

Катерина КЕКАЛО, студентка

ORCID: 0009-0002-0677-3955

e-mail: erozkova777@gmail.com

Київський національний університет імені Тараса Шевченка, Київ, Україна

ЙМОВІРНІСНІ МОДЕЛІ ТА ОПТИМІЗАЦІЯ СТРАТЕГІЙ У СПОРТИВНИХ СТАВКАХ

Анотація. У статті досліджено математичне підґрунтя функціонування ринку спортивних ставок, який в умовах інтенсивної медійної промоції помилково інтерпретується масовою свідомістю як сфера випадкового успіху. Акцентовано увагу на невідповідності між маркетинговим образом «легкого прибутку» та об'єктивною математичною реальністю, що ґрунтується на детермінованих алгоритмах забезпечення дохідності букмекерських організацій.

Метою статті є аналіз ключових математичних інструментів, що пояснюють природу спортивних ставок: маржі букмекера, очікуваного значення (EV), дисперсії, критерію Келлі та розподілу Пуассона. Показано, як маржа забезпечує стабільний прибуток компанії, EV дозволяє визначати «цінність» ставки, дисперсія описує рівень ризику й невизначеності, а критерій Келлі дає оптимальну формулу розміру ставки. Використання розподілу Пуассона розглянуто як приклад статистичного прогнозування результатів у футболі.

Результати дослідження доводять, що математичний підхід може зробити ставку раціональнішою, однак не усуває системної переваги букмекера. Перспективи подальших досліджень полягають у поєднанні класичних ймовірнісних моделей з сучасними методами машинного навчання, а також у міждисциплінарному аналізі впливу реклами та психологічних чинників на поведінку гравців.

Ключові слова: очікуване значення; дисперсія; критерій Келлі; розподіл Пуассона; спортивні ставки; маржа букмекера.

1. Вступ

У сучасному світі спортивні ставки стали важливою складовою спортивної індустрії, що поєднує економіку, статистику та психологію. Їх популярність зростає, зокрема в Україні, де букмекерські сервіси активно рекламуються, формуючи у гравців ілюзію легких грошей та відсутності ризику. Однак за системою ставок стоять не випадковість і удача, а математично обґрунтовані моделі, які забезпечують прибуток організаторам. Актуальність дослідження обумовлена необхідністю точного оцінювання ймовірностей та коригування букмекерських коефіцієнтів для більш раціональної гри.

Фундаментальна проблема аналізу спортивних подій полягає у наявності *маржі букмекера* (overround), яка штучно занижує виплати та спотворює об'єктивну ймовірнісну картину події. Оскільки сума зворотних величин коефіцієнтів у лінії завжди перевищує одиницю, безпосереднє використання котирувань для оцінки реальних шансів є некоректним.

Незважаючи на існування класичних методів деривації «чистих» імовірностей, кожен із них має суттєві методологічні недоліки:

- *адитивні моделі та проста нормалізація* – часто ігнорують нелінійну природу розподілу маржі, що в окремих випадках призводить до математичних аномалій (отримання ймовірностей, що виходять за межі діапазону $[0, 1]$);
- *логарифмічні та ітеративні алгоритми* – попри вищу точність, вони не завжди здатні адекватно нівелювати системні викривлення, такі як ефект «фаворита-аутсайдера» (*favorite-longshot bias*), де ринок схильний недооцінювати фаворитів і переоцінювати малоімовірні результати.

Отже, існує нагальна потреба у розробці та впровадженні вдосконалених методів калібрування, які б дозволяли не лише усувати надлишкову маржу, а й враховувати психологічні та ринкові чинники формування коефіцієнтів. Це є критично важливим для побудови точних прогнозних моделей та досягнення позитивного математичного сподівання у довгостроковій перспективі.

Аналіз останніх досліджень показує, що питання коригування коефіцієнтів активно вивчається. Так, (Clarke, 2017) запропонував метод коригування, який усуває overround без негативних імовірностей і враховує упередження фаворитів та аутсайдерів. Дослідження (Dmochowski, 1977) та (Stefani, 2017) демонструють ефективність математичних моделей, зокрема методу найменших квадратів, для прогнозування результатів футбольних та баскетбольних матчів. Ці роботи підкреслюють важливість статистичного підходу у прийнятті рішень у ставках.

Об'єктом дослідження є методи коригування букмекерських коефіцієнтів для усунення overround та їх вплив на точність прогнозування спортивних подій. Метою роботи є розробка та оцінка ефективності таких методів. Для досягнення мети поставлено завдання: проаналізувати існуючі підходи та порівняти їх ефективність на реальних даних.

Методи дослідження включають статистичний аналіз на прикладі футбольних матчів, що дозволяє оцінити точність прогнозів та оптимізувати прийняття рішень у ставках.

2. Аналіз ключових математичних інструментів прогнозування результатів спортивних подій

2.1. Маржа букмекера

Одним із ключових інструментів отримання прибутку букмекерських компаній є так звана маржа. Вона відображає приховану комісію, яку букмекер закладає у коефіцієнти, та фактично гарантує його вигравш у довгостроковій перспективі незалежно від результату події.

Розглянемо умовний футбольний матч між збірними Англії та Іспанії. Якщо обидві команди мають рівні шанси на перемогу, то «чиста» математична ймовірність вигравшу

кожної з них становитиме 50%. У такому випадку коефіцієнти ставок без врахування маржі дорівнювали б:

$$k_{\text{Англія}} = k_{\text{Іспанія}} = \frac{1}{0,5} = 2. \quad (1)$$

Припустимо, один гравець ставить 1000 грн на Англію, а інший – 1000 грн на Іспанію. У такому випадку після матчу гроші просто переходять від того, хто програв, до того, хто виграв, і букмекер нічого не заробляє. Саме тому у реальності коефіцієнти навмисно занижуються, аби створити «приховану комісію». Наприклад, замість 2,00 на обидві команди, букмекер може запропонувати 1,90 проти 1,90. Тоді сума зворотних величин коефіцієнтів перевищує 1, що утворює маржу.

$$M = \left(\frac{1}{1,9} + \frac{1}{1,9} - 1 \right) \cdot 100\% \approx 5,26\%. \quad (2)$$

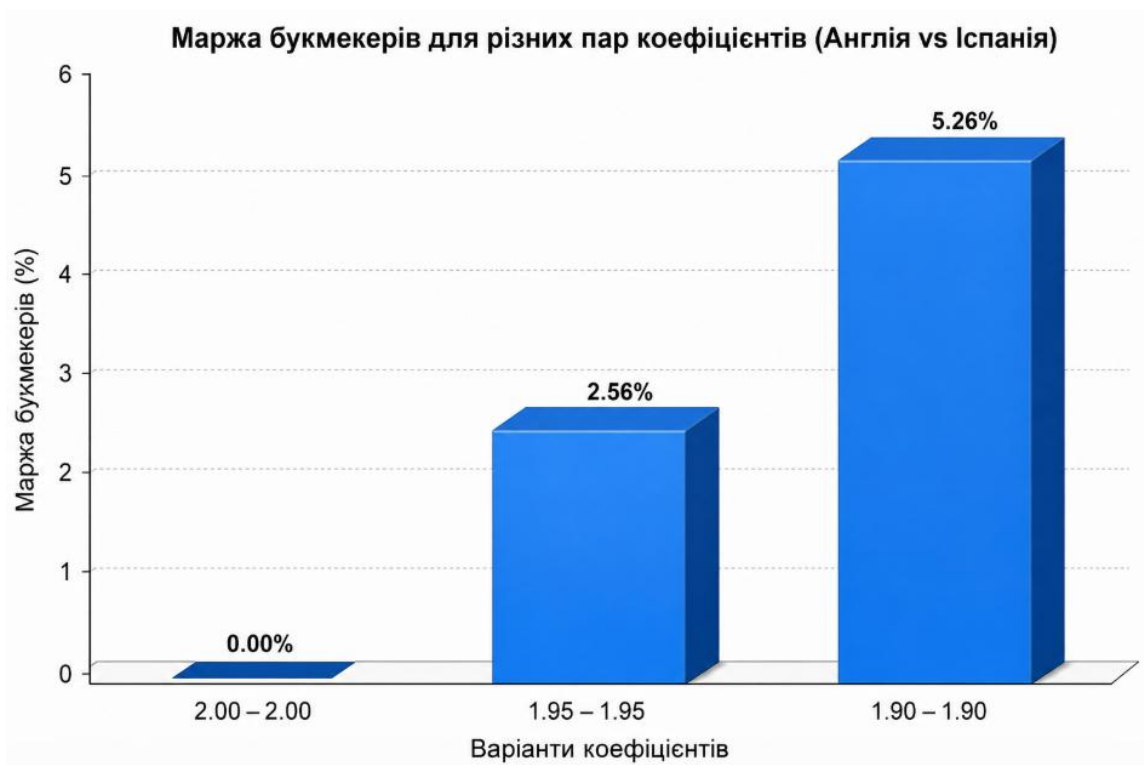


Рис 1. Залежність маржі від виставлених коефіцієнтів

У такій ситуації незалежно від результату букмекер гарантовано отримує близько 5,26% від загальної суми ставок (див. рис. 1).

2.2. Математичне сподівання

Коли гравці роблять ставку, їх цікавить не лише шанс виграшу, а й середній результат у довгостроковій перспективі. Цей показник називається *математичним*

сподіванням (англ. *expected value (EV)*) – це середня сума грошей, яку гравець може очікувати виграти або програти при багаторазовому повторенні ставки з однаковими умовами.

$$EV = s_1 \cdot p_1 - s_2 \cdot p_2, \quad (3)$$

де s_1 – виграна сума за ставку, p_1 – ймовірність виграшу, s_2 – сума програшу за ставку, p_2 – ймовірність програшу.

EV показує чи «вигідна ставка» у довгостроковій перспективі:

- якщо $EV > 0$ – у гравця є перевага виграшу;
- якщо $EV = 0$ – гра між букмекером і гравцем на рівні, ніхто не в плюсі;
- якщо $EV < 0$ – у довгостроковій перспективі гравець програє.

Приклад 1. Зробити ставку у розмірі 1000 грн на перемогу футбольного клубу «Динамо». Букмекерський коефіцієнт на перемогу «Динамо» становить 2.2. Оцінити математичне сподівання EV цієї ставки та визначити, чи є вона вигідною для гравця у довгостроковій перспективі.

Для аналізу вигідності ставки побудуємо таблицю розподілу ймовірностей прибутку та втрат, після чого розрахуємо математичне сподівання.

Таблиця 1

Коефіцієнт на виграш «Динамо» – 2,2

Прибуток/втрата з 1000 грн	Поставили на виграш «Динамо»	Ймовірність
«Динамо» (виграш)	$(2200 - 1000) = +1200$ грн	$\frac{1}{2,2} = 0,45$
«Реал» (прогрaш)	–1000 грн	$1 - 0,45 = 0,55$

$$EV = 1200 \cdot 0,45 - 1000 \cdot 0,55 = -10.$$

Отримане від’ємне значення математичного сподівання свідчить про те, що ставка на «Динамо» в довгостроковій перспективі є невигідною для гравця. Це означає, що середньостатистично з кожної ставки у 1000 грн він буде втрачати близько 10 грн. Таким чином, навіть за привабливих коефіцієнтів букмекер зберігає математичну перевагу над гравцем.

Отже, ставка на «Динамо» в довгостроковій перспективі для гравця є невигідною.

Як розрахувати математичне сподівання у ставках на спорт можна познайомитись (The Mathematical Truth About Sports Betting | Sports betting explained. (2025, September

19). YouTube), (What is Positive EV Sports Betting? (Better than Arbitrage!) (using OddsJam). (2025, September 19). YouTube).

2.3. Дисперсія

Одним із важливих показників оцінки ризику у ставках є *дисперсія*. Вона показує, наскільки сильно можливі результати можуть відхилятися від середнього значення (математичного сподівання). Якщо EV вказує на середній результат у довгостроковій перспективі, то дисперсія описує «розкид» навколо цього результату.

$$D(x) = \sum_{i=1}^n (x_i - EV)^2 \cdot p_i, \quad (4)$$

де x_i – можливі результати випадкової величини (виграш чи програш), EV – математичне сподівання, p_i – ймовірність настання відповідного результату.

Чим більша дисперсія, тим вищі коливання виграшів і програшів у короткостроковій перспективі. Це означає, що високі коефіцієнти (наприклад 5,0 чи 10,0) зазвичай пов'язані з величезною дисперсією: ставка може довго не давати виграшу, але іноді приносити великий виграш.

Приклад 2. Розглянемо ту ж ставку на матч «Динамо» – «Реал» з наступними параметрами:

- коефіцієнт на «Динамо» = 2,2;
- ймовірність перемоги «Динамо» = 0,45;
- ставка = 1000 грн.

Розрахувати дисперсію $D(x)$ та *середньоквадратичне відхилення* $\sigma = \sqrt{D(x)}$ для цієї ставки, а також інтерпретувати отримані результати з погляду ризику у короткостроковій та довгостроковій перспективах.

Можливі результати:

- якщо виграє «Динамо»: чистий прибуток $x_1 = +1200$ грн, з ймовірністю $p_1 = 0,45$;
- якщо виграє «Реал» (або нічия): втрата $x_2 = -1000$ грн, з ймовірністю $p_2 = 0,55$.

Математичне сподівання відповідно до розрахунків прикладу 1, $EV = -10$ грн.

Тепер рахуємо дисперсію:

$$D(x) = (x_1 - EV)^2 \cdot p_1 + (x_2 - EV)^2 \cdot p_2.$$

$$D(x) = (1200 - (-10))^2 \cdot 0,45 + (-1000 - (-10))^2 \cdot 0,55 = 1\,197\,900.$$

Отже, дисперсія становить майже 1,2 млн грн².

Оскільки дисперсія вимірюється у квадратних одиницях (грн²), для зручності аналізу використовують середньоквадратичне відхилення, яке повертає показник до початкових грошових одиниць – грн:

$$\sigma = \sqrt{D(x)} \approx \sqrt{1\,197\,900} \approx 1\,094 \text{ грн.}$$

Це означає, що у короткостроковій перспективі отримане значення $\sigma \approx 1094$ грн, тобто фактичні результати навіть однієї ставки можуть суттєво відхилятися від середнього очікуваного значення (–10 грн) – приблизно на ± 1094 грн.

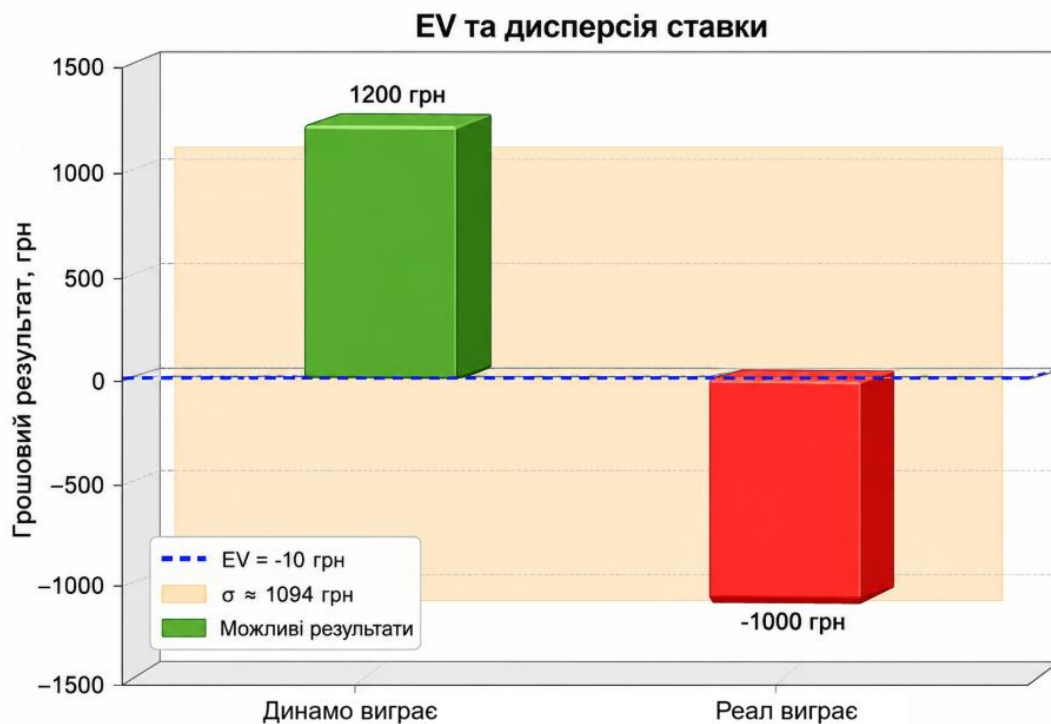


Рис 2. Середній результат і коливання ставки

Дисперсія є ключовим показником *ризик*у. Навіть за умови негативного EV висока дисперсія здатна створювати ілюзію «везіння» або виграшної серії на коротких дистанціях.

Попри будь-які випадкові коливання у короткостроковому періоді на великій кількості ставок (у довгостроковій перспективі) гравець усе одно неминуче залишиться в мінусі через системну перевагу букмекера (див. рис. 2) (How to Use Expert Predictions to Your Advantage. (2020, April 23). Betting.team).

2.4. Критерій Келлі

Критерій Келлі – це математичний метод управління розміром ставки, який допомагає максимізувати зростання капіталу у довгостроковій перспективі, враховуючи ймовірність виграшу і коефіцієнт ставки:

$$C = \frac{K \cdot V - 1}{K - 1}, \quad (5)$$

де K – коефіцієнт букмекера, V – ймовірність виграшу, C – оптимальна ставка (Carta, Conversano, 2020).

Приклад 3. Необхідно визначити оптимальну частку капіталу C для ставки на матч «Динамо» за такими даними:

- коефіцієнт букмекера на виграш «Динамо» $K = 2,2$;
- ймовірність виграшу за власною оцінкою гравця $V = 0,5$ (або 50%).

Для визначення оптимального розміру ставки скористаємося математичною моделлю критерію Келлі (5). Тоді

$$C = \frac{2,2 \cdot 0,5 - 1}{2,2 - 1} = \frac{1}{12} \approx 0,0833.$$

Отримане значення $C \approx 0,0833$ означає, що оптимальний розмір ставки становить 8,33 % від поточного загального капіталу (ігрового банку) гравця.

Отже, оптимальна частка капіталу становить 8,33 %.

Приклад 3 ілюструє, що якщо гравець планує поставити 1000 грн, то за критерієм Келлі варто поставити лише 83 грн. Таким чином гравець максимізує очікуваний приріст своїх грошей для ставок у довгостроковій перспективі, одночасно зменшуючи ризик значних втрат.

Примітка. Критерій Келлі пропонує виділяти частку банку лише тоді, коли чисельник $K \cdot V - 1 > 0$ (тобто коли ставка має додатне математичне сподівання з погляду гравця). Оскільки оцінка ймовірності виграшу гравця $V = 0,5$ вища за оцінку букмекера $\frac{1}{2,2} \approx 0,45$, формула вказує на наявність переваги і розраховує безпечний відсоток для максимізації капіталу.

2.5. Емпіричний аналіз

Для перевірки теоретичних положень щодо застосування ймовірнісних моделей у спортивних ставках було проведено емпіричне дослідження результативності футбольного клубу «Динамо» (Київ) на основі історичних даних матчів. У роботі використано інформацію про кількість голів, забитих командою в офіційних матчах за період з 2018 по 2025 роки.

Загальна вибірка склала 406 матчів, що дозволяє розглядати отримані результати як статистично репрезентативні. Основною досліджуваною випадковою величиною є кількість голів, забитих ФК «Динамо» в одному матчі, яка є дискретною та набуває невід’ємних цілих значень.

З метою аналізу можливих структурних змін у результативності команди всі матчі було поділено на дві підвибірки: мирний період (2018 - 2021 роки) – 206 матчів та період після 2022 року (2022 - 2025 роки) – 200 матчів.

Такий поділ дозволяє оцінити вплив зовнішніх соціально-історичних чинників на статистичні характеристики результативності команди.

На першому етапі дослідження було побудовано емпіричні розподіли кількості забитих голів для кожного з розглянутих періодів. Для цього підраховано кількість матчів, у яких «Динамо» забивало 0, 1, 2, ..., 6 голів.

Таблиця 2

Емпіричний розподіл кількості голів ФК «Динамо» у матчах за періоди 2018 - 2021 та 2022 - 2025 років. (<https://fcdynamo.com/calendar/dynamo-k>)

Голи	2018 - 2021	2022 - 2025
0	37	41
1	62	53
2	50	49
3	28	32
4	16	15
5	11	9
6	2	1
Всього матчів	206	200
Вибіркове середнє	1,830097087	1,79

Для загальної характеристики спортивних результатів ФК «Динамо» проаналізовано співвідношення вигравів, нічиїх та програшів у кожному з досліджуваних періодів. На рис. 3 наведено структуру результатів матчів у період 2018–2021 років та у період 2022–2025 років відповідно.

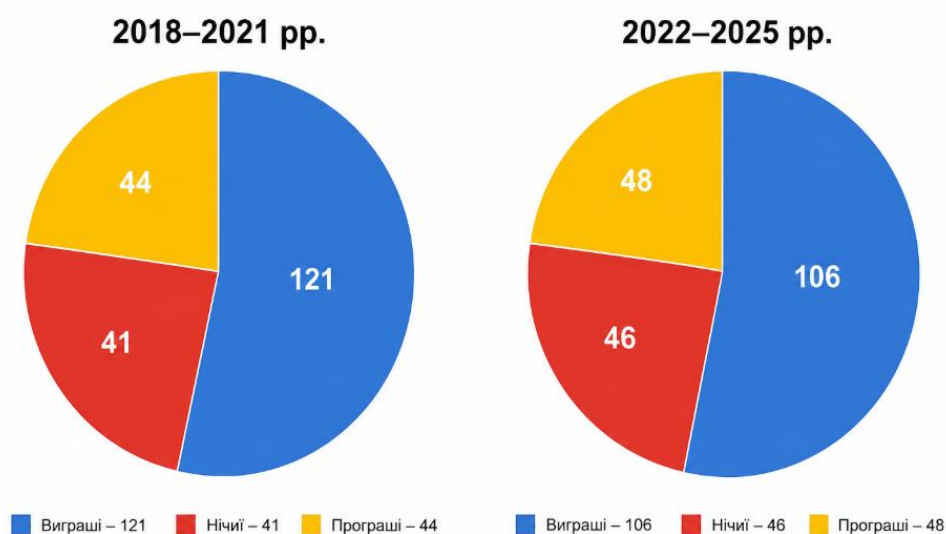


Рис 3. Структура результатів матчів ФК «Динамо» у періоди 2018 - 2021 та 2022 - 2025 років

Аналіз свідчить, що в обох періодах переважають виграші, однак після 2022 року спостерігається відносне зменшення їх частки та одночасне збільшення частки нічиїх і програшів. Це може свідчити про зростання конкурентності матчів та ускладнення умов виступів команди внаслідок воєнного стану в Україні, що є важливим моментом для подальшого аналізу результативності за кількістю забитих голів.

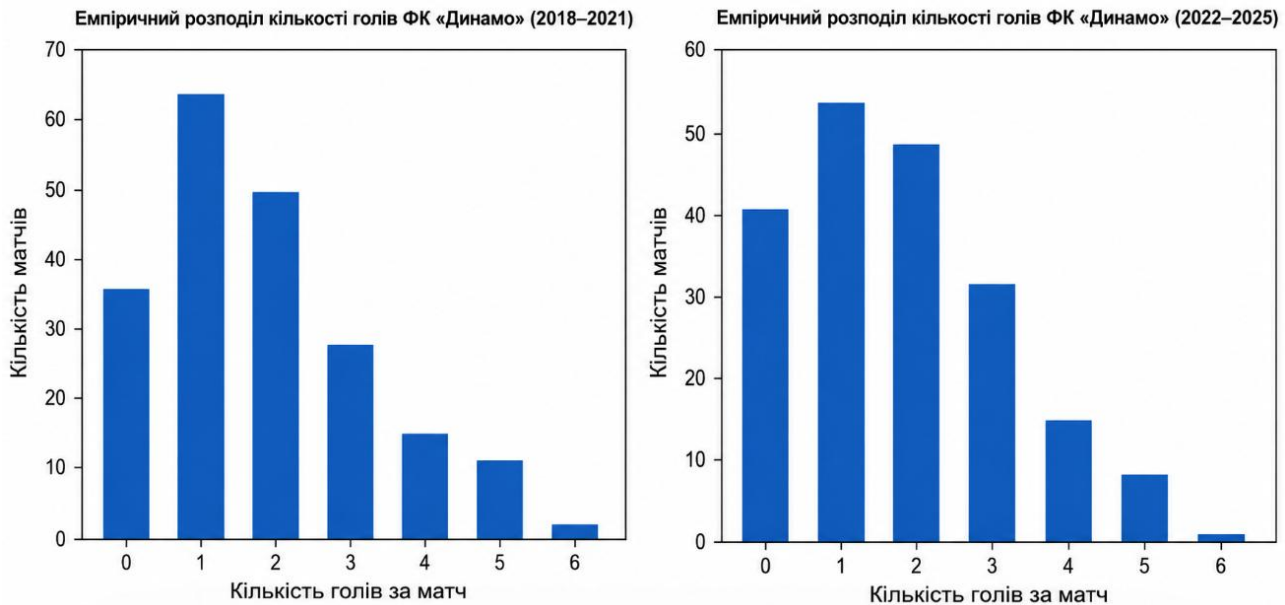


Рис 4. Розподіли кількості голів ФК «Динамо» у періоди 2018 - 2021 та 2022 - 2025 років

Для наочності емпіричні розподіли кількості голів було подано у вигляді гістограм для кожного з періодів окремо (рис. 4). Візуальний аналіз показує, що обидва розподіли є асиметричними з максимумом у зоні малих значень та поступовим спадом для більших значень кількості голів. Така форма характерна для дискретних розподілів рідкісних подій і узгоджується з припущенням про пуассонівський характер досліджуваної випадкової величини.

Кількість голів у футбольному матчі може розглядатися як результат настання незалежних подій протягом обмеженого проміжку часу. За умови відносної сталості середньої інтенсивності забивання голів та незалежності подій, така випадкова величина природно моделюється пуассонівським розподілом.

3. Розподіл Пуассона

3.1. Оцінка параметра λ пуассонівського розподілу

Параметр λ пуассонівського розподілу інтерпретується як середня кількість подій за одиницю часу, у даному випадку – *середня кількість голів*, забитих командою в одному матчі. Параметр λ , в нашому випадку, оцінюватимемо за допомогою вибіркового середнього значення кількості голів.

Використання вибіркового середнього як оцінки λ є статистично обґрунтованим, оскільки для пуассонівського розподілу вибірконе середнє є оцінкою максимальної правдоподібності та є незміщеною і асимптотично ефективною оцінкою. Натомість використання оцінки оберненої величини до вибіркового середнього не має теоретичного обґрунтування в межах пуассонівської моделі та тому не розглядається.

Постає питання, чи могли б бути використані альтернативні оцінки параметра λ . Формально можна розглянути, наприклад, оцінку $\frac{1}{\bar{X}}$, яка також залежить від λ . Однак така оцінка є зміщеною. Щоб показати це, скористаємося розкладом Тейлора функції

$$f(x) = \frac{1}{x}$$

в околі точки λ :

$$\frac{1}{\bar{X}} \approx \frac{1}{\lambda} - \frac{\bar{X} - \lambda}{\lambda^2} + \frac{(\bar{X} - \lambda)^2}{\lambda^3}.$$

Беручи математичне сподівання та враховуючи, що $E[\bar{X}] \approx \lambda$ і

$$\text{Var}(\bar{X}) = \frac{\lambda}{n},$$

отримуємо наближення

$$E\left[\frac{1}{\bar{X}}\right] = \frac{1}{\lambda} + \frac{1}{\lambda^2 n}.$$

Отже, навіть для скінченних вибірок математичне сподівання оцінки $\frac{1}{\bar{X}}$ перевищує $\frac{1}{\lambda}$, що свідчить про її зміщеність. Лише при $n \rightarrow \infty$ це зміщення зникає. Крім того, величина $\frac{1}{\bar{X}}$ не має безпосередньої інтерпретації як оцінка інтенсивності потоку подій у пуассонівській моделі та не виникає з методу максимальної правдоподібності. Тому в подальшому аналізі використовується вибірконе середнє як коректна та статистично обґрунтована оцінка параметра λ .

3.2. Критерій χ^2

З метою формальної перевірки припущення про пуассонівський характер розподілу кількості голів, зробленого на основі емпіричного аналізу, у подальшому застосовується статистичний критерій χ^2 Пірсона. Використання цього тесту дозволяє кількісно оцінити ступінь відповідності емпіричного розподілу теоретичній пуассонівській моделі та підтвердити коректність її застосування для аналізу результативності ФК «Динамо».

Постановка статистичних гіпотез

Для кожного з періодів формувались однакові гіпотези:

- *нульова гіпотеза* H_0 : емпіричний розподіл кількості голів у матчах ФК «Динамо» узгоджується з пуассонівським розподілом; спостережувані відхилення мають випадковий характер;
- *альтернативна гіпотеза* H_1 : емпіричний розподіл не узгоджується з розподілом Пуассона; відхилення є статистично значущими.

Критерій узгодженості встановлено на стандартному рівні значущості $\alpha = 0,05$.

Однією з необхідних умов коректного застосування критерію χ^2 є достатньо великі очікувані частоти в кожній категорії. Згідно з правилом Кокрена, очікуване значення E_i не повинно бути меншим за 5. Під час розрахунку теоретичних частот для категорій із великою кількістю голів було встановлено, що для значення 6 голів очікувана частота є меншою за 2 в обох періодах. Використання таких малих значень могло б призвести до спотворення статистики χ^2 та некоректних висновків.

У зв'язку з цим здійснюємо агрегацію хвостових значень: категорії «5 голів» та «6 голів» об'єднано в одну групу «5 і більше голів», що забезпечує виконання умов застосовності критерію.

Ступені свободи та p -value

Під час застосування критерію згоди χ^2 необхідно визначити кількість ступенів свободи, оскільки від цього залежить критичне значення статистики та p -value. У загальному випадку кількість ступенів свободи обчислюється за формулою:

$$df = k - 1 - m, \quad (6)$$

де k – кількість категорій (груп) у емпіричному розподілі після об'єднання інтервалів, а m – кількість параметрів теоретичного розподілу, які оцінюються за вибіркою.

У даному випадку після агрегації хвостових значень використаємо шість категорій: 0, 1, 2, 3, 4, 5+. Оскільки для пуассонівського розподілу оцінюється один параметр – середнє значення λ , кількість ступенів свободи дорівнює: $df = 6 - 1 - 1 = 4$.

p -value інтерпретується як імовірність отримати значення статистики χ^2 не менше за спостережуване за умови істинності нульової гіпотези. Якщо p -value > 0.05 , нульову гіпотезу не відхиляють, що свідчить про узгодженість емпіричних даних з теоретичною моделлю.

Зокрема, перевіряючи період з 2018 року по 2021 рік вибіркове середнє становить $\lambda = 1,83$, обсяг вибірки – $N = 206$.

Отримане значення статистики $\chi^2 \approx 5,27$ (див. табл. 3), що є меншим за критичне значення $\chi^2_{0,95}(4) = 9,49$.

p -value дорівнює 0,261, що значно перевищує рівень значущості. Таким чином, нульова гіпотеза не відхиляється.

Таблиця 3

Результати перевірки гіпотези (2018 – 2021 рр.)

Голи (k)	Реально (O)	Очікувано (E)	$\frac{(O - E)^2}{E}$
0	37	33,05	0,47
1	62	60,48	0,04
2	50	55,34	0,52
3	28	33,76	0,98
4	16	15,44	0,02
5+	13	7,93	3,24
Всього	206	206	$\chi^2 \approx 5,27$

Для періоду з 2022 по 2025 роки $\lambda = 1,79$, $N = 200$ (табл. 4). Отримали: $\chi^2 = 4,07$, $p\text{-value} = 0,396$. Нульова гіпотеза також не відхиляється, що свідчить про високу узгодженість емпіричних даних з пуассонівською моделлю.

Таблиця 4

Результати перевірки гіпотези (2022 – 2025 рр.)

Голи (k)	Реально (O)	Очікувано (E)	$\frac{(O - E)^2}{E}$
0	41	33,05	1,73
1	53	59,77	0,77
2	49	53,50	0,38
3	32	31,92	0,00
4	15	14,28	0,04
5+	10	7,14	1,15
Всього	200	200	$\chi^2 \approx 4,07$

У ставках на спорт цей розподіл часто використовують для оцінки кількості голів у матчі. Формула розподілу Пуассона:

$$P(X = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}, \quad (7)$$

де параметр λ – середня кількість голів, які забиває за гру команда.

Опишемо логіку розрахунку таблиць. Для кожного періоду збираються реальні дані (стовпчик Реально (O) – *Observed*). Далі, використовуючи формулу розподілу Пуассона (7) з розрахованим середнім λ , знаходимо теоретичну ймовірність для

кожної кількості голів і множимо її на загальну кількість матчів N . Так отримуємо стовпчик Очікувано (E) – *Expected*.

В останньому стовпчику розраховуємо міру розбіжності між теорією та практикою за формулою:

$$\frac{(O - E)^2}{E}.$$

Сума значень цього стовпчика і є значенням статистики χ^2 . Що менша ця сума, то ближчими є реальні дані до теоретичної моделі.

Результати статистичної перевірки підтверджують адекватність використання пуассонівського розподілу для моделювання кількості забитих голів ФК «Динамо» у матчі. Це узгоджується як з теоретичними передумовами, так і з емпіричними спостереженнями.

По-перше, гол у футболі є дискретною подією, що відбувається відносно рідко протягом фіксованого проміжку часу, за умов великої кількості атак та малої ймовірності реалізації кожної окремої спроби. Такі умови є класичними для виникнення пуассонівського потоку подій.

По-друге, у спортивній аналітиці традиційно припускається відносна незалежність моментів забиття голів, що дозволяє застосовувати спрощену стохастичну модель без урахування складних динамічних ефектів.

По-третє, застосовані критерії згоди χ^2 для двох незалежних часових періодів показали відсутність статистично значущих відхилень від пуассонівського розподілу ($p > 0,05$). Це дозволяє вважати модель стійкою до змін зовнішніх умов і використовувати її для подальшого аналізу, зокрема для прогнозування результативності та оцінювання ймовірностей подій у спортивних ставках.

3.3. Використання розподілу Пуассона

Розподіл Пуассона – це математична модель, яка показує ймовірність того, що певна подія відбудеться певну кількість разів за фіксований проміжок часу. Цей підхід часто використовують у спорті.

Аналітичний відділ футбольного клубу (або букмекерська компанія) досліджує результативність ФК «Динамо» за два тривалі часові періоди: з 2018 по 2021 рік ($N_1 = 206$ матчів) та з 2022 по 2025 рік ($N_2 = 200$ матчів). За статистичними даними, вибіркове середнє кількості забитих голів за матч у першому періоді становило $\lambda_{a_1} = 1,83$, а в другому – $\lambda_{a_2} = 1,79$.

Для оптимізації прогнозування майбутніх результатів висунуто припущення, що кількість голів, яку команда забиває в одному матчі, є випадковою величиною, яка підпорядковується закону розподілу Пуассона.

Завдання.

1. *Перевірити статистичну гіпотезу* про відповідність емпіричних (реальних) даних кількості забитих голів теоретичному розподілу Пуассона

для обох періодів за допомогою критерію згоди Хі-квадрат (χ^2) Пірсона на рівні значущості $\alpha = 0,05$. Для цього об'єднати хвостові значення вибірки в категорію «5+» голів та визначити кількість ступенів свободи.

2. Зробити статистичний висновок на основі отриманих значень статистики χ^2 та показника p -value щодо адекватності використання цієї моделі.

Використовуючи побудовану модель, розрахувати теоретичну ймовірність того, що в наступному окремому матчі (наприклад, проти ФК «Реал») команда заб'є рівно 2 голи, якщо її поточна середня результативність λ становить 1,5 гола за гру. Важливо розуміти, що в окремому матчі не можна забити півтора гола, але як середній показник багатьох попередніх ігор це число є базою для прогнозу.

Значення параметра $k = 2$ – це цільова кількість голів у події, ймовірність якої хочемо дізнатися. Тобто шукаємо точну ймовірність того, що команда заб'є рівно 2 голи в конкретному майбутньому матчі проти «Реала». Тоді використовуючи (7), отримаємо:

$$P(X = 2) = \frac{1,5^2 \cdot e^{-1,5}}{2!} \approx 0,25102.$$

Знайдене число 0,251 переводиться у відсотки. Отримуємо приблизно 25,1% (або округлено 25%).

Що це означає для аналітика або гравця на ставках? Якщо «Динамо» зіграє велику кількість матчів із суперниками аналогічного рівня за такої ж форми, то приблизно в кожному четвертому матчі (25% випадків) команда забиватиме рівно 2 голи.

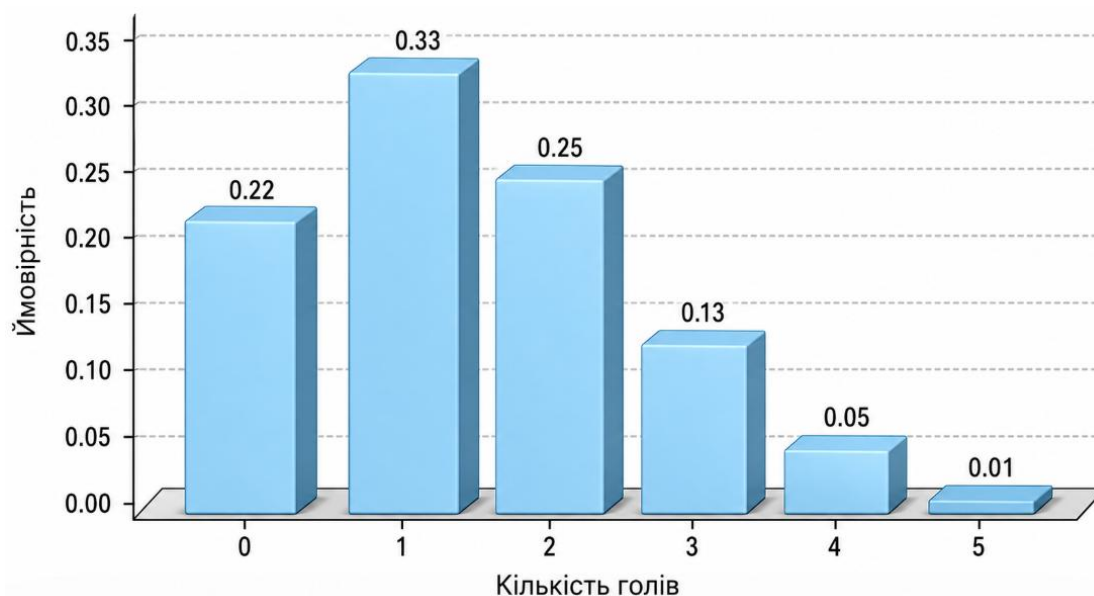


Рис 5. Ймовірність кількості голів «Динамо» за розподілом Пуассона

За допомогою цієї ж формули можна порахувати ймовірність для $k = 0$ (0 голів), $k = 1$ (1 гол), $k = 3$ тощо, щоб отримати повну картину розподілу ймовірностей для цього матчу (див. рис. 5).

У букмекерському бізнесі цей показник одразу перевели б у «чистий» коефіцієнт ($\frac{1}{0,251} \approx 3,98$). Якщо букмекер пропонує на цей результат коефіцієнт, вищий за 4,0, така ставка вважатиметься вигідною.

Використання цього підходу дозволяє кількісно оцінювати ймовірність різних результатів матчу. Можна комбінувати для обох команд, щоб передбачити точний рахунок. (How to Use Expert Predictions to Your Advantage. (2020, April 23). Betting.team).

4. Висновки

Результати дослідження підтверджують, що математичний підхід до спортивних ставок дозволяє оцінювати ризики та приймати більш обґрунтовані рішення. Зокрема, аналіз маржі букмекера свідчить, що вона забезпечує стабільний прибуток гравців незалежно від результату події, що узгоджується з висновками, який наголошує на впливі overround на реальні ймовірності подій. Розрахунки математичного сподівання (EV) і дисперсії демонструють, що навіть вигідні на перший погляд ставки можуть мати високий ризик коливань результатів, що підтверджує висновки щодо важливості врахування варіаційних характеристик у прогнозуванні спортивних подій.

Критерій Келлі показав, що оптимальний розмір ставки дозволяє максимізувати приріст капіталу у довгостроковій перспективі, одночасно зменшуючи ризик значних втрат, що узгоджується з теоретичними розрахунками. Використання розподілу Пуассона для прогнозування кількості голів у матчах підтверджує ефективність статистичних моделей у спортивних прогнозах, що відповідає рекомендаціям практиків щодо застосування ймовірнісних моделей у практиці ставок.

Отже, дослідження відповідає на поставлені питання: математичні моделі дозволяють оцінювати прибутковість ставки, рівень ризику та оптимальний розмір ставки. Практичне значення роботи полягає у тому, що представлені методи можуть бути використані як гравцями для раціональної гри, так і аналітиками для покращення прогнозів спортивних результатів. Перспективи подальших досліджень включають інтеграцію класичних моделей з методами машинного навчання та міждисциплінарний аналіз впливу психологічних та маркетингових чинників на поведінку гравців.

Список використаних джерел

- How to calculate expected value in betting. (n.d.). Smarkets Help Centre. <https://help.smarkets.com/hc/en-gb/articles/214554985-How-to-calculate-expected-value-in-betting>
- How to Use Expert Predictions to Your Advantage. (2020, April 23). Betting.team. <https://betting.team/blog/smart-betting-expertpred>
- The Mathematical Truth About Sports Betting | Sports betting explained. (2025, September 19). YouTube. <https://youtu.be/4dS77MY9wUU>
- What is Positive EV Sports Betting? (Better than Arbitrage!) (using OddsJam). (2025, September 19). YouTube. <https://youtu.be/s6R9xB1RdZU>
- Adjusting bookmaker's odds to allow for overround. (2017). Outlier.Bet. https://outlier.bet/wp-content/uploads/2023/08/2017-clarke-adjusting_bookmakers_odds.pdf
- Carta A and Conversano C (2020) Practical Implementation of the Kelly Criterion: Optimal Growth Rate, Number of Trades, and Rebalancing Frequency for Equity Portfolios. *Front. Appl. Math. Stat.* 6:577050. doi: 10.3389/fams.2020.577050

References

- How to calculate expected value in betting. (n.d.). Smarkets Help Centre. <https://help.smarkets.com/hc/en-gb/articles/214554985-How-to-calculate-expected-value-in-betting>
- How to Use Expert Predictions to Your Advantage. (2020, April 23). Betting.team. <https://betting.team/blog/smart-betting-expertpred>
- The Mathematical Truth About Sports Betting | Sports betting explained. (2025, September 19). YouTube. <https://youtu.be/4dS77MY9wUU>
- What is Positive EV Sports Betting? (Better than Arbitrage!) (using OddsJam). (2025, September 19). YouTube. <https://youtu.be/s6R9xB1RdZU>
- Adjusting bookmaker's odds to allow for overround. (2017). Outlier.Bet. https://outlier.bet/wp-content/uploads/2023/08/2017-clarke-adjusting_bookmakers_odds.pdf
- Carta A and Conversano C (2020) Practical Implementation of the Kelly Criterion: Optimal Growth Rate, Number of Trades, and Rebalancing Frequency for Equity Portfolios. *Front. Appl. Math. Stat.* 6:577050. doi: 10.3389/fams.2020.577050

Отримано редакцією журналу: 20.03.2026

Прийнято до публікації: 19.06.2026

Опубліковано: 30.07.2026

Kateryna KEKALO, Student

ORCID: 0009-0002-0677-3955

e-mail: erozkova777@gmail.com

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

PROBABILITY MODELS AND OPTIMIZATION OF STRATEGIES IN SPORTS BETTING

Abstract. *The article explores the mathematical basis of the functioning of the sports betting market, which, in the conditions of intensive media promotion, is mistakenly interpreted by the mass consciousness as a sphere of random success. The focus is on the discrepancy between the marketing image of "easy profit" and the objective mathematical reality, which is based on deterministic algorithms for ensuring the profitability of bookmaker organizations. The aim of the article is to analyze the key mathematical tools that explain the nature of sports betting: bookmaker's margin, expected value (EV), variance, Cayley criterion and Poisson distribution. It is shown how the margin ensures a stable profit for the company, EV allows you to determine the "value" of the bet, variance describes the level of risk and uncertainty, and Cayley criterion gives the optimal formula for the size of the bet. The use of the Poisson distribution is considered as an example of statistical prediction of results in football. The results of the study prove that a mathematical approach can make betting more rational, but does not eliminate the bookmaker's systemic advantage. Prospects for further research lie in combining classical probability models with modern machine learning methods, as well as in an interdisciplinary analysis of the impact of advertising and psychological factors on player behavior.*

Keywords: *expected value; variance; Cayley criterion; Poisson distribution; sports betting; bookmaker's margin.*

Received: 20.03.2026

Accepted for publication: 19.06.2026

Published: 30.07.2026