УДК 519.866

Д.С. РЕВЕНКО, В.М. ВАРТАНЯН, Ю.А. РОМАНЕНКОВ

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ»

СТАТИСТИЧЕСКАЯ ОЦЕНКА ДИНАМИЧЕСКИХ ПРОЦЕССОВ С НЕОПРЕДЕЛЕННЫМИ ДАННЫМИ

В статье рассмотрены проблемы статистической оценки динамических интервальных данных при исследовании экономических процессов с неопределенностью. Базовыми количественными характеристиками были выбраны: ширина, радиус и длина интервала, а оценочными: корреляция интервальных границ, внутриинтервальная дисперсия, средневзвешенное интервального ряда и коэффициент вариации. В качестве решения проблемы описания внутриинтервальной неопределенности предлагается оценивать экспертным путем распределение вероятностей. Предложено использовать равномерное, треугольное, трапециевидное и равномерное распределения.

Ключевые слова: *интервальный анализ, внутриинтервальная вероятность, прогнозирование, статистика интервальных данных.*

Экономические явления характеризуются постоянной изменчивостью, и в будущем эти изменения будут нарастать. В настоящее время формализация этих явлений основана преимущественно на использовании точных данных. Однако, как отмечалось, методы, основанные на определенности, оказываются неэффективными, потому что действительность не соответствует разработанной схеме. Применение вероятностных методов также не совсем эффективно в случаях, когда используются субъективные данные, а также вероятность не всегда бывает точно достоверно известна [1]. Неадекватность получаемых решений приводит нас к разработке новых методов основанных на параметрическом синтезе необходимых описательных моделей и операций интервального анализа. Эти модели способны дать адекватный подход к малоизученным явлениям и процессам в деятельности предприятия, которые являются неопределенными.

Типичным примером интервальной неопределенности является следующий: результаты наблюдений показывают, что величина x лежит в некотором интервале $[a, b]$. В общем случае о неопределенной характеристике x известно только ограниченное множество значений. При этом неопределенная величина x описывается как нечеткая, у которой функция принадлежности равна 1 на всем отрезке $[a, b]$ и нулю вне его [2].

Главным преимуществом интервального анализа является его способность частично раскрывать знания о неопределенных процессах, т.е. термин «неопределенность» выражает отрицание «определенности», но в большинстве случаев означает незнание (когда любое исследование до чрезмерности затруднилось бы, либо стало совсем невозможным), и состояние частичного знания, когда мы все-таки располагаем какой-то информацией об интересующей нас величине. Необходимо отметить, что интервальная математика не совершенна и в перспективе представляется интерес развития направления в изучении динамических процессов с интервальной неопределенностью.

Целью данной статьи является разработка оценочных статистических характеристик динамических интервальных процессов с целью более глубокого изучения этих процессов.

Прежде чем переходить к разработке указанных выше оценочных характеристик, необходимо обратиться к основной теореме интервального анализа, которая открывает возможность перед исследователем использовать интервальные операции для расширения существующих моделей и оценочных характеристик.

Основная теорема интервального анализа. Пусть $f(x_1, x_2, ..., x_n)$ – рациональная функция вещественных аргументов $x_1, x_2, ..., x_n$ и для нее определен результат $f(x_1, x_2, ..., x_n)$ подстановки вместо аргументов интервалов их измерений $[x_1], [x_2], ..., [x_n] \in IR$ и выполнения всех действий над ними по правилам интервальной арифметики. Тогда

$$\left\{ f(x_1, x_2, ..., x_n) \middle| \, x_1 \in [x_1], ..., x_n \in [x_n] \right\} \subseteq [f]\left([x_1], [x_2], ..., [x_n]\right), \qquad (1)$$

т.е. $[f]\left([x_1], [x_2], ..., [x_n]\right)$ содержит множество значений функции $f(x_1, x_2, ..., x_n)$ на $\left([x_1], [x_2], ..., [x_n]\right)$.

Если выражение для $f(x_1, x_2, ..., x_n)$ содержит не более чем по одному вхождению каждой переменной в первой степени, то вместо включения имеет место точное равенство [3].

Ниже приведены разработанные статистические оценки интервальных динамических данных, которые основываются на базовых понятиях и определениях статистических оценок с «наложением» интервальных операций.

Базовые элементы интервальных статистических характеристик:

Симметричным называется интервал $[x]$, у которого $-\underline{x} = \overline{x}$, где $\underline{x}$ – нижняя граница интервала, а $\overline{x}$ – верхняя граница интервала $[x]$.

Шириной интервала $[x]$ называется величина

$$\mathrm{wid}[x] = \overline{x} - \underline{x}. \tag{2}$$

Радиус интервала $[x]$ называется величина

$$\mathrm{rad}[x] = \frac{\overline{x} - \underline{x}}{2}. \tag{3}$$

Середина интервала $[x]$ есть полусумма его концов

$$\mathrm{mid}[x] = \frac{\underline{x} + \overline{x}}{2}. \tag{4}$$

В интервальной арифметике вещественные числа $x \in R$ отождествляются с интервалами нулевой ширины (вырожденными интервалами) $[x] = [x, x]$.

Взаимосвязь динамики границ интервалов. В качестве определения наличия тесноты связи между верхними и нижними границами интервалов целесообразно использовать коэффициент корреляции. Коэффициент корреляции представляет собой среднее величину из произведений нормированных отклонений для нижних $\underline{x}$ и верхних $\overline{x}$ границ

$$R_{int} = \frac{\sum_{i=1}^{n} (\underline{x_i} - \underline{x_{cp}}) \cdot (\overline{x_i} - \overline{x_{cp}})}{\sqrt{\sum_{i=1}^{n} (\underline{x_i} - \underline{x_{cp}})^2 \cdot \sum_{i=1}^{n} (\overline{x_i} - \overline{x_{cp}})^2}}, \tag{5}$$

где $\underline{x_i}$ – нижняя граница интервала i-го наблюдения;

$\overline{x_i}$ – верхняя граница интервала i -го наблюдения;

n – количество наблюдений в исследуемом временном ряду;

$\underline{x_{cp}}$, $\overline{x_{cp}}$ – средние величины нижней и верхней интервальных границ, которые равны соответственно:

$$\underline{x_{cp}} = \frac{\sum_{i=1}^{n} \underline{x_i}}{n}; \tag{6}$$

$$\overline{x_{cp}} = \frac{\sum_{i=1}^{n} \overline{x_i}}{n}. \tag{7}$$

Коэффициент корреляции может принимать значение от –1 до +1, причем знак определяется в ходе решения.

Если коэффициент корреляции $-1 \le R_{int} \le 0$, это обозначает, что окна интервалов (расстояние между границами интервалов) сужаются или расширяются. Расширение границ интервалов в динамике говорит о возрастающей неопределенности исследуемого явления во времени. А сужение границ говорит о противоположном, т.е. явление «теряет неопределенность».

Противоположный случай: если $0 \le R_{int} \le 1$, это обозначает, что границы интервалов имеют одинаковую тенденцию во времени. Чем ближе R_{int} к 1, тем ближе функциональная зависимость между границами интервалов, что говорит о стационарности и неопределенности процесса. Если R_{int} ближе к 0, это обозначает отсутствие какой либо связи между границами интервалов.

Иногда для качественного анализа тесноты связи можно воспользоваться шкалой Чеддока (табл. 1) [4].

Внутриинтервальная дисперсия. Показывает, насколько разбросаны границы интервалов от среднего каждого интервала:

$$D_{int} = \frac{\sum_{i=1}^{n} \left(\underline{x_i} - \mathrm{mid}[x_i]\right)^2 + \sum_{i=1}^{n} \left(\overline{x_i} - \mathrm{mid}[x_i]\right)^2}{2n}, \tag{8}$$

где $\mathrm{mid}[x_i]$ – середина интервала (среднее интервала), определяется как полусумма концов интервала $[x_i]$ (соотношение 4).

Таблица 1

Шкала Чеддока для оценки качества тесноты связи между признаками

Количественная мера тесноты связи	*Качественная характеристика силы связи*
0,1-0,3	Слабая
0,3-0,5	Умеренная
0,5-0,7	Заметная
0,7-0,9	Высокая
0,9-0,99	Весьма высокая

Средневзвешенная средняя интервального ряда показывает среднее значение интервального ряда с учетом влияния ширины каждого интервала:

$$WA = \sum_{i=1}^{n} \frac{mid[x_i] \cdot win[x_i]}{\sum_{i=1}^{n} win[x_i]}, \quad (9)$$

где $win[x_i]$ – ширина интервала (окно интервала) определяется из соотношения 2.

Средняя интервального ряда. Может использоваться наряду со средневзвешенной средней интервального ряда:

$$Av = \frac{\sum_{i=1}^{n} mid[x_i]}{n}. \quad (10)$$

Коэффициент вариации интервального ряда. По величине коэффициента вариации можно судить о степени разброса интервальных границ вокруг среднего ряда (стационарности интервального динамического ряда):

$$Var_{int} = \frac{\sqrt{D_{int}}}{Av}; \quad (11)$$

$$Var_{int} = \frac{\sqrt{D_{int}}}{WA}. \quad (12)$$

Необходимо обратить внимание, что на данный момент является актуальным решение проблемы неопределенности внутри самого оценочного интервала. Классическая интервальная математика, основывается на том выводе, что мы определяем только границы неопределенности, оставляя саму неопределенность внутри интервала, т.е. прогнозируемые показатели внутри интервала имеют одинаковую вероятность реализации, что равносильно отсутствию каких либо знаний о распределении вероятности. Этот подход не всегда применим к случаям, когда заказчик исследования обладает ограниченным количеством ресурсов и они очень дорогостоящи.

Для дальнейшего решения проблемы раскрытия неопределенности внутри самого интервала было сформулировано допущение: использовать вероятностное распределение внутри интервалов (предложено профессором Севастьяновым П.В. [5]). Он предложил, что вероятностные распределения внутри интервалов в хорошо изученных явлениях известны, и что они не изменяются со временем.

Исходя из заключения Севастьянова, следует, что вероятностное распределение в краткосрочных временных рядах не изменяется (или изменяется, но незначительно). Следовательно, при анализе интервалов мы можем «накладывать» распределения вероятностей (которое может быть получено экспертным путем) на интервал и использовать его для формирования заключения. Еще одним необходимым элементом метода является задание доверительной вероятности (β), которая характеризует уровень вероятности реализации прогноза. Большинство исследователей использует значение доверительной вероятности на уровне 0,95. Следует понимать, что после установления интервального прогноза, вида распределения и выбора доверительной вероятности на уровне 0,95 расчет реально вероятностного диапазона становиться чисто технической задачей. Решение этой задачи заключается в отсечении «лишних» концов реально возможного диапазона соответственно принятой доверительной вероятностью, иначе говоря, нахождением величин:

$$\underline{X} = \underline{x} + \phi, \ \overline{X} = \overline{x} - \phi, \tag{13}$$

где ϕ – величина, зависящая от вида распределения и вероятности неудачи.

Кроме этого необходимо найти и основные численные параметры тех или иных распределений вероятностей.

Теперь рассмотрим каждое из четырех распределений.

Так как мы рассматриваем упрощенный пример распределений вероятностей внутри интервала (т.е. распределения являются симметричными), следовательно, среднее значение $\mathrm{mid}[x]$ рассчитывается как полусумма концов интервала и является одним и тем же, что медиана и мода (соотношение 2).

А. *Нормальное распределение (Закон Гаусса)*. Нормальное распределение является наиболее распространенным распределением в экономике [5]. Известно, что площадь под кривой нормального распределения в пределах $M \pm 3\delta$ ($M = \mathrm{mid}\left[x_R\right]$ – математическое ожидание, δ – стандартное отклонение) примерно равна 99% общей площади, т.е. $\mathrm{win}\left[x_N\right] = \overline{x} - \underline{x} \approx 6\delta$. Пусть z – нормальное отклонение от средней, зависящее от выбранной доверительной вероятности. Тогда нормированное значение искомой величины ϕ составляет

$$\phi_N = u \cdot \delta\partial, \tag{14}$$

где u определяется из соотношения

$$u = 3 - z, \tag{15}$$

z при заданном уровне вероятности равном 0,95 будет составлять 1,6449.

Стандартное отклонение при нормальном распределении составляет

$$\delta = \frac{\mathrm{win}\left[x_N\right]}{6}. \tag{16}$$

Также следует не забывать основное правило распределения вероятности в нормальном распределении (правило трех сигм):

– при $M \pm \delta$, получим вероятность 0,68268;

– при $M \pm 2\delta$, получим вероятность 0,9545;

– при $M \pm 3\delta$, получим вероятность 0,9973 (рис. 1).

Б. *Треугольное распределение (распределение Симпсона)*. Треугольное распределение (рассматриваются случаи только равнобедренного треугольного распределения) чаще всего используется при финансовом моделировании и для моделирования процессов, где отношения между переменными известны, но данных не достаточно (возможно из-за их большой стоимости этих данных).

Для треугольного распределения величина ϕ определяется как

$$\phi_T = \mathrm{win}\left[x_T\right]\sqrt{\frac{\beta}{2}}. \tag{17}$$

Функция плотности вероятности треугольного распределения определяется как

$$f_T = \begin{cases} \dfrac{2(x_T - \underline{x}_T)}{\mathrm{win}[x_T]\cdot(\mathrm{mid}[x_T] - \underline{x}_T)}, & \text{если } \underline{x}_T \le x_T \le \mathrm{mid}[x_T]; \\ \dfrac{2(\overline{x}_T - x_T)}{\mathrm{win}[x_T]\cdot(\overline{x}_T - \mathrm{mid}[x_T])}, & \text{если } \mathrm{mid}[x_T] < x_T \le \overline{x}_T. \end{cases} \quad (18)$$

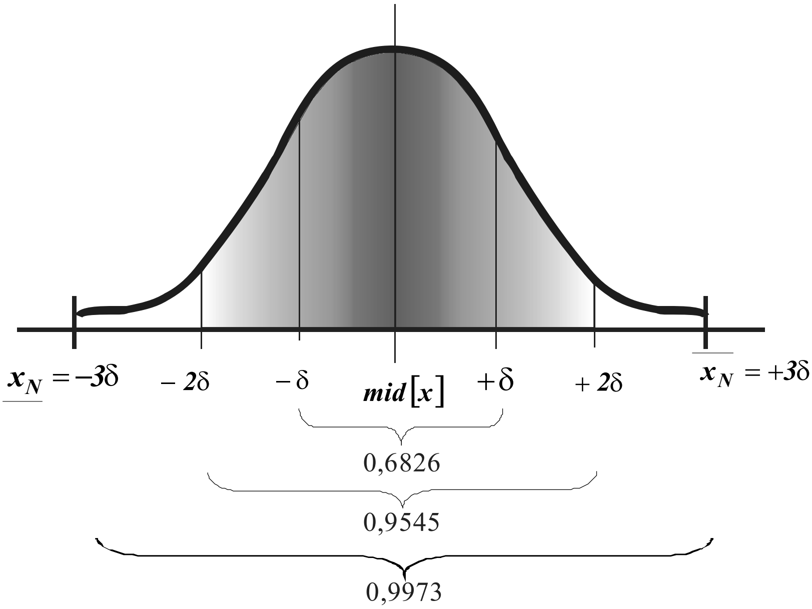

Рис. 1. Нормальное распределение и правило «трех сигм»

Функцию распределения в интервале $[x_T]$ можно определить для любого $\underline{x}_T \le x_T \le \overline{x}_T$ (рис. 2)

$$F_T = \begin{cases} \dfrac{(x_T - \underline{x}_T)^2}{\mathrm{win}[x_T]\cdot(\mathrm{mid}[x_T] - \underline{x}_T)}, & \text{если } \underline{x}_T \le x_T \le \mathrm{mid}[x_T]; \\ 1 - \dfrac{(\overline{x}_T - x_T)^2}{\mathrm{win}[x_T]\cdot(\overline{x}_T - \mathrm{mid}[x_T])}, & \text{если } \mathrm{mid}[x_T] < x_T \le \overline{x}_T. \end{cases} \quad (19)$$

Дисперсия треугольного распределения равна

$$D_T = \frac{(\overline{x}_T - \underline{x}_T)^2}{24}. \quad (20)$$

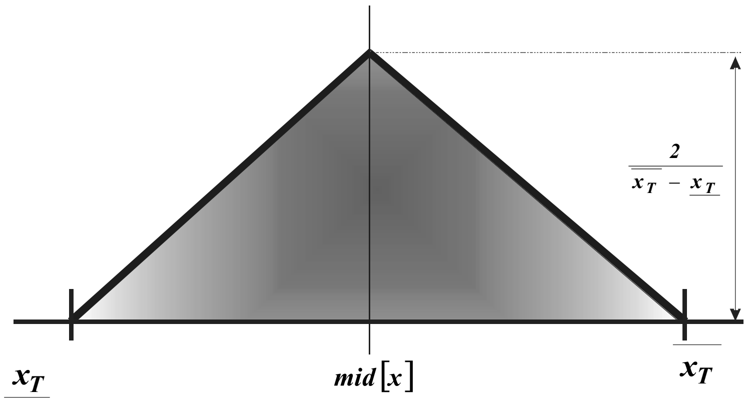

Рис. 2. Треугольное распределение (распределение Симпсона)

В. *Трапециевидное распределение.* Трапециевидное распределение, в основном, используется в теории размытых множеств. Возможны два варианта для определения ϕ:

$$\text{если } \beta \le \frac{1}{2} \cdot \frac{\mathrm{win}[x_{Tr}]\text{-}l}{\mathrm{win}[x_{Tr}]+1}, \text{ то } \phi_{Tr} = \sqrt{\frac{\beta}{2}(\mathrm{win}[x_{Tr}]^2 - l^2)}\,; \qquad (21)$$

$$\text{если } \beta \ge \frac{1}{2} \cdot \frac{\mathrm{win}[x_{Tr}]\text{-}l}{\mathrm{win}[x_{Tr}]+1}, \text{ то } \phi_{Tr} = \frac{\mathrm{win}[x_{Tr}]+1}{2} \cdot \beta + \frac{\mathrm{win}[x_{Tr}]-1}{4}, \qquad (22)$$

где $l = M_2 - M_1$;

M_1, M_2 – наиболее вероятный диапазон трапециевидного распределения (рис. 3).

Функция плотности вероятности трапециевидного распределения определяется как:

$$f_{Tr} = \begin{cases} k\left(\dfrac{x_{Tr} - \underline{x_{Tr}}}{M_1 - \underline{x_{Tr}}}\right), & \text{при } \underline{x_{Tr}} \le x_{Tr} < M_1\,; \\ k, & \text{при } M_1 \le x_{Tr} < M_2\,; \\ k\left(\dfrac{\overline{x_{Tr}} - x_{Tr}}{\overline{x_{Tr}} - M_2}\right), & \text{при } M_2 \le x_{Tr} \le \overline{x_{Tr}}. \end{cases}, \qquad (23)$$

где $k = 2 \cdot (\overline{x_{Tr}} + M_2 - M_1 - \underline{x_{Tr}})^{-1}$, при $x_{Tr} \le M_1 \le M_2 \le x_{Tr}$.

Дисперсия трапециевидного распределения равна

$$D_{Tr} = \frac{win\left[x_{Tr}\right]^2 - 1^2}{24}. \tag{24}$$

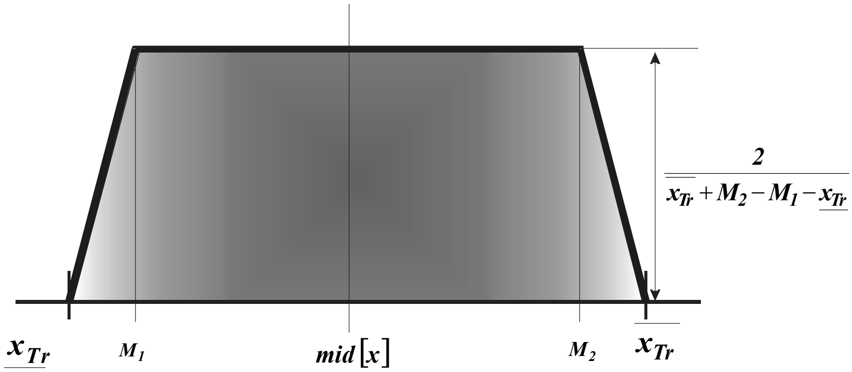

Рис. 3. Трапециевидное распределение

Г. *Равномерное (прямоугольное) распределение*. Величина имеет равномерное распределение на интервале $\left[x_R\right]$, если на этом интервале плотность распределения случайной постоянна, а вне его равна 0.

Для равномерного распределения величина ϕ рассчитывается как

$$\phi = \beta \cdot win\left[x_R\right]. \tag{25}$$

Функцию плотности вероятности равномерного распределения определяется как

$$f = \frac{1}{win\left[x_R\right]}. \tag{26}$$

Это выражение обозначает, что вероятность любого интервала зависит только от его длины. Но если мы имеем дело с дискретными величинами, тогда знаменатель выражения заменяется на конечное число значений этих дискретных величин.

Функцию распределения в интервале $\left[x_R\right]$ можно определить для любого $\underline{x_R} \le x_R \le \overline{x_R}$:

$$F_R = \frac{x_R - \underline{x_R}}{\overline{x_R} - \underline{x_R}} = \frac{x_R - \underline{x_R}}{win\left[x_R\right]}. \tag{27}$$

Дисперсия равномерного распределения равна

$$D_R = \frac{(\overline{x_R} - \underline{x_R})^2}{12}. \quad (28)$$

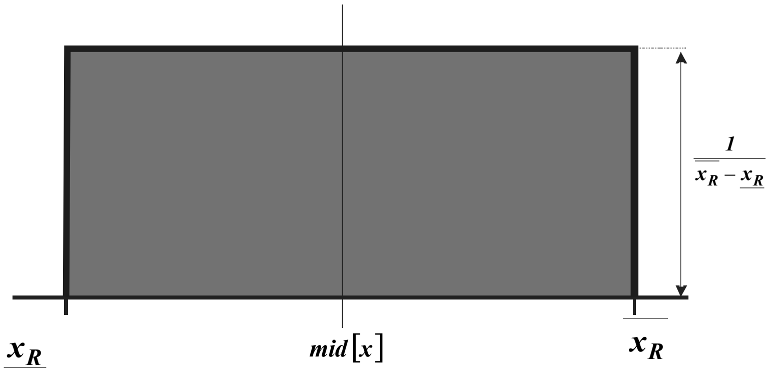

Рис. 4. Равномерное распределение

В заключении следует отметить, что использование оценочных и вероятностных характеристик для изучения динамических процессов с интервальной неопределенностью поможет расширить решение задачи неопределенности в динамике и внутри самих интервалов.

Литература

1. Лафуенте Хил А.М. Финансовый анализ в условиях неопределенности / А.М. Хил Лафуенте. – Минск: Технология, 1998. – 150 с.

2. Клейнер Г.Б. Эконометрические зависимости: принцип и методы построения / Г.Б. Клейнер, С.А. Смоляк. – М.: Наука, 2000. – 104 с.

3. Шарый С.П. Конечномерный интервальный анализ [Электронный ресурс] / С.П. Шарый. – М.: XYZ, 2009. – 570 с. Режим доступа к книге: http://www.sbras.ru/interval/Library/InteBooks/SharyBook.pdf.

4. Общая теория статистики: статистическая методология в изучении коммерческой деятельности: учебник / А.И. Харламов, О. Э. Башина, В.Т. Бабурин, И.А. Ионсен, Т.П. Пройдакова и др. – М.: Финансы и статистика, 1996. – 296 с.

5. Севастьянов П.В. Методика нечетко-интервального имитационного моделирования технико-экономических систем [Электронный ресурс] / П.В. Севастьянов, В.И. Вальковский // Электронное научное издание «Наука и образование: электронное научно-техническое издание». – М., 2006. – № 4. – Режим доступа к изданию: http://www.techno.edu.ru:16001/db/msg/28686.html.

Рецензент: д-р экон. наук, проф. **Н.А. Киним**, директор НИЦ ИПР НАН Украины.

СТАТИСТИЧНА ОЦІНКА ДИНАМІЧНИХ ПРОЦЕСІВ З НЕВИЗНАЧЕНИМИ ДАНИМИ

Д.С. Ревенко, В.М. Вартанян, Ю.О. Романенков

У статті розглянуті проблеми статистичної оцінки динамічних інтервальних даних при дослідженні економічних процесів з невизначеністю. Як вирішення проблеми опису внутрішньоінтервальної невизначеності пропонується оцінювати експертним шляхом розподілу вірогідності. Запропоновано використовувати рівномірний, трикутний, трапецієвидний і рівномірний розподіли.

Ключові слова: інтервальний аналіз, внутрішньоінтервальна вірогідність, прогнозування, статистика інтервальних даних.

STATISTICAL ESTIMATION OF DYNAMIC PROCESSES WITH INDEFINITE INFORMATION

D.S. Revenko, V.M. Vartanyan, Y.A. Romanenkov

In article problems of a statistical estimation of dynamic interval data are considered at research of economic processes with uncertainty. As the decision of a problem of the description inwardly interval uncertainty is offered to estimate expert by distribution of probabilities. It is offered to use uniform, triangular, trapezoid and even distributions.

Key words: the interval analysis, inwardly interval probability, forecasting, statistics of interval data, multiinterval.

Ревенко Даниил Сергеевич – аспират кафедры экономики и маркетинга Национального аэрокосмического университета им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: revenko_dan@ukr.net.

Вартанян Василий Михайлович – д-р техн. наук, проф., заведующий кафедрой экономики и маркетинга Национального аэрокосмического университета им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: vartanyan_vm@ukr.net.

Романенков Юрий Александрович – канд. техн. наук., доцент кафедры менеджмента Национального аэрокосмического университета им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: k602@d6.khai.edu.