$$= \sum_{\omega^{(1)} \in A^{(1)},\, \omega^{(2)} \in A^{(2)}} \mathbf{P}^{(1)}(\omega^{(1)}) \mathbf{P}^{(2)}(\omega^{(2)}) =$$

$$= \sum_{\omega^{(1)} \in A^{(1)}} \mathbf{P}^{(1)}(\omega^{(1)}) \cdot \sum_{\omega^{(2)} \in A^{(2)}} \mathbf{P}^{(2)}(\omega^{(2)}) =$$

$$= \mathbf{P}^{(1)}(A^{(1)}) \mathbf{P}^{(2)}(A^{(2)}) = \mathbf{P}(\tilde{A}^{(1)}) \mathbf{P}(\tilde{A}^{(2)}).$$

Совокупность этих соображений показывает, что в прямом произведении вероятностных пространств пары независимых событий возникают совершенно естественно. Для получения n независимых в совокупности событий естественно взять прямое произведение n вероятностных пространств. Выкладки, основанные на разложении суммы произведений вероятностей в произведение сумм, при этом не изменяются.

Поскольку содержательные законы теории вероятностей относятся к большому числу событий, а наиболее простой способ комбинации событий — предположение их независимости, то простейшая часть этой науки развертывается в прямом произведении пространств, когда число сомножителей стремится к бесконечности. Следует хорошо понять эту ситуацию.

4.2. Испытания Бернулли. Приведем важнейший (по широте употребления) частный пример — испытания Бернулли.

Одно испытание Бернулли — это опыт с двумя исходами, допустим 0 и 1, причем $\mathbf{P}(1)=p$, $\mathbf{P}(0)=q$, $p+q=1$. Единица (допустим) называется успехом, нуль — неудачей. Примером является бросание правильной (тогда $p=q={}^1/_2$) или искривленной монеты (тогда $p\neq q$). Интерес представляют не одно, а n независимых испытаний Бернулли. Итак, «испытания Бернулли — это независимые испытания с двумя исходами и с вероятностью успеха, не меняющейся от испытания к испытанию». Это — определение на уровне здравого смысла, а на математическом уровне определение выглядит так:

$$\Omega = \underbrace{(0,\ 1) \times (0,\ 1) \times \ldots \times (0,\ 1)}_{n \text{ раз}},$$

ω — последовательность нулей и единиц длины n,

$$\mathbf{P}(\omega) = p^{\mu(\omega)} q^{n-\mu(\omega)},$$

где $\mu(\omega)$ — число единиц в последовательности ω, которое можно назвать *числом успехов* в n испытаниях.

Функция от элементарного события в теории вероятностей называется *случайной величиной:* в данном случае $\mu(\omega)$ зависит от случайного исхода n испытаний Бернулли. Возможными значениями величины $\mu(\omega)$ являются числа 0, 1, 2, ..., n. Подсчитаем вероятность события $\{\mu=m\}=\{\omega:\mu(\omega)=m\}$. По определению вероятности события (определение 2 § 1) имеем

$$\mathbf{P}\{\mu = m\} = \sum_{\omega:\mu(\omega)=m} \mathbf{P}(\omega) = C_n^m p^m q^{n-m} \tag{1}$$

(так как число элементов ω, таких, что $\mu(\omega)=m$, равно C_n^m). Набор вероятностей различных значений случайной величины называется *распределением вероятностей* (или просто *распределением*). В частности, формула (1) задает так называемое *биномиальное распределение* (название связано с тем, что правая часть (1) есть член разложения бинома $(p+q)^n$).

Простая формула (1) является удивительно содержательной. Чисто математическое исследование поведения правой части (1) при больших n и различных m и p (возможно, зависящих от n) нетривиально настолько, что в данный момент мы им заниматься не будем, заметив лишь, что оно приводит к весьма интересным результатам. С другой стороны, модель испытаний Бернулли часто с той или иной точностью сопоставляется с реальными явлениями.

Пусть, например, завод собирается выпустить n изделий, причем вероятность выпустить бракованное изделие равна p (известна по прошлым данным). Каково распределение вероятностей для числа μ ожидаемых рекламаций? Оценка этого распределения с помощью формулы (1) будет более или менее грубой (в смысле соответствия фактическим данным) в зависимости от того, насколько точно выполняются предположения модели, т. е. независимость брака для разных изделий и постоянство доли брака p. В связи с возможными колебаниями уровня технологии и качества сырья могут возникнуть отклонения от модели испытаний Бернулли, которые можно интерпретировать как нарушение независимости брака для разных изделий (при сохранении доли брака p), либо как колебания вероятности p (при сохранении независимости), либо как оба явления вместе. Но на практике обычно нет сведений, количественно оценивающих отклонения от модели испытаний Бернулли, известна (да и то по прошлым данным) лишь доля брака p в готовой продукции, и ничего, кроме формулы (1), для практического применения не остается.

Иногда модель испытаний Бернулли является весьма точной (бросания монеты и другие более сложные азартные игры).

Но значительное число применений модели испытаний Бернулли связано не столько с тем, что она хорошо или приемлемо описывает некие явления, сколько с тем, что бывает важно выяснить, что эта модель изучаемого явления не описывает. Испытания Бернулли — это простейшая модель полной случайности. Если полной случайности нет, скажем, число бракованных изделий колеблется не так, как полагалось бы по биномиальному закону, то это может быть важной информацией для совершенствования производства.

Словом, на примере модели испытаний Бернулли целесообразно познакомиться с проблемой проверки статистических гипотез. Статистическая гипотеза — это утверждение о вероятностях тех или иных событий; проверкой гипотезы называется то или иное сопоставление ее с рядом экспериментов.

Пусть, например, двое играют в орлянку, причем монету бросает все время первый игрок (если монета выпадает гербом, то первый игрок получает с второго копейку, в противном случае отдает второму копейку). Второму игроку интересно знать, честно ли бросает монету первый игрок, потому что, возможно, бывают такие специалисты, которые по своему желанию могут выбросить монету любой наперед заданной стороной, хотя на вид монета бросается честно. Пусть для начала монета бросается $n=100$ раз, после чего второй игрок имеет возможность решить, продолжать дальше игру или нет, на основании изучения результатов 100 бросаний. Как ему поступить?

Для понимания реальных возможностей вероятностных методов нужно прежде всего понять, что так поставленный вопрос ответа не имеет. Дело в том, что любой результат 100 бросаний монеты, например последовательность из 100 гербов либо последовательность из 100 цифр, либо, наконец, последовательность ГЦГЦ..., в которой герб и цифра строго чередуются, имеет одну и ту же (при $p=1/2$) вероятность 2^{-100}. Ни одна из последовательностей, которые могут получиться в опыте, не лучше и не хуже, чем другая. Ни одна научная гипотеза (в данном случае — гипотеза испытаний Бернулли) не может быть проверена, если не указать альтернативных гипотез, т. е. каких-то сведений о том, как именно проверяемая гипотеза может нарушаться: гипотеза всегда проверяется против каких-то альтернатив.

В нашем примере простейшая альтернативная гипотеза может состоять в том, что первый игрок, стремясь выиграть побольше, будет выбрасывать герб не с вероятностью $1/2$, а чаше (возможно, не случайно). Тогда число выпадений герба μ будет принимать в каком-то смысле бо́льшие значения, чем было бы при вероятности $p=1/2$. Пусть, скажем, в 100 бросаниях оказалось $\mu=60$ выпадений герба. Мы можем вычислить с помощью таблиц $\mathbf{P}\{\mu\geqslant 60\}$ при гипотезе $p=1/2$ (это будет число порядка 2%). Речь должна идти именно о вероятности $\mathbf{P}\{\mu\geqslant 60\}$, но не о вероятности $\mathbf{P}\{\mu=60\}$: кажется подозрительным не то, что число успехов именно 60, а то, что их значительно больше 50. Дальше рассуждаем так: возможно, что первый игрок бросал монету честно, но произошло выгодное для него случайное событие (позволившее ему выиграть 20 копеек), вероятность которого около 2%; возможно, что первый игрок бросал монету нечестно, следует прекратить игру.

Второй игрок должен принять какое-то решение.

Рассмотрим последствия различных решений второго игрока. В азартные игры играют не ради наживы (теоретические соображения, в том числе довольно сложные, в которые мы сейчас не будем вдаваться, показывают, что в азартных играх с равноправными игроками типа орлянки наживы быть не может), а ради получаемого от игры удовольствия. Если же партнер оказался шулером, то об удовольствии речи быть не может. Таким образом, если второй игрок будет прекращать игру в том случае, если в первых 100 бросаниях его партнер выбросит герб 60 раз или более, то лишь примерно в 2% встреч с честными игроками он напрасно бросит игру (т. е., так сказать, лишится 2% удовольствия).

Иначе можно сказать, что вероятность напрасно обидеть честного партнера составляет около 2%. Различные логические видоизменения этого утверждения типа: «если $\mu<60$, то с вероятностью 0,98 партнер — не шулер» либо «если $\mu\geqslant 60$, то с вероятностью 0,98 партнер — шулер» — будут неверными; несостоятельными в данной постановке задачи будут и вероятностные оценки надежности избавления от шулеров, которая достигается ценой потери 2% удовольствия. Дело в том, что шулер никак не описан: в сущности, о нем сказано лишь то, что он обладает непостижимыми для обычного человека (второго игрока) способностями. Не сказано и какова доля шулеров среди возможных партнеров: например, если их нет вовсе, то партнер — не шулер с вероятностью 1, а не 0,98 или 0,02. Если же шулера имеются и используют свои сверхчеловеческие способности для того, чтобы не попадать под действие критерия $\{\mu\geqslant 60\}$, имитируя при первых 100 бросаниях полную случайность, то применение этого критерия абсолютно ничего не дает в смысле избавления от шулеров.

Мы встретимся в данной книге с ситуацией, когда альтернативная гипотеза описывается не в крайне неопределенной форме «партнер—шулер», а также в виде вероятностной гипотезы, которая приписывает наблюдаемым событиям вероятности, не совпадающие с теми, которые приписывает им проверяемая гипотеза. В этом случае рассмотрение проверки гипотез можно углубить, дав, в частности, вероятностную оценку гарантий, связанных с применением того или иного статистического критерия.

4.3. Распределение Пуассона. Мы видели, что, в частности, задача проверки гипотезы полной случайности потребовала вычисления вероятностей вида $\mathbf{P}\{\mu\geqslant a\}$. Конечно,

$$\mathbf{P}\{\mu\geqslant a\}=\sum_{m\geqslant a}\mathbf{P}\{\mu=m\}, \qquad (2)$$

где вероятности $\mathbf{P}\{\mu=m\}$ даются формулой (1). Громоздкость вычислений при больших значениях m и n побудила ряд математиков искать для вероятностей (1) и (2) какие-то простые приближенные выражения. Оказывается, что асимптотический анализ этих вероятностей приводит к двум наиболее универсальным законам распределения, значение которых выходит далеко за рамки сравнительно убогой модели испытаний Бернулли. Это нормальное распределение (или распределение Гаусса—Лапласа) и распределение Пуассона. Нормальным распределением займемся позже, а сейчас рассмотрим распределение Пуассона.

Это распределение возникает из формулы (1) при предельном переходе, когда $n \to \infty$, $p \to 0$, но так, чтобы $np \to \lambda$, где $\lambda > 0$ — фиксированное число. Так как в пределах одной последовательности испытаний Бернулли p постоянно (и следовательно, не может быть так, чтобы $p \to 0$), нужно рассмотреть несколько более сложную схему: последовательность серий испытаний Бернулли.

В первой серии пусть будет $n=1$ испытание с вероятностью успеха p_1; во второй серии $n=2$ испытания с вероятностью успеха p_2 и т. д., наконец, в n-й серии рассмотрим n испытаний с вероятностью успеха p_n каждое, причем $np_n \to \lambda$. Пусть $\mu(n)$ — число успехов в n-й серии; m — фиксированное целое неотрицательное число.

Теорема Пуассона. *При $n \to \infty$ выполняется предельное соотношение*

$$\mathrm{P}\{\mu(n) = m\} \to \frac{\lambda^m}{m!} e^{-\lambda}.$$

Доказательство. В точном выражении

$$\mathrm{P}\{\mu(n) = m\} = C_n^m p_n^m (1 - p_n)^{n-m}$$

перейдем к пределу при $n \to \infty$, $p_n = \lambda/n + o(1/n)$, m фиксированном. Поскольку

$$\mathrm{P}\{\mu(n) = m\} = \frac{n(n-1)\ldots(n-m+1)}{m!} \times$$

$$\times \left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^m \left(1 - \frac{\lambda}{n} - o\left(\frac{1}{n}\right)\right)^{n-m},$$

то при этом предельном переходе главный член скобки $(\lambda/n + o(1/n))^m$ будет $(\lambda/n)^m$ (так как другие члены разложения этого бинома содержат $o(1/n)$ в некоторой степени и ограниченные — зависящие лишь от m — коэффициенты). Скобка $(1 - \lambda/n - o(1/n))^{n-m}$ имеет пределом $e^{-\lambda}$. Учитывая, что $n(n-1)\ldots(n-m+1)(\lambda/n)^m \to \lambda^m$, получаем утверждение теоремы.

В этой теореме числа $\tilde{p}_m = \lambda^m e^{-\lambda}/m!$ выступают как предельные значения биномиальных вероятностей. Можно ввести случайную величину, для которой вероятности $\tilde{p}_m$ будут точными вероятностями различных значений. Действительно, положим $\Omega = \{0, 1, \ldots, m, \ldots\}$, учитывая, что $\sum\limits_{m=0}^{\infty} \tilde{p}_m = 1$, введем вероятности элементарных событий формулой $\mathbf{P}\{m\} = p_m$ и введем, наконец, случайную величину ξ как функцию на множестве Ω, заданную формулой $\xi(m) = m$. Тогда, очевидно,

$$\mathbf{P}\{\xi = m\} = \frac{\lambda^m}{m!} e^{-\lambda}. \tag{3}$$

Такое распределение вероятностей называется распределением Пуассона.

Распределение Пуассона очень часто находится в разумном согласии с экспериментом — от числа частиц, зарегистрированных счетчиком радиоактивного излучения за какой-то промежуток времени, до числа вызовов, поступивших на телефонную станцию, либо числа отказов какого-то оборудования. Оно необычайно удобно для иллюстрации основных вероятностно-статистических понятий, так как формула (3), зависящая от двух параметров m и λ, несравненно легче поддается табулированию, чем формула (1), зависящая от трех параметров m, n, p. Но для наибольшего удобства обращения с распределением Пуассона нам необходимо понять вероятностный смысл параметра λ, что лучше сделать в рамках общих понятий, связанных со случайными величинами.

§ 5. Случайные величины

Случайной величиной (как уже говорилось в предыдущем параграфе) называется функция $\xi = \xi(\omega)$, определенная на множестве элементарных событий $\Omega = \{\omega\}$. Пока множество Ω не более чем счетно, функция $\xi(\omega)$ совершенно произвольна. Значениями случайной величины могут быть вещественные или комплексные числа, а также кватернионы, матрицы, операторы и т. д. Но для определенности будем говорить о случайных величинах с вещественными значениями.

Точка зрения, с которой функции $\xi = \xi(\omega)$ рассматриваются в теории вероятностей, не совсем похожа на точку зрения математического анализа. На первый план выступают множества уровня функции $\xi(\omega)$: пусть $a_1, a_2, \ldots, a_n, \ldots$ — *различные* значения случайной величины ξ; рассмотрим множества вида $\{\omega : \xi(\omega) = a_i\}$, сокращенно обозначаемые $\{\xi = a_i\}$, и их вероятности вается, а иногда и отвергается, то вторая (математическая) часть есть твердо установленный факт. Первоначально (Лаплас, Пуассон) закон больших чисел выводился из более сложной центральной предельной теоремы теории вероятностей. Впоследствии П. Л. Чебышев нашел элементарное и более общее доказательство, с которым мы и познакомимся. (В оригинальном изложении Чебышева также присутствуют длинные вычисления, с нашей точки зрения совершенно ненужные. Следует иметь в виду, что когда в учебнике сказано, что излагаются результаты того или иного ученого, на самом деле речь всегда идет о некоторой коллективной переработке оригинального изложения, делающей его несравненно более удобопонятным. Можно сравнить, например, изложение электродинамики в трактате Максвелла или лекциях Больцмана с современными нам учебниками.)

Неравенство Чебышева. Мы говорили, что дисперсия случайной величины ξ измеряет в каком-то смысле возможные отклонения ξ от $\mathbf{M}\xi$. Этот смысл уточняется неравенством Чебышева.

Лемма 6. *Пусть существует* $\mathbf{D}\xi$ *и дано число* $\varepsilon>0$. *Тогда*

$$\mathbf{P}\{|\xi-\mathbf{M}\xi|\geqslant\varepsilon\}\leqslant\mathbf{D}\xi/\varepsilon^2.$$

Доказательство. Запишем цепочку равенств и неравенств:

$$\mathbf{D}\xi = M(\xi - M\xi)^2 = \sum_{a_i}(a_i - M\xi)^2\,\mathbf{P}\{\xi = a_i\} \geqslant$$

$$\geqslant \sum_{a_i:|a_i-\mathbf{M}\xi|\geqslant\varepsilon}(a_i - M\xi)^2\,\mathbf{P}\{\xi = a_i\} \geqslant$$

$$\geqslant \varepsilon^2 \sum_{a_i:|a_i-M\xi|\geqslant\varepsilon}\mathbf{P}\{\xi = a_i\} = \varepsilon^2\,\mathbf{P}\{|\xi - M\xi| \geqslant\varepsilon\}.$$

Отсюда вытекает утверждение леммы.

Определение. Говорят, что последовательность случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ сходится к нулю по вероятности, если для любого $\varepsilon>0$

$$\mathbf{P}\{|\xi_n|>\varepsilon\}\to 0 \quad (n\to\infty).$$

Из неравенства Чебышева (лемма 6) вытекает, что для сходимости к нулю по вероятности последовательности $\{\xi_n-\mathbf{M}\xi_n\}$ достаточно, чтобы $\mathbf{D}\xi_n\to 0$.

Теорема (закон больших чисел в форме Чебышева). *Пусть случайные величины* $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ *попарно независимы, причем* $\mathbf{D}\xi_i\leqslant C<\infty$. *Тогда*

$$\mathbf{P}\left\{\left|\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}-\frac{\mathbf{M}\xi_1+\mathbf{M}\xi_2+\ldots+\mathbf{M}\xi_n}{n}\right|>\varepsilon\right\}\to 0$$

при $n\to\infty$ для любого $\varepsilon>0$.

Доказательство. Достаточно установить, что $\mathbf{D}\{(\xi_1+\xi_2+\ldots+\xi_n)/n\}\to 0$. Имеем (в силу леммы 5)

$$\mathbf{D}\left(\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}\right)=\frac{1}{n^2}\mathbf{D}(\xi_1+\xi_2+\ldots+\xi_n)=$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}\xi_i\leqslant\frac{Cn}{n^2}=\frac{C}{n}\to 0\ (n\to\infty).$$

Замечание. Для доказательства закона больших чисел мы применили несложный (но и не тривиальный, поскольку Лаплас и Пуассон его не видели) аппарат, состоящий из неравенства Чебышева и способа вычислять дисперсию суммы $\xi_1+\xi_2+\ldots+\xi_n$. Этот аппарат пригоден и для зависимых случайных величин, если, например, предположить, что $\mathrm{cov}(\xi_i,\xi_j)\to 0$ при $|i-j|\to\infty$ каким-нибудь таким образом, что $\mathbf{D}(\xi_1+\xi_2+\ldots+\xi_n)/n^2\to 0$.

Комментарий. Таким образом, среднее арифметическое $(\xi_1+\xi_2+\ldots+\xi_n)/n$ из большого числа n случайных величин с вероятностью, сколь угодно близкой к 1, не отличается более, чем на ε от неслучайной величины $(\mathbf{M}\xi_1+\mathbf{M}\xi_2+\ldots+\mathbf{M}\xi_n)/n$. Допустим, что $\xi_1, \xi_2, \ldots, \xi_n$ — результаты измерений некоторой физической величины в ансамбле однотипных опытов. Тогда естественно предположить, что все величины $\xi_1, \xi_2, \ldots, \xi_n$ имеют одно и то же распределение вероятностей, в частности,

$$\mathbf{M}\xi_1=\mathbf{M}\xi_2=\ldots=\mathbf{M}\xi_n=a, \qquad (6)$$

и получаем, что $\mathbf{P}\{|(\xi_1+\xi_2+\ldots+\xi_n)/n-a|<\varepsilon\}\to 1$ при $n\to\infty$, т. е. постоянная величина a может быть найдена как среднее арифметическое из результатов наблюдений как угодно точно, лишь бы n было велико. Классики аргументировали, что a и есть истинное значение измеряемой величины. (Их аргументация сводится к тому, что одинаково вероятно наблюдению ξ_i отклониться от истины на данную величину в положительную и отрицательную сторону, а тогда $\mathbf{M}\xi_i=a$ и есть истина). Получается парадоксальный вывод, что можно узнать некую длину с точностью до микрона, пользуясь в качестве измерительного прибора масштабной линейкой. Элементарная ошибка здесь очевидна: измеряя длину 50,1-мм масштабной линейкой, всегда будем получать 50 мм; возникает систематическая ошибка. Более подробное обсуж-