О ВЫЧИСЛЕНИИ ПОКАЗАТЕЛЕЙ АСИММЕТРИИ ДИНАМИЧЕСКИХ КРИВЫХ С КОНЕЧНЫМ ЧИСЛОМ ФИКСИРОВАННЫХ ТОЧЕК

С. И. ШТЕЙН

Для характеристики кривых и полигонов, кроме прочих показателей, применяют и меры, устанавливающие их асимметрию. При распределениях это сводится к определению различий в скученности вариант в симметричных вариациях, а вообще — к определению степени неравенства ординат при симметричных абсциссах или, наоборот, — неравенства абсцисс при симметричных ординатах.

Кривые или полигоны распределений, процессов или состояний построены по координатам, которые могут выражать: а) признаки (свойства), присущие вариантам; б) частоты (численности) вариант; в) положения (размещение) вариант в пространстве или времени.

Линии подобного рода (в дальнейшем — «кривые») можно сгруппировать в два типа:

кривые распределений (в дальнейшем — «статистические»), образованные координатами, определяющими: а) признаки и частоты; б) положения и частоты;

кривые процессов или состояний (в дальнейшем — «динамические»), образованные координатами, определяющими: а) признаки и признаки; б) признаки и положения; в) положения и положения.

Статистические кривые подчинены законам распределения частот вариаций, изучаемым теорией вероятностей. Строение динамических кривых не имеет непосредственного отношения к теории вероятностей, и характер каждой из них есть выражение закономерностей изображенного ею процесса.

Многие из динамических кривых обладают строением, к которому не имеет смысла прилагать представление о симметрии. Это линии, выражающие прямые или обратные функциональные и корреляционные зависимости. Но, наряду с такими, есть кривые, в которых знак зависимости меняется по ходу течения процесса, то есть кривые, имеющие восходящие и нисходящие ветви, а иногда и ветви, неоднократно меняющие направления. К ним относятся, например, кривые изменений: метеорологических элементов во времени; реакции организмов на воздействия факторов среды; эффективности удобрений с увеличением их дозировок; течения многих технологических и экономических процессов; вообще, цикличных процессов в природе и т. п. К подобным кривым вполне приложимо представление о симметрии и, следовательно, имеет значение установление характеристик, определяющих их асимметрию.

Числовое выражение особенностей геометрического соотношения ветвей кривой в одном показателе имеет большое значение, так как при посредстве таких показателей можно сравнивать динамику однородных процессов или состояний в разных условиях. Следствием этого является утверждение, что показатели асимметрии особенно применимы в случаях сравнения кривых однородных процессов и состояний и имеют меньшее значение для характеристики отдельно взятых линий.

В литературе, посвященной математической обработке результатов экспериментов и наблюдений, в подавляющем большинстве случаев нет указаний на применение для динамических кривых показателей асимметрии, отличных от показателей, применяемых при анализе кривых статистических. Между тем нельзя уподоблять статистические и динамические кривые, потому что у них различны характер зависимостей между координатами, закономерности, по которым они формируются и, часто, строение. В связи с этим возникает вопрос о применимости статистических формул и показателей для характеристики асимметрии динамических кривых.

Для того чтобы с помощью показателя можно было охарактеризовать асимметрию кривой, он должен отвечать определенным условиям.

Юл и Кендэл (1960) считают, что мера, определяющая степень отклонения распределения численностей от симметрии, «должна: а) быть числом отвлеченным, иначе она не будет независима от единицы измерения переменной, и б) быть равной нулю, если распределение симметрично».

Круг обязательных условий для показателей асимметрии следует расширить. Но их перечень надо предварить общим замечанием, суть которого не всегда учитывают в статистике: любая статистическая или динамическая кривая — не самостоятельная фигура, а только часть фигуры простого пространства, заключенного между нею, ординатами крайних точек и осью абсцисс. Поэтому рассмотрение симметрии кривой изолировано от остальных элементов всей фигуры неправомерно и может привести к неправильным выводам.

Для того чтобы с помощью показателей можно было охарактеризовать асимметрию любых кривых, формулы, по которым их вычисляют, должны отвечать следующим условиям (кроме указанных Юлом и Кендэлом):

а) возможности вычислять показатель как по измерениям координат кривой, так и непосредственно из числовых значений, по которым она построена;

б) универсальности, то есть применимости к любым типам кривых и числовых рядов;

в) способности отражать не только направленность и размеры отклонений от симметрии, но и полные величины вариаций, с тем чтобы определять и относительное значение асимметрии;

г) геометрической выраженности, поскольку геометрично само понятие асимметрии, а для этого необходимо учитывать масштабы построения кривых по обеим осям.

В статистике предложен ряд формул для вычисления показателей (коэффициентов) асимметрии кривых. Сам факт множественности формул свидетельствует об их эмпирическом характере. Даже для характеристик кривых распределения некоторые из них ограничены известными пределами. В основы этих формул положены соотношения числовых значений, более или менее абстрагированные от положения соответствующих им точек на плоскости, и все они не связаны с масштабами изображения кривых (по крайней мере, по одной из осей).

Среди этих формул можно выделить три типа:

1. Формула Кеппена, выражающая соотношения между количествами вариант плюс и минус — вариаций, вне зависимости от их модулей и размеров отклонений:

$$A_1 = \frac{\overset{-}{n} - \overset{+}{n}}{\overset{-}{n} + \overset{+}{n}} 100,$$

где $\overset{-}{n}$ и $\overset{+}{n}$ — количество случаев соответствующих отклонений от средней. Показатель A_1 не меняет своей величины при умножении $\overset{-}{n}$ и $\overset{+}{n}$ на любое целое число, а также при перемене мест ординат вдоль оси абсцисс в пределах каждой ветви.

2. Формула определения показателя асимметрии по способу моментов. Это классическая, наиболее употребляемая формула учитывает не модули вариаций, а статистические параметры их отклонений от средней. В этой формуле модули вариаций по другой оси (ординат) учитывают только в случае, если они выражают частоты:

$$A_2 = \frac{\mu_3}{\mu_2{}^{3/2}} = \frac{\Sigma p\alpha^3}{N\sigma^3},$$

где: μ_2 и μ_3 — второй и третий моменты;

α — отклонения вариаций от средней;

P — численность соответствующих отклонений;

N — число всех вариант ряда;

σ — стандартное отклонение.

Показатель A_2 не меняет своей величины при изменении всех абсцисс на какую-либо величину, а для кривых, в которых ординаты выражают не частоты, а значения вариант — при изменении величины ординат и перемене их мест в пределах ветви.

3. Формулы, включающие полные значения некоторых характеристик ряда по одной из осей, но учитывающие координаты по другой оси только в случаях, если они выражают частоты. К ним относятся:

формула Пирсона: $A_3 = \frac{\overline{X} - M_0}{\sigma}$;

формула Юла и Кендэла: $A_4 = \frac{Q_1 + Q_3 - M_e}{Q}$,

формула, определяющая показатель асимметрии по соотношению медианы, средней и стандартного отклонения:

$$A_5 = \frac{3(M_e - \overline{X})}{\sigma}.$$

В этих формулах Q_1 и Q_3 — квартили, Q— квартильное отношение $\frac{Q_3 - Q_1}{2}$, $\overline{X}$ — средняя, M_0 — мода, M_e —медиана и σ — стандартное отклонение.

Формулы A_3, A_4 и A_5, примененные к динамическим кривым, не учитывают модулей вариант по одной из осей, и поэтому не учитывают и соотношения между координатами точек кривых.

Приведенные соображения дают основание считать, что:

а) динамические кривые процессов и состояний отличаются от кривых статистических распределений тем, что отражают для каждой варианты ряда не один ее признак (или положение), а два;

б) поскольку статистические и динамические кривые не изолированные линии, а части фигур, их симметрию следует рассматривать с точки зрения симметрии фигур в целом;

в) величины показателей асимметрии имеют служебное значение для сравнения однородных кривых и несущественны как характеристики для отдельно взятых линий;

г) формула для вычисления показателей асимметрии динамических кривых должна давать возможность вычислять их как по геометрическим данным, так и по числовым значениям вариант рядов, отражать не только величины и направленность отклонений, но и модули вариаций, иметь геометрическую выраженность, рассматривая кривые не изолированными, а как части фигур, и обладать универсальностью в части применимости к любым типам кривых;

д) рассмотренные пять наиболее распространенных статистических формул вычисления показателей асимметрии кривых распределения в той или иной мере не отвечают требованиям, предъявляемым к формулам показателей асимметрии динамических кривых и не пригодны для характеристики последних.

Нами предложена формула для аналогичных вычислений, соответствующая перечисленным условиям. В основу формулы положены соотношения величин, являющихся модулями координат точек кривых или числовыми значениями, по которым можно построить графики.

Задачи исследования симметричности ветвей кривых, расположенных по обе стороны оси симметрии (она же одна из нулевых координат), сводятся к установлению наличия или отсутствия зеркальной конгруэнции ее ветвей, а в случае отсутствия таковой — к определению сравнительного значения отклонений ветвей от симметрии.

Наиболее простой способ установления наличия или отсутствия симметрии и размеров отклонений от нее — попарное сравнение координат противоположных точек, равноудаленных от оси симметрии, отнесенное к расстоянию от нее до каждой из точек. Такое сравнение приводит к выражению:

$$\frac{Yr - Yl}{|x|},$$

где Yr — ордината точки на правой (положительной) ветви или число ряда, ей соответствующее;

Yl — то же на левой (отрицательной) ветви;

$|x|$ — абсолютная величина расстояния каждой из противоположных равноудаленных точек от оси симметрии (по оси абсцисс) или соответствующее ей отклонение значения вариации от средней.

Это выражение приобретает знак стороны с большим абсолютным значением ординаты сравниваемых точек.

В случае совмещения оси симметрии не с осью ординат, а с осью абсцисс значение символов меняется.

Рассматривать асимметрию положения точек только в зависимости от расстояния между ними по одной из осей — неправильно, потому что всякая динамическая кривая — не изолированная линия, а линия, составляющая одну из границ фигуры. Иначе говоря, симметрию кривых следует рассматривать не как симметрию линий, а как симметрию фигур на плоскости.

Следовательно, приведенное выше выражение само по себе не может дать достаточного представления об асимметрии, так как при любом изменении ординат ее значение относительно всей фигуры меняется. Поэтому это выражение должно быть отнесено к среднему значению ординат тех же точек $\frac{Yr + Yl}{2}$ и примет вид:

$$\frac{2(Yr - Yl)}{|x|(Yr + Yl)}.$$

Если показатель вычисляют непосредственно по числовому ряду, то значение $|x|$ определяют по отклонению абциссы точки от абсциссы вариации, приходящейся на ось симметрии. При отсутствии на противоположных ветвях кривой фиксированных равноудаленных от оси симметрии точек их значения Yr или Yl вычисляют интерполированием.

Асимметрия кривой складывается из асимметричного расположения всех ее точек. Поэтому суммарная асимметрия кривой, которой мы присваиваем символ As и которую считаем показателем асимметрии, равна

$$As = 2 \sum_{\substack{r=o \\ l=o}}^{\substack{r=m \\ l=n}} \frac{Yr - Yl}{|x|(Yr + Yl)}.$$

В случае, когда фиксированное число точек (данное или интерполированное) нечетно и одна из них находится на оси симметрии, — ее не учитывают, так как она не относится ни к одной из ветвей.

Суммарная асимметрия одной и той же кривой с увеличением числа фиксированных точек может изменяться. Поэтому ее значения следует применять только для сравнения однородных кривых с одинаковым числом сопоставляемых точек и одинаковыми значениями $|x|$.

Показатель As равен нулю при условиях:

а) когда все Yr, соответственно равны Yl, то есть в случае полной симметрии, при которой

$$\Sigma(Yr - Yl) = 0,$$
$$Yr = Yl;$$

б) когда суммарная асимметрия кривой складывается из двух, равных по величине, но противоположных по знаку сумм, из которых одна включает все случаи при $Yr > Yl$, а другая — при $Yr < Yl$:

$$\sum_{Yr > Yl} \frac{Yr - Yl}{|x|(Yr + Yl)} + \sum_{Yr < Yl} \frac{Yr - Yl}{|x|(Yr + Yl)} = 0.$$

Для подобного расположения точек на ветвях кривой мы предлагаем название «компенсированная асимметрия» и символ Ок.

Знак при показателе асимметрии, вычисленный по предлагаемой формуле, соответствует обозначениям, принятым для системы прямоугольных координат: вправо от нуля — плюс, влево — минус. Мы считаем это более удобным, чем употребляемое в математической статистике обратное обозначение (поскольку в статистике не принимают нулевую координату за ось симметрии).

Пределы изменения (размах варьирования) величины As равны

$$\pm 2 \overset{p}{\Sigma} \frac{1}{|x|},$$

где p — число пар сравниваемых точек кривой.

Уравнение предела показателя асимметрии связывает его с величиной гармонической средней:

$$\lim |As| = 2\sum^{p} \frac{1}{|x|} = \frac{2N}{H},$$

где: N — число членов ряда,

H — гармоническая средняя, равная $\frac{N}{\Sigma \frac{1}{x}}$.

Предельное значение показателя асимметрии может быть достигнуто только когда все Yr или Yl равны нулю или при смещении оси симметрии вдоль кривой до полного выноса за ее пределы.

Отсюда можно вывести следствие: всякая кривая, лежащая вне оси симметрии, асимметрична; показателем ее асимметрии служит частное от деления удвоенного числа точек кривой (ряда) на ее гармоническую среднюю.

Предлагаемая формула дает возможность определять асимметрию любых кривых и полигонов (в том числе и статистических), расположенных по обе стороны от нулевой координаты или вариационных рядов, по которым они составлены, и соответствует требованиям, предъявляемым к подобным формулам.

Приводим пример вычисления показателей асимметрии выпадения осадков (в мм) по четырехдекадным периодам (метеостанция Толстовка Амурской области).

	Yl				*Ось сим.*	Yr				
Периоды	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*	
Ср. за 50 лет	3	5	33	111	140	127	36	7	4	466
1951 г.	2	2	10	20	173	330	66	2	5	610
1943 г.	0	0	11	202	116	68	14	0	0	411

Сравнив. периоды	Yr — Yl	Yr + Yl	|x|	$\frac{(Yr - Yl)}{\|x\| (Yr + Yl)}$
		Среднее за 50 лет		
VI—IV	16	238	1	0,067
VII—III	3	69	2	0,022
VIII—II	2	12	3	0,056
IX—I	1	7	4	0,036
				Σ= 0,181 As = 0,181×2= 0,362
		1951 г.		
VI—IV	310	350	1	0,886
VII—III	56	76	2	0,368
VIII—II	0	4	3	0
IX—I	3	7	4	0,107
				Σ= 1,361 As = 1,361×2= 2,722
		1943 г.		
VI—IV	—134	270	1	—0,496
VII—III	3	25	2	0,060
VIII—II	0	0	3	0
IX—I	0	0	4	0
				Σ=—0,436 As =—0,436×2=—0,872