

УДК 519.814: 621.384.668.8

ã И. Е. Леонов, В. Н. Трифанов, В. М. Шубин, М. М. Нестеров

ИНВАРИАНТНАЯ СТАТИСТИКА В МАСС-СПЕКТРОМЕТРИИ

В работе рассматриваются времяпролетные и квадрупольные масс-спектрографы. В первом случае измеряется интенсивность излучения, воспринимаемая детектором, как функция времени пролета частиц. Во втором случае измеряется интенсивность как функция напряженности магнитного поля. Обработку этих функций предлагается осуществлять методом инвариантной статистики. Этот метод позволяет получать статистический резонансный спектр интенсивностей как функций параметров порядка и вероятности возбуждения этих резонансов. Линии статистического спектра должны фиксировать частицы определенной массы. Вероятности возбуждения этих линий определяют относительные количества частиц наблюдаемых масс.

ВВЕДЕНИЕ

Будем рассматривать времяпролетные и квадрупольные масс-спектрографы. Интенсивность регистрируемого детектором излучения есть функция времени пролета частиц — в первом случае. Во втором случае измеряется интенсивность как функция напряженности магнитного поля.

К обработке этих функций предлагается применить метод инвариантной статистики В.Н. Трифанова. Этот метод позволяет получать статистический резонансный спектр интенсивностей как функций параметров порядка (времени пролета, напряженности магнитного поля) и вероятности возбуждения этих резонансов. Линии статистического спектра должны фиксировать частицы определенной массы. Вероятности возбуждения этих линий определяют относительные количества частиц наблюдаемых масс.

В основе такой статистики лежат инварианты. Они не зависят от числа независимых событий в наблюдаемой совокупности. Такие инварианты позволяют найти резонансы параметров порядка и вероятности их возбуждения. Более того, возможна стратификация наблюдаемого резонансного спектра. Эта процедура открыта, она является синтезом эвристики и технологии. Конкретные образцы масс-спектрометрических данных дают возможность применять разные эвристические варианты их обработки по технологиям инвариантной статистики. Это должно позволить обнаруживать слабо проявленные зубцы и давать им соответствующую интерпретацию. Далее для каждой страты первого уровня можно построить стратификацию второго уровня. Такая процедура повторяется до тех пор, пока не исчезает асимметрия в резонансных спектрах, т. е. пока спектры не станут дискретно равномерными с максимальной энтропией и нулевой информацией. Такое иерархическое раскрытие резонансного спектра назовем

ультраметрическим [1]. В отличие от фрактальной иерархии, в которой соблюдается принцип подобия распределений на всех уровнях, в ультраметрической иерархии этот принцип подобия нарушается. Каждый уровень обладает своим неповторимым резонансным спектром.

ТЕХНОЛОГИЯ ИНВАРИАНТНОЙ СТАТИСТИКИ

Прежде чем изложить технологию инвариантной статистики [2], договоримся, что состояние процесса характеризуется параметром порядка, а вероятность его реализации характеризуется интенсивностью процесса в этом состоянии. При этом допущении изложим технологию инвариантной статистики.

Состояния системы, наблюдаемые в процессе опыта, группируются вокруг отдельных точек с состояниями x_k , $k = 1, 2, \dots, N$. Это означает, что практически вся информация о состояниях системы находится в области этих узловых точек. Континуальная система замещается алгебраической с дискретным множеством состояний x_k с их вероятностями p_k .

Статистические моменты состояний вплоть до момента требуемого порядка в системе замещенных точек образуют замкнутую систему, что позволяет получить однозначный закон распределения резонансных состояний. Число и значение замещающих точек, а также вероятности их возбуждения, подлежат определению по наблюдаемой выборке масс-спектрограммы.

Рассмотрим эту технологию подробнее. Пусть x — неделимое состояние-событие, а совокупность n таких событий обозначим

$$x_n = \sum_1^n x.$$

Если неделимые события однородны, то такая сумма будет равна

$$x_n = nx,$$

а ее среднее значение

$$\bar{x}_n = n\bar{x}.$$

Отклонения от средних равны

$$x_n = x_n - \bar{x}_n, \quad x = x - \bar{x}.$$

В силу центрирования событий средние значения отклонений равны нулю:

$$x_n = \langle x_n - \bar{x}_n \rangle = 0, \quad x = \langle x - \bar{x} \rangle = 0.$$

Однако дисперсии и более высокие центральные моменты, как правило, не равны нулю, характеризуются нестабильностью и зависят от числа неделимых событий в совокупности. Поэтому для статистического анализа они мало информативны, поскольку информация об организации наблюдаемых событий (их внутренней связности) скрывается этой нестабильностью. В противоположность им статистические инварианты стабильны и информативны. Именно они проявляют внутреннюю организацию наблюдаемых событий.

Первый (диссипативный) инвариант (G), отражающий по существу основную идею инвариантной статистики, определяет масштаб флуктуаций неделимого события и равен отношению дисперсии к среднему значению наблюдаемого события:

$$G = x_n^2 / \bar{x}_n = x^2 / \bar{x}.$$

Этот инвариант одинаков как для совокупности событий, так и для их неделимой части. Он стабилен и несет информацию о флуктуациях наблюдаемых событий. Для каждой выборки инвариант свой.

Кроме этого инварианта существует счетное множество инвариантов более высоких порядков. Все они также стабильны и информативны. Чем точнее требуется отображать организованность событий, тем больше таких инвариантов требуется привлекать к анализу наблюдаемых процессов. Такие инварианты можно получить, используя разложение моментов высоких порядков по степеням дисперсии наблюдаемых событий.

Формульная запись инвариантов применительно к моментам до шестого порядка включительно имеет вид:

а) по наблюдаемой совокупности событий:

$$G = D_n / \bar{x}_n, \quad D_n = x_n^2 \quad (\text{центрировано}),$$

$$J_1 = \frac{x_n^3}{D_n},$$

$$J_2 = \frac{x_n^4}{D_n} - 3D_n,$$

$$J_3 = \frac{x_n^5}{D_n} - 10J_1D_n,$$

$$J_4 = \frac{x_n^6}{D_n} - (15J_2 + 10J_1^2)D_n - 15D_n^2;$$

б) для неделимых событий:

$$G = D / \bar{x}, \quad D = x^2 \quad (\text{центрировано}),$$

$$J_1 = \frac{x^3}{D},$$

$$J_2 = \frac{x^4}{D} - 3D,$$

$$J_3 = \frac{x^5}{D} - 10J_1D,$$

$$J_4 = \frac{x^6}{D} - (15J_2 + 10J_1^2)D - 15D^2,$$

где x_n, x — совокупность событий и их неделимая часть; D_n, D — их дисперсии.

Построенные инварианты позволяют определить искомые резонансные спектры неделимых и наблюдаемых событий. Можно получить диадные и триадные спектры неделимых событий. Диадные спектры имеют два резонанса x_1, x_2 ($x_1 < x_2$). Триадные спектры имеют три резонанса x_1, x_3, x_2 ($x_1 < x_3 < x_2$). Эти спектры имеют наименьшую информационную избыточность.

По теореме Колмогорова система с нулевой избыточностью имеет размерность $2 < e = 2.718 < 3$. Коль скоро триада ближе к e , то рассмотрим триадную технологию более подробно.

Исследуем два альтернативных случая:

- в первом среднее значение x_3 наблюдаемо, т. е. ($x_3 = \bar{x}$), и отклонение от него

$$x_3 = \langle x_3 - \bar{x} \rangle = 0;$$

- во втором — среднее значение не наблюдаемо, когда $x_3 \neq \bar{x}$, и отклонение от него не равно нулю

$$x_3 = \langle x_3 - \bar{x} \rangle \neq 0.$$

При этом для обоих случаев размах резонансов равен

$$S = x_2 - x_1.$$

Случай первый

Функциональное уравнение таково:

$$x^m x(x^2 - S_1 x + S_2) = 0, \quad x \in (x_1, 0, x_2),$$

для которого

$$S_1 = x^3 / x^2 = x^3 / D = J_1, \quad S_2 = J_2 - J_1^2 + 3D.$$

При $m = 2$ функциональное уравнение приобретает вид

$$x^5 - S_1 x^4 + S_2 x^3 = 0,$$

решая которое получаем резонансы

$$x_1 = (J_1 - \sqrt{12D + 4J_2 - 3J_1^2}) / 2, \\ x_2 = (J_1 + \sqrt{12D + 4J_2 - 3J_1^2}) / 2.$$

При этом размах резонансов ($S = x_2 - x_1$) удовлетворяет условию

$$S^2 = 12D + 4J_2 - 3J_1^2.$$

Дисперсия неделимого события для x^5 равна

$$D = 0; \quad D = (2J_1 J_2 - J_3 - J_1^3) / (4J_1).$$

Число неделимых событий в наблюдаемой совокупности равно $n = D_n / D$, где D_n — дисперсия наблюдаемой совокупности событий. Однако если размах наблюдаемой совокупности неизвестен, то число неделимых событий в этой совокупности определяется так

$$n = S_n^2 / \left(6D_n + \sqrt{36D_n^2 + S_n^2 (4J_2 - 3J_1^2)} \right).$$

При $m = 3$ функциональное уравнение имеет вид

$$x^6 - S_1 x^5 + S_2 x^4 = 0.$$

Решая его, получаем:

$$D = -\frac{3J_2 + J_1^2}{4} + \\ + \sqrt{\frac{(3J_2 + J_1^2)^2}{16} - \frac{(J_4 - J_1 J_3 + J_2 (J_2 - J_1^2))}{6}}.$$

В обоих случаях выбирается то D , которое лучшим образом удовлетворяет условию

$$0 < D < D_n.$$

После определения D находим число событий в наблюдаемой совокупности $n = D_n / D$.

Вероятность возбуждения резонансов неделимого события выражается так:

$$(x = x_1), \quad p_1 = -D / (x_1 S),$$

$$(x = 0), \quad p = p_3 = (D + x_1 x_2) / (x_1 x_2),$$

$$(x = x_2), \quad p_2 = D / (x_2 S), \quad S = x_2 - x_1.$$

Любое (n) разбиваем на триаду ($n = n_1 + n_3 + n_2$), число таких триад равно

$$N = \frac{(n+1)(n+2)}{2};$$

для каждой триады наблюдается резонансная смесь

$$x = n_1 x_1 + n_3 x_3 + n_2 x_2 = n_1 x_1 + n_2 x_2, \quad x_3 = 0.$$

Вероятность возбуждения такой смеси равна

$$p = \frac{n!}{n_1! n_3! n_2!} p_1^{n_1} p_3^{n_3} p_2^{n_2}.$$

Случай второй

В принятых обозначениях получаем необходимое количество функциональных уравнений:

$$x^3 - S_1 x^2 + S_2 x - S_3 = 0, \\ x^4 - S_1 x^3 + S_2 x^2 - S_3 x = 0, \\ x^5 - S_1 x^4 + S_2 x^3 - S_3 x^2 = 0, \\ x^6 - S_1 x^5 + S_2 x^4 - S_3 x^3 = 0.$$

Из первых двух уравнений находим

$$S_3 = x^3 - S_1 x^2, \quad -S_2 = x^4 / x^2 - S_1 x^3 / x^2.$$

Включая два последних уравнения, получаем:

$$D^3 - A_1 D^2 + A_2 D - A_3 = 0, \\ A_1 = 2(J_1^2 - J_2), \\ A_2 = \frac{2(J_4 - J_2^2) + 9(J_2^2 - J_1^4) - 12J_1(J_1 J_2 - J_3)}{12}, \\ A_3 = \frac{(J_3 - J_1 J_2)^2 - (J_4 - J_2^2)(J_2 - J_1^2)}{12}, \\ x^3 - S_1 x^2 + S_2 x - S_3 = 0, \\ S_1 = \frac{J_3 - J_1 J_2 + 6J_1 D}{J_2 - J_1^2 + 2D}, \\ -S_2 = J_2 - J_1 S_1 + 3D, \\ S_3 = (J_1 - S_1)D.$$

Сначала решаем кубическое уравнение для D . Из трех корней выбирается тот, который удовлетворяет условию $0 < D < D_n$. Затем находим полиномиальные инварианты (S_1, S_2, S_3). После чего решаем кубическое уравнение для резонансов независимого события:

$$x = (x_1, x_3, x_2), \quad x_1 < x_3 < x_2.$$

При этом размах резонансов будет равен $S = x_2 - x_1$.

Может оказаться, что D не вписывается в указанные границы. Тогда возможны два случая:

а) $D < 0$, в этом случае задается n и определяется $D = D_n / n$;

б) $D > D_n$, в этом случае принимается $D = D_n$, $n = 1$.

Вероятности резонансов неделимого события:

$$(x = x_1), \quad p_1 = (D + x_2 x_3) / ((x_1 - x_2)(x_1 - x_3)),$$

$$(x = x_3), \quad p_3 = (D + x_1 x_2) / ((x_3 - x_1)(x_3 - x_2)),$$

$$(x = x_2), \quad p_2 = (D + x_1 x_3) / ((x_2 - x_1)(x_2 - x_3)).$$

Число оболочек резонансов наблюдаемых событий равно $n = D_n / D$. Это число разбивается на триаду ($n = n_1 + n_3 + n_2$), число таких триад равно

$$N = \frac{(n+1)(n+2)}{2}.$$

Весь спектр наблюдаемых событий определяется числом таких разбиений в виде смесей

$$x = n_1 x_1 + n_3 x_3 + n_2 x_2 = n_1 x_1 + n_2 x_2, \quad x_3 = 0.$$

Вероятности возбуждения каждой такой смеси равны

$$p = \frac{n!}{n_1! n_3! n_2!} p_1^{n_1} p_3^{n_3} p_2^{n_2}.$$

Таким образом, задача определения резонансного спектра и вероятности возбуждения решена.

Отметим, что резонансные спектры масс-спектрометрии выделяют наиболее вероятные точки сгущения частиц соответствующих масс. Однако наблюдаемые масс-спектрограммы континуальны и, как правило, многомодальны. В этом случае необходимо искать область притяжения каждой резонансной точки сгущения. Это проблема стратификации масс-спектрограмм. Решим ее.

СТРАТИФИКАЦИИ МАСС-СПЕКТРОГРАММ

Рассмотрим два ближайших резонанса наблюдаемой смеси (x_k, x_{k+1}). Каждый из этих резонансов является резонансным центром (точкой сгущения) своей страты (S_k, S_{k+1}), в пределах каждой из которых случайно флуктуируют резонансы простых событий $x = (x_1, x_3, x_2)$ с вероятностями $p = (p_1, p_3, p_2)$. Эти флуктуации создают конусы рассеяния резонансов в пределах своих страт. Это рассеяние в одну сторону характеризуется резонансом x_1 , а в другую сторону резонансом x_2 . Граница между стратами определяется точкой пересечения конусов рассеяния соседних страт (обозначим ее S_k). В результате встречной диссипации резонансов соседних страт получаем $S_k = x_k + n x_2 = x_{k+1} - n x_1$. Из этого выражения находим параметр n и, подставляя его значение в исходное уравнение, находим

$$S_k = (x_2 x_{k+1} - x_1 x_k) / S, \quad S = x_2 - x_1.$$

Следует отметить фундаментальную особенность этого выражения. В нем внутренние граничные условия (x_1, x_2) согласованно сочетаются с внешними граничными условиями, проявленными в наблюдаемых резонансах смесей (x_k, x_{k+1}). Еще отметим, что равномерное распределение получится при условии ($x_2 = -x_1$), тогда $S_k = (x_{k+1} + x_k)/2$. Как правило, такая ситуация возникает в статистически равномерных системах с максимальной энтропией. Но в реальных системах, обычно статистически неравномерных, такие случаи практически не встречаются в силу их неинформативности.

Задача поиска резонансных точек сгущения в масс-спектрограммах для частиц с соответствующими массами требует информации. Такая информация присутствует в несимметричных неравновесных системах. Если нет информации, то и нет частиц с наблюдаемыми массами. Поэтому неравномерность распределения является хорошим критерием иерархического поиска частиц в ультратометрических структурах.

Но рассмотренная ситуация для стратификации масс-спектрограмм не единственная. В общем случае между любыми двумя состояниями можно найти равновесную границу по изложенной выше схеме. При таком подходе возникает возможность классификации потока частиц на отдельные классы с установленными границами между этими классами.

СПИСОК ЛИТЕРАТУРЫ

1. Нестеров М.М., Данилов В.Н., Леонов И.Е. Применение ультратометрической адаптивной статистики для анализа структуры масс-спектрометрического сигнала // Труды Санкт-Петербургского института информатики и автоматизации. 2005. Т. 2, вып. 2. С. 379–385.
2. Трифанов В.Н. Инвариантный статистический анализ и управление в транспортных системах. СПб.: Элмор, 2003. 192 с.

Санкт-Петербургский институт информатики и автоматизации РАН (Леонов И.Е., Нестеров М.М.)

Институт проблем транспорта РАН, Санкт-Петербург (Трифанов В.Н.)

СКБ АП ФГУП "ПО Маяк", г. Озерск Челябинской обл. (Шубин В.М.)

Материал поступил в редакцию 12.11.2007.

INVARIANT STATISTICS IN MASS-SPECTROSCOPY**I. E. Leonov¹, V. N. Trifanov², V. M. Shubin³, M. M. Nesterov¹**¹*Institute for Computer Science and Automation RAS, Saint Petersburg*²*Institute of Transport Problems RAS, Saint Petersburg*³*FSUC SKB AP "PO Mayak", Chelyabinsk oblast, Ozersk town*

The paper considers flight time and quadrupole mass-spectrometers. In the first case, the radiation intensity is measured versus the particle flight time; in the second case, the intensity is measured versus the magnetic field strength. Both functions are supposed to be processed by the invariant statistics technique that enables one to get a statistical resonance spectrum of intensities as a function of the order parameters and the probability of those resonances excitation. The statistical spectrum lines correspond to particles with certain masses. Probabilities of those lines excitation determine relative amounts of particles with the masses measured.