

УДК 574.3, 591.3

МЕРЫ СХОДСТВА И РАЗНООБРАЗИЯ В ОЦЕНКЕ ФЛУКТУИРУЮЩЕЙ АСИММЕТРИИ БИЛАТЕРАЛЬНЫХ ПРИЗНАКОВ

Д.Б. Гелашвили, Е.Н. Солдатов, Е.В. Чупрунов

*Нижегородский государственный университет им. Н.И. Лобачевского
Россия, 603950, Н. Новгород, пр. Гагарина, 23*

Поступила в редакцию 21.05.04 г.

Меры сходства и разнообразия в оценке флуктуирующей асимметрии билатеральных признаков. – Гелашвили Д.Б., Солдатов Е.Н., Чупрунов Е.В. – Предложено и обосновано применение свертки функций как меры сходства при оценке псевдосимметрии биообъектов и ее частного случая – флуктуирующей асимметрии. Приведен алгоритм оценки разнообразия биообъектов по степени флуктуирующей асимметрии билатеральных признаков с помощью информационной энтропии.

Ключевые слова: флуктуирующая асимметрия, свертка функций, информационная энтропия.

Measures of similarity and diversity in the analysis of fluctuating asymmetry of bilateral features. – Gelashvili D.B., Soldatov E.N., Chuprunov E.V. – The usage of function convolution and its particular case (fluctuating asymmetry) as a measure of similarity at estimating the pseudosymmetry of biological objects is proposed and justified. An algorithm to analyze the diversity of biological objects by the fluctuating asymmetry level of bilateral features in terms of information entropy is given.

Key words: fluctuating asymmetry, function convolution, information entropy.

ВВЕДЕНИЕ

Симметрия, точная или приближительная, является важнейшим свойством подавляющего числа живых организмов (Гиляров, 1944; Урманцев, 1974; Марченко, 1999 и др.). С общих естественно-научных позиций симметрия как имманентная характеристика тел материального мира может быть определена как инвариантность физической и геометрической системы по отношению к различного рода преобразованиям (Вейль, 1968; Чупрунов и др., 2000). Вид этих преобразований определяет тип симметрии рассматриваемого объекта. Чаще всего под симметрией понимают инвариантность относительно изометрических преобразований пространства (операций симметрии) – поворотов, отражений в плоскости, инверсии в точке, параллельных переносов пространства, а также всевозможных их комбинаций. Однако можно говорить относительно перестановок отдельных частей системы, о симметрии подобия и т.д. (Шубников, Копчик, 1972).

Будем называть симметричным объект, который состоит из частей, равных относительно какого-либо признака. Проверить равенство объектов или частей системы можно с помощью некоторого преобразования, которое совмещает равные объекты или части одного и того же объекта. В природе чаще всего встречаются лишь приблизительно симметричные (псевдосимметричные) системы, об инвариантности которых относительно операций симметрии также можно говорить лишь приблизительно. Приближительная симметрия может возникать, например, в результате процессов, сопровождающихся нарушением идеальной симметрии.

К такому типу изменений можно отнести флуктуирующую асимметрию (ФА) билатеральных признаков, под которой понимают незначительные и случайные (ненаправленные) отклонения от строгой билатеральной симметрии биообъектов (Захаров, 1987). Авторы статьи считают, что асимметрия, наряду с симметрией, – имманентная дуалистическая характеристика биообъекта, неизбежно проявляющаяся в онтогенезе.

Таким образом, ФА организмов по билатеральным признакам можно рассматривать как случайное макроскопическое событие, заключающееся в независимом проявлении либо на левой, либо на правой, либо на обеих сторонах биообъекта, но в разной степени выраженных признаков, являющихся итогом стохастических микроскопических процессов.

На макроскопическом уровне ФА предлагают использовать в качестве меры в оценке стабильности развития организма (Захаров, 2001; Захаров и др., 2001). В качестве показателей стабильности развития обычно рассматривают нарушения развития (фенодевиации) и онтогенетический шум (Захаров и др., 2001). Следовательно, ФА может служить и мерой онтогенетического шума и, как подчеркивают В.М. Захаров и др. (2001, с. 405), «... преимущество подхода состоит в том, что при этом известна генетическая заданная норма – симметрия, отклонения от которой в ходе развития и представляют собой онтогенетический шум».

Перспективы применения ФА в системе биомониторинга требуют разработки корректных методов количественной оценки ее величины (Кожара, 1985; Захаров, 1987; Гелашвили и др., 2001, 2004; Palmer, Strobeck, 1986, 2003).

Применение в формулировках таких терминов, как «симметрия/асимметрия», «флуктуации», «шум», имеющих вполне содержательное наполнение в физико-математической литературе, дает основание рассмотреть проблему ФА с междисциплинарных позиций физики, теории вероятностей и теории информации.

Таким образом, целью настоящей работы является обоснование возможностей и ограничение применения для количественных оценок ФА алгоритмов и методов, апробированных в кристаллографии, прикладной теории информации и популяционной биологии.

1. Меры сходства для оценки величины флуктуирующей асимметрии

Количественная оценка сходства между различными биосистемами или различными частями одной системы – традиционная задача генетики популяций, популяционной экологии, фенетики, для решения которой вводятся те или иные меры сходства или родственные им меры расстояния (например, генетического) между популяциями (Животовский, 1979).

Рассмотрим в качестве рабочей гипотезы предположение о том, что сходство между биосистемами (популяциями, организмами, частями организмов) можно оценить по равенству этих систем относительно какого-либо признака. Тогда, как указывалось выше, для проверки этого равенства необходимо с помощью некоторого преобразования «совместить» равные по данному признаку объекты. Очевидно, что мы приходим к определению симметрии, а поскольку живая природа характеризуется изменчивостью и разнообразием, речь должна идти о количествен-

ной оценке псевдосимметрии биосистем. Таким образом, псевдосимметрия может рассматриваться как универсальная характеристика биосистем различного масштаба и сложности, а ФА – как ее частный случай.

Критический анализ существующих подходов к оценке ФА, учитывающих либо частоту проявления меристических признаков, либо величину пластических, дан в работе Д.Б. Гелашвили и др. (2004).

Очевидно, что для оценки величины ФА по интегральному показателю имеет значение как частота проявления асимметричного признака (т.е. «пенетрантность»), так и его выраженность (величина), или «экспрессивность».

Алгоритмом, удовлетворяющим не только этим требованиям, но и позволяющим оценивать полностью симметричные и антисимметричные объекты, является свертка функций, успешно применяемая для оценки степени инвариантности атомных структур псевдосимметричных кристаллов относительно изометрических преобразований пространства (Чупрунов и др., 1988) и предложенная нами для оценки степени симметричности биообъектов (Гелашвили и др., 2001, 2004). Свертка может быть интерпретирована как скалярное произведение функций, образующих бесконечномерное пространство векторов, и записана в виде функционала

$$\eta_{\bar{g}}[f(x_1, x_2, x_3)] = \frac{\int_{\Omega} f(x_1, x_2, x_3) f(\bar{g}(x_1, x_2, x_3)) dx_1 dx_2 dx_3}{\int_{\Omega} f(x_1, x_2, x_3)^2 dx_1 dx_2 dx_3}, \quad (1)$$

где интегрирование ведется по всей области Ω определения функции $f(x_1, x_2, x_3)$.

Величина асимметрии в этом случае может быть представлена выражением

$$\bar{A}_1 = 1 - \eta_{\bar{g}} = 1 - \frac{1}{m} \sum_{i=1}^m \frac{2 \sum_{j=1}^{n_i} L_{ij} \cdot R_{ij}}{\sum_{j=1}^{n_i} (L_{ij}^2 + R_{ij}^2)}, \quad (2)$$

где L_{ij} и R_{ij} – значение j -го признака соответственно слева и справа от плоскости симметрии; m – число особей в выборке, n – число признаков.

Метод оценки степени инвариантности биосистем, основанный на использовании функционала (1), является достаточно универсальным и эффективным по следующим причинам:

- величина $\eta_{\bar{g}}[f(x_1, x_2, x_3)]$ всегда находится в интервале от -1 до +1, что позволяет использовать его для оценки симметрии, асимметрии и антисимметрии, а также сравнивать оценки, полученные с помощью батареи тест-организмов;
- придавая разный смысл функции $f(x_1, x_2, x_3)$ – длина, число отверстий, цвет и т.д. – можно оценивать степень симметричности биообъектов практически по любой комбинации признаков по единому алгоритму;
- определяя соответствующим образом оператор преобразования $\bar{g}$, можно оценивать степень симметричности биообъектов не только относительно билатеральных признаков, но и других возможных типов преобразований (поворотной, трансляционной и др.);

- предлагаемый подход дает возможность оценить симметрию (асимметрию) с любой степенью подробности, поскольку учитывает не только альтернативное, но и непрерывное варьирование признака;

- наконец, алгоритм свертки позволяет автоматизировать распознавание степени симметричности биообъектов с помощью компьютерных технологий.

Проверка применимости свертки функций для количественной оценки ФА по меристическим и пластическим признакам у растений, насекомых, амфибий, рептилий, млекопитающих показала ее приемлемость и эффективность (Гелашвили и др., 2001, 2004; Радаев, 2001).

В упоминавшейся работе Л.А. Животовского (1979) приводится несколько мер сходства популяций по полиморфным признакам, предложенных разными авторами:

$$S_1 = \sum_{i=1}^m p_i q_i, \quad S_2 = \frac{\sum_{i=1}^m p_i q_i}{\sqrt{\left(\sum_{i=1}^m p_i^2\right) \left(\sum_{i=1}^m q_i^2\right)}}, \quad S_3 = \frac{\sum_{i=1}^m p_i q_i}{\frac{1}{2} \sqrt{\left(\sum_{i=1}^m p_i^2 + q_i^2\right)}},$$

в том числе предложенная самим Л.А. Животовским (1979) мера $r = \sum_{i=1}^m \sqrt{p_i q_i}$.

Очевидно, что все рассматриваемые меры алгоритмически являются свертками, а мера S_3 полностью аналогична предложенной нами. Заметим, что ранее И.М. Хохуткин и Ю.А. Елькин (1982) также применили алгоритм свертки для оценки видового сходства биотических сообществ наземных моллюсков.

Заметим, что показатель r Л.А. Животовского с точностью до преобразования также соответствует формуле (2). Действительно, запишем выражение (2) и показатель r в следующем виде:

$$\frac{2(L \cdot R)}{L^2 + R^2} = \frac{2}{L^2 + R^2} (L \cdot R) = k(L \cdot R)^n, \quad n=1,$$

$$r \approx \sqrt{p_i q_i} = k(p_i q_i)^n, \quad n=1/2, \quad k=1 \quad \text{или в общем случае} \quad k(a \cdot b)^n.$$

При этом утверждение Л.А. Животовского (1979) об исключительных статистических преимуществах показателя r несколько преувеличено и не согласуется с данными приведенного в статье тестового примера относительно редких морф.

Таким образом, мера сходства, основанная на свертке функций, может считаться достаточно универсальной для оценки псевдосимметрии биообъектов различной сложности и масштаба.

2. Меры разнообразия для оценки ФА билатеральных признаков

Меры сходства, рассмотренные выше, дают количественную оценку степени симметрии/асимметрии, но не отражают вероятностный характер проявления признаков. Между тем «фенотипическая судьба» флуктуирующего признака такова, что он может проявиться или не проявиться у конкретного объекта. В том случае, если признак проявился, он может встречаться в одном из двух альтернативных

состояний: асимметричном (с вероятностью p^1) и симметричном (p^0). Очевидно, что должно выполняться соотношение $p^1 + p^0 = 1$. Оценка вероятностей состояний признаков дает возможность перейти к характеристике их разнообразия. Перефразируя Р. Маргалефа (1992), можно сказать, что разнообразие – мера информации, подразумевающая, что признаки одного типа неразличимы. Традиционно применяемые в биологии меры разнообразия имеют междисциплинарное происхождение. Так, индекс доминирования Симпсона

$$\sum_{i=1}^m p_i^2,$$

вероятно, впервые был использован немецкими криптографами как мера частоты повторов (Хэйек, 2003). В экологии величина

$$\left(\sum_{i=1}^m p_i^2 \right)^{-1}$$

известна как индекс видового разнообразия Симпсона, а в популяционно-генетических исследованиях – как мера эффективного числа аллелей (Животовский, 1980). Л.А. Животовский (1980) предложил для оценки внутривидового разнообразия по комплексу признаков меру

$$\mu = \left(\sum_{i=1}^m \sqrt{p_i} \right)^2,$$

которая принимает максимальное значение

$$\lim_{p_i \rightarrow \frac{1}{n}} \mu \left(\sum_{i=1}^m \sqrt{p_i} \right)^2 = n,$$

равное числу признаков при их равновероятности $p_i = 1/n$.

Широкое распространение в экологических исследованиях получила информационная энтропия К. Шеннона (Shannon, 1948):

$$H = - \sum_{i=1}^m p_i \log_2 p_i.$$

Впоследствии А. Реньи (Rényi, 1961) расширил концепцию информации К. Шеннона и предложил семейство обобщенных энтропий:

$$H_\alpha = \frac{1}{\alpha} \ln \sum_{i=1}^m p_i^\alpha,$$

где $-\infty \leq \alpha \leq +\infty$.

Наконец, М.О. Хилл (Hill, 1973) для оценки биоразнообразия предложил параметрическое семейство индексов, взяв экспоненту от обобщенной энтропии А. Реньи:

$$H_\alpha = \left(\sum_{i=1}^m p_i^\alpha \right)^{1/\alpha}.$$

Детальный анализ этих и других мер разнообразия приведен в работе Ю.А. Песенко (1982). Преимущество той или иной меры разнообразия во многом определяется задачами исследования. Следует согласиться с Л.А. Животовским,

который считает, что «задача полного установления "сфер действия" всех мер разнообразия остается нерешенной» (Животовский, 1980, с. 835), и Ю.А. Песенко (1982, с. 113), указывающим, что «... при более общем подходе следует признать правомерным одновременное существование нескольких индексов разнообразия, что математически связано с неоднозначностью мер плохо организованных множеств». Учитывая вышеизложенное, при оценке разнообразия флуктуирующих признаков мы основывались на классических представлениях теории вероятностей. Сформулируем задачу следующим образом.

Пусть имеем выборку из m объектов, оцениваемых по набору из n признаков, проявляющихся случайным образом в одном из двух альтернативных состояний: асимметричном или симметричном. Обозначим наличие асимметричного состояния 1, а симметричного – 0. Основой для дальнейшего анализа является таблица $\{B_{ij}\}$ с размерами $m \times n = N$, где m – число объектов, обладающих n признаками: $B_{ij} = \{b_{ij} = \{0,1\}, i = \overline{1,m}; j = \overline{1,n}\}$. Таблицу можно рассматривать как выборку из ансамбля сообщений с алфавитом из двух символов. При равной вероятности ($p = \text{const}$) символов 0 и 1 для каждого из n признаков в сообщении возможны варианты $M = 2^N$ их комбинаций (текстов), вероятность каждой из которых $p = M^{-1} = 2^{-N}$.

Мера неопределенности, соответствующая этому ансамблю, задается энтропией Р. Хартли (Shannon, 1948):

$$H = -\log_2 p = -\log_2 2^{-N} = \log_2 2^N = N \log_2 2 = N \text{ бит.}$$

Энтропия Р. Хартли задает максимально возможную неопределенность в системе из N ячеек для двухбуквенного алфавита, равную N бит, или 1 бит/символ.

Количество информации I , получаемой в результате каждого измерения n признаков объекта, равно величине неопределенности (энтропии) с обратным знаком: $I = -H$.

Если вероятности появления символов в сообщении не равновероятны, то для вычисления средней неопределенности в появлении j -го признака в состоянии 1 используется информационная энтропия К. Шеннона, которая в данном контексте может быть записана в следующем виде:

$$H_j = -\sum_{i=1}^m p_{ij} \log_2 p_{ij}.$$

Для получения оценки набора вероятностей, соответствующим вероятностям $\{p_j\}$, следует провести подсчет числа b_{ij}^0 и b_{ij}^1 появления признака в состояниях 0 и 1 по m строкам базовой таблицы для каждого из n признаков. Тогда вероятности появления этого признака в состояниях 0 и 1 вычисляются по формулам

$$p_j^0 = \frac{\sum_{i=1}^m b_{ij}^0}{m} \text{ и } p_j^1 = 1 - p_j^0 = \frac{\sum_{i=1}^m b_{ij}^1}{m}.$$

Неопределенность j -го признака, определяющая его информативность в указанном выше смысле, вычисляется по формуле

$$H_j = -(p_j^0 \log_2 p_j^0 + p_j^1 \log_2 p_j^1).$$

Неопределенность для набора из n признаков (при условии их независимости) равна сумме неопределенностей для каждого из этих признаков:

$$H = \sum_{j=1}^n H_j,$$

а среднее количество неопределенности (информативность на один признак)

$$\bar{H} = \frac{1}{n} \sum_{j=1}^n H_j.$$

Формулу К. Шеннона в этом случае следует трактовать как количественную

Таблица 1
Распределение асимметричных признаков
в модельной выборке

m	n		
	1	2	3
1	1	1	0
2	0	0	1
3	0	0	1
4	0	1	1

оценку средней неопределенности в реализации условия симметрии, которая описывается n признаками этой симметрии.

Пример. Пусть имеем таблицу 1 размером $m \times n = 4 \times 3$, ячейки которой заполнены случайным образом, так что 1 – асимметричное, а 0 – симметричное проявление признака.

Для признака n_1 $p_1^0 = \frac{3}{4}$ и $p_1^1 = \frac{1}{4}$; для n_2 $p_2^0 = \frac{1}{2}$ и $p_2^1 = \frac{1}{2}$; для n_3 $p_3^0 = \frac{1}{4}$ и $p_3^1 = \frac{3}{4}$.

$$H_1 = -\left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4}\right) = 0.8 \text{ бит}, H_2 = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right) = 1.0 \text{ бит},$$

$$H_3 = -\left(\frac{1}{4} \log_2 \frac{1}{4} + \frac{3}{4} \log_2 \frac{3}{4}\right) = 0.8 \text{ бит}, H = 0.8 + 1 + 0.8 = 2.6 \text{ бит},$$

$$\bar{H} = 2.6 : 3 = 0.87 \text{ бит/признак}.$$

В рамках информационной модели ФА следует интерпретировать исходя из основных свойств функции $H(\alpha) = -(p^0 \log_2 p^0 + p^1 \log_2 p^1)$, являющейся мерой неопределенности исхода опыта α .

Максимальное фенотипическое разнообразие достигается при равной вероятности представленности симметричных (p^0) и асимметричных (p^1) состояний признаков, т.е. $p^1 = p^0 = 1/2$, так как $p^1 + p^0 = 1$. Опыт, имеющий все исходы равновероятными, предсказать труднее всего. Поэтому он имеет наибольшую степень неопределенности, или энтропию:

$$h(p) = -p \log_2 p - (1-p) \log_2 (1-p) \text{ при } p^1 = p^0 = \frac{1}{2}, h(p) = \log_2 2 = 1.$$

Живые организмы характеризуются структурной упорядоченностью, но для утверждения, что упорядоченность одной системы выше, чем другой, нужно эти упорядоченности измерить. Такая измеряемая и выражаемая числом упорядоченности может относиться только к макросостояниям системы. В распоряжении физики пока имеется лишь одна величина, характеризующая степень упорядоченности – энтропия. Если ФА организмов по билатеральным признакам рассматривать как случайное макроскопическое событие, заключающееся в независимом прояв-

лении либо на правой, либо на левой, либо на обеих сторонах объекта, но в разной степени выраженных признаков, являющихся итогом стохастических микроскопических процессов, то подходы статистической физики будут вполне оправданы. Флуктуации, рассматриваемые в статистической физике – особый класс процессов, при которых система из более вероятного (менее упорядоченного состояния) переходит в менее вероятное (более упорядоченное состояние и ее энтропия уменьшается). Н. Винер (2003, с. 92) писал: «Описывая организм, мы не пытаемся точно определить в нем каждую молекулу ..., мы стремимся разрешить некоторые вопросы, раскрывающие форму строения организмов – форму строения, которая становится более значимой и менее вероятной по мере того, как организм становится, так сказать, более цельным».

В рассматриваемом контексте снижение уровня ФА как меры отклонения от идеальной билатеральной симметрии следует интерпретировать как поддержание или увеличение биологической упорядоченности, за что приходится «расплачиваться» переходом в менее вероятное состояние, оцениваемое уменьшением энтропии. Действительно, в рамках информационной модели ФА уменьшение вероятности встречаемости асимметричных признаков $p^1 \rightarrow 0$ в случае двухсимвольного «алфавита» влечет $p^0 \rightarrow 1$ и $H(\alpha) \rightarrow 0$. Отметим, что аналогичные рассуждения справедливы при уменьшении вероятности встречаемости симметричных признаков $p^0 \rightarrow 0$, тогда $p^1 \rightarrow 1$ и $H(\alpha) \rightarrow 0$. Эти рассуждения полностью основываются на характере зависимости функции $H(p_1, p_2, \dots, p_n)$ от отдельных вероятностей $p_1, p_2, \dots, p_n$ (Shannon, 1948).

Очевидно, что осуществление редкого события несёт гораздо больше информации, чем заурядного. Однако большое количество информации при многократном повторении опыта мы будем получать редко, поэтому среднее количество информации, содержащееся в одном исходе, оказывается здесь меньше, чем в случае, когда вероятности обоих исходов равны. Именно поэтому в практических задачах оценивается среднее количество информации, представление же о количестве информации, связанное с отдельными исходами опыта, почти никогда не употребляется (Яглом, Яглом, 1973).

Применительно к рассматриваемой информационной модели этот вывод означает «статистическую бесперспективность» оценки ФА одной особи по многим признакам или у многих особей по одному признаку, т.е. ограничение вектор-строкой $\{b_{ij}, j = \overline{1, n}\}$ или вектор-столбцом $\{b_{ij}, i = \overline{1, m}\}$. Для корректных популяционных оценок необходимы репрезентативные выборки, а при фиксированном наборе признаков, участвующих в анализе, объём выборки зависит от числа объектов (организмов). Кроме того, следует подчеркнуть, что максимально симметричный объект ($p^0 \rightarrow 1$), так же как и максимально асимметричный объект ($p^1 \rightarrow 1$), является «эволюционно не перспективным», поскольку не поставляет материал для естественного отбора. Информация как результат выбора порождается разнообразием: «Во вполне однородной популяции количество информации падает до нуля и регуляция её состава становится невозможной» (Шмальгаузен, 1968).

3. Информативность оценки степени асимметрии биообъектов разными алгоритмами

Возможность применения разных алгоритмов (Захаров и др., 2000; Гелашвили и др., 2004; Leary et al., 1983) для оценки степени асимметрии биообъектов приводит к закономерному вопросу об информативности этих подходов.

Рассмотрим этот вопрос на модельном примере. Пусть имеем таблицу значений n признаков на левой (L) и правой (R) сторонах у m объектов (табл. 2).

Таблица 2
Выраженность («экспрессивность») асимметричных признаков в модельной выборке

m	n			
	n_1		n_2	
	L	R	L	R
m_1	3	1	2	2
m_2	3	3	1	1

Оценку степени асимметрии в модельной выборке проведем с помощью алгоритмов, основанных на использовании формул, учитывающих только частоту проявления асимметричных признаков

$$\bar{A}_2 = \frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n b_{ij}, \quad (3)$$

их величину

$$\bar{A}_3 = \frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n \frac{|L_{ij} - R_{ij}|}{(L_{ij} + R_{ij})}, \quad (4)$$

а также свертки функций (2). Поскольку применение формулы (3) предусматривает учет асимметричных признаков в альтернативной форме, преобразуем данные табл. 2 и представим их в виде табл. 3.

Таблица 3
Частота проявления («пенетрантность») асимметричных признаков в модельной выборке

m	n	
	n_1	n_2
m_1	1	0
m_2	0	0

Из четырех учитываемых признаков асимметричен только один (признак n_1 у объекта m_1), следовательно, величина асимметрии, вычисленная по формуле (3), составит $\bar{A}_2 = 0.25$. Для оценки степени асимметрии по формулам (2) и (4) необходимо воспользоваться данными табл. 2, так как алгоритмы этих формул предусматривают оперирование значениями величин признаков на L - и R -сторонах. Очевидно, что степень асимметрии, вычисленная по формуле (4), составит

$$\bar{A}_3 = \frac{1}{4} \left(\frac{|3-1|}{3+1} + 0 + 0 + 0 \right) = 0.125,$$

а вычисленная по формуле (2) –

$$\bar{A}_1 = 1 - \frac{1}{2} \left[\left(\frac{2(3 \cdot 1 + 2 \cdot 2)}{3^2 + 1^2 + 2^2 + 2^2} \right) + \left(\frac{2(3 \cdot 3 + 1 \cdot 1)}{3^2 + 3^2 + 1^2 + 1^2} \right) \right] = 1 - \frac{1}{2} \left(\frac{14}{18} + \frac{20}{20} \right) = 1 - 0.888 = 0.112.$$

В рассматриваемом примере степень асимметрии системы из двух объектов, имеющих по два признака, определяется одним асимметричным признаком n_1 у объекта m_1 . Легко показать, что степень асимметрии самого признака n_1 (m_1), вычисленная по формулам (2), (3) и (4), при $m=1$ и $n=1$ составит соответственно

$\bar{A}_1 = 0.22$, $\bar{A}_2 = 1$ и $\bar{A}_3 = 0.5$. Таким образом, формула (3), не учитывающая выраженность («экспрессивность») асимметрии признака, как в случае оценки одного признака, так и набора признаков, дает наиболее грубую, приближенную оценку степени асимметрии. Интуитивно сделанный вывод понятен, поскольку сравниваемые алгоритмы с разной степенью подробности оценивают асимметрию. Следовательно, правомочен вопрос о количестве информации, получаемой при оценке степени асимметрии с помощью алгоритмов, учитывающих как частоту проявления асимметричного признака, так и его выраженность (величину). Очевидно, что чем более подробно мы оцениваем степень симметрии/асимметрии объекта, максимально полно используя исходные данные, тем большее количество информации об объекте можем получить.

Докажем это, основываясь на свойствах информационной энтропии (Файнштейн, 1960). Воспользуемся следующим условием: если $p_n = q_1 + q_2 > 0$, то

$$H(p_1, p_2, \dots, p_{n-1}, p_n) = H(p_1, p_2, \dots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right).$$

Для данных, приведенных в табл. 2, очевидно, что энтропия признака n_1 H_1 , равновероятно встречающегося в асимметричном и симметричном состояниях ($p_1^0 = p_1^1 = \frac{1}{2}$), равна 1 бит, тогда как для признака n_2 $H_2 = 0$. Таким образом, если

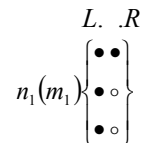
ограничиться только оценкой частоты проявления («пенетрантности») асимметричного признака, количество получаемой информации (I) в силу соотношения $I = -H$ составит 1 бит. Однако единственный асимметричный признак $n_1(m_1)$ определен значениями 3 (L) на левой стороне и 1 (R) – на правой. Другими словами, $n_1(m_1)$ расщеплен на одно симметричное состояние и два асимметричных состояния (рисунок) в соотношении 1/3 и 2/3.

Поскольку вероятность p_1^1 проявления признака $n_1(m_1)$ равна 1/2, то и расщепление p_1^1 необходимо произвести в указанном соотношении (Shannon, 1948):

$$p_1^1 = q_1 + q_2 = \frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.$$

Тогда $H = H_1 + \frac{1}{2} H_1 \left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{6} \log_2 \frac{1}{6} \right) = 1 \text{ бит} + \frac{1}{2} (0.48 + 0.41) \approx 1.445 \text{ бит}.$

Таким образом, как следует из рассмотренного примера, количество информации о степени симметричности / асимметричности исследуемого объекта при учете выраженности («экспрессивности») признака на ~ 0.5 бит больше, чем в случае учета только альтернативных состояний признака, т.е. «пенетрантности». При возможности выбора методологических концепций и методических приемов многое определяется «вкусом и пристрастиями» исследователя. Нам представляется, что привлечение в биоэкологические задачи теоретических положений и математического аппарата смежных естественно-



Расщепление признака $n_1(m_1)$ на симметричные и асимметричные состояния: $\bullet$ – наличие признака, $\circ$ – отсутствие признака

научных дисциплин, в данном случае – теории информации, кристаллографии, физики, дало возможность существенно продвинуться в разработке методов оценки степени симметричности природных объектов.

СПИСОК ЛИТЕРАТУРЫ

- Вейль Г. Симметрия. М.: Наука, 1968. 192 с.
- Винер Н. Кибернетика и общество. Творец и робот. М.: Тайдекс К°, 2003. 248 с.
- Гелашивили Д.Б., Краснов А.К., Логинов В.В., Мокров И.В., Радаев А.А., Силкин А.А., Слепов А.В., Чупрунов Е.В. Методологические и методические аспекты мониторинга здоровья среды государственного природного заповедника «Керженский» // Тр. Гос. природ. заповедника «Керженский». Н. Новгород, 2001. Т. 1. С. 287 – 325.
- Гелашивили Д.Б., Чупрунов Е.В., Иудин Д.И. Структурные и биоиндикационные аспекты флуктуирующей асимметрии билатерально симметричных организмов // Журн. общ. биологии. 2004. Т. 65, №4. С. 377 – 385.
- Гиляров М.С. О функциональном значении симметрии организмов // Зоол. журн. 1944. Т. 23, №5. С. 213 – 215.
- Животовский Л.А. Показатель сходства популяций по полиморфным признакам // Журн. общ. биологии. 1979. Т. 40, №4. С. 587 – 602.
- Животовский Л.А. Показатель внутрипопуляционного разнообразия // Журн. общ. биологии. 1980. Т. 41, №6. С. 828 – 836.
- Захаров В. М. Асимметрия животных. М.: Наука, 1987. 216 с.
- Захаров В.М. Онтогенез и популяция (стабильность развития и популяционная изменчивость) // Экология. 2001. №3. С. 177 – 191.
- Захаров В.М., Баранов А.С., Борисов В.И., Велецкий А.В., Кряжева Н.Г., Чистякова Е.Г., Чубинишвили А.Т. Здоровье среды: методика оценки. Оценка состояния природных популяций по стабильности развития: методологическое руководство для заповедников. М.: Центр экологической политики России, 2000. 66 с.
- Захаров В.М., Жданова Н.П., Кирик Е.Ф., Шкиль Ф.Н. Онтогенез и популяция: оценка стабильности развития в природных популяциях // Онтогенез. 2001. Т. 32, № 6. С. 404 – 421.
- Кожжара А.В. Структура показателя флуктуирующей асимметрии σ_d^2 и его пригодность для популяционных исследований // Биол. науки. 1985. № 6. С. 100 – 103.
- Маргалев Р. Облик биосферы. М.: Наука, 1992. 214 с.
- Марченко А.О. Реализация морфогенетического потенциала растительных организмов: калибровочный подход // Журн. общ. биологии. 1999. Т. 60, № 6. С. 654 – 666.
- Песенко Ю.А. Принципы и методы количественного анализа в фаунистических исследованиях. М.: Наука, 1982. 288 с.
- Радаев А.А. Биоэкологические закономерности стабильности развития пчелы медоносной *Apis mellifera* L. и их применение в биомониторинге: Автореф. дис. ... канд. биол. наук. Н. Новгород, 2001. 23 с.
- Урманцев Ю.А. Симметрия природы и природа симметрии. М.: Мысль, 1974. 228 с.
- Хохуткин И.М., Елькин Ю.А. Опыт применения бинарных отношений для оценки сходства биотических сообществ на примере наземных моллюсков // Фенетика популяций. М.: Наука, 1982. С. 125 – 133.
- Хэйек Л.-Э.С. Анализ данных по биоразнообразию земноводных // Измерение и мониторинг биологического разнообразия. Стандартные методы для земноводных. М.: КМК, 2003. 380 с.
- Чупрунов Е.В., Солдатов Е.А., Тархова Т.Н. О количественных оценках симметричности кристаллических структур // Кристаллография. 1988. Т. 33, № 3. С. 759 – 761.
- Чупрунов Е.В., Хохлов А.Ф., Фадеев М.А. Кристаллография. М.: Высш. шк., 2000. 496 с.

МЕРЫ СХОДСТВА И РАЗНООБРАЗИЯ

- Фанстейн А.* Основы теории информации. М.: Изд-во иностр. лит., 1960. 140 с.
- Шубников А.В., Кончик В.А.* Симметрия в науке и искусстве. М.: Наука, 1972. 340 с.
- Шмальгаузен И.И.* Кибернетические вопросы биологии. Новосибирск: Наука. Сиб. отд-ние, 1968. 222 с.
- Яглом А.М., Яглом И.М.* Вероятность и информация. М.: Наука, 1973. 512 с.
- Hill M.O.* Diversity and evenness: A unifying notation and its consequences // *Ecology*. 1973. Vol. 54. P. 427 – 432.
- Leary R.F., Allendorf F.W., Knudsen K.I.* Developmental stability and enzyme heterozygosity in rainbow trout // *Nature*. 1983. Vol. 54(2). P. 187 – 211.
- Palmer A.R., Strobeck C.* Fluctuating asymmetry: measurement, analysis, patterns // *Ann. Rev. Ecology and Systematics*. 1986. №17. P. 391 – 421.
- Palmer A.R., Strobeck C.* Fluctuating asymmetry analyse revisited // *Developmental instability (DI): cause and consequences*. Oxford: Oxford Univ. Press, 2003. P. 1 – 47.
- Rényi A.* On measures of entropy and information // 4th Berkeley symposium on mathematical statistics and probability / Ed. J. Neyman. Berkeley, 1961. P. 547 – 561.
- Shannon C.E.* The Mathematical Theory of Communication // *Bell System Technical J.* 1948. Vol. 27. July and October. P. 379 – 423, 623 – 656.