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 $\sum_{m=0}^{\infty} \tilde{p}_m = 1$, 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

$$p_i = \mathbf{P}\{\xi = a_i\} = \sum_{\omega:\xi(\omega)=a_i} \mathbf{P}(\omega).$$

Табличка вида

$$\begin{pmatrix} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix} \qquad (1)$$

называется *распределением вероятностей* (или просто *распределением*) случайной величины ξ.

Понятно, что $p_i \geqslant 0$ и $\Sigma p_i = 1$. С другой стороны, для любого распределения вида (1), такого, что $p_i \geqslant 0$ и $\Sigma p_i = 1$, найдется случайная величина с этим распределением. Действительно, достаточно положить $\Omega = (a_1, a_2, \dots, a_n, \dots)$, $\mathbf{P}(a_i) = p_i$ и $\xi(a_i) = a_i$. (Это тривиальное замечание впоследствии разовьется в так называемую «теорему Колмогорова», определяющую вероятностную меру в функциональном пространстве).

Почему в теории вероятностей центральную роль играют множества уровня случайных величин и распределения вероятностей? Дело в том, что даже счетное Ω может быть весьма сложным (прямое произведение большого числа пространств). Задание вероятностной меры в таком Ω, особенно если не предполагается независимости, может быть затруднительным. Но случайная величина может принимать одно и то же значение a_i на многих ω, и тогда определение вероятностей $p_i = \mathbf{P}(\xi = a_i)$ — из опыта на основании частоты наступления события $\{\xi = a_i\}$ или как-нибудь иначе — может быть делом более простым, чем определение вероятностей $\mathbf{P}(\omega)$. Однако при большом числе различных значений a_i даже определение вероятностей p_i может быть сложным. Поэтому распределение вероятностей на практике пытаются характеризовать каким-то небольшим набором чисел, которые называются *параметрами*. Важнейшим таким параметром является *математическое ожидание*.

Определение. Математическим ожиданием $\mathbf{M}\xi$ случайной величины $\xi = \xi(\omega)$ называется сумма

$$\mathbf{M}\xi = \sum_{\omega \in \Omega} \xi(\omega)\mathbf{P}(\omega) \qquad (2)$$

$\left(\text{в предположении, что } \sum_{\omega \in \Omega} |\xi(\omega)|\mathbf{P}(\omega) < \infty\right)$.

Если предположение абсолютной сходимости не выполнено, то говорят, что случайная величина ξ не имеет математического ожидания.

Из формулы (2) очевидным образом вытекают следующие правила исчисления:

$\mathbf{M}(c\xi) = c\mathbf{M}\xi$, если c — константа;
$\mathbf{M}(\xi + \eta) = \mathbf{M}\xi + \mathbf{M}\eta$, если $\mathbf{M}\xi$ и $\mathbf{M}\eta$ существуют

(второе из этих правил вытекает из того, что абсолютно сходящиеся ряды можно складывать почленно). В этих правилах подразумевается, конечно, что $c\xi$ — это случайная величина, определяемая соотношением $(c\xi)(\omega)=c\xi(\omega)$, а сумма $\xi+\eta$ — это случайная величина, определяемая соотношением $(\xi+\omega)(\omega)=\xi(\omega)+\eta(\omega)$. При этом случайные величины ξ и η должны быть определены на одном и том же пространстве элементарных событий (это всегда подразумевается).

Лемма. *Для любой функции $f(x)$ вещественного переменного*

$$Mf(\xi)=\sum_{a_i} f(a_i)p_i, \tag{3}$$

где $p_i=\mathbf{P}\{\xi=a_i\}$, а $f(\xi)(\omega)=f[\xi(\omega)]$.

Доказательство. Лемма, конечно, верна только в том случае, если ряд в правой части (3) сходится абсолютно. Рассмотрим следующие преобразования:

$$Mf(\xi)=\sum_{\omega\in\Omega} f(\xi(\omega))\mathbf{P}(\omega)=$$

$$=\sum_{a_i}\Bigg[\sum_{\omega:\xi(\omega)=a_i} f(\xi(\omega))\mathbf{P}(\omega)\Bigg]=\sum_{a_i}\Bigg[\sum_{\omega:\xi(\omega)=a_i} f(a_i)\mathbf{P}(\omega)\Bigg]=$$

$$=\sum_{a_i}\Bigg[f(a_i)\sum_{\omega:\xi(\omega)=a_i}\mathbf{P}(\omega)\Bigg]=\sum_{a_i} f(a_i)\mathbf{P}\{\xi=a_i\}=$$

$$=\sum_{a_i} f(a_i)p_i.$$

Очевидно, что все ряды, участвующие в этих преобразованиях, одновременно сходятся или не сходятся абсолютно. В предположении же абсолютной сходимости существует некоторое конечное число членов ряда, сумма которых не более чем на заданное $\varepsilon>0$ отличается от суммы ряда, в то время как сумма модулей остальных членов не более ε. Поэтому всевозможные группировки членов ряда сводятся к группировкам слагаемых в конечной сумме и тем самым законны. Следовательно, произведенные преобразования верны, что и доказывает лемму.

В частности, полагая $f(x)=x$, получаем

$$Mf(\xi)=M\xi=\sum_{a_i} a_ip_i, \tag{4}$$

что в виде словесного «заклинания» выглядит так: ***математическое ожидание случайной величины равно сумме ее значений, умноженных на вероятности этих значений.*** Таким образом, математическое ожидание однозначно определяется распределением случайной величины.

Замечание. Но если (4) принять за определение математического ожидания, то будет неудобно доказывать, что $\mathbf{M}(\xi+\eta)=\mathbf{M}\xi+\mathbf{M}\eta$.

Математическое ожидание имеет точную и прозрачную механическую аналогию. Если распределение случайной величины (1) изобразить в виде одномерной механической системы, поместив в точку с абсциссой a_i массу p_i, то, в силу условия $\Sigma p_i=1$, получим, что $\mathbf{M}\xi$ есть абсцисса центра масс такой системы. Механикам, правда, обычно не приходит в голову особо оговаривать, что $\Sigma|a_i|p_i<\infty$, ибо размеры механических систем конечны. Но в теории вероятностей вместе со случайной величиной ξ часто приходится рассматривать, например, $1/\xi$, $\ln\xi$ или $\operatorname{tg}\xi$, так что условие абсолютной сходимости существенно, ибо без него правила исчисления (скажем, $\mathbf{M}(\xi+\eta)=\mathbf{M}\xi+\mathbf{M}\eta$) просто неверны.

Если математическое ожидание есть некий центр, вокруг которого группируются возможные значения случайной величины, то, наверное, должен существовать и параметр распределения, характеризующий величину возможного разброса значений случайной величины вокруг их центра. Мы его немедленно получаем из механической аналогии: в механике это — момент инерции, а в теории вероятностей — *дисперсия.*

Определение. Дисперсией $\mathbf{D}\xi$ случайной величины ξ называется число, определяемое формулой

$$\mathbf{D}\xi = \mathbf{M}(\xi-\mathbf{M}\xi)^2=\sum_{a_i}(a_i-\mathbf{M}\xi)^2 p_i \tag{5}$$

(предполагается, конечно, что ряд в правой части сходится). Проделаем следующую небольшую выкладку:

$$\begin{aligned}\mathbf{D}\xi&=\mathbf{M}(\xi-\mathbf{M}\xi)^2=\mathbf{M}(\xi^2-2\mathbf{M}\xi\cdot\xi+(\mathbf{M}\xi)^2)=\\&=\mathbf{M}\xi^2-\mathbf{M}(2\mathbf{M}\xi\cdot\xi)+\mathbf{M}(\mathbf{M}\xi)^2=\\&=\mathbf{M}\xi^2-2(\mathbf{M}\xi)^2+(\mathbf{M}\xi)^2=\mathbf{M}\xi^2-(\mathbf{M}\xi)^2.\end{aligned}$$

(Для понимания этой выкладки следует иметь в виду, что в ней $\mathbf{M}\xi$ выступает в одних местах как число, а в других — как случайная величина, принимающая единственное значение: например, $\mathbf{M}(\mathbf{M}\xi)^2=(\mathbf{M}\xi)^2$. Так в математическом анализе в выражении $\sin x+1$ единица выступает то как число, то как функция, тождественно равная 1.)

Таким образом, $\mathbf{D}\xi$ выражается через $\mathbf{M}\xi$ и $\mathbf{M}\xi^2$. Поскольку $\mathbf{D}\xi\geqslant 0$, то получаем, что $\mathbf{M}\xi^2\geqslant(\mathbf{M}\xi)^2$. Заметим, что выражение $\mathbf{M}\xi^k$ называется *моментом k-го порядка* случайной величины ξ. Моменты — это тоже параметры распределения.

После этих общих определений рассмотрим *пример* распределения Пуассона. Пусть случайная величина принимает значения 0, 1, 2, . . . , причем $\mathbf{P}\{\xi=m\}=\dfrac{\lambda^m}{m!}e^{-\lambda}$.