motion changes in single-particle experiments. arXiv preprint arXiv:2311.18100, 2023.
2. Kirichenko L., Radivilova T., Bulakh V. Generalized approach to hurst exponent estimating by time series. Informatics Control Measurement in Economy and Environment Protection. 2018. Vol. 8, no. 1. P. 28–31. URL: <https://doi.org/10.5604/01.3001.0010.8639> (date of access: 18.02.2024).
3. Zhang H. Y., Feng Z. Q., Feng S. Y., Zhou Y. A Survey of Methods for Estimating Hurst Exponent of Time Sequence. arXiv preprint arXiv:2310.19051, 2023.
4. John Geweke and Susan Porter-Hudak. The estimation and application of long memory time series models. Journal of time series analysis, 4(4):221–238, 1983.