Исследование вариаций параметров на основе значений энтропийного коэффициента

В. Л. Лазарев
Университет ИТМО
holod25@yandex.ru

УДК 681.5

Аннотация. В работе предложены решения по исследованию вариаций различных параметров на основе анализа значений соответствующих энтропийных коэффициентов. Предложена обобщенная графическая зависимость для описания взаимосвязи между величиной размаха параметра в выборке наблюдений и соответствующим значением энтропийного коэффициента. Даны рекомендации по ее практическому использованию. Предложенные решения могут применяться для решения задач интеллектуального анализа и, в частности, в условиях априорной неопределенности, обусловленных «дефицитом» результатов наблюдений параметров.

Ключевые слова: вариации; состояния неопределенности; энтропийные потенциалы; энтропийный коэффициент

Введение. Вариация какого-либо параметра проявляется разбросом его значений, характеризует состояние неопределенности этого параметра в рассматриваемом процессе или объекте. Исследование вариаций являются одной из основ для «углубления» знаний о природе и сути протекающих процессов. Для изучения этого явления могут использоваться различные подходы, теории и методы. Таковыми, например, являются разработки для проведения статистических исследований, мягких вычислений и измерений, технологий интеллектуального анализа и др. [1–3]. Каждый из подходов имеет свои достоинства и недостатки, предпочтительные области применения, адаптирован к специфике конкретных задач. Одним из перспективных, для исследования вариативных свойств параметров и, особенно в условиях неопределенности, является подход, основанный на использовании разработок теории энтропийных потенциалов (ТЭП) [4, 5]. Возможности методов ТЭП позволяют проводить исследования при наличии ограниченных объемов результатов наблюдений, в условиях ограниченной выборки реализаций параметра.

Отправные положения. Применительно к поставленной проблеме исследование вариаций осуществляется на основе «энтропийного» подхода. При этом для оценки вариативных свойств параметра х используется не сама величина энтропии Нх, а косвенно с ней связанный набор понятий величин энтропийных потенциалов. Эти величины взаимосвязаны соотношением [1, 4, 6]

$$L_{\Delta} = \frac{\Delta_e}{|X_n|} = \frac{K_e \sigma}{|X_n|} . \quad (1)$$

В выражении (1) использованы следующие величины и обозначения. L_{Δ} – величина комплексного энтропийного потенциала, $\Delta_e = K_e\sigma$ – величина энтропийного потенциала, σ – величина среднеквадратического отклонения (СКО), K_e – энтропийный коэффициент, характеризующий вариативные свойства закона распределения результатов наблюдений, X_n – величина нормирующего значения, на базе которого рассматривается состояние неопределенности параметра. Свойства и достоинства величины L_{Δ}, а также целесообразность ее использования в ряде практических задач вместо величины энтропии H_x рассмотрены в [4, 7]. В ней комплексно «интегрированы» свойства и возможности основных «классических» характеристик, используемых для описания вариаций: СКО, базового значения (в частности, например, математического ожидания $X_n = m_x$), коэффициента вариации $v = \frac{\sigma}{m_x}$, (получается из выражения (1) положением $K_e = const$). Следует отметить, что помимо указанных характеристик, для описания вариаций также используются и другие характеристики: размах, линейное отклонение, дисперсия. Использование величины K_e в составе «комплекса» (1) позволяет осуществить учет других вариативных свойств параметра. Рассмотрим этот вопрос подробнее.

Оценки вариаций на основе значений энтропийного коэффициента. По сути дела, энтропийный коэффициент является сомножителем к величине СКО – σ_x, устанавливающим ее взаимосвязь с величиной энтропийного потенциала. Последняя определяется как половина диапазона – *а* равномерного закона распределения, имеющего такую же энтропию – H_x, что и закон распределения рассматриваемого параметра - *x*. Соответствующая зависимость, в общем виде, описывается выражением [4, 6, 8]

$$K_e = \frac{\Delta_e}{\sigma_x} = \frac{0.5e^{H_x}}{\sigma_x} = \frac{0.5e^{-\int_{-\infty}^{\infty} p(x)\ln p(x)dx}}{\sigma_x} . \quad (2)$$

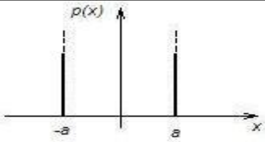
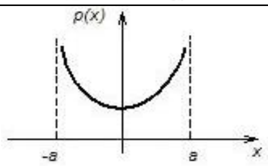
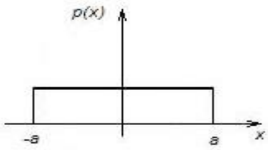
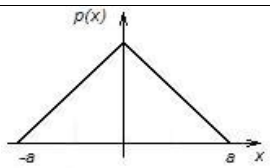
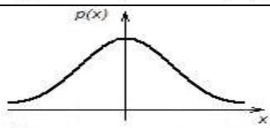
Для реальных законов распределения диапазон изменения величины K_e описывается неравенством [4, 6]

$$1 \le K_e \le 2.07\,. \qquad (3)$$

Для ряда законов, когда имеется аналитическое описание функции плотности распределения – *p(x)*, значение величины K_e может быть определено из выражения (2). (В табл. 1 приведены значения энтропийных коэффициентов для пяти распространенных типовых законов.)

В противном случае, например, когда закон распределения является ассиметричным, многомодальным и пр., определение значения K_e также может быть осуществлено с использованием выражения (2) приближенными численными методами на основании результатов наблюдений (т.н. метод прямого оценивания). Такая ситуация является характерной для большинства практических задач. При этом требуется наличие значительного объема выборки $n \ge 40$, необходимого для построения гистограммы статистического распределения и последующего нахождения интеграла в выражении (2) численными методами.

ТАБЛИЦА I СВОЙСТВА ТИПОВЫХ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Тип закона распределения вероятностей	**Плотность вероятности *p*(*x*)**	**Вид кривой плотности вероятности**	**Величина *a***	**Энтропийный коэффициент K_e**
1	2	3	4	5
1. Дискретное двухзначное распределение	$p(x)=\begin{cases}0.5\ при\ \lvert x\rvert = a;\\ 0\ при\ \lvert x\rvert \ne a\end{cases}$		σ	1.00
2. Арксинусоидальный закон	$p(x) = \dfrac{1}{\pi a\sqrt{1-\left(\dfrac{x}{a}\right)^2}};\ \lvert x \le a\rvert$		$\sigma\sqrt{2}$	$\dfrac{\pi}{2\sqrt{2}} \approx 1.11$
3. Закон равномерной плотности	$p(x)=\begin{cases}\dfrac{1}{2a}\ при\ \lvert x\rvert \le a;\\ 0\ при\ \lvert x\rvert > a.\end{cases}$		$\sigma\sqrt{3}$	$\sqrt{3} \approx 1.73$
4. Треугольный закон. (Распределение Симпсона)	$p(x)=\begin{cases}0\ при\ \lvert x\rvert > a;\\ \dfrac{a-\lvert x\rvert}{a^2}\ при\ \lvert x\rvert \le a.\end{cases}$		$\sigma\sqrt{6}$	$\sqrt{\dfrac{3e}{2}} \approx 2.02$
5. Нормальный закон	$p(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x}{2\sigma^2}}$		∞	$\sqrt{\dfrac{\pi e}{2}} \approx 2.07$

Это обстоятельство является основным препятствием для внедрения «энтропийного» подхода на практике. Особенно остро это проявляется в случаях, когда организация и проведение измерений параметров требует значительных затрат времени и средств. Для получения представительных оценок величины σ_x требуется меньший объем выборки. В настоящее время разработаны методики косвенного оценивания значений величины K_e для различных ситуаций с исходными данными и, в частности, в условиях «дефицита» результатов наблюдений [4, 7, 8]. Таковыми, например, являются метод робастного оценивания, метод косвенного оценивания (на основании характеристик входных воздействий и свойств объекта). Также, в ряде случаев, могут применяться эвристические методики, основанные на использовании аналогий с подобными объектами, явлениями и др. Указанные обстоятельства создают предпосылки для мониторинга вариаций параметров на практике. В качестве «рабочего инструмента» для проведения подобных исследований предлагается использовать зависимость величины половины размаха параметра (или выборки наблюдений) – *a* от значения величины энтропийного коэффициента. В табл. 1, столбцы 4 и 5, приведены соответствующие значения величин *a* и K_e для пяти типовых законов распределений, названия и описания которых приведены в столбцах 1–3. Выбранное множество законов распределений является «представительным» т.к. обеспечивает набор значений энтропийных коэффициентов, «покрывающий» весь реально возможный диапазон их изменения в соответствии с (3).

Из представленных результатов можно сделать следующий вывод. При одинаковой величине СКО (или дисперсии), имеет место тенденция возрастания величины размаха параметра с возрастанием значений величины энтропийного коэффициента и, наоборот. Для унификации представления результатов в качестве показателя вариации параметра введено понятие величины относительного размаха вариации – q, определяемой выражением

$$q = \frac{a}{\sigma}. \tag{4}$$

Зависимость является нелинейной, ее тренд в системе координат K_e; q, основанный на использовании вышеуказанных типовых законов, представлен на рис. 1 пятью реперными точками, соединенными для наглядности аппроксимирующей кривой. Нумерация точек соответствует нумерации типовых законов в табл. 1. Согласно данным табл. 1 и определению (4) координаты этих точек будут: (1.00, 1.00), (1.11, 1.41), (1.73, 1.73), (2.02, 2,45), (2.07, ∞).

Для повышения точности рассматриваемой зависимости количество точек может быть увеличено за счет расширения множества используемых законов распределений. В результате проведенных исследований было установлено, что зависимость является нечеткой или «размытой». Это явление объясняется наличием сюръекции в пространстве вышеуказанных координат для реальных законов распределений (ассиметричных, многомодальных и др.). Другими словами, существуют подмножества законов распределений с одинаковыми значениями величин энтропийного коэффициента, имеющими различные значения величины размаха и, наоборот. На рис. 1 проявления «размытости» условно показаны с помощью затемнений.

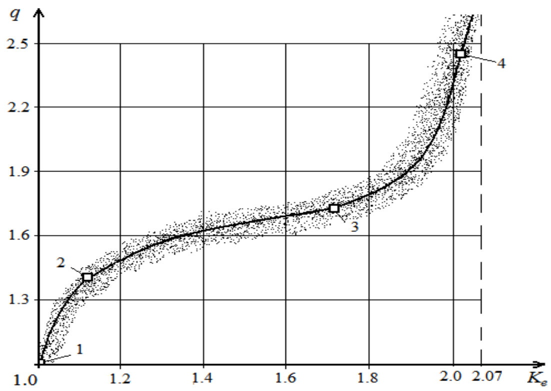


Рис. 1. Зависимость относительного размаха вариации от величины энтропийного коэффициента

Аппроксимация реперных точек может осуществляться различными способами, например, на основе метода наименьших квадратов и др. В ряде случаев весьма удобной является кусочно-линейная аппроксимация зависимости между соседними реперными точками.

Перспективы применения. В значительной мере представленная на рис. 1 зависимость является «обратной» к робастной зависимости для оценки величины K_e по размаху выборки наблюдений, полученной в работе [4]. Ею удобнее пользоваться, например, для решения задач синтеза систем, обеспечивающих попадание выходных параметров в заданную область, когда отклонения различных значений параметров за границы этой области являются одинаково неприемлемыми. Такие задачи имеются в медицине, технике, биологии и др. Обобщенная методика решения будет следующей. Изначально, исходя из назначения и особенностей функционирования системы, задается допустимый диапазон размаха параметра – a. На основании результатов наблюдений находится оценка величины σ и соответствующая оценка величины q из выражения (4). (При этом объем данных, необходимых для нахождения указанных оценок, требуется в разы меньший, чем для нахождения оценки величины K_e с использованием выражения (2) и проведения сопутствующих вспомогательных вычислений и построений.) После чего, с использованием имеющейся зависимости (рис. 1.), определяется оценка предельно-допустимого значения величины энтропийного коэффициента. Если реальное значение K_e превышает допустимое, то проводится коррекция системы. В качестве таковой могут использоваться введение в контур системы специальных корректирующих звеньев, целенаправленная замена отдельных звеньев с целью изменения характеристик и свойств системы и др. Модели и методики трансформации энтропийного коэффициента с использованием системных параметров рассмотрены в [1, 4]. По мере накопления результатов наблюдений, возможно повторение этой процедуры и так далее. В результате реализации итеративного процесса будет осуществляться адаптация системы к изменяющимся условиям ее функционирования. Аналогичным образом полученная зависимость может использоваться для прогнозирования эволюционных процессов и др.

Изложенный подход к исследованию вариаций создает базу для проведения исследований состояний неопределенности различных систем и, в частности, на основе информационных моделей и критериев [1, 5, 8]. Актуальность в проведении таких исследований обусловлена необходимостью повышения качества мониторинга и организации управления системами различной природы [1, 8–10].

Литература

[1] Мягкие измерения и вычисления. Монография в 3-х томах. / Под ред. проф. С.В. Прокопчиной. Том 1. Теоретические основы и методы. М.: ИД «Научная библиотека», 2017. 420 с.

[2] Корн Г., Корн Т. Справочник по математике. Для научных работников и инженеров. М.: Наука. 1978. 832с.

[3] Интеллектуальные системы автоматического управления. /Под ред. И.М. Макарова, В.М. Лохина. М.: Физматлит, 2001. 576 с.

[4] Лазарев В.Л. Теория энтропийных потенциалов. Монография. С-Пб.: Изд-во Политехнического ун-та. 2012. 127 с.

[5] Lazarev V.L. Representative information models for monitoring and control in the conditions of uncertainty // Proceedings of the 18th International Conference on Soft Computing and Measurements, SCM

2015. Saint Petersburg, RUSSIA. Publisher: IEEE. Pp. 54–57. DOI: 10.1109/SCM.2015.7190408.

[6] Электрические измерения неэлектрических величин / А.М. Туричин, П.В. Новицкий, Е.С. Левшина и др. Под ред. П.В. Новицкого. Л.: «Энергия», 1975. 576 с.

[7] Lazarev V.L. Qualimetry of dynamic processes on the basis of entropy potentials concepts // Proceedings of 2017 IEEE 2nd International Conference on Control in Technical Systems, CTS 2017. Saint Petersburg, Russia. Publisher: IEEE. Pp. 87–89. DOI: 10.1109/CTSYS.2017.8109495.

[8] Lazarev V.L. Processing of observations on the basis of information criteria // Proceedings of the 19th International Conference on Soft Computing and Measurements, SCM 2016. Saint Petersburg, Russia. Publisher: IEEE. Pp. 48–50. DOI: 10.1109/SCM.2016.7519679.

[9] Lazarev V.L. Epistemological foundations for generation of perspective competencies in the training of personnel for industrial and economic complex// 4th Forum Strategic Partnership of Universities and Enterprises of Hi-Tech Branches (Science, Education, Innovations). Saint Petersburg, Russia. NOV 11-13, p. 26–28. Publisher: IEEE. DOI: 10.1109/IV Forum.2015.7388242.

[10] Прангишвили И.В. Энтропийные и другие системные закономерности: Вопросы управления сложными системами. М.: Наука, 2003. 428 с.