

В.В. РОМАНЮК, канд. техн. наук, доц. ХНУ (м. Хмельницький)

ОЦІНЮВАННЯ ВІРОГІДНОСТІ РОЗПОДІЛУ СТАТИСТИЧНИХ ЧАСТОТ ВИПАДКОВОЇ ВЕЛИЧИНИ З НЕВІДОМИМ МАТЕМАТИЧНИМ СПОДІВАННЯМ І ДИСПЕРСІЄЮ

Побудовано ядро антагоністичної гри для задачі безумовної оптимізації, за розв'язком якої пропонується приймати рішення про достатність проведених вимірювань випадкової величини для того, щоб її реалізовувати у відповідній математичній моделі у формі розподілу відносних статистичних частот. Представлено програму підтримки прийняття рішення про вірогідність досліджуваного розподілу. Іл.: 4. Бібліогр.: 8 назв.

Ключові слова: антагоністична гра, випадкова величина, розподіл статистичних частот, прийняття рішення.

Постановка проблеми. Вимірювання параметрів випадкових величин є складовою математичного моделювання явищ і процесів, де фігурують ці величини. Зазвичай випадкова величина Θ у математичній моделі реалізуються у формі нормованого частотного розподілу

$$F(\Theta) = [f(\theta_1) \ f(\theta_2) \ \dots \ f(\theta_{n-1}) \ f(\theta_n)] \quad (1)$$

скінченного набору

$$[\theta_1 \ \theta_2 \ \dots \ \theta_{n-1} \ \theta_n] \quad (2)$$

своїх значень, де $f(\theta_k)$ є відносною частотою спостереження значення θ_k

при $\sum_{k=1}^n f(\theta_k) = 1$ та $f(\theta_k) \in (0; 1) \ \forall k = \overline{1, n}$. Якщо вектор-набір (2) подається

відсортованим за зростанням, а про характер розподілу (1) нічого невідомо, то слушним буде перевірка нульової гіпотези про достатність проведених вимірювань над полем значень (2) для того, щоб надалі працювати з даними (1) у досліджуваній математичній моделі. Проте таке оцінювання вірогідності розподілу статистичних частот (1) випадкової величини Θ не може опиратись на відомі методи перевірки гіпотези про той чи інший вид розподілу, адже тут існує декілька проблем. По-перше, нам невідомі математичне сподівання і дисперсія випадкової величини Θ , тому, можливо, проведених вимірювань недостатньо. По-друге, яке є наслідком першого, якщо проведених вимірювань недостатньо, то це означає також і потенційну необхідність розширення відрізка спостережень $[\theta_1; \theta_n]$. Тому перевірку розподілу статистичних частот (1) випадкової величини Θ на вірогідність слід проводити у більш широкому розумінні.

Аналіз останніх досліджень і публікацій. Питання перевірки гіпотез про тип розподілу генеральної сукупності за даними емпіричних розподілів глибоко висвітлюються у всіх посібниках і підручниках з теорії імовірностей та математичної статистики, зокрема, в [1 – 5]. Але там не ставиться питання про достатність проведених вимірювань для того, щоб реалізувати частотний розподіл у досліджуваній математичній моделі. Крім того, якщо для спостережень узято недостатньо широкий інтервал значень випадкової величини, що може бути обумовлено обмеженнями по часу вимірювань або контактом з об'єктом, то це, взагалі кажучи, додатково спотворює припущення про реальний розподіл генеральної сукупності.

Мета і постановка завдань статті. Будемо відштовхуватись від того, що нам невідомі математичне сподівання $M(\Theta)$ і дисперсія $D(\Theta)$ випадкової величини Θ , але після проведення деяких спостережень (прямих чи непрямих вимірювань) ми отримали скінченний відсортований набір (2) з нормованим частотним розподілом (1), де $\theta_k < \theta_{k+1} \quad \forall k = \overline{1, n-1}$. Оскільки $M(\Theta)$ є невідомим, то ми не можемо стверджувати, що $M(\Theta) \approx \arg \max_{\{\theta_k\}_{k=1}^n} f(\theta_k)$, адже

на відрізку спостережень $[\theta_1; \theta_n]$ може трапитись декілька максимальних значень вектора $F(\Theta)$ з однаковими значеннями. Крім того, якщо, скажімо, у нас $F(\Theta) = [0.4 \quad 0.25 \quad 0.25 \quad 0.1]$, то можна припускати, що тут математичне сподівання $M(\Theta)$ у межах відрізка $[\theta_1; \theta_4]$ зміщено ліворуч, і навіть цілком вірогідна ситуація $M(\Theta) < \theta_1$ (котра може бути викликана, наприклад, некоректною постановкою експериментальних спостережень). Тому метою цієї статті є оцінка того, чи ми можемо вважати отриманий у процесі деяких спостережень розподіл (1) скінченного набору (2) значень випадкової величини Θ адекватним для його подальшого застосування у відповідній математичній моделі. Для цього необхідно побудувати модель оцінювання вірогідності розподілу статистичних частот (1) випадкової величини Θ для її значень (2), де математичне сподівання $M(\Theta)$ і дисперсія $D(\Theta)$ є невідомими. Таку модель реалізовуватимемо у формі структурованої задачі прийняття рішень з частково невідомою інформацією, тобто у формі антагоністичної гри.

Побудова ядра антагоністичної гри. Сформулюємо задачу прийняття рішень, у якій ми зацікавлені у такому виборі значень випадкової величини Θ з підмножини $\{\theta_k\}_{k=1}^n$ відрізка $[\theta_1; \theta_n]$, який призведе до найменших втрат, котрі зумовлюються відхиленням прийнятого у даний момент значення цієї випадкової величини від істинного її значення $M(\Theta)$. Значить, у відповідній антагоністичній грі ми будемо виконувати роль другого гравця, намагаючись

мінімізувати свої втрати [6, 7]. Різні випадкові обставини, що, власне, і зумовлюють випадковий характер величини Θ , уособлюватимуть першого гравця [8].

Позначимо через x та y чисті стратегії першого і другого гравців відповідно, де множини $\{x_k\}_{k=1}^n$ та $\{y_k\}_{k=1}^n$ чистих стратегій гравців співпадають, у яких їх k -й елемент відповідає обираю значення θ_k . Тоді ядро $K(x, y)$ антагоністичної гри у точці (x, y) буде пропорційним модулю різниці $x - y$. При цьому коефіцієнт пропорційності є додатною функцією S від модуля різниці $x - y$, і ця функція $S(|x - y|)$ є монотонно неспадною на підмножині $\{\theta_k\}_{k=1}^n$ відрізка $[\theta_1; \theta_n]$. Таким чином, ядром антагоністичної гри, яку ми використаємо для оцінювання вірогідності розподілу статистичних частот (1) випадкової величини Θ для її значень (2), буде квадратна матриця n -го порядку $\mathbf{K} = (a_{ij})_{n \times n}$ з елементами

$$a_{ij} = |\theta_i - \theta_j| S(|\theta_i - \theta_j|) \text{ при } i = \overline{1, n} \text{ та } j = \overline{1, n}. \quad (3)$$

Головна діагональ цієї матриці є, очевидно, нульовою:

$$\mathbf{K} = \begin{pmatrix} 0 & |\theta_1 - \theta_2| S(|\theta_1 - \theta_2|) & \dots \\ |\theta_1 - \theta_2| S(|\theta_1 - \theta_2|) & 0 & \dots \\ \vdots & \vdots & \ddots \\ |\theta_1 - \theta_{n-1}| S(|\theta_1 - \theta_{n-1}|) & |\theta_2 - \theta_{n-1}| S(|\theta_2 - \theta_{n-1}|) & \dots \\ |\theta_1 - \theta_n| S(|\theta_1 - \theta_n|) & |\theta_2 - \theta_n| S(|\theta_2 - \theta_n|) & \dots \\ \dots & |\theta_1 - \theta_{n-1}| S(|\theta_1 - \theta_{n-1}|) & |\theta_1 - \theta_n| S(|\theta_1 - \theta_n|) \\ \dots & |\theta_2 - \theta_{n-1}| S(|\theta_2 - \theta_{n-1}|) & |\theta_2 - \theta_n| S(|\theta_2 - \theta_n|) \\ \ddots & \vdots & \vdots \\ \dots & 0 & |\theta_{n-1} - \theta_n| S(|\theta_{n-1} - \theta_n|) \\ \dots & |\theta_{n-1} - \theta_n| S(|\theta_{n-1} - \theta_n|) & 0 \end{pmatrix}. \quad (4)$$

Також очевидна і симетричність матриці (4): $\mathbf{K} = \mathbf{K}^T$.

Зауважимо, що у матричній грі з матрицею $\mathbf{K} = (a_{ij})_{n \times n}$ обидва гравці володіють однаково потужними множинами своїх змішаних стратегій. Одну з них, а саме – множину змішаних стратегій другого гравця, позначимо як $\mathcal{Q}$, де

$$\mathcal{Q} = \left\{ \mathbf{Q} = [q_1 \quad q_2 \quad \cdots \quad q_{n-1} \quad q_n] \in \mathbb{R}^n : q_k \in [0; 1] \forall k = \overline{1, n}, \sum_{k=1}^n q_k = 1 \right\}. \quad (5)$$

Звісно, у (5) значення q_k є імовірністю обиравання чистої стратегії y_k , тобто імовірністю прийняття рішення про використання у досліджуваній (вихідній) математичній моделі значення θ_k випадкової величини Θ .

Модель прийняття рішення про вірогідність розподілу (1). Взагалі кажучи, в $n \times n$ -грі з матрицею $\mathbf{K} = (a_{ij})_{n \times n}$ завданням другого гравця є мінімізація добутку $\mathbf{P} \cdot \mathbf{K} \cdot \mathbf{Q}^T$ на множині (5), де $\mathbf{P}$ є деякою змішаною стратегією першого гравця (з тієї ж, в принципі, множини). Але тут буде слушним покладання того, що у грі з матрицею (4) перший гравець уже обрав свою змішану стратегію у формі розподілу статистичних частот (1), яка, до речі, є цілком змішаною. Тоді другому гравцю залишається розв'язати задачу

$$\arg \min_{\mathbf{Q} \in \mathcal{Q}} (\mathbf{F}(\Theta) \cdot \mathbf{K} \cdot \mathbf{Q}^T). \quad (6)$$

Покажемо, що ця задача безумовної оптимізації завжди має розв'язок у формі чистої стратегії другого гравця.

Теорема 1. Задача безумовної оптимізації (6) завжди має хоча б один розв'язок виду

$$\bar{\mathbf{Q}} \in \left\{ \mathbf{Q} \in \mathcal{Q} : \exists q_k = 1, k \in \{\overline{1, n}\} \right\} \subset \arg \min_{\mathbf{Q} \in \mathcal{Q}} (\mathbf{F}(\Theta) \cdot \mathbf{K} \cdot \mathbf{Q}^T) \subset \mathcal{Q}. \quad (7)$$

Доведення. Під знаком мінімуму у задачі (6) маємо матричний добуток

$$\mathbf{F}(\Theta) \cdot \mathbf{K} \cdot \mathbf{Q}^T = \mathbf{H} \cdot \mathbf{Q}^T, \quad (8)$$

де

$$\mathbf{H} = [h_1 \quad h_2 \quad \cdots \quad h_{n-1} \quad h_n] \quad (9)$$

є числовим вектором при

$$h_j = \sum_{k=1}^n f(\theta_k) a_{kj} \quad \forall j = \overline{1, n}. \quad (10)$$

Тоді добуток (8) представимо як

$$\mathbf{F}(\Theta) \cdot \mathbf{K} \cdot \mathbf{Q}^T = \mathbf{H} \cdot \mathbf{Q}^T = \sum_{j=1}^n h_j q_j = \sum_{j=1}^n \left(\sum_{k=1}^n f(\theta_k) a_{kj} \right) q_j. \quad (11)$$

Оскільки стратегія $\mathbf{F}(\Theta)$ є цілком змішаною і матриця (4) є невід'ємною,

то вектор (9) є додатним. Тому задача (6) зводиться до мінімізації лінійної комбінації (11) n змінних $\{q_j\}_{j=1}^n$ з додатними коефіцієнтами $\{h_j\}_{j=1}^n$.

Нехай

$$h_l \in \min \left(\{h_j\}_{j=1}^n \right). \quad (12)$$

Оскільки кожна змінна $q_j \in [0; 1]$ при $\sum_{k=1}^n q_k = 1$, то

$$\begin{aligned} \sum_{j=1}^n h_j q_j &= \sum_{j=1}^{l-1} h_j q_j + h_l q_l + \sum_{j=l+1}^n h_j q_j = \sum_{\substack{j=1 \\ j \neq l}}^n h_j q_j + h_l q_l = \\ &= \sum_{\substack{j=1 \\ j \neq l}}^n h_j q_j + h_l \left(1 - \sum_{\substack{j=1 \\ j \neq l}}^n q_j \right) = \sum_{\substack{j=1 \\ j \neq l}}^n (h_j - h_l) q_j + h_l. \end{aligned} \quad (13)$$

Згідно з (12) у виразі (13) кожен коефіцієнт $(h_j - h_l)$ перед q_j є невід'ємним, тому мінімізувати лінійну функцію (13) на множині (5) можна при покладанні $q_j = 0 \quad \forall j = \overline{1, n}$ без $j = l$. Відповідно тоді $q_l = 1$, і знайдені значення $\{q_j\}_{j=1}^n$ утворюватимуть вектор $\tilde{\mathbf{Q}}$ у (7), що і треба було довести.

Зауважимо, що при декількох мінімальних елементах вектора (9)

$$\min \left(\{h_j\}_{j=1}^n \right) = \{h_r\}_{r=1}^m \quad (14)$$

розв'язком задачі (6) буде вектор $\tilde{\mathbf{Q}}$ такий, що

$$\tilde{\mathbf{Q}} \in \left\{ \mathbf{Q} \in \mathcal{Q} : q_k = 0 \quad \forall k \in \{ \overline{1, n} \} \setminus \{l_r\}_{r=1}^m \right\} \subset \arg \min_{\mathbf{Q} \in \mathcal{Q}} \left(\mathbf{F}(\Theta) \cdot \mathbf{K} \cdot \mathbf{Q}^T \right) \subset \mathcal{Q}. \quad (15)$$

Але із усіх векторів виду (15) легко буде відібрати і ті, що відповідатимуть чистим стратегіям другого гравця. І якщо в одній з тих чистих стратегій буде виконана умова

$$q_{l_r} = 1 \quad \text{при} \quad f(\theta_{l_r}) \in \max \left\{ f(\theta_j) \right\}_{j=1}^n, \quad (16)$$

то нам логічно можна буде використати значення θ_{l_r} випадкової величини Θ у вихідній математичній моделі, оскільки воно (експериментально) спостерігалось найбільш часто, і при його обиранні нами очікуються найменші втрати.

Власне, умову (16) назвемо необхідною умовою вірогідності розподілу

статистичних частот (1) випадкової величини Θ для її значень (2), де математичне сподівання $M(\Theta)$ і дисперсія $D(\Theta)$ є невідомими, а функція $S(|x-y|)$ для (3) задана. Якщо ж ні при якому з індексів $\{l_r\}_{r=1}^m$ не вдається виконати умову (16), то розподіл (1) для заданої функції $S(|x-y|)$ слід вважати неприйнятним (або ж неадекватним), і вимагати додаткових вимірювань випадкової величини Θ . Це цілком природно, оскільки найменші втрати при обиранні значення випадкової величини не відповідатимуть найбільш імовірній (за статистичними даними) появи цієї величини.

Програма підтримки прийняття рішення про вірогідність розподілу (1). Перевірку виконання умови (16) зручно виконувати у середовищі MATLAB, у якому реалізуємо програмний модуль **stochparamfreqarr** підтримки прийняття рішення про вірогідність розподілу статистичних частот (1) випадкової величини (стохастичного параметра) Θ для її значень (2), задаючи паралельно функцію S (рис. 1).

Програмний модуль **stochparamfreqarr** має три вхідних параметри: вектор-розподіл (1), вектор відсортованих за зростанням значень (2) і вектор неспадних значень функції S . Можна вказувати тільки два перших вектори, і тоді за умовчанням буде покладено $S=1$. Якщо умова (16) є справедливою, то у середовищі MATLAB повертається вектор $\tilde{\mathbf{Q}}$ як розв'язок задачі (6) та матриця $\mathbf{K}=(a_{ij})_{n \times n}$ як (4). Також користувач бачить вектор (9). За невиконання умови (16) замість вектора $\tilde{\mathbf{Q}}$ повертається повідомлення про неадекватний розподіл (1).

Розглянемо приклад. Нехай є нормований розподіл

$$\mathbf{F}(\Theta)=[0.1 \ 0.1 \ 0.1 \ 0.1 \ 0.4 \ 0.15 \ 0.05] \quad (17)$$

статистичних частот випадкової величини Θ зі значеннями

$$[4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10]. \quad (18)$$

Як видно з рис. 2, для розподілу (17) значень (18) при $S=1$ умова (16) є справедливою.

Але вже для розподілу

$$\mathbf{F}(\Theta)=[0.2 \ 0.2 \ 0.1 \ 0.1 \ 0.3 \ 0.05 \ 0.05] \quad (19)$$

значень (18) буде прийнято рішення про його неприйнятність (рис. 3), оскільки за розв'язком задачі (6) для досягнення найменших втрат треба обирати значення $\theta_3=6$ та $\theta_4=7$, які статистично є найбільш “непопулярними”. Ця алогічність у даному випадку (як і в інших випадках) і є приводом для прийняття рішення про неадекватність розподілу (19).

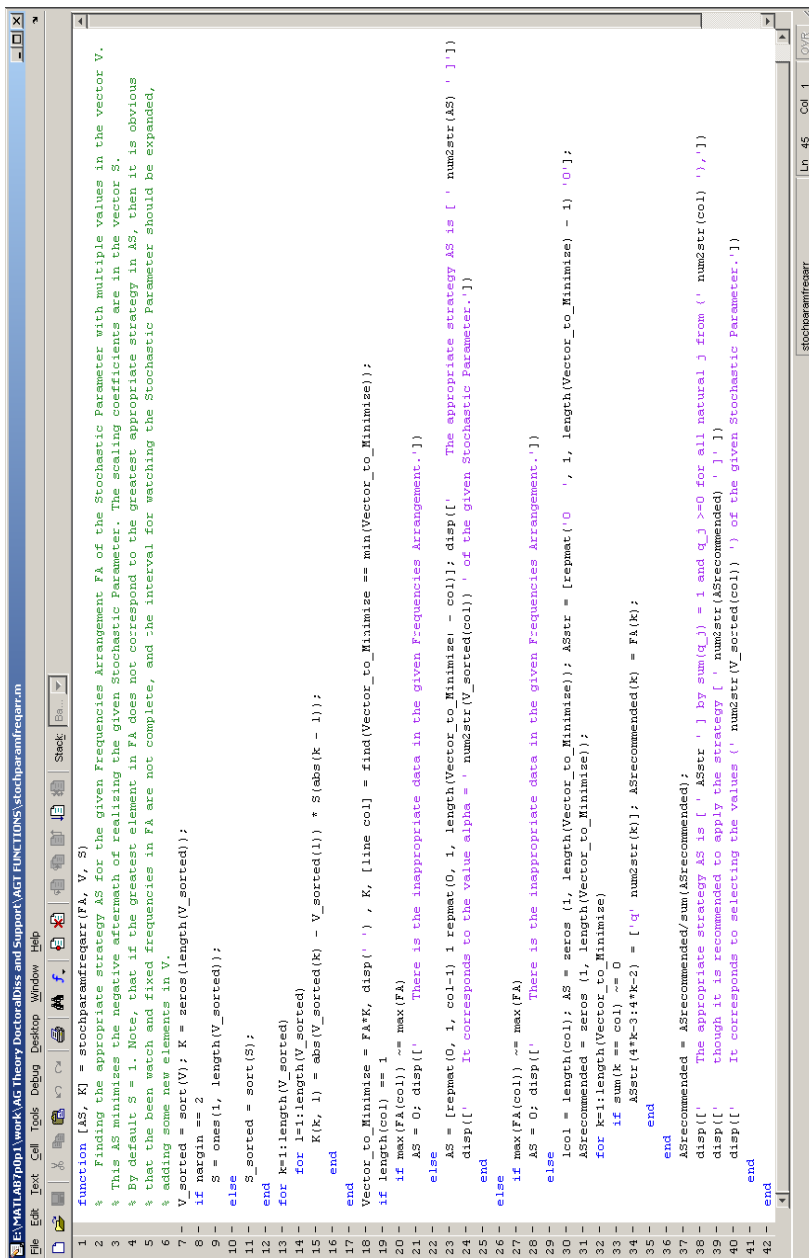


Рис. 1. Повний код програмного модуля **stochparamfreqarr** для підтримки прийняття рішення про вірогідність розподілу (1)

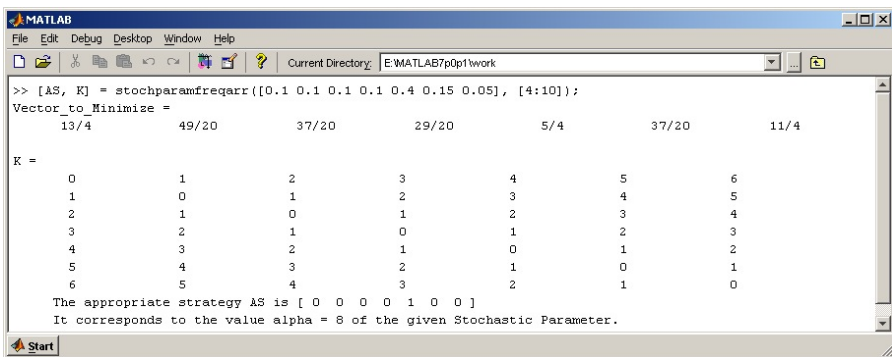


Рис. 2. Запуск модуля **stochparamfreqarr** для прийняття рішення про вірогідність розподілу (17) значень (18)

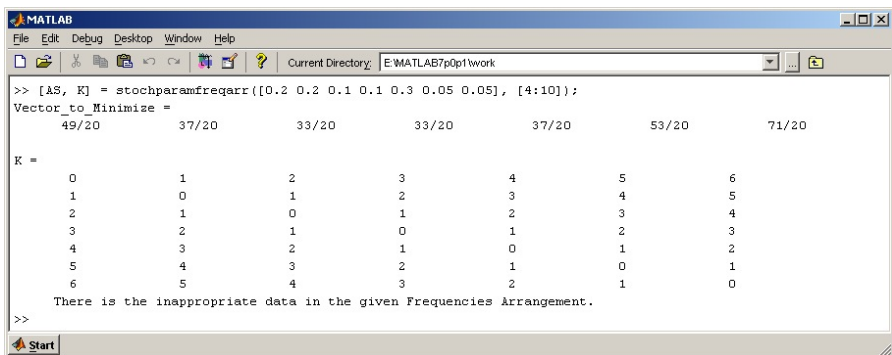


Рис. 3. Прийняття рішення про неадекватність розподілу (19) значень (18)

Висновок та перспектива подальшого дослідження. Звичайно, запропонований метод перевірки достатності проведених вимірювань для того, щоб реалізувати розподіл (1) у досліджуваній математичній моделі, є тільки необхідністю того, щоб для найменших втрат обирались значення випадкової величини Θ з найбільшою відносною частотою. Можливо, у деяких випадках саме це значення і можна використовувати у вихідній математичній моделі, якщо тільки для нього виконано (16). Але можливий випадок, коли розв'язок $\tilde{Q}$ задачі (6) не буде чистою стратегією, як для розподілу (рис. 4)

$$F(\Theta) = [0.1 \ 0.2 \ 0.1 \ 0.1 \ 0.4 \ 0.05 \ 0.05] \quad (20)$$

тих же значень (18), де

$$\tilde{Q} = [0 \ 0 \ 0 \ q_4 \ q_5 \ 0 \ 0] \text{ при } q_4 \in [0; 1], q_5 \in [0; 1], q_4 + q_5 = 1. \quad (21)$$

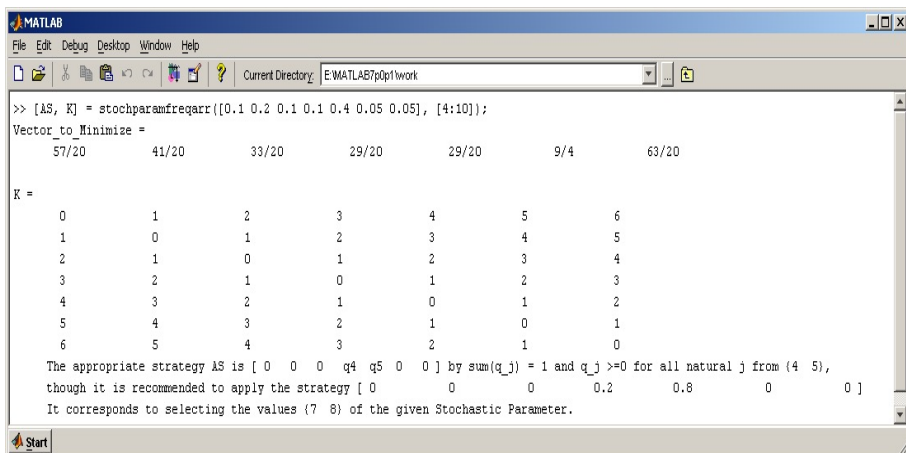


Рис. 4. Прийняття рішення про вірогідність розподілу (20) значень (18) з рекомендацією щодо використання значень $\theta_4 = 7$ та $\theta_5 = 8$ з відповідними імовірностями 0.2 та 0.8

Щоправда, тоді замість (21) рекомендовано використовувати стратегію

$$\tilde{Q} = [0 \ 0 \ 0 \ 0.2 \ 0.8 \ 0 \ 0], \quad (22)$$

співвідношення ненульових елементів котрої відповідає співвідношенню між $f(\theta_4)$ й $f(\theta_5)$ у розподілі (20). Саме для таких розрахунків і створено програмний модуль **stochparamfreqarr** підтримки прийняття рішення про вірогідність розподілу статистичних частот (1) випадкової величини Θ для її значень (2) при заданій функції S , де перевіряється необхідна умова вірогідності (16). А з точки зору подальшого дослідження у цьому напрямку перспективною слід вважати задачу сполучення запропонованої моделі прийняття рішення про вірогідність розподілу (1) із критеріями перевірки статистичних гіпотез про тип розподілу генеральної сукупності та його параметри.

Список літератури: 1. Гмурман В.Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В.Е. Гмурман. – М.: Высшая школа, 1999. – 479 с. 2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для студентов вузов / В.Е. Гмурман. – М.: Высшая школа, 1998. – 400 с. 3. Большаков А.А. Методы обработки многомерных данных и временных рядов: учебное пособие для вузов / А.А. Большаков, Р.Н. Каримов. – М.: Горячая линия – Телеком, 2007. – 520 с. 4. Бочаров П.П. Теория вероятностей. Математическая статистика: учебное пособие / П.П. Бочаров, А.В. Печинкин. – М.: Гардарики, 1998. – 328 с. 5. Вентцель Е.С. Теория вероятностей и её инженерные приложения: учеб. пособие для вузов / Е.С. Вентцель, Л.А. Овчаров. – М.: Высшая школа, 2000. – 480 с. 6. Петросян Л.А. Теория игр: учеб. пособие для ун-тов / Л.А. Петросян, Н.А. Зенкевич, Е.А. Семина. – М.: Высшая школа, Книжный дом "Университет", 1998. – 304 с. 7. Васин А.А. Введение в теорию игр с

приложениями к экономике: учебное пособие / А.А. Васин, В.В. Морозов. – М.: Высшая школа, 2003. – 278 с. 8. Романюк В.В. Модель визначення оптимального рішення проектувальника у задачі про розрахунок поведомжної стійкості двох елементів будівельної конструкції при дії на них нормованого стискаючого зусилля / В.В. Романюк // Проблеми трибології. – 2010. – № 1. – С. 42 – 56.

Стаття представлена д.т.н. проф. Національної академії державної прикордонної служби України імені Б. Хмельницького Катеринчуком І. С.

УДК 519.832.3

Оценивание достоверности распределения статистических частот случайной величины с неизвестным математическим ожиданием и дисперсией / Романюк В. В. // Вестник НТУ "ХПИ". Тематический выпуск: Информатика и моделирование. – Харьков: НТУ "ХПИ". – 2010. – № 21. – С. 152 – 161.

Построено ядро антагонистической игры для задачи безусловной оптимизации, по решению которой предлагается принимать решение о достаточности проведённых измерений случайной величины для того, чтобы её реализовывать в соответствующей математической модели в форме распределения относительных статистических частот. Представлено программу поддержки принятия решения о надёжности исследуемого распределения. Ил.: 4. Библиогр.: 8 назв.

Ключевые слова: антагонистическая игра, случайная величина, распределение статистических частот, принятие решения.

UDC 519.832.3

Evaluating validity of the statistic frequencies distribution of a variate with undefined mathematical expectation and variance / Romanuke V. V. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkiv: NTU "KhPI". – 2010. – № 21. – P. 152 – 161.

There has been constructed the kernel of an antagonistic game for an unconstrained optimization problem, by the solution of which it is being proposed to make the decision on the sufficiency of the carried measurements of a variate for realizing it within the corresponding mathematical model in the form of the distribution of relative statistical frequencies. It has been represented a decision making support program on the reliability of the being investigated distribution. Figs: 4. Refs: 8 titles.

Key words: antagonistic game, random variate, statistic frequencies distribution, decision making.

Поступила в редакцию 18.03.2010