2. МЕТОДОЛОГИЯ РАНГОВОГО АНАЛИЗА

2.1. Общее содержание рангового анализа

Начнем с уже достаточно устоявшегося определения. Ранговый анализ – метод исследования больших технических систем (инфраструктурных объектов, техноценозов), имеющий целью их статистический анализ, а также оптимизацию и полагающий в качестве основного критерия форму видовых и ранговых распределений [83,86,108]. Ранговый анализ как основной инструмент техноценологического подхода при исследовании систем определенного класса базируется на трех фундаментальных основаниях: технократическом подходе к окружающей реальности, восходящем к третьей научной картине мира; негауссовой математической статистике устойчивых безгранично делимых гиперболических распределений; началах термодинамики. Первому основанию рангового анализа была посвящена предыдущая глава. Начала термодинамики лежат в основе закона оптимального построения техноценозов, который будет рассмотрен в третьей главе. Здесь мы основное внимание уделим собственно методологии рангового анализа и его процедурам в сравнительно доступном изложении.

Центром третьей научной картины мира является понятие, дополняющее принципиально новым стратификационным уровнем онтологическое описание окружающей реальности. Это техноценоз (объект техноценологического типа, инфраструктурный объект), главной отличительной чертой которого является специфика связей между техническими элементами-особями. В техноценозах сегодня видится прообраз будущей техносферы, которая по сложности организации и скорости эволюции превзойдет порождающую ее биологическую реальность [98,99,108].

Повторимся и вспомним, что техноценоз – ограниченная в пространстве и времени взаимосвязанная совокупность далее неделимых технических изделий-особей, объединенных слабыми связями. Связи в техноценозе носят особый характер, определяемый конструктивной, а зачастую и технологической независимостью отдельных технических изделий и многообразием решаемых задач; взаимосвязанность техноценоза определяется единством конечной цели, достигаемой с помощью общих систем управления, всестороннего обеспечения и др. [83,86,108,197].

Специфика техноценозов, в том числе, проявляется в методологических основаниях их исследования. Техноценозы не поддаются описанию ни традиционными методами гауссовой математической статистики, оперирующей понятиями среднего и дисперсии как информативно насыщенными свертками больших массивов статистической информации, ни лежащими в основе редукционизма имитационными моделями. Чтобы корректно описать техноценоз, необходимо постоянно оперировать выборкой в целом, как бы велика она ни была, что предполагает построение видовых и ранговых распределений, теоретическая основа которых лежит в области негауссовой математической статистики устойчивых гиперболических безгранично делимых распределений [108,120,185,197,377]. Методология построения видовых и ранговых распределений, а также их последующего использования в целях оптимизации техноценоза составляет основное содержание рангового анализа, который представляет собой новое фундаментальное научное направление, сулящее большие результаты.

В основе рангового анализа лежит весьма сложный математический аппарат. Однако, как и в любой фундаментальной теории, здесь имеется определенный вполне доступный уровень решения задач, фактически граничащий с инженерной методологией. Глубокая теоретическая проработка, всестороннее философское осмысление и многократное апробирование на практике в самых различных областях человеческой деятельности позволяют считать ранговый анализ вполне надежным средством решения задач определенного класса [75-119,148,180,196-214,229,370].

Как представляется, ранговый анализ, позволяя решать задачи оптимального построения техноценозов, занимает своего рода промежуточное положение между имитационным моделированием, с помощью которого осуществляется эффективное проектирование отдельных видов техники (пространственно-технологических кластеров), и методологией исследования операций, применяемой в настоящее время для решения проблем геополитического и макроэкономического планирования (рис. 2.1).

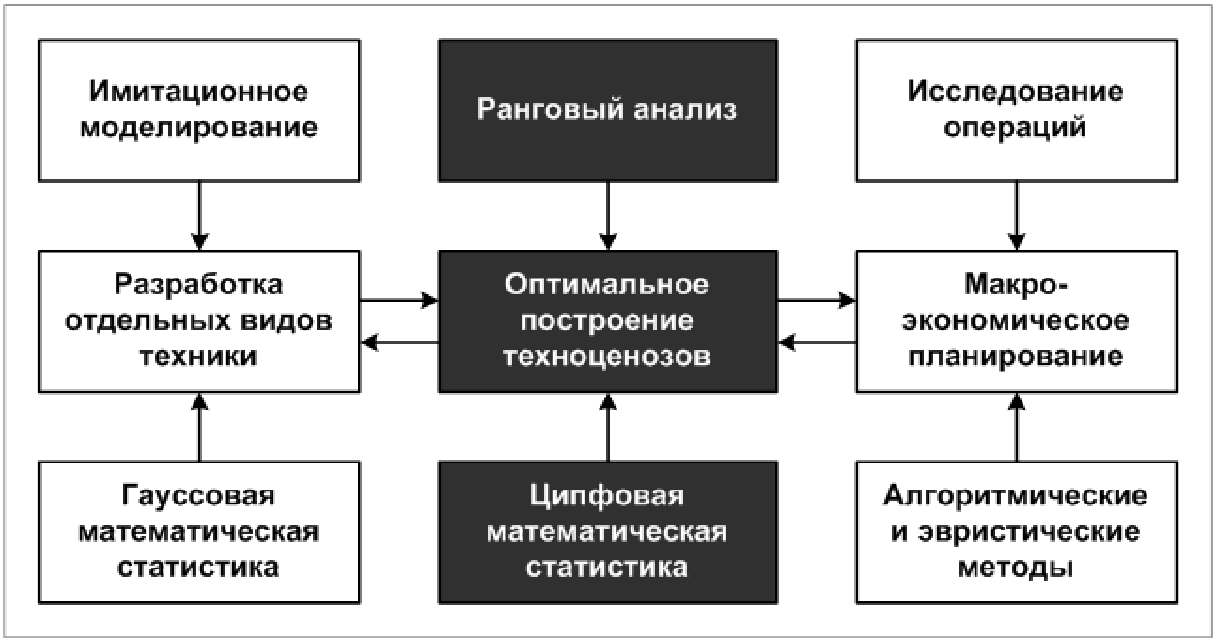

Рис. 2.1. Место рангового анализа в общей методологии решения технических задач

Представляется важным отметить ряд моментов. Во-первых, отсутствие достаточно глубоко разработанной специальной математической методологии делает аппарат исследования операций весьма ненадежным при решении задач соответствующего макроуровня и приводит, с одной стороны, к многочисленным безрезультатным попыткам применения методов имитационного моделирования в сфере геополитики и макроэкономики, а с другой – порождает глубокое недоверие к данной методологии со стороны подавляющего большинства практиков, которые до сих пор предпочитают полагаться в этих вопросах исключительно на свою интуицию.

Во-вторых, все попытки выдвигать требования, основанные на макропрогнозах и эвристических оценках, непосредственно разработчикам отдельных видов техники, равно как и политика последних, заключающаяся в игнорировании геополитических и макроэкономических процессов, с одинаковым успехом приводят к провалу. Думается, именно техноценологическая методология может разрешить проблему органической связи между крайними уровнями современных технических задач (рис. 2.1).

На первом уровне решения инженерных задач осуществляется разработка отдельных видов техники (пространственно-технологических кластеров). Как уже сказано, в качестве основного метода исследования здесь используется имитационное моделирование, базирующееся на классических постулатах гауссовой математической статистики [2,6,10,43,44,51]. Основным критерием, которым руководствуется проектировщик, в конечном итоге является достижение максимального положительного эффекта при минимальных затратах. Формально данный критерий не вызывает сомнений, т.к. полностью соответствует философскому толкованию полезности технического изделия, восходящему к упомянутому выше аристотелевскому «минимаксу». Проблема заключается лишь в том, что здесь закладывается в понятие «положительный эффект» и что – в «затраты».

Подавляющее большинство разработчиков техники трактует эти понятия в узком смысле как некие интегральные параметры, рассчитываемые без глубокого учета того, что произойдет после попытки внедрения спроектированного технического изделия в инфраструктуру техноценоза. В результате может оказаться, что новое техническое изделие вполне хорошо, будучи рассмотрено и испытано как совершенно независимый образец техники. Однако последующие попытки его внедрения в конкретную инфраструктуру могут закончиться полным провалом из-за невозможности адекватного обеспечения жизненного цикла системами управления, восстановления, снабжения, подготовки кадров, утилизации и т.д.

На высшем уровне геополитического и макроэкономического планирования и прогнозирования (рис. 2.1) решения принимаются на основе эвристических и алгоритмических процедур, базирующихся в основном на методологи исследования операций и квалиметрии [6,11,36,165]. Очевидно, что этот уровень является системным по отношению к первому, на котором проектируются отдельные виды технических изделий. Однако «расстояние» между ними настолько велико, что трудно говорить о какой-либо корректной методологии, позволяющей не на словах, а на деле учитывать геополитические интересы при проектировании или модернизации отдельного вида техники. Или, наоборот, в процессе принятия геополитических решений в какой-либо отрасли экономики учитывать параметры техники, представляющей данную отрасль на рынке. Ясно, что подобной методологии собственно на первом и третьем уровнях нет, если конечно мы говорим о научной методологии в полном смысле этого слова.

Таким образом, для решения сформулированных выше задач в области исследования технических систем имеется так называемый средний связующий уровень (рис. 2.1). Здесь применяется специфическая методология, основанная на философском техноценологическом подходе, алгоритмических процедурах рангового анализа и негауссовой (ципфовой) математической статистике гиперболических безгранично делимых распределений. Важнейшей задачей, которая решается на данном уровне, является оптимальное построение техноценозов. В основе же методологии, применяемой при решении данной задачи, лежит ранговый анализ. Рассмотрим его ключевые понятия, математические основы и содержание.

В первую очередь в методологии рангового анализа надо остановиться на понятии распределения. В самом общем случае распределение – это расположение элементов подмножества внутри множества [2,83,108]. В математике рассматриваются статистические и вероятностные распределения. Как правило, исследователь начинает работу с построения статистического распределения, которое возникает при эмпирическом описании выборки конечного объема из генеральной совокупности. Следовательно, оно дискретно на множестве значений случайной величины. Как идеализация статистического распределения в ситуации, когда объем выборки из генеральной совокупности стремится к бесконечности, возникает вероятностное распределение, которое в общем случае является непрерывным на множестве значений случайной величины [108,317,325].

Вторым ключевым моментом является понятие случайной величины, которое базируется на общем представлении о случайности. В современной литературе различают семь причин случайности: 1) непонятая закономерность; 2) скрещение несогласованных процессов; 3) уникальность; 4) неустойчивость движения; 5) относительность знания; 6) имманентная случайность; 7) произвольный выбор [388-391]. Нам представляется, что при исследовании объектов техноценологического типа мы, в той или иной степени, имеем дело с причинами пятого и седьмого типов [83,86,108].

Во-первых, насыщение техноценозов изделиями-особями происходит в условиях одновременного воздействия огромного количества слабосогласованных внешних и внутренних факторов, что делает случайной его номенклатуру или видовую структуру. Также доказано, что видообразование в техноценозе фрактально, а его границы размыты, конвенционны. Кроме того, техноценоз постоянно изменяется во времени, причем это изменение векторизовано и необратимо (однонаправленно). Данные феномены ранее широко обсуждались в литературе [197-199,202,203,206]. Сказано об этом достаточно и здесь в первой главе. Следовательно, можно говорить, что в данный фиксированный момент времени номенклатура техноценоза является случайной. И если описать номенклатуру частотным распределением, то форма последнего будет случайной (его параметры будут случайными величинами). Здесь мы имеем дело в полном смысле этого слова с проявлением трансцендентности техноценозов [99,108,210,211], делающей наши знания относительными, что является фундаментальной причиной случайности (в классификации [388-391] – пятого типа).

Во-вторых, совокупность параметров, описывающих особи техноценоза, составляет двумерное пространство. Оба измерения данного пространства бесконечны, однако, одно из них счетно (перечисляющее особи техноценоза), а второе – континуально (описывающее параметры). Это является следствием другого известного свойства техноценозов, а именно того, что число особей в них бесконечно (точнее, математически счетно) [83,86,108,211-214,243,269]. Кроме того, общее параметрическое пространство делится на два равномощных подпространства: видообразующих и функциональных параметров (об этом подробнее речь будет идти в третьей главе). В любом случае, если осуществлять произвольный выбор особей техноценоза, то параметры выбранных технических изделий составят статистическую выборку случайных величин. Если учесть, что техноценоз трансцендентен, то выбор особей при этом может осуществляться как угодно. Очевидно, что любой выбор из трансцендентной бесконечности будет произвольным и, по сути, случайным (причина седьмого типа). Если полученную выборку обрабатывать методами математической статистики, то можно получить параметрическое распределение.

Таким образом, случайным в широком смысле является сочетание видов, составляющих техноценоз, если мы его рассматриваем среди большого количества других техноценозов. Судить о статистическом (и далее – вероятностном) распределении данных сочетаний можно лишь исследовав поведение техноценозов в более общем таксономическом образовании – метаценозе (доступной для исследования в данный момент времени совокупности техноценозов). В узком смысле случайной является форма видового распределения, описывающего номенклатуру техноценоза, что делает случайной величиной значение соответствующего формального параметра. С другой стороны, если рассматривать совокупность одноименных параметров технических изделий (особей) техноценоза как выборку из параметрического пространства, то значение фиксированного параметра конкретного изделия может рассматриваться как случайная величина, а саму выборку можно описать как статистическое распределение.

Ввиду важности, следует еще раз подчеркнуть принципиальную разницу между видовыми и ранговыми распределениями техноценозов. Видовые распределения случайны в том смысле, что случайны макроскопические параметры их формы. Ранговые же распределения – это распределения случайных величин (параметров, характеризующих особи или объекты). Именно в этом смысле мы и применяем к техноценозам понятие статистического распределения. Последующее особое теоретическое обобщение на континууме техноценоза (эта особенность будет показана ниже) позволяет получать распределение, имеющее смысл вероятностного.

Третьим ключевым моментом в методологии рангового анализа являются понятия негауссовости и ципфовости описывающих техноценозы гиперболических распределений. Как всегда, начнем с определений.

Вероятностное распределение мы называем гауссовым, если для него выполняется центральная предельная теорема: при широких предположениях относительно законов распределения независимых случайных величин с ростом числа слагаемых закон распределения суммы этих величин неограниченно приближается к нормальному. Статистическое распределение называется гауссовым, если зависимость его среднего и дисперсии от объема выборки несущественна, т.е. в условиях данной конкретной исследовательской задачи выполняется закон больших чисел: при достаточно большом числе независимых испытаний среднее арифметическое наблюденных значений случайной величины сходится по вероятности к ее математическому ожиданию [83,86,108,377,383,384]. Очевидно, что, в общем случае, любое распределение, для которого не выполняется хотя бы одно из приведенных выше двух условий, является негауссовым.

Ципфовым мы называем распределение, имеющее при больших значениях переменной вид распределения Ципфа [377,383]:

$$f(x) = \frac{C}{x^{1+\alpha}}, \text{ при } x \geq x_0 > 0 \text{ и } 0 < \alpha < \infty, \qquad (2.1)$$

где $f(x)$ – частота;

C, α – параметры распределения.

Распределение Ципфа ципфово, ципфовое же распределение в общем случае не является распределением Ципфа. Вероятностное ципфовое распределение гауссово при значениях показателя распределения $\alpha \geq 2$ и негауссово при $\alpha < 2$. Статистическое ципфовое распределение с $\alpha > 2$ может быть негауссовым, если зависимость его среднего и дисперсии от объема выборки существенна в рамках данной конкретной задачи [377].

Видовые и ранговые распределения техноценозов относятся к классу так называемых безгранично делимых распределений [120]. В общем случае распределение вероятностей случайной величины X в вероятностном пространстве R^m называется безгранично делимым, если для всякого k можно указать такое распределение x_k, что X представимо в виде k-кратной свертки распределения x_k самого с собой. Безгранично делимые распределения могут быть как устойчивыми, так и неустойчивыми. Говорить, в приложении к техноценозам, о неустойчивых распределениях смысла нет, т.к. последние не предполагают вообще какой-либо фиксированной аппроксимационной формы [83,86,108,120,189,377].

К настоящему времени на обширном эмпирическом материале в различных областях многократно показана одновременно устойчивость и негауссовость ранговых распределений техноценозов [75-119,148,180,196-214,229,370]. Следовательно, для их статистического описания особое значение имеет распределение Ципфа с $\alpha < 2$, которое удовлетворяет предельной теореме Гнеденко – Деблина: для сходимости распределений нормированных сумм одинаково распределенных независимых случайных величин к устойчивым распределениям, отличным от нормального, необходимо и достаточно, чтобы при $x \to \infty$ имело место [120,377]:

$$f(-x) \sim C_1 \frac{h_1(x)}{|x|^\alpha} \text{ и } 1 - f(x) \sim C_2 \frac{h_2(x)}{x^\alpha}, \tag{2.2}$$

где $C_1 \ge 0$, $C_2 \ge 0$, $C_1 + C_2 > 0$ и $0 < \alpha < 2$;

$h_i(x)$ – функции, медленно меняющиеся в смысле Карамата, т.е. такие, что для всех $t > 0$ имеет место равенство:

$$\lim_{x \to \infty} \frac{h_i(tx)}{h_i(x)} = 1.$$

Распределение Ципфа, как и любое другое, имеет частотную и ранговую формы [51,83,86,108,120,377]. Как мы увидим позже, для распределений техноценоза актуальны обе формы. В частотной форме, как правило, представляются видовые распределения, в ранговой – ранговые видовые и ранговые параметрические (по видообразующим или функциональным параметрам). Частотная дифференциальная форма вероятностного распределения Ципфа определяется приведенным выше выражением (2.1). Частотная интегральная его форма для выборки объемом N:

$$F(x) = \frac{C}{\alpha N}\left(\frac{1}{x_0^\alpha} - \frac{1}{x^\alpha}\right) \cong 1 - \frac{x_0^\alpha}{x^\alpha}. \tag{2.3}$$

Ранговая дифференциальная форма вероятностного распределения с параметрами A, B, β и рангом r имеет следующий вид:

$$\varphi(r) = \frac{A}{(r+B)^{\beta}}, \tag{2.4}$$

где B – параметр, учитывающий эффект рангового искажения распределения Хайтуна [83,86,108,377].

Ранговая интегральная форма вероятностного распределения:

$$\Phi(r) = \begin{cases} A \ln \dfrac{r+B}{1+B}, \ \beta = 1; \\ \dfrac{A}{\beta - 1}\left(\dfrac{1}{(1+B)^{\beta-1}} - \dfrac{1}{(r+B)^{\beta-1}} \right), \ \beta \neq 1. \end{cases} \tag{2.5}$$

Четыре параметра распределения Ципфа можно поставить в соответствие параметру α следующим образом [83,86,108,377]:

$$\begin{cases} \beta = \dfrac{1}{\alpha}; \\ C = \dfrac{\alpha(N-1)}{1/x_0^{\alpha} - 1/J^{\alpha}}; \\ A = \left(\dfrac{N-1}{1/x_0^{\alpha} - 1/J^{\alpha}} \right)^{1/\alpha}; \\ B = \dfrac{N-1}{(J/x_0)^{\alpha} - 1} - 1, \end{cases} \tag{2.6}$$

где x_0 – минимальное значение случайной величины на исследуемой эмпирической выборке техноценоза;

J – максимальное значение случайной величины.

Система (2.6) позволяет учесть эффект рангового искажения, который подробно описан в [377]. При этом один из пяти параметров распределения Ципфа α, β, A, B, C (в случае применения системы (2.6) – это, как правило, α) должен быть определен априорно по имеющимся эмпирическим данным методом максимума правдоподобия либо графически.

Теперь мы готовы дать определения, распространяющие ципфовую методологию на прикладную область рангового анализа [83,86,108].

Под видовым понимается распределение Ципфа в частотной дифференциальной форме, устанавливающее непрерывную или дискретную упорядоченную взаимосвязь между множеством значений возможной численности особей техноценоза и количеством популяций, реально представленных в техноценозе данной фиксированной численностью. По сути, видовое распределение устанавливает основополагающую взаимосвязь между массовостью изделий различных видов в техноценозе и их разнообразием. Математически оно относится к гиперболическим устойчивым безгранично делимым распределениям [83,86,108,120,189,377].

В математической статистике под ранговым распределением вообще понимается убывающая последовательность значений параметров, упорядоченная таким образом, что каждое последующее число меньше предыдущего, и поставленная в соответствие рангу (номеру по порядку в данной упорядоченной последовательности). Таким образом, неотъемлемой чертой рангового распределения является целенаправленное ранжирование входящих в него параметров. В этом его коренное отличие от видового распределения, где подобная операция не предусмотрена.

Ранговое распределение техноценоза – полученное в результате процедуры ранжирования видов или особей техноценоза по какому-либо параметру распределение Ципфа в ранговой дифференциальной форме, по сути являющееся невозрастающей последовательностью значений самих параметров, поставленных в соответствие рангу. Различают ранговые распределения, в которых ранжируются: виды по количеству особей, которым они представлены в техноценозе (ранговые видовые); особи по значению видообразующего параметра (ранговые параметрические); особи или объекты по значению параметра, характеризующего процесс их функционирования (ранговые функциональные) [83,86,108,197,377].

Изначально каждое распределение техноценоза в аналитической или графической форме представляет собой совокупность точек, получаемых по имеющимся эмпирическим данным:

$$(x_1, y_1); (x_2, y_2); ...; (x_i, y_i); ...; (x_n, y_n), \qquad (2.7)$$

где i – формальный индекс;
n – общее количество точек.

Точки – результат анализа табулированного рангового распределения техноценоза (о нем подробно будет сказано в следующем параграфе). Для каждого из распределений имеется свое число точек (что есть абсцисса в распределении, а что ордината, мы уже знаем). С точки зрения последующей оптимизации техноценоза, большое значение имеет аппроксимация эмпирических ранговых и видовых распределений, которая в данном случае обладает существенной онтологической спецификой. Учитывая, что в процессе аппроксимации мы фактически без изменения формы обобщаем конечную выборку эмпирических точек техноценоза до континуума генеральной совокупности, можно заключить, что аппроксимационная форма – это и есть соответствующее вероятностное распределение техноценоза.

Формально задача аппроксимации заключается в подборе аналитической зависимости, наилучшим образом описывающей совокупность точек (2.7). Мы задаем в качестве стандартной формы двухпараметрическое гиперболическое аналитическое выражение вида [83,86,108,197]:

$$y(x) = \frac{A}{x^{\alpha}}, \tag{2.8}$$

где A, α – параметры распределения;

x – мощность популяции видового распределения и непрерывный ранг рангового распределения.

Выбор формы (2.8) объясняется традиционно сложившимся подходом среди исследователей, занимающихся ранговым анализом. Безусловно, данная форма далеко не самая совершенная и не учитывает эффект рангового искажения [377], однако она обладает неоспоримым достоинством – сводит задачу аппроксимации к определению всего двух параметров: A, α. Решается эта задача различными методами. В данном случае рассмотрим аппроксимацию методом наименьших квадратов [83].

Суть метода заключается в отыскании таких параметров аналитической зависимости (2.8) A, α, которые минимизируют сумму квадратов отклонений реально полученных эмпирических значений y_i от значений, рассчитанных по аппроксимационной зависимости (2.8), т.е.:

$$S = \sum_{i=1}^{n} (y_i - y(x_i))^2 \to \min. \tag{2.9}$$

Известно, что решение оптимизационной задачи (2.9) сводится к решению следующей системы дифференциальных уравнений [83,86,108]:

$$\begin{cases} \dfrac{\partial S}{\partial A} = 0; \\ \dfrac{\partial S}{\partial \alpha} = 0. \end{cases} \tag{2.10}$$

Для того чтобы перейти от системы дифференциальных уравнений к алгебраическим, прологарифмируем выражение (2.8) и получим:

$$Y(X) = CX + D, \tag{2.11}$$

где $Y = \ln y;\ X = \ln x;$

$C = -\alpha;\ D = \ln A.$

При этом (2.9) сводится к следующему условию:

$$Z = \sum_{i=1}^{n}(Y_i - Y(X_i))^2 \to \min, \tag{2.12}$$

а (2.10) – к системе дифференциальных уравнений:

$$\begin{cases} \dfrac{\partial Z}{\partial D} = 0; \\ \dfrac{\partial Z}{\partial C} = 0 \end{cases} \tag{2.13}$$

и далее, после дифференцирования, к системе:

$$\begin{cases} C\sum_{i=1}^{n} X_i^2 + D\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i Y_i; \\ C\sum_{i=1}^{n} X_i + Dn = \sum_{i=1}^{n} Y_i. \end{cases} \tag{2.14}$$

Параметры в (2.14) определяем по правилу Крамера [189]:

$$C = \frac{\Delta_1}{\Delta_0}; \tag{2.15}$$

$$D = \frac{\Delta_2}{\Delta_0},$$

где определители имеют вид:

$$\Delta_0 = \begin{vmatrix} \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & n \end{vmatrix};$$

$$\Delta_1 = \begin{vmatrix} \sum_{i=1}^{n} X_i Y_i & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} Y_i & n \end{vmatrix}; \qquad \Delta_2 = \begin{vmatrix} \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i Y_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} Y_i \end{vmatrix}.$$

После определения C и D вычислить параметры исходной зависимости (2.8) не составляет труда. При этом можно записать:

$$A = e^{D} \text{ и } \alpha = -C. \tag{2.16}$$

Следует отметить, что здесь весьма подробно разобран лишь метод наименьших квадратов. Имеется целый ряд других методов аппроксимации данных. В любом случае после аппроксимации мы получаем двухпараметрическую зависимость вида (2.8) для каждого из ранговых и видовых распределений. Обычно на графиках ранговых распределений, наряду с эмпирическими точками, показывают аппроксимационные кривые. Это имеет существенное значение с точки зрения последующей оптимизации техноценоза, т.к. применяемые в оптимизационных процедурах критерии во многом строятся на требованиях, предъявляемых к форме гиперболических распределений. Технологически, в конечном итоге, это выливается в систему ограничений к параметрам аппроксимационных кривых.

Практическая реализация рангового анализа состоит в осуществлении следующих этапов-процедур: 1) выделение техноценоза; 2) определение перечня видов в техноценозе; 3) задание функциональных параметров; 4) параметрическое описание техноценоза; 5) построение табулированного рангового распределения; 6) построение графического рангового видового распределения; 7) построение графических ранговых параметрических распределений; 8) построение видового распределения; 9) аппроксимация распределений; 10) оптимизация техноценоза [83,86,108].

В заключение еще раз подчеркнем, что для негауссовых по своей природе видовых и ранговых распределений нет никакого смысла вычислять эмпирические среднее и стандарт, а также выводить теоретические моменты (математическое ожидание и дисперсию). Ранговый анализ предполагает оперирование распределениями, будучи взятыми в целом.