

АКАДЕМИЯ НАУК СССР

МЕТОДЫ  
ПАРАЗИТОЛОГИЧЕСКИХ  
ИССЛЕДОВАНИЙ

Н. А. БРЕЕВ

ПРИМЕНЕНИЕ НЕГАТИВНОГО  
БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ  
ДЛЯ ИЗУЧЕНИЯ ПОПУЛЯЦИОННОЙ  
ЭКОЛОГИИ ПАРАЗИТОВ

prof. Ильинский

АКАДЕМИЯ НАУК СССР  
НАУЧНЫЙ СОВЕТ ПО ПРОБЛЕМЕ  
«БИОЛОГИЧЕСКИЕ ОСНОВЫ ОСВОЕНИЯ, РЕКОНСТРУКЦИИ  
И ОХРАНЫ ЖИВОТНОГО МИРА»

6 МЕТОДЫ ПАРАЗИТОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

К. А. БРЕЕВ

ПРИМЕНЕНИЕ НЕГАТИВНОГО  
БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ  
ДЛЯ ИЗУЧЕНИЯ ПОПУЛЯЦИОННОЙ  
ЭКОЛОГИИ ПАРАЗИТОВ



ИЗДАТЕЛЬСТВО «НАУКА»  
ЛЕНИНГРАДСКОЕ ОТДЕЛЕНИЕ  
Ленинград · 1972

Применение негативного биномиального распределения для изучения популяционной экологии паразитов. Бреев К. А.  
В серии: Методы паразитологических исследований, вып. 6. 1972.  
Изд-во «Наука», Ленингр. отд., Л. 1—72.

В основу пособия положено представление о численности вида как случайной величине. Даны элементарные сведения о дискретных случайных величинах и их основных параметрах — математическом ожидании, дисперсии и распределении. На конкретных паразитологических примерах разобраны три типа распределений: биномиальное, Пуассона и негативное биномиальное, последнее наиболее детально. Показано, какую новую биологическую информацию можно получить, анализируя распределения паразитов: проверка гипотез о вероятностных характеристиках процессов заражения популяции хозяина определенным видом паразита, установление функциональной связи между долей зараженных особей в популяции хозяина и средней численностью паразита, оценка влияния хозяино-паразитных отношений на уровне численности популяций паразита. Илл. — 1, табл. — 30, библ. — 27 назв.

Ответственный редактор

А. С. Мончадский

2-10-6

831-72

#### О Т Р Е Д А К Ц И И

Специальное пособие, посвященное методам паразитологических исследований, было задумано академиком Е. Н. Павловским задолго до его кончины. Ряд обстоятельств помешал завершению этого пособия, хотя потребность в нем в научных и научно-практических учреждениях общепаразитологического, медицинского и ветеринарного направлений и на соответствующих кафедрах высших учебных заведений была и остается очень большой. Выход в свет таких книг, как «Методы изучения природных очагов болезней человека» (изд. «Медицина», 1964), не восполняет этого пробела прежде всего вследствие их узкой специализации.

При исследовании паразитических животных, особенно конкретного медицинского или ветеринарного значения, последовательный зоологический и общебиологический подход и применение соответствующих методов способствуют более полному и быстрому получению важнейших материалов для мероприятий по борьбе с этими паразитами и по ликвидации приносимого ими вреда. Владение методами исследования зоологического и биологического аспектов паразитологических вопросов весьма существенно, что вполне оправдывает выход в свет задуманного акад. Е. Н. Павловским издания.

В условиях полевой работы удобнее пользоваться не объемистым томом, а отдельными выпусками, посвященными сбору, исследованию, лабораторному содержанию и разведению отдельных групп паразитов или их комплексов, объединяемых по группам хозяев или по биоценологической общности.

Первым выпуском этой серии вышло «Паразитологическое исследование рыб» (И. Е. Быховская-Павловская). Вслед за ним были изданы «Методы сбора, исследования и лабораторного

разведения блох и их личинок» (С. О. Высоцкая и А. Н. Кирьякова), «Методы сбора и исследования мокрецов и их личинок» (А. В. Гуцевич и В. М. Глухова), «Паразитологическое исследование птиц» (М. Н. Дубинина), «Методы исследования вшей» (Д. И. Благовещенский). Будет опубликовано «Сбор обитателей гнезд и убежищ мелких млекопитающих» (С. О. Высоцкая и М. К. Даниел) и др.

В последние годы в экологии животных все большее внимание уделяется количественному подходу к управлению природными ресурсами, для которого характерно построение математических моделей взаимодействия факторов, определяющих уровни и изменения численности отдельных видов в экологических системах, и изыскание принципов оптимального регулирования этих сложных процессов. Мы надеемся, что издание пособия «Применение негативного биномиального распределения для изучения популяционной экологии паразитов» (К. А. Бреев), посвященного некоторым частным аспектам количественного исследования хозяино-паразитных отношений на популяционном уровне, будет способствовать развитию такого рода исследований и в паразитологии.

Эту серию авторы посвящают памяти академика Е. Н. Павловского.

## 1. ВВЕДЕНИЕ

Количественные методы в экологии животных получают за последние десятилетия все более широкое и разностороннее развитие. Объясняется это все большей концентрацией внимания и труда экологов к факторам и процессам, определяющим средние уровни и динамику численности популяций различных видов животных. Как справедливо пишет в своей книге Макфедьен (Macfadyen, 1963), популяция животных становится основной единицей исследования в экологии, а статистика — главным аналитическим инструментом. Поскольку в данном случае количественные методы вытекают из самого предмета исследования, они успешно и все более интенсивно разрабатываются уже более 40 лет, используя все более разносторонний математический аппарат и имея целью построение объективных математических моделей популяционных процессов. Представление о достигнутых успехах, современных задачах и методах популяционной экологии можно получить из упомянутого труда Макфедьена и книги Уатта (Watt, 1968).

Однако сказанное относится к свободноживущим видам животных. Иное положение сложилось в паразитологии. Экологическое направление в ней обосновывалось и разрабатывалось преимущественно в Советском Союзе рядом крупных исследователей (Догель, 1941, 1962; Павловский, 1934, 1937, 1945, 1948; Беклемишев, 1970) и их учениками (литературу см.: Быховская-Павловская, 1967). Но главное внимание в этих исследованиях уделялось влиянию среды, в широком смысле этого слова, на формирование паразитофауны, организму как среде обитания, самым различным аспектам хозяино-паразитных отношений и установлению характера связей, существующих в биоценозах между различными хозяевами и переносчиками паразитов. Получаемые закономерности носили преимущественно качественный характер, а количественные ограничивались установлением элементарных соотношений, например экстенсивности и интенсивности заражения, индекса обилия, индекса доминирования и т. п. Уровням и динамике численности популяций паразитов

и факторам их регулирующим, а соответственно и разработке количественных методов, уделялось мало внимания.

Объясняется такое положение в значительной мере методическими трудностями накопления достаточно представительных в статистическом отношении материалов за последовательные ряды лет. Получение таких материалов даже по эктопаразитам сопряжено со многими трудностями, а в отношении эндопаразитов они еще более возрастают, так как для полного учета численности требуется вскрытие хозяев. Применяемые обычно при паразитологических исследованиях выборки для вскрытия в пределах 15—20 особей хозяина часто оказываются недостаточно представительными и, кроме того, проводятся в пределах одной и той же популяции хозяина лишь однократно либо в лучшем случае нерегулярно.

Тенденция паразитологов к более глубокому исследованию хозяино-паразитных отношений вполне обоснована, потому что именно эти отношения определяют основные свойства и само становление паразитических видов, но тем не менее в практических мероприятиях по борьбе с паразитами мы всегда сталкиваемся с их популяциями, существование которых не менее реально, чем у свободноживущих видов. Поэтому создавшееся положение нельзя признать нормальным не только с теоретических, но и с чисто практических позиций. Становится все более очевидным, что доминирующее в последнее время стремление биологизировать борьбу с разного рода вредителями может быть реализовано лишь в той мере, в какой удастся понять, учесть и использовать в практических мероприятиях естественные механизмы, регулирующие численность того или иного вредного вида. Для иллюстрации обоснованности такого подхода в отношении паразитов пока можно сослаться лишь на работы с малярийными комарами (Беклемишев, 1954) и с подкожными оводами (Бреев, 1971). По отношению к растительноядным насекомым сходная точка зрения с наибольшей полнотой изложена Кларком, Гейером, Хьюджесом и Моррисом в их книге «Популяционная экология насекомых в теории и практике» (Clark et al., 1967).

Очевидно, решение такого рода вопросов невозможно без изучения динамики численности популяций паразитов, а соответственно и без развития количественных методов. Их развитие должно быть направлено на разработку, освоение и все более широкое использование методов получения представительных выборок, характеризующих численность того или иного паразита, определение типа его распределения в популяции хозяина, соотношения численности самцов и самок, рождаемости и смертности, распределения смертности по фазам и стадиям, характера распределения в популяции изменчивых признаков вида и физиологического состояния особей, типа роста популяции и колебаний ее численности.

При этом, естественно, должна учитываться специфика паразитизма. В частности, поскольку важнейшую роль в становлении паразитических видов, а в связи с этим, вероятно, и уровней их численности, играют хозяино-паразитные отношения, существенное значение приобретает изучение типов распределения паразитов в популяциях хозяев и анализ факторов их определяющих. Количественные закономерности распределения, проанализированные на достаточном числе представительных выборок при разной плотности популяций одного и того же вида паразита, дают наиболее точное и полное представление об уровне его численности и изменчивости этого уровня, а с другой стороны — непосредственно связаны с процессами его определяющими, так как сам характер распределения представляет итоговый результат этих процессов.

Далее, если подходить к вопросам борьбы с вредными видами с точки зрения оптимального регулирования определенной биологической системы, включающей то или иное число биотических и абиотических компонентов, то очевидно, что система эта носит вероятностный характер. Численность вида, требующая регулирования, зависит, как правило, от сочетания столь разнообразных факторов, что предсказать ее строго определенное значение в каждый заданный момент невозможно, и потому с математической точки зрения численность вида может рассматриваться как случайная величина. Но для полной характеристики случайной величины необходимо знать три параметра: 1) ее математическое ожидание, что соответствует среднему уровню численности, 2) ее дисперсию, т. е. величину, характеризующую степень рассеяния частных значений численности от ее среднего уровня, 3) ее распределение, т. е. совокупность вероятностей появления каждого частного значения. Обычно, хотя и далеко не всегда, мы имеем представление о первых двух параметрах численности для наиболее изученных видов паразитов. Что касается третьего, то он, как правило, остается пока неизвестным, что также подчеркивает важность его исследования.

В перспективе развития методов борьбы оптимальное регулирование численности вредных видов, т. е. удержание ее на заданном экономическими или социальными требованиями уровне, при наименьших нарушениях биоценотических связей в экосистемах, представляет тот пока идеальный предел, в направлении которого развиваются сейчас и популяционная экология вредителей, и так называемые биологические методы борьбы.

Если предлагаемая брошюра, затрагивающая лишь немногие из вопросов, возникающих в ходе решения такого рода задач, будет способствовать развитию аналогичных исследований и в паразитологии, ее основная цель будет достигнута.

Автор глубоко признателен профессору П. В. Терентьеву, просмотревшему рукопись этой работы и много способствовав-

шему ее выполнению своим дружеским вниманием, и искренно благодарен товарищам по работе М. Н. Дубининой и Е. Ф. Сосниной за их постоянную готовность делиться своими материалами и знаниями о процессах заражения паразитами их хозяев.

## 2. СЛУЧАЙНАЯ ВЕЛИЧИНА, ЕЕ МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ, ДИСПЕРСИЯ И РАСПРЕДЕЛЕНИЕ. ДИСКРЕТНЫЕ РАСПРЕДЕЛЕНИЯ — БИНОМИАЛЬНОЕ И ПУАССОНА—БОРТКЕВИЧА

Поскольку брошюра посвящена методике применения в паразитологии одного из типов распределений, необходимо теперь более детально познакомиться с тем, что же такое распределение. Выше мы уже говорили, что с математической точки зрения численность того или иного вида паразита в популяции его хозяина можно рассматривать как случайную величину. Так как мы имеем дело с целыми организмами, то очевидно, что величина эта всегда будет дискретной. Распределением дискретной случайной величины называют сопоставление вероятностей появления каждого из ее возможных частных значений в пределах определенного множества. Таким множеством в нашем случае является популяция хозяина.

Если мы обозначим численность паразита  $\xi$  (кси), ее возможные значения \*  $x_0, x_1, x_2, x_3 \dots x_n$ , а соответствующие им вероятности  $p_0, p_1, p_2, p_3 \dots p_n$ , то таблица распределения примет такой вид:

$$\xi = \begin{pmatrix} x_0 & x_1 & x_2 & x_3 & \dots & x_n \\ p_0 & p_1 & p_2 & p_3 & \dots & p_n \end{pmatrix}.$$

Поскольку вероятности выражаются в долях единицы, а в ряду  $x$  представлены все возможные частные значения  $\xi$ , очевидно, что  $p_0 + p_1 + p_2 + p_3 + \dots + p_n = 1$  или  $\sum_{i=0}^n p_i = 1$ , где  $\Sigma$  обозначает сумму, а  $p_i$  — любое из значений  $p$ , которое может встретиться в пределах  $i$  от нуля до  $n$ .

Для конкретизации этого общего выражения распределения рассмотрим следующий пример. В стаде крупного рогатого скота из 200 голов оказались заражены личинками II и III стадий

\* В рассматриваемых нами случаях надо принимать во внимание и возможное отсутствие паразита, т. е. нулевую численность, у части особей популяции хозяина.

обыкновенного подкожного овода — *Hypoderma bovis* De Geer 50 голов, в том числе одной личинкой 26 голов, двумя — 10, тремя — 6, четырьмя — 4, пятью — 2 и шестью — 2 головы. Общее количество животных в стаде обозначим  $N$ , а размеры групп с равным количеством личинок на каждом животном, или, иначе говоря, частоты повторения одинаковых значений численности личинок, —  $f$ . Тогда  $f_0 = 150, f_1 = 26, f_2 = 10, f_3 = 6, f_4 = 4, f_5 = 2, f_6 = 2$ .

О вероятности появления у отдельных животных в стаде того или иного числа личинок при данном уровне численности овода мы можем судить по доле, которую занимают в стаде отдельные группы животных с равным числом личинок на каждом, т. е.

$$\begin{aligned} p_0 &= \frac{f_0}{N} = \frac{150}{200} = 0.75; & p_4 &= \frac{f_4}{N} = \frac{4}{200} = 0.02; \\ p_1 &= \frac{f_1}{N} = \frac{26}{200} = 0.13; & p_5 &= \frac{f_5}{N} = \frac{2}{200} = 0.01; \\ p_2 &= \frac{f_2}{N} = \frac{10}{200} = 0.05; & p_6 &= \frac{f_6}{N} = \frac{2}{200} = 0.01. \\ p_3 &= \frac{f_3}{N} = \frac{6}{200} = 0.03; \end{aligned}$$

Таблица распределения в этом случае получит такой вид:

$$\xi = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.75 & 0.13 & 0.05 & 0.03 & 0.02 & 0.01 & 0.01 \end{pmatrix}.$$

Таким образом, первая задача при изучении распределения паразитов в популяции хозяев заключается в исследовании достаточно представительной выборки из популяции и определении количества особей изучаемого вида паразита у каждого из членов этой выборки, после чего определяют, какую долю по отношению ко всей выборке составляет каждая группа животных с равным числом особей паразита на каждом. В результате получают так называемое эмпирическое распределение вероятностей, т. е. распределение на основе фактических данных. Дальнейшая задача заключается в том, чтобы определить, какому типу закономерностей подчиняется это распределение. Иначе говоря, возможно ли найти уравнение, определив параметры которого, можно затем с достаточной точностью (в пределах допустимых ошибок, связанных с неполнотой выборки и возможными неточностями при учете числа паразитов) рассчитать вероятность появления любого заданного числа паразитов на одной особи хозяина в пределах изучаемой популяции, т. е. построить полную таблицу распределения уже на основе его модели.

Принципиально возможны два пути решения этой задачи — аналитический и эмпирический. Идя первым из них, мы должны

оценить причины, приводящие к конкретному распределению, и на этой основе сформулировать его количественные закономерности, а затем проверить справедливость наших оценок и формулировок, сопоставляя расчетные данные с фактическими. Такой подход и является математическим моделированием в строгом смысле слова.

Однако в паразитологии мы, как правило, не располагаем еще достаточными знаниями, чтобы определить и характер и степень влияния каждого из факторов, лежащих в основе распределения. В таких случаях можно идти и другим путем. Мы можем воспользоваться уже известными в математической статистике типами распределений случайных величин и определить, какой из них наиболее сходен с изучаемым нами эмпирическим распределением. После этого, зная закономерность, лежащую в основе избранного типа распределения, мы получаем отправную точку для предположений, какие факторы в действительности могут его определять. Естественно, что такие предположения нуждаются в дальнейшем подтверждении путем уже биологического анализа — накопление новых данных о распределении в естественных популяциях, различающихся по интенсивности действия того или иного из предполагаемых факторов, или постановка соответствующих опытов и т. п. Помимо того, эмпирический подход к моделированию распределений может оказаться полезным и для выяснения некоторых новых закономерностей, что и будет показано на конкретных примерах в ходе дальнейшего изложения.

Но каков бы ни был принятый путь, первая задача при изучении случайной величины, в наших случаях — численности определенного вида паразита, состоит в определении ее основных параметров — математического ожидания и дисперсии.

Математическое ожидание дискретной случайной величины определяется суммой произведений всех ее частных значений на вероятность каждого из них:

$$M\xi = \sum x_i p_i. \quad (2.1)$$

В нашем частном примере с личинками овода математическое ожидание численности личинок будет равно (см. таблицу распределения на стр. 9):  $M\xi = 0 \times 0.75 + 1 \times 0.13 + 2 \times 0.05 + 3 \times 0.03 + 4 \times 0.02 + 5 \times 0.01 + 6 \times 0.01 = 0.51$ .

Поскольку мы определяли вероятности появления каждого частного значения на основании частоты случаев его регистрации в определенном изучаемом нами множестве, в нашем примере — число животных в стаде, нетрудно заметить, что математическое ожидание численности паразита в каждой частной выборке равно средней его численности на одну особь хозяина или другую единицу учета — гнездо, нору и т. п., обозначаемую, по терминоло-

гии В. Н. Беклемишева, индексом обилия. В самом деле, если  $p_0 = \frac{f_0}{N}$ ,  $p_1 = \frac{f_1}{N}$  и т. д., то

$$M\xi = \frac{1}{N} \sum x_i f_i, \quad (2.2)$$

т. е. мы получили формулу среднего значения вариационного ряда. В нашем примере:  $f_0 = 150$ ,  $f_1 = 26$ ,  $f_2 = 10$ ,  $f_3 = 6$ ,  $f_4 = 4$ ,  $f_5 = 2$ ,  $f_6 = 2$ . Тогда

$$\begin{aligned} \frac{1}{N} \sum x_i f_i &= \frac{1}{200} (0 \times 150 + 1 \times 26 + 2 \times 10 + 3 \times 6 + 4 \times 4 + \\ &+ 5 \times 2 + 6 \times 2) = \frac{102}{200} = 0.51. \end{aligned}$$

Среднюю численность паразитов в исследуемых выборках мы будем в дальнейшем обозначать  $M$ , в отличие от  $M\xi$ , обозначающего среднюю численность (математическое ожидание) в генеральной совокупности, т. е. во всей популяции хозяина или во всех популяциях данного вида паразита.

Дисперсия случайной величины —  $D\xi$  характеризует степень рассеяния ее частных значений от ее математического ожидания.

$$D\xi = M |(\xi - M\xi)^2|, \quad (2.3)$$

т. е., иначе говоря, дисперсия — это математическое ожидание (или среднее значение) квадрата отклонения случайной величины от ее среднего значения. Так как каждая частная величина отклонения случайной величины от ее среднего значения равна  $x_i - M$ , а частота встречаемости  $x_i$  равна  $f_i$ , мы можем написать формулу (2.3) иначе:

$$D\xi = \frac{1}{N} \sum f_i (x_i - M)^2. \quad (2.4)$$

Но

$$(x_i - M)^2 = x_i^2 - 2x_i M + M^2,$$

откуда

$$\frac{1}{N} \sum f_i (x_i - M)^2 = \frac{1}{N} \sum f_i x_i^2 - 2 \frac{1}{N} \sum f_i x_i M + \frac{1}{N} \sum f_i M^2.$$

В каждом случае применения этой формулы  $M$  и  $M^2$  будут постоянными величинами, и потому их можно вынести за знак суммы, подобно тому, как выносят за скобки постоянный множитель. Тогда

$$2 \frac{1}{N} \sum f_i x_i M = 2M \frac{1}{N} \sum f_i x_i = 2M^2,$$

так как  $\frac{1}{N} \sum f_i x_i = M$  (2.2);

$$\frac{1}{N} \sum f_i M^2 = M^2 \frac{1}{N} \sum f_i = M^2 \frac{1}{N} N = M^2,$$

так как  $\sum f_i = N$ .

В итоге мы получаем

$$\begin{aligned} \frac{1}{N} \sum f_i (x_i - M)^2 &= \frac{1}{N} \sum f_i x_i^2 - 2M^2 + M^2 = \frac{1}{N} \sum f_i x_i^2 - M^2. \\ D\xi &= \frac{1}{N} \sum f_i x_i^2 - M^2, \end{aligned} \quad (2.5)$$

т. е. дисперсия равна среднему квадрату случайной величины минус квадрат ее среднего значения.

Дисперсию случайной величины  $D\xi$ , понимаемую как средний квадрат отклонения\*, в биометрических исследованиях обычно обозначают  $\sigma^2$  (сигма-квадрат, или варианса), причем это обозначение относится к генеральной совокупности исследуемых объектов, т. е. в наших случаях оно означает дисперсию численности паразита в принятой единице учета (особь хозяина, гнездо и т. п.) во всей его популяции. При частных выборках из популяции, с которыми обычно приходится иметь дело, дисперсию (=вариансу) обозначают  $s^2$ . Из математической статистики известно следующее отношение между средним значением дисперсии частных выборок и дисперсией генеральной совокупности:  $s^2 = \sigma^2 \frac{N-1}{N}$ , откуда  $\sigma^2 = s^2 \frac{N}{N-1}$ . Поэтому, чтобы приблизить значение выборочной дисперсии  $s^2$  к значению  $\sigma^2$ , нужно умножить его на  $\frac{N}{N-1}$ . Проделаем эту операцию с полученной нами формулой (2.5). Будем обозначать различные значения случайной величины  $x$ , а частоты каждого отдельного значения  $f$ . Вспомним также, что  $M = \frac{\sum fx}{N}$  (2.2). Тогда

$$\begin{aligned} s^2 &= \left[ \frac{1}{N} \sum f x^2 - \left( \frac{\sum f x}{N} \right)^2 \right] \frac{N}{N-1} = \left( \frac{\sum f x^2}{N} - \frac{(\sum f x)^2}{N^2} \right) \frac{N}{N-1} = \\ &= \frac{\sum f x^2 N}{N(N-1)} - \frac{(\sum f x)^2 N}{N^2(N-1)} = \frac{\sum f x^2}{N-1} - \frac{(\sum f x)^2}{N(N-1)} = \frac{\sum f x^2 - \frac{(\sum f x)^2}{N}}{N-1}. \end{aligned}$$

Итак,

$$s^2 = \frac{\sum f x^2 - \frac{(\sum f x)^2}{N}}{N-1}. \quad (2.6)$$

Этой формулой мы и будем пользоваться в дальнейших расчетах.

\* Иногда дисперсией называют не средний квадрат, а сумму квадратов отклонений случайной величины от ее среднего значения —  $\sum (\xi - M\xi)^2$ .

Для вычисления  $s^2$  необходимо получить значения  $N$ ,  $\sum f x$  и  $\sum f x^2$ . Такие расчеты обычно производят, составляя таблицу последовательных значений  $x$  и  $f$ . Вычислим значение  $s^2$  для вышеупомянутого случая распределения личинок овода в стаде крупного рогатого скота (табл. 1).

Таблица 1

$x$	$f$	$fx$	$fx^2$	$x$	$f$	$fx$	$fx^2$
0	150	—	—	4	4	16	64
1	26	26	26	5	2	10	50
2	10	20	40	6	2	12	72
3	6	18	54				
				$N=200$	$\Sigma fx=102$	$\Sigma fx^2=306$	

$$s^2 = \frac{306 - \frac{(102)^2}{200}}{199} = \frac{306 - \frac{10404}{200}}{199} = \frac{306 - 52.0200}{199} = \frac{253.9800}{199} = 1.2763.$$

Рассмотрим теперь, с какими типами наиболее широко известных распределений дискретных случайных величин приходится встречаться при изучении распределения паразитов. Прежде всего нужно указать на биномиальное распределение, которое возникает в тех случаях, когда появление того или иного значения случайной величины зависит от случайных сочетаний равного количества равных по интенсивности действия положительных и отрицательных факторов, действующих каждый независимо друг от друга.

Для биномиального распределения характерна альтернативная схема изменчивости изучаемого явления или признака, т. е. только его наличие или отсутствие. Типичными примерами могут быть распределение случаев «орла» или «решки» в сериях испытаний с бросанием монеты, распределение случаев выпадения того или иного количества очков при бросании игральной кости и т. п. Изложение принципов и закономерностей этого распределения можно найти в любом руководстве по биометрии (см., например: Урбах, 1964; Лакин, 1968).

Биномиальное распределение имеет два параметра: число испытаний в каждой их серии  $n$  и вероятность осуществления изучаемого явления  $p$ . Вероятность его не осуществления обозначается  $q$ , причем  $p+q=1$ .

Математическое ожидание или среднее значение случайной величины, распределение которой подчиняется биномиальному закону, равно произведению  $n$  на  $p$ :

$$M = np. \quad (2.7)$$

## Дисперсия случайной величины при биномиальном распределении

$$\sigma^2 = npq. \quad (2.8)$$

Разложение распределения, т. е. собственно распределение вероятностей повторного осуществления изучаемого явления или, как принято называть в теории вероятностей, события в сериях из  $n$  испытаний каждой, определяется формулой бинома Ньютона:

$$(q+p)^n = \frac{1}{1} p^0 q^n + \frac{n}{1} p q^{n-1} + \frac{n(n-1)}{2!} p^2 q^{n-2} + \\ + \frac{n(n-1)(n-2)}{3!} p^3 q^{n-3} + \dots + \frac{1}{1} p^n q^0. \quad (2.9)$$

Эту формулу можно трактовать следующим образом. Если мы проводим какое-то определенное количество серий из  $n$  испытаний в каждой, причем в каждом испытании возможны лишь два исхода и вероятность осуществления изучаемого нами события равна  $p$ , то доля серий, в которых это событие не осуществляется ни одного раза, равна  $\frac{1}{1} p^0 q^n$ , появится в одном испытании —  $\frac{n}{1} p q^{n-1}$ , появится в двух испытаниях —  $\frac{n(n-1)}{2!} p^2 q^{n-2}$  и т. д. и появится во всех испытаниях —  $\frac{1}{1} p^n q^0$ .

Для биномиального распределения характерны следующие условия:

- 1) фиксированное число испытаний;
- 2) каждое испытание должно приводить либо к осуществлению, либо к неосуществлению изучаемого события;
- 3) вероятность осуществления во всех испытаниях должна быть одинакова;
- 4) испытания должны быть независимы друг от друга.

Как применить биномиальное распределение для анализа распределения паразитов? Рассмотрим это на нашем примере с личинками овода. Вспомним основные параметры этого распределения:  $N=200$ ,  $M=0.51$ ,  $s^2=1.2763$ .

0	1	2	3	4	5	6
0.75	0.13	0.05	0.03	0.02	0.01	0.01

Изучаемое событие состоит в появлении личинки овода под кожей спины у животного, причем оно может повториться до 6 раз, откуда следует, что мы имеем дело с сериями из 6 испытаний, т. е.  $n=6$ .

Второй параметр биномиального распределения —  $p$ , т. е. вероятность осуществления изучаемого события, мы можем определить из уравнения (2.7):  $M=np$ , откуда  $p=\frac{M}{n}=\frac{0.51}{6}=0.085$ . Далее определяем  $q=1-p=1-0.085=0.915$ .

Зная величины  $n$ ,  $p$  и  $q$ , мы можем по формуле бинома (2.9) рассчитать разложение распределения, т. е. получить вероятность появления серий испытаний, в которых изучаемое событие не произойдет ни одного раза, произойдет один раз, два раза и т. д. до шести раз, или, иными словами, оценить долю животных из общей их совокупности ( $N=200$ ), у которых не будет ни одной личинки, долю животных с одной личинкой и т. д. до шести. Проделаем этот расчет, взяв коэффициенты отдельных членов разложения из таблицы биномиальных коэффициентов (треугольник Паскаля), которую можно найти в приложении к руководствам по биометрии или в сборнике математических таблиц:  $(0.915+0.085)^6=0.915^6+6\times0.085\times0.915^5+15\times0.085^2\times0.915^4+20\times0.085^3\times0.915^3+15\times0.085^4\times0.915^2+6\times0.085^5\times0.915+0.085^6=0.58685+0.32710+0.07596+0.00941+0.00066+0.00002+0.00000$ .

Уже из сопоставления распределения вероятностей отдельных значений изучаемой нами численности личинок овода с полученным значением тех же вероятностей по закономерности биномиального распределения видно, что эти значения далеко не совпадают. Характерно, что и дисперсия биномиального распределения  $\sigma^2=npq=6\times0.085\times0.915=0.466650$  резко отличается от дисперсии эмпирического распределения  $s^2=1.2763$ . Чтобы окончательно убедиться в неприменимости в разбираемом случае модели биномиального распределения, умножим каждое значение вероятностей полученного теоретического ряда на  $N=200$ , т. е. узнаем, какое количество животных в стаде должно было бы не иметь личинок овода, иметь по одной личинке, по две и т. д. в случае, если бы их распределение подчинялось биномиальному закону:

$$0.58685 \times 200 = 117.370; \quad 0.00941 \times 200 = 1.882; \\ 0.32710 \times 200 = 65.420; \quad 0.00066 \times 200 = 0.132; \\ 0.07596 \times 200 = 15.192; \quad 0.00002 \times 200 = 0.004.$$

Затем сопоставим полученный теоретический ряд с эмпирическим и методом хи-квадрат ( $\chi^2$ ) определим вероятность того, что имеющиеся расхождения могут быть отнесены за счет случайных ошибок при выборке.

Как известно, критерий  $\chi^2$  представляет собой сумму отношений между квадратами разностей эмпирических и теоретических частот и теоретическими частотами. В нашем случае его формула

$$\chi^2 = \sum \frac{(f_e - f_B)^2}{f_B}, \quad (2.10)$$

где  $f_e$  — каждая из сравниваемых частот в эмпирическом ряду, а  $f_B$  — соответствующая частота в теоретическом ряду биномиального распределения.

Но прежде чем переходить к вычислению  $\chi^2$  (табл. 2), нужно проделать некоторую предварительную обработку сопоставляемых рядов. Дело в том, что в случаях очень малых теоретических частот

при вычислении  $\chi^2$  могут возникать ошибки, искажающие общий результат. Поэтому принимают за минимальную частоту в теоретическом ряду 5, а все меньшие по размерам частоты объединяют в группы, суммарное значение которых должно быть близким к 5. В нашем примере мы должны объединить для этого частоты с 3-го по 6-й классы, хотя суммарное значение их равно всего 2.02. Объединять эти частоты со 2-м классом мы не можем, потому что тогда мы имели бы всего три члена в каждом ряду для окончательного сравнения и у нас не хватило бы степеней свободы (см. ниже). Соответствующее объединение частот надо произвести и в эмпирическом ряду.

Далее, необходимо уравнять суммарное значение теоретического ряда с суммой эмпирических частот, но в нашем случае суммы обоих рядов совпадают.

Таблица 2

$x$	$f_e$	$f_B$	$f_e - f_B$	$(f_e - f_B)^2$	$\frac{(f_e - f_B)^2}{f_B}$
0	150	117.37	32.63	1064.7169	9.0715
1	26	65.42	39.42	1553.9364	23.7532
2	10	15.19	5.19	26.9361	1.4733
3	6	1.88			
4	4	0.13	2.02	11.98	143.5204
5	2	0.01			
6	2	0.00			
	200	200.00			$\chi^2 = 105.6477$ $v = 4 - 3 = 1$ $P < 0.0005$

Теперь мы можем, проделав необходимые расчеты, определить значение  $\chi^2$ . Получаем:  $\chi^2 = 105.6477$ .

Чтобы определить соответствующую этому значению  $\chi^2$  вероятность случайности различия между теоретическим и эмпирическим рядами, следует найти еще число степеней свободы сравниваемых рядов —  $v$ . Оно определяется вычитанием из числа независимых переменных величин в каждом из рядов числа параметров распределения. Очевидно, что во всех случаях число независимых переменных величин в каждом ряду будет на единицу меньше числа составляющих его величин, потому что сумма всех составляющих не должна меняться. При этом условии значение одной из величин ряда всегда будет зависеть от суммы остальных, и потому эта величина независимой не является.

В нашем примере каждый из рядов после объединения частот в группу состоит из четырех величин, три из которых будут независимыми переменными.

Биномиальное распределение имеет два параметра. Число степеней свободы наших рядов  $v = 3 - 2 = 1$ . На практике число степеней свободы определяют, вычитая из общего числа сопоставляемых величин число параметров распределения, увеличенное на единицу. Правило равенства сумм обоих рядов тоже параметр, ограничивающий степени свободы. Тогда  $v = 4 - 3 = 1$ .

Отсюда понятно, почему мы не могли объединить значения частот 3—6-го классов вместе со 2-м. Тогда в каждом из рядов мы имели бы всего по три величины для сравнения, а  $3 - 3 = 0$ , т. е. применение критерия  $\chi^2$  для определения сходимости рядов становится невозможным. Очевидно, если применять этот критерий для распределений, имеющих два параметра, то число величин в сравниваемых рядах должно быть не менее четырех.

Полученное значение  $\chi^2 = 105.6477$  при одной степени свободы очень велико. По прилагаемой в конце брошюры таблице процентных точек распределения  $\chi^2$  можно убедиться в том, что вероятность случайности расхождения между эмпирическим и теоретическим рядом при таком значении  $\chi^2$  для одной степени свободы много меньше 0.05%. Это означает, что если бы распределение личинок обыкновенного овода на всем поголовье стад, где условия заражения животных оводом были идентичны с исследованным стадом, строго соответствовало бы биномиальному закону, то, беря случайные выборки по 200 голов из этого поголовья, мы получили бы аналогичные расхождения между эмпирическим и теоретическим рядом менее чем в 0.05% случаев. Такая вероятность ничтожно мала, и потому мы вправе заключить, что интересующее нас распределение личинок овода не подчиняется закономерности биномиального распределения. Говоря языком статистики, это означает, что очень мала вероятность нуль-гипотезы, т. е. исходного предположения о возможности соответствия распределения личинок в стаде биномиальному закону, которое мы сделали, начиная наши расчеты.

Но отсюда следует, что биологические процессы, лежащие в основе изучаемого распределения, в чем-то отличаются от приведенных выше условий биномиального распределения. Поскольку первые два из этих условий — фиксированное число испытаний и возможность только двух исходов каждого из них (наличие или отсутствие личинки овода) в нашем случае выполняются, можно предполагать, что не соблюдаются одно или оба из остальных двух условий, т. е. 1) вероятность заражения отдельных животных тем или иным количеством личинок в пределах исследуемого множества (стада) не остается постоянной или (и) 2) наличие одной или нескольких личинок у животного каким-то образом влияет на вероятность заражения или выживания осталь-

ных личинок, что противоречит условию независимости испытаний.

Приведенный ход рассуждений иллюстрирует, каким образом даже при несовпадении исследуемой модели распределения с фактическими данными можно прийти к некоторым, хотя бы и негативным, выводам о характере процессов, определяющих распределение паразита в популяции хозяина.

Рассмотрим теперь еще один широкоизвестный тип распределения, применение которого возможно в паразитологических исследованиях, а именно распределение редких событий, называемое также распределением Пуассона—Борткевича. Оно является предельным случаем биномиального распределения и возникает при очень малых вероятностях изучаемых событий и при очень большом числе испытаний. Но для него также характерны два последних условия, определяющие биномиальное распределение, — постоянство вероятности осуществления события во всех испытаниях в пределах исследуемого множества и полная независимость испытаний друг от друга.

Хорошим примером распределения Пуассона может быть рождение троен в семьях большого города. Случайной величиной здесь будет число троен, родившихся за год, а испытанием — каждый случай рождения ребенка. Число таких испытаний за год достаточно велико, и все они независимы друг от друга. Вероятность рождения тройни очень мала и в то же время постоянна для каждого испытания. Поэтому, если взять статистику рождения троен в городе за достаточно большой промежуток времени, например за 50 лет, и определить число лет, когда не рождалось ни одной тройни, рождалась одна, две, три и т. д., распределение этих чисел будет подчиняться закону Пуассона. Он формулируется так:

$$P_x = \frac{M^x}{x!} e^{-M}, \quad (2.11)^*$$

где  $P_x$  — вероятность появления значения  $x$  случайной величины,  $M$  — математическое ожидание (среднее значение) этой величины,  $e$  — основание натуральных логарифмов ( $e=2.71828$ ).

Из формулы (2.11) очевидно, что распределение Пуассона характеризуется всего одним параметром — средним значением случайной величины  $M$ . В самом деле, если, например, в среднем за год в городе рождается одна тройня ( $M=1$ ), то этого уже достаточно, чтобы определить, какую долю в статистической выборке будут составлять годы, в течение которых не было рожде-

\* Вывод этого уравнения как предельного случая биномиального распределения, когда  $p$  очень мало, а  $n \rightarrow \infty$ , можно найти в книге Урбаха (1964, стр. 84—85) или в любом учебнике по теории вероятностей и математической статистике.

ний троен, родились одна, две и т. д. Проделаем такой расчет, приняв во внимание, что факториал нуля равен  $0!=1$ .\*

$$P_0 = \frac{M^0}{0!} e^{-M} = \frac{1}{1} \cdot \frac{1}{2.71828} = 0.368,$$

$$P_1 = \frac{M^1}{1!} e^{-M} = \frac{1}{1} \cdot \frac{1}{2.71828} = 0.368,$$

$$P_2 = \frac{M^2}{2!} e^{-M} = \frac{1}{2} \cdot \frac{1}{2.71828} = 0.184,$$

$$P_3 = \frac{M^3}{3!} e^{-M} = \frac{1}{6} \cdot \frac{1}{2.71828} = 0.061 \text{ и т. д.}$$

Характерной особенностью распределения Пуассона является также совпадение по абсолютной величине среднего значения случайной величины и ее дисперсии:  $\sigma^2=M$ . Отсюда следует важный вывод, что во всех случаях, когда в изучаемом эмпирическом распределении дисперсия равна среднему значению, оно подчиняется закономерности Пуассона.

Отношение дисперсии к среднему значению случайной величины  $\frac{\sigma^2}{M}$  имеет существенное значение в теории распределений. При равномерно случайному распределению, т. е. в случаях, когда соблюдаются условия, характерные для распределения Пуассона,  $\frac{\sigma^2}{M}=1$ . В случаях, когда  $\frac{\sigma^2}{M}<1$ , распределения называют недорассеянными, а когда  $\frac{\sigma^2}{M}>1$  — перерассеянными.

К недорассеянным, в частности, относятся биномиальное и нормальное распределения, а одному из наиболее часто встречающихся типов перерассеянных распределений — негативному биномиальному — посвящена эта брошюра.

Таким образом, узнав дисперсию и среднее значение случайной величины, по их отношению мы можем судить, к какому типу относится ее распределение, и при поисках закономерности, которой оно подчиняется, выбирать соответствующие модели.

Но вернемся к распределению Пуассона. Если увеличивать значение его единственного параметра —  $M$ , то оно становится весьма сходным с симметричным биномиальным распределением. Последнее возникает в тех случаях, когда  $p=q=0.5$ , т. е. вероятности осуществления и не осуществления изучаемого события равны.

\* Это кажущееся парадоксальное равенство вытекает из следующего рассуждения. Поскольку знаком факториала (!) обозначают произведение натурального ряда чисел, то  $n! = (n-1)! n$ , откуда  $(n-1)! = \frac{n!}{n}$ , и если  $n=1$ , то  $(1-1)! = \frac{1!}{1} = 1$ .

велики, как, например, выпадение «орла» или «решки» при бросании монеты.

В паразитологии распределение паразитов в популяции хозяина подчиняется закону Пуассона, когда вероятность заражения тем или иным числом особей паразита невелика и практически одинакова для всех членов популяции хозяина и когда заражение хозяина одной или несколькими особями паразита практически не влияет на возможность дальнейшего заражения и выживания паразита. Рассмотрим конкретный пример.

Заражение рыб лентециом *Digamma interrupta* Rud. происходит, когда они поедают циклопов с инвазионными процеркоидами этого лентеца, из которых затем в организме рыбы развиваются плероцеркоиды. Известно (Дубинина, 1966), что даже при искусственном заражении циклопов и других ракообразных — первых промежуточных хозяев лентецов — большим количеством онкосфер до стадии инвазионного процеркоида в организме одного рака выживают не более 1—3 особей паразита. Поэтому можно предполагать, что в естественной популяции раков заражение более чем одним процеркоидом случается очень редко. Кроме того, средняя численность процеркоидов в ракообразных бывает обычно очень низкой. Не удается обнаружить зараженных раков даже при обследовании более тысячи экземпляров в очаге заражения. Но у рыб повторное заражение несколькими плероцеркоидами, по-видимому, не сказывается на их развитии. Известны случаи находок до 6—7 плероцеркоидов в одной особи.

Если теперь представить популяцию какого-то вида рыб, в котором могут развиваться плероцеркоиды, обитающую в водоеме, где имеются ракообразные, зараженные процеркоидами, то у одновозрастного молодняка этой популяции распределение заражения плероцеркоидами должно подчиняться закону Пуассона. В самом деле, 1) при малой средней численности раков с процеркоидами вероятность поедания рыбой зараженного рака будет небольшой и одинаковой для всех рыб одного возраста, 2) в случае повторного поедания зараженных раков на развитии новых плероцеркоидов имеющееся заражение не скажется.

Однако если брать популяцию данного вида рыб в целом или только рыб более старшего возраста, то условие постоянства вероятности заражения уже выполниться не будет, потому что 1) рыбы разного возраста поедают в единицу времени неодинаковое количество раков и 2) у взрослых рыб будет наслаждаться заражение за несколько последующих сезонов с разной вероятностью заражения в каждом из них. Поэтому в последних случаях, исследуя распределение плероцеркоидов в выборках из популяции рыб, трудно ожидать соответствия распределению Пуассона.

Проверим теперь, насколько справедливы наши предположения. При обследовании в 1959 г. локальной популяции серебряного

карася в одном из прибайкальских озер на заражение плероцеркоидами лентеца *D. interrupta* M. N. Дубининой были получены следующие результаты (табл. 3).

Таблица 3

Распределение плероцеркоидов *Digamma interrupta* Rud.  
в возрастных группах популяции *Carassius auratus* (Bloch)

Возрастная группа рыб	Число обследованных особей	Не заражены	Заряжены плероцеркоидами					шестью
			одним	двумя	третя	четырьмя	пятью	
Сеголетки . . . .	49	38	9	2	0	0	0	0
Годовики . . . .	41	33	7	1	0	0	0	1
Двухлетки . . . .	89	69	15	3	0	1	0	0
Трехлетки . . . .	102	90	6	3	2	0	1	0
Популяция в целом . . . .	281	230	37	9	2	1	1	1

Вычислим теперь  $M$ ,  $s^2$  и  $\frac{s^2}{M}$  для распределений плероцеркоидов по каждой возрастной группе в отдельности и для всей популяции в целом (табл. 4—8).

Таблица 4

Сеголетки

$x$	$f$	$fx$	$fx^2$
0	38	—	—
1	9	9	9
2	2	4	8
$N = 49$		$\Sigma fx = 13$	$\Sigma fx^2 = 17$
$M = \frac{13}{49} = 0.2653;$			

$$s^2 = \frac{17 - \frac{169}{49}}{48} = \frac{17 - 3.4490}{48} =$$

$$= \frac{13.5510}{48} = 0.2823;$$

$$\frac{s^2}{M} = \frac{0.2823}{0.2653} = 1.064.$$

Таблица 5

Годовики

$x$	$f$	$fx$	$fx^2$
0	33	—	—
1	7	7	7
2	1	2	4
$N = 41$		$\Sigma fx = 9$	$\Sigma fx^2 = 11$
$M = \frac{9}{41} = 0.2195;$			

$$s^2 = \frac{11 - \frac{81}{41}}{40} = \frac{11 - 1.9757}{40} = \frac{9.0243}{40} = 0.2256;$$

$$\frac{s^2}{M} = \frac{0.2256}{0.2195} = 1.028.$$

Двухлетки

Таблица 6

$x$	$f$	$fx$	$fx^2$	$x$	$f$	$fx$	$fx^2$
0	69	—	—	4	1	4	16
1	15	15	15	5	0	0	0
2	3	6	12	6	1	6	36
3	0	0	0				
					$N = 89$	$\Sigma fx = 31$	$\Sigma fx^2 = 79$

$$M = \frac{31}{89} = 0.3520;$$

$$s^2 = \frac{79 - \frac{961}{89}}{88} = \frac{79 - 10.79775}{88} = \frac{68.20225}{88} = 0.77502;$$

$$\frac{s^2}{M} = \frac{0.77502}{0.3520} = 2.202.$$

Трехлетки

Таблица 7

$x$	$f$	$fx$	$fx^2$	$x$	$f$	$fx$	$fx^2$
0	90	—	—	4	0	0	0
1	6	6	6	5	1	5	15
2	3	6	12				
3	2	6	18				
					$N = 102$	$\Sigma fx = 23$	$\Sigma fx^2 = 61$

$$M = \frac{23}{102} = 0.2255;$$

$$s^2 = \frac{61 - \frac{529}{102}}{101} = \frac{61 - 5.1863}{101} = \frac{55.8137}{101} = 0.5526;$$

$$\frac{s^2}{M} = \frac{0.5526}{0.2255} = 2.450.$$

Популяция в целом

Таблица 8

$x$	$f$	$fx$	$fx^2$	$x$	$f$	$fx$	$fx^2$
0	230	—	—	4	1	4	16
1	37	37	37	5	1	5	25
2	9	18	36	6	1	6	36
3	2	6	18				
					$N = 281$	$\Sigma fx = 76$	$\Sigma fx^2 = 168$

$$M = \frac{76}{281} = 0.2705;$$

$$s^2 = \frac{168 - \frac{5776}{281}}{280} = \frac{168 - 20.5552}{280} = \frac{147.4448}{280} = 0.5266;$$

$$\frac{s^2}{M} = \frac{0.5266}{0.2705} = 1.947.$$

Сравнивая теперь полученные значения  $\frac{s^2}{M}$ , мы видим, что они очень близки к единице у сеголеток и годовиков и значительно превышают единицу в остальных группах и по выборке в целом. Следовательно, в первых двух случаях распределение плероцеркоидов должно быть очень близким к распределению Пуассона, а во всех остальных оно этой закономерности явно не подчиняется, что уже подтверждает вышеприведенный ход рассуждений о факторах, обусловливающих характер распределения.

Но проверим подробнее, насколько совпадают распределения плероцеркоидов в двух первых группах с распределением Пуассона (2.11).

1. Сеголетки,  $N = 49$ ,  $M = 0.2653$ .

Нулевой член распределения, т. е. доля рыб в выборке, не зараженных плероцеркоидами, будет определяться значением:

$$P_0 = \frac{M^0}{0!} e^{-M} = \frac{1}{1} 2.71828^{-0.2653}.$$

$$\lg 2.71828^{-0.2653} = \lg 2.71828 \times -0.2653 = 0.4342945 * \times -0.2653 = \\ = -0.1152183 = 1.8847817 = \lg 0.766976.$$

$$P_0 = 0.76698;$$

$$P_1 = \frac{M^1}{1!} e^{-M} = \frac{0.2653}{1} \times 0.766976 = 0.20348;$$

$$P_2 = \frac{M^2}{2!} e^{-M} = \frac{0.2653^2}{2} \times 0.766976 = 0.02699.$$

Чтобы получить теоретический ряд распределения для данной выборки, нужно умножить полученные вероятности  $P_0$ ,  $P_1$  и  $P_2$  на размер выборки, т. е. на  $N$ :  $P_0N = 0.76698 \times 49 = 37.582$ ;  $P_1N = 0.20348 \times 49 = 9.971$ ;  $P_2N = 0.02699 \times 49 = 1.323$ .

\* Лучше использовать семизначные логарифмы, но в крайнем случае можно ограничиться пятизначными. Наиболее удобны для работы «Десятизначные таблицы логарифмов чисел от 1 до 100 000» Петерса (изд. Вычислительного центра АН СССР, М., 1967) или «Таблицы семизначных логарифмов» Г. Вега (Геодезиздат, М., 1960).

Сопоставим теперь эмпирический и теоретический ряды и определим методом  $\chi^2$  вероятность случайности расхождения между ними (табл. 9).

Таблица 9

$x$	$f_e$	$f_P$	$f_e - f_P$	$(f_e - f_P)^2$	$\frac{(f_e - f_P)^2}{f_P}$
0	38	37.58	0.42	0.1764	0.0047
1	9	9.97	0.97	0.9409	0.0944
2	2	1.32 0.13 } 1.45	0.55	0.3025	0.2036
	49	+ 48.87 0.13 — 49.00		$\chi^2 = 0.3077$ $v = 3 - 2 = 1$ $0.50 < P < 0.60$	

Здесь, чтобы сравнивать ряды, необходимо уравнять суммарное значение теоретического ряда с суммой эмпирических частот, так как при вычислении теоретического распределения мы ограничились только тремя членами, опустив все последующие, так называемый «хвост». Суммируя частоты теоретического ряда, получаем  $\sum f_P = 48.87$ , в то время как  $\sum f_e = 49$ . Чтобы уравнять оба ряда, добавим разность  $\sum f_e - \sum f_P = 0.13$  к последней частоте теоретического ряда.

Проделав необходимые расчеты, получаем значение  $\chi^2 = 0.3077$ . Чтобы узнать число степеней свободы, уменьшим на 2 количество сравниваемых величин, так как распределение Пуассона обладает только одним параметром.

По таблице процентных точек распределения  $\chi^2$  (см. «Приложение») определяем, что при одной степени свободы полученное значение  $\chi^2$  лежит между 50 и 60 %. Это означает, что вероятность случайности имеющихся расхождений между эмпирическим и теоретическим рядом находится между 50 и 60 %. Иначе говоря, если распределение плероцеркоидов у сеголеток во всей локальной популяции серебряного карася строго соответствует распределению Пуассона, то, беря выборки сеголеток по 49 экземпляров, мы более чем в 50 % случаев получили бы аналогичные по величине  $\chi^2$  расхождения, какие имеются в нашем случае. Очевидно, что эта вероятность велика, и потому мы вправе утверждать, что распределение плероцеркоидов у сеголеток карася соответствует распределению Пуассона.

2. Годовики,  $N = 41$ ,  $M = 0.2195$ .

$$P_0 = \frac{1}{1} \cdot 2.71828^{-0.2195}.$$

$$\lg 2.71828 \times -0.2195 = 0.4342945 \times -0.2195 = -0.0953276 = 1.9046724 = \\ = \lg 0.80292.$$

$$P_0 = 0.80292;$$

$$P_1 = \frac{0.2195}{1} \cdot 0.80292 = 0.17624;$$

$$P_2 = \frac{0.2195^2}{2} \cdot 0.80292 = 0.01934.$$

$$P_0 N = 0.80292 \times 41 = 32.92;$$

$$P_1 N = 0.17624 \times 41 = 7.226;$$

$$P_2 N = 0.01934 \times 41 = 0.793.$$

В этой возрастной группе вероятность случайности имеющихся расхождений между эмпирическим и теоретическим рядом (табл. 10) еще более велика, чем у сеголеток, т. е. предполагаемая закономерность распределения подтверждается еще более четко.

Таблица 10

$x$	$f_e$	$f_P$	$f_e - f_P$	$(f_e - f_P)^2$	$\frac{(f_e - f_P)^2}{f_P}$
0	33	32.92	0.08	0.0064	0.0002
1	7	7.23	0.23	0.0524	0.0073
2	1	0.79 0.06 } 0.85	0.15	0.0225	0.0265
	41	+ 40.94 0.06 — 41.00		$\chi^2 = 0.0340$ $v = 3 - 2 = 1$ $0.80 < P < 0.90$	

Как было уже сказано, распределение плероцеркоидов в остальных двух возрастных группах и в выборке в целом, уже судя по величине отношения  $\frac{\sigma^2}{M}$ , не подчиняется закону Пуассона. Чтобы окончательно убедиться в этом, рассчитаем теоретический ряд распределения для выборки в целом, поскольку в этом случае  $\frac{\sigma^2}{M} = 1.947$  меньше, чем в группах двухлеток (2.202) и трехлеток (2.450).

Для выборки в целом  $N = 281$ ,  $M = 0.2705$ .

$$P_0 = \frac{1}{1} \cdot 2.71828^{-0.2705}.$$

$$\lg 2.71828 \times -0.2705 = 0.4342945 \times -0.2705 = -0.1174767 = 1.8825233 = \\ = \lg 0.762998.$$

$$P_0 = 0.76300;$$

$$P_1 = \frac{0.2705}{1} \cdot 0.762998 = 0.20639;$$

$$P_2 = \frac{0.2705^2}{2} \cdot 0.762998 = 0.02791;$$

$$P_3 = \frac{0.2705^3}{6} \cdot 0.762998 = 0.00252;$$

$$P_4 = \frac{0.2705^4}{24} \cdot 0.762988 = 0.00017;$$

$$P_5 = \frac{0.2705^5}{120} \cdot 0.762988 = 0.00001;$$

$$P_6 = \frac{0.2705^6}{720} \cdot 0.762988 = 0.00000.$$

$$0.76300 \times 281 = 214.403; \quad 0.00252 \times 281 = 0.708; \\ 0.20639 \times 281 = 57.996; \quad 0.00017 \times 281 = 0.048; \\ 0.02791 \times 281 = 7.843; \quad 0.00001 \times 281 = 0.002.$$

Если теперь сопоставить эмпирический ряд с теоретическим, то и без вычисления  $\chi^2$  видно, что они явно не совпадают, т. е. в данном случае распределение плероцеркоидов не подчиняется закону Пуассона. Очень большая величина  $\chi^2$  при одной степени свободы и ничтожная вероятность случайности имеющихся расхождений подтверждают такой вывод (табл. 11).

Таблица 11

$x$	$f_e$	$f_P$	$f_e - f_P$	$(f_e - f_P)^2$	$\frac{(f_e - f_P)^2}{f_P}$
0	230	214.403	15.60	243.3600	1.1351
1	37	57.996	21.00	441.0000	7.6034
2	9	7.843 0.708 0.048 0.002	8.601 5.40 29.1600	3.3907	
3	2				
4	1				
5	1				
6	1				
	281	281.000		$\chi^2 = 12.1292$ $v = 3 - 2 = 1$ $P < 0.0005$	

Причина, как мы видели выше, заключается в том, что вероятность заражения плероцеркоидами лентеца у отдельных особей в пределах всей популяции, а также у рыб старших возрастов не остается постоянной, а варьирует.

Это же обстоятельство является одной из причин несовпадения распределения личинок подкожного овода крупного рогатого скота с биномиальным распределением (см. стр. 17). Но в последнем случае дело обстоит еще сложнее. При учетах численности подкожных оводов мы обычно имеем дело с личинками II и III стадий, находящимися в подкожных соединительнотканых капсулах на спине животных. По срокам жизни эти стадии относятся уже ко второй половине жизненного цикла овода, и потому распределение личинок II и III стадий представляет результат наложения относительно сложных и разнообразных процессов. Первоначальное заражение скота подкожным оводом происходит при откладке яиц самками на волосы тела животных. Так как численность взрослых оводов очень мала, в стаде средних размеров (100—150 голов) одновременно редко бывает более 1—2 самок овода. У обыкновенного подкожного овода, пример распределения личинок которого мы рассматривали, самки откладывают при каждом подлете к животному только по одному яйцу. Учитывая малое число самок по сравнению с числом животных в стаде, можно предполагать, что распределение яиц среди животных должно носить равномерно-случайный характер, т. е. подчиняться закону Пуассона, а при более интенсивном заражении — биномиальному. Эмбриональная смертность у этого вида овода очень мала, в пределах 5—10%, и потому вряд ли существенно влияет на распределение. Но смертность личинок I стадии, наоборот, очень велика и зависит, с одной стороны, от физиологического состояния хозяина, а с другой — от интенсивности заражения: чем выше интенсивность, тем больше личинок погибает. Поэтому вероятность выживания отдельной личинки до II—III стадии у отдельных животных в пределах стада в конечном счете сильно варьирует, причем пока трудно сказать что-либо определенное о характере этого варьирования.

Итак, и в случаях распределения плероцеркоидов в старших возрастных группах и в популяции рыб в целом, и в случае распределения личинок подкожного овода в стадах крупного рогатого скота, а также и во многих других случаях распределения паразитов в популяциях их хозяев мы будем иметь дело с процессами, приводящими к изменчивости вероятности заражения одной особи хозяина особью паразита в пределах популяции или возрастной группы. Для статистического моделирования такого распределения необходимы и соответствующие типы распределений. Одним из наиболее широко известных в математической статистике распределений с переменной вероятностью является негативное биномиальное распределение.

### 3. НЕГАТИВНОЕ БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И МЕТОДЫ ЕГО ПРИМЕНЕНИЯ

Негативное биномиальное распределение представляет распределение событий в совокупности (определенном множестве) объектов при условии, что математическое ожидание этих событий не фиксировано в пределах этой совокупности, а меняется случайно по закономерности распределения  $\chi^2$ . Иными словами, вероятность наступления событий в этих условиях не остается постоянной в пределах совокупности, а случайно варьирует по тому же закону, что и отклонения параметров строго случайных выборок из генеральной совокупности от параметров этой совокупности.

Поясним примером, что такое распределение  $\chi^2$ . Представим себе, что мы имеем по 5000 штук красных и белых фишек. Разложим их на игральном столе, разделенном на 10 квадратов, помеченных цифрами от 1 до 10, причем на нечетные квадраты положим по 1000 красных, а на четные по 1000 белых фишек. Таким образом, в генеральной совокупности наших фишек красные и белые будут составлять ровно по 50%. Будем теперь бросать шарик рулетки и из того квадрата, на котором он остановится, брать каждый раз по одной фишке. После какого-то количества бросков, например 100, мы получим выборку в 100 фишек, причем количество красных и белых в этой выборке может отличаться от 50.

Как мы уже видели, сумму отношений квадрата разности между фактическим и ожидаемым количеством к ожидаемому называют  $\chi^2$ . Допустим красных было 48, а белых 52. Тогда

$$\chi^2 = \frac{(48 - 50)^2}{50} + \frac{(52 - 50)^2}{50} = 0.16.$$

Если мы будем повторять выборки по 100 фишек и определять  $\chi^2$  для каждой выборки, то получим разные значения, причем сравнительно небольшие отклонения и, соответственно, небольшие величины будут встречаться чаще, а большие реже. Разбив полученные в достаточно большом числе выборок, например 100 или 200, значения  $\chi^2$  на классы, мы сможем оценить частоты каждого из этих классов. Распределение этих частот и будет распределением  $\chi^2$ .

Теория распределения  $\chi^2$  разработана Пирсоном (Pearson, 1900), а в более доступном изложении подробнее с ним можно познакомиться по руководству Снедекора (Snedecor, 1957).

Но вернемся к негативному биномиальному распределению и для лучшего его понимания приведем паразитологический пример. Представим себе популяцию какого-то вида грызунов, хотя бы мышей, обитающую на участке территории, где живет также популяция какого-то вида иксодовых клещей. При передвижении мышей по территории происходят их встречи с личинками клещей

и заражение. Попадание каждой личинки на грызуна мы можем рассматривать с точки зрения теории вероятностей как событие. Отлавливая грызунов и подсчитывая количество личинок на них, мы сможем изучить распределение таких событий в пределах совокупности наших объектов, т. е. популяции данного вида мышей; другими словами — это будет распределение численности личинок клещей на их хозяевах, понимая численность как случайную величину.

Чем же будет определяться в первом приближении характер этого распределения? Мы можем допустить, что мыши передвигаются по занимаемой ими территории относительно равномерно, посещая все участки территории с равной вероятностью. Напомним, что речь идет не об отдельных особях, а о всей популяции. Если при этом личинки клеща рассеяны также равномерно, то вероятность попадания личинки на грызуна может быть большей или меньшей в зависимости от плотности популяции клещей, но останется постоянной для всей территории. В этом случае средняя плотность личинок клещей, а соответственно и вероятность заражения мыши личинкой, в пределах нашей совокупности фиксирована. Изучая распределение личинок на мышах, мы получим распределение Пуассона или биномиальное распределение, при условии, конечно, что события независимы друг от друга, т. е. заражение каждой личинкой не влияет на заражение последующими.

Однако в природе в большинстве случаев зависимости оказываются более сложными. В самом деле, даже если мыши передвигаются относительно равномерно по всей территории, то распределение личинок на ней далеко не равномерно. Оно зависит от распределения яйцекладок самок клещей, в свою очередь зависящего от распределения участков, на которых напитавшиеся самки покидают своих хозяев — крупных млекопитающих. Такие участки концентрируются возле мест лёжек хозяев. Но самки расползаются с этих участков на разные расстояния, некоторые из них, возможно, отпадают после кровососания необязательно на местах лёжек. Помимо того, размеры яйцекладок и процент выживания яиц в каждой из них также неодинаков, как неодинаковы и радиус расползания выплодившихся личинок, и готовность к нападению каждой из них.

Мы имеем, таким образом, сложное сочетание ряда факторов, но в итоге их взаимодействия распределение личинок клеща по территории можно представить в виде большого числа микрочагов, в каждом из которых средняя плотность личинок (на единицу площади территории), а соответственно и вероятность попадания личинки на мышь, когда последняя пробегает через такой участок, будет иметь какое-то свое значение. В результате при передвижении мышей по территории, занимаемой их популяцией, вероятность заражения их личинками будет меняться,

т. е. математическое ожидание события перестанет быть фиксированным.

Если распределение средних плотностей личинок в микроочагах носит случайный характер (распределение  $\chi^2$ ), то по такому же закону будет варьировать и вероятность заражения ими мышей. В этом случае распределение личинок в популяции данного вида мышей будет подчиняться закономерности негативного биномиального распределения.

Разумеется, приведенный пример лишь один из многих, которыми можно было бы проиллюстрировать, что такое негативный бином. Если ограничиться только паразитологическим материалом, то вероятность заражения той или иной особи хозяина особью того или иного вида паразита в пределах популяций чаще бывает изменчива, чем постоянна. Эта изменчивость может быть связана и с непостоянством численности паразита в течение сезона (если анализируется распределение паразита на выборках из популяции хозяина за целый сезон), и с неодинаковой активностью (инфективностью) паразита в пределах его популяции (на территории или во времени), и с различиями в поведении хозяина (неодинаковое поведение разных половозрастных групп, предпочтение одних участков территории по сравнению с другими), и, наконец, с изменчивостью хозяино-паразитных отношений в пределах популяций хозяина и паразита в зависимости от физиологического состояния отдельных особей и от отношения уровней численности обоих популяций, т. е. в конечном счете от интенсивности заражения. Последний случай наиболее интересен с точки зрения изучения факторов, регулирующих средние уровни и динамику численности популяций паразитов.

Поскольку число факторов, определяющих изменчивость вероятности заражения особи хозяина особью паразита, велико, эта изменчивость скорее всего будет носить случайный характер, и потому можно ожидать широкой приложимости негативного биномиального распределения к популяционной экологии паразитов.

Познакомимся теперь подробнее с математическим аппаратом, необходимым для работы с негативным биномиальным распределением. Выше мы видели, что простое биномиальное распределение подчиняется закону бинома Ньютона  $(q+p)^n$ , где  $p$  — вероятность осуществления изучаемого события,  $q$  — вероятность его не осуществления, а  $n$  — число испытаний, в сериях которых происходит это событие.

Негативное биномиальное распределение получило свое название потому, что его классическая формула  $(q-p)^{-k}$  сходна с формулой биномиального распределения и отличается отрицательным знаком у  $p$  и у показателя степени бинома  $k$ . Это распределение имеет два параметра — математическое ожидание  $M$  (в наших случаях средняя численность паразита на особь хо-

зяина в изучаемой выборке) и экспоненту  $k$ , которая характеризует меру дисперсии или меру агрегированности объектов, распределение которых изучается, причем с увеличением агрегированности, а соответственно и дисперсии, абсолютное значение  $k$  убывает. Например, если мы имеем две группы животных по 50 особей в каждой, с равной численностью паразитов в каждой группе — 200 ( $M=200 : 50 = 4$ ), распределение которых подчиняется негативному биному, но в первой группе паразиты распределены более равномерно среди 40 особей, а во второй сконцентрированы на 20 животных, то в первом случае  $k$  будет иметь большее абсолютное значение, чем во втором. Если говорить принятыми в паразитологии терминами, то при равном обилии паразита абсолютная величина  $k$  будет тем меньше, чем меньше экстенсивность и чем больше средняя и максимальная интенсивность заражения.

Величины  $p$  и  $q$  являются производными от основных параметров. Если принять классическую формулу негативного бинома  $(q-p)^{-k}$ , то  $p = \frac{k}{M}$ , а так как  $q-p=1$ , то  $q=1+p$ . Однако при такой системе обозначений формула разложения распределения, с которой придется иметь дело при расчетах теоретического ряда, более сложна, и потому можно воспользоваться несколько иной системой обозначений, а именно — принять, что  $q+p=1$ . Тогда

$$p = \frac{k}{k+M}, \quad (3.1)$$

$$q = 1 - p, \quad (3.2)$$

а формула распределения принимает вид  $p^k(1-q)^{-k}$  и разложение может быть рассчитано по формуле разложения функции  $(1-x)^{-m}$  в степенной ряд.

Каковы методы определения параметров негативного биномиального распределения? Математическое ожидание  $M$  определяется по формуле (2.2):  $M = \frac{\Sigma fx}{N}$ . Определение экспоненты  $k$  значительно более сложно. Сначала определяют начальное значение  $k$  методом моментов, \* исходя из уравнения  $s^2 = M + \frac{M^2}{k}$ , откуда

$$k = \frac{M^2}{s^2 - M}, \quad (3.3)$$

где  $s^2$  — дисперсия случайной величины в изучаемой выборке — см. (2.6):

$$s^2 = \frac{\Sigma fx^2 - \frac{(\Sigma fx)^2}{N}}{N-1}.$$

\* О статистических моментах см.: Урбах, 1964, стр. 48—52.

Из уравнения (3.3), в частности, следует, что чем ближе значение дисперсии  $s^2$  к величине среднего  $M$ , т. е. чем менее агрегированным и чем более равномерно-случайным становится рассеяние событий, распределение которых изучается, тем большим становится значение  $k$ . Иначе говоря, если  $s^2 \rightarrow M$ , то  $k \rightarrow \infty$ . Но  $s^2 = M$  характерно для распределения Пуассона. Таким образом, негативное биномиальное распределение в предельном случае также переходит в распределение Пуассона, т. е. все три анализируемые нами типа распределений составляют единую систему, члены которой логически связаны друг с другом.

Но вернемся к формуле (3.3). Хотя она и выражает общие закономерности негативного биномиального распределения, пользование ею для определения значения  $k$  с любой степенью точности возможно лишь в тех случаях, когда, при малых значениях  $M$ ,  $\frac{k}{M} > 6$ , а при больших значениях  $M$ , когда  $k > 13$ , и при промежуточных значениях  $M$ , когда

$$\frac{(k+M)(k+2)}{M} \geq 15. \quad (3.4)$$

Поэтому после получения начального значения  $k$  необходимо при помощи указанных неравенств оценить, можно ли применять полученное значение  $k$  для дальнейших расчетов или следует использовать другие методы его определения.

Метод пропорций нулевого члена может быть использован в тех случаях, когда

$$(M + 0.17) \left( \frac{f_0}{N} - 0.32 \right) > 0.20, \quad (3.5)$$

где  $f_0$  — количество незараженных животных, а  $\frac{f_0}{N}$  — нулевой член распределения, т. е. доля незараженных животных в выборке. Он заключается в решении уравнения

$$k = \frac{\lg \frac{N}{f_0}}{\lg \left( 1 + \frac{M}{k} \right)}, \quad (3.6)$$

которое легко выводится из уравнений  $\frac{f_0}{N} = p^k$  (см. далее формулу 3.15) и  $p = \frac{k}{k+M}$ .

Уравнение (3.6) сравнительно просто решается методом итерации, т. е. повторных решений со все более точными результатами, если взять за основу начальное значение  $k$ , полученное из формулы (3.3), и последовательно повторять решение уравнения (3.6), подставляя каждый раз в правую часть уравнения значения  $k$ , полученные в предыдущем решении, до тех пор, пока получае-

мые значения  $k$  не начнут совпадать в пределах требуемой точности, т. е. требуемого количества значащих цифр после запятой. Чтобы получить три значащие цифры, обычно требуется не более 5–7 повторений.

Если неравенство (3.5) не имеет места, применяется метод максимального подобия, дающий точное значение  $k$  в любом случае, но и наиболее трудоемкий. Он основан на решении уравнения

$$\sum \left( \frac{A_x}{x+k} \right) = N \ln \left( 1 + \frac{M}{k} \right), \quad (3.7)$$

где  $A_x$  — накопленные для каждого класса  $x$  частоты, т. е. в наших случаях общее количество животных, зараженных паразитами, во всех больших по значению  $x$  классах. Например, если в стаде крупного рогатого скота оказались заражены одной личинкой обыкновенного овода 31 животное, двумя — 8, тремя — 2, четырьмя — 1 и пятью — 1 животное, то для нулевого класса  $x_0$  накопленная частота равна общему количеству всех зараженных животных  $A_{x_0} = 43$ , для 1-го класса (и соответственно для остальных)  $A_{x_1} = 12$ ,  $A_{x_2} = 4$ ,  $A_{x_3} = 2$ ,  $A_{x_4} = 1$ . Если начальное значение  $k$  равно, например, 0.3, то

$$\begin{aligned} \sum \left( \frac{A_x}{x+k} \right) &= \frac{A_{x_0}}{0+k} + \frac{A_{x_1}}{1+k} + \frac{A_{x_2}}{2+k} + \frac{A_{x_3}}{3+k} + \frac{A_{x_4}}{4+k} = \\ &= \frac{43}{0.3} + \frac{12}{1.3} + \frac{4}{2.3} + \frac{2}{3.3} + \frac{1}{4.3}. \end{aligned}$$

Решение уравнения (3.7), по существу также основанное на методе итерации, заключается в последовательном подборе все более приближающихся к истинному значению  $k$ , так, чтобы разность

$$\sum \left( \frac{A_x}{x+k} \right) - N \ln \left( 1 + \frac{M}{k} \right) = z \quad (3.8)$$

стремилась к нулю. Для этого сначала определяют  $z_1$ , подставляя начальное значение  $k_1$ . Затем берут произвольное другое значение  $k_2$ , выбирая его, однако, с таким расчетом, чтобы при подстановке этого второго значения в уравнение (3.8)  $z_2$  имело бы знак, противоположный  $z_1$ . Если  $z_1$  имело положительный знак, т. е.  $\sum \frac{A_x}{x+k}$  была больше  $N \ln \left( 1 + \frac{M}{k_1} \right)$ , то необходимо выбрать новое значение  $k_2$ , так, чтобы  $z_2$  стало отрицательным, т. е.  $N \ln \left( 1 + \frac{M}{k_2} \right)$  стало больше  $\sum \left( \frac{A_x}{x+k} \right)$ . Для этого значение  $k_2$  нужно увеличить по сравнению с начальным. Наоборот, если  $z_2$  оказалось отрицательным, то  $k_2$  нужно брать меньшим, чтобы получить положительное  $z_2$ . Когда это достигнуто, истинное значение  $k$ , очевидно,

будет находиться в интервале между  $k_1$  и  $k_2$ . Далее  $k_3$  определяют уже методом интерполяции, подставляя полученные значения  $k$  и  $z$  в интерполяционную формулу

$$k_3 = k_1 + \frac{z_1(k_2 - k_1)}{z_1 - z_2}, \quad (3.9)$$

составленную из расчета, что  $z=0$  при истинном значении  $k$ . Очевидно, что в случае, когда  $z_1$  отрицательно, а  $z_2$  положительно, формула (3.9) примет вид

$$k_3 = k_2 + \frac{z_2(k_1 - k_2)}{z_2 - z_1}. \quad (3.10)$$

Определив значение  $k_3$ , подставляют его в уравнение (3.8) и снова решают его. Значение  $z_3$  будет по абсолютной величине уже меньшим, чем  $z_1$  и  $z_2$ . После этого снова повторяют всю операцию, добиваясь того, чтобы  $z$  наиболее приближалось к нулю при заданной точности определения истинного значения  $k$ . Дальнейшие пояснения будут даны на конкретном примере.

Следующий этап работы состоит в определении величины ошибок параметров распределения. Ошибка  $M$  вычисляется по обычной формуле для среднего арифметического:

$$m_M = \pm \sqrt{\frac{s^2}{N}}. \quad (3.11)$$

Ошибка  $k$  — как квадратный корень из его вариансы ( $V$ )

$$m_k = \pm \sqrt{V_k}, \quad (3.12)$$

а варианса при вычислении  $k$  методом максимального подобия по формуле

$$V_k = \frac{k_3 - k_4}{z_4 - z_3}, \quad (3.13)$$

где  $k_3$  и  $k_4$  — ближайшие, расположенные по обе стороны от истинного значения  $k$  величины, получаемые в процессе решения уравнения (3.7), а  $z_3$  и  $z_4$  — соответствующие этим значениям разности.

Если  $k$  вычислялось методом пропорции нулевого члена, то

$$V_k = \frac{(1-R)^{-k} - 1 - kR}{N[-\ln(1-R) - R]^2}, \quad (3.14)$$

где  $R = \frac{M}{k+M}$ ,  $\ln$  — натуральный логарифм.

После определения параметров распределения и их ошибок вычисляют теоретический ряд распределения по формуле его разложения:

$$\begin{aligned} p^k(1-q)^{-k} &= p^k + p^k kq + p^k \frac{k(k+1)q^2}{2!} + p^k \frac{k(k+1)(k+2)q^3}{3!} + \dots + \\ &+ p^k \frac{k(k+1)(k+2)\dots(k+n-1)q^n}{n!}, \end{aligned} \quad (3.15)$$

где  $n$  — количество классов распределения, т. е. в наших случаях количество групп животных с одинаковым числом паразитов на каждом. Эти расчеты значительно облегчаются, если учесть, что в каждый последующий член разложения составной частью входит величина предыдущего. Так, если мы обозначим значение нулевого члена  $P_0$ , первого —  $P_1$  и т. д., то

$$P_0 = p^k; \quad P_1 = P_0 kq; \quad P_2 = P_1 (k+1)q \cdot \frac{1}{2}; \quad P_3 = P_2 (k+2)q \cdot \frac{1}{3};$$

$$P_4 = P_3 (k+3)q \cdot \frac{1}{4}; \quad \dots \quad P_n = P_{n-1} (k+n-1)q \cdot \frac{1}{n}. \quad (3.16)$$

Рассчитав теоретический ряд распределения изучаемой величины и перемножив каждый из его членов на размер выборки  $N$ , мы получим уже конкретные количества незараженных животных, зараженных одной особью паразита, двумя и т. д., которые надлежит сравнить с фактически зарегистрированными и оценить методом  $\chi^2$  (стр. 15) вероятность случайности имеющихся расхождений.

Посмотрим теперь на конкретных примерах, как производятся все вышеуказанные расчеты.

1. Возьмем распределение личинок обыкновенного подкожного овода в стаде взрослых коров отделения «Пионер» совхоза им. Тельмана Тосненского района Ленинградской области весной 1965 г. Все стадо, 411 голов, содержавшееся в нескольких скотных дворах, выпасалось в сезон 1964 г. на смежных пастбищах и потому находилось в одинаковых условиях заражения яйцами овода, что позволяет объединить всех животных в одну выборку и изучать распределение личинок по стаду в целом. В этом стаде в июне—июле 1965 г. производился сбор зрелых, выходящих для окукления личинок овода, и потому в течение указанного времени все они находились под постоянным наблюдением, что позволило получить наиболее полные данные о их общем количестве и распределении по стаду. Результаты учетов оказались следующими: незараженных животных в стаде было 211, зараженных одной личинкой — 88 животных, двумя личинками — 48, тремя — 19, четырьмя — 14, пятью — 12, шестью — 5, семью — 2, восьмью — 3, девятью — 4, десятью — 2, одиннадцатью — 2, двенадцатью — 0, тринадцатью личинками — 1 животное. Расположим эти данные в виде таблицы распределения животных по классам интенсивности заражения оводом и определим  $M$ ,  $s^2$  и  $k_1$  (см. табл. 12).

Точность до четвертого знака необходима для того, чтобы избежать накопления ошибок округления при многочисленных операциях с параметрами распределения и вспомогательными величинами. Только для экспоненты  $k$ , учитывая сложность методов вычисления ее точного значения, можно ограничиться тремя знаками.

Таблица 12

$x$	$f$	$fx$	$fx^2$	$x$	$f$	$fx$	$fx^2$
0	211	—	—	8	3	24	192
1	88	88	88	9	4	36	324
2	48	96	192	10	2	20	200
3	19	57	171	11	2	22	242
4	14	56	224	12	0	0	0
5	12	60	300	13	1	13	169
6	5	30	180				
7	2	14	98				
				$N = 411$	$\sum fx = 516$	$\sum fx^2 = 2380$	

$$M = \frac{\sum fx}{N} = \frac{516}{411} = 1.2554;$$

$$s^2 = \frac{\sum fx^2 - (\sum fx)^2 / N}{N-1} = \frac{2380 - \frac{516^2}{411}}{410} = \frac{2380 - \frac{266256}{411}}{410} =$$

$$= \frac{2380 - 647.8248}{410} = \frac{1732.1752}{410} = 4.2248;$$

$$k_1 = \frac{M^2}{s^2 - M} = \frac{1.2554^2}{4.2248 - 1.2554} = \frac{1.57602916}{2.9694} = 0.531.$$

Для анализа распределений крайне необходима счетная машина типа «Земтрон-214» (САР-II) или малая настольная электронная машина типа «Вега», «Орбита», «Искра-11», «Искра-12», «Элка-22» или «Земтрон-220». При наличии машины все расчеты по одному распределению, включающему не более 20—30 членов, занимают не более 1 рабочего дня, конечно при наличии достаточных навыков в обращении с машиной и хорошем владении технической расчетов. Для вычисления квадратов и квадратных корней чисел очень удобны «Таблицы Барлоу».

Проверим теперь, пригоден ли использованный нами метод вычисления  $k$  для данного конкретного случая. Подставим полученные величины в формулу неравенства (3.4):

$$\frac{(0.531 + 1.2554)(0.531 + 2)}{1.2554} = \frac{1.7864 \times 2.531}{1.2554} = 4.52.$$

Поскольку 4.52 много меньше 15, очевидно, что метод моментов здесь непригоден и нужны иные методы для определения истинного значения  $k$ .

Применяя неравенство (3.5), выясним возможность использования в данном случае метода пропорции нулевого члена:

$$(1.26 + 0.17) \left( \frac{211}{411} - 0.32 \right) = 1.43 (0.51 - 0.32) = 0.27.$$

$0.27 > 0.20$  — неравенство соблюдается. Следовательно, указанный метод здесь применим.

Подставим теперь значение  $k_1$  в уравнение (3.6) и решим его, чтобы получить  $k_2$ .

$$k_2 = \frac{\lg \frac{N}{f_0}}{\lg \left( 1 + \frac{M}{k} \right)} = \frac{\lg \frac{411}{211}}{\lg \left( 1 + \frac{1.2554}{0.531} \right)} = \frac{\lg 1.9479}{\lg 3.3642} = \frac{0.2895667}{0.5268818} = 0.550.$$

Полученное значение  $k_2$  подставим снова в уравнение (3.6) и повторим решение.

$$k_3 = \frac{\lg 1.9479}{\lg \left( 1 + \frac{1.2554}{0.550} \right)} = \frac{\lg 1.9479}{\lg 3.2825} = \frac{0.2895667}{0.5162047} = 0.561,$$

и т. д. до тех пор, пока значения  $k$  в пределах трех знаков не совпадут в двух следующих друг за другом решениях:

$$k_4 = \frac{\lg 1.9479}{\lg \left( 1 + \frac{1.2554}{0.561} \right)} = \frac{\lg 1.9479}{\lg 3.2378} = \frac{0.2895667}{0.5102500} = 0.567;$$

$$k_5 = \frac{\lg 1.9479}{\lg \left( 1 + \frac{1.2554}{0.567} \right)} = \frac{\lg 1.9479}{\lg 3.2141} = \frac{0.2895667}{0.5070594} = 0.571;$$

$$k_6 = \frac{\lg 1.9479}{\lg \left( 1 + \frac{1.2554}{0.571} \right)} = \frac{\lg 1.9479}{\lg 3.1986} = \frac{0.2895667}{0.5049599} = 0.573;$$

$$k_7 = \frac{\lg 1.9479}{\lg \left( 1 + \frac{1.2554}{0.573} \right)} = \frac{\lg 1.9479}{\lg 3.1909} = \frac{0.2895667}{0.5039142} = 0.575;$$

$$k_8 = \frac{\lg 1.9479}{\lg \left( 1 + \frac{1.2554}{0.575} \right)} = \frac{\lg 1.9479}{\lg 3.1833} = \frac{0.2895667}{0.5028776} = 0.576;$$

$$k_9 = \frac{\lg 1.9479}{\lg \left( 1 + \frac{1.2554}{0.576} \right)} = \frac{\lg 1.9479}{\lg 3.1795} = \frac{0.2895667}{0.5023588} = 0.576.$$

Мы можем убедиться в том, что новое, более точное значение  $k$ , которое мы получили, значительно отличается от первоначального.

Существует еще один метод решения уравнения (3.6) с применением интерполяции. Он заключается в следующем.

Из уравнения (3.6) следует:

$$k \lg \left( 1 + \frac{M}{k} \right) = \lg \frac{N}{f_0}.$$

Решение тогда заключается в том, чтобы найти такое значение  $k$  в пределах требуемой точности, при котором разность  $k \lg(1 + \frac{M}{k}) - \lg \frac{N}{f_0}$  была бы наименьшей. Вычислим сначала значение правой части уравнения:  $\lg \frac{411}{211} = \lg 1.9479 = 0.2895667$ .

$k = 0.531$ . Тогда  $1 + \frac{M}{k} = 3.3642$ ,  $\lg(1 + \frac{M}{k}) = 0.5268818$ ,  $k \lg(1 + \frac{M}{k}) = 0.2797742$ . Выпишем полученные величины в первую строчку таблицы (табл. 13).

Таблица 13

$k$	$1 + \frac{M}{k}$	$\lg(1 + \frac{M}{k})$	$k \lg(1 + \frac{M}{k})$
0.531	3.3642	0.5268818	0.2797742
0.600	3.0923	0.4902816	0.2941690
0.578	3.1720	0.5013332	0.2897706
0.577	3.1757	0.5018395	0.2895614

Значение  $k \lg(1 + \frac{M}{k})$  оказалось меньшим по сравнению с  $\lg \frac{N}{f_0}$ . Возьмем теперь какое-либо произвольное значение  $k_2$ , при котором  $k_2 \lg(1 + \frac{M}{k})$  было бы большим, чем  $\lg \frac{N}{f_0}$ . Пусть это будет 0.600. Вычислим новые значения всех интересующих нас величин и запишем их в ту же таблицу. Теперь нужно построить интерполяционную формулу для определения величины, на которую нужно увеличить  $k_1$ , чтобы получить наибольшее приближение значения  $k \lg(1 + \frac{M}{k})$  к значению  $\lg \frac{N}{f_0}$ . Обозначим  $k \lg(1 + \frac{M}{k}) = y$ ,  $\lg(\frac{N}{f_0}) = a$ . Тогда на основании простой пропорции получаем:

$$k_3 - k_1 = \frac{(k_2 - k_1)(a - y_1)}{y_2 - y_1}; \quad k_3 = k_1 + \frac{(k_2 - k_1)(a - y_1)}{y_2 - y_1}.$$

Сделаем вычисления:

$$\begin{aligned} k_3 &= 0.531 + \frac{(0.600 - 0.531)(0.2895667 - 0.2797742)}{0.2941690 - 0.2797742} = 0.531 + \\ &+ \frac{0.069 \times 0.0097925}{0.0143948} = 0.531 + \frac{0.0006757}{0.0143948} = 0.531 + 0.047 = 0.578. \end{aligned}$$

Произведем новые расчеты с полученным значением  $k_3$  и запишем их в таблицу. Величина  $k_3 \lg(1 + \frac{M}{k_3})$  уже близка к тре-

буемой, но все-таки превышает ее начиная с четвертого знака, что может дать ошибку в третьем знаке значения  $k$ . Поэтому повторим интерполяцию, беря вместо  $k_2$  и  $y_2$ , соответственно,  $k_3$  и  $y_3$ .

$$\begin{aligned} k_4 &= 0.531 + \frac{(0.578 - 0.531)(0.2895667 - 0.2797742)}{0.2897706 - 0.2797742} = 0.531 + \\ &+ \frac{0.047 \times 0.0097925}{0.0099964} = 0.531 + \frac{0.0004602}{0.0099964} = 0.531 + 0.046 = 0.577. \end{aligned}$$

Проделав необходимые расчеты с  $k_4$ , получаем  $k_4 \lg(1 + \frac{M}{k_4}) = 0.2895614$  с расхождением относительно  $\lg \frac{N}{f_0} = 0.2895667$  в шестом знаке. Очевидно, что в дальнейшем уточнении величины  $k$  нет необходимости, так как оно лежит уже за пределами принятой точности ее вычисления. Значение  $k = 0.577$  более точно по сравнению с величиной  $k = 0.576$ , полученной при первом решении, поскольку из сопоставления полученных величин  $k_3 \lg(1 + \frac{M}{k_3})$  и  $k_4 \lg(1 + \frac{M}{k_4})$  с  $\lg \frac{N}{f_0}$  видно, что при дальнейшем уточнении значения  $k$  оно оказалось бы лежащим между 0.577 и 0.578. Итак, окончательное значение  $\hat{k} = 0.577$ .

Вычислим теперь вариансу и ошибку  $\hat{k}$ , используя для этого формулы (3.14) и (3.12). Поскольку  $m_k = \pm \sqrt{V_k}$ , а точность вычисления ошибки должна быть, во всяком случае, не меньше, чем самой экспоненты  $k$ , то, чтобы получить значение  $m_k$  с точностью до трех знаков, вычисление  $V_k$  надо вести с точностью до шести знаков. Определим сначала

$$R = \frac{1.255474}{0.577 + 1.255474} = \frac{1.255474}{1.832474} = 0.685125.$$

Тогда

$$\begin{aligned} V_k &= \frac{(1 - 0.685125)^{-0.577} - 1 - 0.577 \times 0.685125}{411 [-\ln(1 - 0.685125) - 0.685125]^2} = \\ &= \frac{0.314875^{-0.577} - 1.395317}{411 (-\ln 0.314875 - 0.685125)^2}. \end{aligned}$$

$$\begin{aligned} \lg 0.314875^{-0.577} &= \lg 0.314875 \times -0.577 = 1.4981381 \times -0.577 = -0.5018619 \times \\ &\times -0.577 = 0.2895743 = \lg 1.947935. \end{aligned}$$

Чтобы определить значение  $-\ln 0.314875$ , где  $\ln$  — натуральный логарифм, нужно взять десятичный логарифм данного числа и умножить его на переходный множитель  $\ln 10^* = 2.302585$ . Тогда  $-\ln 0.314875 = -\lg 0.314875 \times 2.302585 = -(1.4981381) \times 2.302585 = -(-0.5018619) \times 2.302585 = 1.155580$ . Итак,

\* Если  $\lg a = x$ , то  $\ln a = \ln 10^x = x \ln 10$ .

$$V_k = \frac{1.947935 - 1.395317}{411(1.155580 - 0.685125)^2} = \frac{0.552618}{411(0.470455)^2} =$$

$$= \frac{0.552618}{411 \times 0.221328} = \frac{0.552618}{90.965808} = 0.005745;$$

$$m_k = \pm \sqrt{0.005745} = \pm 0.075796 \approx 0.076.$$

Перейдем теперь к вычислению теоретического ряда распределения (3.15), для чего сначала необходимо определить вспомогательные параметры  $p$  и  $q$ . Согласно (3.1 и 3.2),

$$p = \frac{k}{k+M} = \frac{0.577}{0.577 + 1.2554} = \frac{0.577}{1.8324} = 0.315;$$

$$q = 1 - p = 0.685.$$

Нулевой член распределения  $P_0 = p^k$ ;  $p^k = 0.315^{0.577}$ .

$$\lg 0.315^{0.577} = \lg 0.315 \times 0.577 = 1.4983106 \times 0.577 = -0.5016894 \times 0.577 =$$

$$= -0.2894748 = 1.7105252 = \lg 0.51348.$$

$$P_0 = 0.51348.$$

По формуле (3.16) определяем остальные члены распределения по 13-й включительно, поскольку в нашем распределении имеется 13 классов.

$$P_1 = P_0 k q = 0.51348 \times 0.577 \times 0.685 = 0.20295;$$

$$P_2 = P_1 (k+1) q \cdot \frac{1}{2} = 0.20295 \times 1.577 \times 0.685 \times 0.5 * = 0.10962;$$

$$P_3 = P_2 (k+2) q \cdot \frac{1}{3} = 0.10962 \times 2.577 \times 0.685 \times 0.3333 = 0.06449;$$

$$P_4 = P_3 (k+3) q \cdot \frac{1}{4} = 0.06449 \times 3.577 \times 0.685 \times 0.25 = 0.03951;$$

$$P_5 = P_4 (k+4) q \cdot \frac{1}{5} = 0.03951 \times 4.577 \times 0.685 \times 0.2 = 0.02477;$$

$$P_6 = P_5 (k+5) q \cdot \frac{1}{6} = 0.02477 \times 5.577 \times 0.685 \times 0.1667 = 0.01578;$$

$$P_7 = P_6 (k+6) q \cdot \frac{1}{7} = 0.01578 \times 6.577 \times 0.685 \times 0.1429 = 0.01016;$$

$$P_8 = P_7 (k+7) q \cdot \frac{1}{8} = 0.01016 \times 7.577 \times 0.685 \times 0.125 = 0.00659;$$

\* При работе на счетной машине гораздо удобнее заменить простые дроби десятичными, так как тогда в случаях, подобных нашему, вычисление сводится к получению последовательного ряда произведений, из которых записываются только интересующие нас результаты. Значения обратных величин натурального ряда чисел, т. е.  $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$  и т. д. в десятичных дробях, можно найти в «Таблицах Барлоу» или других математических таблицах.

$$P_9 = P_8 (k+8) q \cdot \frac{1}{9} = 0.00659 \times 8.577 \times 0.685 \times 0.1111 = 0.00430;$$

$$P_{10} = P_9 (k+9) q \cdot \frac{1}{10} = 0.00430 \times 9.577 \times 0.685 \times 0.1 = 0.00282;$$

$$P_{11} = P_{10} (k+10) q \cdot \frac{1}{11} = 0.00282 \times 10.577 \times 0.685 \times 0.0909 = 0.00186;$$

$$P_{12} = P_{11} (k+11) q \cdot \frac{1}{12} = 0.00186 \times 11.577 \times 0.685 \times 0.0833 = 0.00123;$$

$$P_{13} = P_{12} (k+12) q \cdot \frac{1}{13} = 0.00123 \times 12.577 \times 0.685 \times 0.0769 = 0.00081.$$

Чтобы перейти от распределения вероятностей, к распределению частот изучаемой случайной величины, т. е. к численности личинок овода в стаде, умножим каждую из полученных вероятностей на размер выборки  $N = 411$ .

$$P_0 N = 0.51348 \times 411 = 211.04; *$$

$$P_1 N = 0.20295 \times 411 = 83.41;$$

$$P_2 N = 0.10962 \times 411 = 45.05;$$

$$P_3 N = 0.06449 \times 411 = 26.51;$$

$$P_4 N = 0.03951 \times 411 = 16.24;$$

$$P_5 N = 0.02477 \times 411 = 10.18;$$

$$P_6 N = 0.01578 \times 411 = 6.49;$$

$$P_7 N = 0.01016 \times 411 = 4.18;$$

$$P_8 N = 0.00659 \times 411 = 2.71;$$

$$P_9 N = 0.00430 \times 411 = 1.77;$$

$$P_{10} N = 0.00282 \times 411 = 1.16;$$

$$P_{11} N = 0.00186 \times 411 = 0.76;$$

$$P_{12} N = 0.00123 \times 411 = 0.51;$$

$$P_{13} N = 0.00081 \times 411 = 0.33.$$

Определим теперь  $\chi^2$  и вероятность случайности расхождения между эмпирическим рядом и теоретическим (см. табл. 14).

Для определения числа степеней свободы надо уменьшить число сопоставляемых величин на 3, так как негативное биномиальное распределение имеет два параметра.

Итак,  $\chi^2 = 6.8651$ , а  $P > 30\%$ . Эта вероятность достаточно велика, и потому мы можем заключить, что рассматриваемый пример распределения личинок овода хорошо моделируется негативным биномиальным распределением. Заметим еще, что более одной трети суммарной величины  $\chi^2$  составляет его значение для последней группы классов, т. е. для групп животных с количеством личинок от 9 до 13 на каждом животном. Именно в этой группе расхождение между фактическим и теоретическим распределением было наибольшим. Расхождения могли возникнуть из-за того, что в стаде оказалось несколько животных, выпасавшихся в предыдущем сезоне в другой местности, где плотность популяции овода, а соответственно и заражение животных его яйцами были значительно большими.

Далее, хотя для анализа распределения личинок отбирались лишь взрослые животные, от 4 лет и старше, могли быть допущены

\* Не следует смущаться наличием дробных чисел. Дело в том, что в теоретическом ряду распределения частоты изменяются непрерывно (графически функция распределения имеет вид плавной кривой). Точность в два знака после запятой обязательна для более точного определения  $\chi^2$ .

Таблица 14

$x$	$f_e$	$f_{NB}$	$f_e - f_{NB}$	$(f_e - f_{NB})^2$	$\frac{(f_e - f_{NB})^2}{f_{NB}}$
0	211	211.04	0.04	0.0016	0.0000
1	88	83.41	4.59	21.0681	0.2526
2	48	45.05	2.95	8.7025	0.1932
3	19	26.51	7.51	56.4001	2.1275
4	14	16.24	2.24	5.0176	0.3090
5	12	10.18	1.82	3.3124	0.3254
6	5	6.49	1.49	2.2201	0.3421
7	2	4.18	1.89	3.5721	0.5184
8	3	2.71			
9	4	1.77			
10	2	1.16			
11	2	0.76	+ 4.53	3.81	14.5161
12	0	0.51	+ 0.66		2.7969
13	1	0.33	+ 5.19		
	411	+ 410.34			
		0.66			
		411.00			
				$\chi^2 = 6.8651$	
				$v = 9 - 3 = 6$	
				$0.30 < P < 0.40$	

ошибки в определении возраста, в результате чего оказались включенными более молодые животные, отличающиеся от взрослых меньшей сопротивляемостью к заражению оводом.

Все это означает, что наша выборка могла быть в какой-то мере гетерогенной, что всегда ведет к более или менее существенным ошибкам в определении изучаемых статистических закономерностей. Поэтому при получении исходного материала для статистического анализа и при его первичной обработке следует тщательно взвешивать возможность случайных включений в исследуемую популяцию чужеродных элементов и отсеивать все частные случаи, в отношении которых имеются обоснованные сомнения в их принадлежности к исследуемой совокупности.

Познакомимся теперь на практическом примере с техникой определения экспоненты  $k$  методом максимального подобия.

2. Рассмотрим распределение плероцеркоидов ремненца *Digramma interrupta* Rud. среди годовиков карася *Carassius auratus* (Bloch) по материалам М. Н. Дубининой, полученным в 1958 г. в той же локальной популяции карася, из которой в 1959 г. были сделаны выборки, проанализированные нами выше (стр. 20 и др.). В 1958 г. по условиям работы отлов всех возрастных групп популяции не производился. В выборке оказались практически только годовики и трехлетки. Кроме того, при отлове молодняка исполь-

зовались не сети, а только сачок, что приводило к увеличению в выборке доли менее подвижных зараженных плероцеркоидами ремненца рыб. В итоге среди исследованных 67 годовиков оказались незараженными 27 карасей, зараженными одним плероцеркоидом — 20, двумя — 8, тремя — 8, четырьмя — 2, пятью — 1 и шестью — 1 карась. Определим по обычной схеме  $M$ ,  $s^2$ ,  $\frac{s^2}{M}$  и  $k$  (табл. 15).

Таблица 15

$x$	$f$	$fx$	$fx^2$	$x$	$f$	$fx$	$fx^2$
0	27	—	—	4	2	8	32
1	20	20	20	5	1	5	25
2	8	16	32	6	1	6	36
3	8	24	72			$N = 67$	$\Sigma fx = 79$
							$\Sigma fx^2 = 217$

$$M = \frac{79}{67} = 1.1791;$$

$$s^2 = \frac{217 - \frac{79^2}{67}}{66} = \frac{217 - \frac{6241}{67}}{66} = \frac{217 - 93.1493}{66} = \frac{123.8507}{66} = 1.8765;$$

$$\frac{s^2}{M} = \frac{1.8765}{1.1791} = 1.5914;$$

$$k_1 = \frac{1.1791^2}{1.8765 - 1.1791} = \frac{1.3903}{0.6974} = 1.99.$$

Отношение  $\frac{s^2}{M} = 1.5914$  у годовиков относительно близко к единице по сравнению, например, с данными для просмотренных в том же году трехлеток:  $N = 145$ ,  $M = 0.2345$ ,  $\frac{s^2}{M} = 4.5019$ , но все же отличие от единицы достаточно велико, и поэтому трудно ожидать хорошей сходимости распределения плероцеркоидов у годовиков с распределением Пуассона. Можно предполагать, что негативное биномиальное распределение окажется здесь более подходящей статистической моделью. Некоторые недостатки в методике получения выборки в данном случае не должны нас смущать, потому что они помогут нам понять, как сказываются такие недостатки на параметрах распределения.

Проверим по формуле (3.4) возможность применения метода моментов для определения  $k$ .

$$\frac{(1.18 + 1.99)(1.99 + 2)}{1.18} = \frac{3.17 \times 3.99}{1.18} = \frac{12.65}{1.18} = 10.72.$$

$10.72 < 15$ , неравенство не соблюдается. Следовательно, метод моментов здесь не применим и значение  $k_1 = 1.99$  недостаточно точно.

Определение методом пропорции нулевого члена здесь также непригодно, так как  $(1.18 + 0.17) \left( \frac{27}{67} - 0.32 \right) = 1.35 \times 0.08 = 0.108$ ,

а  $0.108 < 0.20$ . Поэтому мы должны применить метод максимального подобия, основанный на решении уравнения (3.7).

Вычисления начинают с определения накопленных частот  $A_x$  для каждого класса распределения. Что такое накопленные частоты, было объяснено выше. Полученные значения  $A_x$  записывают в таблицу (табл. 16).

Таблица 16

x	$A_x$	$\frac{A_x}{x+k}$			
		$k_1 = 2.0$	$k_2 = 1.85$	$k_3 = 1.882$	$k_4 = 1.880$
0	40	20.00000	21.62162	21.25398	21.27660
1	20	6.66667	7.01754	6.93962	6.94444
2	12	3.00000	3.11688	3.09119	3.09278
3	4	0.80000	0.82474	0.81934	0.81967
4	2	0.33333	0.34188	0.34002	0.34014
5	1	0.14286	0.14599	0.14531	0.14535
$\sum \left( \frac{A_x}{x+k} \right)$		30.94286	33.06865	32.58946	32.61869

Возьмем теперь за исходное  $k_1 = 1.99$ , округлим его до 2.00 и получим  $z_1$  (3.8). Сначала вычислим значение  $\sum \left( \frac{A_x}{x+k} \right)$ . Запишем значения  $\frac{A_x}{x+k}$  для каждого класса в таблицу рядом со значением  $A_x$ . Точность вычисления  $\frac{A_x}{x+k}$  должна быть не менее пяти знаков после запятой, потому что при меньшей точности нельзя обеспечить определение  $k$  с точностью до трех знаков. Суммируя значения  $\frac{A_x}{x+k_1}$  во всех классах, получаем  $\sum \left( \frac{A_x}{x+k_1} \right) = 30.94286$ .

Теперь нужно подсчитать, чему равно  $N \ln \left( 1 + \frac{M}{k_1} \right)$ . Поскольку  $\ln$  — натуральный логарифм, при переходе к десятичным логарифмам оно принимает вид  $N \times 2.302585 \times \lg \left( 1 + \frac{M}{k_1} \right)$ , где  $N \times 2.302585$  будет постоянным множителем при расчетах каждого отдельного распределения. В нашем случае  $N \times 2.302585 = 67 \times 2.302585 = 154.273195$ . Далее,  $\lg \left( 1 + \frac{M}{k_1} \right) = \lg \left( 1 + \frac{1.1791}{2.0} \right) = \lg 1.5896 =$

$= 0.2012879$ , а  $N \ln \left( 1 + \frac{M}{k_1} \right) = 154.273195 \times 0.2012879 = 31.05333$ . Определим теперь  $z_1$  (табл. 17).

Таблица 17

	$k_1 = 2.0$	$k_2 = 1.85$	$k_3 = 1.882$	$k_4 = 1.880$
$\sum \left( \frac{A_x}{x+k} \right)$	30.94286	33.06865	32.58946	32.61869
$N \ln \left( 1 + \frac{M}{k} \right)$	31.05333	33.03835	32.59084	32.61885
$z$	-0.11047	0.03030	-0.00138	0.00013

$z_1 = -0.11047$ . Необходимо теперь подобрать такое значение  $k_2$ , чтобы  $z_2$  было положительным. Для этого, как уже указывалось, величину  $k_2$  нужно уменьшить по сравнению с  $k_1$ . Пусть  $k_2 = 1.85$ . Повторяем все вычисления и определяем  $z_2$ :  $\sum \left( \frac{A_x}{x+k_2} \right) = 33.06865$ ;  $\lg \left( 1 + \frac{M}{k_2} \right) = \lg \left( 1 + \frac{1.1791}{1.85} \right) = \lg 1.6374 = 0.2141548$ ;  $N \ln \left( 1 + \frac{M}{k^2} \right) = 154.273195 \times 0.2141548 = 33.03835$ ;  $z^2 = 0.03030$ .

Итак, мы имеем  $z_1$  и  $z_2$  с разными знаками. Теперь можно приступить к интерполяции. Интерполяционная формула строится исходя из пропорции: если при увеличении величины  $k$  на  $(k_1 - k_2)$  величина  $z$  уменьшается на  $(z_2 - z_1)$ , то на сколько нужно увеличить  $k_2$ , чтобы  $z_2$  уменьшилось на  $z_2$ , т. е. стало равным нулю.

Отсюда  $k_3 = k_2 + \frac{(k_1 - k_2) z_2}{z_2 - z_1}$ .

Подставляя соответствующие величины, получаем:

$$k_3 = 1.85 + \frac{(2.0 - 1.85) 0.03030}{0.03030 - (-0.11047)} = 1.85 + \frac{0.0045450}{0.14077} = 1.85 + 0.032 = 1.882.$$

Теперь повторим все расчеты, используя новое значение  $k$ . Получаем:  $\sum \left( \frac{A_x}{x+k_3} \right) = 32.58946$ ;  $\lg \left( 1 + \frac{M}{k_3} \right) = \lg \left( 1 + \frac{1.1791}{1.882} \right) = \lg 1.6265 = 0.2112541$ ;  $N \ln \left( 1 + \frac{M}{k_3} \right) = 154.273195 \times 0.2112541 = 32.59084$ .  $z_3 = -0.00138$ .

Значение  $z_3$  уже очень невелико по абсолютной величине, но пока нельзя гарантировать, что, изменяя третий знак в значении  $k$ , мы не сможем получить еще большее приближение  $z$  к нулю. Поэтому повторяем интерполяцию:

$$k_4 = k_2 + \frac{(k_3 - k_2)z_2}{z_2 - z_3} = 1.85 + \frac{(1.882 - 1.85)0.03030}{0.03030 - (-0.00138)} = \\ = 1.85 + \frac{0.00096960}{0.03168} = 1.85 + 0.030 = 1.880.$$

Снова повторяем расчеты, чтобы получить  $z_4$ .

$$\sum \left( \frac{A_x}{x+k_4} \right) = 32.61898; \quad \lg \left( 1 + \frac{M}{k_4} \right) = \lg \left( 1 + \frac{1.1791}{1.880} \right) = \\ = \lg 1.62718 = 0.2114356; \quad N \ln \left( 1 + \frac{M}{k_4} \right) = 154.273195 \times 0.2114356 = \\ = 32.61885. \quad z_4 = 0.00013.$$

Теперь уже очевидно, что для дальнейшего уменьшения абсолютной величины  $z$  нам пришлось бы внести изменения уже в четвертый знак после запятой в значении  $k$ , что лежит уже за пределами принятой точности. Поэтому последнее значение  $k=1.880$  будет окончательным.

Вычисление  $k$  методом максимального подобия с первого взгляда может показаться слишком трудоемким и потому практически мало применимым. Следует, однако, иметь в виду, что даже самые обычные электромеханические счетные машины типа «Земтрон 214» позволяют производить наиболее трудоемкую операцию — вычисление  $\sum \left( \frac{A_x}{x+k} \right)$  путем накопления частных, т. е. без записи всех промежуточных результатов, что очень облегчает работу. Затрата времени на полную обработку распределения, подобного рассматриваемому, при элементарной квалификации оператора не превышает 4—5 часов.

Варианса  $V_k$ , необходимая для определения ошибки, вычисляется по формуле (3.13):

$$V_k = \frac{k_3 - k_4}{z_4 - z_3} = \frac{1.882 - 1.880}{0.00013 - (-0.00138)} = \frac{0.00200}{0.00151} = 1.32450; \\ m_k = \pm \sqrt{1.32450} = \pm 1.151.$$

Большая величина ошибки  $k$  связана с недостатком информации в изучаемой выборке. Впрочем, величина ошибки  $k$  не влияет на точность расчета теоретического ряда распределения. Она только означает, что в генеральной совокупности, если брать из последней выборки, аналогичные по размерам исследуемой,  $k$  может меняться в относительно широких пределах, даже при постоянстве факторов, обусловливающих данный характер распределения изучаемого объекта. Для уменьшения ошибки размер выборки пришлось бы значительно увеличить. Но тогда неминуемо пришлось бы изменить и методику ее получения и мы имели бы уже более точное представление о распределении плероцеркоидов

в группе годовиков карася. Возможно, что оно более приближалось бы к распределению Пуассона, как это оказалось в ранее рассмотренном примере. Но не исключено, что в той же локальной популяции карася, но годом раньше, в 1958 г., оно носило характер, более близкий к негативному биному. Как мы видели при анализе вероятностных характеристик процесса заражения популяции рыб ремнеком, для сеголеток всегда будет характерно распределение с постоянной вероятностью заражения, а для двухлеток и более старших групп вероятность будет изменчива, годовики же в этом отношении занимают промежуточное положение. Поэтому изменчив будет и характер распределения в этой возрастной группе, о чем свидетельствует большая величина ошибки  $k$ .

Рассчитаем теперь теоретический ряд распределения, для чего определим сначала величины  $p$  и  $q$  (3.1 и 3.2).

$$p = \frac{1.880}{1.880 + 1.1791} = \frac{1.880}{2.0591} = 0.615; \quad q = 1 - p = 0.385; \\ P_0 = p^k = 0.615^{1.880}. \\ \lg 0.615 \times 1.880 = \lg 1.7888751 \times 1.880 = -0.2111249 \times 1.880 = -0.3969148 = \\ = 1.6030852 = \lg 0.400945. \\ P_0 = 0.40094; * \\ P_1 = 0.40094 \times 1.880 \times 0.385 = 0.29020; \\ P_2 = 0.29020 \times 2.880 \times 0.385 \times 0.5 = 0.16089; \\ P_3 = 0.16089 \times 3.880 \times 0.385 \times 0.333 = 0.08003; \\ P_4 = 0.08003 \times 4.880 \times 0.385 \times 0.25 = 0.03759; \\ P_5 = 0.03759 \times 5.880 \times 0.385 \times 0.2 = 0.01702; \\ P_6 = 0.01702 \times 6.880 \times 0.385 \times 0.1667 = 0.00753; \\ 0.40094 \times 67 = 26.864; \quad 0.03759 \times 67 = 2.519; \\ 0.29020 \times 67 = 19.443; \quad 0.01702 \times 67 = 1.140; \\ 0.16089 \times 67 = 10.780; \quad 0.00753 \times 67 = 0.505. \\ 0.08003 \times 67 = 5.362;$$

Сравним теоретический ряд с эмпирическим, определим значение  $\chi^2$  и вероятность случайности расхождений  $P$  (см. табл. 18).

$P > 30\%$ , и потому мы можем утверждать, что и в этом случае негативное биномиальное распределение достаточно хорошо моделирует распределение плероцеркоидов ремнека в исследуемой выборке, но большая величина ошибки  $k$  заставляет относиться с некоторой осторожностью к этому выводу и обращает внимание на необходимость получения дополнительного материала.

3. Рассмотрим теперь еще один пример расчетов параметров и теоретического ряда негативного биномиального распределения,

\* Чтобы избежать накопления ошибок округления, возможного при всяких значительных по объему расчетах, принимают правило: если за последним знаком округляемой величины стоит 5, то в случаях, когда последний знак нечетное число, его увеличивают на единицу, а четные числа оставляют без изменений.

Таблица 19

$x$	$f$	$fx$	$fx^2$	$x$	$f$	$fx$	$fx^2$
0	0	0	0	14	1	14	196
1	3	3	3	15	5	75	1125
2	4	8	16	16	3	48	768
3	5	15	45	17	1	17	289
4	2	8	32	18	1	18	324
5	5	25	125	19	0	0	0
6	4	24	144	20	0	0	0
7	3	21	147	21	1	21	441
8	3	24	192	22	1	22	484
9	3	27	243	23	1	23	529
10	1	10	100	24	1	24	576
11	2	22	242	25	1	25	625
12	3	36	432				
13	3	39	507				
				$N = 57$	$\Sigma fx = 549$	$\Sigma fx^2 = 7585$	

$$M = \frac{549}{57} = 9.6316;$$

$$s^2 = \frac{7585 - \frac{549^2}{57}}{56} = \frac{7585 - 5287.7368}{56} = \frac{2297.2632}{56} = 41.0226;$$

$$\frac{s^2}{M} = 4.2592;$$

$$k_1 = \frac{9.6316^2}{41.0226 - 9.6316} = \frac{92.7677}{31.3910} = 2.955.$$

Итак, окончательное значение  $k$ , обозначаемое  $\hat{k} = 2.665$ , так как из сопоставления  $z_4$  и  $z_5$  видно, что дальнейшее уточнение находится уже за пределами принятой точности.

Определяем ошибку  $\hat{k}$ :

$$V_{\hat{k}} = \frac{k_4 - k_5}{z_5 - z_4} = \frac{2.666 - 2.665}{0.00104 + 0.00205} = \frac{0.001}{0.00309} = 0.323625;$$

$$m_{\hat{k}} = \pm \sqrt{0.323625} = \pm 0.5689 \approx \pm 0.569.$$

Вычислим теоретический ряд вероятностей:

$$p = \frac{2.665}{2.665 + 9.6316} = \frac{2.665}{12.2966} = 0.217; q = 0.783;$$

$$P_0 = p^k = 0.217^{2.665}; \\ \lg 0.217 \times 2.665 = 1.3364597 \times 2.665 = -0.6635403 \times 2.665 = \\ = -1.7683349 - 2.2316651 = \lg 0.0170477.$$

$$P_0 = 0.01705.$$

$P_1 = 0.01705 \times 2.665 \times 0.783 = 0.03557$  и т. д. по приведенной в предыдущих примерах схеме.

Таблица 18

$x$	$f_e$	$f_{NB}$	$f_e - f_{NB}$	$(f_e - f_{NB})^2$	$\frac{(f_e - f_{NB})^2}{f_{NB}}$
0	27	26.864	0.14	0.0196	0.0001
1	20	19.443	0.56	0.3136	0.0161
2	8	10.780	2.78	7.7284	0.7169
3	3	5.362	2.64	6.9696	1.3003
4	2	2.519			
5	1	1.440			
6	1	0.505			
7					
8					
9					
10					
11					
12					
13					
67		$+ \frac{66.613}{0.387}$		$\chi^2 = 2.1005$ $v = 5 - 3 = 2$ $0.30 < P < 0.40$	

поучительный в том смысле, что он дает возможность убедиться в наличии определенной закономерности в хаотическом, на первый взгляд, фактическом материале.

Возьмем распределение личинок клеща *Ixodes ricinus* L. на взрослых самках лесной мыши *Apodemus sylvaticus* L. по материалам Е. Ф. Сосниной, полученным летом 1961 г. в Крымском заповеднике. Распределение личинок изучалось по отдельным половозрастным группам популяции мыши, потому что предварительный анализ (Соснина, 1969) показал значительные различия в заражении этих групп. Фактические данные о распределении представлены в табл. 19.

Распределение явно перерассеянное, так как  $\frac{s^2}{M}$  много больше единицы. Проверяем возможность использования полученного значения  $k$  для расчета теоретического ряда распределения. Подставляя это значение в неравенство (3.4), получаем:

$$\frac{(2.96 + 9.63)(2.96 + 2)}{9.63} = \frac{12.59 \times 4.96}{9.63} = 6.48.$$

Поскольку  $6.48 < 15$ , полученное значение  $k$  для дальнейших расчетов непригодно.

Метод пропорции нулевого члена в данном случае также неприменим, потому что нулевой член в эмпирическом распределении отсутствует; не зараженных личинками мышей в выборке не было. Следовательно, должен быть применен метод максимального подобия. Запись расчетов для вычисления  $k$  этим методом приведена в табл. 20, в той форме, в какой обычно ведут такие расчеты, и без пояснений, так как они были уже даны выше.

Таблица 20

Личинки *Ixodes ricinus* L. на взрослых самках *Apodemus sylvaticus* L.  
(Крымский заповедник, 1961 г., материалы Е. Ф. Сосниной)

$$N = 57; M = 9.6316; k_1 = 2.955$$

$x$	$A_x$	$k_1 = 2.95$	$k_2 = 2.60$	$k_3 = 2.67$	$k_4 = 2.666$	$k_5 = 2.665$
0	57					
1	54					
2	50					
3	45					
4	43					
5	38					
6	34					
7	31					
8	28					
9	25					
10	24					
11	22					
12	19					
13	16					
14	15					
15	10					
16	7					
17	6					
18	5					
19	5					
20	5					
21	4					
22	3					
23	2					
24	1					
25	—					
$\left  \sum \left( \frac{A_x}{x+k} \right) \right $		82.08972	88.42902	87.06489	87.14154	87.16070

$$N \ln = 57 \times 2.302585 = 131.247345;$$

$$\lg \left( 1 + \frac{9.6316}{2.95} \right) = \lg 4.2699 = 0.6299088;$$

$$\lg \left( 1 + \frac{9.6316}{2.60} \right) = \lg 4.7045 = 0.6725135;$$

$$\lg \left( 1 + \frac{9.6316}{2.67} \right) = \lg 4.6073 = 0.6634465;$$

$$\lg \left( 1 + \frac{9.6316}{2.666} \right) = \lg 4.6128 = 0.6639646;$$

$$\lg \left( 1 + \frac{9.6316}{2.665} \right) = \lg 4.6141 = 0.6640870.$$

Таблица 20 (продолжение)

	$k_1 = 2.95$	$k_2 = 2.60$	$k_3 = 2.67$	$k_4 = 2.666$	$k_5 = 2.665$
$\sum \left( \frac{A_x}{x+k} \right)$	82.08972	88.42902	87.06489	87.14154	87.16070
$- N \ln \left( 1 + \frac{M}{k} \right)$	82.67386	88.26561	87.07559	87.14359	87.15966
$z$	-0.58414	0.16341	-0.01070	-0.00205	0.00104

$$k_3 = 2.60 + \frac{0.35 \times 0.16341}{0.16341 + 0.58414} = 2.60 + \frac{0.0571935}{0.74755} = 2.60 + 0.07 = 2.67;$$

$$k_4 = 2.60 + \frac{0.07 \times 0.16341}{0.16341 + 0.01070} = 2.60 + \frac{0.0114387}{0.17411} = 2.60 + 0.0657 = 2.666;$$

$$k_5 = 2.60 + \frac{0.066 \times 0.16341}{0.16341 + 0.00205} = 2.60 + \frac{0.01078506}{0.16546} = 2.60 + 0.0652 = 2.665.$$

$$\hat{k} = 2.665.$$

В итоге получаем ряд вероятностей, каждую из которых затем умножаем на размер выборки  $N = 57$ , чтобы получить теоретический ряд частот:

$$\begin{aligned}
 P_0N &= 0.01705 \times 57 = 0.972; & P_{13}N &= 0.03965 \times 57 = 2.260; \\
 P_1N &= 0.03557 \times 57 = 2.027; & P_{14}N &= 0.03472 \times 57 = 1.979; \\
 P_2N &= 0.05104 \times 57 = 2.909; & P_{15}N &= 0.03024 \times 57 = 1.722; \\
 P_3N &= 0.06214 \times 57 = 3.542; & P_{16}N &= 0.02612 \times 57 = 1.489; \\
 P_4N &= 0.06891 \times 57 = 3.928; & P_{17}N &= 0.02245 \times 57 = 1.280; \\
 P_5N &= 0.07192 \times 57 = 4.099; & P_{18}N &= 0.01922 \times 57 = 1.096; \\
 P_6N &= 0.07196 \times 57 = 4.102; & P_{19}N &= 0.01636 \times 57 = 0.932; \\
 P_7N &= 0.06977 \times 57 = 3.977; & P_{20}N &= 0.01387 \times 57 = 0.791; \\
 P_8N &= 0.06600 \times 57 = 3.762; & P_{21}N &= 0.01172 \times 57 = 0.668; \\
 P_9N &= 0.06123 \times 57 = 3.490; & P_{22}N &= 0.00986 \times 57 = 0.562; \\
 P_{10}N &= 0.05592 \times 57 = 3.187; & P_{23}N &= 0.00828 \times 57 = 0.472; \\
 P_{11}N &= 0.05041 \times 57 = 2.873; & P_{24}N &= 0.00694 \times 57 = 0.396; \\
 P_{12}N &= 0.04493 \times 57 = 2.561; & P_{25}N &= 0.00580 \times 57 = 0.331.
 \end{aligned}$$

Сопоставляем эмпирический ряд частот с теоретическим, определяем  $\chi^2$  и  $P$  (табл. 21).

Вероятность случайности имеющихся отклонений эмпирического ряда от теоретического превышает 70%, т. е. если во всей популяции лесной мыши распределение личинок *I. ricinus* строго соответствует негативному биномиальному распределению, то при случайных выборках такого же размера, как исследуемая, более чем в 70% случаев мы имели бы аналогичные по величине отклонения. Совпадение очень хорошее.

На приведенном графике (стр. 53), где показано теоретическое и эмпирическое распределение, ясно видно, как вызванная случайными причинами мультимодальность (многовершинность) эмпирического

Таблица 21

$x$	$f_e$	$f_{NB}$	$f_e - f_{NB}$	$(f_e - f_{NB})^2$	$\frac{(f_e - f_{NB})^2}{f_{NB}}$
0	0	0.972			
1	3	2.027			
2	4	2.909			
3	5	3.542			
4	2	3.928	7.470	0.47	0.2209
5	5	4.099			
6	4	4.102	8.201	0.80	0.6400
7	3	3.977			
8	3	3.762	7.739	1.74	3.0276
9	3	3.490			
10	1	3.187	6.677	2.68	7.1824
11	2	2.873			
12	3	2.561	5.434	0.43	0.1849
13	3	2.260			
14	1	1.979			
15	5	1.722	5.961	3.04	9.2416
16	3	1.489			
17	1	1.280			
18	1	1.096			
19	0	0.932	5.588	0.59	0.3481
20	0	0.791			
21	1	0.668			
22	1	0.562			
23	1	0.472	+2.429		
24	1	0.396	+1.593		
25	1	0.331	4.022		
57		+ 55.407			$\chi^2 = 3.6606$
		- 1.593			$v = 9 - 3 = 6$
		57.000			$0.70 < P < 0.80$

распределения выравнивается кривой теоретического ряда, вскрывающей характер закономерности, лежащей в его основе. Этот пример показывает, что в случаях, когда фактические данные о распределении численности паразита в популяции хозяина представляются на первый взгляд хаотичными, не следует отказываться от поисков закономерности. Если таковая будет найдена, она поможет вскрыть и причины, приводящие к тому или иному типу распределения, т. е. глубже понять лежащие в основе биологи-

ческие процессы. О причинах, определяющих характер распределения численности личинок клещей на грызунах, уже говорилось в начале этого раздела, при расшифровке биологического смысла негативного биномиального распределения.

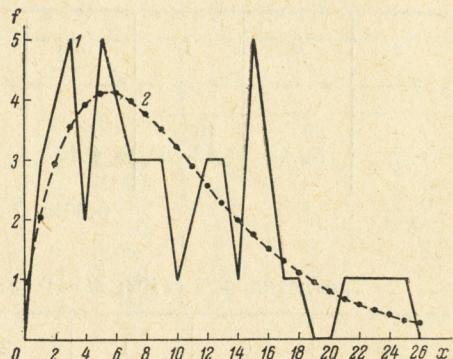
Для лучшего освоения техники расчетов параметров и теоретического ряда негативного бинома примерами для упражнений могут служить данные о распределении плероцеркоидов *D. interrupta* и личинок *I. ricinus*, приведенные в следующем разделе.

Подробнее с теорией и методами использования негативного биномиального распределения в биологической статистике можно познакомиться по работам Бл исса и Фишера (Bliss a. Fisher, 1953), Бл исса (Bliss, 1958), Бл исса и Оуэна (Bliss a. Owen, 1958) и Энскомба (Anscombe, 1950). Для помощи в расчетах очень полезны таблицы негативного биномиального распределения (Williamson a. Bretherton, 1963).

#### 4. ЧТО МОЖЕТ ДАТЬ АНАЛИЗ РАСПРЕДЕЛЕНИЙ ЧИСЛЕННОСТИ ПАРАЗИТОВ В ПОПУЛЯЦИЯХ ИХ ХОЗЯЕВ

Применение математических методов в биологии целесообразно в тех случаях, когда с их помощью можно овладеть биологической информацией, получение которой иными методами менее экономно или вообще невозможно. Что же нового может дать изучение распределений паразитов в популяциях их хозяев?

Прежде всего таким методом могут проверяться гипотезы о вероятностных характеристиках процессов заражения животных паразитами. Примерами могут служить вышеупомянутые данные о распределениях плероцеркоидов ремнеца *Digamma interrupta* Rud. в популяциях серебряного карася и личинок *Ixodes ricinus* L. в популяции лесной мыши. В дополнение можно привести следующие данные. Распределение плероцеркоидов *D. interrupta* у двух- и трехлеток, а также во всей популяции карася моделируется негативным биномиальным распределением со следующими параметрами и рядами (табл. 22—24).



Распределение личинок *Ixodes ricinus* L. на взрослых самках *Apodemus sylvaticus* L. (Крым, 1961 г., материалы Е. Ф. Сосиной).

$N=57$ ;  $M = 9.6316$ ;  $k = 2.665$ ;  $0.70 < P < 0.80$ .  
1 — эмпирическое распределение; 2 — теоретический ряд негативного биномиального распределения.

Таблица 22

Двухлетки —  $N = 90$ ;  $M = 0.3520 \pm 0.0933$ ;  $k = 0.371 \pm 0.180$ 

$x$	0	1	2	3	4	5	6
$f_e$	69	15	3	0	1	0	2
$f_{NB}$	69.48	12.55	4.19	1.61	0.66	0.28	0.12

$$0.30 < P < 0.40$$

Таблица 23

Трехлетки —  $N = 102$ ;  $M = 0.2255 \pm 0.0736$ ;  $k = 0.118 \pm 0.058$ 

$x$	0	1	2	3	4	5
$f_e$	90	6	3	2	0	1
$f_{NB}$	89.93	6.96	2.55	1.18	0.60	0.33

$$0.50 < P < 0.60$$

Таблица 24

Вся популяция —  $N = 281$ ;  $M = 0.2705 \pm 0.0433$ ;  $k = 0.336 \pm 0.111$ 

$x$	0	1	2	3	4	5	6
$f_e$	230	37	9	2	1	1	1
$f_{NB}$	230.42	34.53	10.29	3.57	1.33	0.51	0.21

$$0.50 < P < 0.60$$

Вероятность случайности расхождений между эмпирическим рядом и теоретическим во всех случаях высока. Поэтому можно считать, что ход наших рассуждений об условиях заражения рыб процеркоидами ременца и развитии плероцеркоидов достоверен с высокой степенью вероятности.

Рассмотрим теперь распределение личинок клеща *I. ricinus* в остальных половозрастных группах популяции лесной мыши по материалам Е. Ф. Сосниной. Как уже было показано (стр. 48), распределение личинок этого клеща на самках *Apodemus sylvaticus* хорошо моделируется негативным биномиальным распределением. Такой же результат был получен и при изучении распределения личинок *I. ricinus* на взрослых самцах и молодняке (табл. 25, 26).

Проанализируем эти данные. Прежде всего обращает на себя внимание значительно большая численность личинок на взрослых

Таблица 25

Взрослые самцы

 $N = 97$ ;  $M = 18.9077 \pm 3.0854$ ;  $k = 2.252 \pm 0.346$ 

$x$	$f_e$	$f_{NB}$	$x$	$f_e$	$f_{NB}$	$x$	$f_e$	$f_{NB}$
0	0	0.625	8	6	3.565	16	4	3.186
1	1	1.256	9	5	3.629	17	3	3.056
2	4	1.825	10	2	3.649	18	4	2.924
3	2	2.312	11	5	3.632	19	3	2.783
4	1	2.713	12	4	3.582	20	2	2.642
5	2	3.031	13	4	3.508	21	4	2.501
6	4	3.275	14	3	3.414	22	1	2.364
7	2	3.450	15	3	3.307	23	2	2.229

Таблица 25 (продолжение)

$x$	$f_e$	$f_{NB}$	$x$	$f_e$	$f_{NB}$	$x$	$f_e$	$f_{NB}$
24	1	2.097	36	0	0.887	47	2	0.354
25	1	1.968	37	2	0.819	48	1	0.324
26	2	1.845	38	0	0.755	49	0	0.299
27	2	1.724	39	1	0.694	50	0	0.273
28	0	1.608	40	1	0.640	51	0	0.249
29	1	1.500	41	0	0.590	52	0	0.228
30	2	1.395	42	0	0.543	53	0	0.209
31	0	1.299	43	0	0.500	54	1	0.190
32	0	1.204	44	1	0.459	55	1	0.175
33	2	1.116	45	1	0.421	56	0	0.159
34	0	1.034	46	1	0.386	57	1	0.146
35	0	0.958						

$$0.80 < P < 0.90$$

самцах по сравнению с самками и молодняком, причем это различие статистически достоверно. У молодых мышей различий в интенсивности заражения особей разного пола не было, и поэтому при анализе распределения они были объединены. Большая зараженность самцов объясняется их большей подвижностью и, соответственно, большими размерами посещаемых ими участков, а также, возможно, и гормональными отличиями (Соснина, 1969). Но сравнение второго параметра распределения — экспоненты  $k$  у взрослых самцов и самок не подтверждает последнего предположения. Величины  $k$  (самцы  $2.252 \pm 0.346$ , самки  $2.665 \pm 0.569$ ) практически одинаковы, а если бы у самок имела место повышенная сопротивляемость к заражению личинками, это должно было бы сказаться в изменении значения этого параметра. В самом деле,  $k$  — показатель агрегированности изучаемой случайной величины

Таблица 26

Молодняк, самцы и самки

$$N = 71; M = 10.1268 \pm 1.0329; k = 1.571 \pm 0.306$$

$x$	$f_e$	$f_{NB}$	$x$	$f_e$	$f_{NB}$	$x$	$f_e$	$f_{NB}$
0	1	3.030	14	3	2.402	27	0	0.463
1	6	4.121	15	0	1.882	28	0	0.409
2	6	4.587	16	1	1.694	29	1	0.361
3	7	4.726	17	3	1.515	30	0	0.318
4	5	4.675	18	1	1.355	31	1	0.281
5	1	4.509	19	3	1.207	32	0	0.248
6	2	4.276	20	2	1.074	33	0	0.219
7	6	4.005	21	0	0.956	34	0	0.192
8	5	3.715	22	0	0.849	35	1	0.169
9	2	3.419	23	1	0.759	36	0	0.149
10	3	3.129	24	0	0.669	37	0	0.131
11	4	2.849	25	1	0.552	38	0	0.115
12	2	2.615	26	1	0.525	39	1	0.101
13	1	2.334						

$$0.40 < P < 0.50$$

чины в пределах учетной единицы, т. е. в нашем случае интенсивности скопления личинок на одной особи хозяина. Но скопление зависит от двух факторов: частоты встреч мышей с личинками и защитных реакций, определяющих величину возможного максимального скопления.

Биологический смысл экспоненты  $k$  в данном случае в том и заключается, что она дает возможность статистически характеризовать этот максимум, что было продемонстрировано на примере подкожных оводов (Бреев, 1968б). Поэтому, если бы самки лесной мыши обладали, в отличие от самцов, более выраженной защитной реакцией, это неминуемо привело бы к увеличению абсолютной величины  $k$ , чего мы не наблюдаем. Но у мышат, на которых средняя численность личинок клеща была одного порядка с их численностью на самках, величина  $k = 1.571 \pm 0.306$  заметно ниже по сравнению с  $k$  у взрослых мышей. Следовательно, у молодняка в принципе возможно гораздо большее скопление личинок на одной особи, и поэтому можно предполагать наличие каких-то защитных реакций, усиливающихся с возрастом независимо от пола.

Для проверки этого предположения рассмотрим еще случаи распределения личинок *I. ricinus* и *Haemaphysalis concinna* Koch в популяции полевки *Microtus arvalis* Pall., также на основании материалов Е. Ф. Сосниной, полученных одновременно в той же местности. Поскольку хорошая сходимость распределения личинок клещей на грызунах с негативным биномиальным распределением была продемонстрирована уже достаточно убедительно,

ограничившимся теперь лишь параметрами распределения и вероятностью случайности отличий эмпирического распределения от негативного биномиального.

Таблица 27

Параметры распределения личинок клещей в половозрастных группах популяций грызунов (негативное биномиальное распределение)  
(Крымский заповедник, 1961 г., материалы Е. Ф. Сосниной)

Вид клеща и его хозяина	Параметры распределения	Половозрастные группы популяции хозяина		
		взрослые самцы	взрослые самки	молодняк, самцы и самки
<i>Ixodes ricinus</i> L. на <i>Apodemus sylvaticus</i> L.	<i>N</i>	97	57	71
	<i>M</i>	18.9072 $\pm$ 3.0854	9.6316 $\pm$ 0.8484	10.1268 $\pm$ 1.0329
	<i>k</i>	2.252 $\pm$ 0.346	2.665 $\pm$ 0.569	1.571 $\pm$ 0.306
	<i>P</i>	> 0.80	> 0.70	> 0.40
<i>Ixodes ricinus</i> L. на <i>Microtus arvalis</i> Pall.	<i>N</i>	16	18	54
	<i>M</i>	10.6250 $\pm$ 2.4813	2.2556 $\pm$ 0.7149	2.1111 $\pm$ 0.5636
	<i>k</i>	0.9225 $\pm$ 0.0889	1.034 $\pm$ 0.546	0.3252 $\pm$ 0.0669
	<i>P</i>	> 0.50	> 0.20	> 0.20
<i>Haemaphysalis concinna</i> Koch на <i>Microtus arvalis</i> Pall.	<i>N</i>	29	64	73
	<i>M</i>	0.9655 $\pm$ 0.4016	1.1875 $\pm$ 0.3806	0.8493 $\pm$ 0.2908
	<i>k</i>	0.171 $\pm$ 0.086	0.1197 $\pm$ 0.0303	0.1166 $\pm$ 0.0400
	<i>P</i>	> 0.40	> 0.60	> 0.30

Как видно из данных табл. 27, заражение полевок личинками *I. ricinus* в анализируемых выборках было менее сильным по сравнению с заражением лесной мыши, но все отношения между параметрами распределения в разных половозрастных группах остались такими же. Здесь мы также имеем увеличение значения  $k$  у взрослых полевок, позволяющее предполагать усиление каких-то защитных реакций с возрастом.

Иные отношения складываются при заражении полевок личинками *H. concinna*. Численность личинок здесь наименьшая, что говорит о малой плотности популяции этого клеща в природе в период, когда была получена выборка. Различий в обилии личинок на отдельных половозрастных группах нет, так как имеющиеся небольшие отличия находятся в пределах ошибки. Нет отличий и в значениях  $k$ . Однако делать на этом основании вывод, что у полевок в отношении личинок *H. concinna* отсутствует увеличение сопротивляемости с возрастом, было бы преждевременно, потому что возможно, что при столь малой общей численности личинок такая сопротивляемость просто не проявляется. Чтобы убедиться в допустимости такого предположения, необходимо определить параметры распределения в условиях значительно большей плотности популяции паразита.

Кроме того, окончательный вывод об усилении с возрастом защищенных реакций у грызунов по отношению к личинкам того или иного вида иксодовых клещей можно сделать только на основании экспериментальных данных. Однако статистический анализ дает достаточно информации для обоснования такого эксперимента и вместе с тем позволяет сопоставлять экспериментальные данные с результатами наблюдений в природных популяциях. Таково еще одно из возможных применений анализа распределения численности паразитов в популяциях их хозяев.

Другим примером может быть решение вопросов о функциональной связи между средней численностью паразита и экстенсивностью заражения им популяции хозяина. Установление такой связи существенно прежде всего с точки зрения накопления материалов о динамике численности популяций паразитов и о факто-рах ее обуславливающих. Уже говорилось о малой изученности этих вопросов в паразитологии и о методических трудностях накопления необходимых исходных материалов. Не требует особых разъяснений, что возможность получать данные о численности паразита только на основании процента зараженных особей хозяина значительно упростила бы эту работу, особенно в отношении эндопаразитов.

Принципиальное решение вопроса сводится к определению зависимости между величиной нулевого члена распределения (т. е. долей особей в популяции, не зараженных паразитом, которую легко определить, зная долю зараженных особей) и математическим ожиданием численности паразита. Сама постановка вопроса уже показывает, что он может быть решен только на вероятностной основе, исходя из развитых выше представлений о численности вида как случайной величине и на основе статистической модели распределения. Проиллюстрируем сказанное на примерах рассмотренных нами распределений.

В распределении Пуассона, имеющем всего один параметр, такая связь устанавливается наиболее просто. В самом деле, формула нулевого члена этого распределения (2.11):

$$P_0 = \frac{M^0}{0!} e^{-M} = e^{-M} = \frac{1}{e^M}.$$

Обозначим экстенсивность заражения, выраженную в долях единицы,  $E$ . Тогда,  $P_0 = 1 - E$  и  $1 - E = \frac{1}{e^M}$ , а  $e^M = \frac{1}{1 - E}$ . Но  $e$  — основание натуральных логарифмов, обозначаемых знаком  $\ln$ , в отличие от  $\lg$ , принятого для десятичных логарифмов. Напомним, что логарифмом числа называют показатель степени, в которую нужно возвести основание логарифма, чтобы получить данное число. Следовательно, выражение  $e^M = \frac{1}{1 - E}$  можно записать так:

$$M = \ln\left(\frac{1}{1 - E}\right). \quad (4.1)$$

Но это и будет формула функциональной связи между экстенсивностью заражения  $E$ , выраженной в долях единицы, и средней численностью паразита  $M$  в случаях, когда распределение паразита в популяции хозяина подчиняется закону Пуассона.

Проверим справедливость полученной формулы на рассмотренных выше примерах. Наиболее полное соответствие распределению Пуассона ( $P > 0.90$ ) мы имели в случае распределения плероцеркоидов ремнечка *D. interrupta* в группе годовиков серебряного карася (см. стр. 21). Необходимые для наших расчетов данные об этом распределении следующие:  $N = 41$ ;  $E = \frac{8}{41} = 0.19512$ ;  $1 - E = 0.80488$ ;  $\frac{1}{1 - E} = 1.2424$ . Из формулы (4.1)  $M = \ln 1.2424$ ;  $\ln 1.2424 = \lg 1.2424 \times 2.302585 = 0.0942614 \times 2.302585 = 0.2170$ .

Фактически средняя численность плероцеркоидов в группе годовиков  $M = 0.2195$ , т. е. ошибка нашего определения составляет менее 1%.

В группе сеголеток карася сходимость с распределением Пуассона была меньшей ( $P > 0.50$ ). Исходные данные:  $N = 49$ ;  $E = \frac{11}{49} = 0.22449$ ;  $\frac{1}{1 - E} = 1.2895$ ;  $M = \ln 1.2895 = \lg 1.2895 \times 2.302585 = 0.1104213 \times 2.302585 = 0.2542$ .

Фактическое значение  $M = 0.2653$ . В этом случае, как и следовало ожидать, ошибка оказалась большей, но и здесь она составляет менее 4%.

Выведем теперь формулы связи между экстенсивностью заражения и средней численностью паразита для случаев, когда распределение подчиняется закономерности биномиального и негативного биномиального распределений.

Для биномиального распределения  $(q + p)^n$  характерны следующие зависимости:  $p + q = 1$ ,  $M = np$ . Нулевой член распределения  $P_0 = q^n$ , а так как  $n = \frac{M}{p}$ , то  $P_0 = q^{\frac{M}{p}}$  или  $1 - E = q^{\frac{M}{p}}$ . Логарифмируя это уравнение, получаем:

$$\begin{aligned} \lg(1 - E) &= \frac{M}{p} \lg q; \quad \frac{M}{p} = \frac{\lg(1 - E)}{\lg q}; \quad M = p \frac{\lg(1 - E)}{\lg q}; \\ M &= p \frac{\lg(1 - E)}{\lg(1 - p)}. \end{aligned} \quad (4.2)$$

Поскольку биномиальное распределение имеет два параметра, то и формулу связи экстенсивности заражения со средней численностью помимо величины экстенсивности должен входить второй параметр —  $p$ , т. е. вероятность заражения каждой особи хозяина одной особью паразита. Очевидно, что в условиях, приводящих к разной экстенсивности заражения, эта вероятность не остается постоянной, и возникает новый вопрос о характере ее изменений при меняющейся экстенсивности. Мы остановимся на нем далее.

Негативное биномиальное распределение  $p^k(1-q)^{-k}$ , как уже об этом говорилось, характеризуется такой зависимостью:  $p = \frac{k}{k+M}$ . Нулевой член распределения  $P_0 = p^k$ , но  $P_0 = 1 - E$ , тогда  $1 - E = p^k$ ;  $1 - E = \left(\frac{k}{k+M}\right)^k$ , или, если взять обратные величины,

$$\frac{1}{1-E} = \left(\frac{k+M}{k}\right)^k; \quad \frac{M+k}{k} = \sqrt[k]{\frac{1}{1-E}}; \quad \frac{M}{k} + 1 = \sqrt[k]{\frac{1}{1-E}};$$

$$M = k \left( \sqrt[k]{\frac{1}{1-E}} - 1 \right). \quad (4.3)$$

Чтобы определить среднюю численность паразита на основании величины экстенсивности заражения, здесь также нужно знать еще и величину второго параметра распределения — экспоненты  $k$  и также возникает вопрос о постоянстве или изменчивости этого параметра в условиях разной экстенсивности заражения, т. е. при разных отношениях плотностей популяций хозяина и паразита.

Очевидно, что для решения этого вопроса необходимо исследование параметра в представительных выборках из популяций хозяина во всем спектре существующих плотностей популяций изучаемого паразита — от самых малых до самых больших. Но прежде чем приступить к этой работе, следует убедиться в том, что сам характер распределения при разной интенсивности заражения и в разных половозрастных группах популяции хозяина остается неизменным. На примере заражения серебряного карася плероцеркоидами *D. interrupta* мы видели, что тип распределения не всегда остается постоянным: в группе сеголеток — распределение Пуассона, годовиков — Пуассона или негативное биномиальное, в старших группах — только последнее.

После того как будет установлено, в каких пределах экстенсивности и интенсивности заражения хозяина данным видом паразита или, иначе говоря, в каких пределах плотностей популяции паразита характер распределения его численности остается постоянным, можно приступить к анализу изменения параметров распределения при разных уровнях численности паразита.

Изменения эти неизбежны, так как даже в двух равных по объему и проведенных одновременно выборках из одной и той же популяции хозяина полное совпадение параметров распределения будет скорее исключением, чем правилом. Но в этом случае изменения будут вызваны случайными причинами и различия будут лежать в пределах ошибки. Иное положение возникает, когда дело касается параметров распределения при разных уровнях численности паразита. В принципе здесь возможны два случая.

1. Несмотря на разную численность паразита различия не выходят за пределы ошибок, т. е. параметр остается практически

постоянным, и тогда следует определить его общее значение. Напомним, что параметрами, о которых идет речь, являются: в случае биномиального распределения  $p$ , т. е. вероятность заражения особи хозяина особью паразита и выживания последней до стадии развития, на которой ведется учет; в случае негативного биномиального —  $k$ , т. е. мера агрегированности, мера возможного скопления особей паразита на одной особи хозяина. Такой случай возможен лишь тогда, когда характер и интенсивность действия каждого из факторов, определяющих изучаемый параметр, остаются постоянными независимо от плотности популяции паразита. Поскольку этим постулируется полное отсутствие каких-либо реакций со стороны отдельных особей и популяции хозяина на усиливающееся заражение паразитом, такие случаи, вероятно, будут скорее исключением, чем правилом.

2. При повышении или понижении уровня численности паразита происходит изменение параметров его распределения, связанное с изменением хозяино-паразитных отношений или других факторов, влияющих на численность. Такие случаи будут более обычны уже потому, что в системе хозяин — паразит беспредельное размножение паразита мало вероятно, а обычно наличие какого-то максимума интенсивности заражения. Но если существует такой максимум, то по мере приближения к нему уровня численности паразита вероятность выживания отдельных его особей будет уменьшаться, а сам максимум будет означать какой-то предел агрегированности. Следовательно, будет происходить и закономерное изменение интересующих нас параметров распределений в зависимости от экстенсивности и интенсивности заражения. Количественная формулировка этой закономерности будет вместе с тем статистическим выражением изменения действия совокупности факторов, влияющих на численность паразита в системе хозяин — паразит при разных уровнях численности. Так как существенное значение в определении характера изменения параметра будет играть указанный выше предел интенсивности заражения, а его величина неизбежно специфична для каждой пары хозяина и паразита, можно ожидать такой же специфичности и для статистической характеристики изменения параметра распределения.

Таким образом, изучение распределения численности паразита в популяциях хозяина приводит нас к очень важному и теоретически и практически вопросу о влиянии хозяино-паразитных отношений на динамику численности паразитов и о статистической оценке этого влияния.

Такой подход пока нов, и потому мы можем проиллюстрировать сказанное только одним примером. Изучение распределения личинок II и III стадии обыкновенного подкожного овода в стадах крупного рогатого скота показало, что при самых разных плотностях популяции овода оно хорошо моделируется негативным

биномиальным распределением (Бреев, 1968а, 1968б). Анализ изменений значений параметра распределения  $k$ , с деталями которого можно познакомиться в цитированных статьях, привел к выводу, что закономерность этих изменений может быть выражена уравнением прямолинейной отрицательной регрессии:

$$Y = 4.517 - 0.0436X,$$

где  $Y$  — значение  $\frac{1}{k}$  для экстенсивности заражения  $X$ , выраженной в процентах. Выборочное стандартное отклонение регрессии —  $s_{y \cdot x} = 0.4875$ .

Какую же информацию мы можем извлечь из этого уравнения? Прежде всего можно определить значение экспоненты для любой экстенсивности заражения скота оводом. Возьмем для примера экстенсивность 15%. Тогда  $Y = 4.517 - 0.0436 \times 15 = 4.517 - 0.654 = = 3.863$ . Следовательно,  $\frac{1}{k} = 3.863$ , а  $k = \frac{1}{3.863} = 0.259$ . Но это лишь среднее значение  $k$  для данной экстенсивности заражения, а нам существенно знать и возможные пределы его отклонений в отдельных частных случаях. Иначе говоря, нужно определить доверительный интервал для значений  $k$  с заданной степенью достоверности. Возьмем обычный 95%-й уровень, означающий, что в 95% случаев при выборках из генеральной совокупности при 15% экстенсивности заражения оводом значения  $k$  не будут выходить из пределов этого интервала.

Поскольку при определении доверительного интервала в прямолинейной регрессии используют критерий Стьюдента, отклонение от полученного среднего значения будет определяться произведением названного критерия для 95%-го уровня на ошибку полученного нами значения  $Y = \frac{1}{k}$ . Формула для определения

ошибки  $s_Y = s_{y \cdot x} \sqrt{1 + \frac{1}{n} + \frac{x^2}{\sum x^2}}$  (Snedecor, 1957), где  $s_{y \cdot x}$  — выборочное стандартное отклонение регрессии,  $n$  — число вариантов регрессии, т. е. число отдельных распределений, на основе которых была построена регрессия (в нашем случае  $n = 41$ ),  $x$  — отклонение рассматриваемого случая, т. е. экстенсивности заражения — 15%, от ее среднего значения для регрессии  $\bar{x} = 58.19\%$ ,  $\sum x^2$  — сумма квадратов всех отклонений  $x$  от среднего  $\bar{x}$  —  $= 25192.4587$ .

В данном частном случае

$$\begin{aligned} s_Y &= \pm 0.4875 \sqrt{1 + \frac{1}{41} + \frac{43.192}{25192.4587}} = \\ &= \pm 0.4875 \sqrt{1 + 0.0244 + 0.0740} = \pm 0.4875 \times 1.0480 = \pm 0.5109. \end{aligned}$$

Чтобы определить величину критерия  $t_{0.95}$  из таблицы критериев Стьюдента, надо знать число степеней свободы. В прямолинейной регрессии значение интересующей нас величины  $Y$  ограничено двумя условиями — средним значением  $Y$  и величиной коэффициента регрессии. Следовательно, для определения числа степеней свободы число вариантов регрессии ( $n = 41$ ) нужно уменьшить на  $2 : 41 - 2 = 39$ . При 39 степенях свободы  $t_{0.95} = 2.023$ . Доверительный интервал для  $Y = \frac{1}{k}$  будет равен:  $t_{0.95} \times s_Y = \pm 2.023 \times$

$\times 0.5109 = \pm 1.034$ . Таким образом, величина  $\frac{1}{k}$  может меняться в пределах от  $3.863 - 1.034 = 2.829$  до  $3.863 + 1.034 = 4.897$ , а  $k$ , соответственно, от  $\frac{1}{4.897} = 0.205$  до  $\frac{1}{2.829} = 0.353$ .

Теперь мы знаем и среднее значение  $k = 0.259$  для экстенсивности заражения 15%, и его доверительный интервал. Но ранее мы вывели формулу (4.3) функциональной связи между средней численностью паразита и экстенсивностью заражения для случаев, когда распределение паразита подчиняется закономерности негативного биномиального распределения. Зная, что  $E = 0.15$ , и подставляя полученные значения  $k$ , мы можем определить среднюю численность личинок овода в стаде при данной экстенсивности заражения и пределы ее отклонений. Произведем эти расчеты.

При  $k = 0.259$ :

$$\begin{aligned} M &= 0.259 \left( \sqrt[0.259]{\frac{1}{1 - 0.15}} - 1 \right) = 0.259 \left( \sqrt[0.259]{1.1765} - 1 \right); \\ \lg \sqrt[0.259]{1.1765} &= \lg 1.1765 : 0.259 = 0.0705919 : 0.259 = \\ &= 0.2725556 = \lg 1.8731; M = 0.259 \times 0.8731 = 0.2261. \end{aligned}$$

При  $k = 0.205$ :

$$\begin{aligned} \lg \sqrt[0.205]{1.1765} &= \lg 1.1765 : 0.205 = 0.0705919 : 0.205 = \\ &= 0.3443507 = \lg 2.2098; M = 0.205 \times 2.2098 = 0.2480. \end{aligned}$$

При  $k = 0.353$ :

$$\begin{aligned} \lg \sqrt[0.353]{1.1765} &= \lg 1.1765 : 0.353 = 0.0705919 : 0.353 = \\ &= 0.1999770 = \lg 1.5848; M = 0.353 \times 0.5848 = 0.2064. \end{aligned}$$

В итоге, округляя цифры, мы можем констатировать, что при экстенсивности заражения 15% средняя численность личинок обычновенного овода в стадах крупного рогатого скота будет равна 0.23, меняясь в отдельных случаях в пределах от 0.21 до 0.25 личинки на 1 голову.

Далее, на основе предельных значений  $k$  по формуле  $p = \frac{k}{k + M}$ ,  $q = 1 - p$  можно вычислить  $p$  и  $q$ , затем рассчитать по формуле

разложения распределения (3.15) соответствующие ряды \* и определить по сумме накопленных вероятностей классы, для которых эта сумма равна 0.99 или 0.999. Порядковый номер соответствующего класса будет обозначать максимальное число личинок овода на одном животном, которое можно встретить в стаде с вероятностью, соответственно, 0.01 в стаде из 100 голов или 0.001 в стаде из 1000 голов. Беря за основу таких расчетов доверительные пределы  $k$ , мы получим пределы колебаний максимального количества личинок на 1 голову при разной экстенсивности заражения. Проделаем необходимые вычисления для полученных предельных значений  $k$  0.205 и 0.353, которым соответствуют значения  $M$  0.2480 и 0.2064.

$$1. \ k=0.205; \ M=0.2480.$$

$$0.205 \\ p = \frac{0.205}{0.205 + 0.2480} = 0.453; \ q = 0.547.$$

$$P_0 = p^k = 0.453^{0.205}; \ \lg 0.453 \times 0.205 = 1.6560982 \times 0.205 = -0.3439018 \times 0.205 = -0.0704999 = 1.9295001 = \lg 0.85016. \ P_0 = 0.85016.$$

Будем вычислять теперь по обычной схеме (3.16) остальные члены разложения этого ряда до тех пор, пока сумма накопленных вероятностей не достигнет 0.999 (табл. 28).

Таблица 28

$x$	$P_x$	$\Sigma P_x$	$x$	$P_x$	$\Sigma P_x$
0	0.85016	0.85016	4	0.00554	0.99508
1	0.09533	0.94549	5	0.00255	0.99863
2	0.03142	0.97691	6	0.00121	0.99984
3	0.01263	0.98954			

$$2. \ k=0.353; \ M=0.2064.$$

$$0.353 \\ p = \frac{0.353}{0.353 + 0.2064} = 0.631; \ q = 0.369.$$

$$P_0 = p^k = 0.631^{0.353}; \ \lg 0.631 \times 0.353 = 1.8000294 \times 0.353 = -0.1999706 \times 0.353 = -0.0705896 = 1.9294104 = \lg 0.84998. \ P_0 = 0.84998.$$

Вычисляем остальные члены разложения (табл. 29).

Таблица 29

$x$	$P_x$	$\Sigma P_x$	$x$	$P_x$	$\Sigma P_x$
0	0.84998	0.84998	3	0.00800	0.99624
1	0.11072	0.96060	4	0.00247	0.99871
2	0.02764	0.98824	5	0.00079	0.99950

\* Можно воспользоваться и готовыми таблицами негативного биномиального распределения (Williamson a. Bretherton, 1963).

Сопоставляя значения накопленных вероятностей в полученных рядах, мы видим, что наиболее близкое к 0.99 значение они принимают в одном ряду во 2-м, в другом — в 3-м классе, а к значению 0.999, соответственно, — в 4-м и 5-м классе. Следовательно, максимальное количество личинок овода на одном животном, которое можно встретить при экстенсивности заражения 15% (в 95% случаев) с вероятностью 0.01 будет 2—3, а с вероятностью 0.001 — 4—5.

Если проделать аналогичные расчеты для всей шкалы возможной экстенсивности заражения, мы получим статистическую модель, иллюстрирующую изменения средней численности и пределов интенсивности заражения также по всей шкале их значений. Пример такой модели для личинок обыкновенного подкожного овода крупного рогатого скота приведен в табл. 30. Сопоставление модели с эмпирическими данными,ложенными в ее основу, показало хорошее их совпадение (Бреев, 1968б). Новые данные о распределении личинок того же овода, полученные после построения модели, также в большинстве совпадали с расчетными.

Нетрудно заметить, что, зная характер распределения и закономерность изменения его параметров при разных уровнях численности паразита, мы можем определить среднюю частоту встречаемости (и пределы ее отклонений) любого числа особей паразита на одной особи хозяина при любой заданной экстенсивности или средней интенсивности заражения. Ценность такого рода данных для моделирования разного рода процессов изменения численности паразитов неоспорима. Но помимо того, как мы видели выше, закономерность изменения параметров распределения является статистическим выражением влияния хозяино-паразитных отношений на численность паразита. Рассмотрим с этой точки зрения вышеприведенную зависимость изменения экспоненты негативного биномиального распределения  $k$  от экстенсивности заражения для личинок обыкновенного подкожного овода.

Используя уравнение (4.3) и табл. 30, мы можем экстраполировать статистические параметры распределения при малой плотности личинок на любые другие плотности, вычислить, какой должна быть средняя численность в условиях такой экстраполяции, а затем сравнить полученные результаты с фактически наблюдаемыми. Поскольку параметры распределения являются статистическим выражением условий существования паразита, обеспечивающих тот или иной характер распределения, т. е. в нашем случае с оводом прежде всего хозяино-паразитных отношений, мы таким способом как бы экстраполируем характер этих отношений при малой интенсивности заражения на большую.

Возьмем для примера полученное нами среднее значение экспоненты  $k=0.259$  для экстенсивности заражения 15% и вычислим, какова должна быть средняя численность личинок в стаде при экстенсивности 50 и 80%, если бы  $k$  оставалось постоянным.

1. Для экстенсивности 50%.

$$M = 0.259 \left( \sqrt[0.259]{\frac{1}{1 - 0.50}} - 1 \right) = 0.259 (\sqrt[0.259]{2} - 1);$$

$$\lg 2 : 0.259 = 0.3010300 : 0.259 = 1.1622780 = \lg 14.5304; M = 0.259 \times \\ \times 13.5304 = 3.5044.$$

2. Для экстенсивности 80%.

$$M = 0.259 \left( \sqrt[0.259]{\frac{1}{1 - 0.80}} - 1 \right) = 0.259 (\sqrt[0.259]{5} - 1);$$

$$\lg 5 : 0.259 = 0.6989700 : 0.259 = 2.6987259 = \lg 499.7191; M = 0.259 \times \\ \times 498.7191 = 129.1682.$$

Итак, при экстраполяции параметра распределения, характерного для малой плотности личинок овода (экстенсивность 15%, средняя численность 0.23 личинки на 1 голову), на условия большей плотности мы получили, что при экстенсивности 50% средняя численность должна была бы составлять 3.50, а при 80% — 129.17 личинки на 1 голову. На самом же деле она в среднем равна (табл. 30) при 50% 1.74, а при 80% 4.12 личинки на 1 голову.

Объяснение такого различия, по-видимому, заключается в существовании каких-то факторов, ограничивающих численность паразита, причем по мере возрастания плотности его популяции интенсивность их действия возрастает. В данном частном случае возрастание характеризуется следующими показателями: при увеличении численности с 0.23 до 1.75 личинки на 1 голову — в 2 раза (3.6 : 1.75), а при дальнейшем увеличении до 4.12 личинки на 1 голову — уже в 31 раз (129.17 : 4.12).

В анализируемом нами примере факторы эти скорее всего связаны с хозяино-паразитными отношениями. Подтверждением служат факты наличия у скота защитных реакций по отношению к личинкам оводов, приводящих к гибели 70% и более личинок I стадии и очень большой изменчивости сроков их развития (Бреев, 1967а, 1967б).

Таким образом, анализ изменений параметров распределения паразита при разной плотности его популяций приводит нас к важным выводам о наличии естественных факторов, регулирующих численность, и, что не менее важно, позволяет дать количественную оценку интенсивности их действия.

Приведенные примеры достаточно хорошо показывают, насколько разнообразным может быть анализ распределений численности паразитов в популяции их хозяев, какое обилие новой информации мы можем таким путем получить и к каким важным биологическим выводам эта информация приводит.

Однако как ни интересно и ни перспективно применение прекрасно разработанного математического аппарата для изучения

Таблица 30

Средние показатели и пределы интенсивности заражения крупного рогатого скота личинками II и III стадии, общирновенного подложного овода при различной экстенсивности заражения. Модель (доверительный интервал 95%). (По Брееву, 1968)

Экстенсивность заражения, %	Показатель дисперсии заражения, $k$		Индекс численности овода (личинок на 100 животных)	Интенсивность заражения (личинок на 1 животное)		Максимальное количество личинок на 1 животное с вероятностью до 0.1%
	средний	лим.		средняя	лим.	
5	0.188	0.121—0.255	6	1.18	1.14—1.28	1—2
10	0.188	0.121—0.255	14	1.41	1.30—1.68	2—3
15	0.259	0.205—0.353	23	1.51	1.37—1.65	2—4
20	0.275	0.225—0.381	34	1.72	1.52—1.96	3—5
25	0.292	0.225—0.414	49	1.95	1.66—2.33	4—6
30	0.312	0.237—0.454	69	2.23	1.81—2.77	7—12
35	0.334	0.250—0.503	88	2.51	1.95—3.29	5—9
40	0.360	0.265—0.563	113	2.82	2.08—3.89	6—17
45	0.391	0.282—0.640	144	3.14	2.20—4.59	7—20
50	0.427	0.300—0.742	174	3.48	2.29—5.45	7—25
55	0.471	0.321—0.883	210	3.81	2.36—6.44	8—41
60	0.525	0.345—1.092	248	4.13	2.39—7.61	8—50
65	0.592	0.373—1.462	290	4.45	2.37—9.00	8—55
70	0.684	0.406—2.155	330	4.71	2.30—10.67	7—58
75	0.802	0.444—4.115	372	4.95	2.20—12.84	6—70
80	0.972	0.491—45.454	412	5.45	2.05—15.66	5—86
85	1.233	0.548— $\infty$	454	5.31	1.00—19.91	7—168
90	1.686	0.620— $\infty$	492	5.47	1.00—27.56	4.00—49.42
95	2.667	0.713— $\infty$	554	5.83	1.00—238.77	7.66
99	4.975	0.811— $\infty$	758	99—23639	9.00—238.77	

сложных биологических явлений, следует помнить, что, пользуясь им, мы всегда в той или иной степени схематизируем изучаемые процессы; при этом могут выпадать из поля зрения некоторые детали, учет наличия и влияния которых может быть существен для дальнейшего развития знаний об изучаемом предмете.

В частности, при анализе распределений следует иметь в виду, что факторы, приводящие к тому или иному их типу, могут быть очень разнородными, и потому создание эмпирической статистической модели лишь один из первых шагов в математическом моделировании процессов, регулирующих численность видов. Информацию, полученную при таком первичном моделировании, следует использовать для дальнейшего, уже более глубокого биологического исследования характера этих факторов в каждом частном случае. Новые результаты позволяют создавать математические модели уже на более высоком уровне.

Например, из теории негативного биномиального распределения известно, что изменчивость вероятности осуществления тех или иных событий в рассматриваемой совокупности (в наших случаях в популяции хозяина при заражении ее паразитами) может зависеть от самых разных причин. При построении дедуктивных моделей можно исходить из взаимно исключающих гипотез и тем не менее в итоге прийти к негативному биномиальному распределению. В нашем случае с заражением серебряного карася плероцеркоидами ремненца можно, например, исходить из гипотезы, что вероятность поедания рыбами раков с процеркоидами остается из года в год постоянной, а варьирует вероятность выживания паразитов у рыб разного возраста. Или, наоборот, принять, что вероятность заражения изменчива, а выживаемость остается постоянной. В обоих случаях результирующим будет негативное биномиальное распределение. Пролить свет на истинный характер процессов могут только соответствующие биологические наблюдения и опыты.

Чрезвычайно важное значение, как это неоднократно было показано нами на конкретных примерах, имеет качество исходного материала, подвергаемого статистическому анализу. Чтобы быть уверенными в достоверности получаемых выводов, нужно прежде всего обеспечить всестороннюю объективность методов получения выборок для анализа. Все это элементарные истины, которые тем не менее часто забываются. В целом же, образно говоря, применяя математические методы, мы как бы манипулируем с явлениями, рассматривая их в перевернутый бинокль, и следует время от времени и достаточно часто пользоваться им в нормальном положении, чтобы не потерять руководящую нить исследования.

## ЛИТЕРАТУРА

- Беклемишев В. Н. 1954. Популяционная биология как одна из теоретических основ борьбы с комарами. Научн. сессия АМН СССР совм. с МЗ УзССР, 20—25 сентября, Тез. докл., Ташкент.
- Беклемишев В. Н. 1970. Биоценологические основы сравнительной паразитологии. Изд. «Наука», М. : 7—488.
- Бреев К. А. 1967а. Новые данные о миграции личинок I стадии *Hypoderma bovis* De Geer в организме хозяина. Паразитол. сб. ЗИН АН СССР, 23 : 191—221.
- (Бреев К. А.) Вгеуев К. А. 1967б. On the variability of developmental periods of 1-st instar larvae of *Hypoderma bovis* De Geer and the causes of their migration in host organism. Wiadomosci Parazytologiczne, 13 (4—5) : 579—584.
- Бреев К. А. 1968а. О распределении личинок подкожных оводов в стадах крупного рогатого скота. I. Негативное биномиальное распределение как модель распределения личинок оводов. Паразитология, 2 (4) : 322—333.
- Бреев К. А. 1968б. О распределении личинок подкожных оводов в стадах крупного рогатого скота. II. Экспонента  $k$  негативного биномиального распределения как мера дисперсии заражения животных оводами. Паразитология, 2 (5) : 381—394.
- Бреев К. А. 1971. О некоторых общих принципах построения методов борьбы с паразитическими членистоногими (на примере подкожных оводов). Энтомол. обозр., 50 (2) : 249—266.
- Быховская-Павловская И. Е. 1967. Развитие экологической паразитологии в СССР. Паразитология, 1 (5) : 391—403.
- Догель В. А. 1941. Курс общей паразитологии. Л. : 3—288.
- Догель В. А. 1962. Общая паразитология. Л. : 3—464.
- Дубинина М. Н. 1966. Ремненцы (*Cestoda: Ligulidae*) фауны СССР. Изд. «Наука», Л. : 3—259.
- Лакин Г. Ф. 1968. Биометрия. Изд. «Высшая школа», М. : 3—281.
- Павловский Е. Н. 1934. Организм как среда обитания. Природа, 1 : 80—91.
- Павловский Е. Н. 1937. Учение о биоценозах в приложении к некоторым паразитологическим проблемам. Изв. АН СССР, 4 : 1385—1422.
- Павловский Е. Н. 1945. Экологическое направление в паразитологии. Журн. общ. биол., 6 (2) : 65—92.
- Павловский Е. Н. 1948. Биоценология и паразитология. Зоол. журн., 27 (2) : 97—112.
- Соснина Е. Ф. 1969. Мышевидные грызуны как прокормители иксодовых клещей в лесном поясе горного Крыма. Паразитология, 3 (6) : 505—509.
- Урабах В. Ю. 1964. Биометрические методы. Изд. «Наука», М. : 3—409.

- Anscombe F. J. 1950. Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, 37 (3—4) : 358—382.  
 Bliss Ch. I. 1958. The analysis of insect counts as negative binomial distributions. Proceed. Tenth Intern. Congress Entomol., 2 : 1015—1032.  
 Bliss Ch. I. a. Fisher R. A. 1953. Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial. *Biometrics*, 9 (2) : 176—200.  
 Bliss Ch. I. a. Owen A. R. G. 1958. Negative binomial distribution with a common  $k$ . *Biometrika*, 45 (1—2) : 37—58.  
 Clark L. R., Geier P. W., Hughes R. D., Morris R. F. 1967. The ecology of insect populations in theory and practice. Methuen a. Co. Ltd., London : 3—224.  
 (Macfadyen A. 1963) Макфедден Э. 1965. Экология животных. Цели и методы. Изд. «Мир», М. : 3—357.  
 Pearson K. 1900. On the criterion that a given system of deviation from the probable, in the case of a correlated system of variables, is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag.*, 5th series, 1 : 157.  
 (Snedecor G. W. 1957). Снедекор Д. У. 1961. Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. Сельхозиздат, М. : 3—487.  
 (Watt K. E. F. 1968) Уатт К. 1971. Экология и управление природными ресурсами. Количественный подход. Изд. «Мир», М. : 5—463.  
 Williamson E. a. Bretherton M. H. 1963. Tables of the negative binomial probability distribution. Wiley a. Sons, London : 1—275.

Числи n	%	тр.
074	231	3
408	383	
665	524	5
878	556	
064	781	
	899	
	011	
	119	
	222	
	322	
418	418	8
511	511	
601	601	
689	689	
775	775	28
	858	
	939	
	018	
	096	
	172	
246	246	
319	319	
391	391	
461	461	
530	530	
	598	
	665	
	731	
	795	
	859	
922	922	
984	984	
045	045	
105	105	
165	165	
	224	
	282	
	339	
	396	
	452	
507	507	
562	562	
616	616	
670	670	
723	723	53
	69	

При

К. А

## СОДЕРЖАНИЕ

A п  
B 1  
B 1  
B 1  
C 1  
(M  
P e

(S n  
(W  
W i

	Стр.
От редакции . . . . .	3
1. Введение . . . . .	5
2. Случайная величина, ее математическое ожидание, дисперсия и распределение. Дискретные распределения — биномиальное и Пуассона—Борткевича . . . . .	8
3. Негативное биномиальное распределение и методы его применения	28
4. Что может дать анализ распределений численности паразитов в популяциях их хозяев . . . . .	53
Литература . . . . .	69
Приложение (вклейка между стр. 70—71)	