

# МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ЯВЛЕНИЙ И ПРОЦЕССОВ: принципы построения вероятностных моделей

Межфакультетский курс

Лектор: В. Ю. Королев (факультет ВМК)

## Часть II.

# Испытания Бернулли. 1

Рассмотрим последовательность однотипных случайных экспериментов (испытаний), в каждом из которых возможны лишь два взаимоисключающих исхода. Условимся называть эти исходы “успех” и “неудача”. Пусть вероятность успеха одинакова для каждого испытания и равна  $p$ ,  $0 \leq p \leq 1$ .

С этими испытаниями свяжем случайные величины  $X_1, \dots, X_n$  по следующему правилу. Если в  $k$ -ом испытании зафиксирован успех, то  $X_k = 1$ , а если в  $k$ -ом испытании зафиксирована неудача, то  $X_k = 0$ ,  $k = 1, \dots, n$ . При этом, очевидно, все случайные величины  $X_1, \dots, X_n$  имеют одинаковое распределение:

$$P(X_k = 1) = p = 1 - P(X_k = 0), \quad k = 1, \dots, n. \quad (1)$$

Предположим, что рассматриваемые дихотомические испытания независимы в том смысле, что случайные величины  $X_1, \dots, X_n$  независимы в совокупности. Испытания, удовлетворяющие указанным условиям, называются *испытаниями Бернулли*.

## Испытания Бернулли. 2



Jacob Bernoulli (1655-1705)

# Испытания Бернулли. Биномиальное распределение. 1

Рассмотрим стохастическую ситуацию, заключающуюся в осуществлении  $n$  испытаний Бернулли ( $n \geq 1$ ). Построим соответствующую вероятностную модель. С этой целью заметим, что каждый элементарный исход *последовательности*  $n$  испытаний Бернулли представляет собой *цепочку* вида  $\omega = (\omega_1, \dots, \omega_n)$ , где  $\omega_k$  равно либо 1 либо 0 в зависимости от того, какое значение случайной величины  $X_k$  наблюдается в  $k$ -ом испытании, то есть в зависимости от того, закончилось ли  $k$ -е испытание успехом или неудачей.

Несложно убедиться, что всего имеется  $N = 2^n$  разных цепочек вида  $\omega = (\omega_1, \dots, \omega_n)$ , в которых  $\omega_k$  равно либо 1 либо 0. Отсюда мы соответственно заключаем, что совокупность всех элементарных исходов рассматриваемой стохастической ситуации состоит из  $N = 2^n$  элементов указанного выше вида.

Так как испытания независимы, то вероятность каждой элементарной цепочки  $\omega = (\omega_1, \dots, \omega_n)$  оказывается равной произведению вероятностей соответствующих исходов дихотомических испытаний, составляющих эту цепочку:

$$P(\{\omega\}) = \prod_{k=1}^n p^{\omega_k} (1-p)^{1-\omega_k} = p^{\sum_{k=1}^n \omega_k} (1-p)^{n-\sum_{k=1}^n \omega_k} \quad (2)$$

(число единиц в цепочке  $\omega = (\omega_1, \dots, \omega_n)$  равно  $\sum_{k=1}^n \omega_k$  и, соответственно, число нулей в такой цепочке равно  $n - \sum_{k=1}^n \omega_k$ ).

# Испытания Бернулли. Биномиальное распределение. 3

Пусть  $X$  – случайная величина, равная числу успехов в  $n$  испытаниях Бернулли. Найдем ее распределение.

Очевидно, что возможными значениями случайной величины  $X$  являются числа  $0, 1, \dots, n$ . Несложно видеть, что для каждого элементарного исхода  $\omega = (\omega_1, \dots, \omega_n)$  случайная величина  $X$  равна

$$X(\omega) = \sum_{k=1}^n \omega_k. \quad (3)$$

Пусть  $m$  – целое число,  $0 \leq m \leq n$ . Существует  $C_n^m$  разных цепочек  $\omega = (\omega_1, \dots, \omega_n)$ , удовлетворяющих условию

$$\sum_{k=1}^n \omega_k = m. \quad (4)$$

С учетом соотношения (2) мы замечаем, что каждая из цепочек  $\omega = (\omega_1, \dots, \omega_n)$ , удовлетворяющих условию (4), имеет вероятность  $P(\{\omega\}) = p^m(1-p)^{n-m}$ .

# Испытания Бернулли. Биномиальное распределение. 4

Событие  $\{\omega : X(\omega) = m\}$  состоит из всех цепочек, удовлетворяющих условию (4). Поэтому

$$p_m = P(X = m) = \sum_{\substack{\omega=(\omega_1, \dots, \omega_n): \\ \omega_1 + \dots + \omega_n = m}} P(\{\omega\}) = C_n^m p^m (1-p)^{n-m}, \quad m = 0, \dots, n.$$

Используя формулу бинома Ньютона несложно убедиться, что

$$p_0 + p_1 + \dots + p_n = 1.$$

Набор пар  $\{(m, p_m), m = 0, \dots, n\}$  называется *биномиальным распределением вероятностей*.

# Испытания Бернулли. Биномиальное распределение. 5

Случайную величину  $X$ , равную числу успехов в  $n$  испытаниях Бернулли, можно представить в виде

$$X = X_1 + \dots + X_n, \quad (5)$$

где  $X_1, \dots, X_n$  – введенные в самом начале этого раздела независимые случайные величины. На основании (1) мы заключаем, что

$$EX_k = 1 \cdot p + 0 \cdot (1 - p) = p,$$

$$DX_k = (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) = p(1 - p), \quad k = 1, \dots, n.$$

Поэтому в силу справедливости представления (5) мы заключаем, что если случайная величина  $X$  имеет биномиальное распределение, то

$$EX = EX_1 + \dots + EX_n = n \cdot EX_1 = np,$$

$$DX = DX_1 + \dots + DX_n = n \cdot DX_1 = np(1 - p).$$



# Испытания Бернулли. Биномиальное распределение. 7

Вычисление биномиальных вероятностей  $C_n^k p^k (1-p)^{n-k}$  при больших  $n$  представляет собой очень сложную задачу. Однако оказывается возможным использовать приближенные формулы, которые являются следствиями предельных теорем, о которых пойдет речь ниже.

Пусть  $X$  – случайная величина, имеющая биномиальное распределение с параметрами  $n$  и  $p$ , то есть

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Можно показать, что если  $k - 1 \leq n/4$  и  $p \leq 1/4$ , то

$$P(X = k) = \frac{(np)^k}{k!} \exp\{-np + r_n(k)\}, \quad (6)$$

где

$$kp + 2 \log \frac{4}{3} \cdot \frac{k(1-k)}{6n} - \frac{2(np)^2}{3n} \leq r_n(k) \leq \frac{k(1-k) + 2knp}{2n}.$$

# Испытания Бернулли. Биномиальное распределение. 8

Из соотношения (6) вытекает возможность использования для указанных выше значений параметров приближенной формулы

$$P(X = k) \approx e^{-np} \frac{(np)^k}{k!}, \quad (7)$$

которая имеет хорошую точность даже при умеренно больших  $n$ . Из соотношения (8) также вытекает, что бóльшую точность дает приближенная формула

$$P(X = k) \approx \frac{(np)^k}{k!} \exp \left\{ -np + \alpha \frac{k(1-k)}{n} + kp \right\}, \quad (8)$$

где  $0.5 \leq \alpha \leq 0.5754$ . Однако соотношение (6) показывает, что если  $n \rightarrow \infty$ , но при этом значение  $p$  остается постоянным, то точность приведенных выше аппроксимаций становится неудовлетворительной.

# Испытания Бернулли. Теорема Пуассона. 1

Для того, чтобы при  $n \rightarrow \infty$  формулы (7) и (8) давали приемлемую точность, нужно, чтобы одновременно с неограниченным увеличением  $n$  параметр  $p$  стремился к нулю.

**Теорема.** Пусть  $n \rightarrow \infty$  и одновременно  $p \rightarrow 0$  так, что  $np \rightarrow \lambda$  для некоторого  $\lambda > 0$ . Тогда для каждого  $k = 0, 1, 2, \dots$

$$P(X = k) = C_n^k p^k (1 - p)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}. \quad (9)$$

Это утверждение называется *теоремой Пуассона, теоремой о редких событиях или законом малых чисел*.



Рис.: Siméon Denis Poisson (1781-1840)

Условия теоремы Пуассона можно интерпретировать следующим образом. Если число испытаний Бернулли велико, а вероятность успеха в одном отдельном испытании мала и при этом ожидаемое число успехов умеренно, то биномиальные вероятности можно вычислять по приближенной формуле (7):

$$C_n^k p^k (1-p)^{n-k} \approx e^{-np} \frac{(np)^k}{k!}.$$

## Теорема Пуассона. 3

Просуммировав числа в правой части соотношения (7) по всем  $k$ , получим

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1,$$

то есть набор пар  $\{(k, e^{-\lambda} \lambda^k / k!), k = 0, 1, \dots\}$  представляет собой распределение вероятностей. Это распределение называется *распределением Пуассона*.

Если случайная величина  $X$  имеет распределение Пуассона, то

$$EX = DX = \lambda.$$

## Теорема Пуассона. 4

В предположении, что  $C_n^k h^k (1-p)^{n-k} = 0$  при  $k > n$ , справедлива следующая оценка скорости сходимости в теореме Пуассона:

$$\sum_{k=0}^{\infty} \left| C_n^k h^k (1-p)^{n-k} - e^{-np} \frac{(np)^k}{k!} \right| \leq 2p \min\{1, np\}$$

(*Barbour A. D., Hall P. On the rate of Poisson convergence // Mathematical Proceedings of the Cambridge Philosophical Society, 1984. Vol. 95. P. 473–480*).

# Испытания Бернулли. Теорема Муавра–Лапласа. 1

Если нет оснований считать вероятность успеха в отдельном испытании Бернулли малой, то можно воспользоваться следующим утверждением.

**Теорема.** Пусть  $X$  – число успехов в  $n$  испытаниях Бернулли. Если число испытаний неограниченно увеличивается, а вероятность успеха в одном испытании остается той же самой, то, какими бы ни были числа  $a$  и  $b$ ,  $a \leq b$ ,

$$P\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Это утверждение называется *теоремой Муавра–Лапласа*.

## Испытания Бернулли. Теорема Муавра–Лапласа. 2



Abraham de Moivre  
(1667-1756)



Pierre-Simon de Laplace  
(1749-1827)



# Теорема Муавра–Лапласа. Нормальное распределение. 1

Можно убедиться, что

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1,$$

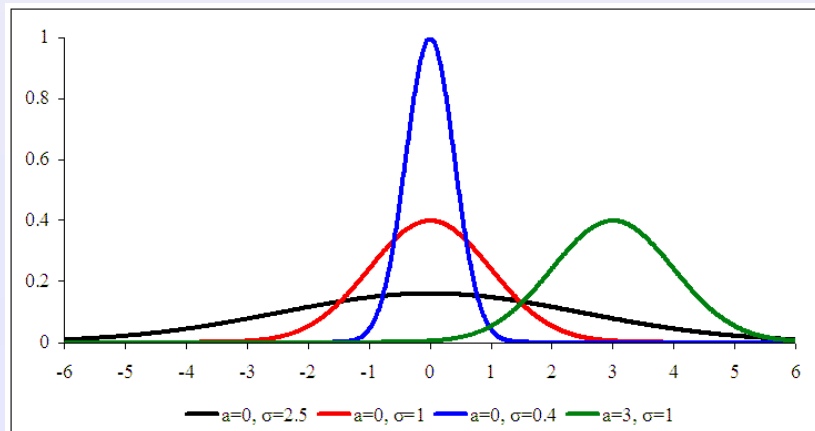
то есть функция  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$  представляет собой плотность распределения. Распределение вероятностей, соответствующее плотности  $\varphi(x)$ , называется *стандартным нормальным*. Соответствующую функцию распределения обозначают  $\Phi(x)$  и называют *стандартной нормальной*. Значения функции  $\Phi(x)$  можно найти в специальных таблицах.

Если случайная величина  $X$  имеет стандартное нормальное распределение, то  $EX = 0$ ,  $DX = 1$ . Можно показать, что если случайная величина  $X$  имеет стандартное нормальное распределение, то случайная величина  $Y = \sigma X + a$  имеет плотность

$$\varphi(x; a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

и функцию распределения  $\Phi\left(\frac{x-a}{\sigma}\right)$ . В таком случае  $EY = a$ ,  $DY = \sigma^2$  и говорят, что случайная величина  $Y$  имеет *нормальное распределение с параметрами  $a$  и  $\sigma^2$* .

## Теорема Муавра–Лапласа. Нормальное распределение. 2



Графики плотности нормального распределения при разных  $a$  и  $\sigma$

Приведем пример применения теоремы Муавра–Лапласа, поясняющий некоторые его свойства.

**Задача.** Общеизвестны книги американского врача Бенджамена Спока о воспитании детей. Доктор Спок известен не только своими книгами, но и своими антивоенными убеждениями. Он неоднократно участвовал в антивоенных акциях, в том числе во время войны, которую США вели во Вьетнаме, и был судим по обвинению в антиконституционной деятельности. Однажды он был арестован в Бостоне за участие в очередной антивоенной демонстрации. Его дело должен был рассматривать суд присяжных. Присяжные отбираются из дееспособного населения Бостона с помощью многоступенчатой процедуры, на очередном этапе которой было отобрано 300 человек (из числа которых потом и должны были быть выбраны 12 присяжных). Однако среди этих 300 человек оказалось лишь 90 женщин. Это обстоятельство послужило причиной того, что адвокаты доктора Спока подали протест (воспитанием детей в США, как и в России, в основном занимаются женщины и поэтому ясно, что в рассматриваемой ситуации при прочих равных женщины с большим сочувствием отнесутся к подсудимому). Насколько обоснован протест?

**Решение.** Доля женщин в дееспособном населении Бостона составляет примерно 50%. Поэтому если вероятность случайного выбора женщины в число 300 кандидатов обозначить  $p$ , то при абсолютно беспристрастной процедуре выбора должно быть  $p = \frac{1}{2}$ . Проверим, так ли это на самом деле. Для этого, предполагая, что  $p = \frac{1}{2}$ , найдем вероятность того, что будет отобрано 90 или еще меньше женщин. Другими словами, найдем вероятность того, что наблюдаемое отклонение числа отобранных женщин от ожидаемого их числа 150 обусловлено чистой случайностью. Если эта вероятность окажется исчезающе малой, то мы придем к выводу о том, что осуществилось событие, которое в сделанном предположении произойти практически не может. Это будет означать, что сделанное предположение не верно. (Такая логика рассуждений характерна для проверки *статистических гипотез*.)

# Теорема Муавра–Лапласа. Нормальное распределение. 5

Итак, пусть  $X$  – количество отобранных женщин. Мы можем считать  $X$  числом успехов в 300 испытаниях Бернулли с вероятностью успеха в одном испытании, равной  $p$ . В предположении полной случайности выбора ( $p = \frac{1}{2}$ ) искомая вероятность имеет вид

$$P(X \leq 90) = \sum_{k=0}^{90} C_{300}^k p^k (1-p)^{n-k} = \frac{1}{2^{300}} \sum_{k=0}^{90} C_{300}^k.$$

Вычисление этой вероятности затруднено. Чтобы приблизительно оценить эту вероятность, применим теорему Муавра–Лапласа и получим

$$P(X \leq 90) = P\left(\frac{X - 150}{\sqrt{75}} \leq \frac{90 - 150}{\sqrt{75}}\right) \approx \Phi\left(\frac{90 - 150}{\sqrt{75}}\right) = \Phi(-6.9282).$$

# Теорема Муавра–Лапласа. Нормальное распределение. 6

Чтобы оценить, насколько мала эта вероятность, зафиксируем произвольное число  $x > 0$  и рассмотрим значение  $\Phi(-x)$ . Так как плотность стандартного нормального распределения  $\varphi(x) \equiv \varphi(x; 0, 1)$  четна, то  $\Phi(-x) = 1 - \Phi(x)$ . Поэтому

$$\Phi(-x) = 1 - \Phi(x) = \int_x^\infty \varphi(y) dy = \int_x^\infty \frac{y}{y} \cdot \varphi(y) dy \leq \frac{1}{x} \int_x^\infty y \cdot \varphi(y) dy.$$

Но  $y dy = d(y^2/2)$ . Поэтому

$$\Phi(-x) \leq \frac{1}{\sqrt{2\pi}x} \int_x^\infty e^{-y} dy = \frac{\varphi(x)}{x}.$$

Следовательно, при  $|x| \rightarrow \infty$  “хвосты” нормального распределение убывают быстрее, чем функция  $\varphi(x)/|x| = e^{-x^2/2}/(\sqrt{2\pi}|x|)$ .

Можно показать, что для  $x > 0$

$$\frac{1}{x} - \frac{1}{x^3} \leq \frac{1 - \Phi(x)}{\varphi(x)} \leq \frac{1}{x}.$$

Таким образом,  $\Phi(-6.9282) < 0.06e^{-24} < 6 \cdot 10^{-12}$ . То есть вероятность того, что среди 300 кандидатов окажется не более 90 женщин, вычисленная в предположении беспристрастности выбора, исчезающе мала. Следовательно, это предположение, не верно. При этом вероятность того, что этот вывод ошибочен, тоже меньше  $6 \cdot 10^{-12}$ .

Теорема Муавра–Лапласа является частным случаем *центральной предельной теоремы*, которая будет подробно обсуждаться ниже.

# Испытания Бернулли. Геометрическое распределение. 1

Рассмотрим стохастическую ситуацию, состоящую в проведении испытаний Бернулли до первой неудачи. В такой ситуации совокупность элементарных исходов состоит из цепочек  $\omega = (\omega_1, \dots, \omega_n)$ , в которых  $\omega_1 = \dots = \omega_{n-1} = 1$ , а  $\omega_n = 0$ ,  $n = 1, 2, \dots$ . Таких цепочек бесконечно много. Так как испытания независимы, то цепочка  $\omega = (\omega_1, \dots, \omega_n)$  указанного вида имеет вероятность

$$p_n = P(\{\omega\}) = p^{n-1}(1-p). \quad (10)$$

По формуле суммы членов бесконечно убывающей геометрической прогрессии можно убедиться, что

$$\sum_{n=1}^{\infty} p_n = (1-p) \sum_{n=1}^{\infty} p^{n-1} = 1.$$

Набор пар  $\{(n, p_n), n = 1, 2, \dots\}$ , где  $p_n$  имеет вид (10), называется *геометрическим* распределением вероятностей.



## Испытания Бернулли. Геометрическое распределение. 2

С рассматриваемой стохастической ситуацией свяжем случайную величину  $X$ , определив ее для каждого элементарного исхода  $\omega = (\omega_1, \dots, \omega_n)$  указанного выше вида как  $X(\omega) = n$ . Случайная величина  $X$  принимает значения  $1, 2, \dots$  и имеет смысл количества испытаний Бернулли, осуществленных до первой неудачи. В таком случае говорят, что случайная величина  $X$  имеет геометрическое распределение.

Можно показать, что если случайная величина  $X$  имеет геометрическое распределение, то

$$EX = \frac{1}{1-p}, \quad DX = \frac{1}{(1-p)^2}.$$

# Теорема Реньи. Показательное распределение

Пусть  $X$  – случайная величина с геометрическим распределением

$$P(X = n) = p^{n-1}(1 - p), \quad k = 1, 2, \dots,$$

$0 < p < 1$ . В этом случае  $EX = 1/(1 - p)$ . Поэтому, если  $p \rightarrow 1$ , то  $EX \rightarrow \infty$ . Однако если одновременно со стремлением величины  $q = 1 - p$  к нулю производить нормировку рассматриваемой случайной величины, деля ее на ее математическое ожидание, то можно убедиться, что, каково бы ни было  $x > 0$ ,

$$P\left(\frac{X}{EX} < x\right) = P(qX < x) = P((1 - p)X < x) \rightarrow 1 - e^{-x}.$$

Это утверждение является частным случаем так называемой *теоремы Реньи*. Легко видеть, что предельная функция является функцией распределения. Такое распределение называется *стандартным показательным*. В общем случае, вводя параметр масштаба  $\lambda > 0$ , можно определить *показательное распределение с параметром  $\lambda$*  как задаваемое соответствующей функцией распределения  $F(x)$ , равной 0 при  $x < 0$  и равной  $1 - e^{-\lambda x}$  при  $x \geq 0$ .

# Показательное распределение

Если  $Y$  – случайная величина с показательным распределением с параметром  $\lambda$ , то

$$EY = \frac{1}{\lambda}, \quad DY = \frac{1}{\lambda^2}.$$

Показательное распределение часто используется в качестве модели распределения продолжительности жизни или длительности безотказной работы. Однако такие модели весьма грубы и, вообще говоря, являются идеальными, поскольку не учитывают эффект старения. Действительно, в предположении о том, что время  $X$  безотказной работы некоторого устройства имеет показательное распределение с некоторым параметром  $\lambda$ , рассмотрим условную вероятность того, что после того как устройство проработало в течение времени  $t$ , устройство проработает еще как минимум время  $\tau$ . Эта вероятность имеет вид

$$\begin{aligned} P(X > t + \tau | X > t) &= \frac{P(X > t + \tau; X > t)}{P(X > t)} = \frac{P(X > t + \tau)}{P(X > t)} = \\ &= \frac{e^{-(t+\tau)}}{e^{-t}} = e^{-\tau} = P(X > \tau), \end{aligned}$$

то есть эта условная вероятность не зависит от  $t$  и совпадает с безусловной вероятностью того, что устройство проработает как минимум время  $\tau$ .

# Испытания Бернулли. Заключение

Итак, мы рассмотрели удобную и простую схему испытаний Бернулли, которую по сути можно описать в терминах случайного процесса с дискретным временем, который в момент времени  $t = n \in \mathbb{N}$  принимает одно из двух значений 0 или 1 независимо от того, какие значения он принял в другие моменты времени.

С этой схемой оказались тесно связанными следующие распределения вероятностей:

- биномиальное,
- геометрическое,
- Пуассона,
- показательное
- нормальное.

Последовательность испытаний Бернулли по сути является аналогом *пуассоновского процесса* для случая дискретного времени.

# Основные предельные теоремы теории вероятностей

При реализации *асимптотического подхода* к построению вероятностных моделей статистических закономерностей, присущих большим массивам данных, в качестве модели принимается асимптотическая аппроксимация – предельное распределение в некоторой предельной теореме теории вероятностей, условия которой можно считать выполненными в рассматриваемой ситуации.

При этом главную роль играет *правильный* (адекватный) выбор структуры наблюдаемой случайной величины, например, аддитивной, мультипликативной или экстремальной. То есть при моделировании первый шаг заключается в осознании того, можно ли считать наблюдаемую случайную величину

- суммой,
- произведением,
- максимумом или минимумом

каких-либо других “элементарных” или “атомических” случайных величин.

Например, во многих ситуациях можно считать, что наблюдаемое значение обусловлено *суммарным* воздействием случайных величин. В частности, приращение случайного процесса на некотором интервале времени может быть представлено в виде суммы его приращений на подынтервалах, на которые делится исходный интервал.

Мультипликативная структура данных возникает, когда наблюдения являются не абсолютными, но относительными приращениями. С помощью перехода к логарифмам мультипликативная схема может быть сведена к аддитивной.

При исследовании данных типа времени жизни, например, в теории надежности, могут возникнуть ситуации, когда агрегат выходит из строя при отказе хотя бы одной из его подсистем. Тогда время его безотказной работы – это наименьшее из времен безотказной работы его подсистем.

В данном курсе мы сосредоточимся на предельных теоремах для вероятностных моделей данных с аддитивной структурой.

Предельные теоремы для сумм случайных величин занимают центральное место в классической теории вероятностей.

Примерами таких теорем являются

- законы больших чисел,
- теорема Пуассона,
- центральная предельная теорема.

Для их правильного применения и, соответственно, получения адекватных результатов необходимо четко понимать, при каких условиях справедливы их утверждения. Описанием этих условий мы и займемся.

# Законы больших чисел. 1

Пусть  $X$  – число успехов в  $n$  испытаниях Бернулли с вероятностью успеха в отдельном испытании, равной  $p$ . Случайная величина  $X$  имеет биномиальное распределение с параметрами  $n$  и  $p$ . Тогда под *частотой успеха* следует понимать дробь  $X/n$ . Еще очень давно было замечено, что при увеличении  $n$  частота успеха сближается с вероятностью  $p$ .

Формально это сближение можно установить, например, с помощью неравенства Чебышева. Зафиксируем какое-либо произвольно малое положительное число  $\varepsilon$  и рассмотрим вероятность

$$P\left(\left|\frac{X}{n} - p\right| > \varepsilon\right)$$

того, что отклонение частоты успеха от вероятности успеха превзойдет  $\varepsilon$  по абсолютной величине.



## Законы больших чисел. 2

Так как  $EX = np$  и  $DX = np(1 - p)$ , то

$$E\frac{X}{n} = p \quad \text{и} \quad D\frac{X}{n} = \frac{DX}{n^2} = \frac{p(1-p)}{n}.$$

Тогда по неравенству Чебышева

$$P\left(\left|\frac{X}{n} - p\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} D\left(\frac{X}{n}\right) = \frac{p(1-p)}{\varepsilon^2 n} \rightarrow 0$$

при  $n \rightarrow \infty$ .

Это утверждение называется *законом больших чисел Бернулли* и является самым простым видом закона больших чисел, который устанавливает сближение среднего арифметического случайных величин со средним арифметическим их математических ожиданий.

# Законы больших чисел. 3

Пусть теперь  $X_1, X_2, \dots$  – независимые не обязательно одинаково распределенные случайные величины. Обозначим

$$EX_j = a_j, \quad DX_j = \sigma_j^2, \quad j = 1, 2, \dots$$

Тогда по неравенству Чебышева для любого  $\epsilon > 0$  имеем

$$P\left(\left|\frac{1}{n} \sum_{j=1}^n X_j - \frac{1}{n} \sum_{j=1}^n a_j\right| > \epsilon\right) \rightarrow 0,$$

если

$$\frac{1}{n^2} \sum_{j=1}^n \sigma_j^2 \rightarrow 0.$$

Это утверждение называется *законом больших чисел Чебышева*.

## Законы больших чисел. 4

В формулировке закона больших чисел Чебышева в *утверждение* входят лишь математические ожидания случайных величин, тогда как *условие* сформулировано в терминах их дисперсий, существование которых является более сильным условием, нежели существование математических ожиданий.

В случае одинаково распределенных слагаемых условие существования дисперсий оказывается излишним. А именно, справедлив *усиленный закон больших чисел Колмогорова*:

**Теорема.** Пусть  $X_1, X_2, \dots$  – независимые одинаково распределенные случайные величины.

I. Если существует  $EX_1 \equiv a$ , то с вероятностью единица

$$\frac{1}{n} \sum_{j=1}^n X_j \longrightarrow a \quad (n \rightarrow \infty). \quad (11)$$

II. Если при некотором  $a \in \mathbb{R}$  с вероятностью единица имеет место сходимость (11), то существует  $EX_1 \equiv a$ .

Условие существования моментов определенного порядка играет критическую роль для справедливости тех или иных предельных теорем теории вероятностей. Для их адекватного применения на практике необходимо быть уверенным в его справедливости.

Усиленный закон больших чисел Колмогорова может быть применен для *практической* проверки конечности моментов.

Пусть, к примеру, нас интересует, конечен ли момент порядка  $\alpha$  случайной величины  $X$ , реализации  $X_1, X_2, \dots$  которой доступны наблюдению. Строим график функции

$$f(n) = \frac{1}{n} \sum_{j=1}^n X_j^\alpha, \quad n \in \mathbb{N},$$

натурального аргумента  $n$ . Если при увеличении  $n$  наблюдается стабилизация этой функции на некотором уровне, то можно сделать вывод о том, что момент  $EX^\alpha$  конечен.

# Суммы случайных индикаторов. Теорема Пуассона. 1

Рассмотрим *схему серий* случайных величин

$$\begin{array}{l} X_{1,1} \\ X_{2,1}, X_{2,2} \\ \dots\dots\dots \\ X_{n,1}, X_{n,2}, \dots, X_{n,n} \\ \dots\dots\dots \end{array}$$

Первый индекс указывает на номер серии. Предположим, что в каждой серии случайные величины  $X_{n,1}, X_{n,2}, \dots, X_{n,n}$ ,  $n = 1, 2, \dots$ , независимы и

$$P(X_{n,i} = 1) = p_n = 1 - q_n.$$

Из теоремы Пуассона вытекает следующий результат. Положим

$$S_n = X_{n,1} + \dots + X_{n,n}.$$

## Суммы случайных индикаторов. Теорема Пуассона. 2

**Теорема.** Если  $np_n \rightarrow \lambda > 0$  при  $n \rightarrow \infty$ , то при каждом фиксированном  $k = 0, 1, 2, \dots$

$$\lim_{n \rightarrow \infty} P(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Естественным обобщением рассмотренной ситуации является та, в которой случайные величины  $X_{n,j}$  могут иметь разное распределение даже в рамках одной и той же серии. Предположим теперь, что

$$P(X_{n,j} = 1) = p_{n,j}, \quad j = 1, \dots, n.$$

Подобную схему серий неодинаково распределенных независимых индикаторов принято называть *схемой Пуассона*. Распределение случайной величины  $S_n$  в такой схеме иногда называют *пуассон-биномиальным*.

# Суммы случайных индикаторов. Теорема Пуассона. 3

**Теорема.** Если

$$\max_{1 \leq j \leq n} p_{n,j} \rightarrow 0, \quad \sum_{j=1}^n p_{n,j} \rightarrow \lambda \in (0, \infty)$$

при  $n \rightarrow \infty$ , то при каждом фиксированном  $k = 0, 1, 2, \dots$

$$\lim_{n \rightarrow \infty} P(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

**Доказательство.** Докажем теорему методом производящих функций. Символ  $f_X(s)$  будет обозначать производящую функцию случайной величины  $X$  в точке  $s$ :

$$f_X(s) = Es^X, \quad |s| \leq 1.$$

Тогда, очевидно,

$$f_{X_{n,j}}(s) = Es^{X_{n,j}} = 1 + p_{n,j}(s - 1).$$

Поскольку случайные величины  $X_{n,1}, \dots, X_{n,n}$  независимы, мы имеем

$$f_{S_n}(s) = Es^{S_n} = \prod_{j=1}^n f_{X_{n,j}}(s) = \prod_{j=1}^n (1 + p_{n,j}(s - 1)).$$

# Суммы случайных индикаторов. Теорема Пуассона. 4

Далее, так как

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n p_{n,j}^2 \leq \lim_{n \rightarrow \infty} \max_{1 \leq j \leq n} p_{n,j} \cdot \lim_{n \rightarrow \infty} \sum_{j=1}^n p_{n,j} = 0$$

и при любом фиксированном  $s$ ,  $|s| \leq 1$ ,

$$\log f_{S_n}(s) = \sum_{j=1}^n \log(1 + p_{n,j}(s-1)) \sim \sum_{j=1}^n p_{n,j}(s-1) \sim \lambda(s-1) \quad (n \rightarrow \infty),$$

то при любом фиксированном  $s$ ,  $|s| \leq 1$ ,

$$f_{S_n}(s) \longrightarrow e^{\lambda(s-1)}.$$

Но последнее выражение представляет собой производящую функцию распределения Пуассона с параметром  $\lambda > 0$ . Теорема доказана.



## Суммы случайных индикаторов. Теорема Пуассона. 5

Приведенные выше версии теоремы Пуассона описывают поведение числа редко наблюдаемых событий при большом числе испытаний. Их иногда называют *теоремами о редких событиях* или *законами малых чисел*. Именно эти теоремы могут рассматриваться как математическое обоснование использования пуассоновского распределения во многих прикладных задачах.

Рассмотрим теперь вопрос о скорости сходимости в законах малых чисел, являющийся ключевым для понимания адекватности пуассоновской аппроксимации.

Для  $k = 0, 1, 2, \dots$  обозначим

$$\pi_k = e^{-\lambda} \frac{\lambda^k}{k!}, \quad b_k = P(S_n = k),$$

$$\Pi = \{\pi_0, \pi_1, \pi_2, \dots\}, \quad B = \{b_0, b_1, b_2, \dots\}.$$

Определим *расстояние по вариации* между распределением случайной величины  $S_n$  и распределением Пуассона по формуле

$$\|B - \Pi\| = \sum_{k=0}^{\infty} |b_k - \pi_k|.$$

## Суммы случайных индикаторов. Теорема Пуассона. 6

Следующие результаты приведем без доказательств, ограничившись лишь соответствующими ссылками.

**Теорема** (Barbour, Hall, 1984). Если  $p_{n,1} = \dots = p_{n,n} = p$ ,  $\lambda = np$ , то справедливо неравенство

$$\|B - \Pi\| \leq 2p \min\{1, \lambda\}.$$

Обозначим

$$\lambda = p_{n,1} + \dots + p_{n,n}, \quad \lambda_2 = p_{n,1}^2 + \dots + p_{n,n}^2.$$

Легко видеть, что

$$ES_n = \lambda, \quad DS_n = \lambda - \lambda_2. \quad (12)$$

**Теорема** (LeCam, 1960). Справедливо неравенство

$$\|B - \Pi\| \leq 2 \min\{9, \lambda\} \max_{1 \leq j \leq n} p_{n,j}.$$

**Теорема** (Пресман, 1985). *Справедливо неравенство*

$$\|B - \Pi\| \leq 2.08 \frac{\lambda_2}{\lambda}.$$

**Теорема** (Deheuvels and Pfeifer, 1987), (Deheuvels and Pfeifer, 1988).  
*Если  $\lambda \rightarrow \infty$  и  $\lambda_2/\lambda \rightarrow 0$ , то*

$$\|B - \Pi\| \sim \frac{\lambda_2}{\sqrt{2\pi e\lambda}}.$$

Из последней теоремы вытекает, что условие  $\max_{1 \leq j \leq n} p_{n,j} \rightarrow 0$  не является необходимым для справедливости пуассоновской аппроксимации для пуассон-биномиального распределения. На самом же деле в качестве *необходимого и достаточного* условия нужно рассматривать условие  $\lambda_2/\lambda \rightarrow 0$ , которое в силу (12) равносильно тому, что

$$\frac{DS_n}{ES_n} \rightarrow 1.$$

# Центральная предельная теорема. 1

Термин *центральная предельная теорема* означает любое утверждение о том, что при выполнении определенных условий функция распределения суммы малых случайных величин с ростом числа слагаемых сходится к нормальной функции распределения. Важность центральной предельной теоремы объясняется тем, что она дает теоретическое объяснение следующему многократно подтвержденному практикой наблюдению: если исход случайного эксперимента определяется большим числом случайных факторов, влияние каждого из которых пренебрежимо мало, то распределение результата такого эксперимента хорошо аппроксимируется нормальным законом с соответствующим образом подобранными математическим ожиданием и дисперсией.

Пусть  $X_1, X_2, \dots$  - независимые одинаково распределенные случайные величины, удовлетворяющие условиям

$$EX_1 = 0, \quad 0 < DX_1 = \sigma^2 < \infty. \quad (1)$$

Функцию распределения случайной величины  $X_1$  обозначим  $F(x)$ .

Функцию распределения нормированной суммы

$S_n = (X_1 + \dots + X_n)/(\sigma\sqrt{n})$  обозначим  $F_n(x)$ .

## Центральная предельная теорема. 2

Функцию распределения и плотность стандартного нормального закона как и ранее будем обозначать  $\Phi(x)$  и  $\varphi(x)$  соответственно,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2} \right\}.$$

Центральная предельная теорема утверждает, что последовательность функций распределения нормированных сумм  $S_n$  случайных величин, удовлетворяющих условию (1), при  $n \rightarrow \infty$  равномерно сходится к стандартной нормальной функции распределения:

$$\rho(F_n, \Phi) \equiv \sup_x |F_n(x) - \Phi(x)| \rightarrow 0.$$

При этом, как мы увидим ниже, второе из условий (1) – условие конечности дисперсии каждого слагаемого – является необходимым и достаточным для указанной равномерной сходимости, если распределения слагаемых одинаковы.

## Центральная предельная теорема. 3

Если же распределения слагаемых различны, то сходимость распределений центрированных и нормированных сумм независимых случайных слагаемых к нормальному закону имеет место, если вклад каждого слагаемого в сумму мал по сравнению с самой суммой, то есть ни одно из слагаемых не играет доминирующей роли. Чтобы формализовать сказанное, обозначим

$$EX_j = a_j, \quad a_1 + \dots + a_n = A_n; \quad DX_j = \sigma_j^2, \quad \sigma_1^2 + \dots + \sigma_n^2 = B_n^2,$$

$$F_n(x) = P(X_1 + \dots + X_n - A_n < xB_n), \quad j \geq 1, \quad n \geq 1.$$

## Центральная предельная теорема. 4

Наиболее хорошо известной версией центральной предельной теоремы для сумм неодинаково распределенных независимых слагаемых является *теорема Линдеберга–Феллера*, которая формулируется следующим образом.

**Теорема** (Lindeberg, 1922), (Feller, 1935). Пусть случайные величины  $X_1, X_2, \dots$  независимы. Для того чтобы

$$\sup_x |F_n(x) - \Phi(x)| \longrightarrow 0 \quad (n \rightarrow \infty)$$

и при каждом  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \max_{1 \leq j \leq n} P(|X_j - a_j| > \epsilon B_n) = 0,$$

необходимо и достаточно, чтобы было выполнено условие Линдеберга: для любого  $\tau > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{j=1}^n \int_{|x-a_j| > \tau B_n} (x - a_j)^2 dP(X_j < x) = 0.$$

## Центральная предельная теорема. 5

Заметим, что все приводимые утверждения формально корректны, если  $\sigma_1^2 > 0$ . Несложно убедиться, что в случае, когда распределения слагаемых в сумме одинаковы, условие Линдеберга оказывается эквивалентным условию конечности дисперсии.

Условие Линдеберга означает, что ни одно слагаемое (ни одна конечная группа) слагаемых в сумме не играет доминирующей роли.

Предположим, что  $\beta_j^3 \equiv E|X_j - a_j|^3 < \infty$ ,  $j \geq 1$ , и обозначим  $\beta_1^3 + \dots + \beta_n^3 = M_n^3$ . Тогда, как несложно видеть,

$$\begin{aligned} & \frac{1}{B_n^2} \sum_{j=1}^n \int_{|x-a_j| > \tau B_n} (x - a_j)^2 dP(X_j < x) \leq \\ & \leq \frac{1}{\tau B_n^3} \sum_{j=1}^n \int_{|x-a_j| > \tau B_n} |x - a_j|^3 dP(X_j < x) \leq \frac{M_n^3}{\tau B_n^3}. \end{aligned}$$



# Центральная предельная теорема. 6

Поэтому в случае существования третьих моментов слагаемых условие Линдеберга вытекает из условия *Ляпунова*:

$$\lim_{n \rightarrow \infty} \frac{M_n^3}{B_n^3} = 0,$$

то есть для случая конечных третьих абсолютных моментов слагаемых условие Ляпунова является достаточным для равномерной сходимости функций распределения центрированных и нормированных сумм независимых случайных слагаемых к стандартной нормальной функции распределения. Это утверждение, доказанное на рубеже XIX–XX веков в работах (Liapunoff, 1900), (Liapunoff, 1901), принято называть *теоремой Ляпунова*.

# Центральная предельная теорема. 7

Если не оговорено противное, ниже подразумевается, что распределения слагаемых одинаковы.

На практике центральная предельная теорема применяется в качестве обоснования возможности аппроксимации распределения случайной величины, имеющей аддитивную структуру, то есть допускающей представление в виде суммы (независимых) случайных величин, нормальным законом. При этом для практических нужд, особенно в задачах оценивания риска, большое значение имеет оценка точности нормальной аппроксимации для распределения суммы  $S_n = X_1 + \dots + X_n$  независимых случайных величин.

Параметрами такой оценки разумно считать число слагаемых  $n$  и простейшие интегральные характеристики распределения слагаемых, например, моменты. Известно, что без дополнительных предположений сходимость распределений нормированных сумм к нормальному закону имеет место, но может быть как угодно медленной (Мацкявичюс, 1983). Поэтому для построения нетривиальных оценок скорости сходимости в центральной предельной теореме необходимы дополнительные условия и дополнительная информация.

# Центральная предельная теорема. 8

Простейшим дополнительным условием такого рода может стать требование существования абсолютного момента

$$\beta_{2+\delta} = E|X_1|^{2+\delta} < \infty$$

с  $0 < \delta \leq 1$ , а дополнительной информацией – конкретное значение  $\beta_{2+\delta}$ . Обозначим

$$\rho(F_n, \Phi) = \sup_x |F_n(x) - \Phi(x)|.$$

Для случая  $\delta = 1$  справедливо *неравенство Берри–Эссеена*

$$\rho(F_n, \Phi) \leq C \frac{\beta_3}{\sigma\sqrt{n}},$$

где  $C$  – абсолютная положительная константа. История поиска числовых оценок этой константы насчитывает около восьмидесяти лет и изобилует замечательными результатами, коллизиями и курьезами. В 1956 г. Эссеев доказал, что

$$C \geq \frac{\sqrt{10} + 3}{6\sqrt{2\pi}} \approx 0.409732.$$

Недавно И. Г. Шевцова показала, что  $C \leq 0.4690$  (Шевцова, 2013).

# Центральная предельная теорема. 9

Для случая  $0 < \delta < 1$  еще в 1967 г. Л. В. Осипов и В. В. Петров показали, что если  $0 < \delta < 1$ , то при  $n \rightarrow \infty$

$$\rho(F_n, \Phi) = o(n^{-\delta/2}).$$

Если же  $\delta = 0$ , то есть известно лишь, что существуют вторые моменты слагаемых, то для любого  $\varepsilon \in (0, \infty)$  справедлива оценка

$$\rho(F_n, \Phi) \leq 1,86(L_n(\varepsilon) + M_n(\varepsilon)) \leq 1,86(L_n(\varepsilon) + \varepsilon) \quad (2),$$

где  $L_n(\varepsilon)$  – это дробь Линдеберга,

$$L_n(\varepsilon) = \frac{1}{B_n^2} \sum_{i=1}^n \mathbb{E} X_i^2 \mathbb{I}(|X_i| \geq \varepsilon B_n), \quad M_n(\varepsilon) = \frac{1}{B_n^3} \sum_{i=1}^n \mathbb{E} |X_i|^3 \mathbb{I}(|X_i| < \varepsilon B_n),$$

см. (Осипов, 1966), (Korolev, Dorofeeva, 2016). Оценки типа (2) разумно считать *естественными*, поскольку они связывают *скорость* сходимости в центральной предельной теореме с *критерием* сходимости.

# Устойчивые и безгранично делимые распределения. 1

Пусть случайные величины  $X_1, X_2, \dots$  независимы и одинаково распределены. Из теоремы Линдберга–Феллера вытекает, что для сходимости распределений их последовательных сумм, нормированных и центрированных соответствующим образом, к нормальному закону необходимо и достаточно существования конечной дисперсии слагаемых. Возникает вполне резонный вопрос: а к каким распределениям могут сходиться распределения нормированных сумм независимых и одинаково распределенных случайных величин, если у слагаемых нет моментов второго порядка? Оказывается, что класс предельных законов для таких сумм совпадает с классом *устойчивых распределений*.

Функция распределения  $G(x)$  и соответствующая ей характеристическая функция  $g(t)$  называются *устойчивыми*, если для любых  $a_1 > 0$  и  $a_2 > 0$  найдутся числа  $a > 0$  и  $b \in \mathbb{R}$  такие, что

$$g(a_1 t)g(a_2 t) \equiv e^{ibt} g(at). \quad (3)$$

Несложно убедиться в том, что условие (3) эквивалентно тому, что для любых  $a_1 > 0$ ,  $a_2 > 0$ ,  $b_1 \in \mathbb{R}$  и  $b_2 \in \mathbb{R}$  существуют числа  $a > 0$  и  $b \in \mathbb{R}$  такие, что

$$G(a_1 x + b_1) * G(a_2 x + b_2) \equiv G(ax + b).$$

## Устойчивые и безгранично делимые распределения. 2

Характеристическая функция  $g(t)$  устойчива тогда и только тогда, когда она может быть представлена в виде

$$g(t) = \exp \left\{ iat - c|t|^\alpha \left( 1 + ib \frac{t}{|t|} Q(t, \alpha) \right) \right\},$$

где

$$Q(t, \alpha) = \begin{cases} \tan \frac{\pi\alpha}{2}, & \text{если } \alpha \neq 1, \\ \frac{2}{\pi} \log |t|, & \text{если } \alpha = 1. \end{cases}$$

Параметр  $\alpha$  при этом называется *характеристическим показателем*.

Все невырожденные устойчивые распределения абсолютно непрерывны.

Примерами устойчивых распределений являются нормальное ( $\alpha = 2$ ), Коши ( $\alpha = 1$ ), Леви ( $\alpha = 1/2$ ) с плотностью

$$p_{1/2}(x) = \begin{cases} 0, & x < 0, \\ \frac{\sigma}{\sqrt{2\pi x^3}} \exp \left\{ -\frac{\sigma^2}{2x} \right\}, & x > 0 \end{cases}$$

и распределение с плотностью  $\bar{p}(x) = p_{1/2}(|x|)$  при  $x < 0$  и  $\bar{p}(x) = 0$  при  $x > 0$ . Другие примеры устойчивых законов, плотности которых выражаются через элементарные функции, неизвестны.

В 1925 г. П. Леви доказал следующую фундаментальную теорему.

**Теорема.** *Для того чтобы функция распределения  $F(x)$  могла быть предельной при  $n \rightarrow \infty$  для распределений сумм*

$$S_n = \frac{X_1 + \dots + X_n - a_n}{b_n}$$

*независимых одинаково распределенных случайных величин при некотором выборе чисел  $a_n \in \mathbb{R}$  и  $b_n > 0$ , необходимо и достаточно, чтобы она была устойчивой.*

Полное доказательство этой теоремы см., например, в (Хинчин, 1938) и (Гнеденко и Колмогоров, 1949).

## Устойчивые и безгранично делимые распределения. 4

Устойчивые законы являются хорошо известными примерами распределений с тяжелыми хвостами. Если  $G_\alpha(x)$  – устойчивая функция распределения с характеристическим показателем  $\alpha \in (0, 2)$ , то

$$G_\alpha(-x) + 1 - G_\alpha(x) \sim \frac{c}{x^\alpha}$$

при  $x \rightarrow \infty$ , где  $c > 0$ . Более того, если  $Z_\alpha$  – случайная величина с функцией распределения  $G_\alpha(x)$ ,  $\alpha \in (0, 2)$ , то  $E|Z_\alpha|^\gamma < \infty$  для любого  $\gamma < \alpha$ , но моменты величины  $Z_\alpha$  порядков, больше или равных  $\alpha$ , не существуют. Таким образом, дисперсии всех устойчивых законов за исключением нормального не определены (иногда говорят, что дисперсии не-нормальных устойчивых законов «бесконечны»).



# Устойчивые и безгранично делимые распределения. 5

Устойчивые законы, а также отрицательное биномиальное, гамма-распределения и распределение Пуассона являются примерами безгранично делимых законов. Функция распределения  $F(x)$  с характеристической функцией  $f(t)$  называются *безгранично делимыми*, если для любого  $n \geq 1$  существует характеристическая функция  $f_n(t)$  такая, что

$$f(t) \equiv (f_n(t))^n$$

Безгранично делимые характеристические функции нигде не обращаются в нуль.

# Устойчивые и безгранично делимые распределения. 6

Рассмотрим предельную схему, обобщающую рассмотренную выше схему *нарастающих сумм*  $S_n$  случайных величин  $X_1, X_2, \dots$ , образующих *одну* последовательность. А именно, предположим, что распределения слагаемых могут изменяться вместе с изменением числа слагаемых в сумме. Пусть  $\{m_n\}_{n \geq 1}$  – последовательность натуральных чисел. Пусть  $\{\xi_{n,j}\}_{j \geq 1, n = 1, 2, \dots}$  – последовательность серий независимых в каждой серии случайных величин. Говорят, что случайные величины  $\{\xi_{n,j}\}$  удовлетворяют *условию равномерной предельной малости*, если для любого положительного числа  $\epsilon$

$$\lim_{n \rightarrow \infty} \sup_{1 \leq j \leq m_n} P(|\xi_{n,j}| > \epsilon) = 0.$$

В 1937 г. А. Я. Хинчин доказал следующую фундаментальную теорему.

**Теорема.** *Для того чтобы функция распределения  $F(x)$  могла быть предельной при  $n \rightarrow \infty$  для распределений сумм  $S_{n,m_n} = \xi_{n,1} + \dots + \xi_{n,m_n}$  независимых случайных величин, удовлетворяющих условию равномерной предельной малости, необходимо и достаточно, чтобы она была безгранично делимой.*

**Доказательство** см., например, в (Гнеденко и Колмогоров, 1949).

# Равномерное распределение. Закон Бенфорда. 1

Равномерное распределение выступает в качестве математической модели максимально неопределенной стохастической ситуации. Рассмотрим стохастическую ситуацию с  $n$  возможными элементарными исходами  $\omega_1, \dots, \omega_n$ . Предположим, что эти исходы равновероятны:

$$P(\{\omega_1\}) = \dots = P(\{\omega_n\}) = \frac{1}{n}.$$

В таком случае говорят, что задано *дискретное равномерное распределение* вероятностей.

## Равномерное распределение. Закон Бенфорда. 2

Непрерывное равномерное распределение вероятностей на некотором отрезке  $[a, b]$  задается с помощью плотности, имеющей вид

$$f(x) = \frac{\mathbf{1}_{[a,b]}(x)}{b-a} = \begin{cases} \frac{1}{b-a}, & \text{если } x \in [a, b], \\ 0, & \text{если } x \notin [a, b]. \end{cases}$$

Из этого определения, в частности, видно, что равномерное распределение не может быть задано на бесконечном интервале.

Если случайная величина  $X$  имеет равномерное распределение на отрезке  $[a, b]$ , то ее функция распределения  $F_X(x)$  имеет вид

$$F_X(x) = \begin{cases} 0, & \text{если } x < a, \\ \frac{x-a}{b-a}, & \text{если } a \leq x \leq b, \\ 1, & \text{если } x > b. \end{cases}$$

При этом

$$EX = \int_a^b \frac{xdx}{b-a} = \frac{1}{b-a} \left( \frac{a+b}{2} \cdot (b-a) \right) = \frac{a+b}{2},$$

$$DX = \frac{(b-a)^2}{12}.$$

**Пример 1.** Пусть  $X$  – случайная величина, функция распределения  $F_X(x)$  которой строго монотонна (напомним, что функция  $F(x)$  называется строго монотонной, если для любых  $x$  и  $y$  таких, что  $x < y$ , выполнено неравенство  $F(x) < F(y)$ ). Построим новую случайную величину  $Y$ , положив

$$Y = F_X(X).$$

Найдем функцию распределения случайной величины  $Y$ . Во-первых, заметим, что  $0 \leq Y \leq 1$ , поскольку любая функция распределения принимает значения между нулем и единицей. Во-вторых, заметим, что поскольку по условию функция  $F_X(x)$  строго монотонна, определена обратная функция  $F_X^{-1}(y)$ , ставящая каждому числу  $y \in (0, 1)$  в соответствие решение уравнения

$$F_X(x) = y.$$

## Равномерное распределение. Закон Бенфорда. 4

В силу строгой монотонности функции  $F_X(x)$  события  $\{\omega : F_X(X(\omega)) < x\}$  и  $\{\omega : X(\omega) < F_X^{-1}(x)\}$  оказываются эквивалентными. Поэтому при  $x \in [0, 1]$

$$\begin{aligned} F_Y(x) &= P(Y < x) = P(F_X(X) < x) = P(X < F_X^{-1}(x)) = \\ &= F_X(F_X^{-1}(x)) \equiv x. \end{aligned}$$

Таким образом,

$$F_X(x) = \begin{cases} 0, & \text{если } x < 0, \\ x, & \text{если } 0 \leq x \leq 1, \\ 1, & \text{если } x > 1, \end{cases}$$

что соответствует равномерному распределению на отрезке  $[0, 1]$  с плотностью

$$f_X(x) = \mathbf{1}_{[0,1]}(x).$$

**Пример 2.** Пусть случайная величина  $X$  имеет равномерное распределение на отрезке  $[0, 1]$ ,  $F(x)$  – произвольная строго монотонная функция распределения. Как мы видели в предыдущем примере, в таком случае определена обратная функция  $F^{-1}(x)$ . Построим новую случайную величину  $Y$ , положив  $Y = F^{-1}(X)$ . Найдем функцию распределения  $F_Y(x)$  случайной величины  $Y$ . Имеем

$$\begin{aligned} F_Y(x) &= P(Y < x) = P(F^{-1}(X) < x) = \\ &= P(X < F(x)) = F_X(F(x)) \equiv F(x). \end{aligned}$$

# Равномерное распределение. Закон Бенфорда. 6

Как известно, в любой современной системе программирования имеются стандартные процедуры, генерирующие так называемые псевдослучайные числа. Такие псевдослучайные числа можно считать независимыми реализациями случайных величин, имеющих равномерное распределение на отрезке  $[0, 1]$ . В то же время может возникнуть потребность в имитировании не равномерно распределенных случайных величин, а случайных величин, имеющих заданную функцию распределения. Приведенные выше рассуждения показывают, как это можно сделать. В частности, случайная величина  $Y$ , построенная из равномерно на  $[0, 1]$  распределенной случайной величины  $X$  с помощью преобразования

$$Y = -\frac{1}{\lambda} \log X,$$

где  $\lambda$  – некоторое положительное число, будет иметь показательное распределение с плотностью  $f_Y(x) = \lambda e^{-\lambda x}$  и функцией распределения  $F_Y(x) = 1 - e^{-\lambda x}$  ( $x \geq 0$ ).



**Пример 3.** *Равномерное распределение на окружности.* Свернем отрезок  $[0, 1]$  в окружность единичной длины. Будем говорить, что случайно выбираемая точка на окружности имеет равномерное распределение, если вероятность ее попадания в любую дугу окружности равна длине этой дуги. Такое распределение должно иметь место при игре в рулетку. Однако чтобы это было так, нужно, чтобы было выполнено некоторое дополнительное условие. А именно, рассмотрим угол, на который повернулось колесо рулетки до полной остановки, как непрерывную случайную величину  $X$  с плотностью  $f_X(x)$  (мы допускаем углы поворота, превосходящие  $360^\circ$ ). Введем новую единицу измерения угла, которую назовем, скажем, *оборот*. Тогда случайная величина  $X$  может быть представлена в виде

$$X = [X] + \{X\},$$

где  $[X]$  – число полных оборотов колеса рулетки, повернувшегося на угол  $X$ , а  $\{X\} = X - [X]$ . Так как длина окружности считается равной единице, то, очевидно,  $0 \leq \{X\} < 1$ . Величина  $\{X\}$  называется *дробной частью* величины  $X$ , а  $[X]$  называется *целой частью*.

# Равномерное распределение. Закон Бенфорда. 8

Можно показать (см., например, (Феллер, 1984), т. 2, с. 79), что для  $x \in [0, 1]$  плотность  $\tilde{f}(x)$  случайной величины  $\{X\}$  имеет вид

$$\tilde{f}(x) = \sum_{n=-\infty}^{+\infty} f_X(x+n) \left( = \lim_{N \rightarrow \infty} \sum_{n=-N}^N f_X(x+n) \right) \quad (1)$$

и  $\tilde{f}(x) = 0$  для  $x \notin [0, 1]$ . Оказывается, что для того, чтобы плотность  $\tilde{f}(x)$  была “примерно равномерной”, нужно, чтобы плотность  $f_X(x)$  исходной случайной величины  $X$  имела малый максимум, то есть чтобы распределение случайной величины  $X$  было “сильно размазано” по положительной полуоси. А именно, пусть максимум плотности  $f_X(x)$  равен, скажем,  $m$  и достигается, скажем, в точке  $x = a$ , причем при  $x < a$  функция  $f_X(x)$  возрастает, а при  $x > a$  – убывает. Тогда можно показать, что

$$|\tilde{f}(x) - 1| \leq m. \quad (2)$$

Неравенство (2) означает, что чем меньше  $m$  (то есть чем сильнее размазано распределение случайной величины  $X$ ), тем ближе плотность  $\tilde{f}(x)$  к равномерной, то есть тем “случайнее” результат вращения колеса рулетки.

**Пример 4.** *Закон Бенфорда.* Пусть  $X$  – случайная величина, равная первой значащей (ненулевой) цифре наугад выбранного числа из произвольно взятых таблиц (логарифмов, степеней, нормального распределения, длин рек, площади озер, численности населения, и т. п.). Кажется бы, в условиях полной случайности выбора таблиц и числа в таблицах случайная величина  $X$  должна иметь дискретное равномерное распределение на множестве чисел  $\{1, 2, \dots, 9\}$ :

$$P(X = k) = \frac{1}{9} \approx 0.111, \quad k = 1, 2, \dots, 9.$$

Однако на практике эта закономерность отнюдь не наблюдается. Напротив, почти всегда наблюдается другая закономерность:

$$P(X = k) \approx p_k = \log_{10}(k + 1) - \log_{10} k, \quad k = 1, 2, \dots, 9.$$

# Равномерное распределение. Закон Бенфорда. 10

Для наглядности приведем конкретные значения чисел  $p_k$ :

$$\begin{array}{lll} p_1 = 0.3010 & p_2 = 0.1761 & p_3 = 0.1249 \\ p_4 = 0.0969 & p_5 = 0.0792 & p_6 = 0.0669 \\ p_7 = 0.0580 & p_8 = 0.0512 & p_9 = 0.0458 \end{array}$$

При этом вероятность выпадения одной из цифр 1, 2, 3 или 4 совсем не равна ожидаемому числу  $\frac{4}{9} \approx 0.4444$ , а оказывается почти равной 0.7 (0.6992 – для любителей точности). Это обстоятельство может быть использовано при заключении пари подобно тому, как это с успехом делал некий известный прикладной математик, о чем упоминается в (Феллер, 1984), т. 2, с. 80 со ссылкой на статью Ф. Бенфорда, опубликованную в 1938 г. С тех пор описанная закономерность носит название *закона Бенфорда*. Эта закономерность настолько устойчива, что известен случай, когда на основании результатов ревизии прокурор Нью-Йорка возбудил уголовное дело в отношении некоей фирмы, в финансовом отчете которой данные не удовлетворяли закону Бенфорда.

# Равномерное распределение. Закон Бенфорда. 11

Известны многие попытки теоретического обоснования этой эмпирической закономерности. В частности, такое обоснование может быть дано с помощью рассуждений, приведенных в предыдущем примере. А именно, наудачу выбранное число из наугад взятых таблиц может рассматриваться как положительная случайная величина с некоторым (неизвестным) распределением. Первая значащая цифра  $X$  такой случайной величины  $Y$  равна  $k$  в том и только в том случае, когда при некотором целом  $n$  выполнено соотношение

$$k \cdot 10^n \leq Y < (k + 1) \cdot 10^n. \quad (4)$$

Введем случайную величину  $Z = \log_{10} Y$ . Тогда соотношение (4) эквивалентно неравенствам

$$n + \log_{10} k \leq Z < n + \log_{10}(k + 1) \quad (5)$$

Как показано в предыдущем примере, если распределение случайной величины  $Y$  очень сильно размазано по положительной полуоси, то распределение случайной величины  $\{Z\}$  (дробной части величины  $Z$ ) является приближенно равномерным. Но тогда для каждого  $k = 1, 2, \dots, 9$  вероятность неравенств (10.1.5) оказывается приближенно равной  $p_k = \log_{10}(k + 1) - \log_{10} k$ , как и должно быть в соответствии с законом Бенфорда.

# Случайные процессы. 1

**Определение 1.** Случайным процессом называется семейство случайных величин  $X(\tau) = X(\Omega, \tau)$ , заданных на одном (базовом) вероятностном пространстве  $(\Omega, \mathcal{A}, P)$  и зависящих от параметра  $\tau$ , принимающего значения из некоторого множества  $T$ .

Из определения 1 следует, что случайный процесс имеет двоякую сущность. С одной стороны, если  $\Omega \in \Omega$  фиксировано, то  $X(\tau) = X(\Omega, \tau)$  есть некоторая функция от  $\tau \in T$ . С другой стороны, если  $t \in T$  фиксировано, то  $X(t) = X(\Omega, t)$  – случайная величина, определенная на  $(\Omega, \mathcal{A}, P)$ . Эту случайную величину иногда называют *проекцией* случайного процесса в точке  $t$ .

Договоримся обозначать параметр случайного процесса греческой буквой, если речь идет о процессе как об элементе пространства случайных функций (строгое определение случайной функции мы дадим чуть позже), и латинской буквой, если речь идет о случайной величине, являющейся проекцией процесса в выделенной точке. Так, например,  $\{X(\tau), \tau \in T\}$  или  $X(\tau)$  есть случайный процесс, определенный на  $T$ , а  $\{X_n(t_n)\}_{n \geq 1}$ ,  $t_n \in T$ , – последовательность случайных величин, являющихся проекциями процессов  $X_n(\tau)$  в некоторых точках  $t_n \in T$ .

## Случайные процессы. 2

Очевидно, что последовательности случайных величин  $X_1, X_2, \dots$  являются случайными процессами с  $T = \{1, 2, \dots\}$ . Процессы, определенные на множестве  $T = \{\dots, -1, 0, 1, \dots\}$  или его подмножестве, обычно называются *процессами с дискретным временем* или *случайными последовательностями*. Если же множество  $T$  представляет из себя интервал (конечный или бесконечный), то семейство случайных величин  $X(\tau) = X(\Omega, \tau)$  принято называть *случайным процессом с непрерывным временем*. Тем не менее, параметр  $\tau$  в определении случайного процесса, конечно, не обязан иметь смысл времени.

Рассмотрим случайный процесс  $X(\tau) = X(\Omega, \tau)$ . Как уже было отмечено, при фиксированном  $\Omega_0 \in \Omega$  мы получаем функцию  $X(\tau) = X(\Omega_0, \tau)$ ,  $\tau \in T$ , которую принято называть *траекторией* случайного процесса  $X(\tau)$ . Пусть  $\mathcal{S}$  – функциональное пространство всех возможных траекторий случайного процесса  $X(\tau)$ . В этом случае случайный процесс  $X(\tau)$  можно рассматривать как случайный элемент, принимающий значения в пространстве  $\mathcal{S}$ . Остановимся на данной трактовке определения 1 более подробно.

## Случайные процессы. 3

Итак, пусть  $\mathcal{S}$  – некоторое метрическое пространство. Рассмотрим вероятностные меры, определенные на классе  $\Sigma$  всех борелевских подмножеств пространства  $\mathcal{S}$ . Класс  $\Sigma$  представляет собой  $\sigma$ -алгебру, порожденную всеми открытыми подмножествами пространства  $\mathcal{S}$ , то есть минимальную  $\sigma$ -алгебру, содержащую все открытые подмножества. Вероятностная мера  $P$  на  $\Sigma$  – это неотрицательная счетно-аддитивная функция множеств  $P : \Sigma \rightarrow [0, 1]$  такая, что  $P(\mathcal{S}) = 1$ .

Пусть  $X$  есть отображение базового вероятностного пространства  $(\Omega, \mathcal{A}, P)$  в метрическое пространство  $\mathcal{S}$ . Если отображение  $X$  является измеримым, то есть  $\{\omega \mid X(\omega) \in B\} \in \mathcal{A}$  для любого множества  $B \in \Sigma$ , то  $X$  называется *случайным элементом*. Причем, если пространство  $\mathcal{S}$  является пространством вещественнозначных функций, то  $X$  принято называть *случайной функцией* или *случайным процессом*. Далее в текущем разделе мы будем иметь дело только со случайными процессами.



**Определение 2.** Распределением случайного процесса  $X$  называется вероятностная мера  $P_X$ , заданная на измеримом пространстве  $(S, \Sigma)$ , определенная для всех множеств  $A \in \Sigma$  следующим образом:

$$P_X(A) = P(\{\omega \mid X(\omega) \in A\}) \equiv P(X \in A).$$

При фиксированных значениях  $t_1, \dots, t_k$ ,  $t_i \in T$ ,  $i = 1, \dots, k$ , мы получаем  $k$ -мерный случайный вектор  $(X(\Omega, t_1), \dots, X(\Omega, t_k))$ . Распределения этих случайных векторов для различных  $k \geq 1$  и  $t_1, \dots, t_k$  называются *конечномерными распределениями* процесса  $\{X(\tau), \tau \in T\}$ . Обозначим

$$F_{t_1, \dots, t_k}(x_1, \dots, x_k) = P(\{\omega \mid X(\Omega, t_1) < x_1, \dots, X(\Omega, t_k) < x_k\}).$$

**Определение 3.** Случайный процесс  $\{X(\tau), \tau \in T \subseteq \mathbb{R}\}$  называется процессом с независимыми приращениями, если для любых  $n \geq 1$ , любых  $t_i \in T, i = 0, \dots, n$ , таких, что  $t_0 < t_1 < \dots < t_n$ , случайные величины

$$X(t_0), X(t_1) - X(t_0), \dots, X(t_n) - X(t_{n-1})$$

независимы в совокупности.

**Определение 4.** Случайный процесс  $\{X(\tau), \tau \in T\}$  называется однородным (по времени), если

$$X(t+h) - X(t) \stackrel{d}{=} X(s+h) - X(s)$$

для всех  $t, s$  и  $h \geq 0$  таких, что  $t \in T, t+h \in T, s \in T, s+h \in T$ .

**Определение 5.** Случайный процесс  $\{X(\tau), \tau \in T\}$  называется стохастически непрерывным, если для любых  $t \in T$  и  $\varepsilon > 0$

$$\lim_{s \rightarrow t, s \in T} P(|X(s) - X(t)| > \varepsilon) = 0.$$

**Определение 6.** Однородный стохастически непрерывный случайный процесс с независимыми приращениями, определенный на  $T = [0, \infty)$ , P-почти все траектории которого выходят из нуля, непрерывны справа и имеют конечные пределы слева, называется процессом Леви.

Важным примером случайного процесса с непрерывными траекториями, который можно рассматривать как модель непрерывного хаотического движения, является винеровский процесс, служащий для описания координаты частицы, подверженной броуновскому движению.

**Определение 7.** Процесс Леви  $W_{a,\sigma^2}(\tau)$ ,  $a \in \mathbb{R}$ ,  $\sigma > 0$ , такой, что при  $t > 0$  проекция  $W_{a,\sigma^2}(t)$  имеет нормальное распределение с параметрами  $at$  и  $\sigma^2 t$ , называется винеровским процессом или процессом броуновского движения с коэффициентами сноса  $a$  и диффузии  $\sigma^2$ . Винеровский процесс  $W_{0,1}(\tau)$  называется стандартным винеровским процессом.

Еще одним случайным процессом, широко используемым в качестве математической модели хаотических потоков событий, является пуассоновский процесс.

**Определение 8.** Процесс Леви  $N_\lambda(\tau)$  называется пуассоновским процессом с параметром  $\lambda > 0$ , если при  $t > 0$  проекция  $N_\lambda(t)$  имеет пуассоновское распределение с параметром  $\lambda t$ .

Можно показать, что пуассоновский процесс является целочисленным и обладает неубывающими кусочно постоянными траекториями. Заметим, что  $\lambda$  имеет смысл среднего числа скачков пуассоновского процесса за единицу времени. Параметр  $\lambda$  называется *интенсивностью* пуассоновского процесса. Пуассоновский процесс с единичной интенсивностью договоримся называть *стандартным пуассоновским процессом*.

Ниже эти два типа процессов будут рассмотрены подробнее в качестве базовых моделей многих реальных явлений.

# Энтропийный (информационный) подход к построению вероятностных моделей

Как уже было сказано, в рамках энтропийного (информационного) подхода в качестве вероятностной модели используется наиболее неблагоприятная модель, обладающая свойствами наибольшей неопределенности при заданных предположениях. По сути этот подход реализует минимаксный принцип выбора наилучшего метода в наихудших условиях.

Для строгой формализации базовых принципов такого подхода нам необходимо подробно обсудить понятия (модели) информации и энтропии.

Интуитивно ясно, что понятие информации тесно связано с понятием неопределенности. В свою очередь, в 20-30-х годах XX столетия Хартли и Шеннон предложили связать понятия информации и вероятности. Такой подход оказался вполне разумным и плодотворным.

# Информация и энтропия. 1

Пусть  $A$  и  $B$  – события, вероятности которых положительны.

**Определение 1.** Информацией (по Шеннону), содержащейся в событии  $B$  относительно события  $A$ , называется

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

Если  $B = A$ , то, очевидно,  $I(A|A) = -\log P(A)$ . Таким образом, мы приходим к следующему важному определению.

**Определение 2.** Информацией (по Шеннону), содержащейся в событии  $A$ , называется

$$I(A) = -\log P(A). \quad (1)$$

# Информация и энтропия. 2

Смысл этих определений легко пояснить, рассмотрев простейшие свойства введенных понятий. Эти свойства оказываются аналогичными тем, которые должны быть присущи информации с точки зрения здравого смысла.

- 1 Осуществление события, вероятность которого невелика, как правило, несет в себе больше информации, нежели осуществление события, вероятность которого значительна. Например, в начале нашего столетия представлялось практически невероятно обнаружить живой экземпляр считавшегося давно вымершим вида кистеперых рыб. Когда же такая рыба – целакант – была поймана, фактически произошла революция в ихтиологии. В то же время, поимка любого экземпляра рыбы такого распространенного вида как, скажем, треска, может нести в себе информацию разве что о месте прохождения косяка этих рыб и никакой научной революции не вызывает.

- 2 Если события  $A$  и  $B$  независимы, то, очевидно, осуществление события  $B$  не дает никакой информации о событии  $A$ . Действительно, в этом случае мы имеем

$$I(A|B) = \log 1 = 0.$$

- 3 Если события  $A$  и  $B$  независимы, то их одновременное осуществление несет в себе столько же информации, сколько содержится в каждом из них в отдельности. Действительно, в этом случае мы имеем

$$\begin{aligned} I(AB) &= -\log P(AB) = -\log(P(A)P(B)) = \\ &= -\log P(A) - \log P(B) = I(A) + I(B). \end{aligned}$$



Логарифмическое определение информации (1) восходит к Хартли (Hartley, 1928). Основание логарифма не играет определяющей роли и существенно лишь для выбора единицы измерения информации. Обычно используют логарифмы по основанию 2, для которых единица информации содержится в событии, вероятность которого  $1/2$ . Такая единица информации называется бит (от английского bit (кусочек) или как аббревиатура термина Binary digiT (двоичный разряд)). Единица информации, порожденная натуральными логарифмами, называется нат (nat).

## Информация и энтропия. 4

Пусть  $\mathcal{E}$  – эксперимент, в котором может осуществиться лишь один из  $n$  исходов  $A_1, \dots, A_n$ . Обозначим  $P(A_i) = p_i$  (очевидно, что  $p_1 + \dots + p_n = 1$ ). Тогда мы можем считать информацию, полученную в результате этого эксперимента, случайной величиной, принимающей значения  $I(A_1), \dots, I(A_n)$  соответственно с вероятностями  $p_1, \dots, p_n$ . Обозначим эту случайную величину  $Q(\mathcal{E})$ . Введем следующую интегральную информационную характеристику  $\mathcal{E}$ .

**Определение 3.** Энтропией  $H(\mathcal{E})$  эксперимента  $\mathcal{E}$  называется величина

$$H(\mathcal{E}) = EQ(\mathcal{E}) = - \sum_{i=1}^n p_i \log p_i. \quad (2)$$

Энтропия эксперимента может служить мерой его неопределенности, что подтверждается совпадением свойств формально введенной величины  $H(\mathcal{E})$ , приводимых в следующей теореме, с ожидаемыми с позиций здравого смысла свойствами разумной меры неопределенности.

**Теорема 1.** Величина  $H(\mathcal{E})$  обладает следующими свойствами.

1°.  $H(\mathcal{E}) \geq 0$ , причем равенство достигается тогда и только тогда, когда существует  $i \in \{1, \dots, n\}$  такое, что  $p_i = 1$ .

2°. Пусть  $\mathcal{E}_0$  – эксперимент с  $n$  равновероятными исходами. Тогда  $H(\mathcal{E}) \leq H(\mathcal{E}_0)$ , каким бы ни был эксперимент  $\mathcal{E}$  с таким же числом  $n$  возможных исходов.

3°. Пусть  $\mathcal{E}_1$  – эксперимент с  $n - 1$  исходом, построенный из эксперимента  $\mathcal{E}$  с помощью объединения двух исходов, скажем,  $A_i$  и  $A_j$  ( $i \neq j$ ), и пусть  $\mathcal{E}_2$  – эксперимент с исходами  $A_i$  и  $A_j$ , вероятности которых (в рамках  $\mathcal{E}_2$ ) соответственно равны  $p_i/(p_i + p_j)$  и  $p_j/(p_i + p_j)$ . Тогда

$$H(\mathcal{E}) = H(\mathcal{E}_1) + (p_i + p_j)H(\mathcal{E}_2).$$

4°. Энтропия  $H(\mathcal{E})$  зависит не от  $A_1, \dots, A_n$ , а от  $p_1, \dots, p_n$ , будучи симметрической функцией переменных  $p_1, \dots, p_n$ .

5°. Энтропия  $H(\mathcal{E})$  – непрерывная функция от  $p_1, \dots, p_n$ .

**Доказательство.** 1°. Доопределив функцию  $f(p) = -p \log p$  по непрерывности нулем при  $p = 0$ , заметим, что  $f(p) \geq 0$ , причем  $f(p) = 0$  тогда и только тогда, когда  $p = 0$  или  $p = 1$ .

2°. Обозначим  $g(x) = x \log x$ . Легко убедиться, что  $g''(x) \geq 0$  при  $x \in [0, 1]$ , то есть  $g(x)$  выпукла при указанных  $x$ . Это означает, что для любых неотрицательных  $\alpha_1, \dots, \alpha_n$  таких, что  $\alpha_1 + \dots + \alpha_n = 1$ , при любых неотрицательных  $y_1, \dots, y_n$

$$g\left(\sum_{i=1}^n \alpha_i y_i\right) \leq \sum_{i=1}^n \alpha_i g(y_i).$$

Положим  $\alpha_i = \frac{1}{n}$ ,  $y_i = p_i$ . Тогда в силу последнего неравенства

$$H(\mathcal{E}_0) = -\log \frac{1}{n} = -n \left[ \sum_{i=1}^n \frac{p_i}{n} \log \left( \sum_{i=1}^n \frac{p_i}{n} \right) \right] \geq -n \sum_{i=1}^n \frac{p_i}{n} \log p_i = H(\mathcal{E}),$$

что и требовалось доказать.

# Информация и энтропия. 7

3°. Во-первых, покажем, что  $H(\mathcal{E}_1) \leq H(\mathcal{E})$ . Действительно,

$$\begin{aligned} H(\mathcal{E}_1) &= - \sum_{k \neq i, k \neq j} p_k \log p_k - (p_i + p_j) \log(p_i + p_j) = \\ &= - \sum_{k \neq i, k \neq j} p_k \log p_k - p_i \log(p_i + p_j) - p_j \log(p_i + p_j) \leq \\ &\leq - \sum_{k \neq i, k \neq j} p_k \log p_k - p_i \log p_i - p_j \log p_j = H(\mathcal{E}_1). \end{aligned}$$

Во-вторых, вычислим

$$\begin{aligned} H(\mathcal{E}) - H(\mathcal{E}_1) &= -p_i \log p_i - p_j \log p_j + (p_i + p_j) \log(p_i + p_j) = \\ &= (p_i + p_j) \left[ -\frac{p_i}{p_i + p_j} \log p_i - \frac{p_j}{p_i + p_j} \log p_j + \right. \\ &\quad \left. + \left( \frac{p_i}{p_i + p_j} + \frac{p_j}{p_i + p_j} \right) \log(p_i + p_j) \right] = \\ &= (p_i + p_j) \left[ -\frac{p_i}{p_i + p_j} \log \frac{p_i}{p_i + p_j} - \frac{p_j}{p_i + p_j} \log \frac{p_j}{p_i + p_j} \right] = (p_i + p_j) H(\mathcal{E}_2), \end{aligned}$$

что и требовалось доказать.

Наконец, пункты 4° и 5° очевидны. Теорема доказана.



# Информация и энтропия. 8

Из свойства 2° энтропии видно, что энтропия эксперимента с максимальной неопределенностью максимальна, а из свойства 1° вытекает, что энтропия минимальна в полностью “определенном” эксперименте. Свойство 3° показывает, что энтропия в определенном смысле аддитивна: при увеличении неопределенности за счет увеличения числа исходов энтропия возрастает, причем прирост энтропии пропорционален вероятности дополнительных исходов. Эти свойства показывают, что энтропия, определяемая соотношением (2), является вполне разумной мерой неопределенности стохастического эксперимента с конечным числом исходов. Более того, Д. К. Фаддееву удалось показать, что система свойств 1° – 5° однозначно определяет функционал (2) (Фаддеев, 1956) (см. также (Яглом и Яглом, 1973)). Другими словами, если мы захотим сконструировать какую-либо меру неопределенности, которая должна обладать вполне естественно ожидаемыми от такой характеристики свойствами 1° – 5°, то мы неизбежно придем к энтропии.

# Информация и энтропия. 9

Очевидно, вместо экспериментов с  $n$  исходами мы можем рассматривать дискретные случайные величины, принимающие какие-либо значения  $x_1, \dots, x_n$  с вероятностями  $p_1, \dots, p_n$ , считая, что значение  $x_i$  случайной величины  $X$  наблюдается в результате осуществления исхода  $A_i$  эксперимента  $\mathcal{E}$ . Поэтому по аналогии с (3.4.2) мы можем определить энтропию простой случайной величины  $X$  как

$$H(X) = - \sum_{i=1}^n p_i \log p_i,$$

или, вводя плотность  $p(x)$  распределения  $X$  относительно считающей меры

$$p(x) = \begin{cases} p_i, & x = x_i, \quad i = 1, \dots, n; \\ 0, & x \notin \{x_1, \dots, x_n\}, \end{cases}$$

как

$$H(X) = -E \log p(X). \quad (3)$$

# Информация и энтропия. 10

По формальной аналогии с (3), если  $X$  – абсолютно непрерывная случайная величина с лебеговой плотностью  $p(x)$ , то определим энтропию  $H(X)$  величины  $X$  так же, как (3). Однако, аналогия таких определений энтропии дискретной и абсолютно непрерывной случайных величин оказывается чисто формальной. Как известно, каждая случайная величина  $X$  как функция элементарного исхода может быть представлена в виде поточечного предела последовательности простых случайных величин  $X_n$ . Тогда, конечно же,  $X_n \Rightarrow X$ . Но если мы попытаемся использовать предельный переход при неограниченно увеличивающемся числе значений аппроксимирующих простых случайных величин для того, чтобы получить энтропию предельной абсолютно непрерывной случайной величины  $X$  в виде (3), мы неизбежно потерпим неудачу, так как предел энтропий аппроксимирующих простых случайных величин оказывается бесконечным. Тем не менее, выражение в правой части (3.4.3) оказывается пределом для “стандартизованных” энтропий аппроксимирующих простых случайных величин в следующем смысле. По сути энтропия (3) абсолютно непрерывной случайной величины определяет среднюю информацию, содержащуюся в  $X$  по сравнению с “бесконечно большой аддитивной постоянной”.



# Информация и энтропия. 11

Чтобы пояснить сказанное, разобьем область значений величины  $X$  на непересекающиеся интервалы  $\Delta_i$  равной длины  $\delta$ . Определим соответствующую дискретную аппроксимацию  $X_\delta$  случайной величины  $X$ , полагая  $X_\delta$  равной некоторому фиксированному элементу интервала  $\Delta_i$ , если  $X$  попадает в этот интервал. Тогда  $X_\delta \Rightarrow X$  ( $\delta \rightarrow 0$ ), и при некоторых условиях регулярности на плотность  $p(x)$  величины  $X$  для некоторых точек  $x_i^* \in \Delta_i$  мы имеем

$$\begin{aligned} H(X_\delta) &= - \sum_i P(X \in \Delta_i) \log P(X \in \Delta_i) = - \sum_i p(x_i^*) \delta \log(p(x_i^*) \delta) = \\ &= - \sum_i p(x_i^*) \delta \log p(x_i^*) - \sum_i p(x_i^*) \delta \log \delta = - \sum_i p(x_i^*) \delta \log p(x_i^*) - \log \delta. \end{aligned} \tag{4}$$

Первое слагаемое в правой части (4) представляет собой интегральную сумму Дарбу для величины  $H(X)$ , определяемой соотношением (3).

Поэтому мы можем записать

$$H(X) = \lim_{\delta \rightarrow 0} [H(X_\delta) + \log \delta]. \tag{5}$$

Таким образом, величину  $H(X)$ , определенную для абсолютно непрерывной случайной величины  $X$  соотношением (3) и являющуюся пределом нормированных мер неопределенности, саму можно считать мерой неопределенности стохастического эксперимента, в результате которого наблюдается случайная величина  $X$ , то есть мерой неопределенности ее распределения.

В соответствии с определением 1, величину  $-\log \delta$  в соотношении (5) можно интерпретировать как информацию, содержащуюся в событии, вероятность которого равна  $\delta$ . Таким образом, эта величина может характеризовать рост неопределенности распределения случайной величины  $X_\delta$ , вызванный квантованием. Поэтому можно считать, что  $H(X)$  характеризует неопределенность абсолютно непрерывной случайной величины, обусловленную формой ее распределения. При этом, однако, соотношение (3) имеет разный смысл для дискретных и абсолютно непрерывных случайных величин.

**Определение 4.** Величина  $H(X)$ , определенная для абсолютно непрерывной случайной величины  $X$  соотношением (3), называется дифференциальной энтропией случайной величины  $X$ .

# Информация и энтропия. 13

Рассмотрим теперь вопрос о том, какие абсолютно непрерывные распределения имеют наибольшую дифференциальную энтропию. Ответ на него укажет на наиболее неопределенные (непредсказуемые) абсолютно непрерывные случайные величины.

В вариационном исчислении хорошо известен следующий метод решения так называемой изопериметрической задачи.

Пусть  $p(x)$  – некоторая функция вещественного аргумента  $x$ ,  $a$  и  $b$  – некоторые числа,  $a < b$ . Предположим, что заданы функции  $F(x, p(x))$ ,  $\varphi_i(x, p(x))$  и числа  $a_i$ ,  $i = 1, \dots, n$ . Известно, что функция  $p(x)$ , обращающая в максимум функционал

$$J(p) = \int_a^b F(x, p) dx$$

при условиях

$$\int_a^b \varphi_i(x, p) dx = a_i, \quad i = 1, \dots, n, \quad (6)$$

находится из уравнения

$$\frac{\partial F}{\partial p} + \lambda_1 \frac{\partial \varphi_1}{\partial p} + \dots + \lambda_n \frac{\partial \varphi_n}{\partial p} = 0. \quad (7)$$

Здесь  $\lambda_1, \dots, \lambda_n$  – неопределенные коэффициенты, определяемые подстановкой решения уравнения (7) в условия (6) (см., например, (Эльсгольц, 1969)).

**Теорема 2.** 1°. Пусть случайная величина  $X$  имеет равномерное распределение на некотором отрезке  $[-a, a]$ . Тогда  $H(X) \geq H(Y)$  для любой случайной величины  $Y$ , удовлетворяющей условию  $P(|Y| \leq a) = 1$ .

2°. Пусть случайная величина  $X$  имеет показательное распределение с некоторым параметром  $\mu > 0$ . Тогда  $H(X) \geq H(Y)$  для любой случайной величины  $Y$ , удовлетворяющей условиям  $P(Y \geq 0) = 1$  и  $EY = \mu^{-1}$ .

3°. Пусть случайная величина  $X$  имеет нормальное распределение с некоторыми параметрами  $a$  и  $\sigma^2$ . Тогда  $H(X) \geq H(Y)$  для любой случайной величины  $Y$ , удовлетворяющей условиям  $EY = a$ ,  $DY = \sigma^2$ .

**Доказательство.** 1°. Положим  $F(x, p) = -p \log p$ ,  $\varphi_1(x, p) = p$ .  
Тогда мы имеем задачу

$$J(p) = - \int_{-a}^a p \log p dx \rightarrow \max_p$$

при условии

$$\int_{-a}^a \varphi_1(x, p) dx = 1. \quad (8)$$

Уравнение (7) для этой задачи принимает вид

$$\frac{\partial F}{\partial p} + \lambda \frac{\partial \varphi_1}{\partial p} = -(1 + \log p) + \lambda = 0.$$

Решением этого уравнения является функция  $p(x) = e^{\lambda-1}$ . Эта функция постоянна по  $x \in [-a, a]$ . С учетом условия (8) мы заключаем, что  $p(x) = \frac{1}{2a} \mathbf{1}(-a \leq x \leq a)$ , то есть  $p(x)$  – это плотность распределения, равномерного на  $[-a, a]$ .

2°. Положим  $F(x, p) = -p \log p$ ,  $\varphi_1(x, p) = p$ ,  $\varphi_2(x, p) = xp$ . Тогда мы имеем задачу

$$J(p) = - \int_0^{\infty} p \log p dx \rightarrow \max_p$$

при условиях

$$\int_0^{\infty} \varphi_1(x, p) dx = 1, \quad \int_0^{\infty} \varphi_2(x, p) dx = \mu^{-1}. \quad (9)$$

Уравнение (7) для этой задачи принимает вид

$$\frac{\partial F}{\partial p} + \lambda_1 \frac{\partial \varphi_1}{\partial p} + \lambda_2 \frac{\partial \varphi_2}{\partial p} = -(1 + \log p) + \lambda_1 + \lambda_2 x = 0.$$

Решением этого уравнения является функция  $p(x) = e^{\lambda_1 - 1 + \lambda_2 x}$ . Подставляя эту функцию в условия (9), находим, что  $e^{\lambda_1 - 1} = -\lambda_2$  и  $\lambda_2 = -\mu$ , то есть  $p(x) = \mu e^{-\mu x} \mathbf{1}(x \geq 0)$ .

# Информация и энтропия. 17

3°. Не ограничивая общности, будем считать, что  $a = 0$ . Положим  $F(x, p) = -p \log p$ ,  $\varphi_1(x, p) = p$ ,  $\varphi_2(x, p) = x^2 p$ . Тогда мы имеем задачу

$$J(p) = - \int_{-\infty}^{\infty} p \log p dx \rightarrow \max_p$$

при условиях

$$\int_{-\infty}^{\infty} \varphi_1(x, p) dx = 1, \quad \int_{-\infty}^{\infty} \varphi_2(x, p) dx = \sigma^2. \quad (10)$$

Уравнение (7) для этой задачи принимает вид

$$\frac{\partial F}{\partial p} + \lambda_1 \frac{\partial \varphi_1}{\partial p} + \lambda_2 \frac{\partial \varphi_2}{\partial p} = -(1 + \log p) + \lambda_1 + \lambda_2 x^2 = 0.$$

Решением этого уравнения является функция  $p(x) = e^{\lambda_1 - 1 + \lambda_2 x^2}$ . Подставляя эту функцию в условия (10), находим, что  $e^{\lambda_1 - 1} = \sqrt{-\lambda_2 / \pi}$  и  $\lambda_2 = -\frac{1}{2}\sigma^{-2}$ , откуда  $e^{\lambda_1 - 1} = \frac{1}{\sigma\sqrt{2\pi}}$ , то есть  $p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$ ,  $-\infty < x < \infty$ . Теорема доказана.

# Пуассоновский процесс как модель дискретного хаоса. 1

Здесь и в дальнейшем мы делаем акцент на тех свойствах пуассоновского процесса, которые позволяют считать его основной математической моделью потока событий, которые абсолютно хаотично рассредоточены во времени.

Приведем еще одно определение пуассоновского процесса (по Хинчину).

**Определение 1.** Случайный процесс  $\{\xi(t), t \geq 0\}$ , называется *пуассоновским*, если он обладает следующими свойствами:

- 1  $\xi(t)$  – процесс с независимыми приращениями, то есть для любых  $n, t_0, \dots, t_n$  ( $n$  – натуральное,  $t_0 \leq t_1 \leq \dots \leq t_n$  – вещественные) случайные величины  $\xi(t_1) - \xi(t_0), \dots, \xi(t_n) - \xi(t_{n-1})$  независимы;
- 2  $\xi(t)$  – однородный процесс, то есть для любых  $s, t$  и  $h > 0$  случайные величины  $\xi(t+h) - \xi(t)$  и  $\xi(s+h) - \xi(s)$  одинаково распределены;
- 3  $\xi(0) = 0$ ;
- 4 при  $h \downarrow 0$  и некотором  $\lambda > 0$

$$P(\xi(h) = 0) = 1 - \lambda h + o(h);$$

$$P(\xi(h) = 1) = \lambda h + o(h);$$

$$P(\xi(h) \geq 2) = o(h).$$



## Пуассоновский процесс как модель дискретного хаоса. 2

Найдем распределение случайной величины  $\xi(t)$  при произвольном  $t > 0$ . С этой целью рассмотрим ещ производящую функцию  $\psi_t(s) = \mathbb{E}s^{\xi(t)}$ , определенную при  $|s| \leq 1$ . По свойствам 1, 2 и 3 для произвольного  $h > 0$  мы имеем

$$\begin{aligned}\psi_{t+h}(s) &= \mathbb{E}s^{\xi(t+h)} = \mathbb{E}s^{\xi(t+h)-\xi(t)+\xi(t)} = \\ &= \mathbb{E}s^{\xi(t)}\mathbb{E}s^{\xi(t+h)-\xi(t)} = \mathbb{E}s^{\xi(t)}\mathbb{E}s^{\xi(h)} = \psi_t(s)\psi_h(s).\end{aligned}\quad (1)$$

По свойству 4 при  $h \downarrow 0$  в силу ограниченности  $s$

$$\psi_h(s) = (1 - \lambda h + o(h)) + s(\lambda h + o(h)) + o(h) = 1 - \lambda h(1 - s) + o(h).$$

Подставляя это выражение в (1), мы получаем

$$\psi_{t+h}(s) = \psi_t(s)(1 + \lambda h(s - 1) + o(h)),$$

откуда

$$\frac{\psi_{t+h}(s) - \psi_t(s)}{h} = \lambda(s - 1)\psi_t(s) + o(1).\quad (2)$$

# Пуассоновский процесс как модель дискретного хаоса. 3

Устремляя в (2)  $h \downarrow 0$ , мы приходим к дифференциальному уравнению

$$\frac{d\psi_t(s)}{dt} = \lambda(s-1)\psi_t(s),$$

решением которого, удовлетворяющим начальному условию  $\psi_0(s) \equiv 1$  (см. свойство 3), является функция

$$\psi_t(s) = \exp\{\lambda t(s-1)\}. \quad (3)$$

Раскладывая функцию (3) в ряд по степеням  $s$ , мы получаем

$$\psi_t(s) = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} s^k. \quad (4)$$

Функция (4), будучи степенным рядом, однозначно определяется коэффициентами при степенях  $s$ . Но с учетом определения производящей функции это означает, что

$$P(\xi(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k = 0, 1, \dots$$

Другими словами, случайная величина  $\xi(t)$  имеет распределение Пуассона с параметром  $\lambda t$ .

Итак, пуассоновский процесс принимает только целые неотрицательные значения. При этом свойство 4 означает, что траектории пуассоновского процесса не убывают, кусочно-постоянны, непрерывны справа, а их скачки имеют одинаковую величину, равную единице.

Мы сразу же замечаем, что

$$E\xi(t) \equiv D\xi(t) \equiv \lambda t.$$

Поэтому  $\lambda$  имеет смысл среднего числа скачков пуассоновского процесса за единицу времени. Параметр  $\lambda$  называется *интенсивностью пуассоновского процесса*.

Пусть  $\tau_0 = 0, \tau_1, \dots$  – точки скачков пуассоновского процесса. Большой интерес представляет вопрос о том, каковы распределения случайных величин  $\tau_n - \tau_{n-1}$ ,  $n \geq 1$ , то есть о распределении длин интервалов времени между скачками пуассоновского процесса. Вначале мы дадим не вполне строго обоснованный ответ на этот вопрос, отложив строгое доказательство до следующего раздела.

Итак, пуассоновский процесс однороден, поэтому величины  $\tau_n - \tau_{n-1}$ ,  $n \geq 1$ , распределены одинаково. Пуассоновский процесс имеет независимые приращения, поэтому эти величины независимы. Вследствие однородности, чтобы определить тип распределения этих величин, нам достаточно рассмотреть лишь одну из них, скажем,  $\tau_1 - \tau_0 = \tau_1$ . Событие  $\{\tau_1 > t\}$  эквивалентно событию  $\{\xi(t) = 0\}$ . Поэтому

$$P(\tau_1 \leq t) = 1 - P(\tau_1 > t) = 1 - P(\xi(t) = 0) = 1 - e^{-\lambda t}.$$

Другими словами, случайные величины  $\tau_n - \tau_{n-1}$ ,  $n \geq 1$ , независимы и имеют одно и то же показательное распределение с параметром  $\lambda$ . Если точки скачков пуассоновского процесса отождествить с моментами регистрации некоторых однотипных событий, то  $\xi(t)$  принимает смысл общего количества событий, зарегистрированных до момента  $t$ . В этом смысле оказывается очень полезно рассматривать пуассоновский процесс как точечный процесс на полупрямой  $t \geq 0$ .

**Определение 2.** Последовательность случайных величин  $\{\tau_n\}_{n \geq 1}$  называется *точечным процессом*, если она удовлетворяет следующим условиям:

1° если  $\tau_n < \infty$ , то  $\tau_{n+1} > \tau_n$ ;

2° для всякого  $t < \infty$  найдется такое  $n$ , что  $\tau_n > t$ .

Со всяким точечным процессом  $\{\tau_n\}_{n \geq 1}$  можно связать случайную целочисленную неотрицательную (считающую) меру  $\nu(A)$ , определенную на борелевских множествах  $A$ , положив

$$\nu(A) = \sum_{k=1}^{\infty} \mathbf{1}(\tau_k \in A).$$

Реализацией случайной считающей меры является обычная считающая мера. Каждая траектория точечного процесса однозначно определяет реализацию случайной считающей меры и наоборот. Поэтому иногда точечные процессы определяют как случайные меры. Например, если  $A = [0, t)$ , а  $\{\tau_n\}_{n \geq 1}$  – точки скачков пуассоновского процесса  $\xi(t)$ , то, очевидно,

$$\nu([0, t)) = \xi(t).$$

Пусть  $n \geq 1$  – произвольное целое число,  $[a, b]$  – произвольный непустой интервал,  $\{\tau_j\}_{j \geq 1}$  – точки скачков пуассоновского процесса с некоторой интенсивностью  $\lambda > 0$ . Найдем условное совместное распределение точек скачков пуассоновского процесса, попавших в интервал  $[a, b]$ , при условии, что на этом интервале пуассоновский процесс имеет ровно  $n$  скачков. Перенумеруем точки скачков так, чтобы  $\tau_1$  оказалась наименьшей из точек скачков, попавших в  $[a, b]$ .

**Теорема 3.** *Условное совместное распределение случайных величин  $\tau_1, \dots, \tau_n$  при условии  $\xi(b) - \xi(a) = n$  совпадает с совместным распределением вариационного ряда, построенного по выборке объема  $n$  из равномерного распределения на  $[a, b]$ .*

**Доказательство.** Если  $(\eta_1, \dots, \eta_n)$  – случайный вектор с совместной плотностью  $f(x_1, \dots, x_n)$ ,  $h > 0$ , то при  $h \downarrow 0$

$$P(\eta_j \in [x_j, x_j + h), j = 1, \dots, n) = f(x_1, \dots, x_n)h^n + o(h^n).$$

Мы будем использовать это соотношение. Пусть  $a = t_0 + h < t_1 < \dots < t_n < t_{n+1} = b$ , а  $h > 0$  столь мало, что  $h < \min_{0 \leq j \leq n} (t_{j+1} - t_j)$ . Найдем условную вероятность  $P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n \mid \xi(b) - \xi(a) = n)$ . Очевидно, что

$$\begin{aligned} & P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n \mid \xi(b) - \xi(a) = n) = \\ &= \frac{P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n; \xi(b) - \xi(a) = n)}{P(\xi(b) - \xi(a) = n)} = \\ &= \frac{P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n)}{P(\xi(b) - \xi(a) = n)}. \end{aligned} \tag{1}$$

Рассмотрим событие, стоящее под знаком вероятности в числителе правой части соотношения (1). Оно эквивалентно тому, что в интервалы  $[t_{j-1} + h, t_j)$ ,  $j = 1, \dots, n + 1$ , не попала ни одна точка скачков, а в каждый из интервалов  $[t_j, t_j + h)$ ,  $j = 1, \dots, n$ , попало ровно по одной точке скачков. Поэтому, используя однородность пуассоновского процесса и независимость его приращений мы будем иметь

$$\begin{aligned}
 & P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n) = \\
 & = P(\xi(t_j) - \xi(t_{j-1} + h) = 0, j = 1, \dots, n + 1; \\
 & \quad \xi(t_j + h) - \xi(t_j) = 1, j = 1, \dots, n) = \\
 & = (\lambda h)^n \exp\{-n\lambda h\} \times \\
 & \quad \times \exp\{-\lambda[t_1 - a + t_2 - t_1 - h + \dots + t_n - t_{n-1} - h + b - t_n - h]\} = \\
 & \quad = (\lambda h)^n \exp\{-\lambda(b - a)\}. \quad (2)
 \end{aligned}$$



Далее,

$$P(\xi(b) - \xi(a) = n) = \exp\{-\lambda(b-a)\} \frac{\lambda^n (b-a)^n}{n!} \quad (3)$$

Подставляя (2) и (3) в (1), получаем

$$P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n \mid \xi(b) - \xi(a) = n) = \frac{n!}{(b-a)^n} h^n.$$

С учетом сказанного в самом начале доказательства мы заключаем, что совместная условная плотность случайных величин  $\tau_1, \dots, \tau_n$  при условии  $\xi(b) - \xi(a) = n$  равна  $n!/(b-a)^n$ . Но, как известно, именно такой вид имеет совместная плотность порядковых статистик, построенных по выборке объема  $n$  из равномерного распределения на  $[a, b]$ . Теорема доказана.

Как мы обещали ранее, вернемся к вопросу о совместном распределении случайных величин  $\zeta_j = \tau_j - \tau_{j-1}$ ,  $j = 1, \dots, n$ , в пуассоновском точечном процессе. В соотношении (2) положим  $a = 0$ ,  $b = t_n$ . Тогда

$$P(\tau_j \in [t_j, t_j + h), j = 1, \dots, n) = \lambda^n \exp\{-\lambda t_n\} h^n.$$

Как и при доказательстве теоремы 1, отсюда мы заключаем, для совместной плотности  $p_{\tau_1, \dots, \tau_n}(t_1, \dots, t_n)$  случайных величин  $\tau_1, \dots, \tau_n$  справедливо соотношение

$$p_{\tau_1, \dots, \tau_n}(t_1, \dots, t_n) = \lambda^n \exp\{-\lambda t_n\}.$$

Но якобиан преобразования  $\tau_j \mapsto \zeta_j = \tau_j - \tau_{j-1}$ ,  $j = 1, \dots, n$ , равен единице ( $\tau_0 = 0$ ). Поэтому, обозначая  $s_j = t_j - t_{j-1}$ ,  $j = 1, \dots, n$ , для совместной плотности  $p_{\zeta_1, \dots, \zeta_n}(s_1, \dots, s_n)$  случайных величин  $\zeta_1, \dots, \zeta_n$  мы получаем представление

$$p_{\zeta_1, \dots, \zeta_n}(s_1, \dots, s_n) = \lambda^n \exp\{-\lambda(s_1 + \dots + s_n)\} = \prod_{j=1}^n \lambda e^{-\lambda s_j},$$

что означает, что случайные величины  $\zeta_1, \dots, \zeta_n$  независимы и одинаково показательно распределены с параметром  $\lambda$ .

Как мы уже установили, равномерное распределение является “наиболее непредсказуемым” среди всех распределений, сосредоточенных на конечном интервале (оно имеет максимальную энтропию среди всех таких распределений). Другими словами, равномерное распределение лучше других соответствует представлению об абсолютно хаотичном расположении точек на отрезке. Поэтому доказанная выше теорема 1 убедительно свидетельствует о том, что пуассоновский точечный процесс является как нельзя более подходящей моделью потока случайных событий, абсолютно хаотично рассредоточенных во времени.

В теореме 3 концы интервала  $[a, b]$  выбирались неслучайно. Оказывается, что если концы рассматриваемого интервала отождествить с какими-либо точками скачков пуассоновского процесса, сделав их тем самым случайными, то свойство равномерности распределения точек скачков, расположенных внутри такого случайного интервала, сохранится.

Действительно, очень легко убедиться, что условная плотность  $p_{\zeta_j | \zeta_j + \zeta_{j+1} = s}(x)$  случайной величины  $\zeta_j$  при условии  $\zeta_j + \zeta_{j+1} = s$ , где  $s > 0$  произвольно, равна  $s^{-1} \mathbf{1}(0 \leq x \leq s)$ , что соответствует равномерному распределению на интервале  $[0, s]$ .

Из пункта 2° теоремы 2 вытекает, что пуассоновский процесс является наиболее неопределенным, наиболее хаотичным среди всех процессов восстановления с конечными математическими ожиданиями и абсолютно непрерывными распределениями длин  $\xi_j$ ,  $j \geq 1$ , промежутков времени между последовательными “восстановлениями”, так как его характеризует показательное распределение случайных величин  $\xi_j$ ,  $j \geq 1$ .

Таким образом, установлена связь пуассоновского процесса с двумя наиболее неопределенными распределениями – равномерным и показательным.

Подводя промежуточный итог, можно сделать следующий вывод. Теоремы 1 и 2 являются убедительными аргументами в пользу того, что пуассоновский процесс является самой подходящей математической моделью однородных и абсолютно хаотичных потоков однотипных событий.

Оказывается, есть и асимптотическая связь этого процесса с третьим наиболее непредсказуемым распределением – нормальным. А именно, пуассоновский процесс обладает свойством асимптотической нормальности.

Пусть  $N(t)$ ,  $t \geq 0$ , – пуассоновский процесс с интенсивностью  $\lambda > 0$ . Как мы видели выше,  $EN(t) \equiv DN(t) \equiv \lambda t$ . Наша цель – показать, что пуассоновский процесс является асимптотически нормальным в том смысле, что

$$P\left(\frac{N(t) - \lambda t}{\sqrt{\lambda t}} < x\right) \implies \Phi(x) \text{ при } \lambda t \rightarrow \infty. \quad (1)$$

На самом деле справедливо более сильное утверждение, из которого будет вытекать (1). Оказывается, что сходимость функций распределения, участвующих в (1), равномерна по  $x$  и справедлива следующая оценка скорости этой равномерной сходимости.

**Теорема 4** (Korolev, Shevtsova, 2012). *При любых  $\lambda > 0$ ,  $t > 0$*

$$\sup_x \left| P\left(\frac{N(t) - \lambda t}{\sqrt{\lambda t}} < x\right) - \Phi(x) \right| \leq \frac{0.3041}{\sqrt{\lambda t}}.$$

Обратим внимание, что пуассоновский процесс обладает свойством асимптотической нормальности при  $\lambda t \rightarrow \infty$ , что выполнено, например, при фиксированной интенсивности, но при неограниченно растущем времени или при фиксированном времени, но при неограниченно растущей интенсивности.

# Принципы выбора моделей в рамках энтропийного подхода. 1

Распределения, участвующие в теореме 2 в качестве распределений с максимальной энтропией, то есть в качестве наиболее неопределенных, возникают в предельных теоремах теории вероятностей.

- **Нормальное распределение** возникает в центральной предельной теореме и ее модификациях, описывающих асимптотическое поведение распределений случайных величин с аддитивной структурой при неограниченном увеличении числа случайных аддитивных составляющих.
- **Показательное распределение** возникает в предельных теоремах для случайных величин типа времени жизни (например, в теореме Реньи), в зависимости которых от “атомических случайных величин” операция суммирования заменена операцией взятия максимума (или минимума). Также оно, будучи тесно связанным с пуассоновским процессом, появляется и в предельных теоремах для сумм случайных величин, описывающих потоки редких событий.
- **Равномерное распределение** возникает в предельных теоремах для цензурированных данных, описывая поведение погрешностей округления.

# Принципы выбора моделей в рамках энтропийного подхода. 2

Формальное совпадение энтропии по Шеннону, описанной выше, с хорошо известной из курса физики больцмановской энтропией наводит на заключение о том, что предельные теоремы теории вероятностей сами по себе являются математическими (вероятностными) моделями проявления универсального принципа неубывания энтропии в энергетически или информационно замкнутых системах.

Из этого немедленно вытекает, что три типа распределений, упомянутых выше, отнюдь не исчерпывают все распределения с условно максимальной энтропией. Ближайшими примерами служат устойчивые законы, возникающие в предельных теоремах Леви–Хинчина для сумм независимых одинаково распределенных случайных величин с бесконечными дисперсиями.

Поэтому, если какая-либо вероятностная модель, отличная от упомянутых в теореме 2, оказалась адекватной, следует исследовать те условия, при которых она обладает свойствами максимума энтропии. Описание таких условий может дать дополнительное понимание сути наблюдаемых закономерностей.

# Принципы выбора моделей в рамках энтропийного подхода. 3

Универсальный принцип неубывания энтропии описывает поведение замкнутых систем. Информационная или энергетическая “замкнутость” системы, поведение которой моделируется на языке теории вероятностей, характеризуется тем, что в используемой модели рассматривается только один источник случайности – элементарные (атомические) случайные величины. На практике ни одна из реальных систем замкнутой не является. Поэтому в модели разумно вводить дополнительные источники случайности, “отвечающие” за влияние внешней среды на рассматриваемую систему. При этом естественно возникают смешанные вероятностные модели, основанные на уже описанных предельных распределениях с максимальной энтропией, но параметры которых становятся случайными. На современном языке такие модели называют *байесовскими*.