

В.П. Малайчук, С.В. Клименко, Н.О. Лисенко

**ІНФОРМАТИВНІСТЬ СТАТИСТИЧНОЇ ОБРОБКИ
ЕКСПЕРИМЕНТАЛЬНИХ ВИМІРЮВАНЬ
МОДИФІКОВАНИМ КРИТЕРІЕМ БУША-ВІНДА**

Анотація. Класична математична статистика пропонує перевіряти статистичну однорідність вибірок нормальних випадкових величин комбінованим критерієм Буша Вінда та оцінювати їх зсуви та масштаби формулами апроксимації законів та функцій розподілу Гауса. Пропонується більш простий аналог критерію Буша Вінда, який формується за формулами логістичних випадкових величин. Шляхом проведення комп'ютерних обчислювальних експериментів оцінюється їх ефективність та факторний аналіз впливу розмірів вибірок і видів статистичних закономірностей досліджуваних випадкових величин, їх асиметричності.

Ключові слова: непараметрична статистика, захист інформації, кібербезпека, однорідність експериментальних вибірок вимірювань, критерій Буша Вінда, обчислювальний експеримент.

Постановка проблеми. Помилки комп'ютерної обробки експериментальних вимірювань, які містять інформацію про стан природних і технічних об'єктів та технологічних процесів, порушують їх кібербезпеку. Дуже важливим є застосування ефективних критеріїв прийняття рішень, особливо, коли йдеться про забезпечення інформаційної безпеки. Контрольовані ознаки, наприклад, характеристики клавіатурного почерку, інтенсивність мережевих атак та багато інших, описуються випадковими величинами, закони розподілу яких, як правило, є невідомими [1].

Аналіз останніх досліджень та публікацій. Класична непараметрична статистика пропонує порівнювати вибірки випадкових величин за ранговими критеріями однорідності, вільних від видів розподілу. Це критерії зсуву вибірок, у яких однакові дисперсії та критерії масштабів (розкиду), якщо у них рівні математичні сподівання. Таких критеріїв існує велика кількість [2]. Використовуючи критерій зсуву Ван-дер-Вардена та критерій масштабу Клотца, Буш і Вінд запропонували комбінований критерій Буша-Вінда. Це асимптотично

оптимальна непараметрична статистика для рівноправного тестування двох нормальних середніх і вибірових дисперсій в генеральній сукупності [3].

Ван-дер-Варден запропонував алгоритм перевірки гіпотези (припущення) про те, що дві вибірки вимірювань $x_1(k)$ і $x_2(k)$ однорідні шляхом їх порівняння. Спочатку формується єдина вибірка розміром $2n$ $x_{12}(i) = x_1(i) + x_2(i - n)$ і потім усі три вибірки перетворюються в упорядковані за зростанням $x_1^*(k)$, $x_{12}^*(i)$ і $x_2^*(k)$.

Ранги впорядкованих вибірок $x_1^*(k)$ і $x_2^*(k)$, як номери їх в вибірці $x_{12}^*(i)$, визначаються за формулами

$$R_{12}(x_1^*(k/j)) = \sum_{i=1}^{2n} \text{sgn}(x_1^*(k/j) - x_{12}^*(i/j)), k = 1, 2, \dots, n;$$

$$R_{12}(x_2^*(k/j)) = \sum_{i=1}^{2n} \text{sgn}(x_2^*(k/j) - x_{12}^*(i/j)), k = 1, 2, \dots, n,$$

де j – номер обчислювального експерименту, $\text{sgn}(x - a)$ – функція одиничного стрибка, дорівнює 1, якщо $x \geq a$, та дорівнює 0, якщо $x < a$.

Якщо вибірки $x_1(k)$ і $x_2(k)$ однорідні, то однорідні вони також і з впорядкованою вибіркою $x_{12}^*(i)$. Ван-дер-Варден пропонує перевірку того, що у них однакова статистика зсувів. Його критерій формується на апроксимаціях функції нормальних (гаусових) випадкових величин

$$F_a(x) = 1 - 0,852 \exp\left(-\frac{(x + 1,574)^{2,34}}{2,0637}\right), x \geq 0.$$

Якщо $x \leq 0$, то $F_a(-x) = 1 - F_a(x)$. Її зворотня функція $x = \psi(F_a(x))$ описується формулою

$$\psi(F_a(x)) = 4,91 \left[(F_a(x))^{0,14} - (1 - F_a(x))^{0,14} \right].$$

Емпіричні неперервні функції розподілу ймовірності $F^*(x)$ експериментальних вибірок формуються за формулою

$$F(x) = \frac{1}{n} \sum_{i=1}^n \text{sgn}(x - x^*(i)).$$

Внаслідок, ранги $R(x^*(k))$ функціонально зв'язані з функцією розподілу ймовірності

$$F^*(x^*(k)) = \frac{1}{n} R(x^*(k)).$$

Використовуючи експериментальні ранги $R_{12}(x_1^*/j)$ і $R_{12}(x_2^*/j)$, Ван-дер-Варден запропонував критерій перевірки рівності зсувів визначати за зворотною емпіричною функцією

$$\psi(F_a^*(x(k))) = 4,91 \left[(F_a^*(x(k)))^{0,14} - (1 - F_a^*(x(k)))^{0,14} \right].$$

Це випадкові величини з нульовим математичним сподіванням і дисперсією

$$D_a[\psi_a(x^*(k))] = \frac{1}{2n-1} \sum_{i=1}^{2n} \psi_a^2\left(\frac{i}{2n+1}\right).$$

Критерій Ван-дер-Вардена описується за формулою

$$B_a(j) = \frac{1}{n} \sum_{k=1}^n \psi_a\left(\frac{R(x^*(k/j))}{2n+1}\right) \left(\sqrt{\frac{1}{2n-1} \sum_{i=1}^{2n} \psi_a^2\left(\frac{i}{2n+1}\right)} \right)^{-1}$$

Клотц запропонував критерії для перевірки гіпотези про рівність в двох вибірках показників масштабу (вибіркових дисперсій) на основі емпіричних оцінок квадрату зворотної функції

$$K_a(j) = \frac{1}{n} \sum_{k=1}^n \psi_a^2\left(\frac{R(x(k/j))}{2n+1}\right) - \frac{1}{2n} \sum_{i=1}^{2n} \psi_a^2\left(\frac{i}{2n+1}\right) \left(\sqrt{\frac{1}{2n-1} \sum_{i=1}^{2n} \psi_a^4\left(\frac{i}{2n+1}\right)} \right)^{-1},$$

де під радикалом дисперсія $\psi_a^2\left(\frac{R(x(k/j))}{2n+1}\right)$.

Буш і Вінд запропонували перевіряти гіпотезу про рівність зсувів і масштабів двох вибірок $x_1(k/j)$ і $x_2(k/j)$ за двомірною функцією модулів критеріїв Ван-дер-Вардена та Клотца

$$B_a(j) = 2 \left[\ln \left(\frac{1}{2(1 - F_a(|B_a(j)|))} \right) + \ln \left(\frac{1}{2(1 - F_a(|K_a(j)|))} \right) \right].$$

Предбачається, що якщо у досліджуваних вибірках $x_1(k)$ і $x_2(k)$ однакові математичні сподівання та дисперсії, тоді критерій Буша-Вінда – це випадкові величини, які описуються законом розподілу хі-квадрат з чотирма ступенями свободи. Якщо $n < 30$, тоді значення порогів рівняння B_0 представлені у таблиці 1.

Значення порогових значень B_0 для критерія Буша-Вінда

$n \backslash n$	110	115	220	225	330
00,45	88,05	88,97	99,05	99,09	99,24

Формулювання мети дослідження. Розглянемо задачу перевірки гіпотези про статистичну однорідність двох експериментальних вибірок вимірювань, якщо критерії Ван-дер-Вардена і Клотца, які формуються за апроксимаціями зворотних функцій Гауса, замінити їх аналогами – зворотними функціями логістичних випадкових величин. Проведемо обчислювальні експерименти та дослідимо інформативність класичного критерію Буша-Вінда і його аналога, який формується на логістичній зворотній функції розподілу.

Виклад основного матеріалу. Сформуємо вибірки випадкових величин з трьома видами статистичних закономірностей – логістичні, релєєвські і експоненціальні з однаковими математичними очікуваннями. Оцінимо за критерієм Буша-Вінда їх однорідність. Функції розподілу ймовірності описуються формулами

$$F_1(x) = \left(1 + \exp \left(- \frac{\pi}{\sqrt{3}} \frac{x-a}{\sigma} \right) \right)^{-1}, \quad F_2(x) = 1 + \exp \left(- \frac{x^2}{2b^2} \right), \quad F_3(x) = 1 + \exp \left(- \frac{x}{\lambda} \right).$$

Математичні очікування таких вибірок дорівнюють

$$M[x_1] = a, \quad D[x_1] = \sigma^2; \quad M[x_2] = \sqrt{\frac{\pi}{2}}b, \quad D[x_2] = \frac{4-\pi}{2}b^2; \quad M[x_3] = \lambda, \quad D[x_3] = \lambda^2.$$

Використовуючи програмні генератори $\xi_1(k)$, $\xi_2(k)$, $\xi_3(k)$ вибірок випадкових величин з рівномірним законом розподілу $w(\xi) = 1$ і функцією розподілу вибірок $F(\xi) = \xi$, сформуємо вибірки з логістичним $x_1(k)$, релєєвським $x_2(k)$ і експоненціальним $x_3(k)$ законами розподілу ймовірності.

Закон збереження ймовірності при перетворенні випадкових величин $\xi(k)$ в $x(k)$ описується диференціальним рівнянням $w[\xi]d\xi = W(x)dx$. Вирішивши його, отримаємо формулу функціональної залежності $\xi(k) = F(x(k))$. Використовуючи її обернену функцію $x(k) = \Psi(\xi(k))$, отримаємо формули перетворення вибірки $\xi(k)$ у вибірку $x(k)$. Це математичні моделі програмних генераторів вибірок випадкових величин із заданими статистичними закономірностями

$$x_1(k) = a + \sigma \frac{\sqrt{3}}{\pi} \ln \left(\frac{\xi_1(k)}{1 - \xi_1(k)} \right), \quad x_2(k) = \sqrt{2b^2 \ln \left(\frac{1}{1 - \xi_2(k)} \right)}, \quad x_3(k) = \lambda \ln \left(\frac{1}{1 - \xi_3(k)} \right).$$

Проведемо обчислювальні експерименти і оцінимо інформативність критерію Буша - Вінда і його модифікованого аналогу. Класичний критерій Буша-Вінда формується по зворотній функції Гауса. Аналогом її може бути зворотна логістична функція

$$\psi_x(F(x)) = \frac{\sqrt{3}}{\pi} \ln \left(\frac{F(x)}{1 - F(x)} \right),$$

яка записується у вигляді

$$\psi_x \left(\frac{R \left(x^* \left(\frac{k}{j} \right) \right)}{2n+1} \right) = \frac{\sqrt{3}}{\pi} \ln \left(\frac{R \left(x^* \left(\frac{k}{j} \right) \right)}{2n+1 - R \left(x \left(\frac{k}{j} \right) \right)} \right).$$

Якщо в формулах $B_a(j)$ и $K_a(j)$ функцію $\psi_a \left(\frac{R \left(x \left(\frac{k}{j} \right) \right)}{2n+1} \right)$ замінити на

функцію $\psi_x \left(\frac{R \left(x \left(\frac{k}{j} \right) \right)}{2n+1} \right)$, то отримаємо

$$B_x(j) = \frac{\frac{1}{n} \sum_{k=1}^n \psi_x \left(\frac{R \left(x^* \left(\frac{k}{j} \right) \right)}{2n+1} \right)}{\sqrt{\frac{1}{2n-1} \sum_{i=1}^{2n} \psi_x^2 \left(\frac{i}{2n+1} \right)}}, \quad K_x(j) = \frac{\frac{1}{n} \sum_{k=1}^n \psi_x^2 \left(\frac{R \left(x \left(\frac{k}{j} \right) \right)}{2n+1} \right) - \frac{1}{2n} \sum_{i=1}^{2n} \psi_x^2 \left(\frac{i}{2n+1} \right)}{\sqrt{\frac{1}{2n-1} \sum_{i=1}^{2n} \psi_x^4 \left(\frac{i}{2n+1} \right)}}.$$

В формулі $B_a(j)$ функції $F_a(B(j))$ і $F_a(K(j))$ на функції $F_x(B_x(j))$ і $F_x(K_x(j))$, то аналог критерію Буша-Вінда $B_x(j)$ визначимо за формулами

$$F_x(|B_x(j)|) = \frac{1}{1 + \frac{\pi}{\sqrt{3}} \exp(-|B_x(j)|)},$$

$$F_x(|K_x(j)|) = \frac{1}{1 + \frac{\pi}{\sqrt{3}} \exp(-|K_x(j)|)},$$

$$B_{\lambda}(j) = 2 \left[\ln \left(\frac{0,5 \left(1 + \frac{\pi}{\sqrt{3}} \exp(-|B_{\lambda}(j)|) \right)}{\frac{\pi}{\sqrt{3}} \exp(-|B_{\lambda}(j)|)} \right) + \ln \left(\frac{0,5 \left(1 + \frac{\pi}{\sqrt{3}} \exp(-|K_{\lambda}(j)|) \right)}{\frac{\pi}{\sqrt{3}} \exp(-|K_{\lambda}(j)|)} \right) \right].$$

Проведемо N обчислювальних експериментів та оцінимо ефективність критеріїв $B_a(j)$ та $B_{\lambda}(j)$. Отримаємо кількості N_a і N_{λ} виконання нерівностей $B_a(j) < B_0$ і $B_{\lambda}(j) < B_0$. Їх відношення $\frac{N_a}{N} = P_a^*$ і $\frac{N_{\lambda}}{N} = P_{\lambda}^*$ - оцінки ймовірностей прийняття вірних рішень про статистичну однорідність досліджуваних вибірок вимірювань.

Перший обчислювальний експеримент. Проведемо дослідження статистичних закономірностей показників $B_a(j)$ та $B_{\lambda}(j)$. Обчислимо мінімальні $B_{\min}$, максимальні $B_{\max}$ та середні значення $\bar{B}$, вибіркові дисперсії D_B^* , оцінки емпіричної ймовірності вірного розпізнавання однорідності вибірок P_{BP}^* та побудуємо гістограми досліджуваних показників $B_a(j)$ та $B_{\lambda}(j)$.

Статистичні показники критерію Буша-Вінда $\bar{B}_a(j)$ та його аналогу $\bar{B}_{\lambda}(j)$ для трьох видів однорідних некорельованих випадкових величин довжиною $n = 100$ представлені в таблиці 2.

Таблиця 2

Статистичні показники $\bar{B}_a(j)$ та $\bar{B}_{\lambda}(j)$ для трьох видів однорідних некорельованих випадкових величин довжиною $n = 100$

Показники	Логістичні в.в.		Релеєвські в.в.		Експоненціальні в.в.	
	$\bar{B}_a(j)$	$\bar{B}_{\lambda}(j)$	$\bar{B}_a(j)$	$\bar{B}_{\lambda}(j)$	$\bar{B}_a(j)$	$\bar{B}_{\lambda}(j)$
$B_{\max}$	14.116	12.624	15.386	13.824	13.857	12.941
$B_{\min}$	0.08	0.039	0.155	0.036	0.016	0.114
$\bar{B}$	3.41	3.748	3.449	3.754	3.394	3.736
D_B^*	5.441	5.283	5.736	5.349	4.989	4.871
P_{BP}^*	0.992	0.997	0.987	0.992	0.993	0.998

Модифікований показник Буша-Вінда $\bar{B}_{\lambda}(j)$ незначно відрізняється від класичного показника $\bar{B}_a(j)$, спостерігається зміщення середнього значення в більшу сторону (на 9%). Вибіркова дисперсія $\bar{B}_{\lambda}(j)$ менше від вибіркової

дисперсії $\bar{B}_a(j)$ (в середньому на 4%), а ймовірність прийняття вірного рішення більше на 0,5 %.

При довжині вибірок $n > 30$ для класичного критерію Буша-Вінда справедлива апроксимація розподілом хі-квадрат з $f = 4$ ступенями свободи. Наведено на рис.1 гістограми показників класичного критерія Буша-Вінда $\bar{B}_a(j)$ та його аналога $\bar{B}_л(j)$ разом з кривою закону розподілу хі-квадрат з чотирма ступенями свободи. Гістограми показників $\bar{B}_a(j)$ та $\bar{B}_л(j)$ побудуємо для трьох видів досліджуваних однорідних некорельованих вибірок вимірювань довжиною $n = 100$.

При великих розмірах вибірок $n > 100$ запропонований аналог критерію Буша-Вінда не дуже значно, але більш ефективний, ніж класичний критерій.

Другий обчислювальний експеримент. При $n > 30$ показники критерію Буша-Вінда – це випадкові величини, які описуються законом розподілу хі-квадрат з чотирма ступенями свободи. Перевіримо гіпотези про закони розподілу ймовірності досліджуваних показників $B_a(j)$ та $B_л(j)$ за критерієм хі-квадрат [4].

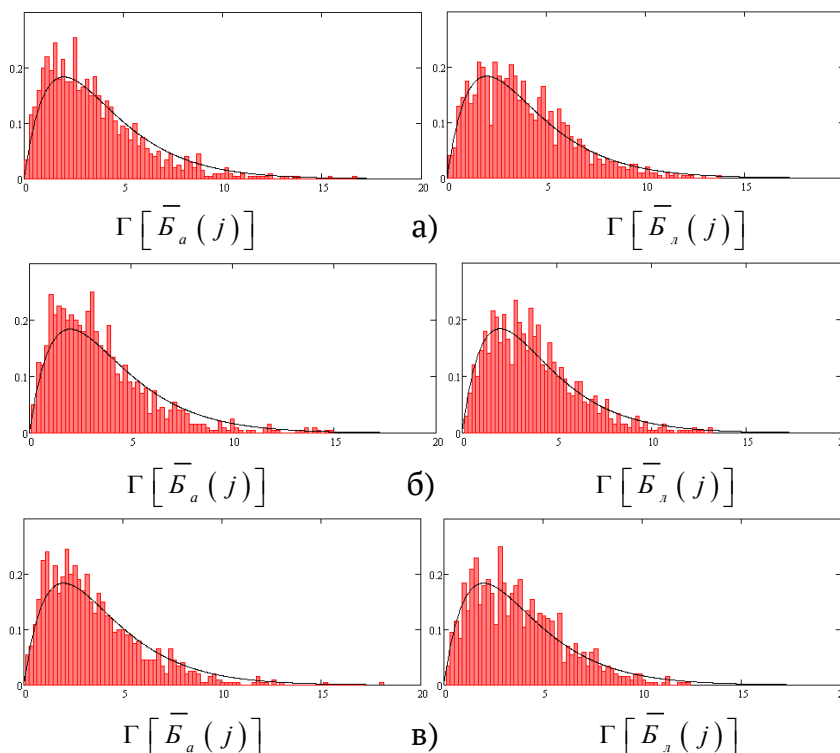


Рисунок 1 - Крива закону розподілу хі-квадрат з чотирма ступенями свободи та гістограми показників $\bar{B}_a(j)$ і $\bar{B}_л(j)$ для вибірок а) логістичних, б) релеєвських та в) експоненціальних випадкових величин ($n = 100$)

Задамо кількість інтервалів гістограми L та обчислимо інтервал дискретизації

$$\Delta x = \frac{x_{\max} - x_{\min}}{L},$$

де $x_{\min}$ - мінімальне значення вибірки вимірювань та $x_{\max}$ - максимальне значення вибірки вимірювань.

Гістограмні числа $k_1, k_2, \dots, k_L$ - це кількість вимірювань досліджуваної вибірки, які влучили в інтервали дискретизації $x_{0i} \pm \Delta x / 2$, где $x_{0i} = x_{01} + (i-1)\Delta x$, $x_{01} = x_{\min} + \Delta x / 2$. Якщо використати функцію Хевісайда (одиничного стрибка) для визначення гістограмних чисел, то алгоритм їх визначення буде мати вигляд

$$k_i = \sum_{k=1}^n [\operatorname{sgn}(x_k - x_{0i} + \Delta x / 2) - \operatorname{sgn}(x_k - x_{0i} - \Delta x / 2)],$$

де $\operatorname{sgn}(x)$ - функція одиничного стрибка приймає лише два значення: $\operatorname{sgn}(x) = 0$, якщо $x < 0$, та $\operatorname{sgn}(x) = 1$, якщо $x \geq 0$.

Якщо закон розподілу вимірювань $w(x)$, то ймовірності влучення вимірювань в i -тий інтервал гістограми p_i визначаються за формулою

$$p_i = \int_{x_{0i} - \frac{\Delta x}{2}}^{x_{0i} + \frac{\Delta x}{2}} W(x) dx \approx W(x_{0i}) \Delta x, \quad i = 1, 2, 3, \dots, L,$$

де x_{0i} - центр i -го інтервалу гістограми.

Очевидно, що відносні гістограмні числа $p_i^* = \frac{k_i}{n}$ - є оцінками ймовірностей p_i . Якщо ймовірності p_i визначені за гіпотетичним законом розподілу $w(x)$, то зважена сума квадратів різниці p_i и p_i^* може бути показником близькості $w(x)$ та невідомого закону досліджуваної вибірки вимірювань

$$z = \sum_{i=1}^L \left(\frac{p_i^* - p_i}{\sqrt{D_i}} \right)^2 = \sum_{i=1}^L \frac{(k_i - np_i)^2}{np_i(1 - p_i)}.$$

Рішення про вірність висунутої гіпотези слід приймати шляхом порівняння обчисленого значення показника z з деяким граничним значенням z_0 : якщо $z \leq z_0$, то слід вважати, що гіпотеза вірна, якщо $z > z_0$, то гіпотезу слід відхилити та запропонувати іншу. Поріг порівняння z_0 визначають за правилом Неймана - Пірсона із умови: ймовірність помилки 1-го роду не повинна

перевищувати деяке граничне значення P_0 (наприклад, $P(H/B) \leq 0.05$). Приблизне значення z_0 можна оцінити за формулою

$$z_0 = L \left(1 - \frac{2}{9L} + \Psi(P) \sqrt{\frac{2}{9L}} \right)^3,$$

де $\Psi(x)$ - функція, зворотна інтегралу ймовірності Гауса. Якщо декілька гіпотез задовільняють нерівності $z \leq z_0$, то треба обирати гіпотезу, для якої z мінімально [4].

Перевіримо гіпотезу про закон розподілу хі-квадрат з чотирма ступенями свободи для показників класичного критерія Буша-Вінда $B_a(j)$ та його модифікованого аналога $B_n(j)$.

Таблиця 3

Значення показників близькості z , граничні значення z_0 та результати роботи критерію хі-квадрат для $B_a(j)$ та $B_n(j)$ за трьома видами однорідних некорельованих випадкових величин довжиною $n = 100$

Довжина вибірок	Показники близькості та граничні значення	Логістичні в.в.		Релесівські в.в.		Експоненціальні в.в.	
		$B_a(j)$	$B_n(j)$	$B_a(j)$	$B_n(j)$	$B_a(j)$	$B_n(j)$
100	z	16.951	10.998	22.34	17.107	23.409	18.808
	z_0	21.013		22.78		28.86	
50	z	17,462	6,858	21,332	8,734	16,555	11,945
	z_0	19,66		22,35		27,58	
30	z	10,462	8,804	19,004	12,011	16,273	14,53
	z_0	21,43		23,07		24,98	

Як видно з таблиці 3, показники Буша-Вінда $B_a(j)$ та $B_n(j)$ задовільняють умові про те, що вони описуються законом розподілу хі-квадрат з чотирма ступенями свободи (за критерієм хі-квадрат $z \leq z_0$). Аналог критерію Буша-Вінда демонструє менші показники близькості z , що свідчить

про те, що висунута гіпотеза про закон розподілу ймовірності підходить йому краще, ніж класичному показнику $B_a(j)$. Отже, модифікований показник $B_{\lambda}(j)$ з більшою ймовірністю описується законом розподілу хі-квадрат з чотирма ступенями свободи, ніж класичний варіант критерію, з використанням апроксимації нормальної функції розподілу.

Третій обчислювальний експеримент. Практична значущість критерія Буша-Вінда щільно пов'язана з довжиною вибірок вимірювань. Якщо критерій характеризується високими значеннями ймовірностей вірного розпізнавання навіть в умовах обмеженості вхідних даних, то його можна рекомендувати для практичного застосування. Перевіримо ефективність класичного та модифікованого критеріїв Буша-Вінда, застосовуючи короткі вибірки вимірювань $n = 10; 20$.

Обчислимо статистичні характеристики показників $B_a(j)$ та $B_{\lambda}(j)$ та представимо їх в таблиці 4. Побудуємо гістограми досліджуваних показників $\bar{B}_a(j)$ та $\bar{B}_{\lambda}(j)$ для трьох видів однорідних некорельованих коротких вибірок вимірювань. На рис. 2 представлені гістограми показників класичного критерія Буша-Вінда $\bar{B}_a(j)$ та його аналога $\bar{B}_{\lambda}(j)$ разом з кривою закону розподілу хі-квадрат з чотирма ступенями свободи для довжини вибірок $n = 10$.

Таблиця 4

Статистичні показники $\bar{B}_a(j)$ та $\bar{B}_{\lambda}(j)$ критерію Буша-Вінда для трьох видів однорідних некорельованих випадкових величин довжиною $n = 10$

	Логістичні в.в.		Релеєвські в.в.		Експоненціальні в.в.	
	$\bar{B}_a(j)$	$\bar{B}_{\lambda}(j)$	$\bar{B}_a(j)$	$\bar{B}_{\lambda}(j)$	$\bar{B}_a(j)$	$\bar{B}_{\lambda}(j)$
$B_{\max}$	11,147	9,33	9,797	8,628	8,463	7,958
$B_{\min}$	0,159	0,039	0,209	0,345	0,293	0,45
$\bar{B}$	3,225	3,494	3,402	3,657	3,523	3,807
D_B^*	4,197	3,859	4,805	4,435	4,376	4,263
P_{BP}^*	0,992	0,997	0,991	0,997	0,992	0,999

Таблиця 5

Статистичні показники $\bar{B}_a(j)$ та $\bar{B}_l(j)$ критерію Буша-Вінда для трьох видів
однорідних некорельованих випадкових величин довжиною $n = 20$

	Логістичні в.в.		Релеєвські в.в.		Експоненціальні в.в.	
	$\bar{B}_a(j)$	$\bar{B}_l(j)$	$\bar{B}_a(j)$	$\bar{B}_l(j)$	$\bar{B}_a(j)$	$\bar{B}_l(j)$
$B_{\max}$	14,47	11,191	13,689	11,684	13,497	12,353
$B_{\min}$	0,036	0,059	0,15	0,005	0,029	0,115
$\bar{B}$	3,464	3,789	3,376	3,668	3,395	3,693
D_B^*	4,822	4,409	4,962	4,557	4,77	4,572
P_{BP}^*	0,992	0,998	0,991	0,997	0,995	0,997

При порівнянні даних таблиці 2 (для довгих вибірок вимірювань, $n = 100$) з даними таблиць 3 та 4 (для коротких вибірок вимірювань, $n = 10$ та 20) можна зробити наступні висновки.

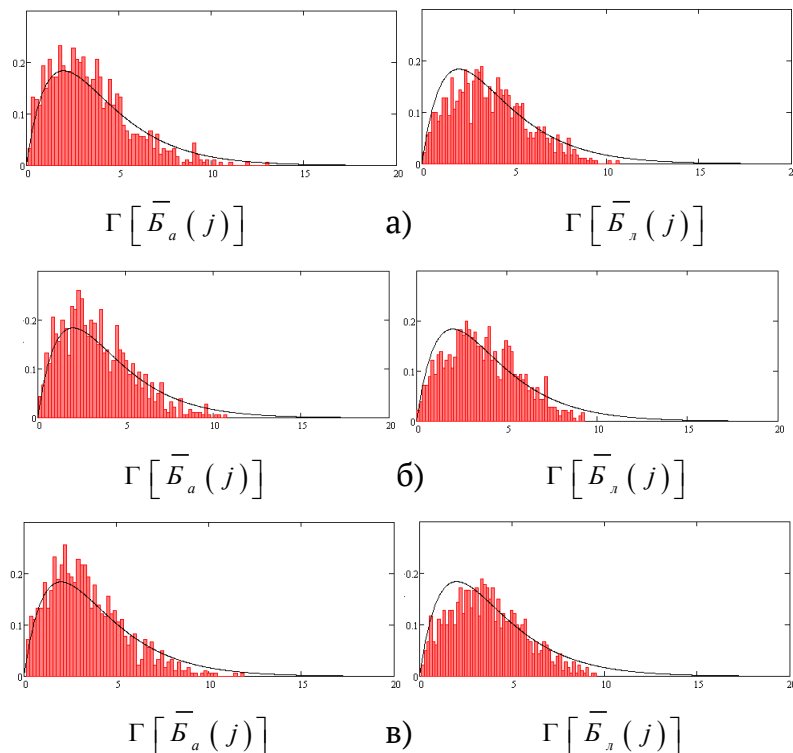


Рисунок 2 - Графік закону розподілу хі-квадрат з чотирма ступенями свободи та гістограми показників $\bar{B}_a(j)$ і $\bar{B}_l(j)$ для вибірок а) логістичних, б) релеєвських та в) експоненціальних випадкових величин ($n = 10$)

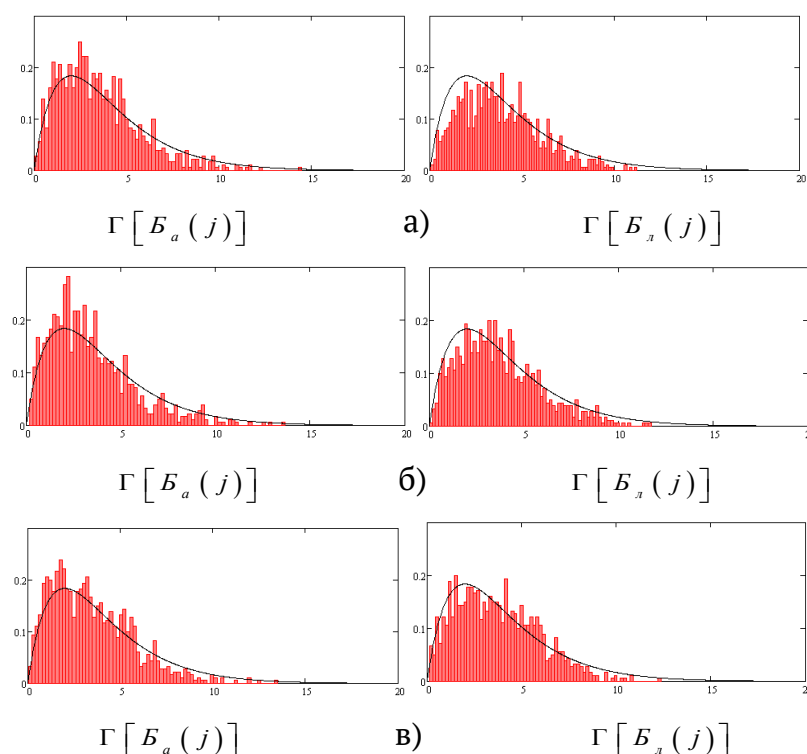


Рисунок 3 - Графік закону розподілу хі-квадрат з чотирма ступенями свободи та гістограми показників $\bar{B}_a(j)$ і $\bar{B}_n(j)$ для вибірок а) логістичних, б) релеєвських та в) експоненціальних випадкових величин ($n = 20$)

При використанні вибірок вимірювань довжиною $n = 10; 20$ середні значення класичного і модифікованого показників критерію Буша-Вінда залишаються без суттєвих змін і відхиляються від середніх значень, отриманих при використанні вибірок довжиною $n = 100$, не більше ніж на 7%. При зменшенні довжини вибірок вимірювань звужується діапазон розмаху $B_{\max} - B_{\min}$, а вибіркові дисперсії показників $\bar{B}_a(j)$ і $\bar{B}_n(j)$ зменшуються в середньому на 16%.

Особливо слід звернути увагу на те, що значення ймовірностей прийняття вірного рішення, як за класичним критерієм Буша-Вінда, так і за його аналогом, при застосуванні коротких вибірок вимірювань ($n = 10; 20$) залишаються на такому ж високому рівні, як і при застосуванні довгих вибірок вимірювань випадкових величин ($n = 100$).

Отже, можна зробити висновок про ефективність класичного та модифікованого критеріїв Буша-Вінда при застосуванні коротких вибірок вимірювань ($n = 10; 20$).

Висновки. Запропонований в роботі аналог критерію Буша-Вінда відрізняється від класичного критерію зменшенням обчислювальної складності зі збереженням ефективності. Емпіричні ймовірності розпізнавання однорідності вибірок, які отримані шляхом проведення обчислювальних експериментів для вибірок логістичних, релеевських та експоненціальних випадкових величин, свідчать про непараметричність, високу чутливість та можливість застосування критерію в умовах обмеженості експериментальних даних.

Модифікований критерій Буша-Вінда характеризується високою інформативністю та може бути рекомендований для статистичної обробки експериментальних вимірювань.

ЛІТЕРАТУРА

1. Хиценко В.Е. Ранговые критерии в задачах защиты информации // Доклады ТУСУРа, № 1 (25), часть 2, июнь 2012, С. 29-33.
2. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006. – 816 с.
3. Bush J.R., Wieand H.S. An asymptotically optimal nonparametric statistic for testing equally of two normal population means and variances // Commun. Stat.-Theor. Meth. 1982. V.11, №1. P. 1-12.
4. Малайчук В.П. Статистична теорія оцінювання [Текст]: Навчальний посібник / В.П. Малайчук, Н.О. Лисенко. – Дніпропетровськ: РВВ ДНУ, 2009. – 84 с.

REFERENCE

1. Khitsenko V.E. Rank criteria in information security tasks // Reports of TUSUR, No. 1 (25), part 2, June 2012, pp. 29-33.
2. Kobzar A. I. Applied mathematical statistics/ A. And yu Kobzar - M. : FIZMATLIT, 2006. - 816 p.
3. Bush J.R., Wieand H.S. An asymptotically optimal nonparametric statistic for testing equally of two normal population means and variances // Commun. Stat.-Theor. Meth. 1982. V.11, №1. P. 1-12.
4. Malaychuk V.P. Statistical theory of estimation [Text]: Manual / V.P. Malaychuk, N.O. Lysenko. - Dnipropetrovsk: RVV DNU, 2009. - 84 p.

Received 10.11.2022.

Accepted 14.11.2022.

Informativeness of statistical processing of experimental measurements by the modified Bush-Wind criterion

The use of effective decision-making criteria is very important, especially when it comes to ensuring information security. Controlled attributes, such as keyboard handwriting characteristics, intensity of network attacks, and many others, are described by

random variables whose distribution laws are usually unknown. Classical nonparametric statistics suggests comparing samples of random variables by rank-based homogeneity criteria that are independent of the type of distribution. Using the Van der Warden shift criterion and the Klotz scale criterion, Bush and Wind proposed the combined Bush-Wind criterion. It is an asymptotically optimal nonparametric statistic for equal testing of two normal means and sample variances in a population. The article considers the problem of testing the hypothesis of statistical homogeneity of two experimental measurement samples if the Van der Warden and Klotz criteria, which are formed by approximations of the inverse Gaussian functions, are replaced by their analogues - the inverse functions of logistic random variables. Computational experiments are carried out and the informativeness of the classical Bush-Wind criterion and its analog, which is formed on the logistic inverse distribution function, is investigated. The analog of the Bush-Wind criterion proposed in this paper differs from the classical criterion by reducing computational complexity while maintaining efficiency. The empirical probabilities of recognizing the homogeneity of samples, obtained by conducting computational experiments for samples of logistic, Rayleigh and exponential random variables, indicate nonparametricity, high sensitivity and the possibility of applying the criterion in conditions of limited experimental data. The modified Bush-Wind criterion is characterized by high information content and can be recommended for statistical processing of experimental measurements.

Малайчук Валентин Павлович - професор кафедри кібербезпеки та комп'ютерно-інтегрованих технологій, фізико-технічний факультет, Дніпровський національний університет імені Олеся Гончара.

Клименко Світлана Володимирівна - доцент кафедри кібербезпеки та комп'ютерно-інтегрованих технологій, фізико-технічний факультет, Дніпровський національний університет імені Олеся Гончара.

Лисенко Наталія Олександрівна - доцент кафедри кібербезпеки та комп'ютерно-інтегрованих технологій, фізико-технічний факультет, Дніпровський національний університет імені Олеся Гончара.

Malaichuk Valentin Pavlovich - professor of the department of cybersecurity and computer-integrated technologies, faculty of physics and technology, Oles Honchar Dnipro national university.

Klymenko Svitlana Volodimirivna - docent of the department of cybersecurity and computer-integrated technologies, faculty of physics and technology, Oles Honchar Dnipro national university.

Lysenko Nataliia Oleksandrivna - docent of the department of cybersecurity and computer-integrated technologies, faculty of physics and technology, Oles Honchar Dnipro national university.