Math-Net.Ru

Общероссийский математический портал

А. Н. Колмогоров, Теория вероятностей, *Теория вероятн. и ее примен.*, 2003, том 48, выпуск 2, 211–248

DOI: <https://doi.org/10.4213/tvp282>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.196.121.252

26 апреля 2023 г., 18:36:49

Из наследия Андрея Николаевича Колмогорова

© 2003 г.

КОЛМОГОРОВ А. Н.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ*

В 1956 г. издательством Академии наук СССР были изданы подготовленные Математическим институтом им. В. А. Стеклова три книги монографии «Математика. Ее содержание, методы и значение». Редакционная коллегия состояла из А. Д. Александрова, А. Н. Колмогорова и М. А. Лаврентьева. На правах рукописи монография была напечатана в 1953 г. тиражом 350 экземпляров с целью ее обсуждения математической общественностью.

Точка зрения А. Н. Колмогорова относительно замысла этого издания состояла в том, что хорошо было бы иметь, на самом деле, две книги: первая, «которая неофициально мыслилась в качестве “Антикуранта”» (см. предисловие к 3-му русскому изданию книги: *Р. Курант, Г. Роббинс. Что такое математика? Элементарный очерк идей и методов. 3-е изд., испр. и доп. М.: МЦНМО, 2001*), т.е. книга для всех желающих познакомиться в живой и доступной форме с элементами высшей математики, проверить уровень своих математических способностей, а в случае молодых читателей подумать о целесообразности выбора математики в качестве своей профессии, и вторая книга, «рассчитанная на более подготовленных читателей, включая и нас самих математиков, часто очень беспомощных в оценке перспектив развития своей науки в целом».

В конечном итоге, вышедшие три тома, содержащие 20 глав, более соответствовали первому из указанных вариантов, что следует и из предисловия, в котором сказано, что «коллектив авторов при составлении этой книги исходил из намерения ознакомить достаточно широкие круги советской интеллигенции с содержанием и методами отдельных математических дисциплин, их материальными основами и путями развития».

Во втором томе этой монографии (глава XI) была опубликована статья А. Н. Колмогорова, которая воспроизводится в настоящем юбилейном выпуске вместе с отзывом А. Я. Хинчина и избранными местами из переписки А. Д. Александрова (как фактически главного редактора издания) и А. Н. Колмогорова, интересной как их взглядами на содержание представленного Колмогоровым варианта его статьи, так и философско-методологическими аспектами теории вероятностей.

А. Н. Ширяев

* Работа была опубликована в сборнике «Математика. Ее содержание, методы и значение», т. 2. М.: Издательство Академии наук СССР, 1956, с. 252–284.

§ 1. Вероятностные закономерности

Простейшие закономерности, устанавливаемые естествознанием, заключаются в указании условий, при которых какое-либо интересующее нас событие заведомо происходит или заведомо не происходит, т.е. эти условия могут быть выражены по одной из следующих двух схем:

1) если осуществляется комплекс (т.е. совокупность) условий S , то с достоверностью происходит событие A ;

2) если осуществляется комплекс условий S , то событие A произойти не может.

В первом случае событие A по отношению к комплексу условий S называется «достоверным» или «необходимым» событием, а во втором — «невозможным» событием. Например, при атмосферном давлении и температуре t , лежащей в пределах $0^\circ < t < 100^\circ$ (комплекс условий S), вода с достоверностью находится в жидком состоянии (достоверное событие A_1), а в газообразном или в твердом состоянии находиться не может (невозможные события A_2 и A_3).

Событие A , которое при осуществлении комплекса условий S иногда происходит, а иногда не происходит, называется по отношению к этому комплексу условий *случайным*. Возникает вопрос: означает ли случайность события A отсутствие всякой закономерной связи между комплексом условий S и событием A ? Пусть, например, установлено, что лампы определенного типа, производимые определенным заводом (условия S), иногда горят более 2000 час. (событие A), а иногда до истечения этого срока перегорают и приходят в негодность. Могут ли тем не менее опыты с проверкой способности ламп гореть 2000 час. служить для характеристики качества продукции завода? Или следует ограничиться указанием того срока (скажем, 500 час.), в течение которого практически все лампы безотказно работают, и того срока, после которого практически все лампы перегорают (скажем, срока в 10 000 час.)? Ясно, что характеристика срока работы ламп только неравенствами $500 \leq T \leq 10\,000$ мало удовлетворит потребителя. Потребитель получит значительно более полную информацию, если ему будет сказано, что приблизительно в 80% случаев лампы служат не менее 2000 час. Еще более полная характеристика качества ламп будет содержаться в указании для любого T процента $\nu(T)$ ламп, которые служат не менее T часов, хотя бы в виде графика, изображенного на рис. 1.

Кривая $\nu(T)$ практически находится при помощи испытаний с пробной партией из достаточно большого числа (100–200) ламп. Естественно, что найденная таким образом кривая имеет действительную ценность лишь в том случае, если она правильно отражает реальную закономерность, действующую не только для данной пробной партии, но и вообще для ламп, производимых при заданном качестве материа-

лов и при установленной на заводе технологии производства, т.е. если испытания, произведенные с другими пробными партиями, состоящими из ламп, изготовленных в тех же общих условиях, приводят к близким результатам (т.е. к кривым $\nu(T)$, мало отличающимся от кривой, полученной в результате испытаний первой пробной партии). Это значит, что *статистическая* закономерность, выражаемая кривыми $\nu(T)$ в пробных партиях, является лишь отражением *вероятностной* закономерности, связывающей срок службы лампы с качеством материалов, из которых она изготовлена, и с технологией ее изготовления.

Рис. 1

Эта вероятностная закономерность задается при помощи функции $P(T)$, где $P(T)$ есть вероятность того, что отдельная лампа (произведенная при заданных условиях) будет гореть не менее T часов.

Утверждение о существовании у события A определенной вероятности

$$P(A/S) = p$$

при условиях S заключается в том, что в различных достаточно длинных сериях испытаний (т.е. осуществлений комплекса условий S) получаемые частоты появления события A

$$\nu_r = \frac{\mu_r}{n_r}$$

(где n_r — число испытаний r -й серии, μ_r — число тех испытаний этой серии, при которых произошло событие A) будут приблизительно одинаковы и близки к p .

Гипотеза о существовании такой константы $p = P(A/S)$ (объективно обусловленной характером связи между комплексом условий S и событием A), к которой частоты ν оказываются, вообще говоря, тем ближе, чем больше число испытаний n , хорошо оправдывается для широкого класса явлений. Такого рода явления естественно называть *вероятностно-случайными*¹⁾.

¹⁾ Иногда их называют *статистическими*.

Рассмотренный выше пример относится к области вероятностных закономерностей массового производства. Реальность такого рода закономерностей не подлежит никакому сомнению. На них основаны весьма важные практические приемы статистического выборочного контроля массовой продукции. Близкой по способу образования вероятностных закономерностей является область вероятностных законов рассеивания снарядов, имеющих основное значение для теории стрельбы. Так как исторически это один из первых примеров реального применения методов теории вероятностей к техническим задачам, то к некоторым простейшим задачам теории стрельбы мы еще вернемся далее.

Сказанное выше о «близости» вероятности p и частоты ν при большом числе испытаний n несколько расплывчато; мы ничего не сказали о том, насколько мала разность $\nu - p$ при том или ином n . Степень близости ν к p получит количественную оценку в § 3. Интересно отметить, что полностью исключить некоторую неопределенность в этом вопросе нельзя. Само утверждение о близости ν и p , как это обнаруживается при уточнении вопроса, имеет лишь вероятностный характер.

§ 2. Аксиомы и основные формулы элементарной теории вероятностей

Поскольку большая роль статистических закономерностей несомненна, возникает вопрос о методах их изучения. Прежде всего возникает мысль о возможности чисто эмпирического, экспериментального их исследования. Так как вероятностная закономерность проявляется в массовых процессах, то представляется естественным, что для ее обнаружения необходимо произвести массовый эксперимент.

Такое представление, однако, истинно лишь отчасти. Установив некоторые вероятностные закономерности экспериментально, можно вывести из них новые вероятностные закономерности логическим или вычислительным путем при помощи некоторых общих допущений. Прежде чем показать, как это делается, мы должны перечислить некоторые основные определения и формулы теории вероятностей.

Из представления о вероятности как о нормальном значении частоты $\nu = m/n$, где $0 \leq m \leq n$, и, следовательно, $0 \leq \nu \leq 1$, вытекает, что вероятность $P(A)$ любого события A естественно считать лежащей между нулем и единицей¹⁾

$$0 \leq P(A) \leq 1. \quad (1)$$

Два события называются *несовместимыми*, если они не могут (при осуществлении комплекса условий S) произойти оба. Например, при

¹⁾ Запись $P(A/S)$ мы для краткости изменим теперь на $P(A)$.

бросании игральной кости выпадение четного числа очков и выпадение тройки являются событиями несовместимыми. Событие A называется соединением событий A_1 и A_2 , если оно состоит в том, что происходит хотя бы одно из событий A_1 или A_2 . Например, при бросании игральной кости событие A , состоящее в выпадении 1, 2 или 3, является соединением событий A_1 и A_2 , где A_1 состоит в выпадении 1 или 2, а A_2 — в выпадении 2 или 3. Легко видеть, что для чисел появлений m_1, m_2 и m двух несовместимых событий A_1 и A_2 и их соединения $A = A_1 \cup A_2$ имеет место равенство $m = m_1 + m_2$, что дает для соответствующих частот $\nu = \nu_1 + \nu_2$.

Это приводит к естественности следующей аксиомы сложения вероятностей:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2), \quad (2)$$

если события A_1 и A_2 несовместимы и $A_1 \cup A_2$ обозначает их соединение.

Далее, для достоверного события U естественно принять

$$P(U) = 1. \quad (3)$$

Вся математическая теория вероятностей строится на простых аксиомах такого типа, как (1), (2) и (3). С точки зрения чистой математики *вероятность* является числовой функцией от «события», обладающей рядом аксиоматически фиксированных свойств. Свойства вероятностей, выражаемые формулами (1), (2) и (3), служат достаточной основой для построения так называемой элементарной теории вероятностей, если не настаивать на том, что в аксиоматизации нуждаются и сами понятия события, соединения событий и определяемого ниже совмещения событий. Начинаящему полезнее держаться наглядного понимания терминов «событие» и «вероятность», но полезно знать, что не поддающийся полной формализации реальный смысл этих понятий не влияет на полную формальную отчетливость аксиоматизированного чисто математического изложения теории вероятностей.

Будем называть соединением событий $A_1, A_2, \dots, A_s$, данных в любом числе, событие A , заключающееся в том, что происходит хотя бы одно из этих событий. Тогда из аксиомы сложения легко получаем для любого числа попарно несовместимых событий $A_1, A_2, \dots, A_s$, и их соединения A

$$P(A) = P(A_1) + P(A_2) + \dots + P(A_s)$$

(так называемая теорема сложения вероятностей).

Если соединение этих событий есть достоверное событие (т.е. при каждом осуществлении комплекса условий S происходит какое-либо из событий A_k), то

$$P(A_1) + P(A_2) + \dots + P(A_s) = 1.$$

В этом случае систему событий $A_1, A_2, \dots, A_s$ называют *полной системой* событий.

Рассмотрим теперь два, вообще говоря, совместимых события A и B . Событие C назовем *совмещением* событий A и B ($C = AB$), если событие C состоит в том, что происходят оба события A и B ¹⁾.

Например, если событие A состоит в том, что число очков, выпадающее при бросании игральной кости, четно, а B — в том, что оно кратно трем, то событие C состоит в том, что число очков равно шести.

Пусть при большом числе n повторных испытаний событие A появилось m раз, а событие B появилось l раз, причем k раз вместе с событием A . Отношение k/m естественно назвать условной частотой события B при условии A . Связывающей частоты k/m , m/n и k/n формуле

$$\frac{k}{m} = \frac{k}{n} : \frac{m}{n}$$

естественно сопоставить следующее определение:

Условной вероятностью $P(B|A)$ события B при условии A называется отношение

$$P(B | A) = \frac{P(AB)}{P(A)}.$$

Здесь предполагается, конечно, что $P(A) \neq 0$.

Если события A и B по существу никак не связаны друг с другом, то естественно предполагать, что событие B не должно появляться при условии наступления события A ни существенно чаще, ни существенно реже, чем при рассмотрении всех вообще испытаний, т.е. что приближенно $k/m \sim l/n$ или

$$\frac{k}{m} = \frac{k}{n} \frac{m}{n} \sim \frac{l}{n} \frac{m}{n}.$$

В последнем приближенном равенстве $m/n = \nu_A$ есть частота события A , а $l/n = \nu_B$ — частота события B , наконец, $k/n = \nu_{AB}$ — частота совмещения событий A и B .

Мы видим, что эти частоты связаны соотношением

$$\nu_{AB} \sim \nu_A \nu_B.$$

Для вероятностей событий A, B и AB естественно поэтому принять соответствующее точное равенство

$$P(AB) = P(A) P(B). \quad (4)$$

Равенство (4) служит определением *независимости* двух событий A и B .

¹⁾ Аналогично совмещение C любого числа событий $A_1, A_2, \dots, A_s$ состоит в наступлении всех указанных событий.

Аналогично можно определить независимость любого числа событий. Кроме того, можно дать определение независимости любого числа испытаний (последнее, грубо говоря, сводится к тому, что тот или иной исход части этих испытаний никак не влияет на исход остальных¹⁾).

Вычислим теперь вероятность P_k ровно k появлений некоторого события A в n независимых испытаниях, в каждом из которых вероятность p появления этого события одна и та же. Обозначим через $\bar{A}$ событие, заключающееся в неоявлении события A . Очевидно, что

$$P(\bar{A}) = 1 - P(A) = 1 - p.$$

Из определения независимости испытаний нетрудно усмотреть, что вероятность какой-либо определенной последовательности, составленной из k появлений A и $n - k$ неоявлений A , равна

$$p^k(1 - p)^{n-k}. \quad (5)$$

Так, например, при $n = 5$ и $k = 2$ вероятность получить последовательность исходов $A\bar{A}A\bar{A}A$ будет $p(1 - p)p(1 - p)(1 - p) = p^2(1 - p)^3$.

По теореме сложения вероятность P_k равна сумме вероятностей всех последовательностей с k появлениями и $n - k$ неоявлениями события A , т.е. в силу (5) равна произведению числа таких последовательностей на $p^k(1 - p)^{n-k}$. Число таких последовательностей, очевидно, равно числу сочетаний из n по k , поскольку k положительных исходов могут занимать в ряду из n испытаний любые k мест.

Окончательно получаем

$$P_k = C_n^k p^k (1 - p)^{n-k} \quad (k = 0, 1, \dots, n) \quad (6)$$

(так называемое биномиальное распределение).

Чтобы увидеть, как применяются приведенные выше определения и формулы, рассмотрим пример, относящийся к теории стрельбы.

Пусть для поражения цели достаточно пяти попаданий. Нас интересует вопрос, имеем ли мы право рассчитывать на то, что необходимые пять попаданий получатся в результате 40 выстрелов. Чисто эмпирический метод решения этой задачи заключался бы в следующем. При

¹⁾ Точнее *независимость испытаний* означает следующее. Разобьем n испытаний каким-либо образом на две группы; пусть событие A заключается в том, что все испытания первой группы заканчиваются какими-либо наперед заданными исходами, а событие B — в том, что все испытания второй группы заканчиваются какими-либо исходами, также наперед заданными. Испытания называются *независимыми* (в совокупности), если при любом разбиении и любом задании исходов определенные выше события A и B независимы в смысле (4).

К обсуждению реального смысла понятия независимости мы еще вернемся в § 4.

заданных размерах цели и заданной дистанции стрельбы производится много (скажем, 200) стрельб по 40 выстрелов в каждой и определяется, в каком числе стрельб получилось не менее пяти попаданий в цель. Если этот результат был достигнут, например, в 195 стрельбах из 200, то вероятность P равна приблизительно

$$P = \frac{195}{200} = 0,975.$$

По рассмотренному чисто эмпирическому рецепту исследования мы потратили бы 8000 снарядов для решения крайне специальной задачи. Так на практике, конечно, не поступают. Вместо этого начинают с исследования рассеивания снарядов при данной дистанции стрельбы независимо от размеров цели. Оказывается, что отклонения по дальности и боковые отклонения от средней точки падения подчиняются в смысле частоты, с которой встречаются отклонения различных размеров, закону, изображенному на рис. 2.

Рис. 2

Буквой B здесь обозначено так называемое вероятное отклонение. Вероятное отклонение, вообще говоря, различно для отклонений по дальности и боковых отклонений и, кроме того, увеличивается с увеличением дистанции стрельбы. Вероятные отклонения для различных дистанций для каждого типа орудия и снаряда находятся эмпирически при помощи опытных стрельб на артиллерийском полигоне. После же этого решение всевозможных специальных задач такого типа, как поставленная выше, производится расчетным путем.

Предположим для простоты, что интересующая нас цель имеет вид прямоугольника, одна сторона которого направлена вдоль линии стрельбы и имеет размеры в два вероятных отклонения по дальности, а другая сторона, перпендикулярная линии стрельбы, равна двум вероятным боковым отклонениям. Предположим, далее, что цель хорошо пристреляна и средняя траектория полета снарядов проходит через ее центр (рис. 3).

Рис. 3

Предположим еще, что боковое отклонение и отклонение по дальности независимы¹⁾. Тогда для попадания в цель данным снарядом необходимо и достаточно, чтобы его отклонения по дальности и боковое не превосходили соответствующих вероятных отклонений. В соответствии с рис. 2 каждое из этих событий будет наблюдаться примерно для 50% выпущенных снарядов, т.е. с вероятностью $\frac{1}{2}$. Совмещение обоих этих событий будет происходить примерно для 25% выпущенных снарядов, т.е. вероятность попадания отдельного снаряда в цель будет равна

$$p = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4},$$

а вероятность промаха при отдельном выстреле будет равна

$$q = 1 - p = 1 - \frac{1}{4} = \frac{3}{4}.$$

Предполагая, что попадания при отдельных выстрелах представляют собой независимые события, и применяя биномиальную формулу (6), мы находим, что вероятность получить при 40 выстрелах ровно k попаданий будет равна

$$P_k = C_{40}^k p^k q^{40-k} = \frac{40 \cdot 39 \cdots (39-k)}{1 \cdot 2 \cdots k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{40-k}.$$

Интересующая нас вероятность получить не менее пяти попаданий выразится теперь формулой $P = \sum_{k=5}^{40} P_k$. Но ее проще вычислить по формуле $P = 1 - Q$, исходя из вероятности $Q = \sum_{k=0}^4 P_k$ получить менее пяти попаданий.

Можно подсчитать, что

$$\begin{aligned} P_0 &= \left(\frac{3}{4}\right)^{40} \sim 0,00001, \\ P_1 &= 40 \left(\frac{3}{4}\right)^{39} \frac{1}{4} \sim 0,00013, \\ P_2 &= \frac{40 \cdot 39}{2} \left(\frac{3}{4}\right)^{38} \left(\frac{1}{4}\right)^2 \sim 0,00087, \\ P_3 &= \frac{40 \cdot 39 \cdot 38}{2 \cdot 3} \left(\frac{3}{4}\right)^{37} \left(\frac{1}{4}\right)^3 \sim 0,0037, \\ P_4 &= \frac{40 \cdot 39 \cdot 38 \cdot 37}{2 \cdot 3 \cdot 4} \left(\frac{3}{4}\right)^{36} \left(\frac{1}{4}\right)^4 \sim 0,0113, \end{aligned}$$

откуда

$$Q = 0,016, \quad P = 0,984.$$

¹⁾ Эти предположения независимости подтверждаются опытом.

Полученная вероятность P даже несколько ближе к единице, чем это обычно признается достаточным в теории стрельбы при назначении числа снарядов, способного обеспечить выполнение поставленной задачи. Чаще всего считают возможным указывать число снарядов, которое гарантирует выполнение поставленной задачи с вероятностью 0,95.

Рассмотренный пример несколько схематичен, но он достаточно убедительно показывает важность вероятностных расчетов. Установив из опыта зависимость вероятных отклонений от дистанции стрельбы, для чего достаточно совсем небольшого числа стрельб на полигоне, мы можем потом при помощи несложных расчетов получать ответы на самые разнообразные вопросы. Так же дело обстоит и во всех других областях, где совместное действие большого числа случайных факторов приводит к статистическим закономерностям. При непосредственной обработке массовых наблюдений выясняются лишь самые простые из этих статистических закономерностей, т.е. находятся лишь некоторые исходные вероятности. Затем, при помощи законов теории вероятностей, отправляясь от этих исходных вероятностей, вычисляют вероятности более сложных явлений и на основе этих вычислений делают выводы о статистических закономерностях, управляющих интересующими нас сложными явлениями.

Иногда удается и совсем обойтись без собирания массового статистического материала, так как исходные вероятности могут быть определены из достаточно убедительных соображений симметрии. Например, традиционный вывод о том, что игральная кость, т.е. куб, вырезанный из однородного материала, при бросании с достаточной высоты падает на каждую из своих граней с одинаковой вероятностью $\frac{1}{6}$, был сделан несомненно раньше, чем вполне систематически был собран достаточный для оправдания этого вывода наблюдательный материал. Систематические опыты такого рода производились в XVIII–XX вв. преимущественно составителями учебников по теории вероятностей, когда теория вероятностей уже была разработанной наукой. Результат этой проверки был удовлетворителен, но распространение подобной деятельности на новые аналогичные случаи вряд ли представляет интерес. Например, насколько нам известно, никто не производил достаточно обширных опытов с бросанием вырезанного из однородного материала правильного двенадцатигранника. Но нет никаких сомнений, что если бы было произведено, скажем, 12 000 таких бросаний, то двенадцатигранник упал бы на каждую из своих граней приблизительно в тысяче случаев.

Получение исходных вероятностей из соображений симметрии или однородности играет также большую роль во многих серьезных научных задачах, например во всех задачах на столкновения или сближения беспорядочно двигающихся молекул газа или — с таким же успехом — звезд

галактики. Конечно, в таких более деликатных случаях предпочитают хотя бы косвенно проверять сделанные допущения путем сравнения вытекающих из них выводов с опытом.

§ 3. Закон больших чисел и предельные теоремы

Вполне естественна потребность количественно уточнить утверждение о том, что в «больших» сериях испытаний частоты появления события «близки» к его вероятности. Следует ясно представить себе известную деликатность этой задачи. В наиболее типичных для теории вероятностей случаях дело обстоит так, что в сколь угодно длинных сериях испытаний остаются теоретически возможными оба крайних значения частоты

$$\frac{\mu}{n} = \frac{n}{n} = 1 \quad \text{и} \quad \frac{\mu}{n} = \frac{0}{n} = 0.$$

Поэтому, каково бы ни было число испытаний n , нельзя утверждать с полной достоверностью, что будет выполнено, скажем, неравенство

$$\left| \frac{\mu}{n} - p \right| < \frac{1}{10}.$$

Например, если событие A заключается в выпадении при бросании игральной кости шестерки, то при n бросаниях с вероятностью $(\frac{1}{6})^n > 0$ мы все время будем получать одни шестерки, т.е. с вероятностью $(\frac{1}{6})^n$ получим частоту появления шестерок, равную единице, а с вероятностью $(1 - \frac{1}{6})^n > 0$ шестерка не выпадает ни одного раза, т.е. частота появления шестерок окажется равной нулю.

Во всех подобных задачах любая нетривиальная оценка близости между частотой и вероятностью действует не с полной достоверностью, а лишь с некоторой меньшей единицы вероятностью. Можно, например, доказать, что в случае независимых испытаний¹⁾ с постоянной вероятностью p появления события неравенство

$$\left| \frac{\mu}{n} - p \right| < 0,02 \tag{7}$$

для частоты μ/n будет выполняться при $n = 10\,000$ (и любом p) с вероятностью

$$P > 0,9999. \tag{8}$$

Здесь мы прежде всего хотим подчеркнуть, что в приведенной формулировке количественная оценка близости частоты μ/n к вероятности p связана с введением новой вероятности P .

¹⁾ См. сноску на с. 219. Доказательство оценки (8) пояснено на с. 225.

Реальный смысл оценки (8) таков: если произвести N серий по n испытаний и сосчитать число M серий, в которых выполняется неравенство (7), то при достаточно большом N приближенно будет

$$\frac{M}{N} \approx P > 0,9999. \quad (9)$$

Но если мы захотим уточнить соотношение (9) как в отношении степени близости M/N к P , так и в отношении надежности, с которой можно утверждать, что такая близость будет иметь место, то придется обратиться к рассмотрениям, аналогичным тем, которые мы уже провели в применении к близости μ/n и p . При желании такое рассуждение можно повторять неограниченное число раз, но вполне понятно, что это не позволит нам совсем освободиться от необходимости на последнем этапе обратиться к вероятностям в примитивном грубом понимании этого термина.

Не следует думать, что подобного рода затруднения являются какой-то особенностью теории вероятностей. При математическом изучении реальных явлений мы всегда их схематизируем. Отклонения хода действительных явлений от теоретической схемы можно, в свою очередь, подвергнуть математическому изучению. Но для этого сами эти отклонения надо уложить в некоторую схему и этой последней пользоваться уже без формального математического анализа отклонений от нее.

Заметим, впрочем, что при реальном применении оценки¹⁾

$$P \left\{ \left| \frac{\mu}{n} - p \right| < 0,02 \right\} > 0,9999 \quad (10)$$

к единичной серии из n испытаний мы опираемся и на некоторые соображения симметрии: неравенство (10) указывает, что при очень большом числе N серий соотношение (7) будет выполняться не менее чем в 99,99% случаев; естественно с большой уверенностью ожидать, что, в частности, неравенство (7) осуществится в интересующей нас определенной серии из n испытаний, если мы имеем основания считать, что эта серия в ряду других серий занимает рядовое, ничем особенным не отмеченное положение.

Вероятности, которыми принято пренебрегать в различных практических положениях, различны. Выше уже отмечалось, что при ориентировочных расчетах расхода снарядов, гарантирующего выполнение поставленной задачи, удовлетворяются нормой расхода снарядов, при которой поставленная задача решается с вероятностью 0,95, т.е. пренебрегают вероятностями, не превышающими 0,05. Это объясняется тем, что переход на расчеты, исходящие из пренебрежения, скажем, лишь

¹⁾ Так записывают оценку (8) для вероятности неравенства (7).

вероятностями, меньшими 0,01, приводил бы к большому увеличению норм расхода снарядов, т.е. практически во многих случаях к выводу о невозможности выполнить поставленную задачу за тот короткий промежуток времени, который для этого имеется, или с фактически могущим быть использованным запасом снарядов.

Иногда и в научных исследованиях ограничиваются статистическими приемами, рассчитанными исходя из пренебрежения вероятностями в 0,05. Но это следует делать лишь в случаях, когда собирание более обширного материала очень затруднительно. Рассмотрим в виде примера таких приемов следующую задачу. Допустим, что в определенных условиях употребительный препарат для лечения какого-либо заболевания дает положительный результат в 50%, т.е. с вероятностью 0,5. Предлагается новый препарат и для проверки его преимуществ над старым планируется применить его в десяти случаях, выбранных беспристрастно из числа больных, находящихся в том же положении, что и те, для которых установлена эффективность старого препарата в 50%. При этом устанавливается, что преимущество нового препарата будет считаться доказанным, если он даст положительный результат не менее чем в восьми случаях из десяти. Легко подсчитать, что такое решение связано с пренебрежением вероятностью получить ошибочный вывод (т.е. вывод о доказанности преимущества нового препарата, в то время как он равноценен или даже хуже старого) как раз порядка 0,05. В самом деле, если в каждом из десяти испытаний вероятность положительного исхода равна p , то вероятности получить при десяти испытаниях 10, 9 или 8 положительных исходов, равны соответственно

$$P_{10} = p^{10}, \quad P_9 = 10 p^9(1 - p), \quad P_8 = 45 p^8(1 - p)^2.$$

В сумме для случая $p = \frac{1}{2}$ получаем

$$P = P_{10} + P_9 + P_8 = \frac{56}{1024} \sim 0,05.$$

Таким образом, в предположении, что на самом деле новый препарат точно равноценен старому, мы рискуем сделать ошибочный вывод о том, что новый препарат превосходит старый, с вероятностью порядка 0,05. Чтобы свести эту вероятность приблизительно к 0,01, не увеличивая числа испытаний $n = 10$, пришлось бы установить, что преимущество нового препарата будет считаться доказанным лишь тогда, когда его применение даст положительный результат не менее чем в девяти случаях из десяти. Если это требование покажется сторонникам нового препарата слишком суровым, то придется назначить число испытаний n значительно большим, чем 10. Если, например, при $n = 100$ установить, что преимущества нового препарата будут считаться доказанными при $\mu > 65$, то вероятность ошибки будет лишь $P \approx 0,0015$.

Если норма в 0,05 для серьезных научных исследований явно недостаточна, то вероятностью ошибки в 0,001 или в 0,003 по большей части принято пренебрегать даже в столь академических и обстоятельных исследованиях, как обработка астрономических наблюдений. Впрочем, иногда научные выводы, основанные на применении вероятностных закономерностей, обладают и значительно большей достоверностью (т.е. построены на пренебрежении значительно меньшими вероятностями). Об этом еще будет сказано далее.

В рассмотренных примерах мы уже неоднократно применяли частные случаи биномиальной формулы (6)

$$P_m = C_n^m p^m (1-p)^{n-m}$$

для вероятности P_m получить ровно m положительных исходов при n независимых испытаниях, в каждом из которых положительный исход имеет вероятность p . Рассмотрим при помощи этой формулы вопрос, поставленный в начале этого параграфа, о вероятности

$$P = \mathbf{P}\left\{\left|\frac{\mu}{n} - p\right| < \varepsilon\right\}, \quad (11)$$

где μ — фактическое число положительных исходов¹⁾. Очевидно, эта вероятность может быть записана в виде суммы тех P_m , для которых m удовлетворяет неравенству

$$\left|\frac{m}{n} - p\right| < \varepsilon, \quad (12)$$

т.е. в виде

$$P = \sum_{m=m_1}^{m_2} P_m, \quad (13)$$

где m_1 — наименьшее из значений m , удовлетворяющих неравенству (12), а m_2 — наибольшее из таких m .

Формула (13) при сколько-нибудь больших n мало пригодна для непосредственных вычислений. Поэтому имело очень большое значение открытие Муавром для случая $p = \frac{1}{2}$ и Лапласом при любом p асимптотической формулы, которая позволяет очень просто находить и изучать поведение вероятностей P_m при больших n . Формула эта имеет вид

$$P_m \sim \frac{1}{\sqrt{2\pi np(1-p)}} e^{-(m-np)^2/[2np(1-p)]}. \quad (14)$$

¹⁾ μ принимает с вероятностями P_m значения $m = 0, 1, \dots, n$, т.е. $\mathbf{P}\{\mu = m\} = P_m$.

Если p не слишком близко к нулю или единице, то она достаточно точна уже при n порядка 100. Если положить

$$t = \frac{m - np}{\sqrt{np(1-p)}}, \quad (15)$$

то формула (14) приобретет вид

$$P_m \sim \frac{1}{\sqrt{2\pi np(1-p)}} e^{-t^2/2}. \quad (16)$$

Из (13) и (16) можно вывести приближенное представление вероятности (11)

$$P \sim \frac{1}{\sqrt{2\pi}} \int_{-T}^T e^{-t^2/2} dt = F(T), \quad (17)$$

где

$$T = \varepsilon \sqrt{\frac{n}{p(1-p)}}. \quad (18)$$

Разность между левой и правой частями в (17) при постоянном и отличном от нуля и единицы p стремится при $n \rightarrow \infty$ равномерно относительно ε к нулю. Для функции $F(T)$ составлены подробные таблицы. Вот краткая выдержка из них

T	1	2	3	4
F	0,68269	0,95450	0,99730	0,99993

При $T \rightarrow \infty$ значение функции $F(T)$ стремится к единице.

Произведем при помощи формулы (17) оценку вероятности

$$\mathbf{P}\left\{\left|\frac{\mu}{n} - p\right| < 0,02\right\}$$

при $n = 10\,000$. Так как $T = 2/\sqrt{p(1-p)}$, то

$$P \approx F\left(\frac{2}{\sqrt{p(1-p)}}\right).$$

Так как функция $F(T)$ монотонно возрастает с возрастанием T , то для не зависящей от p оценки P снизу надо взять наименьшее возможное (при различных p) значение T . Такое наименьшее значение получится при $p = \frac{1}{2}$, и оно будет равно 4. Поэтому приближенно

$$P \geq F(4) = 0,99993. \quad (19)$$

В неравенстве (19) не учтена ошибка, происходящая из-за приближенного характера формулы (17). Производя оценку связанной с этим обстоятельством погрешности, можно во всяком случае установить, что $P > 0,9999$.

В связи с рассмотренным примером применения формулы (17) следует отметить, что оценки остаточного члена формулы (17), дававшиеся в теоретических сочинениях по теории вероятностей, долго оставались мало удовлетворительными. Поэтому применения формулы (17) и ей подобных к расчетам при не очень больших n или при вероятностях p , очень близких к 0 или к 1 (а такие вероятности во многих случаях и имеют особенно большое значение) часто основывались лишь на опыте проверок такого рода результатов для ограниченного числа примеров, а не на достоверно установленных оценках возможной ошибки. Более подробное исследование, кроме того, показало, что во многих практически важных случаях приведенные выше асимптотические формулы нуждаются не только в оценке остаточного члена, но и в уточнении (так как без такого уточнения остаточный член слишком велик). В обоих направлениях наиболее полные результаты принадлежат С. Н. Бернштейну.

Соотношения (11), (17) и (18) можно переписать в виде

$$P \left\{ \left| \frac{\mu}{n} - p \right| < t \sqrt{\frac{p(1-p)}{n}} \right\} \sim F(t). \quad (20)$$

Для достаточно больших t правая часть формулы (20), не содержащая n , сколь угодно близка к единице, т.е. к значению вероятности, которое соответствует полной достоверности. Мы видим, таким образом, что, как правило, отклонения частоты μ/n от вероятности p имеют порядок $1/\sqrt{n}$. Такая пропорциональность точности действия вероятностных закономерностей квадратному корню из числа наблюдений типична и для многих других вопросов. Иногда говорят даже в порядке несколько упрощенной популяризации о «законе квадратного корня из n » как основном законе теории вероятностей. Полную отчетливость эта мысль получила благодаря введению великим русским математиком П. Л. Чебышевым в систематическое употребление метода сведения различных вероятностных задач к подсчетам «математических ожиданий» и «дисперсий» для сумм и средних арифметических «случайных величин».

Случайной величиной называется величина, которая в данных условиях S может принимать различные значения с определенными вероятностями. Для нас достаточно рассмотреть случайные величины, могущие принимать лишь конечное число различных значений. Чтобы указать, как говорят, *распределение вероятностей* такого рода случайной величины ξ , достаточно указать возможные ее значения $x_1, x_2, \dots, x_s$ и

вероятности

$$P_r = \mathbf{P}\{\xi = x_r\},$$

В сумме эти вероятности по всем различным возможным значениям величины ξ всегда равны единице:

$$\sum_{r=1}^s P_r = 1.$$

Примером случайной величины может служить изучавшееся выше число μ положительных исходов при n испытаниях.

Математическим ожиданием величины ξ называется выражение

$$\mathbf{M}(\xi) = \sum_{r=1}^s P_r x_r,$$

а дисперсией величины ξ называют математическое ожидание квадрата отклонения $\xi - \mathbf{M}(\xi)$, т.е. выражение

$$\mathbf{D}(\xi) = \sum_{r=1}^s P_r (x_r - \mathbf{M}(\xi))^2.$$

Корень квадратный из дисперсии

$$\sigma_\xi = \sqrt{\mathbf{D}(\xi)}$$

называется *средним квадратическим отклонением* (величины от ее математического ожидания $\mathbf{M}(\xi)$).

В основе простейших применений дисперсий и средних квадратических отклонений лежит знаменитое неравенство Чебышева

$$\mathbf{P}\{|\xi - \mathbf{M}(\xi)| \leq t\sigma_\xi\} \geq 1 - \frac{1}{t^2}. \quad (21)$$

Оно показывает, что отклонения ξ от $\mathbf{M}(\xi)$, значительно превышающие σ_ξ , встречаются редко.

При образовании сумм случайных величин

$$\xi = \xi^{(1)} + \xi^{(2)} + \dots + \xi^{(n)}$$

для их математических ожиданий всегда имеет место равенство

$$\mathbf{M}(\xi) = \mathbf{M}(\xi^{(1)}) + \mathbf{M}(\xi^{(2)}) + \dots + \mathbf{M}(\xi^{(n)}). \quad (22)$$

Аналогичное равенство для дисперсий

$$\mathbf{D}(\xi) = \mathbf{D}(\xi^{(1)}) + \mathbf{D}(\xi^{(2)}) + \dots + \mathbf{D}(\xi^{(n)}) \quad (23)$$

верно только при некоторых ограничениях. Для справедливости равенства (23) достаточно, например, чтобы величины $\xi^{(i)}$ и $\xi^{(j)}$ с различными номерами не были, как говорят, «коррелированы» между собой, т.е. чтобы при $i \neq j$ выполнялось равенство¹⁾

$$\mathbf{M}\{(\xi^{(i)} - \mathbf{M}(\xi^{(i)}))(\xi^{(j)} - \mathbf{M}(\xi^{(j)}))\} = 0. \quad (24)$$

В частности, равенство (24) соблюдается, если величины $\xi^{(i)}$ и $\xi^{(j)}$ независимы между собой²⁾. Таким образом, для взаимно независимых слагаемых всегда действует равенство (23). Для средних арифметических

$$\zeta = \frac{1}{n} (\xi^{(1)} + \xi^{(2)} + \dots + \xi^{(n)})$$

из (23) вытекает

$$\mathbf{D}(\zeta) = \frac{1}{n^2} (\mathbf{D}(\xi^{(1)}) + \mathbf{D}(\xi^{(2)}) + \dots + \mathbf{D}(\xi^{(n)})). \quad (25)$$

Предположим теперь, что для всех слагаемых дисперсии не превосходят некоторой постоянной

$$\mathbf{D}(\xi^{(i)}) \leq C^2.$$

Тогда по (25)

$$\mathbf{D}(\zeta) \leq \frac{C^2}{n},$$

и в силу неравенства Чебышева при любом t

$$\mathbf{P}\left\{|\zeta - \mathbf{M}(\zeta)| \leq \frac{tC}{\sqrt{n}}\right\} \geq 1 - \frac{1}{t^2}. \quad (26)$$

Неравенство (26) содержит в себе так называемый закон больших чисел в форме, установленной Чебышевым: если величины $\xi^{(i)}$ взаимно

¹⁾ Коэффициентом корреляции между величинами $\xi^{(i)}$ и $\xi^{(j)}$ называется выражение

$$R = \frac{\mathbf{M}\{(\xi^{(i)} - \mathbf{M}(\xi^{(i)}))(\xi^{(j)} - \mathbf{M}(\xi^{(j)}))\}}{\sigma_{\xi^{(i)}} \sigma_{\xi^{(j)}}}.$$

Если $\sigma_{\xi^{(i)}} > 0$ и $\sigma_{\xi^{(j)}} > 0$, то условие (24) равносильно тому, что $R = 0$.

Коэффициент корреляции R характеризует степень зависимости между случайными величинами. Всегда $|R| \leq 1$, причем $R = \pm 1$ только при наличии линейной связи

$$\eta = a\xi + b \quad (a \neq 0).$$

Для независимых величин $R = 0$.

²⁾ Независимость двух случайных величин ξ и η , способных принимать, соответственно, значения $x_1, x_2, \dots, x_m$ и $y_1, y_2, \dots, y_n$, по определению обозначает, что при любых i и j события $A_i = \{\xi = x_i\}$ и $B_j = \{\eta = y_j\}$ независимы в смысле определения, данного в § 2.

независимы и имеют ограниченные дисперсии, то при возрастании n их средние арифметические ζ все реже заметно отклоняются от своих математических ожиданий $M(\zeta)$.

Более точно говорят, что *последовательность величин*

$$\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(n)}, \dots$$

подчиняется закону больших чисел, если для соответствующих средних арифметических ζ и при любом постоянном $\varepsilon > 0$

$$P\{|\zeta - M(\zeta)| \leq \varepsilon\} \rightarrow 1 \quad (27)$$

при $n \rightarrow \infty$.

Чтобы получить из неравенства (26) предельное соотношение (27), достаточно положить

$$t = \varepsilon \frac{\sqrt{n}}{C}.$$

Большой ряд исследований А. А. Маркова, С. Н. Бернштейна, А. Я. Хинчина и других посвящен вопросу возможно большего расширения условий применимости предельного соотношения (27), т.е. условий применимости закона больших чисел. Эти исследования имеют принципиальное значение. Однако еще более важным является точное исследование распределения вероятностей отклонений $\zeta - M(\zeta)$.

Великой заслугой русской классической школы в теории вероятностей является установление того факта, что при очень широких условиях асимптотически (т.е. со все большей точностью при неограниченно растущих n) справедливо равенство

$$P\{t_1\sigma_\zeta < \zeta - M(\zeta) < t_2\sigma_\zeta\} \sim \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-t^2/2} dt. \quad (28)$$

Чебышев дал почти полное доказательство этой формулы для случая независимых и ограниченных слагаемых. Марков восполнил недостающее звено в рассуждениях Чебышева и расширил условия применимости формулы (28). Еще более общие условия были даны Ляпуновым. Вопрос о распространении формулы (28) на суммы зависимых слагаемых с особенной полнотой был изучен С. Н. Бернштейном.

Формула (28) охватила столь большое число частных задач, что долгое время ее называли центральной предельной теоремой теории вероятностей. Хотя при новейшем развитии теории вероятностей она оказалась включенной в ряд более общих закономерностей, ее значение трудно переоценить и в настоящее время.

Если слагаемые независимы и их дисперсии одинаковы и равны:

$$D(\xi^{(i)}) = \sigma^2,$$

то формуле (28) удобно, учитывая соотношение (25), придать вид

$$\mathbf{P}\left\{\frac{t_1\sigma}{\sqrt{n}} < \zeta - \mathbf{M}(\zeta) < \frac{t_2\sigma}{\sqrt{n}}\right\} \sim \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-t^2/2} dt. \quad (29)$$

Покажем, что соотношение (29) содержит в себе решение задачи об отклонениях частоты μ/n от вероятности p , которой мы занимались ранее. Для этого введем случайные величины $\xi^{(i)}$, определяя их следующим условием:

$$\xi^{(i)} = \begin{cases} 0, & \text{если } i\text{-е испытание имело отрицательный исход,} \\ 1, & \text{если } i\text{-е испытание имело положительный исход.} \end{cases}$$

Легко проверить, что тогда

$$\mu = \xi^{(1)} + \xi^{(2)} + \dots + \xi^{(n)}, \quad \frac{\mu}{n} = \zeta,$$

$$\mathbf{M}(\xi^{(i)}) = p, \quad \mathbf{D}(\xi^{(i)}) = p(1-p), \quad \mathbf{M}(\zeta) = p,$$

и формула (29) дает

$$\mathbf{P}\left\{t_1\sqrt{\frac{p(1-p)}{n}} < \frac{\mu}{n} - p < t_2\sqrt{\frac{p(1-p)}{n}}\right\} \sim \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-t^2/2} dt,$$

что при $t_1 = -t$, $t_2 = t$ снова приводит к формуле (20).

§ 4. Дополнительные замечания об основных понятиях теории вероятностей

Говоря о случайных явлениях, которым свойственна устойчивость частот, т.е. тенденция при большом числе повторений известных условий происходить с частотами, группирующимися вокруг некоторого *нормального уровня* — вероятности $\mathbf{P}(A/S)$, мы допустили в § 1 некоторую неточность и расплывчатость формулировок в двух отношениях. Первая из допущенных неточностей заключается в том, что мы не указали, насколько велики должны быть по своей численности серии испытаний n_r для того, чтобы устойчивость частот уже обязана была проявиться, и каковы именно допустимые отклонения частот μ_r/n_r друг от друга и от нормального их уровня p при тех или иных численностях отдельных серий $n_1, n_2, \dots, n_s$. Эта неточность на первом этапе формирования понятий новой науки неизбежна. Она несколько не больше, чем известная расплывчатость, свойственная простейшим геометрическим понятиям точки или прямой в их *физическом* понимании. Эта сторона дела была затем уточнена в § 3.

Существенное другая скрытая в наших формулировках неясность, относящаяся к способу формирования тех серий, в которых должна наблюдаться устойчивость частот появления события A .

Мы уже видели, что к статистическим и вероятностным методам исследования обращаются тогда, когда точное индивидуальное предсказание хода событий оказывается неосуществимым. Желая же искусственно создать по возможности чисто случайные явления, специально заботятся о том, чтобы никакими доступными средствами нельзя было заранее выделить те случаи, в которых явление A будет иметь тенденцию появляться чаще, чем с некоторой нормальной для него частотой.

Так организуются, например, тиражи государственных займов. Если в данном тираже из общего числа N облигаций на M из них выпадает выигрыш, то вероятность выигрыша для отдельной облигации равна $p = M/N$. Это значит, что каким бы образом мы ни выделили заранее до тиража совокупность облигаций достаточно большой численности n , мы можем быть практически уверены, что отношение μ/n числа μ выигравших облигаций в выделенной совокупности к общей численности n этой совокупности окажется близким к p . Например, лица, предпочитающие приобретать четные номера облигаций, не получают никакого систематического преимущества перед теми, которые предпочитают приобретать нечетные номера; точно так же не получают никакого преимущества лица, которые исходили бы из убеждения, что лучше всего приобретать облигации с номерами, разлагающимися ровно на три простых множителя, или облигации, номера которых являются соседними с теми, на которые упали выигрыши в предшествующем тираже, и т.п.

Точно так же при стрельбе из исправного орудия данного образца, с хорошо обученным обслуживающим персоналом и при нормальном способе получения снарядов, прошедших обычный для выпускаемой продукции контроль, мы будем получать отклонения от средней точки падения меньше определенного заранее вероятного отклонения B приблизительно в половине случаев. Эта пропорция сохранится в ряде последовательных стрельб, сохранится она и в том случае, если мы отдельно подсчитаем число отклонений, меньших B , для четных или нечетных (по порядку их во времени) выстрелов и т.п. Но вполне возможно, что, произведя отбор особенно однородных (в отношении их веса и т. п.) снарядов, мы могли бы рассеивание несколько уменьшить, т.е. получить серию снарядов, для которой доля отклонений, больших стандартного B , окажется существенно меньше *половины*.

Итак, говорить о том, что событие A является «вероятностно-случайным» и приписывать ему определенную вероятность

$$p = P(A/S)$$

можно только тогда, когда указан класс допустимых способов формирования серий испытаний. Указание этого класса мы будем считать включенным в условия S .

При заданных условиях S свойство события A быть вероятностно-случайным и иметь вероятность $p = P(A/S)$ выражает *объективный* характер связи между условиями S и событием A . Иначе говоря, не существует событий абсолютно случайных, события являются случайными или необходимыми в зависимости от того, в какой связи они рассматриваются, но в определенных условиях событие может быть случайным совершенно объективно, и это его свойство не зависит от состояния знаний какого бы то ни было наблюдателя. Если вообразить наблюдателя, который мог бы улавливать во всех деталях отличительные свойства и особые обстоятельства полета снарядов и, следовательно, предсказывать индивидуальные для каждого из них отклонения от средней траектории, то его присутствие не помешало бы снарядам рассеиваться по законам теории вероятностей (если, конечно, стрельба будет производиться обычным способом, а не по указаниям нашего воображаемого наблюдателя).

Заметим по этому поводу, что обсуждавшееся выше формирование серий, в которых проявляется тенденция к постоянству частот в смысле их группирования вокруг нормального значения — вероятности, тоже происходит в реальной обстановке совершенно независимо от нашего вмешательства. Например, именно в силу вероятностно-случайного характера движений молекул в газе число молекул, ударяющихся даже за очень малые промежутки времени о какую-либо площадку стенки сосуда или поверхности помещенных в газе тел, оказывается с большой точностью пропорциональным площади этой площадки и длине промежутка времени. Отклонения от этой пропорциональности в тех случаях, когда число ударов невелико, тоже следуют законам теории вероятностей и вызывают явления типа броуновского движения, о чем будет идти речь далее.

Обратимся к выяснению реального смысла понятия независимости. Напомним, что условная вероятность события A при условии B определялась формулой

$$P(A | B) = \frac{P(\dot{A}B)}{P(B)}. \quad (30)$$

Напомним также, что события A и B назывались независимыми, если, согласно (4),

$$P(AB) = P(A)P(B).$$

Из независимости событий A и B и $P(B) > 0$ следует

$$P(A | B) = P(A).$$

Все теоремы математической теории вероятностей, говорящие о независимых событиях, применимы к любым событиям, удовлетворяющим

условию (4), или его обобщениям на случай взаимной независимости многих событий. Эти теоремы имели бы, однако, мало интереса, если бы это определение не находилось в связи со свойствами реально независимых (в причинном смысле) явлений.

Известно, например, что вероятность новорожденному оказаться мальчиком имеет довольно устойчивое значение $P(A) = \frac{22}{43}$. Если B обозначает условие, что рождение происходит в день соединения Юпитера с Марсом, то в предположении, что расположение планет не определяет индивидуальных судеб людей, условная вероятность $P(A|B)$ имеет то же самое значение: $P(A|B) = \frac{22}{43}$, т.е. фактический подсчет частоты рождения мальчиков при таких специальных астрологических условиях привел бы именно к частоте $\frac{22}{43}$. Хотя такой подсчет в достаточно обширных размерах, возможно, никем не производился, нет оснований сомневаться в его результате.

Мы привели этот несколько устарелый по содержанию пример для того, чтобы показать, что развитие человеческого познания состоит не только в установлении истинных связей между явлениями, но и в опровержении связей воображаемых, т.е. установлении в надлежащих случаях тезиса о независимости каких-либо двух кругов явлений. Разоблачение бессмысленности попыток астрологов связать между собою два круга явлений, друг с другом не связанных, является одним из классических тому примеров.

Естественно, что такого рода независимость не следует абсолютизировать. Например, в силу закона всемирного тяготения несомненно, что перемещение спутников Юпитера оказывает некоторое влияние, скажем, на полет артиллерийского снаряда. Но очевидно, что на практике с этими влияниями мы можем не считаться. С философской стороны, быть может, было бы правильнее вместо независимости говорить о несущественности в данной конкретной обстановке тех или иных зависимостей. Но как бы то ни было, независимость событий в объявленном сейчас конкретном и относительном понимании этого термина ни в какой мере не противоречит принципу всеобщей связи всех явлений, она лишь его необходимое дополнение.

Подсчеты вероятностей по формулам, вводимым из допущений о независимости тех или иных событий, имеют реальный интерес в том случае, когда события, бывшие сначала независимыми, ходом самих явлений приводятся затем в связь. Например, можно рассчитывать вероятности столкновения частиц космического излучения с частицами пронизываемой ими среды, исходя из допущения, что движение частиц среды до появления вблизи них быстро двигающейся частицы космического излучения происходит независимо от перемещения этой частицы. Можно рассчитывать вероятность попадания вражеской пули в лопасть вращающегося пропеллера, исходя из допущения, что его положение от-

носителю оси не зависит от траектории пули (конечно, это допущение было бы ошибочно по отношению к собственной пуле, выпускаемой при стрельбе через пропеллер согласованно с его вращением) и т.п.; число таких примеров может быть неограниченно продолжено.

Можно даже сказать, что всюду, где достаточно ясно проявляются вероятностные закономерности, мы имеем дело с влиянием большого числа если не совсем независимых между собой, то в том или ином смысле слабо связанных друг с другом факторов.

Это совсем не означает, что следует всюду некритически вводить допущения о тех или иных независимостях. Наоборот, это заставляет, во-первых, особенно тщательно разрабатывать критерии для проверки гипотез о независимости, а во-вторых, особенно тщательно исследовать пограничные случаи, когда зависимости между факторами уже необходимо учитывать, но эти зависимости таковы, что вероятностные закономерности в измененном и осложненном виде еще могут проявиться. Выше отмечалось, что уже классическая русская школа теории вероятностей широко развернула исследования во втором из этих направлений.

Заканчивая рассмотрение вопроса о независимости, отметим, что как определение независимости двух событий формулой (4), так и формальное определение независимости нескольких случайных величин значительно шире понятия реальной независимости в смысле принадлежности к причинно не связанным кругам явлений.

Предположим, например, что точка ξ падает на отрезок $[0, 1]$ так, что при

$$0 \leq a < b \leq 1$$

вероятность ее попадания на отрезок $[a, b]$ равна длине этого отрезка $b - a$. Легко доказать, что тогда в разложении

$$\xi = \frac{\alpha_1}{10} + \frac{\alpha_2}{100} + \frac{\alpha_3}{1000} + \dots$$

абсциссы точки ξ в десятичную дробь знаки α_k будут взаимно независимы, хотя они и связаны по своему происхождению¹⁾. (Из этого обстоятельства вытекает много теоретически, а частью и практически интересных следствий.)

Такая гибкость формального определения независимости не должна рассматриваться как его недостаток. Наоборот, она лишь расширяет

¹⁾ То же имеет место при любом n для знаков α_k разложения числа ξ в дробь

$$\xi = \frac{\alpha_1}{n} + \frac{\alpha_2}{n^2} + \frac{\alpha_3}{n^3} + \dots$$

область применимости теорем, установленных при тех или иных допущениях о независимости. Эти теоремы применимы одинаково и в случаях, где независимость постулируется в силу реальных соображений, и в случаях, где независимость доказывается расчетом, исходя из ранее принятых допущений о распределениях вероятностей исследуемых событий и случайных величин.

Вообще исследование формальной структуры математического аппарата теории вероятностей привело к интересным результатам. Оказалось, что этот аппарат занимает вполне определенное и очень простое место в намечающейся постепенно классификации основных объектов изучения современной математики.

Мы уже говорили о понятии совмещения AB двух событий и о понятии соединения $A \cup B$ события A и B . Напомним, что события называются несовместимыми, если их совмещение невозможно, т.е. если $AB = N$, где N — знак невозможного события.

Основной аксиомой элементарной теории вероятностей является требование (см. § 2), чтобы при условии $AB = N$ соблюдалось равенство

$$P(A \cup B) = P(A) + P(B).$$

Свойства основных понятий теории вероятностей — случайных событий и их вероятностей — вполне аналогичны, например, свойствам плоских фигур и их площадей. Достаточно понимать под AB пересечение (общую часть) двух фигур, под $A \cup B$ — их соединение, под N — условно вводимую «пустую» фигуру, а под $P(A)$ — площадь фигуры A и аналогия в намеченных пределах будет полной.

Таковыми же свойствами обладают объемы трехмерных фигур.

Наиболее общей теорией образований подобного типа, охватывающей как частные случаи теории объемов и площадей, является сейчас общая теория меры, некоторые сведения о которой можно найти в главе XV (том 3)*, посвященной теории функций действительного переменного.

Следует только отметить, что в теории вероятностей имеются по сравнению с общей теорией меры или специально с теорией площадей и объемов некоторые специфические черты: вероятность никогда не бывает больше единицы. Эту максимальную вероятность имеет необходимое событие U

$$P(U) = 1.$$

Аналогия не является поверхностной. Оказалось, что вся математическая теория вероятностей с формальной стороны может быть постро-

* См. сб. «Математика. Ее содержание, методы и значение», т. 3. М.: Издательство Академии наук СССР, 1956, с. 3–36. — *Прим. ред.*

ена как теория меры, специализированная допущением, что мера «всего пространства» U равна единице¹⁾.

Такой подход к делу не только внес полную ясность в формальное строение математической теории вероятностей, но и способствовал вполне конкретному прогрессу как самой теории вероятностей, так и смежных с ней по формальному строению математических теорий. В теории вероятностей с успехом были использованы тонкие методы, разработанные в метрической теории функций действительного переменного. Одновременно вероятностные методы оказались применимыми в вопросах смежных областей математики не «по аналогии», а путем формально строгого их перенесения в новые области. Всюду, где оказываются выполненными аксиомы теории вероятностей, применимы и следствия из этих аксиом, хотя бы данная область не имела ничего общего с реальной случайностью.

Наличие аксиоматизированной теории вероятностей избавляет нас от соблазна «определять» вероятность способами, претендующими на соединение их непосредственной естественно-научной убедительности с приспособленностью к построению на их основе формально строгой математической теории. Такие определения приблизительно соответствовали бы в геометрии «определению» точки как того, что получится, если бесконечное число раз обрезать со всех сторон физическое тело, уменьшая каждый раз, скажем, вдвое его диаметр.

К такого рода определениям относится определение вероятности как предела частот при неограниченном увеличении числа испытаний. Допущение о вероятностном характере испытаний, т.е. о тенденции частот группироваться вокруг постоянного значения, само по себе бывает верно (как и допущение о «случайности» какого-либо явления) лишь при сохранении некоторых условий, которые не могут сохраняться неограниченно долго и с неограниченной точностью. Поэтому точный переход к пределу

$$\frac{\mu}{n} \rightarrow p$$

не может иметь реального значения. Формулировка принципа устойчивости частот при обращении к такому предельному переходу требует определения допустимых способов отыскания бесконечных последовательностей испытаний, которое тоже может быть лишь математической фикцией. Все это нагромождение понятий могло бы еще подлежать серьезному рассмотрению, если бы в результате получилось построение

¹⁾ Тем не менее по существу решаемых задач теория вероятностей остается самостоятельной математической дисциплиной; основные для теории вероятностей результаты (подобные изложенным в § 3) кажутся искусственными и ненужными с точки зрения чистой теории меры.

теории столь своеобразной, что иными путями до ее строгого обоснования нельзя было бы дойти. Но, как указано выше, обоснование математической теории вероятностей при современном состоянии теории меры производится просто добавлением условия

$$P(U) = 1.$$

Вообще, при реальном анализе понятия вероятности вовсе не обязательно стремиться к его формальному определению. По-видимому, с чисто формальной стороны о вероятности нельзя сказать ничего больше следующего: вероятность $P(A/S)$ есть число, вокруг которого, при условиях S и при предусмотренных этими условиями способах формирования серий, имеют тенденцию группироваться частоты, причем при возрастании численности этих серий в разумных пределах, не нарушающих однородности условий, эта тенденция проявляется со все большей отчетливостью и точностью, достигая достаточных в данной конкретной обстановке надежности и точности при достижимых численностях серий.

Действительно, важной задачей является не формальное уточнение этого определения, а возможно более широкое выяснение условий, при которых такого типа вероятностная случайность должна проявляться. Надо ясно понимать то, что в действительности гипотеза о вероятностном характере какого-либо явления лишь очень редко обосновывается непосредственной статистической проверкой. Только при первом проникновении вероятностных методов в какую-либо новую область дело часто начиналось с того, что чисто эмпирически отмечалось постоянство частот. В силу сказанного в § 3, для того чтобы статистически обнаружить постоянство частот с точностью до ϵ , необходимо пользоваться сериями примерно по $n = 1/\epsilon^2$ испытаний. Например, для того чтобы установить, что в данном конкретном вопросе имеет смысл считать вероятность определенной с точностью до 0,0001, необходимо произвести ряд серий испытаний примерно по 100 000 000 испытаний в каждой.

Гораздо чаще гипотеза вероятностной случайности вводится на основании соображений симметрии или на основании соображений о практической независимости отдельных, приходящих в соприкосновение рядов явлений и т.д. Затем эта гипотеза проверяется косвенным путем. Например, благодаря тому, что число молекул в конечных объемах газа выражается числами порядка 10^{20} и более, число $\sqrt{n}$, соответствующее вероятным выводам кинетической теории газов, бывает очень велико и, действительно, многие из этих выводов оправдываются с большой точностью. Например, давление на противоположные стороны подвешенной в спокойном воздухе пластинки даже микроскопических для нашего глаза размеров оказывается строго одинаковым, хотя уже превышение давления на одну из сторон над давлением на другую в тысячные доли процента при надлежащей постановке опыта могло бы быть замечено.

§ 5. Детерминированные и случайные процессы

Принцип причинной обусловленности всех явлений находит наиболее простое математическое выражение в методе изучения реальных процессов при помощи дифференциальных уравнений; этот метод продемонстрирован на ряде примеров в § 1 главы V.

Пусть состояние изучаемой системы в момент времени t определяется при помощи n параметров

$$x_1, x_2, \dots, x_n.$$

Скорости изменения этих параметров, как известно, выражаются их производными по времени

$$\dot{x}_k = \frac{dx_k}{dt}.$$

Если допустить, что эти скорости являются функциями от значений параметров, то мы получим систему дифференциальных уравнений

$$\dot{x}_1 = f_1(x_1, x_2, \dots, x_n),$$

$$\dot{x}_2 = f_2(x_1, x_2, \dots, x_n),$$

$$\dots\dots\dots$$

$$\dot{x}_n = f_n(x_1, x_2, \dots, x_n).$$

Большая часть законов природы, открытых в эпоху зарождения математического естествознания, начиная с галилеевского закона падения тел, выражается именно таким образом. Галилей не мог облечь свое открытие в указанную стандартную форму из-за того, что в его время соответствующие математические понятия еще не были разработаны. Это было сделано Ньютоном.

Обычное в механике и многих других областях физики обращение к дифференциальным уравнениям второго порядка с принципиальной стороны не вносит ничего нового, так как, обозначая скорости $\dot{x}_k$ новыми знаками

$$v_k = \dot{x}_k,$$

мы получаем для вторых производных величин x_k выражение

$$\frac{d^2 x_k}{dt^2} = \dot{v}_k,$$

и уравнения второго порядка для n величин $x_1, x_2, \dots, x_n$ сводятся к уравнениям первого порядка для $2n$ величин $x_1, \dots, x_n, v_1, v_2, \dots, v_n$.

В виде примера рассмотрим задачу падения тяжелого тела в земной атмосфере. Предполагая, что в рассмотрение входят лишь небольшие высоты над земной поверхностью, будем считать сопротивление среды зависящим только от скорости, а не от высоты. Состояние изучаемой

системы характеризуется двумя параметрами: расстоянием тела от земной поверхности z и его скоростью v . Изменение этих двух величин во времени определяется двумя дифференциальными уравнениями

$$\begin{aligned}\dot{z} &= -v, \\ \dot{v} &= g - f(v),\end{aligned}\tag{31}$$

где g — ускорение силы тяжести, $f(v)$ — некоторый «закон сопротивления» для рассматриваемого тела.

Если скорости невелики и тело достаточно массивно, например в случае падения камня средних размеров с высоты в несколько метров, то сопротивлением воздуха можно пренебречь, и уравнения (31) превращаются в уравнения

$$\begin{aligned}\dot{z} &= -v, \\ \dot{v} &= g.\end{aligned}\tag{32}$$

Если предположить, что в начальный момент времени t_0 величины z и v имеют значения z_0 и v_0 , то легко решить уравнения (32) и получить формулу

$$z = z_0 - v(t - t_0) - g\left(\frac{t - t_0}{2}\right)^2,$$

описывающую весь процесс падения. Например, в предположении $t_0 = 0$, $v_0 = 0$ получаем формулу

$$z = z_0 - \frac{gt^2}{2},$$

открытую Галилеем.

В общем случае интегрирование уравнений (31) несколько сложнее, но принципиальный результат (при весьма общих ограничениях, наложенных на функцию $f(v)$) остается тем же: по значениям z_0 и v_0 в начальный момент времени t_0 однозначно вычисляются значения z и v для всех дальнейших моментов времени t вплоть до падения тела на поверхность земли. Можно мысленно снять и это последнее ограничение, предполагая, что падение продолжается и при отрицательных значениях z . Для схематизированной таким образом задачи можно установить следующее: если функция $f(v)$ при возрастании v монотонно возрастает и стремится к бесконечности при $v \rightarrow \infty$, то при неограниченном продолжении падения, т.е. при неограниченном возрастании переменного t , скорость v стремится к постоянному предельному значению c , которое является корнем уравнения

$$g = f(c).$$

С наглядной стороны этот результат математического анализа поставленной задачи понятен: скорость падения возрастает до тех пор,

пока ускорение силы тяжести не уравновесится сопротивлением воздуха. При прыжке с открытым парашютом стационарная скорость v около пяти метров в секунду устанавливается¹⁾ довольно скоро. При затяжном прыжке с нераскрытым парашютом сопротивление воздуха меньше, и в соответствии с этим стационарная скорость оказывается больше и устанавливается лишь после того, как парашютист пролетит очень большое расстояние.

При падении легких тел, подобных брошенному в воздух перу или пушинке, начальный период заметным образом ускоренного движения очень мал и иногда совсем незаметен. Стационарная скорость падения устанавливается очень быстро, и с известным приближением можно принять, что все время $v = c$. В этом случае остается одно дифференциальное уравнение

$$\dot{z} = -c,$$

которое интегрируется очень просто:

$$z = z_0 - c(t - t_0).$$

Так будет падать пушинка в совершенно спокойном воздухе.

В совершенно общем виде подчеркнутая нами выше детерминистическая концепция трактуется в современной теории динамических систем, которой посвящен ряд важных работ советских математиков — Н. Н. Боголюбова, В. В. Степанова и многих других. Эта общая теория охватывает как частные случаи и такие математические схемы реальных явлений, в которых состояние системы определяется уже не конечным числом параметров, а заданием одной или нескольких функций, как это типично, например, для механики непрерывных сред. В таких случаях элементарные закономерности изменения состояния за «бесконечно малый» промежуток времени задаются уже не обыкновенными дифференциальными уравнениями, а уравнениями в частных производных или иными средствами. Но общим во всех детерминированных математических схемах реальных процессов является то, что, во-первых, состояние изучаемой системы считается исчерпывающим образом определенным посредством задания некоторого математического объекта ω (системы n действительных чисел, одной или нескольких функций и т.п.); во-вторых, последующие значения для моментов времени $t > t_0$ однозначно определяются по значению ω_0 , соответствующему начальному моменту времени t_0

$$\omega = F(t_0, \omega_0, t).$$

Как мы уже видели, в случае процессов, описываемых дифференциальными уравнениями, нахождение функции F' требует интегрирования

¹⁾ Имеется в виду, что v становится практически достаточно близким к c .

этих дифференциальных уравнений с начальными условиями $\omega = \omega_0$ при $t = t_0$.

Представители механистического материализма считали, что описанная схема является точным и прямым выражением детерминированности реальных явлений — физического принципа причинности. По идее Лапласа, состояние мира в данный момент времени определяется бесконечным числом параметров, подчиненных бесконечному числу дифференциальных уравнений. Если бы некий «всеобъемлющий ум» мог записать все эти уравнения и их проинтегрировать, то, по Лапласу, он мог бы с полной точностью предвидеть всю эволюцию мира на протяжении бесконечного времени.

Однако на самом деле количественная математическая бесконечность крайне груба по сравнению с качественной неисчерпаемостью действительного мира. Ни введение бесконечного числа параметров, ни описание состояния непрерывных сред при помощи функций от точки пространства не являются адекватным отображением бесконечной сложности реальных явлений.

Как подчеркивалось в § 3 главы V, исследование реальных явлений не всегда идет в направлении увеличения числа вводимых в задачу параметров; вообще далеко не всегда целесообразно усложнять ту характеристику ω , которая определяет отдельное «состояние системы» в математической схеме, служащей для расчета данного явления. Искусство исследователя скорее состоит в том, чтобы найти очень простое фазовое пространство Ω (т.е. множество значений ω , или, иначе говоря, множество различных возможных состояний системы)*, которое позволяло бы тем не менее, заменяя реальный процесс процессом детерминированного перемещения точки ω в этом пространстве, ухватить все *существенные* стороны реального процесса.

Но, выделив из реального процесса его существенные черты, мы получаем некоторый остаток, который приходится считать случайным. Неучтенные случайные факторы всегда оказывают некоторое влияние на течение процесса. Существует очень мало явлений, подвергающихся математическому изучению, для которых при сопоставлении теории с наблюдениями нельзя было бы заметить влияния неучтенных случайных факторов. Таково или почти таково положение с теорией движения планет под действием силы тяготения: расстояния между планетами так велики по сравнению с их размерами, что идеализированное представление их материальными точками действует почти безотказно;

* В приведенном выше примере с падением тел фазовое пространство Ω есть система пар чисел (z, v) , т.е. плоскость. Вообще о фазовом пространстве см. главу XVII и главу XVIII (том 3), § 3. (См. сб. «Математика. Ее содержание, методы и значение», т. 3. М.: Издательство Академии наук СССР, 1956, с. 93–212. — *Прим. ред.*)

пространство, в котором они двигаются, заполнено столь разреженной материей, что ее сопротивление исчезающе мало; массы планет так велики, что световое давление при их движении почти не играет роли. Эти исключительные обстоятельства и приводят к тому, что математическое решение задачи о движении системы из n материальных точек, «состояние» которой описывается $6n$ параметрами¹⁾, а изменение состояния рассчитывается с учетом только силы тяготения, так поразительно хорошо совпадает с наблюдениями над движением планет.

Несколько приближается к случаю движения планет случай полета артиллерийского снаряда в предположении, что сопротивление воздуха введено в уравнения движения. Это тоже одна из классических областей, в которых математический метод исследования сравнительно легко и быстро одержал большие победы. Но здесь роль возмущающих случайных факторов уже значительно больше и рассеивание снарядов, т.е. их отклонения от теоретической траектории, соответствующей нормальным, назначенным для данного выстрела начальным условиям и среднему состоянию атмосферы во время стрельбы, достигают десятков, а на больших дистанциях и сотен метров. Отклонения эти частично вызываются случайными отклонениями начального направления и начальной скорости от нормы, частично случайными отклонениями от нормы массы и коэффициента сопротивления снаряда, частично же неравномерностью и порывами ветра и прочими случайными обстоятельствами, связанными с чрезвычайно сложным и подвижным режимом, господствующим в реальной земной атмосфере.

Рассеивание снарядов подробно изучается методами теории вероятностей, и результаты этого исследования весьма существенны для практики стрельбы.

Но что означает, собственно говоря, исследование случайных явлений? Казалось бы, когда случайный «остаток» при данной схематизации явления оказался настолько велик, что им нельзя пренебречь, то единственная возможность состоит в том, чтобы усложнить описание явления введением новых параметров и в усложненной схеме исследовать явление более подробно по той же математической схеме детерминированных явлений.

Такой путь во многих случаях практически неосуществим. Например, при исследовании падения материального тела в атмосфере с учетом неравномерной, порывистой (или, как обычно говорят, турбулентной) структуры ветрового потока вместо двух параметров z и u пришлось бы ввести совершенно необозримый математический аппарат полного описания этой структуры.

¹⁾ Тремя координатами и тремя составляющими скорости каждой точки.

Но этот сложный путь на самом деле необходим только в тех случаях, когда нам во что бы то ни стало нужно проследить влияние остаточных «случайных» факторов на течение нашего процесса во всех деталях и отдельно для каждого индивидуального случая. К счастью, в действительности очень часто наши реальные потребности заключаются совсем в другом: требуется лишь оценить суммарный эффект действия случайных факторов за большой промежуток времени или в большом числе повторений изучаемого процесса.

В качестве примера рассмотрим перенос песка водным потоком в реке или в том или ином искусственном гидротехническом сооружении. Обычно этот перенос происходит таким образом, что большинство песчинок спокойно лежит на дне и лишь изредка особенно сильные завихрения вблизи дна выхватывают отдельные песчинки и переносят их сразу на довольно значительные расстояния вплоть до внезапной остановки в каком-либо новом месте. Чисто теоретически движение каждой такой песчинки могло бы быть рассчитано индивидуально по законам гидромеханики. Но для этого нам было бы необходимо знать начальное состояние дна и потока во всех деталях и рассчитывать его шаг за шагом, отмечая те моменты времени, когда давление на какую-либо покоящуюся песчинку окажется достаточным, чтобы привести ее в движение, и прослеживая перемещение сдвинутых с места песчинок вплоть до их остановки. Абсурдность постановки такой задачи для реального научного исследования очевидна. И несмотря на это, средние или, как принято говорить, статистические закономерности передвижения донных наносов водными потоками вполне поддаются изучению.

Примеры, в которых действие большого числа случайных факторов приводит к вполне отчетливым статистическим закономерностям, было бы легко умножить. Один из самых увлекательных по широте перспектив и в то же время самых известных такого рода примеров доставляет кинетическая теория газов, которая показывает, как из совместного действия множества случайных столкновений молекул возникают точные закономерности, которым подчинены давление газа, как целого, на стенки, диффузное распространение одного газа в другом и т.д.

§ 6. Случайные процессы марковского типа

А. А. Маркову принадлежит заслуга построения вероятностной схемы непосредственно обобщающей детерминистическую схему, записанную в § 5 при помощи уравнения

$$\omega = F(t_0, \omega_0, t).$$

Правда, Марков рассмотрел только случай, когда фазовое пространство изучаемой системы состоит лишь из конечного числа состояний

$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, и изучал изменение состояния системы лишь при разделении времени t на дискретные шаги. Но в этой крайне схематизированной модели он сумел уловить ряд фундаментальных закономерностей.

Вместо функции F , однозначно определяющей состояние ω в момент времени $t > t_0$ по состоянию ω_0 в момент времени t_0 , Марков вводит вероятности

$$P(t_0, \omega_i; t, \omega_j)$$

получения состояния ω_j в момент времени t при условии, что в момент времени t_0 имело место состояние ω_i . Эти вероятности Марков связывает для любых трех моментов времени

$$t_0 < t_1 < t_2$$

соотношением, которое можно назвать основным уравнением марковских процессов

$$P(t_0, \omega_i; t_2, \omega_j) = \sum_{k=1}^n P(t_0, \omega_i; t_1, \omega_k) P(t_1, \omega_k; t_2, \omega_j). \quad (33)$$

Когда фазовое пространство является непрерывным многообразием, наиболее типичен случай существования *плотностей вероятности* $p(t_0, \omega_0; t, \omega)$ перехода из состояния ω_0 в состояние ω за промежуток времени (t_0, t) . В этом случае вероятность перехода за промежуток времени между моментами t_0 и t из состояния ω_0 в какое-либо из состояний ω , принадлежащих области G фазового пространства Ω , записывается в виде

$$P(t_0, \omega_0; t, G) = \int_G p(t_0, \omega_0; t, \omega) d\omega, \quad (34)$$

где $d\omega$ — элемент объема в фазовом пространстве¹⁾. Для плотностей вероятности $p(t_0, \omega_0; t, \omega)$ основное уравнение (33) приобретает вид

$$p(t_0, \omega_0; t_2, \omega_2) = \int_{\Omega} p(t_0, \omega_0; t_1, \omega) p(t_1, \omega; t_2, \omega_2) d\omega. \quad (35)$$

Уравнение (35) решить довольно трудно, но при известных ограничениях из него можно вывести дифференциальные уравнения в частных производных, которые легче поддаются изучению. Некоторые из этих уравнений были получены из нестрогих физических соображений физиками Фоккером и Планком. В полном виде теория так называемых

¹⁾ Собственно говоря, равенство (34) служит определением плотности вероятности. Величина $p d\omega$ равна (с точностью до бесконечно малых высшего порядка) вероятности перейти за время от t_0 до t из состояния ω_0 в элемент объема $d\omega$.

стохастических дифференциальных уравнений была построена советскими авторами (С. Н. Бернштейн, А. Н. Колмогоров, И. Г. Петровский, А. Я. Хинчин и др.).

Выписывать здесь эти уравнения мы не будем.

Метод стохастических дифференциальных уравнений позволяет, например, легко решить задачу о движении в спокойной атмосфере очень малого тела, для которого средняя скорость его падения s значительно меньше скорости его «броуновского движения», вызываемого тем, что из-за малости его размеров толчки молекул воздуха на противоположные его стороны не вполне уравниваются.

Пусть s — средняя скорость падения, D — так называемый коэффициент диффузии. Если предположить, что на земной поверхности ($z = 0$) частица не задерживается, а «отражается», т.е. под действием броуновских сил вновь отправляется в путешествие по атмосфере, и допустить, что в момент времени t_0 частица находится на высоте z_0 , то плотность вероятности $p(t_0, z_0; t, z)$ ее нахождения в момент времени t на высоте z выразится формулой

$$p(t_0, z_0; t, z) = \frac{1}{2\sqrt{\pi D(t-t_0)}} \left[e^{-(z-z_0)^2/[4D(t-t_0)]} + e^{-(z+z_0)^2/[4D(t-t_0)]} \right] \times e^{-c(z-z_0)/(2D)-c^2(t-t_0)/(4D)} + \frac{c}{D\sqrt{5}} e^{-cz/D} \int_{(z+z_0-c(t-t_0))/(2\sqrt{D(t-t_0)})}^{\infty} e^{-z^2} dz.$$

На рис. 4 изображено, как могут меняться кривые $p(t_0, z_0; t, z)$ в последовательные моменты времени t .

Рис. 4

Рис. 5

Мы видим, что в среднем высота частицы убывает, а положение ее становится все более неопределенным (все более «случайным»). Самое интересное заключается в том, что при любых t_0 и z_0 и при $t \rightarrow \infty$

$$p(t_0, z_0; t, z) \rightarrow \frac{c}{D} e^{-cz/D}, \quad (36)$$

т.е. существует предельное распределение для высот частицы, и математическое ожидание высоты расположения частицы с возрастанием t стремится к положительному пределу

$$z^* = \frac{c}{D} \int_0^\infty z e^{-cz/D} dz = \frac{D}{c}. \quad (37)$$

Таким образом, несмотря на то, что наша частица, находясь над поверхностью земли, все время имеет тенденцию падать под действием силы тяжести, она при неограниченном продолжении этого процесса (блуждания в атмосфере) будет в среднем находиться на определенной положительной высоте. Если бы мы взяли начальное z_0 меньшим, чем z^* , то оказалось бы, что по истечении достаточно большого промежутка времени среднее положение частицы будет выше начального, как это изображено на рис. 5, где $z = 0$.

Для отдельной частицы средние значения z^* , о которых идет речь, являются лишь математическими ожиданиями, но в силу закона больших чисел для большого числа частиц они будут осуществляться реально: плотность расположения такого рода частиц по высоте будет следовать указанным закономерностям и, в частности, по истечении достаточно большого промежутка времени стабилизируется в соответствии с формулой (36).

Все сказанное непосредственно применимо лишь к примешанным к воздуху в малой концентрации газам, дымам и т.п., так как величины c и D предполагались определенными заранее заданным состоянием атмосферы. Однако с некоторыми усложнениями теория применима к взаимной диффузии газов, составляющих атмосферу, и к возникающим на основе этой диффузии распределениям их плотностей по высоте.

С увеличением размеров частиц отношение c/D возрастает, и благодаря этому процесс их перемещения вместо диффузионного характера приобретает характер закономерного падения по законам, рассмотренным в § 5. Теория позволяет проследить все переходы между чисто диффузным движением и таким закономерным падением.

Задача движения взвешенных в атмосфере частиц под действием турбулентного перемешивания более сложна, но в принципе тоже может быть подчинена аналогичным вероятностным методам.

ЛИТЕРАТУРА

Популярная литература

Гнеденко Б. В., Хинчин А. Я. Элементарное введение в теорию вероятностей. Изд. 3-е, Гостехиздат, 1952.

Современные университетские курсы

Бернштейн С. Н. Теория вероятностей. Изд. 4-е, Гостехиздат, 1946.

Гнеденко Б. В. Курс теории вероятностей. Изд. 2-е, Гостехиздат, 1954.

Специальные монографии

Чандрасекар С. Перенос лучистой энергии. М., 1953.

Эйнштейн А., Смолуховский М. Броуновское движение. Сборник статей, 1936.

Отзыв о статье «Теория вероятностей»

Статью «Теория вероятностей» следует признать вполне удачной во всех отношениях. Написанная на очень высоком научном уровне, без всяких скидок на неподготовленность читателя, она в то же время доступна пониманию широкого круга лиц, интересующихся математикой. Самое ценное в этой статье, однако, представляет собой никогда еще в такой мере не достигнутая в принципиальных освещениях теории вероятностей связь общей теории с практическими применениями; автор умеет показать, как наиболее абстрактные разделы теоретического построения неразрывно связаны с конкретными запросами практики. Поэтому в его изложении теория вероятностей предстает перед читателем одновременно как стройная и законченная логическая конструкция и как мощный аппарат естествознания и техники, причем эти две картины так тесно проникают друг в друга, что образуют собою органическое, неразрывное единство.

10.X.1951 г.

Член-корр. АН СССР А. Я. Хинчин

Из письма А. Д. Александрова

Глубокоуважаемый Андрей Николаевич!

Так как на меня возложена ответственность за завершение работы над монографией «Математика, ее содержание, методы и значение», обращаюсь к Вам по поводу Вашей статьи.

В настоящем виде статья трудна, и с ее появлением может даже создаться представление, будто основы теории вероятностей и нельзя понять «простому смертному».

Я прочел Вашу статью «Вероятность» в БСЭ. Она произвела на меня очень большое впечатление. Я, конечно, не специалист и занимался вероятностью лишь в связи с физикой, но решаюсь высказать свое мнение. Ничего подобного по простоте и вместе с тем глубине изложения я не видел. В своих занятиях основами квантовой механики я подошел к тем же взглядам, но не смог сформулировать их с такой глубиной и четкостью. Ваша формулировка о связи явления с условиями имеет очень большое значение для квантовой механики, для преодоления чисто статистического ее толкования, отвечающего мизесовскому подходу к вероятности (кстати, В. А. Фок специально сказал мне, что считает Ваше толкование вероятности крайне важным).

Для того чтобы Ваши глубокие и важные, а вместе с тем простые идеи стали достоянием возможно более широких кругов читателей, я просил бы Вас перестроить Вашу статью, с тем чтобы, начав с указания на объективный характер статистических закономерностей, дав краткую критику субъективизма, сразу перейти к указанию на связь явления с условиями и развернуть Ваше понимание вероятности, чтобы оно не затерялось за детерминистической схемой или какими-либо выкладками. В соединении глубины с простотой изложения, Ваша статья станет лучшим украшением нашей монографии. Очень прошу Вас именно упростить изложение, чтобы сделать Ваши идеи максимально доступными. Я позволил бы себе посоветовать Вам испытать потом понятность изложения на каком-нибудь среднем читателе, имеющем смутное представление о предмете.

Еще одно замечание. Когда Вы говорите о связи теории вероятностей с теорией меры, естественно оговорить, что различие не сводится к тому, что $P(U) = 1$, а включает своеобразие вероятностных постановок задач, что выражается, скажем, в понятии независимых событий, по-видимому, чуждом учению о площадях и объемах в обычном понимании.

15.II.1952

С глубоким уважением
А. Д. Александров

Из письма А. Н. Колмогорова

Глубокоуважаемый Александр Данилович!

Благодарю Вас за Ваше письмо по поводу моей статьи для монографии о математике. Из Ваших пожеланий я заведомо могу

1) пояснить на примерах в момент их введения понятия случайной величины, математического ожидания, дисперсии и на возможно более простом примере — марковскую схему;

2) более полно осветить специфичность проблем, когда говорится о возможности формального включения теории вероятностей в теорию меры (я говорил об этом в моих «Основных понятиях теории вероятностей» — в популярной статье, тем более, следовало сказать нечто аналогичное, и я просто позабыл это сделать).

Основная трудность при освещении философских вопросов теории вероятностей начинается с ясного изложения и диалектического объединения двух положений:

- 1) существует объективная случайность,
- 2) не существует ничего абсолютно случайного.

Я понимаю, что с точки зрения отстаивания права на существование теории вероятностей (а, как Вы знаете, эта тема не совсем беспредметна) наиболее существенно первое. Но мне представляется, что в сколь угодно популярной статье необходимо говорить об обеих сторонах дела.

Во всяком случае, я учту то обстоятельство, что, по Вашему мнению, статья не освещает достаточно выпукло именно объективный характер вероятностных высказываний.

Раз уже Вы поинтересовались моей статьей «Вероятность» в БСЭ, мне интересно знать еще Ваше суждение о статье «Больших чисел закон», где мне принадлежит только общая (не соци(ально)-экономическая) часть. Вопрос этот имеет некоторую актуальность, так как со стороны многих наших экономистов поступили резкие возражения и против моей общей части. Я же пока единственно в чем готов признать некоторую неудачу, это в том, что Б.ч.з. назван «принципом», чего, м(ожет) б(ыть), и не следовало делать.

26.II.1952 г.

Искренне Вас уважающий
А. Колмогоров