Емельянов И.Г., Межжерин В.А., Михалевич О.А. Методы интегральной оценки организмов // Вестн. зоологии. — 1986. — N 3. — с. 46—57.

Зайцев Г. Н. Математическая статистика в экспериментальной ботанике. — М.: Наука, 1984. — 424 с.

Лакин Г. Ф. Биометрия. М.: Высш. школа, — 1990. — 352 с.

Межжерин В. А., Кальниш В. В., Ищенко А. И. Единство фенотипических проявлений в генетически различных группах живых организмов // Докл. АН СССР. — 1975. — 225, N 1. — С. 205—206.

Межжерин В. А., Емельянов И. Г., Михалевич О. А. Комплексные подходы в изучении популяций мелких млекопитающих. — Киев: Наук. думка, 1991. — 204 с.

Плохинский Н. А. Биометрия. — Новосибирск: Изд-во Сибирск. отд-ния АН СССР, 1961. — 364 с.

Терентьев П. В. Метод корреляционных плеяд // Вестн. Ленингр. ун-та — 1959. — **9**, вып. 2. — С. 137—144.

Урбах В. Ю. Математическая статистика для биологов и медиков. — М.: Изд-во АН СССР, 1963. — 323 с.

Фролов И. Т. (ред.) Философский словарь. — М.: Политиздат, 1980. — 444 с.

Albrecht G. H. Some comments on the use of ratios // Syst. Zool.— 1978. — 27, N 1.— P. 67—71.

Atchley W. R., Anderson D. Ratios and the analysis of biological data // Ibid. — 1978.— 27, N 1.— P. 71—78.

Atchley W. R., Gaskins C. T., Anderson D. Statistical properties of ratios. I. Empirical results // Ibid.— 1976.— **25**, N 2.— P. 137—148.

Dodson P. On the use of ratios in growth studies // Ibid.— 1978.— 27, N 1.— P. 62—67.

Hills M. On ratios — A response to Atchley, Gaskins, and Anderson // Ibid. — P. 61—62.

Jackson D.A., Harvey H.H., Somerth K.M. Ratios in aquatic sciences: statistical shortcoming with mean depth and morphoedaphic index // Can. J. Fish. Aquat. Sci.— 1990.— **47**.— P. 1788—1795.

Pearson K. On a form of spurious correlation which may arise when indices are used in the measurement of organs. // Proc. R. Soc. London. — 1897. — **60**. — P. 489 — 502.

Phillips R.B. Shape characters in numerical taxonomy and problems with ratios // Taxon.— 1983.— 32.— P. 535—544.

Pimentel R.A. A comparative study of data and ordination techniques based on a hybrid swarm of sand verbenas (Abronia Juss.). // Syst. Zool. — 1981. — 30, N 3. — P. 250—267.

Somers K.M. Multivariate allometry and removal of size with principal components analysis // Systematic Zoology. — 1986. — **35**, N 3. — P. 359—368.

Somers K.M. Allometry, isometry and shape in principal components analysis // Ibid. — 1989. — **38**, N 2. — P. 169—173.

Thorpe R.S. Quantitative handling of characters useful in snake systematics with particular reference to intraspecific variation in the Ringed Snake Natrix natrix (L.) // Biol. J. Lin. Soc. — 1975.— 7.— P. 27—43.

Thorpe R.S. A comparative study of ordination techniques in numerical taxonomy in relation to racial variation in the ringed snake Natrix natrix (L.) // Ibid. — 1980. — **13**. — P. 7—40.

Институт зоологии НАН Украины
(252601 Киев)

Получено 23.02.94

УДК 591.5

Ю. В. Белкин

ФОРМУЛА ИСТИННОЙ ОЦЕНКИ РАЗНООБРАЗИЯ

Формула істинної оцінки різноманітності. Белкін Ю. В. — Запропонована формула розрахунку індексу різноманітності, яка долає труднощі вирішення існуючої проблеми. Вживання формули дозволяє уникнути методу апроксимації, який не вирішує з достатньою точністю розв'язання специфічних питань.

К л ю ч о в і с л о в а: максимальна (абсолютна) вирівнянність, порушена вирівнянність, індекс різноманітності, ряди розподілення.

On a Real Diversity Estimation Formula. Belkin Ju. V. — A formula of diversing index calculation is proposed to overcome existing difficulties in problem solution. The use of this formula provides to reject the approximative method which does not resolve specific questions with sufficient confidence.

K e y w o r l d s: maximal (absolute) equalization, infringement equalization, diversity index, distribution rows.

Еще в 1971 г. А.Хендриксон и П.Эрлих, анализируя основные подходы в оценке видового разнообразия (индексы Патрика и Хатчинсона, Маргалефа, Фишера, Макартура, Макинтоша) и предлагая в качестве более надежного средства модифицированный ими индекс Симпсона, отмечали неэффективность современных подходов в качестве точного количественного критерия для оценки стабильности сообщества и структуры экосистемы (Hendrickson, Ehrlich, 1971).

Анализируя особенности различных подходов к решению проблемы оценки разнообразия, их уязвимость, Ю.А.Песенко (1982), ссылаясь также на Е.Пиэлоу (Pielou, 1969) и С.Хэрлберта (Hurlbert,1971), доказывает, что все известные показатели видового разнообразия, будучи основаны на разных теоретических предпосылках и построены с использованием аппарата целого ряда математических направлений (математической статистики, теории вероятностей, теории информации, комбинаторики, теории множеств, геометрии многомерных пространств), по-разному выражают особенности структуры сообществ или отражают разные стороны разнообразия, что вызывает серьезную критику всей концепции разнообразия. И делает вывод, что существующие математические средства оценки разнообразия "имеют определенные недостатки и не подходят для широкого использования" (с.84-85). Ту же сентенцию (со ссылкой на многих авторов, критически относящихся к концепции разнообразия) высказывает и Б. Клауснитцер (1990).

О неблагополучном состоянии проблемы оценки разнообразия можно судить уже из факта многочисленности попыток ее решения. Среди причин, обусловивших такую ситуацию, можно выделить, по крайней мере, два обстоятельства, имевших негативное влияние на ход решения вопроса. Пожалуй, главная ошибка допущена в методологическом плане. Она нашла выражение в использовании идеологии аппроксимации, подрывающей принцип самодостаточности внутреннего закона рассматриваемого нами явления — в данном случае феномена разнообразия.

С другой стороны (и, конечно же, находясь под детерминирующим воздействием первой причины), ущербность подходов к проблематике разнообразия была усилена игнорированием того факта, что вычислительные формулы производили показатели, не ограниченные лимитами роста (или убывания). И это при всей очевидности несоответствия таких показателей системам с конечными параметрами, каковыми являются статистические выборки. Более того, диапазон вычисляемых значений разнообразия должен колебаться в стандартизованных лимитах — таких, как нуль и единица, — позволяющих получить точку отсчета для относительной оценки. Справедливости ради следует отметить, что в рамках метода аппроксимации преодоление этого затруднения является задачей не просто трудной, а, вероятно, и неразрешимой. В этой связи процитируем Ю. А. Песенко: "...разнообразие зависит от двух не связанных между собой свойств коллекции (читай: выборки — Ю. Б.) и двусмысленность неизбежна, так как коллекция с немногими видами и высокой выравненностью может иметь одинаковое разнообразие с другой коллекцией, содержащей много видов, но с низкой выравненностью их обилия". Именно отсутствие возможности разграничения влияний количества видов и степени выравненности обилий этих же видов в одном ряду распределения и, таким образом, придания формуле способности учитывать оба эти фактора соразмерно их роли не в отдельности, а во взаимосвязи, как раз и выросло в самостоятельную проблему.

С таким пониманием ситуации становится ясно, что должно существовать отношение двух величин одной природы, в каждой из которых взаимосвязанно отражались бы и количество видов и их обилие, а показатель разнообразия при этом не должен был бы реагировать ни на один из этих факторов отдельно. Одной из таких величин, естественно, должно быть выражение гипотетически возможного состояния выборки с количеством видов S и количеством особей N — такого, когда обилие видов было бы одинаковым: $n_1 = n_2 = ...n^i$ — речь идет о единственно возможном варианте распределения указанного числа особей среди указанного числа видов, определенных для выборки того же объема, то есть с теми же параметрами S и N. В этом случае показатели обилия всех видов будут равны средней по выборке: N/S. В таком ряде распределения разброс значений обилия видов (или же весов элементов, или же частот различных событий — для выборок иной физической природы) равен нулю, а выборка имеет максимальную выравненность значений обилия видов. В сравнение с таким гипотетическим рядом распределения, обладающим максимальной, абсолютной выравненностью, и должен вводиться реально существующий ряд распределения с нарушенной выравненностью.

Хотелось бы отметить, что такой подход в вопросе оценки разнообразия созвучен позиции Е.Фейджера (Fager, 1972), рассматривавшего абсолютную выравненность в качестве отправного условия в своих рассуждениях. Он стоял близко к решению обсуждаемой проблемы, однако прошел мимо статистической природы феномена разнообразия и остановился на попытке решения проблемы средствами комбинаторики, применив весьма остроумный прием, смысл которого состоит в установлении количества последовательных минимальных изменений распределения обилий в выборке, приводящих последнюю от состояния абсолютной выравненности к реально данной, т. е. с фактической степенью вырожденности этой выравненности. К сожалению, в этом ракурсе решить проблему не удалось.

Принципиально удовлетворяющий предъявляемым требованиям индекс разнообразия D (от лат. diversitas) может быть получен отношением E_a/E_v, где E_a (a от лат. absolutus) — абсолютная выравненность, E_v (v от лат. variatio — изменение, отклонение) — степень вырождения реального ряда распределения. Физический смысл этого отношения состоит в том, что происходит вычисление количества абсолютной выравненности, приходящейся на единицу вырожденности в реально существующем ряду распределения. Остается лишь символы E_a и E_v заменить математическими выражениями, благодаря которым формула обретает вычислительную функцию.

С учетом сказанного необходимо отразить однотипную взаимосвязь N и S как в числителе, так и в знаменателе. При простой подстановке значений N и S мы во всех случаях получаем в результате единицу (NS/NS). Подобного рода затруднение возникает при сравнении сумм отклонений от средней — отрицательных с одной и положительных с другой стороны, но оно преодолевается с помощью дисперсии. В нашем случае единственной возможностью преодоления возникшего препятствия является операция с квадратами значений обилия. В этом случае всегда будет иметь место различие между суммой квадратов значений обилия видов при абсолютной выравненности и того же рода суммой, подсчитанной для ряда распределения с нарушенной выравненностью. Это произойдет потому, что наименьшая сумма квадратов значений обилия всех видов в выборке будет наименьшей для ряда распределения с максимальной выравненностью. Из этой аксиомы следует, что иное распределение даст большую сумму квадратов значений обилия — тем большую, чем выше степень вырождения, т. е. чем больше разброс значений обилия видов. Для знаменателя таким выражением будет Σn_i^2. В числитель же следует поставить выражение ($n^2_m \cdot S$), чтобы отразить физический смысл ситуации, когда обилия всех видов равны между собой. В таком случае мы получим формулу, являющуюся мерой разнообразия:

$$D = (n_m^2 \cdot S) / \Sigma n_i^2,$$

где n_m есть средняя обилия для выборки.

Однако пользоваться этой формулой нельзя, поскольку установление нижнего порога колебания значений задается только числом видов. Это приводит к искажению масштаба интервалов между отдельными вычисленными значениями по мере приближения их к нижнему пределу, т. е. при больших степенях разброса данных. Кроме того к очень сильным искажениям приводит существование в выборке обилий со значением равным единице, которые после возведения в квадрат входят в диспропорцию с другими элементами, испытавшими на себе ту же процедуру. Как установлено опытным путем, вычисленный по этой формуле ряд значений для какого-то списка выборок, состоящих из одинаковых количеств видов и особей, но имеющих различные степени разброса данных, выведенный на график, показывает несвойственную нелинейность. Характер изменения крутизны графика по мере приближения к нижнему порогу, свидетельствует об утере формулой с определенного момента заложенной в нее функции. Для нормальной работы формулы в нее должен быть введен дополнительный механизм учета степени разброса данных, выравнивающий масштаб всего диапазона вычисляемых показателей. Эту роль эффективно выполняет отношение средней величины обилий, меньших средней для всей выборки к средней величине обилий, больших чем средняя по выборке, то есть

$$(\Sigma n_x : S_x) : (\Sigma n_y : S_y) = (\Sigma n_x \cdot S_y) : (\Sigma n_y \cdot S_x).$$

Еще раз вернемся к понятию абсолютной выравненности. Чтобы нарушить это состояние, необходимы изменения обилий по крайней мере двух элементов (видов) выборки, поскольку в конечном определенном множестве всякие перегруппировки имеют обратно пропорциональную зависимость - какой-то элемент (или группа таковых) может быть увеличен во столько раз, во сколько раз уменьшен другой элемент (или группа таковых). На первый взгляд может показаться, что степень разброса данных и, следовательно, индекс разнообразия зависят только от соотношения двух величин — с одной стороны, суммы обилий тех видов, чьи обилия оказались меньшими средней, и, с другой стороны, суммы обилий тех видов, обилия которых превышают среднюю. Однако очевидно, что может существовать неограниченное количество выборок, в которых одно и то же соотношение создается одними и теми же количествами соотносящихся элементов, доли которых, однако, различны в разных выборках. То есть имеет место различие в количестве элементов с обилиями, значения которых равны средней. Естественно, их роль в выборке игнорировать нельзя. Напрашивается решение этой проблемы — необходимость указания соответствующих долей от общего числа видов как для совокупности обилий, каждое из которых равно средней, так и для той совокупности, которая образует соотношение вышеупомянутых групп с обилиями меньшими и большими средней. Тогда мы получим конечное выражение следующего вида:

$$D = \sqrt{\frac{(n_m)^2 \cdot (S_x + S_y)}{\Sigma n_x^2 + \Sigma n_y^2}} \cdot \sqrt{\frac{\Sigma n_x \cdot S_y}{\Sigma n_y \cdot S_x}} \cdot \frac{S_x + S_y}{S} + \frac{S - S_x - S_y}{S},$$

где n_m — среднее обилие вида в выборке; S — общее число видов, S_x — число видов с обилиями, меньшими средней; S_y — число видов с обилием выше средней; n_i — элемент выборки (любой); n_x — элемент выборки с обилием ниже средней; n_y — элемент выборки с обилием выше средней.

Механизм работы этой вычислительной конструкции описан ниже. Как мы видим, формула состоит из двух плеч-слагаемых, действия которых равнозначны. Рассмотрим сначала первое (слева). Оно представлено тремя сомножителями, в первом из которых мы узнаем оговоренное выше отношение абсолютной гипотетической выравненности к выравненности нарушенной (данной). Но здесь этой зависимостью охвачены не все элементы (виды) выборки, а лишь те, обилия которых не равны средней. В знаменателе дроби собрана сумма квадратов обилий именно таких видов. Выражение n_m^2 (т. е. средней, возведенной в квадрат) повторено столько раз, сколько имеется элементов, обилия которых не равны средней. Таким образом на этом этапе мы исключаем из процесса вычислений все элементы n_m-группы — то есть те, значения обилия которых равны средней каждое в отдельности. Второй сомножитель выполняет очень важную корректирующую функцию, устанавливая соотношение между средними величинами обилия элементов x-группы и y-группы. Дело в том, что основная дробь (принципиальное отношение), как уже было отмечено выше, начинает плохо работать у нижнего предела значений вычисляемых показателей. Особенно сильные искажения результатов происходят при обработке рядов распределения, содержащих одно или даже несколько данных, обилия которых равны единице каждое. Оговариваемый корректирующий сомножитель, напротив, наиболее эффективно проявляет свое действие именно в таких экстремальных ситуациях и отдает приоритет в этом процессе основной дроби по мере удаления от низшего порога значений вычисляемых показателей. Совместное действие обоих сомножителей обеспечивает равномерность масштаба интервалов между различными вычисленными показателями по всему диапазону вычислений. И, наконец, третий сомножитель $(S_x + S_y)/S$ устанавливает долю от общего числа видов, для которой и производится вычисление в левом плече формуле. Следует заметить, что извлечение квадратного корня из основной дроби не случайно и предназначено для востановления отношения из дисперсного состояния и приведения его к общему для всей формулы масштабу. Сходную роль выполняет второй радикал.

Правое плечо состоит из одной дроби, также устанавливающей долю от общего числа видов, но для тех элементов выборки, обилия которых, каждого в отдельности, равны средней. Имплицитно здесь присутствует тот же, что и в левом плече, набор вычислительных элементов, но из-за равенства числителя и знаменателя в каждой такой дроби, вырабатывающих единицу, они опущены. Физический смысл второго слагаемого состоит в вычислении доли всех элементов (видов) выборки со значениями обилия, равными средней.

Предлагаемая формула вычисляет показатели в диапазоне $0<d\leq 1$, что позволяет сравнивать степени разнообразия неодинаковых по объему (общему количеству видов и полной сумме значений их обилия) выборок, или же различные структуры выборки одного объема. Кроме того, появляется возможность отображения показателей в процентном выражении.

Работа формулы проверена на большом количестве тестовых задач с нормальными и разнообразными экстремальными условиями. Удовлетворительные результаты получены по всему диапазону вычислений. Однако объем статьи не позволяет представить документацию процесса разработки данной методики. Более подробное освещение данной работы предполагается сделать в последующих публикациях.

Клауснитцер Б. Экология городской фауны. — М. : Мир, 1990. — С. 177—181.

Песенко Ю. А. Принципы и методы количественного анализа в фаунистических исследованиях. — М. : Наука, 1982.— С. 59—118.

Fager E.W. Diversity: a sampling study // Amer. Natur.— 1972. — **106**, N 949.— P. 293—310.

Hendrickson A. Jr., Ehrlich Paul R. An expanded concept of "species diversity" // Notulae Naturae.— 1971.— N 439. — P. 1—6.

Hurlbert S.H. The nonconcept of species diversity: A critique and alternative parameters // Ecology. — 1971.— **52**, N 4. — P.577—586.

Pielou E.C. An introduction to mathematical ecology. N.Y.; London : Wiley, 1969. — 286 p.

Институт зоологии НАН Украины
(252601 Киев)

Получено 26.04.94