

УДК 321.01:351

Качинський Анатолій Броніславович

Добровольський Євген Леонідович

Ланде Дмитро Володимирович

ВИВЧЕННЯ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК ГРОМАДСЬКОЇ АКТИВНОСТІ НАСЕЛЕННЯ ЯК ІНДИКАТОРА ЗАГРОЗ НАЦІОНАЛЬНІЙ БЕЗПЕЦІ УКРАЇНИ

Постановка проблеми. Громадська активність – діяльність соціальних суб'єктів, яка має на меті впливати на прийняття політичних рішень, реалізацію інтересів суспільства, буває усвідомленою і стихійною [1]. Відповідно, кількісна оцінка громадської активності може бути використана у процесі моніторингу інформаційного простору з метою виявлення загроз національній безпеці України у внутрішньополітичній сфері.

Ця стаття є другою в серії робіт авторів із вивчення активності громадян. Наразі пропонується розглянути динаміку її розвитку за три роки в Україні, вивчивши статистичні закономірності розподілу кількісних показників громадської активності. На думку авторів, найбільш повна класифікація проявів активності громадян держави викладена у праці Томаша Жиро [2], детальний опис її складових наведено у попередній роботі авторів.

Головна **мета** – розкрити особливості статистичних характеристик громадської активності населення України за 2008–2010 роки. У ході цього дослідження вирішуються **завдання** вимірювання випадкових величин (в.в.), що відповідають динаміці змін інтенсивності визначених видів громадської активності, виконання статистичних процедур встановлення та вивчення параметрів розподілу в.в., а також перевірки пуасоново-сті виходів випадкових величин за граничні значення.

На сьогодні вивчення активності громадян проводиться політологами “в ручному режимі”. Тобто лише експерти, аналізуючи

останні новини політичного та суспільного життя, можуть доходити висновку про зміни політичного настрою суспільства, визначати взаємозв'язки різноманітних подій, прогнозувати їх розвиток. Недоліком зазначеного методу є те, що на отримані результати дослідження накладає свій відбиток суб'єктивність думки дослідника: одні події отримують більш позитивний характер, інші – навпаки негативний. Дослідник не завжди має змогу охопити всі останні новини, а виокремлює лише ті, що, на його думку, є найбільш інформативними, цікавими та правдоподібними. Автори статті пропонують концептуально новий підхід до оцінювання досліджуваних явищ – повністю автоматизовану процедуру аналізу останніх новин суспільного життя, розраховуючи параметри математичної моделі громадської активності населення.

Проведене дослідження громадської активності населення України за 2008–2010 роки можна розділити на два етапи: 1) створення математичної моделі даних досліджуваних явищ та отримання результатів вимірювань якісних показників громадської активності; 2) проведення статистичних розрахунків, дослідження розподілів в.в., виявлення закономірностей та фактів.

На першому етапі виконано вимірювання випадкових величин, що відповідають видам громадської активності. З цією метою в інформаційно-аналітичному комплексі Infostream [3] розроблено лінгвістичні фільтри (табл. 1), котрі дають змогу оцінювати кількість повідомлень із описом одного з видів

громадської активності в розрізі кожного дня. Наведені фільтри виокремлюють лише ті повідомлення, які відносяться до подій в Україні (умова “country.UA”), а також джерелом новин є українські ЗМІ (“geo.UA”). Якщо перша умова дозволяє обмежити досліджувані явища у межах однієї держави, то

друга умова – відхилити “хибні” потоки новин (наприклад, ті новини закордонних ЗМІ, що використовуються з метою зовнішнього інформаційного впливу) [4].

Таблиця 1

Запити в інформаційно-аналітичній системі InfoStream

Вид громадської активності	Запит в інформаційній системі
Демонстрація	(митинг пикет мітин пикет)&(country.UA geo.UA*)
Лобізм	(Лобі Лобби)&(country.UA geo.UA*)
Судовий процес	((Судов~процес) (Судебн~процесс))&(country.UA geo.UA*)
Страйк	(забастов страйк)&(country.UA geo.UA*)
Тиск та протест	((давлен тиск)&(протест))&(country.UA geo.UA*)
Бойкот	(мовчанка бойкот)&(country.UA geo.UA*)
Громадська непокора	((граждан~неповиновен) (громад~непокора))&(country.UA geo.UA*)
Делегітимізація	(Делегітиміз Делегитимиз)&(country.UA geo.UA*)
Політичне насилля	((полити~насил) (політич~насил))&(country.UA geo.UA*)

Статистична модель. У результаті проведених вимірювань отримано вибірку випадкових величин ξ_1, \dots, ξ_n для кожного окремого виду громадської активності, де ξ_i становить кількість повідомлень, що були

зафіксовані у ЗМІ за i -ий день дослідження із певною тематикою. Розподіл таких в.в. не відомий, ξ_i – цілі невід’ємні числа. Нашим першочерговим завданням є встановлення закону розподілу отриманої вибірки.

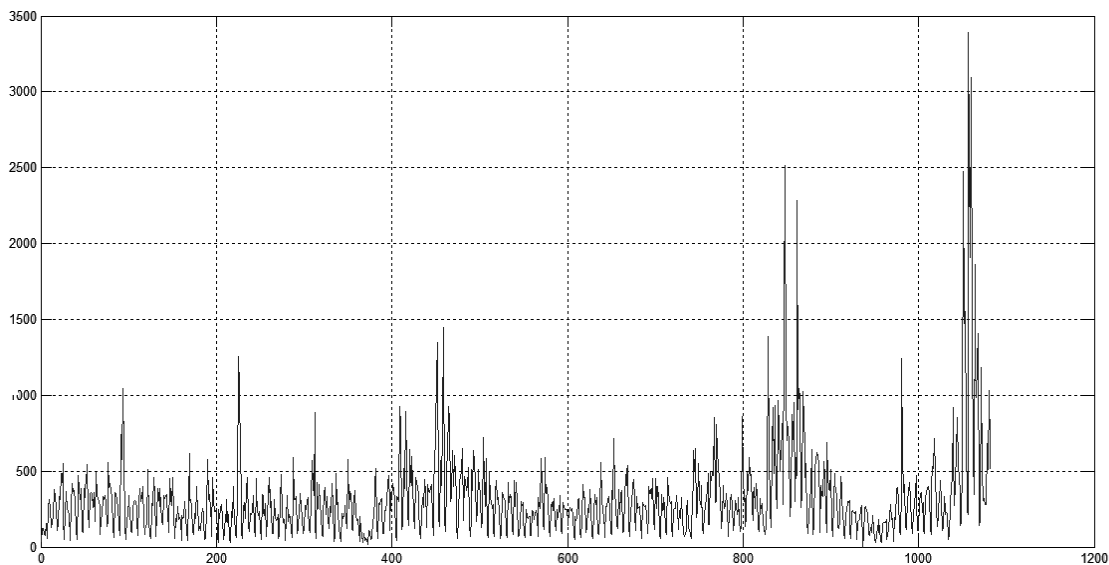


Рис. 1. Кількість повідомлень, що описують вид громадської активності “Демонстрація”, за період із 1 січня 2008 р. по 31 грудня 2010 р. (вісь ординат відображає номер дня за зазначений період)

Оскільки ξ_i – кількість подій у певний проміжок часу, то виконаємо перевірку на відповідність ξ_i закону Пуасона. Для цього скористаємося критерієм хі-квадрат. Необхідно розрахувати середнє $\bar{\xi}_i$ та розглянути розподіл Пуасона із середнім $\lambda = \bar{\xi}_i$, далі необхідно розрахувати кількість очікуваних випадків у послідовних інтервалах та власне провести процедуру перевірки підбору розподілу за критерієм хі-вадрат. Зауважимо, що будь-яке інформаційне повідомлення одночасно може відноситись відразу до кількох видів громадської активності. Проведені дослідження охоплюють період із 1 січня 2008 р. по 31 грудня 2010 р., а отже $i=1, \dots, 1096$. Як приклад, отримані результати для виду громадської активності «Демонстрації» відображено на рис. 1. (надалі буде-

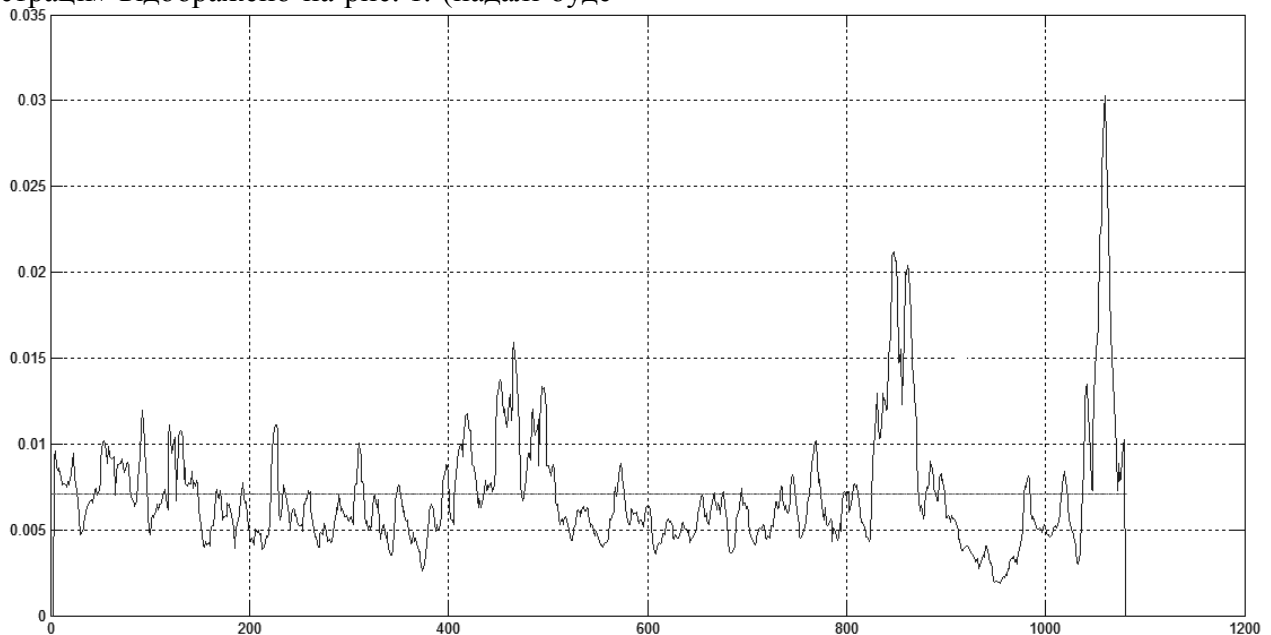


Рис. 2. Відсоток повідомлень від загальної кількості, що описують вид громадської активності «Демонстрація», за період із 1 січня 2008 р. по 31 грудня 2010 р. (вісь абсцис відображає номер дня за зазначений період)

Обробку даних легко описати за формулою:

$$\xi'_i = \frac{1}{7} \cdot \sum_{j=i-3}^{j=i+3} \frac{\xi_j}{N_j} \quad (1),$$

де ξ_j – кількість повідомлень із визначеною тематикою за j -ий день дослідження,

мо відображати результати дослідження лише для цього виду громадської активності, як для найбільш цікавого та інформативного з погляду дослідження, однак в кінці наведемо результати вивчення й інших видів громадської активності).

Обробка статистичних даних. Отримані результати, як і будь-які інші статистичні вимірювання, потребують попередньої обробки та очищення від аномальних даних. Для цього, по-перше, виконано згладжування за допомогою рухомого інтервалу довжиною 7 днів. Саме такий розмір інтервалу обрано у зв'язку з циклічністю інтенсивності потоку новин (він мінімальний у вихідні дні, а максимальний у середині тижня) [5]. По-друге виконано нормування отриманого значення на загальну кількість повідомлень, що надійшли до системи за один окремий день.

N_j – загальна кількість повідомлень, що надійшли до інформаційно-аналітичної системи за j -ий день дослідження. Зауважимо, що ξ'_i мають бути цілими невід'ємними числами, тому виконується їх масштабування таким чином, щоб всі ξ'_i знаходилися в діапазоні (0; 100) та були цілими числами. Ро-

бота ведеться саме з такими значеннями, однак у статті будемо відображати не масштабовані значення.

Отримані результати обробки вхідних даних відображено на рис. 2.

Теоретична модель (розподіл Пуасона). З теорії математичної статистики відомо, що кількість випадкових подій за одиницю часу, що описує закон Пуасона, має функцію розподілу (2). Досліджувана модель відображає кількість певних подій (демонстрацій, страйків, проявів лобізму тощо) за одиницю часу, отже теоретичною моделлю має бути:

$$F(\xi = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!} \quad (2)$$

$$\lambda > 0, k = 0, 1, 2, 3 \dots$$

Гіпотези. Висунемо гіпотезу, що досліджувані випадкові величини розподілені за законом Пуасона:

$H_0: \{\xi_i\}$ розподілені за законом Пуасона

$H_1: \{\xi_i\}$ розподілені не за законом Пуасона

Перевірка гіпотез. Закон розподілу Пуасона має лише один параметр – λ , який

визначає середню кількість подій, що настають в одиницю часу. У такому випадку, розрахована оцінка параметра λ для прояву громадської активності “Демонстрація” становить 295,63 повідомлень за день (без згладжування), або ж 0,0067% (зі згладжуванням рухомим інтервалом та нормуванням) від загальної кількості повідомлень за добу. На основі оцінки вхідного параметра отримано теоретичну функцію розподілу для досліджуваного процесу, яку і буде подано на вхід до статистики хі-квадрат. Однак повернімося на крок назад і побудуємо графік емпіричної щільності розподілу (рис. 3.): відмітимо, що отримана функція має “важкий хвіст” у своїй правій частині. Саме цей факт першим спонукає нас прийняти до сумніву запропоновану гіпотезу H_0 . На графіку можна відмітити три локальних максимуми на значеннях 0,0225, 0,03 та 0,035, тобто можна дійти висновку, що у випадку загострення громадської активності, що відповідає тематичному інформаційному потоку “Демонстрація”, відповідні відсотки повідомлень із визначеною тематикою будуть становити саме такі значення.

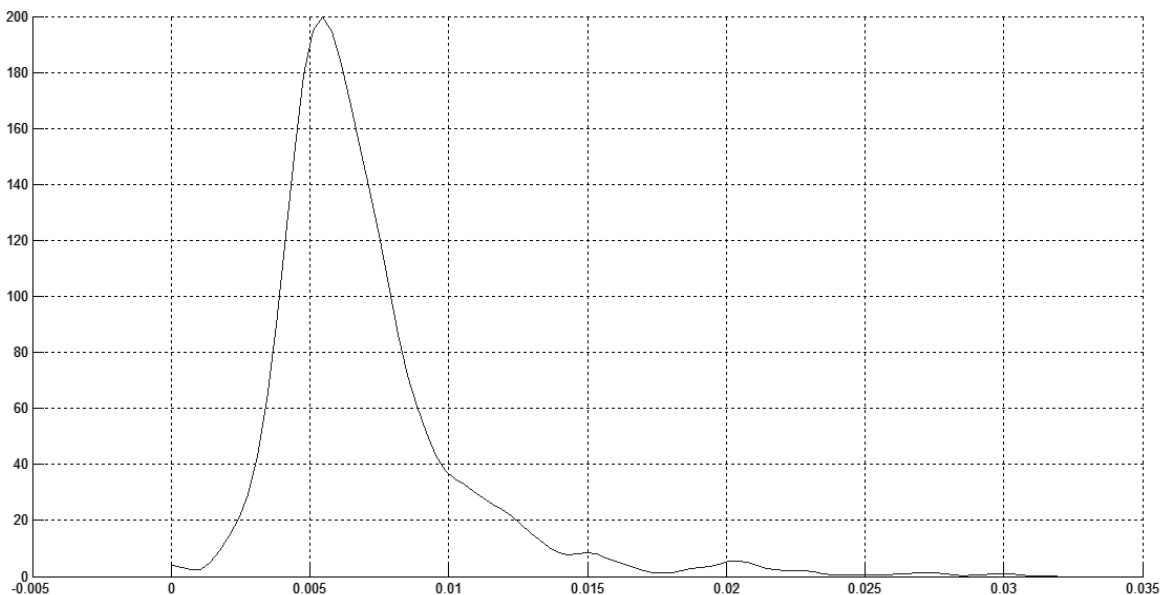


Рис. 3. Емпірична функція щільності розподілу випадкової величини, що відповідає відсотку повідомлень від загальної кількості, які описують вид громадської активності “Демонстрація”, за період із 1 січня 2008 р. по 31 грудня 2010 р. (вісь ординат відображає номер дня за зазначений період)

Наступний крок нашого дослідження – це застосування статистики хі-квадрат для встановлення кореляції емпіричної та теоретичної функції розподілу досліджуваної випадкової величини. На рис. 4. відображено обидві функції: суцільна лінія – емпірична, штрихова – теоретична.

Факт того, що емпірична функція розподілу ймовірності є більш пологою пояснюється природністю досліджуваної в.в. – тобто не існує чіткої межі між гострим проявом громадської активності та її відсутністю. Саме така різниця швидкості зростання функцій розподілу ймовірності є другим фактом, що змушує поставити під сумнів запропоновану гіпотезу.

Застосування самої статистики хі-квадрат (3) дає негативну відповідь на питання чи є запропонована гіпотеза H_0 прави-

льною, підтверджуючи наші попередні висновки.

$$\chi^2 = N \cdot \sum \frac{(P_i^{theor} - P_i^{emp})^2}{P_i^{theor}} \quad (3),$$

де P_i^{theor} и P_i^{emp} – теоретична та емпірична ймовірності потрапляння випадкової величини в i -ий інтервал.

Отже, доходимо до першого висновку – випадкова величина, що визначає кількість повідомлень за добу, котрі описують окремих вид громадської активності “Демонстрація”, не підпорядковується закону розподілу Пуасона. Аналогічні результати отримано для інших видів громадської активності населення.

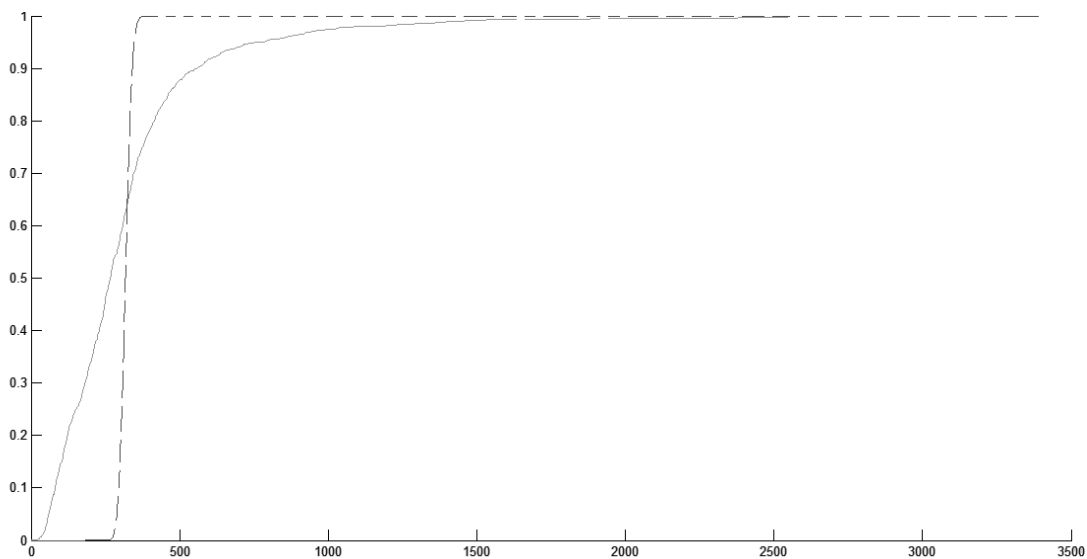


Рис. 4. Функції розподілу в.в. “Демонстрація” для перевірки гіпотези на відповідність закону Пуасона (суцільна лінія – емпірична, штрихова – теоретична)

Теоретична модель (біноміальний розподіл). Аналогічні розрахунки для встановлення закону розподілу випадкових величин ξ_i було виконано з огляду на припущення про їх біноміальний розподіл:

$$F(\xi = k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \quad (4),$$

де ξ виступає як сума n випадкових величин, що розподілені за законом Бернуллі, тобто

$$\xi = \sum_{i=1}^n y_i \quad (5),$$

де $y_i = \begin{cases} 1, & \text{з ймовірністю } p \\ 0, & \text{з ймовірністю } 1-p \end{cases} \quad i=1, \dots, n.$

Рішення про виконання перевірки саме для такого розподілу прийнято на основі вигляду графіку щільності розподілу випадкових величин ξ_i' (рис. 3).

Гіпотези. Висуємо гіпотезу, що досліджувані випадкові величини розподілені за згаданим вище законом, отже маємо дві гіпотези:

H_0 : $\{\xi_i'\}$ розподілені за біноміальним законом

H_1 : $\{\xi_i'\}$ розподілені не за біноміальним законом

Перевірка гіпотез. Оцінку параметра p для біноміального розподілу та розрахунок теоретичної та емпіричної функції розподілу виконано за допомогою стандартних функцій пакету MATLAB. У такий спосіб отримано, що $p=0,2349$, функції розподілу відображені на рис. 5.

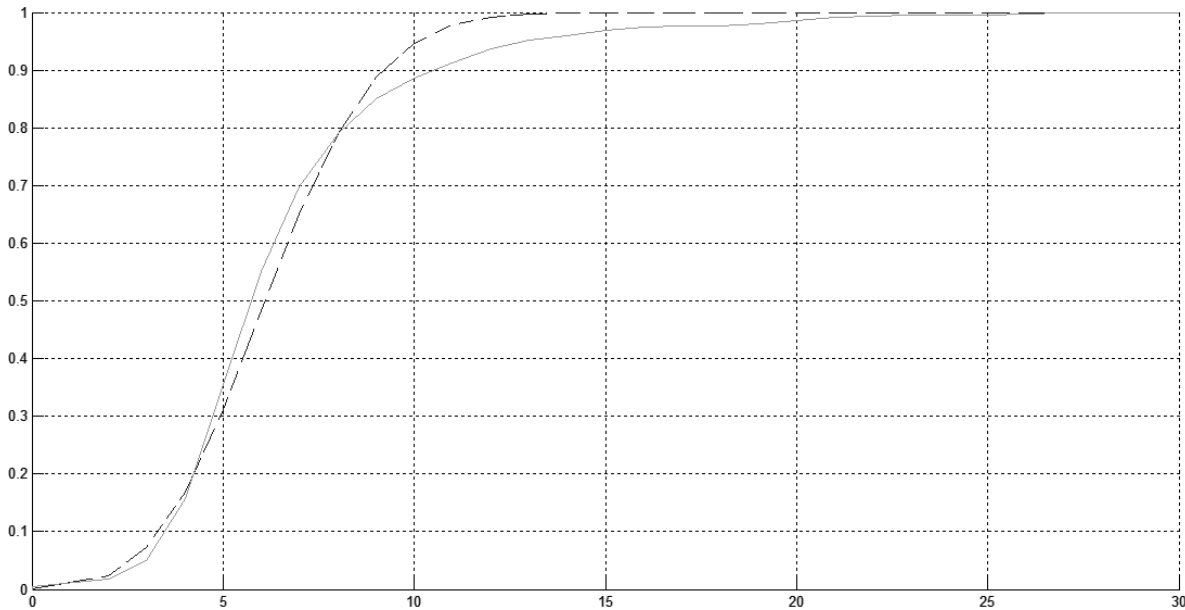


Рис. 5. Функції розподілу в.в. “Демонстрація” для перевірки гіпотези на відповідність біноміальному закону розподілу (суцільна лінія – емпірична, штрихова лінія – теоретична)

Критерій хі-квадрат дає позитивний результат щодо кореляції двох функцій розподілу. При дослідженні інших видів громадської активності встановлено анало-

гічні закономірності розподілу випадкових величин, їх статистичні характеристики занесені до табл. 2.

Таблиця 2

Статистичні характеристики видів громадської активності населення України за 2008–2010 роки

Вид громадської активності	$\bar{\xi}_i$	$\max(\bar{\xi}_i)$	$\bar{\xi}_i'$	$\max(\bar{\xi}_i')$	p
Демонстрація	296	3393	0,0067	0,0303	0,2349
Лобізм	107	380	0,0022	0,0053	0,4508
Судовий процес	45	438	0,0010	0,0050	0,2011
Страйк	77	797	0,0016	0,0067	0,2351
Делегітимізація влади	0,8	68	0,00002	0,0008	0,0065

На другому етапі досліджувалися виходи випадкової величини ξ_i' за граничне значення: відповідно до теорії екстремумів випадкових величин розподіл кількості таких виходів за одиницю часу має підпорядковуватися закону Пуасона (2), отже виконаємо перевірку цієї гіпотези. Для цього необхідно визначитися із двома параметрами: граничним значенням та періодами досліджень. Щодо першого параметра, визначимо можливе граничне значення та перевіримо його узгодженість із роботами інших науковців. Щодо другого показника – періодів досліджень, то весь інтервал із 1 січня 2008 р. по 31 грудня 2010 р. буде поділено на періоди терміном один місяць із початком кожного 1-го та 15-го числа місяця. Досліджуваний інтервал тривалістю в один місяць обраний у зв'язку з тим, що, на думку психологів, люди часто вибудовують свої плани на початку та в кінці календарного місяця, який, у свою чергу, приблизно рівний періоду обертання Місяця

довкола Землі [6]. Додаткові інтервали, що починаються кожного 15-го числа місяця, введено для врахування подій, котрі не співпали з початком календарного місяця.

Статистична модель. Таким чином отримуємо 72 досліджуваних інтервали, а отже і випадкові величини $\{v_k\}$ $k=1, \dots, 72$, що відповідають кількості виходів ξ_i' за граничне значення у відповідний період досліджень, тобто

$$v_k = \sum_{i \in \{k\text{-ий діапазон дослідження}\}} 1_{\xi_i' > \text{граничне значення}} \quad (6),$$

де 1 – функція-індикатор точки виходу ξ_i' за граничне значення.

На рис. 6 зображено залежність емпіричного значення $E[v_k]$ від граничного значення для дослідження громадської активності “Демонстрації”.

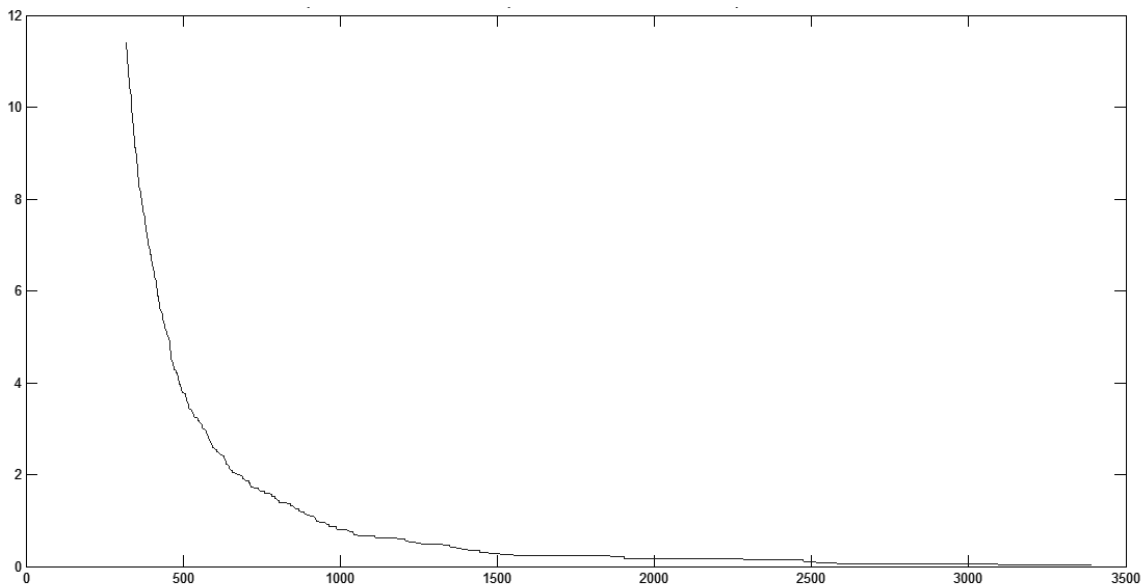


Рис. 6. Залежність емпіричного значення $E[v_k]$ від граничного значення для дослідження громадської активності “Демонстрації”

Вісь ординат – граничне значення (відлік починається зі значення середньої кількості повідомлень за добу з визначеною тематикою громадської активності), вісь абсцис – кількість виходів за граничне значення.

Гіпотези. Висунемо гіпотезу, що досліджувані випадкові величини розподілені за законом Пуасона, отже маємо дві гіпотези:
 H_0 : $\{v_k\}$ розподілені за законом Пуасона
 H_1 : $\{v_k\}$ розподілені не за законом Пуасона

Перевірка гіпотез. Для всіх значень V_k виконується розрахунок статистики χ^2 квадрат (3) в залежності від рівня відсікання (графік залежності відображено на рис. 7). Нагадаємо, що гіпотеза, яку розглядаємо, буде підтверджена, якщо значення статистики χ^2 не буде перевищувати відповідне табличне значення. Так для нашого випадку (29 ступенів волі) при рівні значимості 0,3 статистика χ^2 має бути меншою ніж 31,39. Отже, на основі отриманих результатів можемо дійти висновку, що гіпотеза про пуасоновість виходів випадкової величини за граничне значення для виду громадської активності “Демонстрація” приймається для всіх рівнів відсікання в інтервалі від 295 до 2500.

Проте досі не дали відповіді на запитання, яке значення вважати за граничне. Звернемо увагу, що мінімальне значення статистики χ^2 досягається при рівні відсікання 711, саме за цього граничного значення запропонована гіпотеза має підтверджуватися з най-

більшим рівнем довіри. Але чи можна вважати його за істинне?

Підраховавши дні за період із 1 січня 2008 р. по 31 грудня 2010 р., протягом яких зафіксовано понад 711 повідомлень із описом громадської активності “Демонстрація”, з’ясовано, що відсоток таких днів становить 5,7 % від загальної тривалості досліджуваного періоду. Як приклад, у праці А.А.Давидова [7] йдеться про різноманітні відхилення у соціумі та природі: від алкоголізму та вживання наркотиків до генетичних змін людини та максимумів сонячної активності. Автор роботи визначає «константу необхідної дисгармонії», що відповідає відсотку відхилень та аномалій показників різноманітних процесів від норми. Розраховане А.А.Давидовим значення змінної коливається у межах від 4 % до 8 %, що дуже добре підтверджується і нашими результатами роботи.

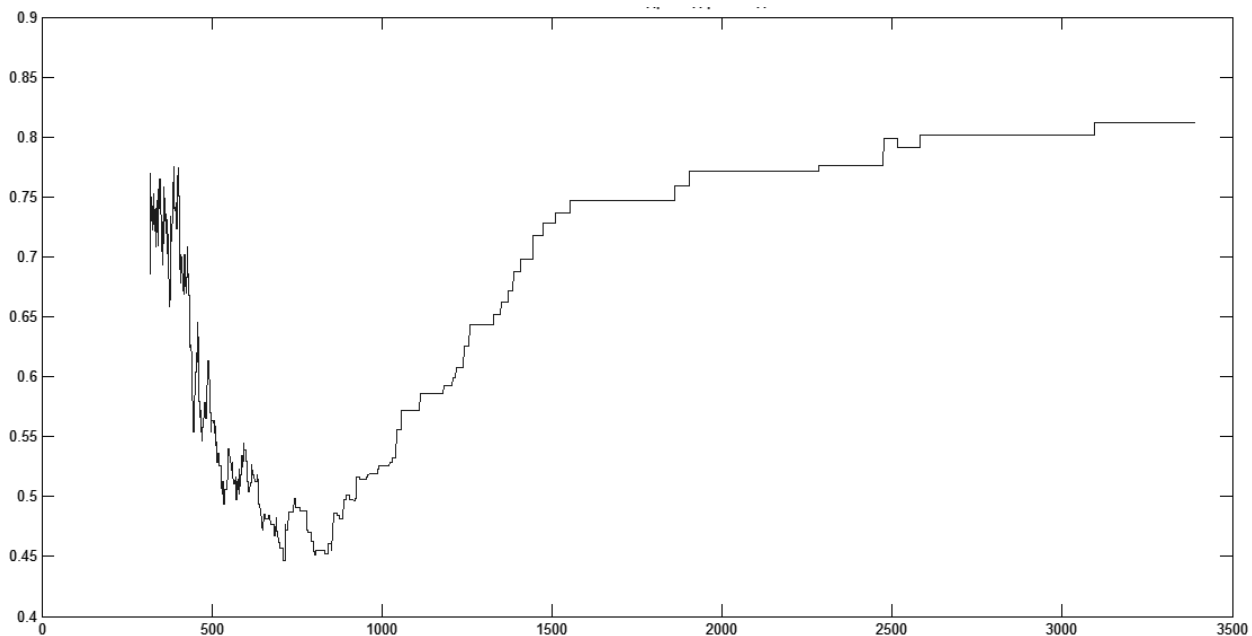


Рис. 7. Залежність статистики χ^2 для перевірки гіпотези про відповідність в.в. u_k закону Пуасона в залежності від граничного значення для дослідження громадської активності “Демонстрації”

Вісь ординат – граничне значення (відлік починається зі значення середньої кількості повідомлень за добу з визначеною тематикою громадської активності), вісь абсцис – значення статистики χ^2 .

Провівши аналогічні розрахунки для інших видів громадської активності, встановлено, що відсоток днів, які відмічені проявами громадської активності, знаходиться у межах 5–7 %.

Висновки. У ході дослідження встановлено, що випадкові величини, котрі відповідають кількості повідомлень із тематикою в розрізі видів активності громадян України,

розподілені у часі за біноміальним законом, кількість виходів таких величин за граничне значення – за законом Пуасона, а кількість днів, що характеризуються значними проявами громадської активності, складають 5–7 %. На думку авторів, отримані результати можуть бути використані для своєчасного виявлення загроз національній безпеці України у внутрішньополітичній сфері. Підкреслимо, що наведені оцінки розраховані на основі інформаційних повідомлень ЗМІ України, а не думок окремих експертів, що дає змогу говорити про більш високий ступінь їх об'єктивності.

Список використаних джерел

1. Шемчушенко Ю.С. Політологічний енциклопедичний словник / Ю.С.Шемчушенко. – К. : Генеза, 2001. – 723 с.
2. Жиро Т. Политология / Т.Жиро ; [науч. ред., авт. предисл. С.В.Быков, И.И.Царьков]. – Х. : Гуманит. центр, 2006. – 425 с.
3. Документація інформаційно-аналітичної системи [Електронний ресурс]. – Режим доступу : <http://online.infostream.ua/>.
4. Горбулін В.П. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання : моногр. / В.П.Горбулін, О.Г.Додонов, Д.В.Ланде. – К. : Інтертехнологія, 2009. – 164 с.

5. Основы моделирования и оценки электронных информационных потоков / Д.В.Ландэ, В.Н.Фурашев, С.М.Брайчевский, А.Н.Григорьев. – К. : ООО “Инжиниринг”, 2006. – 90 с.
6. Всесвітня англomовна Інтернет енциклопедія “Вікіпедія” [Електронний ресурс]. – Режим доступу : <http://ru.wikipedia.org/wiki/Месяц>.
7. Давыдов А.А. Константы в социальных системах / А.А.Давыдов // Вестник РАН. – 1993. – № 8. – С. 733–736.

В статье описаны результаты статистических исследований гражданской активности населения Украины за 2008–2010 годы, выполнен анализ распределения интенсивности ее проявлений во времени в разрезе видов активности, определены характеристики такого распределения. Выполнен анализ выходов качественных показателей гражданской активности за предельные значения, определено процентную долю времени, отмеченного сверхвысокими проявлениями гражданской активности населения.

Ключевые слова: гражданская активность населения, статистические характеристики политической активности.

This article describes the results of statistical researching of the social activity of population of Ukraine for 2008–2010 years, the analysis of distribution of its manifestations in time in the types of activity, characteristics of such distribution. This article contains the analysis of overrunning social activity quality indicators, it is defined percentage share of time, denoted as ultra-high manifestations of social activity of population.

Key words: social participation of the population, the statistical properties of political activity.