Конспект лекций
по дополнительным главам теории вероятностей
мехмат МГУ, осень 2023 г. лектор А.М.Зубков

Оглавление

§ 1. Производящие функции и их свойства. Неравенства для вероятности объединения событий

Определение. *Производящей функцией* $f_\nu(s)$ распределения случайной величины ν, принимающей целые неотрицательные значения, называется степенной ряд

$$f_\nu(s) = \mathbf{E}s^\nu = \sum_{r=0}^{\infty} \mathbf{P}\{\nu = r\} s^r.$$

Производящая функция

$$f_\nu(s) = \mathbf{E}s^\nu = \sum_{r=0}^{\infty} \mathbf{P}\{\nu = r\} s^r$$

как степенной ряд с неотрицательными коэффициентами, сумма которых равна 1 является аналитической функцией в круге радиуса 1, а возможно, и в более широкой области.

Свойства производящих функций:

1) $f_\nu(1) = 1, f_\nu(0) = \mathbf{P}\{\nu = 0\},\ \frac{d}{ds^k} f_\nu(s)\Big|_{s=0} = k!\mathbf{P}\{\nu = k\},\ k = 1, 2, \ldots$

2) $f_\nu(s)$ и все ее производные монотонно не убывают на отрезке [0,1].

3) $\frac{d}{ds^k} f_\nu(s)\Big|_{s=1-} = \mathbf{E}\nu^{[k]} = \mathbf{E}\nu(\nu-1)\ldots(\nu-k+1),\ k = 1, 2, \ldots$

Доказательство свойства 3) проводится непосредственно:

$$\frac{d^k}{ds^k} f_\nu(s)\Big|_{s=1-} = \frac{d^k}{ds^k} \sum_{r=0}^{\infty} \mathbf{P}\{\nu = r\} s^r \Big|_{s=1-} =$$

$$= \sum_{r=0}^{\infty} r^{[k]} \mathbf{P}\{\nu = r\} s^{r-k} \Big|_{s=1-} = \sum_{r=0}^{\infty} r^{[k]} \mathbf{P}\{\nu = r\} = \mathbf{E}\nu^{[k]}.$$

Величины $\mathbf{E}\nu^{[k]} = \mathbf{E}\nu(\nu-1)\ldots(\nu-k+1)$ называются *факториальными моментами* целочисленной случайной величины ν.

Факториальные и степенные моменты связаны линейными соотношениями, коэффициенты в которых называются числами Стирлинга первого или второго рода:

$$\mathbf{E}\nu^{[2]} = \mathbf{E}\nu(\nu-1) = \mathbf{E}\nu^2 - \mathbf{E}\nu,\ \mathbf{E}\nu^{[3]} = \mathbf{E}\nu(\nu-1)(\nu-2) = \mathbf{E}\nu^3 - 3\mathbf{E}\nu^2 + 2\mathbf{E}\nu, \ldots$$

$$\mathbf{E}\nu^2 = \mathbf{E}\nu^{[2]} + \mathbf{E}\nu,\ \mathbf{E}\nu^3 = \mathbf{E}\nu^{[3]} + 3\mathbf{E}\nu^2 + \mathbf{E}\nu, \ldots.$$

4) Если $\xi_1, \ldots, \xi_n$ – независимые неотрицательные целочисленные случайные величины, то

$$\mathbf{E}s^{\xi_1 + \ldots + \xi_n} = \mathbf{E}s^{\xi_1} \ldots s^{\xi_n} = \prod_{k=1}^{n} \mathbf{E}s^{\xi_k}.$$

Для многих стандартных дискретных распределений производящие функции имеют простой аналитический вид. Например, если ν имеет распределение Пуассона с параметром λ:

$$\mathbf{P}\{\nu = k\} = \frac{\lambda^k}{k!}\, e^{-\lambda}, \quad k = 0, 1, \ldots,$$

то

$$\mathbf{E}s^{\nu} = \sum_{k=0}^{\infty} \mathbf{P}\{\nu = k\} s^k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\, e^{-\lambda} s^k = e^{\lambda(s-1)}.$$

Факториальные моменты случайной величины ξ, имеющей распределение Пуассона с параметром λ, вычисляются по простым формулам:

$$\mathbf{E}\nu^{[m]} = \frac{d^m}{ds^m}\, e^{\lambda(s-1)}|_{s=1} = \lambda^m, \quad m = 1, 2, \ldots$$

Если ν имеет геометрическое распределение с параметром p:

$$\mathbf{P}\{\nu = k\} = pq^{k-1}, \quad k = 1, 2, \ldots, \quad p, q > 0, p + q = 1,$$

то

$$\mathbf{E}s = \sum_{k=1}^{\infty} \mathbf{P}\{\nu = k\} s^k = \sum_{k=1}^{\infty} pq^{k-1} s^k = \frac{ps}{1 - qs} = \frac{p}{q}\left(\frac{1}{1 - qs} - 1\right).$$

Если ν имеет биномиальное распределение:

$$\mathbf{P}\{\nu = k\} = C_n^k p^k q^{n-k}, \quad k = 0, 1, \ldots, n, \quad p, q > 0, p + q = 1,$$

то

$$\mathbf{E}s^{\nu} = \sum_{k=0}^{n} \mathbf{P}\{\nu = k\} s^k = \sum_{k=0}^{n} C_n^k p^k q^{n-k} s^k = (q + ps)^n.$$

Пусть проводятся независимые однородные испытания Бернулли с вероятностью успеха p и неудачи $q = 1 - p$. Случайная величина ν_r имеет *отрицательное биномиальное распределение* с параметрами p и r, если она равна порядковому номеру испытания, при котором происходит r-й успех, т. е.

$$\mathbf{P}\{\nu_r = k\} = C_{k-1}^{r-1} p^r q^{k-r}.$$

Чтобы найти производящую функцию отрицательного биномиального распределения, заметим, что промежутки времени между моментами появления соседних успехов независимы и имеют такое же распределение, как ν_1, а эта случайная величина имеет геометрическое распределение с параметром p. Поэтому ν_r имеет такое же распределение, как сумма r независимых случайных величин, имеющих геометрическое распределение с параметром p. По свойству 4) производящих функций тогда $\mathbf{E}s^{\nu_r}$ есть r-я степень производящей функции геометрического распределения:

$$\mathbf{E}s^{\nu_r} = \frac{p^r s^r}{(1 - qs)^r},$$

т. е. *отрицательной степенью бинома.*

В различных вероятностных задачах возникают суммы зависимых случайных величин, в частности, суммы зависимых индикаторов. (Например, любую целочисленную неотрицательную случайную величину ξ можно представить в виде суммы индикаторов: $\xi = \sum_{k\geqslant 1} \mathbb{I}\{\xi \geqslant k\}$.) Удобный способ вычисления факториальных моментов сумм зависимых индикаторов дает следующая лемма.

Лемма о факториальных моментах. *Если* $\xi = \chi_1+\ldots+\chi_n$ *— сумма случайных индикаторов, то при любом натуральном* k

$$\mathbf{E}\xi^{[k]} = k! \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathbf{P}\{\chi_{i_1} = \ldots = \chi_{i_k} = 1\}. \tag{1}$$

Замечание. Если суммирование индикаторов проводится по конечному или счетному множеству индексов N, элементы которого не являются линейно упорядоченными, то формулу (1) удобнее использовать в следующем эквивалентном виде:

$$\mathbf{E}\xi^{[k]} = \sum_{\substack{i_1,\ldots,i_k\in N\\ i\alpha\neq i_\beta\,(\alpha\neq\beta)}} \mathbf{P}\{\chi_{i_1} = \ldots = \chi_{i_k} = 1\}. \tag{2}$$

Доказательство леммы. По определению

$$\mathbf{E}\xi^{[k]} = k!\mathbf{E}\,\frac{\xi^{[k]}}{k!} = k!\mathbf{E}C_\xi^k. \tag{3}$$

Покажем, что сумма в правой части (1) равна $\mathbf{E}C_\xi^k$. Для этого представим ее в виде

$$\begin{aligned}\sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathbf{P}\{\chi_{i_1} = \ldots = \chi_{i_k} = 1\} &= \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathbf{E}I\{\chi_{i_1} = \ldots = \chi_{i_k} = 1\} = \\ &= \mathbf{E} \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} I\{\chi_{i_1} = \ldots = \chi_{i_k} = 1\} = \\ &= \mathbf{E}\sum_{m=0}^{m} \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} I\{\xi = m, \chi_{i_1} = \ldots = \chi_{i_k} = 1\}.\end{aligned} \tag{4}$$

Заметим теперь, что если $\xi(\omega) = \chi_1(\omega) + \ldots + \chi_n(\omega) = m$, то в сумме

$$\sum_{1\leqslant i_1<\ldots<i_k\leqslant n} I\{\xi = m, \chi_{i_1} = \ldots = \chi_{i_k} = 1\}$$

ровно C_m^k слагаемых равны 1, а остальные слагаемые равны нулю, т. е. сумма равна $C_m^k I\{\xi = m\}$. Из этого замечания, (4) и (3) следует, что

$$\sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathbf{P}\{\chi_{i_1}=\ldots=\chi_{i_k}=1\} = \mathbf{E}\sum_{m=0}^{m} C_m^k I\{\xi = m\} = \sum_{m=0}^{m} C_m^k \mathbf{P}\{\xi = m\},$$

что по определению и есть $\mathbf{E}C_\xi^k$. Лемма доказана.

Пример: число пустых ячеек в равновероятной схеме размещения частиц по ячейкам. Пусть $\xi_1, \xi_2, \ldots, \xi_T$ — номера ячеек, в которые попали 1-я, 2-я,..., T-я частицы; пусть эти случайные величины независимы и имеют равномерное распределение на множестве $\{1, \ldots, N\}$. Пусть $\eta_j(T) = \sum_{t=1}^T I\{\xi_t = j\}$ — число частиц, попавших в j-ю ячейку. Число $\mu_0(T, N)$ ячеек, оставшихся пустыми после размещения T частиц, есть сумма индикаторов:

$$\mu_0(T, N) = I\{\eta_1(T) = 0\} + \ldots + I\{\eta_N(T) = 0\}.$$

Следствие 1. *При любом натуральном k*

$$\mathbf{E}\mu_0^{[k]}(T, N) = k!\, C_N^k \left(1 - \frac{k}{N}\right)^T = N^{[k]} \left(1 - \frac{k}{N}\right)^T. \tag{5}$$

Доказательство. Согласно лемме о факториальных моментах

$$\mathbf{E}\mu_0^{[k]}(T, N) = k! \sum_{1\leqslant i_1<\ldots<i_k\leqslant N} \mathbf{P}\{I\{\eta_{i_j}(T) = 0\} = 1,\, j = 1, \ldots, k\} =$$
$$= k! \sum_{1\leqslant i_1<\ldots<i_k\leqslant N} \mathbf{P}\{\eta_{i_j}(T) = 0\},\, j = 1, \ldots, k\},$$

Но при любых попарно различных $i_1, \ldots, i_k \in \{1, \ldots, N\}$

$$\mathbf{P}\{\eta_{i_j}(T) = 0, j = 1, \ldots, k\} =$$
$$= \mathbf{P}\{\xi_t \notin \{i_1, \ldots, i_k\},\, t = 1, \ldots, T\} = \left(1 - \frac{k}{N}\right)^T.$$

Лемма о связи вероятностей и факториальных моментов. *Для любой целочисленной неотрицательной случайной величины ξ и любого целого $r \geqslant 1$ справедливо равенство*

$$\mathbf{P}\{\xi = 0\} = 1 + \sum_{k=1}^{r-1} \frac{(-1)^k}{k!}\mathbf{E}\xi^{[k]} + \frac{(-1)^r}{r!}\,\theta\,\mathbf{E}\xi^{[r]}, \quad \textit{где} \quad \theta \in [0,\, 1]. \tag{6}$$

Для любых целых $m, r \geqslant 1$ при некоторых $\theta, \theta' \in [0,1]$ справедливы соотношения

$$\mathbf{P}\{\xi = m\} = \frac{1}{m!} \sum_{k=0}^{r-1} \frac{(-1)^k}{k!} \mathbf{E}\xi^{[m+k]} + \frac{(-1)^r}{m!\, r!} \theta_m \mathbf{E}\xi^{[m+r]}, \tag{7}$$

$$\mathbf{P}\{\xi > m\} = \frac{1}{m!} \sum_{k=0}^{r-1} \frac{(-1)^k}{(m+k+1)\, k!} \mathbf{E}\xi^{[m+k+1]} + \frac{(-1)^r \mathbf{E}\xi^{[m+r+1]}}{m!\, (m+r)\, r!} \theta'_m. \tag{8}$$

Доказательство. Разложим производящую функцию $f(s) = \mathbf{E}s^{\xi}$ по формуле Тейлора в точке $s = 1$:

$$f(s) = f(1) + \sum_{k=1}^{r-1} \frac{1}{k!} f^{(k)}(1)(s-1)^k + \frac{1}{r!} f^{(r)}(u_s)(s-1)^r, \tag{9}$$

где $u_s \in [s, 1]$. Полагая в этом равенстве $s = 0$ и замечая, что

$$f(0) = \mathbf{E}s^{\xi}|_{s=0} = \mathbf{P}\{\xi = 0\}, \quad f^{(k)}(u) \leqslant f^{(k)}(1) = \mathbf{E}\xi^{[k]},\ s \in [0,1],$$

так как все производные производящей функции монотонно возрастают на отрезке [0, 1], получаем равенство (6).

Для доказательства равенства (7) в общем случае рассмотрим m-ю производную производящей функции распределения ξ:

$$\frac{d^m}{ds^m} f(s) = \sum_{j=m}^{\infty} \mathbf{P}\{\xi = j\} j^{[m]} s^{j-m} =$$

$$= f^{(m)}(1) + \sum_{k=1}^{r-1} \frac{1}{k!} f^{(m+k)}(1)(s-1)^k + \frac{1}{r!} f^{(m+r)}(u_s)(s-1)^r, \quad u_s \in [s, 1].$$

Остается положить здесь $s = 0$, заменить производные в 1 факториальными моментами и разделить обе части на $m!$

Для доказательства (8) введем производящую функцию вероятностей $\mathbf{P}\{\xi > m\}$:

$$g(s) = \sum_{m=0}^{\infty} \mathbf{P}\{\xi > m\} s^m = \sum_{m=0}^{\infty} s^m \sum_{k=m+1}^{\infty} \mathbf{P}\{\xi = k\} = \sum_{k=1}^{\infty} \mathbf{P}\{\xi = k\} \sum_{m=0}^{k-1} s^m.$$

Производные функции $g(s)$ имеют вид

$$g^{(r)}(s) = \sum_{k=1}^{\infty} \mathbf{P}\{\xi = k\} \sum_{m=r}^{k-1} m^{[r]} s^{m-r}, \quad r = 0, 1, \ldots$$

Из равенства $(k+1)^{[r+1]} - k^{[r+1]} = k^{[r]}(k+1-(k-r)) = (r+1)k^{[r]}$ следует, что

$$\sum_{m=r}^{k-1} m^{[r]} = \frac{k^{[r+1]}}{r+1}\,;$$

используя это тождество, находим явные выражения для производных $g(s)$ в 1:

$$g^{(r)}(1) = \frac{1}{r+1}\sum_{k=1}^{\infty}\mathbf{P}\{\xi = k\}k^{[r+1]} = \frac{1}{r+1}\mathbf{E}\xi^{[r+1]}.$$

Теперь аналогично предыдущему разлагаем $g^{(m)}(s)$ на отрезке [0,1] по формуле Тейлора в точке $s = 1$:

$$\begin{aligned} g^{(m)}(s) &= \mathbf{P}\{\xi > m\}m! + \sum_{k=m+1}^{\infty}\mathbf{P}\{\xi > k\}k^{[m]}s^{k-m} = \\ &= \sum_{k=m}^{m+r-1}\frac{g^{(k)}(1)}{(k-m)!}\,(s-1)^{k-m} + \frac{g^{(m+r)}(u_s)}{r!}\,(s-1)^r = \\ &= \sum_{k=m}^{m+r-1}\frac{\mathbf{E}\xi^{[k+1]}}{(k+1)(k-m)!}(s-1)^{k-m} + \theta'_m\frac{\mathbf{E}\xi^{[m+r+1]}}{(m+r+1)r!}(s-1)^r. \end{aligned}$$

Остается подставить сюда $s = 0$ и разделить обе части на $m!$:

$$\mathbf{P}\{\xi > m\} = \frac{1}{m!}\sum_{k=0}^{r-1}(-1)^k\frac{\mathbf{E}\xi^{[m+k+1]}}{(m+k+1)k!} + \theta'_m(-1)^r\frac{\mathbf{E}\xi^{[m+r+1]}}{m!(m+r+1)r!},$$

что и требовалось доказать.

Следствие 2 (Формула включения-исключения и неравенства Бонферрони). *Если $A_1, \ldots, A_N$ — события, определенные на одном и том же вероятностном пространстве, и $\xi = \sum\limits_{k=1}^{N}\chi\{A_k\}$ — число одновременно происходящих событий, а $S_k = \sum\limits_{1\leqslant i_1<\ldots<i_k\leqslant N}\mathbf{P}\{A_{i_1}\cap A_{i_2}\cap\ldots\cap A_{i_k}\} = \frac{1}{k!}\,\mathbf{E}\xi^{[k]}$, то*

$$\mathbf{P}\{A_1\cup\ldots\cup A_N\} = 1 - \mathbf{P}\{\xi = 0\} = \sum_{k=1}^{N}(-1)^{k+1}S_k, \tag{10}$$

$$\mathbf{P}\{\xi = m\} = \sum_{k=0}^{N-m}(-1)^k C_{m+k}^{m}S_{m+k}, \tag{11}$$

и при любом $r = 1, 2, \ldots$

$$\sum_{k=1}^{2r} (-1)^{k+1} S_k \leqslant \mathbf{P}\{A_1 \cup \ldots \cup A_N\} = \mathbf{P}\{\xi > 0\} \leqslant \sum_{k=1}^{2r+1} (-1)^{k+1} S_k. \qquad (12)$$

Доказательство. Если $N < \infty$, то производящая функция $\mathbf{E}s^{\xi}$ является многочленом степени не выше N. Поэтому формулы включения-исключения (10) — (11) следуют из (9) при $r = N$ и того, что в силу леммы о факториальных моментах $\mathbf{E}\xi^{[k]} = k!\, S_k$, а неравенства Бонферрони (12) следуют из (6) и из того, что в силу неотрицательности всех производных производящей функции на отрезке [0,1] знак остаточного члена зависят только от четности r.

Формула (12) справедлива и при $N = \infty$, а формулы (10)–(11) при $N = \infty$ справедливы только если ряды в их правых частях сходятся.

Отметим еще, что из формул (6), (10) и (5) следует формула

$$\mathbf{P}\{\mu_0(T, N) = 0\} = \sum_{k=0}^{N} (-1)^k C_N^k \left(1 - \frac{k}{N}\right)^T .$$

Неравенства Бонферрони удобны тем, что в них входят только факториальные моменты, которые сравнительно легко вычисляются, если случайная величина ξ представлена в виде суммы индикаторов. Однако неравенства Бонферрони могут давать бессодержательные результаты, если моменты случайной величины велики. В качестве примера рассмотрим распределение Пуассона: если ξ имеет распределение Пуассона с параметром λ, то ее производящая функция равна $e^{\lambda(s-1)}$, поэтому k-й факториальный момент ξ равен λ^k, и неравенства Бонферрони для ξ принимают вид

$$\sum_{k=1}^{2m} (-1)^{k-1} \frac{\lambda^k}{k!} \leqslant \mathbf{P}\{\xi \geqslant 1\} = 1 - e^{-\lambda} \leqslant \sum_{k=1}^{2m+1} (-1)^{k-1} \frac{\lambda^k}{k!}.$$

В частности,

$$\lambda - \frac{\lambda^2}{2} \leqslant \mathbf{P}\{\xi \geqslant 1\} = 1 - e^{-\lambda} \leqslant \lambda - \frac{\lambda^2}{2} + \frac{\lambda^3}{6}, \qquad (13)$$

$$\lambda - \frac{\lambda^2}{2} + \frac{\lambda^3}{6} - \frac{\lambda^4}{24} \leqslant \mathbf{P}\{\xi \geqslant 1\} = 1 - e^{-\lambda} \leqslant \lambda - \frac{\lambda^2}{2} + \frac{\lambda^3}{6} - \frac{\lambda^4}{24} + \frac{\lambda^5}{120}.$$

Левые части этих неравенств отрицательны соответственно при $\lambda > 2$ и $\lambda > 2.8$, а правые части становятся больше 1 при $\lambda > 1.6$ и $\lambda > 2.2$.

Естественную задачу уточнения неравенств Бонферрони (без использования характеристик, отличных от моментов случайной величины), можно сформулировать как экстремальную задачу: найти точные верхние и нижние границы для $\mathbf{P}\{\xi > 0\}$ при условии, что фиксировано несколько ее первых моментов $\mathbf{E}\xi, \mathbf{E}\xi^2, \ldots, \mathbf{E}\xi^r$. Далее будет получено решение этой задачи в классе $\mathcal{P}(0, [1, \infty))$ распределений, носителем которых является множество $0 \cup [1, \infty)$. Этот класс шире класса $\mathcal{P}(\mathbb{Z}^+)$ распределений целочисленных неотрицательных случайных величин, и поэтому решения экстремальной задачи классе $\mathcal{P}(0, [1, \infty))$ являются оценками для решений в классе $\mathcal{P}(\mathbb{Z}^+)$.

Пусть $a_0, a_1, \ldots$ – действительные числа; для $r = 1, 2, \ldots$ построим матрицы $A_r = \|a_{i+j}\|_{i,j=0}^r$ и введем обозначение

$$D(a_0, a_1, \ldots, a_{2r}) = \det A_r$$

для определителя матрицы A_r.

Теорема*Если* $\mathbf{P}\{\xi \geqslant 0\} = 1, \mathbf{E}\xi^j = m_j < \infty, j = 1, \ldots, 2r,$ *и* $D(m_2, m_3, \ldots, m_{2r}) \neq 0,$ *то*

$$\mathbf{P}\{\xi > 0\} \geqslant -\frac{D(0, m_1, m_2, \ldots, m_{2r})}{D(m_2, \ldots, m_{2r})} \geqslant 0.$$

Если, кроме того, $\mathbf{E}\xi^{2r+1} = m_{2r+1} < \infty, \mathbf{P}\{0 < \xi < 1\} = 0$ *и выполнено условие* $D(\Delta m_3, \Delta m_4, \ldots, \Delta m_{2r+1}) \neq 0,$ *где* $\Delta m_j = m_j - m_{j-1},$ *то*

$$\mathbf{P}\{\xi \geqslant 1\} \leqslant m_1 + \frac{D(0, \Delta m_2, \ldots, \Delta m_{2r+1})}{D(\Delta m_3, \ldots, \Delta m_{2r+1})}.$$

Нижняя оценка для $\mathbf{P}\{\xi > 0\}$ всегда неотрицательна. Верхняя оценка может оказываться больше 1, но можно показать,что в таком случае в классе $\mathcal{P}(0, [1, \infty))$ существуют распределения с моментами $m_1, \ldots, m_{2r+1}$, носитель которых содержится в множестве $[1, \infty)$.

Доказательство теоремы. Для доказательства первого неравенства рассмотрим квадратичную форму

$$Q_1(z_0, \ldots, z_r) = \mathbf{E}\chi\{\xi > 0\}(z_0 + z_1\xi + \ldots + z_r\xi^r)^2 \geqslant 0.$$

Раскрывая скобки, получим:

$$Q_1(z_0, \ldots, z_r) = \sum_{0 \leqslant i,j \leqslant r} z_i z_j \mathbf{E}\chi\{\xi > 0\}\xi^{i+j} = z_0^2\mathbf{P}\{\xi > 0\} + \sum_{\substack{0 \leqslant i,j \leqslant r \\ i+j>0}} z_i z_j \mathbf{E}\xi^{i+j}.$$

Так как форма Q_1 неотрицательно определена, то определитель ее матрицы и определители всех ее главных миноров неотрицательны, в частности,

$$\begin{vmatrix} \mathbf{P}\{\xi > 0\} & m_1 & m_2 & \dots & m_{r-1} & m_r \\ m_1 & m_2 & m_3 & \dots & m_r & m_{r+1} \\ m_2 & m_3 & m_4 & \dots & m_{r+1} & m_{r+2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ m_{r-1} & m_r & m_{r+1} & \dots & m_{2r-2} & m_{2r-1} \\ m_r & m_{r+1} & m_{r+2} & \dots & m_{2r-1} & m_{2r} \end{vmatrix} \geqslant 0. \qquad (14)$$

Этот определитель является линейной функцией по $\mathbf{P}\{\xi > 0\}$, поэтому (14) можно переписать в виде

$$\mathbf{P}\{\xi > 0\} D(m_2, \dots, m_{2r}) + D(0, m_1, \dots, m_{2r}) \geqslant 0.$$

Далее, так как $D(m_2, \dots, m_{2r})$ – определитель матрицы неотрицательно определенной квадратичной формы $\mathbf{E}(z_1\xi + \ldots + z_r\xi^r)^2$, то $D(m_2, \dots, m_{2r}) \geqslant 0$ и – согласно условию $D(m_2, m_3, \dots, m_{2r}) \neq 0$ – он положителен. Значит,

$$\mathbf{P}\{\xi > 0\} \geqslant -\frac{D(0, m_1, m_2, \dots, m_{2r})}{D(m_2, m_3, \dots, m_{2r})}.$$

Стоящий в числителе определитель $D(0, m_1, m_2, \dots, m_{2r})$ отрицателен, поскольку он является определителем матрицы не положительно определенной квадратичной формы

$$\mathbf{E}\chi\{\xi > 0\}(z_0 + z_1\xi + \ldots + z_r\xi^r)^2 - z_0^2\mathbf{P}\{\xi > 0\} \qquad (15)$$

(она принимает отрицательные значения, если $z_1, \dots, z_r$ фиксированы и отрицательны, а $z_0 \to \infty$). Все главные миноры матрицы квадратичной формы (15), не содержащие первую строку и первый столбец, неотрицательны; следовательно, отрицательным должен быть определитель $D(0, m_1, m_2, \dots, m_{2r})$ полной матрицы формы (15).

Следовательно, полученная нижняя оценка для $\mathbf{P}\{\xi > 0\}$ всегда положительна.

Аналогично, рассматривая неотрицательно определенную квадратичную форму

$$Q_2(z_0, \dots, z_r) = \mathbf{E}\chi\{\xi \geqslant 1\}(\xi - 1)(z_0 + z_1\xi + \ldots + z_r\xi^r)^2 \geqslant 0,$$

находим, что

$$Q_2(z_0, \dots, z_r) = \sum_{0 \leqslant i,j \leqslant r} z_i z_j \mathbf{E}\chi\{\xi \geqslant 1\}(\xi - 1)\xi^{i+j} =$$
$$= z_0^2(m_1 - \mathbf{P}\{\xi \geqslant 1\}) + \sum_{\substack{0 \leqslant i,j \leqslant r \\ i+j>0}} z_i z_j (m_{i+j+1} - m_{i+j}).$$

Из неотрицательности определителя

$$\begin{vmatrix} m_1 - \mathbf{P}\{\xi \geqslant 1\} & m_2 - m_1 & m_3 - m_2 & \dots & m_{r+1} - m_r \\ m_2 - m_1 & m_3 - m_2 & m_4 - m_3 & \dots & m_{r+2} - m_{r+1} \\ m_3 - m_2 & m_4 - m_3 & m_5 - m_4 & \dots & m_{r+3} - m_{r+2} \\ \dots & \dots & \dots & \dots & \dots \\ m_r - m_{r-1} & m_{r+1} - m_r & m_{r+2} - m_{r+1} & \dots & m_{2r} - m_{2r-1} \\ m_{r+1} - m_r & m_{r+2} - mr + 1 & m_{r+3} - m_{r+2} & \dots & m_{2r+1} - m_{2r} \end{vmatrix} =$$
$$= (m_1 - \mathbf{P}\{\xi \geqslant 1\}) D(\Delta m_3, \dots, \Delta m_{2r+1}) + D(0, \Delta m_2, \dots, \Delta m_{2r+1}) \geqslant 0$$

и из положительности $D(\Delta m_3, \dots, \Delta m_{2r+1})$ следует, что

$$\mathbf{P}\{\xi \geqslant 1\} \leqslant m_1 + \frac{D(0, \Delta m_2, \dots, \Delta m_{2r+1})}{D(\Delta m_3, \dots, \Delta m_{2r+1})}.$$

Теорема доказана.

§2. Производящие функции и предельные теоремы. Метод моментов

Производящие функции часто используются при доказательстве предельных теорем. Основу для этого дает следующее утверждение, аналогичное теореме непрерывности для характеристических функций.

Теорема непрерывности для производящих функций. *Пусть* $f_n(s) = \sum\limits_{k \geqslant 0} p_{nk} s^k = \mathbf{E} s^{\nu_n}$ $(n = 1, 2, \dots)$ *и* $f(s) = \sum\limits_{k \geqslant 0} p_k s^k = \mathbf{E} s^{\nu}$ *– производящие функции неотрицательных случайных величин. Условия*

$$p_{nk} \to p_k \quad \text{при} \quad n \to \infty \quad \text{при всех} \quad k \geqslant 0 \tag{16}$$

и

$$f_n(s) \to f(s) \quad \text{при} \quad n \to \infty \quad \text{при всех} \quad s \in [0, 1) \tag{17}$$

эквивалентны.

Доказательство. Пусть выполнено (16). Тогда для любого $s \in [0,1)$ и для любого натурального N имеем

$$|f_n(s) - f(s)| \leqslant \sum_{k \geqslant 0} |p_{nk} - p_k| s^k \leqslant$$

$$\leqslant \sum_{k=0}^{N} |p_{nk} - p_k| + \sum_{k \geqslant N} s^k \leqslant \sum_{k=0}^{N} |p_{nk} - p_k| + s^N/(1-s).$$

Пусть $\varepsilon > 0$ – произвольное малое число. Выберем N так, чтобы выполнялось неравенство $s^N/(1-s) < \varepsilon/2$, а затем выберем $n_0 < \infty$ так, что $\sum\limits_{0 \leqslant k < N} |p_{nk} - p_k| < \varepsilon/2$ при всех $n > n_0$. Тогда при всех $n > n_0$ выполняется неравенство $|f_n(s) - f(s)| \leqslant \varepsilon$, что и означает выполнение (17).

Пусть теперь выполнено (17). Допустим, что (16) не выполнено хотя бы при одном значении k. Пусть для определенности (16) не выполняется при $k = 0$ (в других случаях рассуждения аналогичны).

Каждая из последовательностей $\{p_{nk},\ n = 1, 2, \ldots\}$ ограничена, поэтому из них можно выбирать сходящиеся подпоследовательности. Выберем возрастающую последовательность $\{n_0(t),\ t = 1, 2, \ldots\}$ так, чтобы последовательность $\{p_{n_0(t),0}\}$ сходилась к какому-то неотрицательному числу $q_0 \neq p_0$ (это возможно, так как по предположению p_{n0} не сходится к p_0). Затем из $\{n_0(t)\}$ выберем подпоследовательность $\{n_1(t)\}$ так, чтобы последовательность $\{p_{n_1(t),1}\}$ сходилась к некоторому неотрицательному числу q_1 (при этом, конечно, $p_{n_1(t),0} \to q_0$), и т.д. Тогда для построенной «диагональным методом Кантора» последовательности $\{n_t(t),\ t = 1, 2, \ldots\}$ будет выполняться условие

$$p_{n_t(t),k} \to q_k \quad \text{при} \quad t \to \infty \quad \text{для всех} \quad k \geqslant 0.$$

По доказанному тогда

$$f_{n_t(t)}(s) \to \sum_{k=0}^{\infty} q_k s^k \quad \text{при} \quad t \to \infty \quad \text{при всех} \quad s \in [0,1),$$

причем $q_0 \neq p_0$, и поэтому $\sum\limits_{k \geqslant 0} q_k s^k \neq f(s)$. Но из условия (17) следует, что $f_{n_t(t)}(s) \to f(s)$ при $t \to \infty$. Полученное противоречие доказывает, что $q_0 = p_0$. В общем случае, предположив, что (16) не выполняется, можно найти значение $K = \min\{k : p_{n,k} \not\to p_k\ (n \to \infty)\}$ и повторить проведенные выше рассуждения, начиная с подпоследовательности $p_{n_K(t),K}$, выбрав ее сходящейся к числу $q_k \neq p_k$. Тогда построенная диагональным методом последовательность

$$f_{n_t(t)}(s) \to f_q(s) = \sum_{k=0}^{\infty} q_k s^k \quad \text{при} \quad t \to \infty \quad \text{при всех} \quad s \in [0,1),$$

где $q_j = p_j$ при $j < K$ и $q_K \neq p_K$, и поэтому $f_q(s) \neq f(s)$, т. е. $f_{n_t(t)}(s) \not\to f_q(s)$, $t \to \infty$, что противоречит условию (17). Теорема доказана.

Таким образом, если $\nu_1, \nu_2, \ldots$ – последовательность неотрицательных целочисленных случайных величин с производящими функциями $f_n(s) = \mathbf{E}s^{\nu_n}$, то сходимость $f_n(s) \to f(s)$, $0 \leqslant s < 1$, означает, что распределения ν_n сходятся к распределению целочисленной случайной величины с производящей функцией $f(s)$.

В качестве первого примера применения этого свойства докажем предельную теорему Пуассона для сумм независимых неодинаково распределенных индикаторов.

Теорема Пуассона. *Если $\xi_n = \chi_1^{(n)} + \ldots + \chi_n^{(n)}$ – последовательность целочисленных случайных величин, представленных в виде сумм независимых индикаторов $\chi_j^{(n)}$, $1 \leqslant j \leqslant n, n = 1, 2, \ldots,$*

$$\mathbf{P}\{\chi_j^{(n)} = 1\} = p_{nj}, \quad \mathbf{P}\{\chi_j^{(n)} = 0\} = 1 - p_{nj},$$

и выполняются условия

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

то

$$\mathbf{P}\{\xi_n = k\} \to \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots$$

Доказательство. Так как по условию слагаемые, образующие ξ_n, являются индикаторами и независимы, то

$$\mathbf{E}s_n^{\xi} = \prod_{j=1}^{n} \mathbf{E}s^{\chi_j(n)} = \prod_{j=1}^{n} (1 - p_{nj} + p_{nj}s) = \prod_{j=1}^{n} (1 - p_{nj}(1-s))$$

и в силу равенства

$$\ln(1-x) = -x - \delta_x, \text{ где } 0 \leqslant \delta_x = \sum_{k=2}^{\infty} \frac{x^k}{k} \leqslant \sum_{k=2}^{\infty} \frac{x^k}{2} = \frac{x^2}{2(1-x)} < x^2, \ x \in [0, \tfrac{1}{2}],$$

при некотором $\theta = \theta_s \in [0,\, 1]$

$$\ln \mathbf{E} s^{\xi_n} = \sum_{j=1}^{n} \ln(1 - p_{nj}(1-s)) = -\sum_{j=1}^{n} p_{nj}(1-s) + \theta \sum_{j=1}^{n} (1-s)^2 (p_{nj})^2.$$

Отсюда и из оценки

$$\sum_{j=1}^{n} (p_{nj})^2 < (\max_{1 \leqslant j \leqslant n} p_{nj}) \sum_{j=1}^{n} p_{nj} = (\max_{1 \leqslant j \leqslant n} p_{nj})(\lambda + o(1))$$

следует, что если выполняются условия теоремы, то при любом $s \in [0,1]$

$$\ln \mathbf{E} s^{\xi_n} = -(1-s)(\lambda + o(1)) + \theta\lambda(1-s)^2 \max_{1 \leqslant j \leqslant n} p_{nj} \to (s-1)\lambda, \quad n \to \infty.$$

Следовательно, $\mathbf{E} s^{\xi_n} \to e^{\lambda(s-1)} = \sum\limits_{k \geqslant 0} \frac{\lambda^k}{k!}\, s^k e^{-\lambda}$, и (согласно теореме непрерывности для производящих функций) распределения ξ_n сходятся к распределению Пуассона с параметром λ.

Эквивалентность сходимости производящих функций и сходимости распределений целочисленных случайных величин позволяет сформулировать удобные для проверки условия на моменты случайных величин, достаточные для сходимости последовательностей распределений.

Теорема (метод моментов). *Пусть $\nu_1, \nu_2, \dots$ – последовательность неотрицательных целочисленных случайных величин, $m_k^{(n)} = \mathbf{E}\nu_n^{[k]}$, $n, k = 1, 2, \dots$ Если существуют пределы*

$$m_k = \lim_{n \to \infty} m_k^{(n)}, \quad k = 1, 2, \dots, \tag{18}$$

и

$$m_k = O(k!) \quad \text{при} \quad k \to \infty, \tag{19}$$

то существует такая случайная величина ν, что

$$\lim_{n \to \infty} \mathbf{P}\{\nu_n = j\} = \mathbf{P}\{\nu = j\}, \quad j = 0, 1, \dots,$$

и $\mathbf{E}\nu^{[k]} = m_k$, $k = 1, 2, \dots$

Доказательство. Из условия $m_k = O(k!)$, $k \to \infty$, следует, что ряд $f(s) = 1 + \sum_{k=1}^{\infty} \frac{m_k}{k!} (s-1)^k$ сходится при любом $s \in (0,1]$ (поэтому $f(s)$ непрерывна на $(0,\, 1]$); при этом $f^{(k)}(1) = m_k$, $k = 1, 2, \dots$ Покажем, что

$$f_n(s) = \mathbf{E} s^{\nu_n} \to f(s),\ n \to \infty, \quad \text{при любом } s \in [0,1].$$

Фиксируем произвольные $s \in (0,\,1)$ и $\varepsilon > 0$. Выберем $N = N(\varepsilon) < \infty$ так, чтобы выполнялось неравенство $\sum_{k\geqslant N} \frac{m_k}{k!}(s-1)^k < \frac{\varepsilon}{4}$. Тогда при всех $x \in (0,\,s]$

$$f(x) = 1 + \sum_{k=1}^{N-1} \frac{m_k}{k!}\,(x-1)^k + \theta\frac{\varepsilon}{4}\,,\ \theta = \theta(x) \in [-1,\,1]. \tag{20}$$

Выпишем аналогичную формулу Тейлора для производящей функции $f_n(x)$ с остаточным членом в форме Лагранжа: для некоторого $u_{n,x} \in (x,\,1)$

$$\begin{aligned} f_n(x) &= 1 + \sum_{k=1}^{N-1} \frac{m_k^{(n)}}{k!}\,(x-1)^k + \frac{f_n^{(N)}(u_{n,x})}{N!}(s-1)^N = \\ &= 1 + \sum_{k=1}^{N-1} \frac{m_k^{(n)}}{k!}\,(x-1)^k + \theta_n\frac{m_N^{(n)}}{N!}\,(x-1)^N\,, \end{aligned} \tag{21}$$

где $\theta_n = \theta_n(x) \in [0,\,1]$, поскольку производные $f_n(s)$ неотрицательны, монотонно не убывают на отрезке $[0,\,1]$ и $f_n^{(k)}(1) = m_k^{(n)}$, $k = 1, \ldots, N$. Так как $m_k^{(n)} \to m_k$, $n \to \infty$, $k = 1, 2, \ldots$, то для каждого фиксированного N при всех достаточно больших n

$$\sum_{k=1}^{N-1} \frac{|m_k^{(n)} - m_k|}{k!} < \frac{\varepsilon}{3} \quad \text{и} \quad \frac{m_N^{(n)}}{N!}\,|s-1|^N < \frac{\varepsilon}{3}.$$

Из этих оценок, (20), (21) и произвольности выбора $s \in (0,\,1]$ и $\varepsilon > 0$ следует, что $f_n(s) \to f(s)$, $n \to \infty$, при любом $s \in (0,\,1]$.

Далее, производящие функции $f_n(s)$ на любом отрезке $[0, a] \subset [0,\,1)$ равностепенно непрерывны (точнее, удовлетворяют условию Липшица с коэффициентом $\frac{1}{1-a}$), поскольку они определяются как степенные ряды с неотрицательными коэффициентами, при $s = 1$ все они равны 1, их производные на отрезке $[0, a]$ неотрицательны, монотонны и поэтому не могут быть больше $\frac{f_n(1)-f_n(a)}{1-a} \leqslant \frac{1}{1-a}$. Из равностепенной непрерывности и монотонности функций $f_n(s)$ и их сходимости на $(0,\,1]$ к функции $f(s)$ следует, что функция $\lim_{n\to\infty} f_n(s)$ непрерывна и монотонна на $(0, 1]$, поэтому ее можно доопределить $f(0) = \lim_{n\to\infty} f_n(0)$ по непрерывности.

Рассуждая так же, как при доказательстве теоремы непрерывности для производящих функций, можно показать, что непрерывным пределом последовательности производящих функций (рядов с неотрицательными коэффициентами) может быть только производящая функция. Последнее утверждение теоремы следует из построения производящей функции $f(s)$. Теорема доказана.

Метод моментов часто используется при доказательстве сходимости к распределению Пуассона; факториальные моменты случайной величины ν, имеющей распределение Пуассона с параметром λ, имеют простой вид:

$$\mathbf{E}\nu^{[k]} = \frac{d^k}{ds^k}\mathbf{E}s^{\nu}\Big|_{s=1} = \frac{d^k}{ds^k}e^{\lambda(s-1)}\Big|_{s=1} = \lambda^k e^{\lambda(s-1)}\Big|_{s=1} = \lambda^k.$$

Теорема 2. *Если $\nu_n, n = 1, 2, \ldots,$ — такая последовательность целочисленных неотрицательных случайных величин, что $\mathbf{E}\nu_n^{[k]} \to \lambda^k$ при $n \to \infty$ для любого $k = 1, 2, \ldots,$ то распределения случайных величин ν_n сходятся к распределению Пуассона с параметром λ:*

$$\lim_{n\to\infty} \mathbf{P}\{\nu_n = j\} = \frac{\lambda^j}{j!}\, e^{-\lambda}, \quad j = 0, 1, \ldots$$

Доказательство. При указанных в следствии условиях

$$\mathbf{E}s^{\nu} = \sum_{k\geqslant 0} \frac{\lambda^k}{k!}\,(s-1)^k = e^{\lambda(s-1)},$$

т. е. предельная случайная величина ν имеет распределение Пуассона с параметром λ.

§ 3. Предельная теорема Б. А. Севастьянова для сумм индикаторов

Многие комбинаторно-вероятностные задачи сводятся к исследованию сумм зависимых индикаторов в схеме серий. Одной из нескольких типичных ситуаций являются такие последовательности сумм зависимых индикаторов $\xi^{(n)} = \eta_1^{(n)} + \ldots + \eta_N^{(n)}$ с числом слагаемых $N = N(n) \to \infty$, что математическое ожидание каждого отдельного слагаемого стремится к 0, а математическое ожидание самих сумм ограничено. Пусть $\widetilde{\xi}^{(n)} = \widetilde{\eta}_1^{(n)} + \ldots + \widetilde{\eta}_N^{(n)}$ — сумма *независимых* индикаторов, каждый из которых имеет такое же распределение, как соответствующий индикатор исходной суммы:

$$\mathbf{P}\{\widetilde{\eta}_k^{(n)} = 1\} = \mathbf{P}\{\eta_k^{(n)} = 1\}, \quad k, n = 1, 2, \ldots$$

Если при $n \to \infty$ распределения $\xi^{(n)}$ и $\widetilde{\xi}^{(n)}$ сближаются и распределение $\widetilde{\xi}^{(n)}$ сходится к распределению Пуассона, то и распределение $\xi^{(n)}$ сходится к тому же распределению Пуассона. Однако непосредственно доказать сближение распределений $\xi^{(n)}$ и $\widetilde{\xi}^{(n)}$, как правило, сложно.

Удобный прием доказательства теорем о сходимости к распределению Пуассона методом моментов, отлаженный на доказательстве ряда предельных теорем, был описан Б.А.Севастьяновым в 1972 г.

Пусть $\xi_n = \eta_1^{(n)} + \ldots + \eta_n^{(n)}$ — сумма зависимых индикаторов,

$$\mathbf{P}\{\eta_j^{(n)} = 1\} = b_j^{(n)}, \quad \mathbf{P}\{\eta_j^{(n)} = 0\} = 1 - b_j^{(n)}, \quad j = 1, \ldots, n.$$

Для любого набора

$$\mathbf{i} = (i_1, \ldots, i_m) \in \{1, \ldots, n\}^m, \quad i_r \neq i_s \ (r \neq s),$$

положим

$$\mathbf{P}\{\eta_{i_j}^{(n)} = 1, j = 1, \ldots, m\} = b_{i_1,\ldots,i_m}^{(n)} = b_{\mathbf{i}}^{(n)}.$$

Теорема. *Если вероятности* $b_{i_1,\ldots,i_m}^{(n)} = b_{\mathbf{i}}^{(n)}$ *удовлетворяют следующим условиям:*

а) $\lim\limits_{n\to\infty} \max\limits_{1 \le i \le n} b_i^{(n)} = 0, \quad \lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} b_i^{(n)} = \lambda \in [0, \infty),$

б) *для любых* $m, n \in \{2, 3, \ldots\}$ *существуют такие исключительные множества* $I_m(n)$, *состоящие из наборов* $\mathbf{i} = (i_1, \ldots, i_m)$ *попарно различных элементов из* $\{1, \ldots, n\}$, *что*

$$\lim_{n\to\infty} \sum_{\mathbf{i} \in I_m(n)} b_{\mathbf{i}}^{(n)} = \lim_{n\to\infty} \sum_{\mathbf{i} \in I_m(n)} b_{i_1}^{(n)} \ldots b_{i_m}^{(n)} = 0, \tag{22}$$

$$\lim_{n\to\infty} \max_{\mathbf{i} \notin I_m(n)} \left| \frac{b_{\mathbf{i}}^{(n)}}{b_{i_1}^{(n)} \ldots b_{i_m}^{(n)}} - 1 \right| = 0, \tag{23}$$

то

$$\lim_{n\to\infty} \mathbf{P}\{\xi_n = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots \tag{24}$$

В этой теореме условия а) обеспечивают сходимость к распределению Пуассона сумм независимых индикаторов, а условия б) — сближение распределений сумм зависимых и независимых индикаторов.

Доказательство. Достаточно показать, что

$$\lim_{n\to\infty} \mathbf{E}\xi_n^{[m]} = \lambda^m, \quad m = 1, 2, \ldots$$

Докажем сначала теорему при дополнительном условии, что индикаторы независимы (тем самым получим еще одно доказательство теоремы Пуассона). Будем отмечать соответствующие этому условию величины знаком $\widetilde{\ }$. По формуле для факториальных моментов сумм индикаторов

$$\mathbf{E}\widetilde{\xi}_n^{[m]} = m! \sum_{1\leqslant i_1<\ldots<i_m\leqslant n} \mathbf{P}\{\widetilde{\eta}_{i_j}^{(n)} = 1,\ j = 1,\ldots,m\} =$$

$$= \sum_{\substack{1\leqslant i_1,\ldots,i_m\leqslant n\\ i_j\neq i_k (j\neq k)}} b_{i_1}^{(n)}\ldots b_{i_m}^{(n)} = \left(\sum_{i=1}^n b_i^{(n)}\right)^m - \sum_{\substack{i_1,\ldots,i_m=1\\ \exists r,s\colon i_r=i_s}}^n b_{i_1}^{(n)}\ldots b_{i_m}^{(n)}.$$

Заметим, что последняя сумма неотрицательна, и оценим ее сверху:

$$\sum_{\substack{i_1,\ldots,i_m=1\\ \exists r,s\colon i_r=i_s}}^n b_{i_1}^{(n)}\ldots b_{i_m}^{(n)} \leqslant \sum_{1\leqslant r<s\leqslant m}\ \sum_{\substack{i_1,\ldots,i_m=1\\ i_r=i_s}}^n \prod_{k=1}^m b_{i_k}^{(n)} \leqslant C_m^2 \max_{1\leqslant i\leqslant n} b_i^{(n)} \sum_{i_1,\ldots,i_{m-1}=1}^n \prod_{k=1}^{m-1} b_{i_k}^{(n)} =$$

$$= C_m^2 \max_{1\leqslant i\leqslant n} b_i^{(n)} \left(\sum_{i=1}^n b_i^{(n)}\right)^{m-1} = C_m^2 \max_{1\leqslant i\leqslant n} b_i^{(n)} (\lambda + o(1))^{m-1},$$

так как $\sum_{i=1}^n b_i^{(n)} \to \lambda$ при $n\to\infty$. Следовательно,

$$\mathbf{E}\widetilde{\xi}_n^{[m]} = \left(\sum\nolimits_{i=1}^n b_i^{(n)}\right)^m - O\left(\max_{1\leqslant i\leqslant n} b_i^{(n)}\right) \to \lambda^m \quad \text{при } n\to\infty. \tag{25}$$

Рассмотрим теперь суммы зависимых индикаторов. Пусть $I_m(n)$ — указанные в условии теоремы исключительные множества. Согласно формуле для факториальных моментов суммы индикаторов

$$\mathbf{E}\xi_n^{[m]} = \mathbf{E}(\eta_1^{(n)} + \ldots + \eta_n^{(n)})^{[m]} = \sum_{\substack{1\leqslant i_1,\ldots,i_m\leqslant n:\\ i_j\neq i_k (j\neq k)}} \mathbf{E}\eta_{i_1}^{(n)}\ldots\eta_{i_m}^{(n)} = \tag{26}$$

$$= \sum_{\mathbf{i}\notin I_m(n)} b_{\mathbf{i}}^{(n)} + \sum_{\mathbf{i}\in I_m(n)} b_{\mathbf{i}}^{(n)}, \quad \text{где} \quad \sum_{\mathbf{i}\notin I_m(n)} = \sum_{\substack{\mathbf{i}=(i_1,\ldots,i_m):\\ 1\leqslant i_1,\ldots,i_m\leqslant n\\ i_j\neq i_k (j\neq k)}}.$$

Сравним это разложение с аналогичным разложением для факториальных моментов суммы независимых индикаторов:

$$\mathbf{E}\widetilde{\xi}_n^{[m]} = \mathbf{E}(\widetilde{\eta}_1^{(n)} + \ldots + \widetilde{\eta}_n^{(n)})^{[m]} = \sum_{\substack{1\leqslant i_1,\ldots,i_m\leqslant n:\\ i_j\neq i_k (j\neq k)}} \mathbf{E}\widetilde{\eta}_{i_1}^{(n)}\ldots\widetilde{\eta}_{i_m}^{(n)} =$$

$$= \sum_{\mathbf{i}\notin I_m(n)} b_{i_1}^{(n)}\ldots b_{i_k}^{(n)} + \sum_{\mathbf{i}\in I_m(n)} b_{i_1}^{(n)}\ldots b_{i_k}^{(n)}. \tag{27}$$

Последние слагаемые в правых частях (26) и (27) есть $o(1)$ согласно условию (22). Из условия (23) следует, что для любого $\varepsilon > 0$ существует такое $n_0 < \infty$, что при любом $n > n_0$ и любом наборе $\mathbf{i} = (i_1, \ldots, i_m) \notin I_m(n)$

$$(1-\varepsilon) b_{i_1}^{(n)} \ldots b_{i_m}^{(n)} \leqslant b_{\mathbf{i}}^{(n)} \leqslant (1+\varepsilon) b_{i_1}^{(n)} \ldots b_{i_m}^{(n)}.$$

Суммируя эти неравенства, находим:

$$(1-\varepsilon) \sum_{\mathbf{i} \notin I_m(n)} b_{i_1}^{(n)} \ldots b_{i_m}^{(n)} \leqslant \sum_{\mathbf{i} \notin I_m(n)} b_{\mathbf{i}}^{(n)} \leqslant (1+\varepsilon) \sum_{\mathbf{i} \notin I_m(n)} b_{i_1}^{(n)} \ldots b_{i_m}^{(n)}. \qquad (28)$$

Из (25) — (28) следует, что для любого $\varepsilon > 0$

$$(1-\varepsilon) \lim_{n\to\infty} \mathbf{E}\widetilde{\xi}_n^{[m]} \leqslant \lim_{n\to\infty} \mathbf{E}\xi_n^{[m]} \leqslant (1+\varepsilon) \lim_{n\to\infty} \mathbf{E}\widetilde{\xi}_n^{[m]},$$

т. е. $\lim\limits_{n\to\infty} \mathbf{E}\xi_n^{[m]} = \lambda^m$. Тем самым теорема доказана.

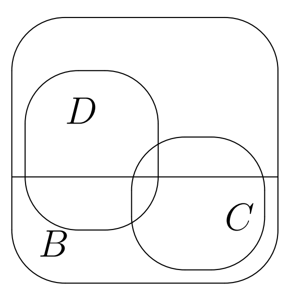
§ 4. Условные математические ожидания

Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$ и заданную на нем случайную величину ξ. Условная вероятность события $C \in \mathcal{F}$ при условии, что происходит некоторое событие $B \in \mathcal{F}$, определяется равенством

$$\mathbf{P}\{C|B\} = \frac{\mathbf{P}\{CB\}}{\mathbf{P}\{B\}} = \mathbf{P}\{CB|B\}.$$

Если событие B фиксировано, то совокупность условных вероятностей $\mathbf{P}\{C|B\}, C \in \mathcal{F}$, задает на измеримом пространстве $(\Omega, \mathcal{F})$ новую вероятностную меру $\mathbf{P}_B, \mathbf{P}_B(C) = \mathbf{P}\{C|B\}$, сосредоточенную на множестве B:

$$\mathbf{P}_B\{B\} = 1, \quad \mathbf{P}_B\{\Omega \backslash B\} = 0.$$



Пусть ξ — случайная величина, заданная на $(\Omega, \mathcal{F}, \mathbf{P})$ (т. е. функция, определенная в каждой точке $\omega \in \Omega$ и принимающая значения из конечного или счетного множества $X = \{x_1, x_2, \ldots\}$); ей соответствует разбиение

$\{A_1, A_2, \ldots\}$ пространства Ω на непересекающиеся $\mathcal{F}$-измеримые множества $A_k = \{\omega \in \Omega\colon \xi(\omega) = x_k\} \in \mathcal{F}, k \geq 1$, порождающее σ-алгебру $\sigma(\xi)$. Математическое ожидание ξ можно вычислять как относительно основного распределения $\mathbf{P}$:

$$\mathbf{E}\xi = \sum_{\omega\in\Omega} \xi(\omega)\mathbf{P}(\omega) = \sum_{k\geq 1} x_k \mathbf{P}\{A_k\},$$

так и относительно условного распределения $\mathbf{P}_B$:

$$\mathbf{E}\{\xi|B\} = \sum_{\omega\in\Omega} \xi(\omega)\mathbf{P}\{\omega|B\}) = \sum_{k\geq 1} x_k \mathbf{P}\{A_k|B\},$$

Например, если $\xi(\omega) = \mathbb{I}_A(\omega)$, то

$$\mathbf{E}\{\mathbb{I}_A \,|\, B\} = \sum_{\omega\in A} \mathbb{I}_A(\omega)\mathbf{P}\{\omega \,|\, B\} = \sum_{\omega\in A} \mathbf{P}\{\omega \,|\, B\} = \mathbf{P}\{A \,|\, B\}. \tag{29}$$

В формуле полной вероятности используются условные вероятности относительно совокупностей событий, образующих разбиение пространства элементарных событий Ω: если $B_1, B_2, \ldots \in \mathcal{F}$ — попарно несовместные события, $\cup_{k\geqslant 1} B_k = \Omega$, то для любого события $A \in \mathcal{F}$

$$\mathbf{P}\{A\} = \sum_{k\geqslant 1} \mathbf{P}\{B_k\}\mathbf{P}\{A \,|\, B_k\}.$$

Если вероятности $\mathbf{P}\{B_k\}$ и $\mathbf{P}\{A \,|\, B_k\}$ для всех k вычисляются просто, то вычисление вероятности события A может существенно упроститься.

Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство и $\mathcal{B} = \{B_1, B_2, \ldots\}$ — измеримое разбиение Ω на множества положительной меры. (Далее все разбиения мы будем предполагать измеримыми, не оговаривая этого специально.) Тогда для каждого события C определены условные вероятности $\mathbf{P}\{C|B_k\}, k = 1, 2, \ldots$ Введем теперь случайную величину (т. е. функцию на Ω), которая на каждом множестве B_k принимает значение $\mathbf{P}\{C|B_k\}$:

$$\xi_{\mathcal{B}} = \xi_{\mathcal{B}}(\omega) = \sum_{k\geq 1} \mathbf{P}\{C|B_k\}\mathbb{I}_{B_k}(\omega) = \mathbf{P}\{C|\mathcal{B}\}.$$

Случайная величина $\xi_{\mathcal{B}}(\omega)$ называется *условной вероятностью* события C относительно разбиения $\mathcal{B}$.

В терминах условных вероятностей стандартная формула полной вероятности

$$\mathbf{P}\{C\} = \sum_{k\geq 1} \mathbf{P}\{C|B_k\}\mathbf{P}\{B_k\}, \quad B_k \cap B_j = \varnothing\,(k \neq j), \quad \bigcup_{k\geq 1} B_k = \Omega,$$

принимает другой вид. Заменим в ней $\mathbf{P}\{B_k\}$ на $\mathbf{E}\chi_{B_k}(\omega)$ и воспользуемся аддитивностью математического ожидания:

$$\mathbf{P}\{C\} = \sum_{k\geq 1} \mathbf{P}\{C|B_k\}\mathbf{P}\{B_k\} = \sum_{k\geq 1} \mathbf{P}\{C|B_k\}\mathbf{E}\mathbb{I}_{B_k}(\omega) =$$
$$= \mathbf{E}\sum_{k\geq 1} \mathbf{P}\{C|B_k\}\mathbb{I}_{B_k}(\omega) = \mathbf{E}\mathbf{P}\{C|\mathcal{B}\},$$

где $\mathcal{B} = \{B_1, B_2, \ldots\}$ — разбиение, или

$$\mathbf{P}\{C\} = \mathbf{E}\mathbf{P}\{C|B\}. \tag{30}$$

Если разбиение $\mathcal{B}$ порождается случайной величиной η, то говорят об *условной вероятности относительно случайной величины* η или относительно порожденной ею σ-алгебры $\sigma(\eta)$:

$$\mathbf{P}\{C|\eta\} = \mathbf{P}\{C|\sigma(\eta)\} = \mathbf{P}\{C|\eta\}(\omega).$$

Рассмотрим теперь аналогичную конструкцию для случайных величин. Пусть на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$ заданы: разбиение $\mathcal{B} = \{B_1, B_2, \ldots\}$ и случайная величина ξ с конечным или счетным множеством значений $x_1, x_2, \ldots \in \mathbb{R}$, порождающая разбиение $\mathcal{A} = \{A_1, A_2, \ldots\}$ пространства Ω на непересекающиеся множества $A_k = \{\omega \in \Omega\colon \xi(\omega) = x_k\} \in \mathcal{F}, k \geq 1$.

Условные вероятности

$$\mathbf{P}\{A_k|\mathcal{B}\} = \mathbf{P}\{A_k|\mathcal{B}\}(\omega) = \mathbf{P}\{\xi = x_k|\mathcal{B}\}(\omega),\ \omega \in \Omega,$$

определены для всех $k \geq 1$, и для каждого $\omega \in \Omega$ они образуют вероятностное распределение на множестве $\{x_1, x_2, \ldots\}$ значений ξ, зависящее от ω.

Определим *условное математическое ожидание* ξ *относительно разбиения* $\mathcal{B}$ как функцию, отображающую Ω в $\mathbb{R}$, т. е. как случайную величину (разумеется, если ряды в определении сходятся абсолютно):

$$\mathbf{E}\{\xi|\mathcal{B}\}(\omega) = \sum_{k\geq 1} x_k \mathbf{P}\{A_k|\mathcal{B}\}(\omega) = \sum_{k\geq 1} x_k \mathbf{P}\{\xi = x_k|\mathcal{B}\}(\omega).$$

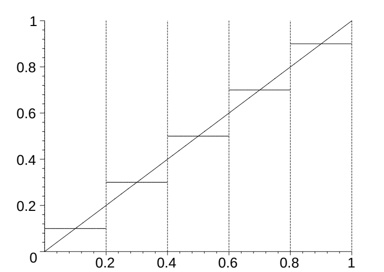
Преобразуем эту формулу (изменения порядка суммирований законны, так как все элементы рядов неотрицательны):

$$\mathbf{E}\{\xi|\mathcal{B}\}(\omega) = \sum_{k\geq 1} x_k \mathbf{P}\{A_k|\mathcal{B}\}(\omega) = \sum_{k\geq 1} x_k \sum_{j\geq 1} \mathbf{P}\{A_k|B_j\}\mathbb{I}_{B_j}(\omega) =$$
$$= \sum_{j\geq 1} \mathbb{I}_{B_j}(\omega) \sum_{k\geq 1} x_k \mathbf{P}\{A_k|B_j\} = \sum_{j\geq 1} \mathbb{I}_{B_j}(\omega)\mathbf{E}\{\xi|B_j\}.$$

Таким образом, условное математическое ожидание ξ относительно разбиения $\mathcal{B}$ есть случайная величина, принимающая на каждом элементе B_j этого разбиения постоянное значение, равное $\mathbf{E}\{\xi|B_j\} \in \mathbb{R}$. Значит, $\mathbf{E}\{\xi|\mathcal{B}\}(\omega)$ — функция, измеримая относительно разбиения $\mathcal{B}$ (для любого $\mathcal{B}$-измеримого множества D событие $\{\omega : \mathbf{E}\{\xi|\mathcal{B}\} \in D\}$ является объединением множеств B_j, т. е. принадлежит σ-алгебре, порожденной разбиением $\mathcal{B}$.

Пример. Пусть $\Omega = [0,1)$, $\mathcal{F}$ — борелевская σ-алгебра, $\mathbf{P}$ — мера Лебега, $\xi(\omega) = \omega$ и σ-алгебра $\mathcal{B} \subset \mathcal{F}$ порождается множествами $[0,\frac{1}{5}), [\frac{1}{5},\frac{2}{5}), [\frac{2}{5},\frac{3}{5}), [\frac{3}{5},\frac{4}{5}), [\frac{4}{5},1)$. Тогда случайную величину $\mathbf{E}\{\xi|\mathcal{B}\}$ можно описать таблицей

B_j	$[0,\frac{1}{5})$	$[\frac{1}{5},\frac{2}{5})$	$[\frac{2}{5},\frac{3}{5})$	$[\frac{3}{5},\frac{4}{5})$	$[\frac{4}{5},1]$
$\mathbf{E}\{\xi|\mathcal{B}\}(\omega)$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{7}{10}$	$\frac{9}{10}$



Утверждение 1. *Для любой случайной величины ξ с конечным математическим ожиданием и для любого события C, измеримого относительно разбиения $\mathcal{B} = \{B_1, B_2, \ldots\}$, справедлива формула*

$$\mathbf{E}\mathbb{I}_C\xi = \mathbf{E}\{\mathbb{I}_C\mathbf{E}\{\xi|\mathcal{B}\}\}. \tag{31}$$

Замечание. Условие конечности $\mathbf{E}\xi$ существенно: например, если $\mathbf{P}\{\xi = k\} = \mathbf{P}\{\xi = -k\} = \frac{1}{2k(k+1)}$, $k = 1, 2, \ldots$, то $\mathbf{E}\xi$ не определено, так как ряд $\sum_{k=1}^{\infty} \frac{k}{2k(k+1)}$ не сходится абсолютно, но $\mathbf{E}\{\xi|B_k\} = 0$ для разбиения $\mathcal{B} = \{B_1, B_2, \ldots\}$, $B_k = \{k, -k\}$, $k \geqslant 1$.

Доказательство. Так как математическое ожидание ξ конечно, то для любого события $C \in \mathcal{B}$ величину $\mathbf{E}\mathbb{I}_C\xi$ можно разбить на слагаемые вида $\mathbf{E}\mathbb{I}_{B_j}\xi$. Так как C измеримо относительно $\mathcal{B}$, то $C = \bigcup\limits_{j: B_j \subset C} B_j$ и ввиду абсолютной сходимости ряда $\sum_{j \geqslant 1} \mathbf{E}\mathbb{I}_{B_j}\xi$

$$\mathbf{E}\mathbb{I}_C\xi = \sum_{j:\, B_j \subset C} \mathbf{E}\mathbb{I}_{B_j}\xi = \sum_{j: B_j \subset C} \frac{\mathbf{E}\mathbb{I}_{B_j}\xi}{\mathbf{P}\{B_j\}} \mathbf{P}\{B_j\} =$$

$$= \sum_{j: B_j \subset C} \mathbf{E}\{\xi | B_j\}\mathbf{P}\{B_j\} = \mathbf{E} \sum_{j: B_j \subset C} \mathbf{E}\{\xi | B_j\}\mathbb{I}_{B_j} = \mathbf{E}\mathbb{I}_C\mathbf{E}\{\xi|\mathcal{B}\}.$$

Следствие 1. *Для любой случайной величины ξ с конечным математическим ожиданием и любого разбиения $\mathcal{B}$ справедлива формула полного математического ожидания*:

$$\mathbf{E}\xi = \mathbf{E}\mathbf{E}\{\xi|\mathcal{B}\}.$$

Отметим несколько свойств условных математических ожиданий.

1) Линейность: для любых случайных величин ξ, η с конечными математическими ожиданиями и чисел a, b справедливо равенство

$$\mathbf{E}\{a\xi + b\eta|\mathcal{B}\} = a\mathbf{E}\{\xi|\mathcal{B}\} + b\mathbf{E}\{\eta|\mathcal{B}\}.$$

Для случайных величин $\xi_1, \xi_2, \ldots$ с конечными математическими ожиданиями и чисел $a_1, a_2, \ldots$ равенство

$$\mathbf{E}\sum_{k=1}^{\infty} a_k\xi_k = \sum_{k=1}^{\infty} a_k\mathbf{E}\xi_k$$

справедливо, если $\mathbf{P}\left\{\sum_{k=1}^{\infty} |a_k\xi_k| < \infty\right\} = 1$ и ряд в правой части сходится абсолютно.

2) Если существует $\mathbf{E}\xi$, то $\mathbf{E}\{\xi|\{\varnothing, \Omega\}\} = \mathbf{E}\xi$.

3) $\mathbf{P}\{\xi = C = \mathrm{const}\} = 1 \Rightarrow \mathbf{E}\{\xi|\mathcal{B}\} = \mathrm{C}$.

4) $\mathbf{E}\{\mathbb{I}_A|\mathcal{B}\} = \mathbf{P}\{A|\mathcal{B}\}$ для любого $A \in \mathcal{F}$.

Действительно, если $\mathcal{B}$ порождается разбиением $\{B_1, B_2, \ldots\}$, то в силу свойства 1) и неотрицательности слагаемых

$$\mathbf{E}\{\mathbb{I}_A|\mathcal{B}\} = \mathbf{E}\left\{\sum_{i \geq 1} \mathbb{I}_{AB_i}|\mathcal{B}\right\} = \sum_{i \geq 1} \mathbf{E}\{\mathbb{I}_{AB_i}|\mathcal{B}\} =$$

$$= \sum_{i \geq 1} \mathbf{E}\{\mathbb{I}_{AB_i}|B_i\}\mathbb{I}_{B_i}(\omega) = \sum_{i \geq 1} \mathbf{P}\{A|B_i\}\mathbb{I}_{B_i}(\omega) = \mathbf{P}\{A|\mathcal{B}\}.$$

5) Если случайная величина η измерима относительно разбиения $\mathcal{B} = \{B_1, B_2, \ldots\}$, то

$$\mathbf{E}\{\xi\eta|\mathcal{B}\} = \eta\mathbf{E}\{\xi|\mathcal{B}\},$$

т.е. *из-под знака условного математического ожидания можно выносить случайный множитель, измеримый относительно σ-алгебры в условии.*

Доказательство. Будем считать, что η принимает на множестве B_j значение $b_j, j \geq 1$. По определению

$$\mathbf{E}\{\xi\eta|\mathcal{B}\}(\omega) = \sum_{j\geq 1} \mathbb{I}_{B_j}(\omega)\mathbf{E}\{\xi\eta|B_j\} = \sum_{j\geq 1} \mathbb{I}_{B_j}(\omega)\mathbf{E}\{\xi b_j|B_j\} =$$
$$= \sum_{j\geq 1} b_j\mathbb{I}_{B_j}(\omega)\mathbf{E}\{\xi|B_j\} = \sum_{j\geq 1} \eta\mathbf{E}\{\xi|B_j\}\mathbb{I}_{B_j}(\omega),$$

т.е. $\mathbf{E}\{\xi\eta|\mathcal{B}\}$ — это случайная величина, которая на B_j равна $\eta\mathbf{E}\{\xi|B_j\}$ при любом $j \geq 1$. Значит, она равна $\eta\mathbf{E}\{\xi|\mathcal{B}\}$.

6) Функция $\varphi(a) = \mathbf{E}(\xi - a)^2$ минимальна при $a = \mathbf{E}\xi$:

$$\mathbf{E}(\xi - a)^2 = \mathbf{E}(\xi - \mathbf{E}\xi + \mathbf{E}\xi - a)^2 =$$
$$= \mathbf{E}(\xi - \mathbf{E}\xi)^2 + 2(\mathbf{E}\xi - a)\mathbf{E}(\xi - \mathbf{E}\xi) + (\mathbf{E}(\mathbf{E}\xi - a)^2 =$$
$$= \mathbf{E}(\xi - \mathbf{E}\xi)^2 + (\mathbf{E}(\mathbf{E}\xi - a)^2 \geqslant \mathbf{E}(\xi - \mathbf{E}\xi)^2.$$

Аналогичное свойство верно для условных математических ожиданий

Утверждение. *Если ξ — случайная величина, определенная на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$, и $\mathcal{B} \subset \mathcal{F}$ — σ-алгебра, то нижняя грань значений $\mathbf{E}(\xi - a(\omega))^2$ по множеству $\mathcal{B}$-измеримых функций $a(\omega)$ достигается на функции $a(\omega) = \mathbf{E}\{\xi|\mathcal{B}\}$.*

Доказательство. Проведем рассуждения, аналогичные доказательству для безусловных математических ожиданий:

$$\mathbf{E}(\xi - a(\omega))^2 = \mathbf{E}(\xi - \mathbf{E}\{\xi|\mathcal{B}\} + \mathbf{E}\{\xi|\mathcal{B}\} - a(\omega))^2 =$$
$$= \mathbf{E}(\xi - \mathbf{E}\{\xi|\mathcal{B}\})^2 + 2\mathbf{E}(\xi - \mathbf{E}\{\xi|\mathcal{B}\})(\mathbf{E}\{\xi|\mathcal{B}\} - a(\omega)) +$$
$$+ \mathbf{E}(\mathbf{E}\{\xi|\mathcal{B}\} - a(\omega))^2 =$$
$$= \mathbf{E}(\xi - \mathbf{E}\{\xi|\mathcal{B}\})^2 + \mathbf{E}(\mathbf{E}\{\xi|\mathcal{B}\} - a(\omega))^2 \geqslant \mathbf{E}(\xi - \mathbf{E}\{\xi|\mathcal{B}\})^2,$$

так как по свойствам условных математических ожиданий $\mathbf{E}\eta = \mathbf{E}\mathbf{E}\{\eta|\mathcal{B}\}$ и $\mathbf{E}\{(\xi - \mathbf{E}\{\xi|\mathcal{B}\})|\mathcal{B}\} = 0$:

$$\mathbf{E}(\mathbf{E}\{\xi|\mathcal{B}\} - a(\omega))(\xi - \mathbf{E}\{\xi|\mathcal{B}\}) =$$
$$\mathbf{E}\mathbf{E}\{(\mathbf{E}\{\xi|\mathcal{B}\} - a(\omega))(\xi - \mathbf{E}\{\xi|\mathcal{B}\})|\mathcal{B}\} =$$
$$= \mathbf{E}(\mathbf{E}\{\xi|\mathcal{B}\} - a(\omega))\mathbf{E}\{(\xi - \mathbf{E}\{\xi|\mathcal{B}\})|\mathcal{B}\} = 0.$$

В рассмотренном выше случае конечных разбиений определения условных распределений и математических ожиданий конструктивны и довольно наглядны. В общем случае, когда разбиения $\mathcal{B}$ порождаются, например, случайными величинами, имеющими непрерывные распределения, условное математическое ожидание относительно σ-алгебры $\mathcal{B}$ определяется неконструктивно с помощью аналога формулы (31). *Условным математическим ожиданием* случайной величины ξ, определенной на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$, относительно σ-алгебры $\mathcal{B} \subset \mathcal{F}$ называется случайная величина $\mathbf{E}\{\xi|\mathcal{B}\}$, измеримая относительно $\mathcal{B}$ и удовлетворяющая тождеству

$$\int\limits_C \xi d\mathbf{P} = \int\limits_C \mathbf{E}\{\xi|\mathcal{B}\} d\mathbf{P} \quad \text{для любого события } C \in \mathcal{B}. \tag{32}$$

Это равенство аналогично определению плотности абсолютно непрерывного распределения. Так же, как плотность определяется с точностью до значения на множестве меры 0, так и условное математическое ожидание в общем случае определяется с точностью до значений на множествах $\mathbf{P}$-меры 0.

Доказательство существования такой случайной величины нетривиально и основывается на теореме Радона–Никодима.

Теорема Радона–Никодима. *Пусть $(\Omega, \mathcal{F})$ — измеримое пространство, P — σ-конечная мера на $\mathcal{F}$ и λ — абсолютно непрерывная относительно P мера со знаком (т. е. $\lambda(A) = 0$ для любого измеримого множества A с $P(A) = 0$ и $\lambda = \lambda_1 - \lambda_2$, где меры λ_1 и λ_2 неотрицательны и хотя бы одна из них конечна). Тогда существует такая $\mathcal{F}$-измеримая функция $f = f(\omega)\colon \Omega \to [-\infty, \infty]$ (плотность меры λ относительно меры P), что*

$$\lambda(C) = \int\limits_C f(\omega) P(d\omega) \quad \textit{для любого} \quad C \in \mathcal{F}. \tag{33}$$

С точностью до множества P-меры нуль функция f единственна: если h — другая функция, удовлетворяющая (33), то $P\{\omega\colon f(\omega) \neq h(\omega)\} = 0$.

В нашем случае мера $\lambda(C) = \int\limits_C \xi d\mathbf{P}, C \in \mathcal{B}$, а $f(\omega) = \mathbf{E}\{\xi|\mathcal{B}\}$ — плотность меры λ относительно меры $\mathbf{P}$. Теорема Радона–Никодима относится к теории меры; ее доказательство можно найти, например, в учебнике А.А.Боровкова «Теория вероятностей» [1].

Пример. Пусть $\Omega = \{\omega = (\omega_1, \omega_2)\} = [0,\ 1]^2$, $\eta = \eta(\omega) = \omega_1,\ \xi = \xi(\omega) = \omega_1 + \omega_2$. Случайная величина η порождает σ-алгебру $\mathcal{B} = \sigma(\eta) = \{A \times [0,\ 1] : A \subseteq [0,\ 1]$ измеримо$\}$. Найдем условное математическое ожидание $\mathbf{E}\{\xi|\mathcal{B}\}$ случайной величины ξ относительно η. По определению условного математического ожидания для каждого $C = A \times [0,\ 1] \in \mathcal{B}$ должно выполняться равенство

$$\int_C \xi(\omega) P(d\omega) = \int_C \mathbf{E}\{\xi|\mathcal{B}\} P(d\omega),$$

где функция $\mathbf{E}\{\xi|\mathcal{B}\}$ должна зависеть только от η, т.е. только от ω_1. В нашем случае

$$\int_C \xi(\omega) P(d\omega) = \int_A \int_0^1 (\omega_1 + \omega_2) d\omega_2 d\omega_1 = \int_A (\omega_1 + \tfrac{1}{2}) d\omega_1,$$

поэтому $\mathbf{E}\{\xi|\mathcal{B}\} = \omega + 1 + \frac{1}{2} = \eta + \frac{1}{2}$.

§ 5. Условные распределения

Условным математическим ожиданиям соответствуют условные распределения относительно σ-алгебр или порождающих эти σ-алгебры случайных величин. Для дискретных вероятностных пространств условные распределения (условные вероятности) были определены в предыдущем параграфе, и по ним определялись условные математические ожидания.

В общем случае понятие условного математического ожидания определяется с помощью теоремы Радона–Никодима. Вероятностная мера любого измеримого множества A равна математическому ожиданию индикатора этого множества. Таким образом, по условному математическому ожиданию можно построить условную вероятностную меру как совокупность значений условных математических ожиданий индикаторов измеримых множеств.

Пусть ξ и η — случайные величины, определенные на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$, со значениями в $S = \mathbb{R}^s$ и $T = \mathbb{R}^t$ соответственно; в этих пространствах будем рассматривать обычные борелевские σ-алгебры измеримых подмножеств $\mathcal{B}^s$ и $\mathcal{B}^t$, пусть P_ξ и P_η — меры на S и T, порожденные случайными величинами ξ и η.

Определение 1. Функция $P_\xi(B, y)$, $y \in T = \mathbb{R}^t, B \in \mathcal{B}^s$, называется *условным распределением ξ при условии $\eta = y$*, если а) при каждом $B \subset S, B \in \mathcal{B}^s$ для любого $A \subset T, A \in \mathcal{B}^t$,

$$\mathbf{E}_\eta P_\xi(B,\eta)\mathbb{I}\{\eta \in A\} \stackrel{\text{def}}{=} \int_A P_\xi(B,y)P_\eta(dy) = \mathbf{P}\{\xi \in B, \eta \in A\},$$

б) при каждом $y \in T$ функция $P_\xi(\cdot, y)$ задает распределение вероятностей на S.

Замечание. Функция $P_\xi(B,\eta(\omega)) = \mathbf{P}\{\xi \in B \mid \eta\}$ фактически есть условная вероятность события $\{\xi \in B\}$ относительно η.

В силу теоремы Радона–Никодима для каждого $B \in \mathcal{B}^s$ существует такая измеримая функция $g_B(y)$, что $g_B(y) = \mathbf{P}\{\xi \in B|\eta = y\}$. Значения измеримой функции можно произвольно изменять в конечном или счетном множестве точек, не изменяя интегралов от нее (вероятностей событий). Поэтому функция $g_B(y)$ удовлетворяет условию а), но она может не удовлетворять условию б), так как, например, значения $g_B(y)$ можно при разных B «испортить» при одном и том же y, так что функция $g_B(y)$ при этом y не будет вероятностной мерой (например, будет принимать отрицательные значения или не будет аддитивной).

Однако в случае евклидовых пространств, борелевских σ-алгебр и абсолютно непрерывных распределений совокупность функций $g_B(y)$ можно выбрать так, чтобы выполнялись как условие а), так и условие б).

Нетрудно показать, что условное математическое ожидание является математическим ожиданием по условному распределению.

Теорема 1. *Если функция $g(x) : S = \mathbb{R}^s \to \mathbb{R}$ измерима и $\mathbf{E}|g(\xi)| < \infty$, то*

$$\mathbf{E}\{g(\xi)|\eta\} = \int g(x)P(dx|\eta).$$

Доказательство. Если $g(x) = \mathbb{I}_A(x)$ — индикатор измеримого множества A, то левая и правая части равенства равны $\mathbf{P}\{A|\eta\}$. Из аддитивности математического ожидания следует, что утверждение теоремы верно для любой простой функции (принимающей конечное число значений). Для неотрицательных функций утверждение теоремы вытекает из теоремы о монотонной сходимости, а для функций произвольного знака — из аддитивности математического ожидания.

Определение 2. Если распределения случайных величин ξ и η абсолютно непрерывны и условное распределение $P(B|y)$ при каждом $y \in T =$

$\mathbb{R}^t$ абсолютно непрерывно относительно некоторой меры μ на $S = \mathbb{R}^s$:

$$\mathbf{P}\{\xi \in B | \eta = y\} = \int_B f(x|y)\mu(dx),$$

то плотность $f(x|y)$ называется *условной плотностью ξ относительно меры μ при условии $\eta = y$.*

Более подробно это определение можно переформулировать так: измеримая по паре переменных $x \in S, y \in T$ функция $f(x|y)$ есть условная плотность ξ при условии $\eta = y$, если

1) для любых измеримых множеств $A \subset S$, $B \subset T$

$$\int_{y \in A} \int_{x \in B} f(x|y)\mu(dx)\mathbf{P}\{\eta \in dy\} = \mathbf{P}\{\xi \in A,\, \eta \in B\},$$

2) при каждом $y \in T$ функция $f(x|y)$ является плотностью распределения вероятностей.

Из определения и теоремы следует, что если существует условная плотность распределения случайной величины ξ при условии η, то

$$\mathbf{E}\{g(\xi)|\eta\} = \int g(x) f(x|\eta)\mu(dx).$$

Если распределение случайной величины η имеет плотность $q(y)$ относительно некоторой меры λ на $T = \mathbf{R}^t$, то условие 1) можно переписать в виде

$$\int_{y \in A} \int_{x \in B} f(x|y)q(y)\mu(dx)\lambda(dy) = \mathbf{P}\{\xi \in A,\, \eta \in B\}. \tag{34}$$

На пространстве $S \times T = \mathbf{R}^s \times \mathbf{R}^t$ можно задать прямое произведение мер $\mu \times \lambda$; тогда (34) означает, что совместное распределение ξ и η в $S \times T$ имеет плотность относительно $\mu \times \lambda$, равную

$$f(x, y) = f(x|y)q(y).$$

Верно и обратное.

Теорема 2. *Если совместное распределение ξ и η в $S \times T$ имеет плотность $f(x, y)$ относительно $\mu \times \lambda$, то функция*

$$f(x|y) = \frac{f(x,y)}{q(y)}, \quad \textit{где } q(y) = \int f(x, y)\mu(dx),$$

есть условная плотность ξ при условии $\eta = y$, а функция $q(y)$ — плотность η относительно меры λ.

Доказательство. Так как

$$\int_A q(y)\lambda(dy) = \int_A \int f(x,y)\mu(dx)\lambda(dy) = \mathbf{P}\{\eta \in A\},$$

то $q(y)$ — плотность распределения η относительно λ. Далее, $\int f(x|y)\mu(dx) = \frac{1}{q(y)}\int f(x,y)\mu(dy) = 1$, так что $f(x|y)$ — плотность распределения вероятностей, и условие (34), эквивалентное условию 1), тоже выполняется. Значит, функция $f(x|y)$ удовлетворяет всем условиям определения условной плотности и поэтому является условной плотностью.

Замечание. Меняя местами ξ и η в теореме, получаем, что при тех же условиях существует условная плотность

$$q(y|x) = \frac{f(x,y)}{f(x)}, \quad \text{где } f(x) = \int f(x,y)\lambda(dy),$$

случайной величины η при условии $\xi = x$.

§ 6. Многомерное нормальное и связанные с ним распределения

Напомним основные определения.

Случайная величина $\zeta_{0,1}$ имеет стандартное нормальное распределение с параметрами $(0,1)$ (математическим ожиданием 0 и дисперсией 1), если ее плотность распределения равна $\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$, $-\infty < x < \infty$; функция стандартного нормального распределения есть

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du,$$

а его характеристическая функция

$$\psi(t) = \mathbf{E}e^{it\zeta_{0,1}} = e^{-t^2/2}, \quad -\infty < t < \infty.$$

Распределение случайной величины $\zeta_{a,\sigma^2} = a + \sigma\zeta_{0,1}$, являющейся линейным преобразованием $\zeta_{0,1}$, называется нормальным распределением с параметрами (a,σ^2). Это распределение имеет плотность

$$\varphi_{a,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

функцию распределения

$$\Phi_{a,\sigma^2}(x) = \mathbf{P}\{\zeta_{a,\sigma^2} \leqslant x\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(u-a)^2}{2\sigma^2}} du,$$

и характеристическую функцию

$$\psi_{a,\sigma^2}(t) = \mathbf{E} e^{it\zeta_{a,\sigma^2}} = e^{ita} e^{-\frac{\sigma^2 t^2}{2}}, \quad -\infty < t < \infty.$$

Случайный вектор $\overline{\zeta}_{0,E} = (\zeta_1, \dots, \zeta_d)$ имеет стандартное многомерное нормальное распределение в $\mathbb{R}^d$, если его компоненты $\zeta_1, \dots, \zeta_d$ — независимые случайные величины, имеющие стандартное нормальное распределение. Такой случайный вектор имеет нулевой вектор математических ожиданий и единичную матрицу ковариаций:

$$\mathbf{E}\overline{\zeta}_{0,E} = \overline{0}, \quad \mathrm{Cov}(\overline{\zeta}_{0,E}) = \|\mathrm{cov}(\zeta_i, \zeta_j)\| = E = \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix}.$$

В силу независимости компонент плотность и характеристическая функция распределения вектора $\overline{\zeta}_{0,E} = (\zeta_1, \dots, \zeta_d)$ имеют вид

$$\varphi(x_1, \dots, x_d) = \prod_{k=1}^{d} \varphi(x_k) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{x_1^2 + \dots + x_d^2}{2}},$$

$$\psi(t_1, \dots, t_d) = \mathbf{E} e^{i(\overline{t}, \overline{\zeta}_{0,E})} = \mathbf{E} e^{i(t_1\zeta_1 + \dots + t_d\zeta_d)} = \mathbf{E} \prod_{k=1}^{d} e^{it_k\zeta_k} = e^{-\frac{t_1^2 + \dots + t_d^2}{2}}.$$

Плотность $\varphi_{0,E}(\overline{x})$ распределения $\overline{\zeta}_{0,E}$ зависит только от расстояния точки $\overline{x} = (x_1, \dots, x_d)$ от начала координат, т. е. инвариантна относительно ортогональных преобразований $\mathbb{R}^d$, оставляющих точку $\overline{0}$ на месте. Такие распределения называются *сферически симметричными*. Если случайный вектор имеет сферически симметричное распределение, то ортогональное преобразование с неподвижной точкой $\overline{0}$ переводит этот случайный вектор в другой случайный вектор, имеющий то же самое распределение.

Замечание. Согласно определению случайный вектор $\overline{\zeta}_{0,E}$ с $\mathbf{E}\overline{\zeta}_{0,E} = \overline{0}$, $\mathrm{Cov}(\overline{\zeta}_{0,E}) = E$ можно представить в виде $\overline{\zeta}_{0,E} = \zeta_1\overline{e_1} + \dots + \zeta_d\overline{e_d}$, где $\zeta_1, \dots, \zeta_d$ — независимые случайные величины, имеющие нормальное распределение с параметрами $(0, 1)$, а $\overline{e_1}, \dots, \overline{e_d}$ — единичные векторы координатных осей. Из сферической симметричности распределения $\overline{\zeta}_{0,E}$ следует, что для любого другого ортонормированного базиса $\overline{e'_1}, \ldots, \overline{e'_d}$ в соответствующем разложении $\overline{\zeta}_{0,E} = \zeta'_1\overline{e'_1} + \ldots + \zeta'_d\overline{e'_d}$ случайные величины $\zeta'_1, \ldots, \zeta'_d$ тоже независимы и имеют то же самое нормальное распределение с параметрами $(0, 1)$.

Определение. Распределение случайного d-мерного вектора $\overline{\kappa}$ называется *многомерным нормальным*, если его можно представить в виде линейного преобразования случайного вектора $\overline{\zeta}_{\overline{0},E}$ с нулевым математическим ожиданием и единичной матрицей ковариаций: $\overline{\kappa} = B\overline{\zeta}_{\overline{0},E} + \overline{a}$, где $\overline{a} \in \mathbb{R}^d$ — неслучайный вектор и B — $(d \times d)$-матрица.

Замечание 1. Так как множество линейных преобразований образует полугруппу относительно операции суперпозиции, то *любое линейное преобразование случайного вектора, имеющего многомерное нормальное распределение, является многомерным нормальным распределением.* В частности, такое линейное преобразование может понижать размерность, например, переводить многомерный случайный вектор в скалярную линейную комбинацию его компонент.

Если переход от вектора $\overline{\zeta}_{\overline{0},E}$ к вектору $\overline{\kappa}$ осуществляется линейным преобразованием $g(\overline{x}) = B\overline{x} + \overline{a}$, то $\overline{\kappa} = \left(\sum_{k=1}^d b_{1k}\zeta_k, \ldots, \sum_{k=1}^d b_{dk}\zeta_k\right)^\top$,

$$\mathbf{E}\overline{\kappa} = \mathbf{E}g(\overline{\zeta}_{\overline{0},E}) = \mathbf{E}(B\overline{\zeta}_{\overline{0},E} + \overline{a}) = \overline{a}$$

и, так как компоненты вектора $\overline{\zeta}$ независимы и имеют нулевое среднее и единичную дисперсию, то

$$\operatorname{cov}\left(\sum_{k=1}^d b_{ik}\zeta_k, \sum_{m=1}^d b_{jm}\zeta_m\right) = \sum_{k=1}^d\sum_{m=1}^d \operatorname{cov}(b_{ik}\zeta_k, b_{jm}\zeta_m) =$$
$$= \sum_{k=1}^d b_{ik}b_{jk}\mathbf{E}\zeta_k^2 = \sum_{k=1}^d b_{ik}b_{jk} = (BB^\top)_{ij}$$

т. е.

$$\operatorname{Cov}(\overline{\kappa}) = B\operatorname{Cov}(\overline{\zeta}_{\overline{0},E})B^\top = BB^\top.$$

В частности, отсюда следует, что $\det\operatorname{Cov}(\overline{\kappa}) = (\det B)^2 \geqslant 0$. Если матрица B вырождена, то распределение вектора $\overline{\kappa}$ вырождено и его носителем является гиперплоскость размерности меньше d.

По формуле для преобразования плотности при взаимно однозначном линейном отображении с невырожденной $d \times d$-матрицей B плотность распределения вектора $\overline{\kappa}$ имеет вид

$$\varphi_{\overline{\kappa}}(\overline{x}) = \frac{\varphi_{\overline{\zeta}_{\overline{0},I}}(B^{-1}(\overline{x}-\overline{a}))}{|\det B|} =$$
$$= \frac{1}{\sqrt{(2\pi)^d \det C(\overline{\zeta})}} \exp\left\{-\frac{1}{2}(\overline{x}-\overline{a})C^{-1}(\overline{\kappa})(\overline{x}-\overline{a})^\top\right\}. \qquad (35)$$

Так как плотность $\varphi_{\overline{\kappa}}(\overline{x})$ выражается через $\overline{a} = \mathbf{E}\zeta$ и $\mathrm{Cov}(\overline{\zeta})$, то *многомерное нормальное распределение однозначно определяется вектором математических ожиданий и матрицей ковариаций.*

Если случайный вектор $\kappa = (\kappa_1, \ldots, \kappa_n)$ имеет многомерное нормальное распределение и все его компоненты $\kappa_1, \ldots, \kappa_n$ независимы, то ковариация любых двух его компонент κ_i, κ_j, $i \neq j$, равна 0, значит, ковариационная матрица вектора κ — диагональная.

Обратно, *если случайный вектор κ имеет многомерное нормальное распределение с некоррелироваными компонентами, то его матрица ковариаций диагональна и (в силу однозначности определения нормального распределения вектором математических ожиданий и матрицей ковариаций) его компоненты — независимые случайные величины.*

Характеристическая функция случайного вектора $\overline{\kappa} = B\overline{\zeta}_{\overline{0},E} + \overline{a}$, имеющего многомерное нормальное распределение с плотностью (35), равна

$$\psi_{\overline{\kappa}}(\overline{t}) = \mathbf{E}e^{i(\overline{t},\overline{\kappa})} = \mathbf{E}e^{i(\overline{t},B\overline{\zeta}_{\overline{0},E}+\overline{a})} = e^{i(\overline{t},\overline{a})}\mathbf{E}e^{i(\overline{t},B\overline{\zeta}_{\overline{0},E})} = e^{i(\overline{t},\overline{a})}\mathbf{E}e^{i(B^\top\overline{t},\overline{\zeta}_{\overline{0},E})} =$$
$$= e^{i(\overline{t},\overline{a})}e^{-\frac{1}{2}(B^\top\overline{t},B^\top\overline{t})} = e^{i(\overline{t},\overline{a})}e^{-\frac{1}{2}\overline{t}^\top BB^\top\overline{t}} = e^{i(\overline{t},\overline{a})}e^{-t\mathrm{Cov}(\overline{\kappa})\overline{t}^\top/2}, \quad \overline{t} \in \mathbb{R}^d.$$

Для случайного вектора с математическим ожиданием a и матрицей ковариаций C мы будем использовать обозначение $\zeta_{a,C}$.

Центральная предельная теорема. *Пусть $\overline{\xi^{(n)}} = (\xi_1^{(n)}, \ldots, \xi_d^{(n)})$, $n = 1, 2, \ldots,$ — независимые одинаково распределенные случайные векторы с $\mathbf{E}\overline{\xi^{(1)}} = \overline{a} \in \mathbb{R}^d$ и матрицей ковариаций $C(\overline{\xi})$; пусть $\overline{\zeta_n} = \overline{\xi^{(1)}} + \ldots + \overline{\xi^{(n)}}$. Последовательность случайных векторов*

$$\overline{\zeta_n'} = \frac{(\overline{\zeta_n} - n\overline{a})}{\sqrt{n}}$$

слабо сходится к случайному вектору $\overline{\zeta}$, который имеет многомерное нормальное распределение с $\mathbf{E}\overline{\zeta} = \overline{0}$, $C(\overline{\zeta}) = C(\overline{\xi})$.

Лемма. *Пусть случайный вектор* $\overline{\xi} = (\xi_1, \ldots, \xi_d)$ *имеет конечные математическое ожидание* $\overline{a} = (a_1, \ldots, a_d) = (\mathbf{E}\xi_1, \ldots, \mathbf{E}\xi_d)$ *и матрицу ковариаций* $C(\overline{\xi})$. *Тогда при* $|\overline{t}| = |t_1| + \ldots + |t_d| \to 0$

$$\mathbf{E}e^{i(\overline{t},\overline{\xi})} = e^{i(\overline{t},\overline{a})} \left(1 - \frac{1}{2}\,\overline{t}C(\overline{\xi})\overline{t}^{\top} + o(|\overline{t}|^2)\right).$$

Пример. Пусть случайный вектор (ξ_1, ξ_2) имеет двумерное нормальное распределение с вектором средних $\mathbf{a} = (a_1, a_2)$, где $\mathbf{E}\xi_1 = a_1$, $\mathbf{E}\xi_2 = a_2$, и ковариациями $\sigma_{ij} = \mathbf{E}(\xi_i - a_i)(\xi_j - a_j)$, $i, j = 1, 2$. Пусть $\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$ — коэффициент корреляции между ξ_1 и ξ_2. Если $|\rho| \neq 1$, то матрица $\Sigma = \|\sigma_{ij}\|$ вторых моментов не вырождена $|\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1 - \rho^2)$ и

$$\Sigma^{-1} = \frac{1}{|\Sigma|} \begin{Vmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{Vmatrix} = \frac{1}{1-\rho^2} \begin{Vmatrix} \frac{1}{\sigma_{11}} & -\frac{\rho}{\sqrt{\sigma_{11}\sigma_{12}}} \\ -\frac{\rho}{\sqrt{\sigma_{11}\sigma_{12}}} & \frac{1}{\sigma_{22}} \end{Vmatrix}.$$

Плотность совместного распределения ξ_1, ξ_2 равна

$$\varphi_{a,\Sigma}(x, y) = \frac{1}{2\pi\sigma_{11}\sigma_{22}} \times \tag{36}$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-a_1)^2}{\sigma_{11}} - \frac{2\rho(x-a_1)(y-a_2)}{\sqrt{\sigma_{11}\sigma_{22}}} + \frac{(y-a_2)^2}{\sigma_{22}}\right]\right\}. \tag{37}$$

Одномерные плотности ξ_1 и ξ_2 равны

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_{11}}} \exp\left\{-\frac{(x-a_1)^2}{\sigma_{11}}\right\}, \quad q(y) = \frac{1}{\sqrt{2\pi\sigma_{22}}} \exp\left\{-\frac{(y-a_2)^2}{\sigma_{22}}\right\}.$$

Следовательно, условная плотность ξ_1 при условии $\xi_2 = y$ равна

$$f(x|y) = \frac{f(x, y)}{q(y)} =$$

$$= \frac{1}{\sqrt{2\pi\sigma_{11}(1-\rho^2)}} \exp\left\{-\frac{1}{2\sigma_{11}(1-\rho^2)}\left(x - a_1 - \rho\sqrt{\tfrac{\sigma_{11}}{\sigma_{22}}}\,(y - a_2)\right)^2\right\}.$$

Это плотность нормального распределения со средним $a_1 + \rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\,(y - a_2)$ и дисперсией $\sigma_{11}(1 - \rho^2)$:

$$\mathbf{E}\{\xi_1|\xi_2\} = a_1 + \rho\sqrt{\tfrac{\sigma_{11}}{\sigma_{22}}}\,(\xi_2 - a_2).$$

Прямая $x = a_1 + \rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\,(y - a_2)$ называется *линией регрессии* ξ_1 *на* ξ_2. Она представляет собой график наилучшего среднеквадратического приближения ξ_1 при заданном значении $\xi_2 = y$.

§ 7. Семейство распределений χ^2

Если ζ имеет стандартное нормальное распределение, то случайная величина ζ^2 имеет функцию распределения

$$\mathbf{P}\{\zeta^2 \leqslant x\} = \mathbf{P}\{|\zeta| \leqslant x\} = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}), \quad x \geqslant 0,$$

и плотность

$$p_1(x) = 2\frac{\varphi(\sqrt{x})}{2\sqrt{x}} = \frac{x^{-1/2}}{\sqrt{2\pi}}\, e^{-x/2}, \quad x \geqslant 0. \tag{38}$$

Если ζ_1, ζ_2 независимы и имеют стандартные нормальные распределения, т. е. вектор (ζ_1, ζ_2) имеет двумерное сферически симметричное нормальное распределение, то сумма $\zeta_1^2 + \zeta_2^2$ имеет функцию распределения

$$\mathbf{P}\{\zeta_1^2 + \zeta_2^2 \leqslant x\} = \int_{u^2+v^2 \leqslant x} \frac{1}{2\pi} e^{-(u^2+v^2)/2} du dv = \int_0^{\sqrt{x}} \int_0^{2\pi} \frac{1}{2\pi} e^{-r^2/2} r\, d\varphi dr =$$

$$= \int_0^{\sqrt{x}} r e^{-r^2/2} dr = \int_0^{x/2} e^{-u} du = 1 - e^{-x/2}, \quad x \geqslant 0,$$

и плотность

$$p_2(x) = \frac{1}{2}\, e^{-x/2}, \quad x \geqslant 0. \tag{39}$$

Из формул (38), (39) и формулы свертки по индукции выводится формула

$$p_{2k+1}(x) = \frac{x^{k-\frac{1}{2}}}{2^{k+\frac{1}{2}}\Gamma\left(k+\frac{1}{2}\right)}\, e^{-x/2}, \quad x \geqslant 0,$$

для плотности распределения суммы нечетного числа квадратов независимых случайных величин $\zeta_1, \ldots, \zeta_{2k+1}$ ($k \geqslant 1$):

$$p_{2k+1}(x) = \int_0^x p_{2k-1}(u) p_2(x-u)\, du = \int_0^x \frac{u^{k-\frac{3}{2}}}{2^{k-\frac{1}{2}}\Gamma\left(k-\frac{1}{2}\right)}\, e^{-u/2}\, \frac{e^{-(x-u)/2}}{2}\, du =$$

$$= \frac{x^{k-\frac{1}{2}}}{2^{k+\frac{1}{2}}\Gamma\left(k+\frac{1}{2}\right)}\, e^{-x/2}, \quad x \geqslant 0, \tag{40}$$

в записи этой формулы использована *гамма-функция Эйлера* $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$, $\alpha > 0$, обладающая следующими свойствами:

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha), \quad \Gamma(n+1) = n!\, (n = 0, 1, \ldots), \quad \Gamma(\tfrac{1}{2}) = \sqrt{2\pi}.$$

Аналогично из (39) и формулы свертки выводится формула

$$p_{2k}(x) = \frac{x^{k-1}}{2^k \Gamma(k)}\, e^{-x/2}, \quad x \geqslant 0.$$

для плотности распределения суммы четного числа квадратов независимых случайных величин $\zeta_1, \ldots, \zeta_{2k}$ $(k \geqslant 2)$:

$$p_{2k}(x) = \int_0^x p_{2k-2}(u)p_2(x-u)\,du = \int_0^x \frac{u^{k-2}}{2^{k-1}\Gamma(k-1)}\,e^{-u/2}\,\frac{e^{-(x-u)/2}}{2}\,du = \\ = \frac{x^{k-1}}{2^k\Gamma(k)}\,e^{-x/2}, \quad x \geqslant 0. \tag{41}$$

Формулы (40) и (41) приводятся к общему виду.

Определение. Распределение суммы квадратов n независимых случайных величин $\xi_1, \ldots, \xi_n$, имеющих стандартное нормальное распределение, называется *распределением хи-квадрат с n степенями свободы* и имеет плотность

$$p_n(x) = \frac{x^{\frac{n}{2}-1}}{2^{n/2}\Gamma(\frac{n}{2})}\,e^{-x/2}, \quad x \in [0, \infty).$$

Пользуясь тем, что $\mathbf{E}\xi_k^2 = 1$, $\mathbf{E}\xi_k^4 = 3$, $\mathbf{D}\xi_k^2 = 2$, нетрудно проверить, что

$$\mathbf{E}(\xi_1^2 + \ldots + \xi_n^2) = n, \quad \mathbf{D}(\xi_1^2 + \ldots + \xi_n^2) = 2n. \tag{42}$$

Лемма о проекции нормального распределения. *Если $\overline{\xi}$ — случайный вектор, имеющий сферически симметричное нормальное распределение в $\mathbb{R}^N$ с нулевым математическим ожиданием и единичной матрицей ковариаций E, а $\overline{c} = (c_1, \ldots, c_N) \in \mathbb{R}^N$, $|\overline{c}| = 1$, то проекция $\overline{\xi}^{(\overline{c})}$ вектора $\overline{\xi}$ на гиперплоскость $L_{\overline{c}} = \{\overline{x} \in \mathbb{R}^N : (\overline{x}, \overline{c}) = 0\}$, ортогональную вектору $\overline{c}$, имеет нормальное распределение с математическим ожиданием 0 и матрицей ковариаций*

$$C(\overline{\xi}^{(\overline{c})}) = E - \|c_i c_j\|_{i,j=1}^N;$$

квадрат длины вектора $\overline{\xi}^{(\overline{c})}$ имеет распределение χ^2 с $N-1$ степенями свободы.

Доказательство. Рассмотрим наряду с исходным базисом $\overline{e_1}, \ldots, \overline{e_N}$ в $\mathbb{R}^N$ ортонормированный базис $\overline{e'_1}, \ldots, \overline{e'_N}$, в котором $\overline{e'_N} = \overline{c}$. Вектор $\overline{\xi}$ можно разложить как по исходному, так и по новому базисам:

$$\overline{\xi} = \xi_1\overline{e_1} + \ldots + \xi_N\overline{e_N} = \xi'_1\overline{e'_1} + \ldots + \xi'_N\overline{e'_N}. \tag{43}$$

Так как по условию $\overline{\xi}$ имеет сферически симметричное нормальное распределение с единичной матрицей ковариаций, то координаты $(\xi_1, \ldots, \xi_N)$

независимы и имеют стандартное нормальное распределение и координаты $\xi'_1, \ldots, \xi'_N$ тоже независимы и имеют стандартное нормальное распределение.

Проекцию вектора $\overline{\xi}$ на гиперплоскость $L_{\overline{c}}$ можно представить в виде $\overline{\xi}^{(\overline{c})} = \xi'_1 \overline{e'_1} + \ldots + \xi'_{N-1} \overline{e'_{N-1}}$. Единичные векторы $\overline{e'_1}, \ldots, \overline{e_{N-1}}'$ ортогональны, значит, $|\overline{\xi}^{(\overline{c})}|^2 = (\xi'_1)^2 + \ldots + (\xi'_{N-1})^2$ имеет распределение χ^2 с $N-1$ степенями свободы.

Далее, в силу (43)

$$\overline{\xi} = \overline{\xi}^{(\overline{c})} + \xi'_N \overline{e'_N},$$

и векторы $\overline{\xi}^{(\overline{c})}$ и $\xi'_N \overline{e'_N}$ независимы (и ортогональны).

Очевидно, $\mathbf{E}\overline{\xi}^{(\overline{c})} = \overline{0}$. Найдем ковариационную матрицу $\overline{\xi}^{(\overline{c})}$ в исходном базисе. Так как ковариационная матрица суммы независимых случайных векторов с конечными вторыми моментами равна сумме ковариационных матриц этих векторов, то

$$C(\overline{\xi}) = E = C(\overline{\xi}^{(\overline{c})} + \xi'_N \overline{e'_N}) = C(\overline{\xi}^{(\overline{c})}) + C(\xi'_N \overline{e'_N})$$

и

$$C(\xi'_N \overline{e'_N}) = C(\xi'_N \overline{c}) = C((\xi'_N c_1, \ldots, \xi'_N c_N)) = \|\mathrm{Cov}(\xi'_N c_i, \xi'_N c_j)\| = \|c_i c_j\|,$$

т. е.

$$C(\overline{\xi}^{(\overline{c})}) = E - \|c_i c_j\|. \tag{44}$$

Лемма доказана.

Пусть проводится n независимых испытаний с исходами $1, \ldots, N$ и вероятностями исходов $p_1, \ldots, p_N$, и в векторе $\overline{\nu^{(n)}} = (\nu_1^{(n)}, \ldots, \nu_N^{(n)})$ координата $\nu_k^{(n)}$, $k = 1, \ldots, N$, равна числу появлений исхода k в n испытаниях. Распределение вектора $\overline{\nu^{(n)}}$ называют *полиномиальным распределением* с параметрами n и $\overline{p} = (p_1, \ldots, p_N)$:

$$\mathbf{P}\{\nu_1^{(n)} = n_1, \ldots, \nu_N^{(n)} = n_N\} = \begin{cases} \frac{n!}{n_1! \ldots n_N!} p_1^{n_1} \ldots p_N^{n_N}, & \text{если } \sum_{j=1}^N n_j = n, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Утверждение. *Пусть случайный вектор $\overline{\nu^{(n)}} = (\nu_1^{(n)}, \ldots, \nu_N^{(n)})$ имеет полиномиальное распределение с параметрами n и $\overline{p} = (p_1, \ldots, p_N)$.*

Тогда распределение вектора $\frac{1}{\sqrt{n}}(\overline{\nu^{(n)}} - n\overline{p})$ *при* $n \to \infty$ *сходится к нормальному распределению с нулевым средним и матрицей ковариаций* $\mathrm{diag}(p_1, \ldots, p_N) - \|p_i p_j\|_{i,j=1}^N$.

Доказательство. Случайный вектор $\overline{\nu^{(n)}}$ имеет такое же распределение, как вектор чисел появлений исходов в n независимых испытаниях с N исходами, имеющими вероятности $(p_1, \ldots, p_N)$. Поэтому $\overline{\nu^{(n)}}$ можно представить в виде суммы n независимых одинаково распределенных случайных векторов: $\overline{\nu^{(n)}} = \overline{\gamma^{(1)}} + \ldots + \overline{\gamma^{(n)}}$, где вектор $\overline{\gamma^{(k)}} = (\gamma_1^{(k)}, \ldots, \gamma_N^{(k)})$ описывает исход k-го испытания: $\gamma_j^{(k)} = 1$, если в k-м испытании появился исход j, и $\gamma_j^{(k)} = 0$ в остальных случаях. Тогда

$$\mathbf{E}\overline{\gamma^{(k)}} = \sum_{j=1}^N p_j \overline{e_j} = (p_1, \ldots, p_N),$$

$$\mathbf{D}\gamma_j^{(k)} = \mathbf{E}(\gamma_j^{(k)} - \mathbf{E}\gamma_j^{(k)})^2 = p_j(1-p_j)^2 + (1-p_j)(p_j)^2 = p_j(1-p_j),$$

$$\mathrm{Cov}(\gamma_i^{(k)}, \gamma_j^{(k)}) = \mathbf{E}\gamma_i^{(k)}\gamma_j^{(k)} - \mathbf{E}\gamma_i^{(k)}\mathbf{E}\gamma_j^{(k)} = -p_i p_j, \quad i \neq j. \tag{45}$$

Таким образом, матрицу ковариаций единичного случайного вектора $\overline{\gamma^{(k)}}$ можно представить в виде

$$C(\overline{\gamma^{(k)}}) = \mathrm{diag}(p_1, \ldots, p_N) - \|p_i p_j\|_{i,j=1}^N. \tag{46}$$

Отсюда и из центральной предельной теоремы вытекает утверждение следствия.

Отметим, что ковариационная матрица $\mathrm{diag}(p_1, \ldots, p_N) - \|p_i p_j\|_{i,j=1}^N$ — вырожденная: сумма элементов в любой ее строке равна 0. Вырожденность ковариационной матрицы означает, что распределение сосредоточено на какой-то гиперплоскости. Действительно сумма компонент вектора $\overline{\nu^{(n)}}$ равна n (общему числу исходов, появившихся в n испытаниях), т. е. вектор $\overline{\nu^{(n)}} - np$ с вероятностью 1 лежит в гиперплоскости

$$x_1 + \ldots + x_N = 0.$$

Многомерное нормальное распределение с вырожденной ненулевой ковариационной матрицей является сингулярным и не имеет плотности; если же ковариационная матрица нулевая, то распределение сосредоточено в одной точке.

Пусть случайный вектор $\overline{\nu^{(n)}} = (\nu_1^{(n)}, \ldots, \nu_N^{(n)})$ имеет полиномиальное распределение с параметрами n и $\overline{p} = (p_1, \ldots, p_N)$. Случайная величина

$$X_n^2 = \sum_{j=1}^{N} \frac{(\nu_j^{(n)} - np_j)^2}{np_j}$$

называется *статистикой Пирсона.*

Теорема о предельном распределении статистики Пирсона. *Пусть случайный вектор $\overline{\nu^{(n)}} = (\nu_1^{(n)}, \ldots, \nu_N^{(n)})$ имеет полиномиальное распределение с параметрами n и $\overline{p} = (p_1, \ldots, p_N)$. Если вектор $\overline{p} = (p_1, \ldots, p_N)$ фиксирован, а $n \to \infty$, то распределение статистики Пирсона X_N^2 сходится к распределению χ^2 с $N-1$ степенями свободы.*

Доказательство. Согласно следствию из многомерной центральной предельной теоремы распределение вектора

$$\left(\frac{\nu_1^{(n)} - np_1}{\sqrt{n}}, \ldots, \frac{\nu_N^{(n)} - np_N}{\sqrt{n}}\right) \tag{47}$$

при $n \to \infty$ сходится к нормальному распределению с нулевым средним и матрицей ковариаций $\operatorname{diag}(p_1, \ldots, p_N) - \|p_i p_j\|_{i,j=1}^N$, где δ_{ij} — символ Кронекера. Статистику Пирсона можно интерпретировать как квадрат длины вектора

$$\left(\frac{\nu_1^{(n)} - np_1}{\sqrt{np_1}}, \ldots, \frac{\nu_N^{(n)} - np_N}{\sqrt{np_N}}\right). \tag{48}$$

Он получается из вектора (47) линейным преобразованием: делением j-й ($j = 1, \ldots, N$) координаты на $\sqrt{p_j}$. При этом математическое ожидание каждой координаты (47) остается нулевым, и при любых $i, j = 1, \ldots, N$

$$\operatorname{Cov}\left(\frac{\nu_i^{(n)} - np_i}{\sqrt{np_i}}, \frac{\nu_j^{(n)} - np_j}{\sqrt{np_j}}\right) = \frac{1}{\sqrt{p_i p_j}} \operatorname{Cov}\left(\frac{\nu_i^{(n)} - np_i}{\sqrt{n}}, \frac{\nu_j^{(n)} - np_j}{\sqrt{n}}\right) = \begin{cases} 1 - p_i, & \text{если } i = j, \\ -\sqrt{p_i p_j}, & \text{если } i \neq j. \end{cases}$$

Поэтому при $n \to \infty$ распределение векторов (48) сходится к нормальному распределению с нулевым вектором средних и матрицей ковариаций $E - \|\sqrt{p_i p_j}\|$. Согласно лемме о проекции нормального распределения распределение квадрата длины случайного вектора с таким распределением имеет распределение χ^2 с $N-1$ степенями свободы. Поскольку квадрат длины — непрерывная функция, распределение ее значений от допредельных векторов (48) сходится к распределению ее значения от предельного случайного вектора. Теорема доказана.

Утверждение. *Если случайные величины $\xi_1, \ldots, \xi_n$ независимы и $\xi_k \sim N(a_k, 1), k = 1, \ldots, n$, то распределение суммы $\xi_1^2 + \ldots + \xi_n^2$ зависит только от n и от $a_1^2 + \ldots + a_k^2$.*

Доказательство. Вектор $\xi = (\xi_1, \ldots, \xi_n)$ имеет n-мерное нормальное распределение с математическим ожиданием $\mathbf{a} = (a_1, \ldots, a_n)$ и единичной матрицей ковариаций E. Ортогональное линейное преобразование $\mathbb{R}^n$, которое переводит вектор $\mathbf{a}$ в вектор $\mathbf{a}' = \left(0, 0, \ldots, 0, \sqrt{\sum_{k=1}^n a_k^2}\right) = (0, 0, \ldots, 0, |\mathbf{a}|)$, переводит случайный вектор ξ в случайный вектор $\xi' = (\xi_1', \ldots, \xi_n')$, имеющий n-мерное нормальное распределение с математическом ожиданием $\mathbf{a}'$ и единичной матрицей ковариаций. Поэтому случайные величины $\xi_1', \ldots, \xi_{n-1}' \sim N(0, 1), \xi_n' \sim N(|\mathbf{a}|, 1)$ независимы. В силу ортогональности преобразования существует детерминированная связь между случайными величинами:

$$(\xi_1')^2 + \ldots + (\xi_n')^2 = \xi_1^2 + \ldots + \xi_n^2. \tag{49}$$

Но $(\xi_1')^2 + \ldots + (\xi_{n-1}')^2 \sim \chi_{n-1}^2$, $(\xi_n')^2$ — квадрат случайной величины, имеющей нормальное распределение $N(|\mathbf{a}|, 1)$, и эти случайные величины независимы. Таким образом, распределение их суммы зависит только от n и от $|\mathbf{a}| = \sqrt{\sum_{k=1}^n a_k^2}$. Утверждение доказано.

Это утверждение обосновывает корректность следующего определения.

Определение 24.2. Распределение суммы $\xi_1^2 + \ldots + \xi_n^2$, где $\xi_1^2, \ldots, \xi_n^2$ независимы и $\xi_k \sim N(a_k, 1), k = 1, \ldots, n$, называется *нецентральным распределением хи-квадрат с n степенями свободы и параметром нецентральности $|\mathbf{a}|^2 = a_1^2 + \ldots + a_n^2$.*

Если параметр нецентральности равен 0, то нецентральное распределение хи-квадрат совпадает с обычным распределением хи-квадрат.

Легко проверить, что если случайные величины $\eta_1, \ldots, \eta_k$ независимы и имеют нецентральные распределения хи-квадрат с параметрами $(n_1, b_1^2), \ldots, (n_k, b_k^2)$ соответственно, то сумма $\eta_1 + \ldots + \eta_k$ имеет нецентральное распределение хи-квадрат с параметрами $\left(\sum_{j=1}^k n_j, \sum_{j=1}^k b_j^2\right)$.

Замечание. Несложно найти первые два момента нецентрального распределения хи-квадрат с n степенями свободы и параметром нецентральности a^2. Если $\zeta_{a,\sigma^2} \sim N(a,\sigma^2)$, то $\zeta_{a,\sigma^2} = \sigma\zeta_{0,1} + a$ и

$$\begin{gathered}
\mathbf{E}\zeta_{a,\sigma^2} = a, \mathbf{E}\zeta_{a,\sigma^2}^2 = \mathbf{D}\zeta_{a,\sigma^2} + (\mathbf{E}\zeta_{a,\sigma^2})^2 = \sigma^2 + a^2, \\
\mathbf{E}\zeta_{a,\sigma^2}^4 = \mathbf{E}(\zeta_{0,1} + a)^4 = \mathbf{E}\zeta_{0,1}^4 + 4a\mathbf{E}\zeta_{0,1}^3 + 6a^2\mathbf{E}\zeta_{0,1}^2 + 4a^3\mathbf{E}\zeta_{0,1} + a^4 = \\
= 3\sigma^4 + 6a^2\sigma^2 + a^4, \\
\mathbf{D}(\zeta_{a,\sigma^2})^2 = \mathbf{E}\zeta_{a,\sigma^2}^4 - (\mathbf{E}\zeta_{a,\sigma^2}^2)^2 = 2\sigma^4 + 4a^2\sigma^2.
\end{gathered}$$

Из этих равенств и из (49) следует, что если $\xi_k \sim N(a_k, 1), k = 1, \ldots, n$, независимы и $|\mathbf{a}|^2 = a_1^2 + \ldots + a_n^2$, то

$$\begin{gathered}
\mathbf{E}(\xi_1^2 + \ldots + \xi_n^2) = n + |\mathbf{a}|^2, \\
\mathbf{D}(\xi_1^2 + \ldots + \xi_n^2) = 2(n-1) + 2 + 4|\mathbf{a}|^2 = 2n + 4|\mathbf{a}|^2 = 2(n + 2|\mathbf{a}|^2).
\end{gathered}$$

Сравнение этих формул с (42) показывает, что отличие нецентрального распределения хи-квадрат от центрального увеличивается с увеличением параметра нецентральности.

В следующей теореме показано, что нецентральное распределение хи-квадрат является предельным для статистики Пирсона, построенных по частотам исходов в некоторых схемах серий независимых испытаний с конечным числом исходов.

Теорема. (Условия сходимости распределения статистики Пирсона к нецентральному распределению хи-квадрат.) *Пусть* $p_1, \ldots, p_n > 0$, $a_1, \ldots, a_n$ *фиксированы*, $\sum_{k=1}^n p_k = 1, \sum_{k=1}^n a_k = 0$ *и для любого натурального* T *случайный вектор* $\nu^{(T)} = (\nu_1^{(T)}, \ldots, \nu_n^{(T)})$ *имеет полиномиальное распределение* $\mathrm{Poly}\left(T; p_1 + \frac{a_1}{\sqrt{T}}, \ldots, p_n + \frac{a_n}{\sqrt{T}}\right)$. *Тогда при* $T \to \infty$ *распределение статистики*

$$\sum\nolimits_{k=1}^n \frac{(\nu_k^{(T)} - Tp_k)^2}{Tp_k}$$

сходится к нецентральному распределению хи-квадрат с $n-1$ *степенями свободы и параметром нецентральности* $\lambda^2 = \sum_{k=1}^n \frac{a_k^2}{p_k}$.

Доказательство. Чтобы упростить формулы, будем использовать обозначения $\mathbf{p} = (p_1, \ldots, p_n)$, $\mathbf{a} = (a_1, \ldots, a_n)$. Так как вектор

$$\begin{gathered}
\tfrac{1}{\sqrt{T}} \left(\nu^{(T)} - T\left(\mathbf{p} + \tfrac{\mathbf{a}}{\sqrt{T}}\right)\right) = \\
= \left(\tfrac{1}{\sqrt{T}} \left(\nu_1^{(T)} - T\left(p_1 + \tfrac{a_1}{\sqrt{T}}\right)\right), \ldots, \tfrac{1}{\sqrt{T}} \left(\nu_n^{(T)} - T\left(p_n + \tfrac{a_n}{\sqrt{T}}\right)\right)\right)
\end{gathered}$$

имеет центрированное и нормированное полиномиальное распределение, то его математическое ожидание равно 0, а матрица ковариаций имеет вид

$$\operatorname{diag}\left(\mathbf{p}+\frac{\mathbf{a}}{\sqrt{T}}\right)-\left\|\left(p_i+\frac{a_i}{\sqrt{T}}\right)\left(p_j+\frac{a_j}{\sqrt{T}}\right)\right\|=\operatorname{diag}(\mathbf{p})-\|p_ip_j\|+O\left(\frac{1}{\sqrt{T}}\right).$$

Доказательство использует лемму, громоздкое доказательство которой проведем чуть позже.

Лемма. *При выполнении условий теоремы и $T\to\infty$ распределения векторов*

$$\frac{1}{\sqrt{T}}\left(\nu^{(T)}-T\left(\mathbf{p}+\frac{\mathbf{a}}{\sqrt{T}}\right)\right)$$

сходятся к многомерному нормальному распределению с нулевым средним и матрицей ковариаций $D=\operatorname{diag}(\mathbf{p})-\|p_ip_j\|$.

Так как

$$\frac{\nu_1^{(T)}-T\left(p_k+\frac{a_k}{\sqrt{T}}\right)}{\sqrt{T}}=\frac{\nu_1^{(T)}-Tp_k}{\sqrt{T}}-a_k,\quad k=1,\ldots,n,$$

то в силу леммы 24.1 при $T\to\infty$ распределение вектора

$$\frac{1}{\sqrt{T}}(\nu^{(T)}-T\mathbf{p})=\left(\frac{\nu_1^{(T)}-Tp_1}{\sqrt{T}},\ldots,\frac{\nu_n^{(T)}-Tp_n}{\sqrt{T}}\right)$$

сходится к нормальному распределению $N(\mathbf{a},D)$. Распределения векторов $\nu^{(T)}-T\mathbf{p}$ и распределение $N(\mathbf{a},D)$ являются вырожденными в силу того, что $\sum_{k=1}^n\nu_n^{(T)}=\sum_{k=1}^nTp_n=T$, сосредоточены на гиперплоскости $\sum_{k=1}^nx_k=0$. Следовательно, при $T\to\infty$ распределение вектора

$$\left(\frac{\nu_1^{(T)}-Tp_1}{\sqrt{Tp_1}},\ldots,\frac{\nu_n^{(T)}-Tp_n}{\sqrt{Tp_n}}\right)$$

тоже сосредоточено на гиперплоскости, проходящей через 0, и при $T\to\infty$ сходится к распределению $N\left(\left(\frac{a_1}{\sqrt{p_1}},\ldots,\frac{a_n}{\sqrt{p_n}}\right),E-\left\|\sqrt{p_ip_j}\right\|\right)$.

При доказательстве теоремы о сходимости распределения статистики Пирсона к распределению хи-квадрат было показано, что нормальное распределение с нулевым средним и матрицей ковариаций $E-\left\|\sqrt{p_ip_j}\right\|$ является вырожденным распределением и что ортогональным преобразованием его можно перевести в нормальное распределение, у которого $n-1$ координат независимы и имеют стандартные нормальные распределения, а n-я координата тождественно равна нулю. В нашем случае вектор средних может быть ненулевым, но он лежит в $(n-1)$-мерной гиперплоскости, содержащей носитель распределения. Поэтому после такого же ортогонального преобразования мы получим нормальное распределение, у которого $n-1$ координат независимы и имеют нормальные распределения с единичными дисперсиями, а n-я координата тождественно равна нулю. Отсюда и из леммы о нецентральном распределении хи-квадрат следует утверждение теоремы.

Доказательство леммы 6.4.1. Случайный вектор $\nu^{(T)}$, имеющий полиномиальное распределение $\mathrm{Poly}(T; q_1, \ldots, q_n)$, можно представить в виде суммы T независимых одинаково распределенных единичных векторов, принимающих значения в множестве $\mathbf{e}_1 = (1, 0, \ldots, 0), \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots, \mathbf{e}_n = (0, 0, \ldots, 0, 1)$:

$$\nu^{(T)} = \sum_{t=1}^{T} \eta_t, \quad \mathbf{P}\{\eta_t = \mathbf{e}_k\} = p_k + \tfrac{a_k}{\sqrt{T}},\ k = 1, \ldots, n.$$

Тогда

$$\tfrac{1}{\sqrt{T}} \left(\nu^{(T)} - T\left(\mathbf{p} + \tfrac{\mathbf{a}}{\sqrt{T}}\right)\right) = \tfrac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\eta_t - \left(\mathbf{p} + \tfrac{\mathbf{a}}{\sqrt{T}}\right)\right).$$

Для доказательства леммы достаточно показать, что характеристические функции этих векторов

$$f_\nu(\mathbf{u}) \stackrel{\mathrm{def}}{=} \mathbf{E} \exp\left\{i\left(\mathbf{u}, \tfrac{1}{\sqrt{T}} \left(\nu^{(T)} - T\left(\mathbf{p} + \tfrac{\mathbf{a}}{\sqrt{T}}\right)\right)\right)\right\}$$

при $T \to \infty$ сходятся к характеристической функции нормального закона с нулевым средним и матрицей ковариаций $D = \mathrm{diag}(\mathbf{p}) - \|p_i p_j\|$, т. е. к

$$\exp\left\{-\frac{1}{2}\, \mathbf{u} D \mathbf{u}^{\mathsf{T}}\right\} = \exp\left\{-\frac{1}{2}\left(\sum\nolimits_{j=1}^{n} p_j u_j^2 - \sum\nolimits_{i,j=1}^{n} p_i p_j u_i u_j\right)\right\}.$$

Так как случайные векторы $\eta_1, \ldots, \eta_T$ независимы и одинаково распределены, то

$$\begin{gathered} f_\nu(\mathbf{u}) = \mathbf{E} \exp\left\{i\left(\mathbf{u}, \tfrac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} \left(\eta_t - \left(\mathbf{p} + \tfrac{\mathbf{a}}{\sqrt{T}}\right)\right)\right)\right\} = \\ = \left(\mathbf{E} \exp\left\{i\left(\tfrac{\mathbf{u}}{\sqrt{T}}, \eta_1\right)\right\} \exp\left\{-i\left(\tfrac{\mathbf{u}}{\sqrt{T}}, \mathbf{p} + \tfrac{\mathbf{a}}{\sqrt{T}}\right)\right\}\right)^T. \end{gathered}$$

Используя формулу полной вероятности и формулу Тейлора, находим, что при $\mathbf{u} = \mathrm{const}$, $T \to \infty$

$$\mathbf{E} \exp\left\{i\left(\tfrac{\mathbf{u}}{\sqrt{T}}, \eta_1\right)\right\} = \sum\nolimits_{k=1}^{n} \left(p_k + \tfrac{a_k}{\sqrt{T}}\right) \exp\left\{i\tfrac{u_k}{\sqrt{T}}\right\} =$$
$$= \sum\nolimits_{k=1}^{n} \left(p_k + \tfrac{a_k}{\sqrt{T}}\right)\left(1 + i\tfrac{u_k}{\sqrt{T}} - \tfrac{u_k^2}{2T} + o\left(\tfrac{1}{T}\right)\right) =$$
$$= 1 + i\tfrac{(\mathbf{u}, \mathbf{p}+\mathbf{a}/\sqrt{T})}{\sqrt{T}} - \sum\nolimits_{k=1}^{n} \tfrac{p_k u_k^2}{2T} + o\left(\tfrac{1}{T}\right)$$

и

$$\exp\left\{-i\left(\tfrac{\mathbf{u}}{\sqrt{T}}, \mathbf{p} + \tfrac{\mathbf{a}}{\sqrt{T}}\right)\right\} = 1 - i\tfrac{(\mathbf{u}, \mathbf{p}+\mathbf{a}/\sqrt{T})}{\sqrt{T}} - \tfrac{(\mathbf{u}, \mathbf{p}+\mathbf{a}/\sqrt{T})^2}{2T} + o\left(\tfrac{1}{T}\right).$$

Из последних двух формул следует, что

$$\mathbf{E} \exp\left\{i\left(\tfrac{\mathbf{u}}{\sqrt{T}}, \eta_1\right)\right\} \exp\left\{-i\left(\tfrac{\mathbf{u}}{\sqrt{T}}, \mathbf{p} + \tfrac{\mathbf{a}}{\sqrt{T}}\right)\right\} =$$
$$= \left(1 - i\tfrac{(\mathbf{u}, \mathbf{p}+\mathbf{a}/\sqrt{T})}{\sqrt{T}} - \tfrac{(\mathbf{u}, \mathbf{p}+\mathbf{a}/\sqrt{T})^2}{2T}\right)\left(1 + i\tfrac{(\mathbf{u}, \mathbf{p}+\mathbf{a}/\sqrt{T})}{\sqrt{T}} - \sum_{k=1}^{n} \tfrac{p_k u_k^2}{2T}\right) + o\left(\tfrac{1}{T}\right) =$$
$$= 1 - \tfrac{1}{2T}\left(\sum\nolimits_{k=1}^{n} p_k u_k^2 - \sum\nolimits_{k,j=1}^{n} u_k u_j p_k p_j\right) + o\left(\tfrac{1}{T}\right) =$$
$$= 1 - \tfrac{1}{2T}\,\mathbf{u} D \mathbf{u}^{\mathsf{T}} + o\left(\tfrac{1}{T}\right).$$

Следовательно, при $\mathbf{u} = \mathrm{const}$, $T \to \infty$

$$f_\nu(\mathbf{u}) = \left(\mathbf{E} \exp\left\{i\left(\tfrac{\mathbf{u}}{\sqrt{T}}, \eta_1\right)\right\} \exp\left\{-i\left(\tfrac{\mathbf{u}}{\sqrt{T}}, \mathbf{p} + \tfrac{\mathbf{a}}{\sqrt{T}}\right)\right\}\right)^T =$$
$$= \left(1 - \tfrac{1}{2T}\,\mathbf{u} D \mathbf{u}^{\mathsf{T}} + o\left(\tfrac{1}{T}\right)\right)^T = \exp\left\{-\tfrac{1}{2}\,\mathbf{u} D \mathbf{u}^{\mathsf{T}}\right\}(1 + o(1)),$$

что и требовалось доказать. Лемма доказана.

§7. Длина отрезка апериодичности случайного отображения

Пусть по конечному множеству $S = \{1, \ldots, N\}$ и функции $f : \{1, \ldots, N\} \to \{1, \ldots, N\}$ построена рекуррентная последовательность

$$x_{n+1} = f(x_n), \quad n = 0, 1, \ldots$$

Последовательности такого вида используются, например, в датчиках псевдослучайных чисел в ЭВМ. Так как $N < \infty$, то рано или поздно последовательность $\{x_n\}$ зацикливается. Например, если $f(x) = x + 1 \mod N$, то период такой последовательности равен N. Если $f(x) = ax + c \mod N$, то период равен N тогда и только тогда, когда выполнены три условия: а) c и N взаимно просты, б) $b = a - 1$ кратно p для каждого простого p, являющегося делителем N, в) b кратно 4,если N кратно 4. Что можно сказать о моменте зацикливания, если функция f выбирается случайно?

Рассмотрим множество всех функций $f : S \to S$. Число таких функций равно N^N. Припишем каждой функции вероятность $1/N^N$. Будем считать, что случайная функция $\varphi : S \to S$ имеет равномерное распределение на множестве всех отображений S в себя.

Функция $\varphi : S \to S$ определяется набором значений $\varphi(1), \ldots, \varphi(N)$. Если φ имеет равномерное распределение на множестве всех отображений $S \to S$, то случайные величины $\varphi(1), \ldots, \varphi(N)$ независимы и каждая из них имеет равномерное распределение на S:

$$\mathbf{P}\{\varphi(k) = n\} = N^{N-1}/N^N = 1/N \quad \text{при любых} \quad k, n \in S.$$

Верно и обратное: если случайные величины $\varphi(1), \ldots, \varphi(N)$ независимы и равномерно распределены на S, то определяемая ими функция φ имеет равномерное распределение на множестве всех отображений $S \to S$: для каждой фиксированной функции $f : S \to S$

$$\mathbf{P}\{\varphi = f\} = \mathbf{P}\{\varphi(k) = f(k), k \in S\} = \prod_{k \in S} \mathbf{P}\{\varphi(k) = f(k)\} = \frac{1}{N^N}.$$

Функции $\varphi : S \to S$ можно сопоставить ориентированный граф с множеством вершин S, в котором каждое ребро (i, j), ведущее из i в j, существует тогда и только тогда, когда $\varphi(i) = j$. При любом x_0 траектория $x_{n+1} = \varphi(x_n), n = 0, 1, \ldots$, рано или поздно попадает в уже пройденную вершину графа, и отрезок апериодичности τ_N — это число ребер, пройденных до первого попадания в уже встречавшуюся вершину:

$$\tau_N = \min\{t \geqslant 1 : \varphi^t(x_0) \in \{x_0, \varphi(x_0), \ldots, \varphi^{t-1}(x_0)\}\}.$$

Введем также случайную величину κ_N, равную числу точек графа, лежащих на его циклах:

$$\kappa_N = |\{x \in S : x = \varphi^t(x) \quad \text{при некотором} \quad t \geqslant 1\}|.$$

Из нашего предположения о распределении значений функции φ следует, что для любого $n \geq 0$ и любых попарно различных $a_0, a_1, \ldots, a_{n-1} \in S$ условное распределение $x_n = \varphi(x_{n-1})$ при условии, что $x_0 = a_0, \ldots, x_{n-1} =$

a_{n-1}, не зависит от $a_0, a_1, \ldots, a_{n-1}$ и является равномерным на S. А это значит, что *до момента зацикливания* τ_N *включительно последовательность* $\{x_n\}$ *имеет те же свойства, что последовательность* $\{\xi_n\}$ *независимых случайных величин, имеющих равномерное распределение на* S. Однако после случайного момента τ_N свойства последовательностей $\{x_n\}$ и $\{\xi_n\}$ различаются: элементы $\{x_n\}$ начинают циклически повторяться, а элементы $\{\xi_n\}$ являются независимыми равномерно распределенными на S случайными величинами.

Теорема о числе шагов до зацикливания. *Если* φ *– случайное равновероятное отображение конечного множества* $S = \{1, \ldots, N\}$ *в себя, то при* $N \to \infty$

$$\lim \mathbf{P}\{\tau_N > x\sqrt{2N}\} = e^{-x^2} \quad (x \geqslant 0).$$

и

$$\mathbf{E}\kappa_N < 1 + \sqrt{\frac{\pi N}{2}}, \quad \mathbf{E}\kappa_N = (1 + o(1))\sqrt{\frac{\pi N}{2}}.$$

Таким образом, при случайном равновероятном выборе отображения N-элементного множества в себя длина отрезка апериодичности при $N \to \infty$ с вероятностью, стремящейся к 1, имеет порядок $O(\sqrt{N})$ (т. е. существенно меньше максимально возможного периода), и среднее число циклических точек имеет такой же порядок. Однако для конкретных, специально выбранных функций распределение длины отрезка апериодичности и числа циклических точек могут быть другими. Например, если функция f взаимно однозначна, то все точки лежат на циклах.

Доказательство. Пусть $v \geqslant 1$ — целое число. Если все v значений

$$x_0, \varphi(x_0), \varphi^2(x_0) = \varphi(\varphi(x_0)), \ldots, \varphi^{v-1}(x_0)$$

различны, то $\varphi^v(x_0) = \varphi(\varphi^{v-1}(x_0))$ не зависит от них и имеет равномерное распределение на S (поскольку аргумент $\varphi^{v-1}(x_0)$ отличен от $x_0, \varphi(x_0), \varphi^2(x_0), \ldots, \varphi^{v-2}(x_0)$), т. е.

$$\begin{gathered}\mathbf{P}\{\tau_N > v |\, \tau_N > v-1\} = \\ \mathbf{P}\{\varphi^v(x_0) \notin \{x_0, \varphi(x_0), \ldots, \varphi^{v-1}(x_0)\} \,|\, \varphi^i(x_0) \neq \varphi^j(x_0), 0 \leqslant i < j < v\} = \\ = 1 - \tfrac{v}{N}\,.\end{gathered}$$

Значит,

$$\mathbf{P}\{\tau_N > t\} = \prod_{v=1}^{t-1} \mathbf{P}\{\tau_N > v |\, \tau_N > v-1\} = \prod_{v=1}^{t-1} \left(1 - \tfrac{v}{N}\right).$$

Если $0 < x < \frac{1}{2}$, то $\ln(1-x) = -\sum_{k\geqslant 1} \frac{x^k}{k} < -x$ и $\ln(1-x) = -\sum_{k\geqslant 1} \frac{x^k}{k} > -x - \sum_{k\geqslant 2} \frac{x^k}{2} = -x - \frac{x^2}{2(1-x)} > -x - x^2$. Поэтому при $0 \leqslant t < \frac{N}{2}$

$$\mathbf{P}\{\tau_N > t\} \leqslant \exp\left\{-\sum\nolimits_{v=1}^{t-1} \frac{v}{N}\right\} = \exp\left\{\frac{t(t-1)}{2N}\right\}, \tag{50}$$

$$\mathbf{P}\{\tau_N > t\} \geqslant \exp\left\{-\sum\nolimits_{v=1}^{t-1} \left(\frac{v}{N} + \frac{v^2}{N^2}\right)\right\} = \exp\left\{-\frac{t(t-1)}{2N}\left(1 + \frac{2t-1}{3N}\right)\right\}. \tag{51}$$

Полагая $t = \left[x\sqrt{2N}\right]$ и переходя к пределу при $N \to \infty$, завершаем доказательство первого утверждения теоремы.

Для доказательства второго утверждения заметим, что если $\chi_N(x)$ — индикатор того, что точка $x \in S$ лежит на цикле графа, соответствующего функции φ, то $\varkappa_N = \sum\limits_{x=1}^{N} \chi_N(x)$ и

$$\mathbf{E}\varkappa_N = N\mathbf{E}\chi_N(x)$$

в силу равноправия всех точек из S. Далее, по формуле полной вероятности аналогично предыдущим рассуждениям получаем:

$$\mathbf{P}\{\chi_N(x) = 1\} = \sum_{t=1}^{N} \mathbf{P}\{x \text{ лежит на цикле длины } t\} = \sum_{t=1}^{N} \frac{1}{N} \prod_{k=1}^{t-1} \left(1 - \frac{k}{N}\right), \tag{52}$$

т. е.

$$\mathbf{E}\varkappa_N = N\mathbf{P}\{\chi_N(x) = 1\} = \sum_{t=1}^{N} \prod_{k=1}^{t-1} \left(1 - \frac{k}{N}\right).$$

Заметим, что из (50) следуют немного более грубые неравенства

$$\mathbf{P}\{\tau_N \geqslant t\} \leqslant \exp\left\{-\frac{(t-1)^2}{2N}\right\}, \quad 1 \leqslant t < N,$$

$$\mathbf{P}\{\tau_N \geqslant t\} \geqslant \exp\left\{-\frac{t^2}{2N} - \frac{t^3}{3N^2}\right\}, \quad 1 \leqslant t < \tfrac{N}{2}. \tag{53}$$

Отсюда и из (52), используя замену переменных $z = \frac{1}{2}y^2$ и то, что интеграл от плотности $e^{-x^2/2}/\sqrt{2\pi}$ стандартного нормального распределения по интервалу $[0,\infty)$ равен $\frac{1}{2}$, находим:

$$\mathbf{E}\kappa_N = \sum_{t=1}^{N}\prod_{k=1}^{t-1}\left(1-\frac{k}{N}\right) < \sum_{t=1}^{N}\exp\left\{-\frac{(t-1)^2}{2N}\right\} < 1+\sum_{t=1}^{\infty}e^{-\frac{t^2}{2N}} <$$

$$< 1+\int\limits_{y>0} e^{-\frac{y^2}{2N}}dy = 1+\sqrt{N}\int\limits_{z>0} e^{-\frac{z^2}{2}}dz = 1+\sqrt{N}\frac{\sqrt{2\pi}}{2} = 1+\sqrt{\frac{\pi N}{2}}.$$

Аналогично (пользуясь тем, что $\frac{t^3}{N^2} < \frac{1}{N^{1/4}}$ при $0<t<N^{7/12}$) получаем оценку снизу:

$$\mathbf{E}\kappa_N = \sum_{t=1}^{N}\prod_{k=1}^{t-1}\left(1-\frac{k}{N}\right) > \sum_{t=1}^{N^{7/12}}\exp\left\{-\frac{t^2}{2N}-\frac{t^3}{3N^2}\right\} >$$

$$> \sum_{t=1}^{N^{7/12}}\exp\left\{-\frac{t^2}{2N}-\frac{1}{N^{1/4}}\right\} > e^{-N^{-1/4}}\int\limits_{1}^{N^{7/12}} e^{-\frac{y^2}{2N}}dy =$$

$$= e^{-N^{-1/4}}\sqrt{N}\int\limits_{1/\sqrt{N}}^{N^{1/6}} e^{-\frac{z^2}{2}}dz = (1+o(1))\sqrt{\frac{\pi N}{2}}, \quad N\to\infty.$$

Теорема доказана.

Покажем, как этот результат можно применить при оценке сложности метода Полларда поиска делителя натурального числа N.

Алгоритм тривиального деления N на все натуральные (или все простые) числа, меньшие $\sqrt{N}$, имеет при $N\to\infty$ сложность порядка $O(\sqrt{N})$; такой алгоритм практически неприменим уже к числам порядка $10^{30}-10^{40}$.

Первый универсальный асимптотически более эффективный алгоритм был предложен в середине 70-х годов Поллардом; он предназначается для поиска «малых» делителей числа N и часто называется «ρ-алгоритмом», поскольку основан на поиске отрезка апериодичности некоторых отображений, а форма греческой буквы ρ напоминает цикл с подходом.

Метод Полларда состоит в следующем. Пусть N — натуральное число, имеющее простые делители $p_1<p_2<\ldots$, которые нам неизвестны, и требуется найти хотя бы один делитель числа N. Выберем какой-нибудь многочлен $f(x)$ с целыми коэффициентами, выберем начальное значение $y_0\in Z_N$ и будем строить последовательность пар $\{(y_n,y_{2n}),n=0,1,\ldots\}$

по формулам

$$y_{n+1} \equiv f(y_n) \pmod N, \quad y_{2(n+1)} \equiv f(f(y_{2n})) \pmod N, \quad n = 0, 1, \ldots$$

Для каждого n будем вычислять $d_n = \text{НОД}(y_{2n} - y_n, N)$ до тех пор, пока не окажется, что $d_n > 1$ (и при этом $y_{2n} - y_n \neq 0$). Понятно, что значение $d_n > 1$ является делителем N.

Обоснование того, что метод Полларда лучше метода тривиального деления, использует несколько разнородных соображений.

Лемма 1. *Пусть* $A = \{1, \ldots, N\}, N < \infty$, *и по функции* $f : A \to A$ *построена рекуррентная последовательность*

$$y_0 \in A,\ y_{n+1} = f(y_n),\ n \geqslant 0, \tag{54}$$

которая имеет отрезок апериодичности $s = \min\{n : y_n \in \{y_0, y_1, \ldots, y_{n-1}\}\}$. *Тогда*

$$\min\{n : y_n = y_{2n}\} \leqslant s.$$

Доказательство. Пусть T — минимальный период последовательности $\{y_n\}$. По определению отрезка апериодичности имеем $s \geqslant T,\ y_s = y_{s-T}$ и $y_{s+k} = y_{s+k-T}$ при всех $k \geqslant 0$. Поэтому если $s - T + 1 \leqslant n \leqslant s$ и $n = mT$, m – натуральное, то точка y_n лежит на цикле, $T \geqslant 1$ и

$$y_{2n} = y_{n+n} = y_{n+mT} = y_n.$$

Лемма доказана.

Следовательно, период любой рекуррентной последовательности вида (54) на конечном множестве можно найти, вычисляя

$$(y_{n+1}, y_{2(n+1)}) = (f(y_n), f(f(y_{2n}))), \quad n = 0, 1, \ldots,$$

до тех пор, пока не будет выполнено условие $y_{2n} = y_n$. Такое значение n и будет периодом (возможно, не минимальным). При использовании этого метода нужно помнить не всю цепочку $y_0, y_1, \ldots, y_n$, а только текущую пару (y_n, y_{2n}) и ее порядковый номер n.

Лемма 2. *Пусть* $N = pv$ *– произведение двух натуральных чисел,* $f(x) = a_0 + a_1 x + \ldots + a_m x^m$ *– многочлен с целыми коэффициентами, и последовательности* $\{y_n\}$ *со значениями в* $Z_N = \{0, 1, \ldots, N-1\}$ *и* $\{y_n(p)\}$ *со значениями в* $Z_p = \{0, 1, \ldots, p-1\}$ *удовлетворяют условиям*

$$y_0 \equiv y_0(p) \pmod p,$$
$$y_{n+1} \equiv f(y_n) \pmod N, \quad y_{n+1}(p) \equiv f(y_n(p)) \pmod p, \quad n = 0, 1, \ldots$$

Тогда при всех натуральных n

$$y_n(p) \equiv y_n \pmod p.$$

Доказательство проводится индукцией по n и следует из того, что если $a \in Z_N, a' \in Z_p, a \equiv a' \pmod p$ и $f(x)$ – многочлен с целыми коэффициентами, то

$$f(a) \equiv f(a') \pmod p.$$

Последнее равенство следует из двух простых замечаний: а) вычисление значения многочлена сводится к последовательному выполнению операций умножения и сложения, б) если $a' \equiv a \pmod p$, $b' \equiv b \pmod p$, то

$$(a' + b') \pmod p = ((a + b) \pmod N) \pmod p,$$
$$(a'b') \pmod p = ((ab) \pmod N) \pmod p.$$

Получим теперь эвристическую (не вполне строгую) оценку сложности метода Полларда, заменяя детерминистический алгоритм его вероятностной моделью и оценивая математическое ожидание числа шагов в этой вероятностной модели.

Согласно теореме о распределении длины отрезка апериодичности случайного отображения конечного множества $\{1, \ldots, M\}$ случайная длина отрезка апериодичности имеет порядок $O(\sqrt{M})$. Пусть p — любой делитель числа N. Если считать последовательность $y_n(p)$, в которой

$$y_0(p) \equiv y_0 \pmod p \quad \text{и} \quad y_{n+1}(p) \equiv f(y_n(p)) \pmod p,$$

случайной, то длина ее отрезка апериодичности должна иметь порядок $O(\sqrt{p})$. Согласно лемме 1 тогда условие $y_{2n}(p) = y_n(p)$ с большой вероятностью будет выполняться при некотором $n_p = O(\sqrt{p})$.

Так как делители N нам неизвестны, то построить последовательность $\{y_n(p)\}$ мы не можем. Однако согласно лемме 2 при любом n выполняется соотношение $y_n(p) \equiv y_n \pmod p$. Поэтому

$$y_{2n_p} - y_{n_p} \equiv 0 \pmod p$$

и, значит, НОД$(y_{2n_p} - y_{n_p}, N)$ делится на p. Значит, хотя бы один делитель N с большой вероятностью будет найден за $n_p = O(\sqrt{p})$ шагов.

Число арифметических операций при вычислении значения многочлена f имеет не меньший порядок, чем число $|f|$ его ненулевых коэффициентов.

Вычисление НОД по алгоритму Евклида проводится за $O(\ln N)$ арифметических операций. Таким образом, сложность каждого шага имеет порядок $O(|f| + \ln N)$. Поскольку для любого составного N его минимальный делитель не превосходит $\sqrt{N}$, постольку какой-нибудь делитель N при использовании описанного алгоритма с большой вероятностью будет найден за

$$O(\sqrt[4]{N}(|f| + \ln N))$$

арифметических операций. Если $|f| = O(\ln N)$, то оценкой числа арифметических операций оказывается $O(\sqrt[4]{N} \ln N)$, что по порядку существенно меньше, чем $O(\sqrt{N})$ операций в алгоритме тривиального деления.

Замечание. Можно ли считать использование теоремы о распределении длины отрезка апериодичности математически строгим? В этой теореме предполагается, что рассматриваемое случайное отображение выбирается равновероятно из множества всех N^N отображений N-элементного множества в себя. Как следует из оценок числа операций, не имеет смысла выбирать в качестве f многочлен с числом ненулевых коэффициентов, превышающим $\ln N$, поскольку это увеличит порядок оценки сложности алгоритма. Однако число многочленов степени не выше N с $O(\ln N)$ ненулевыми коэффициентами не превосходит $C_N^{\ln N} N^{\ln N} = O(N^{2 \ln N})$ при $N \to \infty$, что по порядку значительно меньше общего числа N^N отображений N-элементного множества в себя. Значит, считать такие многочлены реализациями случайного отображения с равномерным распределением не вполне корректно. Однако практически даже совсем простые многочлены порождают отображения, подходящие для использования в методе Полларда.

Такая ситуация часто возникает при анализе сложных алгоритмов, и поэтому при их исследовании используются «правдоподобные» предположения (например, предположение о том, что даже просто вычисляемые многочлены, как правило, обладают — с точки зрения этого алгоритма — свойствами, близкими к свойствам случайных отображений, имеющих равномерное распределение). Не нужно думать, что эта не очень приятная для математика ситуация возникает только в таких прикладных задачах. По сути дела, теория пределов является по существу методом получения простых, но неточных (приближенных) формул в случаях, когда точные формулы либо слишком громоздки, либо не могут быть записаны с помощью элементарных функций.

§ 8. Примеры применения метода моментов

В качестве первого примера применения метода моментов докажем предельную теорему для числа пустых ячеек $\mu_0(T,N)$ в равновероятной схеме размещения частиц по ячейкам.

Теорема 3. *Если* $T, N \to \infty$ *так, что* $T = N\ln N - N\ln\lambda + o(N)$, *то*

$$\lim_{n\to\infty} \mathbf{P}\{\mu_0(T,N) = k\} = \frac{\lambda^k}{k!}\, e^{-\lambda}, \quad k = 0, 1, \ldots$$

Доказательство. Согласно (5)

$$\mathbf{E}\mu_0^{[m]}(T,N) = N^{[m]}\left(1 - \frac{m}{N}\right)^T.$$

При указанных в теореме условиях для каждого фиксированного $m = 1, 2, \ldots$

$$\ln N^{[m]} = \ln\left(N^m \prod_{j=0}^{m}\left(1 - \frac{j}{N}\right)\right) = m\ln N + O\left(\frac{1}{N}\right),$$

$$\ln\left(1-\frac{m}{N}\right)^T = T\ln\left(1-\frac{m}{N}\right) = -\left(\frac{m}{N} + O\left(\frac{1}{N^2}\right)\right)(N\ln N - N\ln\lambda + o(N)) =$$
$$= -\left(1 + O\left(\frac{1}{N}\right)\right) m(\ln N - \ln\lambda + o(1)),$$

откуда следует, что

$$\ln \mathbf{E}\mu_0^{[m]}(T,N) = m\ln\lambda + O(N^{-1}\ln N) = m\ln\lambda + o(1),$$

т. е.

$$\lim_{N,T\to\infty} \mathbf{E}\mu_0^{[m]}(T,N) = \lambda^m,\ m = 1, 2, \ldots$$

Отсюда и из теоремы 2 вытекает утверждение теоремы 3.

Для равновероятной схемы размещения T частиц по N ячейкам нетрудно найти математическое ожидание числа $\mu_r(T,N)$ при любом r: если $\eta_k(T)$ — число частиц, попавших в k-ю ячейку после размещения T частиц, то

$$\mathbf{E}\mu_r(T,N) = \mathbf{E}\sum_{k=1}^{N} I\{\eta_k(T) = r\} =$$
$$= NC_T^r \frac{1}{N^r}\left(1 - \frac{1}{N}\right)^{T-r} = \frac{T^{[r]}}{r!N^{r-1}}\left(1 - \frac{1}{N}\right)^{T-r}.$$

В схеме серий, когда $r = \text{const} \geqslant 1$, а $N, T \to \infty$, возможно несколько типов регулярного поведения $\mathbf{E}\mu_r(T,N)$:

а) если $T = \alpha N^{1-\frac{1}{r}}(1+o(1))$, $\alpha = \text{const}$, то

$$\mathbf{E}\mu_r(T,N) = (1+o(1))\frac{\alpha^r}{r!},$$

т. е. среднее число ячеек, в которые попало r частиц, стремится к константе;

б) если $T = \alpha N(1+o(1))$, то

$$\mathbf{E}\mu_r(T,N) = (1+o(1))N\frac{\alpha^r}{r!}\,e^{-\alpha},$$

т. е. среднее число ячеек, в которые попало r частиц, растет по N и T линейно;

в) если $T = N\ln N + rN\ln\ln N - N\ln\alpha + o(N)$, то

$$\ln\frac{T^{[r]}}{N^r} = r\ln T + O\left(\tfrac{r^2}{T}\right) - r\ln N = r\ln\ln N + O\left(\tfrac{\ln\ln N}{\ln N} + \tfrac{1}{T}\right)$$

и

$$\ln\left(N\left(1-\frac{1}{N}\right)^{T-r}\right) = \ln N + (T-r)\left(-\tfrac{1}{N} + O\left(\tfrac{1}{N^2}\right)\right) =$$
$$= -r\ln\ln N + \ln\alpha + O\left(\tfrac{r}{N} + \tfrac{T}{N^2}\right),$$

откуда следует, что $\mathbf{E}\mu_r(T,N) = (1+o(1))\frac{\alpha}{r!}$, т. е. среднее число ячеек, в которые попало r частиц, стремится к константе.

Если выполняются условия а), то говорят, что параметры N, T равновероятной схемы размещения частиц по ячейкам изменяются в *левой* области, если выполняются условия б) — то в *центральной* области, и если выполняются условия в) — то в *правой* области.

Для последовательностей полиномиальных (неравновероятных) схем размещения аналогичные понятия левой, центральной и правой областей можно вводить только если вероятности ячеек изменяются не слишком нерегулярно. В качестве примера докажем методом моментов теорему о сходимости распределений $\mu_r(T,N)$ в полиномиальной схеме к распределению Пуассона в левой области.

Теорема 4. *Если целое $r \geqslant 2$ фиксировано, а последовательность полиномиальных схем размещения $T \to \infty$ частиц по $N \to \infty$ ячейкам удовлетворяет условиям*

$$T\max_{1\leqslant k\leqslant N} p_k \to 0, \quad C_T^r\sum_{k=1}^{N} p_k^r \to \lambda < \infty,$$

то

$$\lim_{N,T\to\infty}\mathbf{P}\{\mu_r(T,N)=k\}=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\ldots$$

Доказательство. Покажем, что при условиях теоремы факториальный момент $\mu_r(T,N)$ порядка m стремится к λ^m при любом натуральном m. По лемме о факториальных моментах

$$\mathbf{E}\mu_r^{[m]}(T,N)=m!\sum_{1\leqslant j_1<\ldots<j_m\leqslant N}\mathbf{P}\{\eta_{j_k}=r,k=1,\ldots,m\}=$$

$$=\sum_{\substack{j_1,\ldots,j_m=1\\ j_k\neq j_l\,(k\neq l)}}^{N}\frac{T^{[mr]}}{(r!)^m}p_{j_1}^r\ldots p_{j_m}^r(1-p_{j_1}-\ldots-p_{j_m})^{T-mr}.$$

Первый коэффициент под знаком суммы не зависит от индексов суммирования и

$$\tfrac{1}{(r!)^m}T^{[mr]}=(\tfrac{T^r}{r!})^m(1+o(1))\quad\text{при}\quad m=\mathrm{const},\ r=\mathrm{const},\ T\to\infty.$$

При условиях теоремы последний множитель не больше 1 и стремится к 1 равномерно по области суммирования:

$$1\geqslant(1-p_{j_1}-\ldots-p_{j_m})^{T-mr}\geqslant\left(1-m\max_{1\leqslant k\leqslant N}p_k\right)^{T-mr}\geqslant 1-Tm\max_{1\leqslant k\leqslant N}p_k=1+o(1).$$

Наконец,

$$\sum_{\substack{j_1,\ldots,j_m=1\\ j_k\neq j_l\,(k\neq l)}}^{N}p_{j_1}^r\ldots p_{j_m}^r=\left(\sum_{j=1}^{N}p_j^r\right)^m-\sum_{\substack{j_1,\ldots,j_m=1\\ \exists k\neq l:\ j_k=j_l}}^{N}p_{j_1}^r\ldots p_{j_m}^r,$$

и

$$\sum_{\substack{j_1,\ldots,j_m=1\\ \exists k\neq l:\ j_k=j_l}}^{N}p_{j_1}^r\ldots p_{j_m}^r\leqslant\sum_{1\leqslant k<l\leqslant m}\sum_{\substack{j_1,\ldots,j_m=1\\ j_k=j_l}}^{N}\prod_{w=1}^{m}p_{j_w}^r=C_m^2\sum_{j_1,\ldots,j_{m-1}=1}^{N}p_{j_{m-1}}^r\prod_{w=1}^{m-1}p_{j_w}^r\leqslant$$

$$\leqslant C_m^2\left(\max_{1\leqslant j\leqslant N}p_j\right)^r\sum_{j_1,\ldots,j_{m-1}=1}^{N}\prod_{w=1}^{m-1}p_{j_w}^r=C_m^2\left(\max_{1\leqslant j\leqslant N}p_j\right)^r\left(\sum_{j=1}^{N}p_j^r\right)^{m-1}.$$

Заметим теперь, что по условию теоремы $C_T^r\sum\limits_{k=1}^{N}p_k^r\to\lambda$, т. е.

$$\left(\sum_{j=1}^{N}p_j^r\right)^{-1}=\frac{C_T^r}{\lambda+o(1)}=O(T^r),$$

значит,

$$\sum_{\substack{j_1,\ldots,j_m=1\\ \exists k\neq l: j_k=j_l}}^{N} p_{j_1}^r \ldots p_{j_m}^r = O\left(\left(T\max_{1\leqslant j\leqslant N} p_j\right)^r \left(\sum_{j=1}^N p_j^r\right)^m\right) = o\left(\left(\sum_{j=1}^N p_j^r\right)^m\right).$$

Из этих оценок следует, что при условиях теоремы

$$\mathbf{E}\mu_r^{[m]}(T,N) = (1+o(1))\left(\frac{T^r}{r!}\right)^m \left(\sum_{j=1}^N p_j\right)^m = (1+o(1))\lambda^m.$$

Теорема доказана.

При применении теоремы Б.А.Севастьянова исключительные множества часто выбирают так, чтобы включить в них все совокупности случайных величин, между которыми могут быть зависимости.

Пусть $\xi_1, \xi_2, \ldots$ – последовательность независимых одинаково распределенных случайных величин, $\mathbf{P}\{\xi_t = k\} = p_k$, $k = 0, 1, \ldots$, $p_0, p_1 > 0$. Будем через $\Xi_{t,s}$ обозначать цепочку $(\xi_t, \ldots, \xi_{t+s-1})$ из s последовательных величин, начиная с t-й, и введем сокращенные обозначения для цепочек из s знаков: $0^s = (0, \ldots, 0), 0^{s-1}1 = (0, \ldots, 0, 1)$. Очевидно,

$$\mathbf{P}\{\Xi_{t,s} = 0^s\} = p_0^s, \quad \mathbf{P}\{\Xi_{t,s} = 0^{s-1}1\} = p_0^{s-1}p_1.$$

Найдем асимптотические формулы для распределений двух похожих случайных величин:

$$\tau(0^s) = \min\{t \geqslant 1 : \Xi_{t,s} = 0^s\}, \quad \tau(0^{s-1}1) = \min\{t \geqslant 1 : \Xi_{t,s} = 0^{s-1}1\},$$

Утверждение. *Если $T, s \to \infty$ так, что*

$$T \cdot \mathbf{P}\{\Xi_{1,s} = 0^{s-1}1\} = Tp_0^{s-1}p_1 \to \lambda,$$

то

$$\mathbf{P}\{\tau(s)p_0^{s-1}p_1 > \lambda\} \to e^{-\lambda}. \tag{55}$$

Если $T, s \to \infty$ так, что

$$T \cdot \mathbf{P}\{\Xi_{1,s} = 0^s\} = Tp_0^s \to \lambda,$$

то при $s \to \infty$

$$\mathbf{P}\{\tau(0^s)p_0^s(1-p_0) > \lambda\} \to e^{-\lambda}. \tag{56}$$

Доказательство. Положим

$$\sigma_T(0^{s-1}1)=\sum_{t=1}^{T}\chi\{\Xi_{t,s}=0^{s-1}1\},\quad \sigma_T(0^s)=\sum_{t=1}^{T}\chi\{\Xi_{t,s}=0^s\}.$$

Тогда $\{\tau(0^{s-1}1)>T\}=\{\sigma_T(0^{s-1}1)=0\}$ и $\{\tau(0^s)>T\}=\{\sigma_T(0^s)=0\}$. Поэтому формулы (55) и (56) можно получить как следствия предельных теорем для распределений $\sigma_T(0^{s-1}1)$ и $\sigma_T(0^s)$.

Убедимся сначала в том, что при условиях утверждения распределение $\sigma_T(0^{s-1}1)$ сходится к распределению Пуассона с параметром λ. Для этого воспользуемся теоремой Б.А.Севастьянова. Слагаемые в сумме $\sigma_T(0^{s-1}1)$ одинаково распределены и при $s\to\infty$

$$\mathbf{P}\{\Xi_{t,s}=0^{s-1}1\}=p_0^{s-1}p_1\to 0;$$

кроме того, по условию

$$\mathbf{E}\sigma_T(0^{s-1}1)=Tp_0^{s-1}p_1\to\lambda.$$

По формуле для факториальных моментов суммы индикаторов имеем при любом $m=2,3,\ldots$:

$$\mathbf{E}\sigma_T^{[m]}(0^{s-1}1)=m!\sum_{1\leqslant t_1<\ldots<t_m\leqslant T}\mathbf{P}\{\Xi_{t_j,s}=0^{s-1}1,j=1,\ldots,m\}.$$

Но

$$\mathbf{P}\{\Xi_{t_j,s}=0^{s-1}1,\,j=1,\ldots,m\}=\begin{cases}0 \quad\text{при}\quad \min_{1\leqslant k<m}|t_{k+1}-t_k|<s,\\ \left(p_0^{s-1}p_1\right)^m \quad\text{в противном случае.}\end{cases}$$

Поэтому если в качестве исключительного множества $I_m=I_m(T)$ выбрать

$$I_m=\{(t_1,\ldots,t_m)\subset\{1,\ldots,T\}:\min_{1\leqslant j<m}|t_{j+1}-t_j|<s\},$$

то условия

$$\sum_{I_m}\mathbf{P}\{\Xi_{t_j,s}=0^{s-1}1,\,j=1,\ldots,m\}=0$$

и

$$\max_{(t_1,\ldots,t_m)\notin I_m}\left|\frac{\mathbf{P}\{\Xi_{t_j,s}=0^{s-1}1,\,j=1,\ldots,m\}}{\prod\limits_{1\leqslant j\leqslant m}\mathbf{P}\{\Xi_{t_j,s}=0^{s-1}1\}}-1\right|=0$$

будут выполняться при любом $T > ms$. Остается убедиться в том, что

$$\sum_{I_m} \prod_{j=1}^{m} \mathbf{P}\{\Xi_{t_j,s} = 0^{s-1}1\} = |I_m| \left(p_0^{s-1} p_1\right)^m \to 0, \quad T \to \infty,$$

т. е. что $|I_m| = o(T^m)$. Но

$$I_m \subset \bigcup_{j=1}^{m-1} \{(t_1, \ldots, t_m) \subset \{1, \ldots, T\} : \, t_{j+1} - t_j < s\}$$

и $|\{(t_1, \ldots, t_m) \subset \{1, \ldots, T\} : t_{j+1} - t_j < s\}| \leqslant sT^{m-1}$. Значит, $|I_m| < msT^{m-1} = o(T^m)$, что и требовалось доказать. Таким образом, распределение $\sigma_T(0^{s-1}1)$ сходится к распределению Пуассона с параметром λ, и формула (55) доказана.

Рассмотрим теперь случайную величину

$$\sigma_T(0^s) = \sum_{j=1}^{T} \chi\{\Xi_{t,s} = 0^s\}.$$

В этой сумме слагаемые тоже одинаково распределены, вероятность p_0^s того, что каждое фиксированное слагаемое отлично от 0, стремится к 0, и $\mathbf{E}\sigma_T(0^s) = Tp_0^s \to \lambda$ по условию.

Вычислим второй факториальный момент $\sigma_T(0^s)$:

$$\mathbf{E}\sigma_T^{[2]}(0^s) = 2! \sum_{1 \leqslant t_1 < t_2 \leqslant T} \mathbf{P}\{\Xi_{t_1,s} = \Xi_{t_2,s} = 0^s\} =$$

$$= 2 \sum_{k=1}^{s-1} \sum_{t_1=1}^{T-k} \mathbf{P}\{\Xi_{t_1,s} = \Xi_{t_1+k,s} = 0^s\} + 2 \sum_{1 \leqslant t_1 < t_1+s \leqslant t_2 \leqslant T} \mathbf{P}\{\Xi_{t_1,s} = \Xi_{t_2,s} = 0^s\}.$$

Вторая сумма содержит C_{T-s+1}^2 слагаемых, каждое из которых равно $(p_0)^{2s}$, т. е. согласно условию она стремится к λ^2. Значения слагаемых в первой сумме тоже легко вычисляются:

$$\mathbf{P}\{\Xi_{t_1,s} = \Xi_{t_1+k,s} = 0^s\} = \mathbf{P}\{\xi_{t_1} = \xi_{t_1+1} = \ldots = \xi_{t_1+k+s-1} = 0\} = (p_0)^{k+s-1},$$

поэтому при $s \to \infty$

$$2 \sum_{k=1}^{s-1} \sum_{t_1=1}^{T-k} \mathbf{P}\{\Xi_{t_1,s} = \Xi_{t_1+k,s} = 0^s\} = 2 \sum_{k=1}^{s-1} (T-k)(p_0)^{k+s-1} =$$

$$= 2Tp_0^s \frac{1 - p_0^{s-1}}{1 - p_0} + 2p^s \left(\frac{sp_0^{s-1}}{1 - p_0} - \frac{1 - p_0^s}{(1 - p_0)^2} \right) \to \frac{2\lambda}{1 - p_0}.$$

Таким образом, $\mathbf{E}\sigma_T^{[2]}(0^s)$ стремится к $\lambda^2 + \frac{2\lambda}{1-p_0} \neq \lambda^2$, и методом моментов доказать сходимость распределения $\sigma_T(0^s)$ к распределению Пуассона нельзя.

Более того, предельное распределение $\sigma_T(0^s)$ и не является пуассоновским. Чтобы убедиться в этом, рассмотрим вспомогательную случайную величину

$$\sigma_T(\overline{0}0^s) = \sum_{t=1}^{T} \chi\{\Xi_{t,s} = 0^s,\, \xi_{t-1} \neq 0\},$$

считая для упрощения записи формул, что $\xi_0 = 1$. Тогда

$$\mathbf{P}\{\Xi_{t,s} = 0^s, \xi_{t-1} \neq 0\} = \begin{cases} (p_0)^s, & t = 1, \\ (1-p_0)(p_0)^s, & t > 1. \end{cases}$$

При наших условиях все эти вероятности стремятся к 0, а

$$\mathbf{E}\sigma_T(\overline{0}0^s) = (p_0)^s(1 + (T-1)(1-p_0)) \to (1-p_0)\lambda.$$

Вычисление факториальных моментов случайной величины $\sigma_T(\overline{0}0^s)$ проводится так же, как для цепочек $0^{s-1}1$, поскольку события

$$\{\Xi_{t_j,s} = 0^s, \xi_{t_j-1} \neq 0\}, \quad j = 1, \ldots, m,$$

несовместны, если $\min\limits_{1 \leqslant k < m} |t_{k+1} - t_k| < s$. Поэтому

$$\mathbf{E}\sigma_T^{[k]}(\overline{0}0^s) \to ((1-p_0)\lambda)^k \quad \text{при любом} \quad k = 2, 3, \ldots,$$

и распределение $\sigma_T(\overline{0}0^s)$ сходится к распределению Пуассона с параметром $(1-p_0)\lambda$. Так как

$$\tau(0^s) = \min\{t \geqslant 1 : \Xi_{t,s} = 0^s\} = \min\{t \geqslant 1 : \Xi_{t,s} = 0^s, \xi_{t-1} \neq 0\},$$

то отсюда следует формула (56).

Сравнение формул (55) и (56) показывает, что, например, при $p_0 = p_1 = \frac{1}{2}$ время ожидания появления цепочки 0^s вдвое больше времени ожидания цепочки $0^{s-1}1$. Этот эффект объясняется тем, что математические ожидания чисел появлений цепочек 0^s и $0^{s-1}1$ одинаковы, но цепочки 0^s появляются группами случайного размера: при условии, что $(\xi_t, \ldots, \xi_{t+s-1}) = 0^s$, вероятность события $\{(\xi_{t+1}, \ldots, \xi_{t+s}) = 0^s\}$ равна p_0, а не p_0^s. Значит, каждое событие $\{\Xi_{t,s} = 0^s, \xi_{t-1} \neq 0\}$ влечет за собой появление случайного числа событий $\{\Xi_{t,s} = 0^s\}$, причем число дополнительных появлений цепочки

0^s имеет геометрическое распределение с параметром p_0. Поэтому предельное распределение случайной величины $\sigma_T(0^s)$ можно представить в виде суммы случайного числа независимых случайных величин, имеющих геометрическое распределение с параметром p_0: число слагаемых совпадает с числом появления событий $\{\Xi_{t,s} = 0^s,\, \xi_{t-1} \neq 0\}$ и имеет распределение Пуассона с параметром $(1-p_0)\lambda$.

Производящая функция суммы k независимых случайных величин с геометрическим распределением с параметром p_0 равна $\left(\frac{(1-p_0)s}{1-p_0 s}\right)^k$. Используя формулу полной вероятности по числу слагаемых, находим, что производящая функция предельного распределения $\sigma_T(0^s)$ в нашей схеме серий имеет вид

$$\sum_{k=0}^{\infty} \frac{((1-p_0)\lambda)^k}{k!} e^{-(1-p_0)\lambda} \left(\frac{(1-p_0)s}{1-p_0 s}\right)^k = \exp\left(-(1-p_0)\lambda \frac{1-s}{1-p_0 s}\right).$$

Это пример сложного пуассоновского распределения, которое часто возникает при исследовании сумм зависимых индикаторов.

Применим теорему Севастьянова для доказательства сходимости распределения $\mu_r(T,N)$ к распределению Пуассона в правой области для неравновероятной полиномиальной схемы.

Теорема 4. *Если целое $r \geqslant 0$ фиксировано, а последовательность полиномиальных схем размещения $T \to \infty$ частиц по $N \to \infty$ ячейкам удовлетворяет условиям*

$$T \min_{1 \leqslant k \leqslant N} p_k \to \infty, \quad \mathbf{E}\mu_r(T,N) = \sum_{k=1}^{N} C_T^r p_k^r (1-p_k)^{T-r} \to \lambda < \infty,$$

то

$$\lim_{N,T\to\infty} \mathbf{P}\{\mu_r(T,N) = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots$$

Доказательство. Пусть $\eta_k = \eta_k(T)$ — число частиц, попавших в k-ю ячейку после размещения T частиц. Тогда

$$\mu_r(T,N) = \mathbb{I}\{\eta_1 = r\} + \ldots + \mathbb{I}\{\eta_N = r\},$$

$$b_{i_1,\ldots,i_m} = \mathbf{P}\{\eta_{i_k} = r,\, k = 1,\ldots,m\} = \frac{T^{[rm]}}{(r!)^m} \left(1 - \sum_{k=1}^{m} p_{i_k}\right)^{T-rm} \prod_{k=1}^{m} p_{i_k}^r.$$

Проверим выполнение условий а): во-первых,

$$b_k = C_T^r p_k^r (1-p_k)^{T-r} \leqslant \frac{(Tp_k)^r}{r!} e^{-Tp_k} \to 0,$$

так как $x^r e^{-x} \to 0$ при $x \to \infty$ при любом $r \geqslant 0$, а $Tp_k \to \infty$, и во-вторых, $\sum\limits_{1 \leqslant k \leqslant N} b_k = \mathbf{E}\mu_r(T, N) \to \lambda$ по условию.

Проверим выполнение условий п.б). Так как $p_1 + \ldots + p_N = 1$, то $\min\limits_{1 \leqslant k \leqslant N} p_k \leqslant \frac{1}{N}$; по условию $T \min\limits_{1 \leqslant k \leqslant N} p_k \to \infty$, значит, $T/N \to \infty$, т. е. $N = o(T)$. Убедимся в том, что исключительные множества можно выбирать следующим образом:

$$I_1(T) = \left\{ i \in \{1, \ldots, N\} : p_i > \frac{1}{T^{3/4}} \right\},$$
$$I_m(T) = \{(i_1, \ldots, i_m) : \{i_1, \ldots, i_m\} \cap I_1(T) \neq \varnothing, i_r \neq i_s (r \neq s)\}.$$

(Если, например, все $p_i = \frac{1}{N}$, то $I_m(T) = \varnothing$ при всех $m \geqslant 1$.) Очевидно,

$$\sum_{i \in I_1(T)} b_i = \sum_{\substack{i=1 \\ i : p_i > T^{-3/4}}}^{N} C_T^r p_i^r (1 - p_i)^{T-r} \leqslant C_T^r N \sup_{T^{-3/4} \leqslant x \leqslant 1} x^r (1-x)^{T-r}.$$

Заметим, что функция $x^r (1-x)^{T-r}$ на $[0, 1]$ неотрицательна, равна 0 при $x = 1$, а ее производная, равная $-T(1-x)^{T-1}$ при $r = 0$ и равная

$$\frac{d}{dx} x^r (1-x)^{T-r} = x^{r-1} (1-x)^{T-r-1} (r(1-x) - (T-r)x) = x^{r-1} (1-x)^{T-r-1} (r - Tx)$$

при $r \geqslant 1$, положительна при $0 < x < \frac{r}{T}$ и отрицательна при $\frac{r}{T} < x < 1$, т. е. максимальна при $x = \frac{r}{T}$. Далее, при $T > r^4$ (т. е. при $r < T^{1/4}$) имеем $\frac{r}{T} < \frac{1}{T^{3/4}}$, поэтому при $r = \mathrm{const}$ и $T \to \infty$ функция $x^r (1-x)^{T-r}$ на отрезке $[T^{-3/4}, 1]$ убывает,

$$\sup_{T^{-3/4} \leqslant x \leqslant 1} x^r (1-x)^{T-r} = \left(\frac{1}{T^{3/4}} \right)^r \left(1 - \frac{1}{T^{3/4}} \right)^{T-r}$$

и при $T \to \infty, N = o(T)$

$$\sum_{i \in I_1(T)} b_i \leqslant C_T^r N \left(\frac{1}{T^{3/4}} \right)^r \left(1 - \frac{1}{T^{3/4}} \right)^{T-r} \leqslant$$
$$\leqslant \frac{T^r}{r!\, T^{3r/4}} e^{-(T-r)/T^{3/4}} N = o\left(T^{1+\frac{r}{4}} e^{-T^{1/4}} \right) = o(1).$$

Пусть теперь $m > 1$. Разбивая $I_m(T)$ на подмножества

$$I_{m,k}(T) = \{(i_1, \ldots, i_m) \in I_m(T) \colon i_k \in I_1(T),\ i_j \notin I_1(T)\, (1 \leqslant j < k)\},\ k = 1, \ldots, m,$$

по номеру первого элемента в $i_1, \ldots, i_m$, принадлежащего $I_1(T)$, находим, что при $T \to \infty$ и $m = \mathrm{const}$

$$\sum_{(i_1,\ldots,i_m)\in I_m(T)} b_{i_1}\ldots b_{i_m} = \sum_{k=1}^{m} \sum_{(i_1,\ldots,i_m)\in I_{m,k}(T)} b_{i_1}\ldots b_{i_m} \leqslant$$
$$\leqslant \sum_{k=1}^{m} \sum_{i_k\in I_1(T)} b_{i_k} \sum_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_m=1}^{N} b_{i_1}\ldots b_{i_{k-1}} b_{i_{k+1}}\ldots b_{i_m} \leqslant$$
$$\leqslant m \left(\sum_{i\in I_1(T)} b_i\right)\left(\sum_{i=1}^{N} b_i\right)^{m-1} = m\, o(1)\, \lambda^{m-1}.$$

Чтобы оценить сумму по $I_m(T)$ вероятностей $b_{i_1,\ldots,i_m}$, заметим, что если $(i_1,\ldots,i_m) \in I_m(T)$, то $\max\{p_{i_1},\ldots,p_{i_m}\} > T^{-3/4}$ и

$$b_{i_1,\ldots,i_m} = \frac{T^{[rm]}}{(r!)^m}\left(1 - \sum_{k=1}^{m} p_{i_k}\right)^{T-rm} \prod_{k=1}^{m} p_{i_k}^r \leqslant$$
$$\leqslant T^{rm}\left(1 - \sum_{k=1}^{m} p_{i_k}\right)^{T-rm} < T^{rm}\left(1 - \frac{1}{T^{3/4}}\right)^{T-rm} < T^{rm} e^{-T^{1/4} + rmT^{-3/4}}.$$

Так как $N = o(T)$ и $|I_m(T)| < N^m$, то при $T \to \infty$

$$\sum_{(i_1,\ldots,i_m)\in I_m(T)} b_{i_1,\ldots,i_m} \leqslant N^m T^{rm} e^{-T^{1/5}} \to 0.$$

Для завершения проверки условий теоремы Севастьянова осталось убедиться в том, что

$$\max_{(i_1,\ldots,i_m)\notin I_m(T)} \left|\frac{b_{i_1,\ldots,i_m}}{b_{i_1}\ldots b_{i_m}} - 1\right| \to 0. \tag{57}$$

Если $(i_1,\ldots,i_m) \notin I_m(T)$, то $p_{i_1} + \ldots + p_{i_m} \leqslant \frac{m}{T^{3/4}} \to 0$ при $m = \mathrm{const}$ и $T \to \infty$, поэтому

$$\frac{b_{i_1,\ldots,i_m}}{b_{i_1}\ldots b_{i_m}} = \frac{\frac{T^{[mr]}}{(r!)^m} p_{i_1}^r \ldots p_{i_m}^r (1 - p_{i_1} - \ldots - p_{i_m})^{T-rm}}{\left(\frac{T^{[r]}}{(r!)}\right)^m p_{i_1}^r \ldots p_{i_m}^r (1 - p_{i_1})^{T-r} \ldots (1 - p_{i_m})^{T-r}} \sim$$
$$\sim \left(\frac{1 - p_{i_1} - \ldots - p_{i_m}}{(1 - p_{i_1})\ldots(1 - p_{i_m})}\right)^T.$$

Далее воспользуемся неравенствами

$$\ln(1-x) \leqslant -x \quad \text{при} \quad 0 \leqslant x \leqslant 1,$$

$$\ln(1-x) = -\sum_{k=1}^{\infty} \frac{x^k}{k} \geqslant -x - \frac{1}{2}\sum_{k=1}^{\infty} x^k = -x - \frac{x^2}{2(1-x)} \geqslant -x - x^2, \quad 0 \leqslant x \leqslant \frac{1}{2},$$

из них следует, что при достаточно больших T

$$-(p_{i_1} + \ldots + p_{i_m}) - (p_{i_1} + \ldots + p_{i_m})^2 \leqslant \ln(1 - p_{i_1} - \ldots - p_{i_m}) \leqslant -(p_{i_1} + \ldots + p_{i_m}),$$
$$-(p_{i_1} + \ldots + p_{i_m}) - (p_{i_1}^2 + \ldots + p_{i_m}^2) \leqslant \ln(1 - p_{i_1}) \ldots (1 - p_{i_m}) \leqslant -(p_{i_1} + \ldots + p_{i_m}).$$

Почленное вычитание дает:

$$-\frac{m^2}{T^{3/2}} \leqslant -(p_{i_1}+\ldots+p_{i_m})^2 \leqslant \ln \frac{1 - p_{i_1} - \ldots - p_{i_m}}{(1-p_{i_1})\ldots(1-p_{i_m})} \leqslant p_{i_1}^2+\ldots+p_{i_m}^2 \leqslant \frac{m}{T^{3/2}}.$$

Значит, при любом $m = \mathrm{const}$ и $T \to \infty$

$$\left| \ln \left(\frac{1 - p_{i_1} - \ldots - p_{i_m}}{(1-p_{i_1})\ldots(1-p_{i_m})} \right)^T \right| \leqslant T\frac{m^2}{T^{3/2}} = \frac{m}{T^{1/2}} \to 0$$

равномерно вне исключительного множества $I_m(T)$. Отсюда следует (57). Теорема доказана.

§9. Центральная предельная теорема для числа пустых ячеек в равновероятной схеме

Рассмотрим равновероятную схему размещения T частиц по N ячейкам. Как и ранее, будем через $\mu_0(T, N)$ обозначать число ячеек, оставшихся пустыми после размещения T частиц.

Предельные теоремы о пуассоновской аппроксимации распределения $\mu_r(T, N)$ мы доказывали методом моментов. Центральную предельную теорему для числа пустых ячеек в равновероятной схеме мы докажем другим, красивым и коротким способом, предложенным в [2]. Для неравновероятных схем доказательство значительно более громоздко.

Теорема. *В равновероятной схеме размещения T частиц по N ячейкам распределение случайной величины $\mu_0(T, N)$ удовлетворяет соотношению*

$$\mathbf{P}\left\{ \frac{\mu_0(T,N) - \mathbf{E}\mu_0(T,N)}{\sqrt{\mathbf{D}\mu_0(T,N)}} \leqslant x \right\} \to \Phi(x), \quad x \in (-\infty, \infty),$$

если $T, N \to \infty$ так, что $\frac{T}{N} \to C \in (0, \infty)$.

Замечание. Для равновероятной полиномиальной схемы мы доказали, что распределение $\mu_0(T,N)$ сходится к распределению Пуассона с параметром λ, если $N,T \to \infty$, $T = N\ln N - N\ln\lambda + o(N)$ (в правой области). С другой стороны, в левой области при $N,T \to \infty$, $\frac{T^2}{2N} \to \lambda$ распределение $\mu_2(T,N)$ сходится к распределению Пуассона с параметром λ, и $\mathbf{P}\{\exists r \geqslant 3 : \mu_r(T,N) > 0\} \to 0$. Поэтому в левой области

$$\mathbf{P}\{\mu_0(N,T) + \mu_1(N,T) + \mu_2(N,T) = N, \mu_1(N,T) + 2\mu_2(N,T) = T\} \to 1,$$

значит, $\mathbf{P}\{\mu_0(N,T) = N - T + \mu_2(N,T)\} \to 1$, т. е. распределение $\mu_0(N,T) - (N-T)$ сходится к распределению Пуассона:

$$\mathbf{P}\{\mu_0(N,T) - (N-T) = k\} \to \frac{\lambda^k}{k!}\, e^{-\lambda}, \quad k = 0, 1, \ldots$$

Доказательство разобьем на несколько лемм.

Лемма 1. *Пусть $\{\xi_{n1}, \ldots, \xi_{nn}\}_{n=1}^{\infty}$ – последовательность серий независимых случайных величин,*

$$\mathbf{P}\{\xi_{nk} = 1\} = p_{nk},\, \mathbf{P}\{\xi_{nk} = 0\} = 1 - p_{nk}, \quad k, n \geq 1,$$

и $\zeta_n = \xi_{n1} + \ldots + \xi_{nn}$. Если $\mathbf{D}\zeta_n \to \infty$ при $n \to \infty$, то

$$\mathbf{P}\left\{\frac{\zeta_n - \mathbf{E}\zeta_n}{\sqrt{\mathbf{D}\zeta_n}} \leqslant x\right\} \to \Phi(x), \quad x \in (-\infty, \infty).$$

Доказательство леммы 1. Покажем, что для сумм ζ_n выполняются условия теоремы Ляпунова об асимптотической нормальности, т. е. что отношение суммы третьих центральных абсолютных моментов слагаемых $\left(\sum_{k=1}^{n} \mathbf{E}|\xi_{nk} - p_{nk}|^3\right)$ к кубу среднеквадратичного отклонения $\left((\mathbf{D}\zeta_n)^{3/2}\right)$ стремится к 0 при $n \to \infty$. Действительно,

$$\mathbf{E}\xi_{nk} = p_{nk}, \quad \mathbf{D}\xi_{nk} = p_{nk}(1 - p_{nk}),$$
$$\mathbf{E}|\xi_{nk} - \mathbf{E}\xi_{nk}|^3 = p_{nk}(1 - p_{nk})(1 - 2p_{nk} + 2p_{nk}^2) < \mathbf{D}\xi_{nk}.$$

Таким образом,

$$\sum_{k=1}^{n} \mathbf{E}|\xi_{nk} - \mathbf{E}\xi_{nk}|^3 < \sum_{k=1}^{n} \mathbf{D}\xi_{nk} = \mathbf{D}\zeta_n = o((\mathbf{D}\zeta_n)^{3/2}),$$

если $\mathbf{D}\zeta_n \to \infty$. Лемма 1 доказана.

Лемма 2. *Если производящая функция* $\varphi(z) = \mathbf{E}z^{\mu}$ *целочисленной неотрицательной случайной величины* μ *– многочлен степени* N*, все корни* $x_1, \ldots, x_N$ *которого – действительные числа, то распределение* μ *совпадает с распределением суммы* N *независимых индикаторов* $\xi_1, \ldots, \xi_N$*, причем*

$$\mathbf{P}\{\xi_j = 0\} = -\frac{x_j}{1-x_j}, \quad \mathbf{P}\{\xi_j = 1\} = \frac{1}{1-x_j}, \quad j = 1, \ldots, N.$$

Доказательство леммы 2. Так как $\varphi(z) = \sum\limits_{k=0}^{N} \mathbf{P}\{\mu = k\} z^k$ — ряд с неотрицательными коэффициентами и $\varphi(1) = 1$, то $\varphi(z)$ отлично от 0 при всех $z > 0$. Значит, все корни $x_1, \ldots, x_N$ неположительны. Тогда

$$\varphi(z) = \mathbf{E}s^{\mu} = C(z - x_1) \ldots (z - x_N), \quad \text{где} \quad C = \frac{1}{(1-x_1)\ldots(1-x_N)}$$

и

$$\varphi(z) = \prod_{k=1}^{N} \left(-\frac{x_k}{1-x_k} + \frac{z}{1-x_k} \right). \tag{58}$$

В правой части этого равенства стоит произведение линейных функций с неотрицательными коэффициентами, которые при $s = 1$ принимают значение 1. Каждый сомножитель $\frac{z}{1-x_k} - \frac{x_k}{1-x_k}$ является производящей функцией случайной величины ξ_k, принимающей значение 0 с вероятностью $-\frac{x_j}{1-x_j}$ и значение 1 с вероятностью $\frac{1}{1-x_j}$. Отсюда и из равенства (58) следует, что распределения μ и суммы независимых случайных величин $\xi_1, \ldots, \xi_N$ совпадают. Лемма 2 доказана.

Лемма 3. *Производящая функция распределения* $\mu_0(T, N)$ *имеет только действительные корни.*

Доказательство леммы 3. Пусть $f_{T,N}(s) = \mathbf{E}s^{\mu_0(T,N)}$ — производящая функция распределения $\mu_0(T, N)$. Ранее было доказано, что

$$\mathbf{E}\mu_0^{[k]}(T, N) = \frac{d^k}{ds^k} f_{T,N}^{(k)}(s) \Big|_{s=1} = N^{[k]} \left(1 - \frac{k}{N}\right)^T,$$

и $f_{T,N}(s)$ – многочлен степени не выше N. Разлагая $f_{T,N}(s)$ в ряд Тейлора в точке $s = 1$, получаем:

$$f_{T,N}(s) = 1 + \sum_{k=1}^{N} \frac{N^{[k]}}{k!} \left(1 - \frac{k}{N}\right)^T (s-1)^k = \frac{1}{N^T} \sum_{k=0}^{N} C_N^k (s-1)^k (N-k)^T.$$

Достаточно доказать, что действительны все корни многочлена

$$R(z) = \sum_{k=0}^{N} C_N^k z^k (N-k)^T,$$

или (поскольку $R(0) = N^T f_{T,N}(1) \neq 0$) что действительны все корни многочлена

$$z^N R\left(\frac{1}{z}\right) = \sum_{k=0}^{N} C_N^k z^{N-k} (N-k)^T.$$

Введем оператор D, действующий на произвольную дифференцируемую функцию h по формуле

$$Dh(z) = zh'(z).$$

Оператор D переводит многочлен в многочлен той же степени, и если $h(z) = (1+z)^N = \sum_{k=0}^{N} C_N^k z^{N-k}$, то

$$D(1+z)^N = z((1+z)^N)'_z = z\sum_{k=0}^{N} C_N^k z^{N-k-1}(N-k) = \sum_{k=0}^{N} C_N^k z^{N-k}(N-k).$$

По индукции находим:

$$D^T(1+z)^N = \sum_{k=0}^{N} C_N^k z^{N-k}(N-k)^T = z^N R\left(\frac{1}{z}\right).$$

Легко проверить, что если все корни многочлена $f(z)$ действительны, то действительны все корни многочлена $zf'(z)$. Остается заметить, что у многочлена $(1+z)^N$ все корни действительны и поэтому действительны все корни многочлена $D^T(1+z)^N$. Лемма 3 доказана.

Лемма 4. *Если выполнены условия теоремы, то*

$$\mathbf{D}\mu_0(T,N) = N(N-1)\left(1-\frac{2}{N}\right)^T - N^2\left(1-\frac{1}{N}\right)^{2T} + N\left(1-\frac{1}{N}\right)^T \to \infty.$$

Доказательство леммы 4. Точная формула для $\mathbf{D}\mu_0(T,N)$ следует из того, что в равновероятной схеме размещения частиц по ячейкам

$$\mathbf{E}\mu_0(T,N) = N\left(1-\frac{1}{N}\right)^T,$$

$$\mathbf{E}\mu_0^{[2]}(T,N) = N(N-1)\left(1-\frac{2}{N}\right)^T$$

и что $\mathbf{D}\mu_0(T,N) = \mathbf{E}\mu_0^{[2]}(T,N) + \mathbf{E}\mu_0(T,N) - (\mathbf{E}\mu_0(T,N))^2$. Для доказательства соотношения $\mathbf{D}\mu_0(T,N) \to \infty$ преобразуем указанное в лемме выражение для дисперсии:

$$\mathbf{D}\mu_0(T,N) =$$

$$= \left(1 - \frac{2}{N}\right)^T \left[N^2\left(1 - \left(1 + \frac{1}{N(N-2)}\right)^T\right) + N\left(\left(1 + \frac{1}{N-2}\right)^T - 1\right)\right].$$

В случае, когда $\frac{T}{N} \to C \in (0,\infty)$, доказательство того, что $\mathbf{D}\mu_0(T,N) \to \infty$, несложно:

$$\mathbf{D}\mu_0(T,N) \sim e^{-2C}\left[N^2\left(1 - \exp\left\{\frac{C}{N-2}\right\}\right) + N(e^C - 1)\right] \sim$$

$$\sim Ne^{-2C}\left[e^C - 1 - \frac{NC}{N-2}\right] \to \infty,$$

$$\mathbf{D}\mu_0(T,N) \sim e^{-2C}\left[N^2\left(1 - \left(1 + \frac{T}{N(N-2)} + O\left(\frac{T^2}{N^4}\right)\right)\right) + N(e^C - 1)\right]$$

поскольку $e^C > 1 + C$ при $C > 0$.

Лемма доказана.

Из лемм 1 – 4 следует утверждение теоремы.

Замечание. С помощью немного более сложных оценок можно доказать, что лемма 4 и теорема справедливы при более широких условиях $\min\{\frac{T^2}{N}, \ln N - \frac{T}{N}\} \to \infty$ и что поэтому в случае равновероятной схемы размещения число $\mu_0(T,N)$ пустых ячеек при $T,N \to \infty$ (и соответствующих центрировке и нормировке) может иметь в качестве предельных только нормальное или пуассоновские распределения.

§ 10. Неравенства для вероятностей больших уклонений

Большим уклонением случайной величины ξ обычно называют попадание ее значения в область, которая удалена от $\mathbf{E}\xi$ на такое расстояние, что вероятность попадания в эту область близка к нулю. Необходимость исследования и оценивания вероятностей больших уклонения возникает в математической статистике при построении критериев с малой вероятностью ошибки.

Иногда оценки вероятностей больших уклонений удается получать в виде неравенств, чаще — в виде асимптотических формул с оценкой порядка остаточного члена.

Примером оценки вероятностей больших уклонений в виде неравенств является неравенство Хефдинга.

Теорема 1 (Неравенство Хефдинга). *Пусть* $S_n = X_1 + \ldots + X_n$, *где* $X_1, \ldots, X_n$ *независимы и* $\mathbf{P}\{0 \leq X_k \leq 1\} = 1$ *для всех* $k = 1, \ldots, n$. *Положим* $\mathbf{E}S_n = \mu n$. *Тогда при любом* $z \in [0, 1-\mu)$

$$\mathbf{P}\left\{\frac{1}{n}S_n - \mu \geq z\right\} \leq \left(\left(\frac{\mu}{\mu+z}\right)^{\mu_+z}\left(\frac{1-\mu}{1-\mu-z}\right)^{1-\mu-z}\right)^n \leq e^{-nz^2 g(\mu)} \leq e^{-2nz^2}, \tag{59}$$

где

$$\frac{1}{1-2\mu}\ln\frac{1-\mu}{\mu} \text{ при } \mu \in \left(0, \frac{1}{2}\right), \quad g(\mu) = \frac{1}{2\mu(1-\mu)} \text{ при } \mu \in \left[\frac{1}{2}, 1\right).$$

Лемма 1. *Пусть случайная величина* ν *такова, что* $\mathbf{E}e^{t\nu} < \infty$ *на полуинтервале* $[0, b), 0 < b \leq \infty$. *Тогда при любом* $t \in [0, b)$

$$\mathbf{P}\{\nu \geq x\} \leq e^{-tx}\mathbf{E}e^{t\nu}. \tag{60}$$

Доказательство. Так как $e^{tx} > 0$ при всех действительных x и t и $e^{t(\nu - x)} \geqslant 1$ при $t > 0$ и $\nu \geqslant x$, то

$$\mathbf{E}e^{t\nu} \geqslant e^{tx}\mathbf{E}e^{t(\nu-x)}\chi\{\nu \geqslant x\} \geqslant e^{tx}\mathbf{E}\chi\{\nu \geqslant x\} = e^{tx}\mathbf{P}\{\nu \geqslant x\}.$$

Умножив обе части неравенства на e^{-tx}, получим (7).

Лемма 2. *Если* $\mathbf{P}\{a \leqslant X \leqslant b\} = 1$, *то для любого действительного* h

$$\mathbf{E}e^{hX} \leqslant \frac{b - \mathbf{E}X}{b-a}e^{ha} + \frac{\mathbf{E}X - a}{b-a}e^{hb}. \tag{61}$$

Доказательство. Так как функция e^{hx} выпукла, то на любом отрезке $[a, b]$ ее график лежит под хордой, соединяющей значения функции на концах отрезка, т. е.

$$e^{hX} \leqslant \frac{b-X}{b-a}\,e^{ha} + \frac{X-a}{b-a}e^{hb} \text{ при любом значении } X \in [a, b].$$

Вычисляя математические ожидания от обеих частей неравенства, получаем (9).

Доказательство неравенства Хефдинга. Согласно лемме 1 и свойствам мультипликативности математического ожидания

$$\mathbf{P}\left\{\frac{1}{n}\,S_n-\mu\geq z\right\}=\mathbf{P}\{S_n\geqslant n(\mu+z)\}\leqslant$$
$$\leqslant e^{-hn(\mu+z)}\mathbf{E}e^{hS_n}=e^{-hn(\mu+z)}\mathbf{E}e^{h(X_1+\ldots+X_n)}=e^{-hn(\mu+z)}\prod_{k=1}^{n}\mathbf{E}e^{hX_k}.$$

По условию $0\leqslant X_k\leqslant 1$. Положим $\mu_k=\mathbf{E}X_k$. Тогда $n\mu=\mu_1+\ldots+\mu_n$. По лемме 2 с $[a,b]=[0,1]$ находим:

$$\prod_{k=1}^{n}\mathbf{E}e^{hX_k}\leqslant\prod(1-\mu_k+\mu_k e^h).$$

Так как среднее геометрическое не больше среднего арифметического, то

$$\prod_{k=1}^{n}(1-\mu_k+\mu_k e^h)^{1/n}\leqslant\frac{1}{n}\sum_{k=1}^{n}(1-\mu_k+\mu_k e^h)=1-\mu+\mu e^h.$$

Следовательно,

$$\mathbf{P}\left\{\frac{1}{n}\,S_n-\mu\geqslant z\right\}\leqslant\left\{e^{-h(\mu+z)}(1-\mu+\mu e^h)\right\}^n. \tag{62}$$

Так как

$$\frac{\partial}{\partial h}\left\{e^{-h(\mu+z)}(1-\mu+\mu e^h)\right\}=e^{-h(\mu+z)}(\mu e^h-(\mu+z)(1-\mu+\mu e^h)),$$

то минимум правой части (62) достигается при

$$e^h=\frac{(1-\mu)(\mu+z)}{(1-\mu-z)\mu}>\frac{(1-\mu)\mu}{(1-\mu)\mu}=1,\quad\text{т.е. при}\quad h=\ln\frac{(1-\mu)(\mu+z)}{(1-\mu-z)\mu}>0.$$

Подстановка этих значений в (62) завершает доказательство первого неравенства теоремы:

$$\mathbf{P}\left\{\frac{1}{n}\,S_n-\mu\geqslant z\right\}\leqslant\left\{\frac{(1-\mu)(\mu+z)}{(1-\mu-z)\mu}\right\}^{(\mu+z)n}\left\{\frac{1-\mu}{1-\mu-z}\right\}^n=$$
$$=\left\{\left(\frac{\mu}{\mu+z}\right)^{\mu+z}\left(\frac{1-\mu}{1-\mu-z}\right)^{1-\mu-z}\right\}^n.$$

Неравенство

$$\left\{\left(\frac{\mu}{\mu+z}\right)^{\mu+z}\left(\frac{1-\mu}{1-\mu-z}\right)^{1-\mu-z}\right\}^n \leqslant e^{-2nz^2}$$

при $0 \leqslant z < 1-\mu$ эквивалентно неравенству

$$(\mu+z)\ln\frac{\mu}{\mu+z} + (1-\mu-z)\ln\frac{1-\mu}{1-\mu-z} \leqslant -2z^2.$$

При $z = 0$ последнее неравенство выполнено; поэтому для его доказательства достаточно показать, что производная по z левой части не больше производной правой:

$$\ln\frac{\mu}{\mu+z} - \frac{\mu+z}{\mu+z} - \ln\frac{1-\mu}{1-\mu-z} + \frac{1-\mu-z}{1-\mu-z} = \ln\frac{\mu}{\mu+z} - \ln\frac{1-\mu}{1-\mu-z} \leqslant -4z.$$

При $z = 0$ последнее неравенство выполнено; поэтому для его доказательства достаточно показать, что производная по z левой части не больше производной правой:

$$-\frac{1}{\mu+z} - \frac{1}{1-\mu-z} = -\frac{1}{(\mu+z)(1-\mu-z)} \leqslant -4.$$

Последнее неравенство выполняется при всех z, так как $u(1-u) \leqslant \frac{1}{4}$ при всех $u \in \mathbb{R}$.

Неравенство Хёфдинга – Азумы

Для случайной величины X, принимающей действительные значения, будем использовать обозначение $\|X\|_\infty = \inf\{c : \mathbf{P}\{|X| \leqslant c\} = 1\}$. Если $\|X\|_\infty < \infty$, то случайная величина X ограничена.

Теорема (неравенство Хёфдинга – Азумы). *Если $X_1, X_2, \ldots, X_n$ — ограниченные случайные величины и*

$$\mathbf{E}X_{i_1}\ldots X_{i_k} = 0 \quad \textit{для всех} \quad k \geqslant 1,\ 1 \leqslant i_1 < \ldots < i_k \leqslant n, \tag{63}$$

то

$$\mathbf{P}\left\{\sum\nolimits_{j=1}^n X_j \geqslant z\right\} \leqslant \exp\left\{-\frac{z^2}{2\sum_{j=1}^n \|X_j\|_\infty^2}\right\}.$$

Замечание. Условие (63) выполняется для независимых ограниченных случайных величин $X_1, X_2, \ldots, X_n$ с нулевыми математическими ожиданиями, и к некоторым функциям от них, например, к случайным величинам

$Y_1 = X_1, Y_2 = X_1X_2, Y_3 = X_2X_3, \ldots, Y_n = X_{n-1}X_n$: если $1 \leqslant i_1 < \ldots < i_k \leqslant n$, то при любом целом $k > 1$

$$\begin{gathered}\mathbf{E}Y_{i_1}\ldots Y_{i_k} = \mathbf{E}Y_{i_1}(X_{i_2-1}X_{i_2})\ldots(X_{i_k-1}X_{i_k}) = \\ = \mathbf{E}(Y_{i_1}(X_{i_2-1}X_{i_2})\ldots Y_{i_{k-1}}X_{i_k-1})\mathbf{E}X_{i_k} = 0.\end{gathered}$$

В более общем случае условие (63) выполняется, если $\mathbf{E}X_1 = 0$ и для любого целого $k > 1$

$$\mathbf{E}\{X_k \,|\, X_1, \ldots, X_{k-1}\} = 0,$$

так как по свойствам условного математического ожидания при $1 \leqslant i_1 < \ldots < i_k \leqslant n$

$$\begin{gathered}\mathbf{E}X_{i_1}\ldots X_{i_k} = \mathbf{E}\{\mathbf{E}\{X_{i_1}\ldots X_{i_k} \,|\, X_1, \ldots, X_{i_k-1}\}\} = \\ = \mathbf{E}\{X_{i_1}\ldots X_{i_{k-1}}\mathbf{E}\{X_{i_k} \,|\, X_1, \ldots, X_{i_k-1}\}\} = 0.\end{gathered}$$

Доказательство. Гиперболические синус и косинус определяются равенствами

$$\cosh(x) = \frac{e^x + e^{-x}}{2}, \quad \sinh(x) = \frac{e^x - e^{-x}}{2}.$$

Так как e^x — выпуклая функция, то при $x \in [-1, 1]$

$$\begin{gathered}e^{ax} = \exp\left\{a\left(\frac{1+x}{2}\right) - a\left(\frac{1-x}{2}\right)\right\} \leqslant \\ \leqslant \frac{1+x}{2}\,e^{a} + \frac{1-x}{2}\,e^{-a} = \cosh(a) + x\sinh(a).\end{gathered} \tag{64}$$

Из условия теоремы следует, что при любых константах a_i, b_i, $i = 1, \ldots, n$

$$\mathbf{E}\prod_{j=1}^{n}(b_jX_j + a_j) = \prod_{j=1}^{n} a_j, \tag{65}$$

так как математическое ожидание каждого отличного от константы одночлена равно 0. Положим $Y_j = \frac{X_j}{\|X_i\|_\infty}$ и $a = t\|X_j\|_\infty$. Из (64) следует, что

$$e^{tX_j} = e^{aY_j} \leqslant \cosh(a) + Y_j\sinh(a) = \cosh(t\|X_j\|_\infty) + X_j\,\frac{\sinh(t\|X_j\|_\infty)}{\|X_j\|_\infty}.$$

Далее, используя (64) и (65), находим:

$$\begin{gathered}\mathbf{E}\exp\left\{t\sum\nolimits_{j=1}^{n} X_j\right\} = \mathbf{E}\prod\nolimits_{j=1}^{n} e^{tX_j} \leqslant \\ \leqslant \mathbf{E}\prod\nolimits_{j=1}^{n}\left(\cosh(t\|X_j\|_\infty) + X_j\,\frac{\sinh(t\|X_j\|_\infty)}{\|X_j\|_\infty}\right) = \\ = \prod\nolimits_{j=1}^{n}\cosh(t\|X_j\|_\infty) \leqslant \exp\left\{\frac{t^2}{2}\sum\nolimits_{j=1}^{n}\|X_j\|_\infty^2\right\},\end{gathered} \tag{66}$$

здесь в последнем переходе использована оценка

$$\cosh(x) = \sum\nolimits_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} \leqslant \sum\nolimits_{k=0}^{\infty} \frac{x^{2k}}{2^k k!} = e^{x^2/2}. \tag{67}$$

Теперь можно завершить доказательство теоремы с помощью неравенства Маркова:

$$\mathbf{P}\left\{\sum\nolimits_{j=1}^{n} X_j \geqslant z\right\} = \mathbf{P}\left\{\exp\left\{t\sum\nolimits_{j=1}^{n} X_j\right\} \geqslant e^{tz}\right\} \leqslant$$

$$\leqslant \frac{1}{e^{tz}}\mathbf{E}\exp\left\{t\sum\nolimits_{j=1}^{n} X_j\right\} \leqslant \exp\left\{\frac{t^2}{2}\sum\nolimits_{j=1}^{n} \|X_j\|_\infty^2 - tz\right\} =$$

$$= \exp\left\{-\frac{z^2}{2\sum_{j=1}^{n}\|X_j\|_\infty^2}\right\}, \quad \text{если положить} \quad t = \frac{z}{\sum_{j=1}^{n}\|X_j\|_\infty^2}.$$

§11. Вероятности больших уклонений

Общую теорему о вероятностях больших уклонений рассмотрим в наиболее простой форме, в доказательстве которой основные идеи не загромождаются техническими деталями.

Теорема 2. *Пусть случайные величины $X_1, X_2, \ldots$ независимы, одинаково распределены и удовлетворяют условию*

$$\varphi(h) = \mathbf{M}e^{hX_1} < \infty \quad \textit{для всех} \quad h \in (-\infty, \infty). \tag{68}$$

Пусть $S_n = X_1 + ... + X_n$. Тогда для любого $a > \mathbf{E}X_1$

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbf{P}\left\{S_n \geq an\right\} = -I(a),$$

где

$$I(z) = \sup_{h\in R}\left\{zh - \ln\varphi(h)\right\}.$$

Доказательство. Без ограничения общности можно считать, что $a = 0$ и, значит, $\mathbf{E}X_1 < 0$. Действительно, при $a \neq 0$ можно рассмотреть случайные величины $X_1^* = X_1 - a, X_2^* = X_2 - a, \ldots$, для которых $\mathbf{E}X_1^* < 0$,

$$\{S_n \geqslant an\} = \{X_1^* + \ldots + X_n^* \geqslant 0\}$$

и $\varphi^*(h) = \mathbf{M}e^{hX_1^*} = \mathbf{M}e^{h(X_1-a)} = e^{-ah}\mathbf{M}e^{hX_1}$, так что $\ln\varphi^*(h) = \ln\varphi(h) - ah$; тогда

$$I(a) = \sup_h\{ah - \ln\varphi(h)\} = \sup_h\{-\ln\varphi^*(h)\} = I^*(0).$$

Мы будем предполагать далее, что распределение X_1 невырождено, поскольку в противном случае вырождены и распределение S_n, и функция $I(z)$. (Если $\mathbf{P}\{X_1 = b\} = 1$, то $\varphi(h) = e^{hb}$, $I(b) = 0, I(z) = \infty$ при $z \neq b$ и $\mathbf{P}\{S_n > an\} = 0$ при $a > b$.) Положим

$$\rho = \inf\{\varphi(h)\colon h \in \mathbb{R}\} \geqslant 0.$$

Тогда $I(0) = -\ln \rho$ ($I(0) = \infty$, если $\rho = 0$), так что нам нужно доказать, что

$$\limsup_{n\to\infty} \tfrac{1}{n} \ln \mathbf{P}\{S_n \geq 0\} = \ln \rho. \qquad (69)$$

Пусть $F(x) = \mathbf{P}\{X_1 \leq x\}$ — функция распределения X_1. Из условия (68) следует, что функция φ бесконечно дифференцируема (как интеграл от бесконечно дифференцируемых функций) и

$$\varphi'(h) = \int\limits_R x e^{hx} dF(x), \quad \varphi''(h) = \int\limits_R x^2 e^{hx} dF(x).$$

Следовательно, $\varphi''(h) > 0$, т. е. φ строго выпукла, и $\varphi'(0) = \mathbf{E}X_1 < 0$.

Рассмотрим три случая в соответствии с носителем распределения X_1.

а) $\mathbf{P}\{X_1 < 0\} = 1$. Тогда $\varphi'(h) = \mathbf{E}X_1 e^{hX_1} < 0$, т. е. φ строго убывает и

$$\rho = \lim_{h\to\infty} \varphi(h) = \lim_{h\to\infty} \mathbf{E}e^{hX_1} = \lim_{h\to\infty} \int_{-\infty}^{0-} e^{hy} d\mathbf{P}\{X_1 \leqslant y\} = 0.$$

Так как $\mathbf{P}\{S_n \geq 0\} = 0$, то в этом случае (69) выполнено.

б) $\mathbf{P}\{X_1 \leq 0\} = 1$ и $\mathbf{P}\{X_1 = 0\} \in (0, 1)$. Тогда $\varphi'(h) = \mathbf{E}X_1 e^{hX_1} < 0$, φ опять строго убывает и $\rho = \lim\limits_{h\to\infty} \varphi(h) = \lim\limits_{h\to\infty} \mathbf{E}e^{hX_1} = \mathbf{P}\{X_1 = 0\} > 0$. Так как

$$\mathbf{P}\{S_n \geq 0\} = \mathbf{P}\{X_1 = ... = X_n = 0\} = \rho^n,$$

то в этом случае (69) тоже выполняется.

в) $\mathbf{P}\{X_1 < 0\} > 0$ и $\mathbf{P}\{X_1 > 0\} > 0$. Тогда

$$\lim_{|h|\to\infty} \varphi(h) = \lim_{|h|\to\infty} (\mathbf{E}e^{hX_1}\mathbb{I}\{X_1 \leqslant 0\} + \mathbf{E}e^{hX_1}\mathbb{I}\{X_1 > 0\}) = \infty,$$

и, поскольку $\varphi(h)$ строго выпукла, существует единственная точка минимума $\beta \in (0, \infty)$, в которой

$$\varphi(\beta) = \rho, \quad \varphi'(\beta) = 0. \qquad (70)$$

Тогда
$$\mathbf{P}\{S_n \geqslant 0\} = \mathbf{E}\mathbb{I}\{\beta S_n \geqslant 0\} \leqslant \mathbf{E}e^{\beta S_n} = [\varphi(\beta)]^n = \rho^n.$$
Следовательно,
$$\limsup_{n\to\infty} \tfrac{1}{n} \ln \mathbf{P}\left\{S_n \geq 0\right\} \leq \ln \rho.$$

При выводе нижней оценки используется полезный прием (так называемая замена меры). Введем новую последовательность независимых одинаково распределенных случайных величин $\hat{X}_n$ с функциями распределения
$$\hat{F}(x) = \frac{1}{\rho}\int\limits_{-\infty}^{x} e^{\beta u} dF(u).$$
Функцию $\hat{F}(x)$ называют *преобразованием Крамера* функции F. Заметим, что
$$\rho = \varphi(\beta) = \int\limits_{-\infty}^{\infty} e^{\beta u} dF(u).$$
Докажем три леммы.

Лемма 3. *Справедливы соотношения*
$$\mathbf{E}\hat{X}_1 = \hat{\mu} = 0, \quad \mathbf{D}\hat{X}_1 = \hat{\sigma}^2 \in (0, \infty).$$

Доказательство. Пусть $\hat{\varphi}(h) = \mathbf{E}e^{h\hat{X}_1}$. Тогда при любом $h \in R$
$$\hat{\varphi}(h) = \int\limits_{R} e^{hx} d\hat{F}(x) = \frac{1}{\rho}\int\limits_{R} e^{hx} e^{\beta x} dF(x) = \frac{1}{\rho}\,\varphi(h+\beta) < \infty. \tag{71}$$
Значит, $\hat{\varphi}$ тоже бесконечно дифференцируема и в силу (71)
$$\mathbf{E}\hat{X}_1 = \hat{\varphi}'(0) = \frac{1}{\rho}\,\varphi'(\beta) = 0, \quad \mathbf{D}\hat{X}_1 = \hat{\varphi}''(0) = \frac{1}{\rho}\,\varphi''(\beta) \in (0, \infty).$$

Лемма 4. *Если*
$$\hat{S}_n = \sum_{k=1}^{n} \hat{X}_k,$$
то
$$\mathbf{P}\left\{S_n \geqslant 0\right\} = \rho^n \mathbf{E}\left\{e^{-\beta \hat{S}_n} \chi\left\{\hat{S}_n \geqslant 0\right\}\right\}.$$

Доказательство. Имеем
$$d\hat{F}(x) = \rho^{-1} e^{\beta x} dF(x)$$

и

$$\mathbf{P}\left\{S_n \geqslant 0\right\} = \int\limits_{\{x_1+\ldots+x_n\geqslant 0\}} dF(x_1)\ldots dF(x_n) =$$

$$= \int\limits_{x_1+\ldots+x_n\geqslant 0} \left[\rho e^{-\beta x_1} d\hat{F}(x_1)\right]\ldots\left[\rho e^{-\beta x_n} d\hat{F}(x_n)\right] =$$

$$= \rho^n \int\limits_{x_1+\ldots+x_n\geqslant 0} e^{-\beta(x_1+\ldots+x_n)} d\hat{F}(x_1)\ldots d\hat{F}(x_n) = \rho^n \mathbf{E} e^{-\beta\hat{S}_n}\mathbb{I}\{\hat{S}_n \geqslant 0\},$$

что совпадает с утверждением леммы 4.

Лемма 5. *Справедливо неравенство*

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mathbf{E}\left\{e^{-\beta\hat{S}_n}\chi\left\{\hat{S}_n \geqslant 0\right\}\right\} \geqslant 0.$$

Доказательство. Из леммы 3 следует, что для $\hat{S}_n$ справедлива центральная предельная теорема:

$$\mathbf{P}\left\{\hat{S}_n \leq x\hat{\sigma}\sqrt{n}\right\} = \mathbf{P}\left\{\frac{\hat{S}_n}{\hat{\sigma}\sqrt{n}} \leq x\right\} \to \Phi(x) \quad \forall x \in R.$$

Выберем число $C > 0$ так, чтобы

$$\Phi(C) - \Phi(0) = \frac{1}{\sqrt{2\pi}} \int\limits_0^C e^{-x^2/2} dx = \frac{1}{4};$$

тогда

$$\mathbf{E}\left\{e^{-\beta\hat{S}_n}\chi\left\{\hat{S}_n \geqslant 0\right\}\right\} \geqslant \mathbf{E}\left\{e^{-\beta\hat{S}_n}\mathbb{I}\left\{\frac{\hat{S}_n}{\hat{\sigma}\sqrt{n}} \in [0,\, C]\right\}\right\} \geqslant$$

$$\geqslant \mathbf{E}\left\{e^{-\beta C\hat{\sigma}\sqrt{n}}\mathbb{I}\left\{\frac{\hat{S}_n}{\hat{\sigma}\sqrt{n}} \in [0,\, C]\right\}\right\} = \mathbf{P}\left\{\frac{\hat{S}_n}{\hat{\sigma}\sqrt{n}} \in [0, C]\right\} e^{-\beta C\hat{\sigma}\sqrt{n}}.$$

При достаточно больших n вероятность в правой части близка к $\frac{1}{4}$, поэтому

$$\ln \mathbf{E}\left\{e^{-\beta\hat{S}_n}\chi\left\{\hat{S}_n \geqslant 0\right\}\right\} \geqslant C' - \beta C\hat{\sigma}\sqrt{n}, \quad C' = \ln \mathbf{P}\left\{\frac{\hat{S}_n}{\hat{\sigma}\sqrt{n}} \in [0, C]\right\} \to \ln\frac{1}{4},$$

значит,

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mathbf{E}\left\{e^{-\beta \hat{S}_n}\chi\left\{\hat{S}_n \geqslant 0\right\}\right\} \geqslant \liminf_{n\to\infty} \frac{C' - \beta C\hat{\sigma}\sqrt{n}}{n} \geqslant 0,$$

и лемма 5 доказана.

Объединяя утверждения лемм 4 и 5, получаем:

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbf{P}\left\{S_n \geqslant 0\right\} \geqslant \liminf_{n\to\infty} \frac{1}{n} \ln \rho^n \mathbf{E}\left\{e^{-\beta \hat{S}_n}\chi\left\{\hat{S}_n \geqslant 0\right\}\right\} \geqslant \ln \rho.$$

Теорема доказана.

§12. Расстояние по вариации между распределениями и его применения

В теории вероятностей все построения проводятся на основном вероятностном пространстве $(\Omega, \mathcal{F}, P)$; в частности, случайные величины – это измеримые функции, отображающие пространство элементарных событий Ω в другое измеримое пространство. Естественно возникает вопрос о мере близости случайных величин. Такие меры можно вводить разными способами. Например, самое простое расстояние между случайными величинами X и Y – это индикаторная метрика

$$\iota(X,Y) = \mathbf{E}\chi\{X \neq Y\} = \mathbf{P}\{X \neq Y\}.$$

Она обладает основными свойствами метрики:

$$\begin{gathered} \iota(X,Y) = 0 \Leftrightarrow \mathbf{P}\{X = Y\} = 1, \\ \iota(X,Y) = \iota(Y,X), \\ \iota(X,Y) \leqslant \iota(X,Z) + \iota(Z,Y). \end{gathered}$$

Первые два свойства очевидны, а последнее является следствием соотношения $\{X \neq Y\} \subset \{X \neq Z\} \cup \{Z \neq Y\}$.

Другим примером метрики в пространстве случайных величин является метрика $\mu(X,Y) = \mathbf{E}|X - Y|$. Для нее свойства метрики тоже легко проверяются, в частности, неравенство треугольника следует из того, что $|X - Y| \leqslant |X - Z| + |Z - Y|$.

Как правило, в теории вероятностей интересуются только распределениями случайных величин, и поэтому чаще всего сравниваются не сами случайные величины, а их распределения.

Один из важных видов расстояния между распределениями (который применим к случайным величинам, принимающим значения произвольной природы) – это расстояние по вариации:

$$\rho(X,Y) = \rho(P_X, P_Y) = \sup_A |P_X(A) - P_Y(A)| = \sup_A \left| \int_A P_X(d\omega) - \int_A P_Y(d\omega) \right|,$$

где sup берется по всем измеримым подмножествам A пространства B. Свойства расстояния для этого функционала тоже легко проверяются:

$$\rho(X,Y) = 0 \Leftrightarrow P_X = P_Y,$$
$$\rho(X,Y) = \rho(Y,X),$$
$$\rho(X,Y) = \sup_A |P_X(A) - P_Y(A)| = \sup_A |P_X(A) - P_Z(A) + P_Z(A) - P_Y(A)| \leqslant$$
$$\leqslant \sup_A |P_X(A) - P_Z(A)| + \sup_A |P_Z(A) - P_Y(A)| = \rho(X,Z) + \rho(Z,Y).$$

Лемма 1. *Справедливо равенство*

$$\rho(X,Y) = \frac{1}{2} \int\limits_B |P_X(dx) - P_Y(dx)|,$$

которое в случае счетного множества B принимает вид

$$\rho(X,Y) = \frac{1}{2} \sum_{b \in B} |P_X(b) - P_Y(b)|,$$

а в случае, когда меры P_X и P_Y абсолютно непрерывны с плотностями $p_X(x)$, $p_Y(x)$, $x \in B$, – вид

$$\rho(X,Y) = \frac{1}{2} \int\limits_B |p_X(x) - p_Y(x)| dx.$$

Доказательство. По теореме Хана о знакопеременных мерах существует такое измеримое множество $A_0 \subset B$, что

$$P_X(A) - P_Y(A) \geqslant 0 \quad \text{для любого измеримого множества} \quad A \subseteq A_0$$

и

$$P_X(A) - P_Y(A) \leqslant 0 \quad \text{для любого измеримого множества} \quad A \subseteq B \backslash A_0.$$

Тогда для любого измеримого множества A

$$P_X(A) - P_Y(A) = \int\limits_{A \cap A_0} (P_X(dx) - P_Y(dx)) + \int\limits_{A \setminus A_0} (P_X(dx) - P_Y(dx)).$$

Подынтегральная функция в первом интеграле неотрицательна, а во втором неположительна, поэтому максимальное значение $P_X(A) - P_Y(A)$ достигается при $A = A_0$, а минимальное – при $A = B \setminus A_0$. Так как

$$P_X(A_0) - P_Y(A_0) = \int\limits_{A_0} (P_X(dx) - P_Y(dx)) = \int\limits_{A_0} |P_X(dx) - P_Y(dx)| =$$

$$= P_Y(B \setminus A_0) - P_X(B \setminus A_0) = \int\limits_{B \setminus A_0} (P_X(dx) - P_Y(dx)) = \int\limits_{B \setminus A_0} |P_X(dx) - P_Y(dx)|,$$

то

$$\int\limits_{A_0} |P_X(dx) - P_Y(dx)| = \int\limits_{B \setminus A_0} |P_X(dx) - P_Y(dx)| = \frac{1}{2} \int\limits_{B} |P_X(dx) - P_Y(dx)|,$$

т. е.

$$\sup_A |P_X(A) - P_Y(A)| = \frac{1}{2} \int\limits_{B} |P_X(dx) - P_Y(dx)|.$$

В общем случае не обязательно дискретных мер множество A_0 следует определять как такое измеримое подмножество пространства B, что

$$P_X(A) - P_Y(A) \geqslant 0 \quad \text{для любого измеримого множества} \quad A \subseteq A_0$$

и

$$P_X(A) - P_Y(A) \leqslant 0 \quad \text{для любого измеримого множества} \quad A \subseteq B \setminus A_0.$$

Существование такого множества A_0 следует из теоремы Хана – Банаха о разложении счетно-аддитивной меры на произвольном измеримом пространстве на положительную и отрицательную части (см., например, [1]).

Замечание. Из леммы 1 следует, что для любых случайных величин X, Y, принимающих действительные значения,

$$\sup_{t \in R} |\mathbf{P}\{X \leqslant t\} - \mathbf{P}\{Y \leqslant t\}| \leqslant \rho(X, Y)$$

и что расстояние по вариации между дискретным и абсолютно непрерывным распределениями принимает максимально возможное значение, равное 1.

Лемма 2. *Если $f : B \to C$ — измеримая функция, то*

$$\rho(f(X), f(Y)) \leqslant \rho(X, Y).$$

Доказательство. Имеем:

$$\rho(f(X), f(Y)) = \sup_{A \subseteq C} |\mathbf{P}\{f(X) \in A\} - \mathbf{P}\{f(Y) \in A\}| =$$
$$= \sup_{A \subseteq C} |\mathbf{P}\{X \in f^{-1}(A)\} - \mathbf{P}\{Y \in f^{-1}(A)\}| \leqslant$$
$$\leqslant \sup_{D \subseteq B} |\mathbf{P}\{X \in D\} - \mathbf{P}\{Y \in D\}| = \rho(X, Y).$$

Лемма 3. *Пусть X, Y – независимые случайные величины и X', Y' – тоже независимые случайные величины, принимающие значения в одном и том же пространстве B. Тогда*

$$\rho((X, Y), (X', Y')) \leqslant \rho(X, X') + \rho(Y, Y').$$

Доказательство. Рассмотрим вспомогательную пару независимых случайных величин (X^0, Y^0), в которой X^0 имеет такое же распределение, как X', а Y^0 – такое же распределение, как Y. Тогда в силу неравенства треугольника для расстояния по вариации и предположений о независимости случайных величин

$$\rho((X, Y), (X', Y')) \leqslant \rho((X, Y), (X^0, Y^0)) + \rho((X^0, Y^0), (X', Y')) =$$
$$= \frac{1}{2} \left[\int\limits_B |P_{(X,Y)}(dw) - P_{(X^0,Y^0)}(dw)| + \int\limits_B |P_{(X^0,Y^0)}(dw) - P_{(X',Y')}(dw)| \right] =$$
$$= \frac{1}{2} \left[\int\limits_B P_Y(dy) |P_X(dx) - P_{X'}(dx)| + \int\limits_B P_{X'}(dx) |P_Y(dy) - P_{Y'}(dy)| \right] =$$
$$\leqslant \rho(X, X') + \rho(Y, Y').$$

Замечание 1. Приведем пример пар зависимых случайных величин (X, Y) и (X', Y'), для которых утверждение леммы 3 неверно. Действительно, пусть X имеет стандартное нормальное распределение; тогда

$$1 = \rho((X, X), (X, -X)) > \rho(X, X) + \rho(X, -X) = 0.$$

Замечание 2. Утверждение леммы 3 по индукции распространяется на любое конечное число пар независимых случайных величин.

Лемма 4. *Пусть X и Y — случайные величины, принимающие значения в одном и том же измеримом пространстве, $Q(X,Y)$ — совокупность всех таких пар случайных величин (X',Y'), определенных на вероятностном пространстве $(\Omega,\mathcal{F},\mathbf{P})$ с мерой $\mathbf{P}$, не имеющей атомов, что X' имеет такое же распределение, как X, а Y' – такое же распределение, как Y. Тогда*

$$\rho(X,Y)=\inf_{(X',Y')\in Q(X,Y)}\mathbf{P}\{X'\neq Y'\}.$$

Доказательство. Сначала покажем, что $\rho(X,Y)\leqslant\mathbf{P}\{X'\neq Y'\}$ для любой пары $(X',Y')\in Q(X,Y)$. Для любого измеримого множества $A\subset B$

$$\begin{gathered}|\mathbf{P}\{X\in A\}-\mathbf{P}\{Y\in A\}|=|\mathbf{P}\{X'\in A\}-\mathbf{P}\{Y'\in A\}|=\\ =|\mathbf{P}\{X'\in A,\,Y'=X'\}+\mathbf{P}\{X'\in A,\,Y'\neq X'\}-\\ -\mathbf{P}\{Y'\in A,\,Y'=X'\}-\mathbf{P}\{Y'\in A,\,Y'\neq X'\}|=\\ =|\mathbf{P}\{X'\in A,Y'\neq X'\}-\mathbf{P}\{Y'\in A,Y'\neq X'\}|\leqslant\mathbf{P}\{Y'\neq X'\},\end{gathered}$$

так как $\mathbf{P}\{X'\in A,\,Y'=X'\}=\mathbf{P}\{Y'\in A,\,Y'=X'\}$. Следовательно, $\rho(X,Y)\leqslant\mathbf{P}\{Y'\neq X'\}$ при любых $(X',Y')\in Q(X,Y)$.

Покажем, как можно построить такую пару случайныз величин $(X',Y')\in Q(X,Y)$, для которой в последнем неравенстве имеет место знак равенства. Чтобы не усложнять рассуждения, рассмотрим только случай, когда множество значений случайных величин X и Y дискретно; тогда можно считать, что оно совпадает с $S=\{0,1,\ldots\}$. В этом случае меры P_X и P_Y определяются последовательностями вероятностей $P_X(k)$, $P_Y(k)$, $k\in S$. Положим

$$\mu_{XY}(k)=\min\{P_X(k),\,P_Y(k)\},\quad k\in S.$$

Заметим, что

$$\begin{gathered}P_{XY}\overset{\mathrm{def}}{=}\sum_{k\geqslant0}\mu_{XY}(k)=\sum_{k\geqslant0}\left\{\frac{P_X(k)+P_Y(k)}{2}-\frac{|P_X(k)-P_Y(k)|}{2}\right\}=\\ =1-\frac{1}{2}\sum_{k\geqslant0}|P_X(k)-P_Y(k)|=1-\rho(X,Y).\end{gathered}$$

Заметим еще, что при любом $k\in S$ одно из чисел

$$P_X(k)-\mu_{XY}(k),\quad P_Y(k)-\mu_{XY}(k)$$

равно 0, а другое неотрицательно. Таким образом, множества

$$A_X = \{k \in S : P_X(k) - \mu_{XY}(k) > 0\}, \quad A_Y = \{k \in S : P_Y(k) - \mu_{XY}(k) > 0\}$$

не пересекаются. В пространстве элементарных событий Ω выделим измеримое множество C, $\mathbf{P}\{C\} = P_{XY}$, и его дополнение $C^* = \Omega \backslash C$. Теперь построим такую пару случайных величин (X', Y'), что $X'(\omega) = Y'(\omega)$ при $\omega \in C$ и $X'(\omega) \neq Y'(\omega)$ при $\omega \in C^*$. Разобьем C на непересекающиеся подмножества C_k, $k \in S$, так, чтобы $\mathbf{P}\{C_k\} = \mu_{XY}(k)$, $k \in S$, и положим

$$X'(\omega) = Y'(\omega) = k \quad \text{для всех} \quad \omega \in C_k,\ k \in S.$$

Чтобы доопределить X' и Y' на C^*, разобьем C^* двумя способами на подмножества

$$C^*_{X,k} \colon \mathbf{P}\{C^*_{X,k}\} = P_X(k) - \mu_{XY}(k), \quad k \in A_X,$$

и подмножества

$$C^*_{Y,k} \colon \mathbf{P}\{C^*_{Y,k}\} = P_Y(k) - \mu_{XY}(k), \quad k \in A_Y.$$

Положим

$$X'(\omega) = k \quad \text{для всех} \quad \omega \in C^*_{X,k}, \quad k \in A_X,$$
$$Y'(\omega) = k \quad \text{для всех} \quad \omega \in C^*_{Y,k}, \quad k \in A_Y.$$

При таком построении

$$\mathbf{P}\{X' = k\} = \mathbf{P}\{C_k\} + \mathbf{P}\{C^*_{X,k}\} = P_X(k), \quad k \in S,$$
$$\mathbf{P}\{Y' = k\} = \mathbf{P}\{C_k\} + \mathbf{P}\{C^*_{Y,k}\} = P_Y(k), \quad k \in S,$$

и

$$\mathbf{P}\{X' \neq Y'\} = \mathbf{P}\{C^*\} = 1 - P_{XY} = \rho(X, Y),$$

так как на множестве C^* случайные величины X' и Y' принимают значения из не пересекающихся множеств A_X и A_Y, а на множестве C случайные величины X и Y по определению совпадают. Лемма 4 доказана.

Согласно лемме 4 расстояние по вариации $\rho(X, Y)$ есть нижняя грань значений $\mathbf{P}\{X' \neq Y'\}$ по множеству $Q(X, Y)$ всех пар (X', Y') случайных векторов с такими же маргинальными распределениями, как у вектора (X, Y).

Расстояние по вариации является естественной мерой возможности различения простых статистических гипотез.

Пусть по наблюдению случайной величины Z нужно различить две простые гипотезы:

H_0: Z имеет распределение P_0,

H_1: Z имеет распределение P_1,

с ошибками первого и второго родов α и β.

Пусть в качестве множества принятия гипотезы H_0 выбрано множество C, так что при $Z \in C$ принимается гипотеза H_0, а при $Z \notin C$ принимается гипотеза H_1. Тогда вероятности ошибок первого и второго родов равны соответственно $\alpha = \mathbf{P}_0\{Z \notin C\}$ и $\beta = \mathbf{P}_1\{Z \in C\}$. Из определения расстояния по вариации следует, что

$$\mathbf{P}_0\{Z \in C\} - \mathbf{P}_1\{Z \in C\} = 1 - \alpha - \beta, \quad \mathbf{P}_0\{Z \in C\} - \mathbf{P}_1\{Z \in C\} \leqslant \rho(P_0, P_1).$$

Значит,

$$1 - \alpha - \beta \leqslant \rho(P_0, P_1), \quad \text{или} \quad \alpha + \beta \geqslant 1 - \rho(P_0, P_1).$$

Таким образом, если расстояние по вариации между P_0 и P_1 мало́, то сумма вероятностей ошибок при различении двух простых гипотез не может быть малой, т. е. не существует эффективных статистических критериев, которые различают гипотезы, близкие в смысле расстояния по вариации.

Обратно, если расстояние по вариации между мерами P_0 и P_1 равно Δ, то по теореме Хана существует такое измеримое множество C в пространстве значений Z, что $P_0(D) > P_1(D)$ для любого измеримого $D \subseteq C$ и $P_0(D) \leqslant P_1(D)$ для любого измеримого $D \subseteq \Omega \backslash C$. Тогда $P_0(C) - P_1(C) = P_1(\Omega \backslash C) - P_0(\Omega \backslash C) = \Delta$. Поэтому если принимать гипотезу H_0 тогда и только тогда, когда $Z \in C$, то сумма вероятностей ошибок первого и второго родов будет равна

$$P_0\{Z \notin C\} + P_1\{Z \in C\} = 1 - P_0\{Z \in C\} + P_1\{Z \in C\} = 1 - \Delta.$$

Расстояние по вариации дает универсальные верхние оценки для разности между вероятностями, которые приписываются разными вероятностными мерами одному и тому же множеству. Поэтому существует довольно много уточнений разных предельных теорем, в которых изучается расстояние по вариации между допредельными и предельным распределениями. Доказательства таких теорем обычно основываются на лемме 4: для оценки расстояния по вариации между распределениями P_1 и P_2 выбирается достаточно богатое вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$ и на нем строятся случайные величины X с распределением P_1 и Y с распределением P_2 так, чтобы вероятность их несовпадения (как функций на Ω) была как можно меньше. Тогда в силу леммы 4 вероятность несовпадения X и Y не меньше расстояния по вариации между распределениями P_1 и P_2.

В качестве первого примера получим оценки для расстояния по вариации в теореме Пуассона.

Получим оценки для расстояния по вариации в теореме Пуассона.

Теорема 1. *Пусть $\xi_1, \ldots, \xi_n$ – независимые индикаторы,*

$$\mathbf{P}\{\xi_k = 1\} = p_k, \quad \mathbf{P}\{\xi_k = 0\} = 1 - p_k, \quad k = 1, 2, \ldots, n, \tag{72}$$

и $\zeta_n = \xi_1 + \ldots + \xi_n$, а κ – случайная величина, имеющая распределение Пуассона с параметром $\lambda = p_1 + \ldots + p_n$. Тогда

$$\rho(\zeta_n, \kappa) \leqslant \sum_{k=1}^{n} p_k^2 \leqslant \lambda \max\{p_1, \ldots, p_n\}.$$

Доказательство. Так как семейство распределений Пуассона замкнуто относительно свертки, то случайную величину κ можно представить в виде суммы независимых случайных величин $\kappa_1, \ldots, \kappa_n$, имеющих распределения Пуассона с параметрами $p_1, \ldots, p_n$ соответственно. Если $(\xi_1^*, \kappa_1), \ldots, (\xi_n^*, \kappa_n)$ — такие независимые пары случайных величин, что $\xi_1^*, \ldots, \xi_n^*$ имеют распределения (72), то $\zeta_n^* = \xi_1^* + \ldots + \xi_n^*$ имеет такое же распределение, как ζ_n, $\kappa = \kappa_1 + \ldots + \kappa_n$ имеет распределение Пуассона с параметром λ и

$$\{\zeta_n^* \neq \kappa\} = \{\xi_1^* + \ldots + \xi_n^* \neq \kappa_1 + \ldots + \kappa_n\} \subseteq$$
$$\subseteq \{(\xi_1^*, \ldots, \xi_n^*) \neq (\kappa_1, \ldots, \kappa_n)\} = \bigcup_{k=1}^{n} \{\xi_k^* \neq \kappa_k\}.$$

Из леммы 4 следует, что

$$\rho(\zeta_n, \kappa) \leqslant \mathbf{P}\{\zeta_n^* \neq \kappa\} \leqslant \mathbf{P}\{(\xi_1^*, \ldots, \xi_n^*) \neq (\kappa_1, \ldots, \kappa_n)\} \leqslant \sum_{k=1}^{n} \mathbf{P}\{\xi_k^* \neq \kappa_k\}. \tag{73}$$

Согласно лемме 4 минимальное значение $\mathbf{P}\{\xi_k^* \neq \kappa_k\}$ (по всем парам случайных векторов (ξ_k^*, κ_k) с указанными выше распределениями компонент) равно

$$\rho(\xi_k^*, \kappa_k) =$$
$$= \tfrac{1}{2}\left(|\mathbf{P}\{\kappa_k = 0\} - \mathbf{P}\{\xi_k^* = 0\}| + |\mathbf{P}\{\kappa_k = 1\} - \mathbf{P}\{\xi_k^* = 1\}| + \mathbf{P}\{\kappa_k \geqslant 2\}\right) =$$
$$= \tfrac{1}{2}\left((e^{-p_k} - (1-p_k)) + (p_k - p_k e^{-p_k}) + 1 - (1+p_k)e^{-p_k}\right) =$$
$$= p_k(1 - e^{-p_k}) < p_k^2,$$

отсюда и из (73) следует утверждение теоремы.

Замечание. Точность аппроксимации распределения суммы независимых индикаторов пуассоновским распределением можно немного улучшить, если приближать распределение суммы $\zeta_n = \xi_1 + \ldots + \xi_n$ распределением суммы независимых случайных величин $\tilde{\kappa}_1, \ldots, \tilde{\kappa}_n$, где $\tilde{\kappa}_k$ имеет распределение Пуассона с параметром $\lambda_k = -\ln(1-p_k)$. Тогда

$$\mathbf{P}\{\tilde{\kappa}_k = 0\} = e^{-\lambda_k} = 1 - p_k = \mathbf{P}\{\xi_k = 0\},$$
$$\mathbf{P}\{\tilde{\kappa}_k = 1\} = \lambda e^{-\lambda_k} = -(1-p_k)\ln(1-p_k) = \sum_{m\geqslant 1}(1-p_k)\frac{p_k^m}{m} =$$
$$= p_k + \sum_{m=2}^{\infty} p_k^m \left(\frac{1}{m} - \frac{1}{m-1}\right) = p_k - \sum_{m=2}^{\infty}\frac{p_k^m}{m(m-1)}$$

и $\tilde{\kappa} = \tilde{\kappa}_1 + \cdots + \tilde{\kappa}_n$ имеет распределение Пуассона с параметром $\Lambda = -\sum\limits_{k=1}^{n}\ln(1-p_k)$. Оценка расстояния по вариации между распределениями ζ_n и $\tilde{\kappa}$ проводится точно так же за исключением последнего неравенства:

$$\rho(\xi_k^*, \tilde{\kappa}_k) =$$
$$= \tfrac{1}{2}\left(|\mathbf{P}\{\tilde{\kappa}_k = 0\} - \mathbf{P}\{\xi_k^* = 0\}| + |\mathbf{P}\{\tilde{\kappa}_k = 1\} - \mathbf{P}\{\xi_k^* = 1\}| + \mathbf{P}\{\tilde{\kappa}_k \geqslant 2\}\right) =$$
$$= \frac{1}{2}\left(\left(p_k - \left(p_k - \sum_{m=2}^{\infty}\tfrac{p_k^m}{m(m-1)}\right)\right) + \left(1 - (1-p_k) - \left(p_k - \sum_{m=2}^{\infty}\tfrac{p_k^m}{m(m-1)}\right)\right)\right) =$$
$$= \sum_{m=2}^{\infty}\frac{p_k^m}{m(m-1)} < \frac{p_k^2}{2} + \frac{p_k^3}{6(1-p_k)} < \frac{p_k^2 + p_k^3}{2} \quad \text{при} \quad p_k < \frac{2}{3}\,.$$

Таким образом, $\rho(\zeta_n, \tilde{\kappa}) < \frac{1}{2}\sum_{k=1}^{n}(p_k^2 + p_k^3)$, если $\max_k p_k < \frac{2}{3}$, т. е. изменение параметра пуассоновского распределения почти в 2 раза уменьшает оценку расстояния по вариации.

Важный и полезный пример оценки расстояния по вариации – сравнение схемы выборки с возвращением и схемы выборки без возвращения.

Теорема 2. *Пусть в каждой из двух урн содержится N шаров с номерами $1, \dots, N$. Пусть $\xi_1, \dots, \xi_n$ – номера шаров, извлеченных из первой урны по схеме равновероятного выбора без возвращения, а $\gamma_1, \dots, \gamma_n$ – номера шаров, извлеченных из второй урны по схеме равновероятного выбора с возвращением. Тогда при $n, N \to \infty$*

$$\rho((\xi_1, \dots, \xi_n), (\gamma_1, \dots, \gamma_n)) \leqslant \frac{C_n^2}{N}.$$

Доказательство. Пусть $A(n, N)$ – множество всех таких $\mathbf{x} = (x_1, \dots, x_n) \in \{1, \dots, N\}^n$, что $x_i = x_j$ при некоторых i и j, $1 \leqslant i < j \leqslant n$. Так как

$$A(n, N) = \bigcup_{1 \leqslant i < j \leqslant n} \{\mathbf{x}\colon x_i = x_j\},$$

то

$$|A(n, N)| \leqslant \sum_{1 \leqslant i < j \leqslant n} |\{\mathbf{x} : x_i = x_j\}| = C_n^2 N^{n-1}.$$

Для любого $\mathbf{x} = (x_1, \dots, x_n) \in \{1, \dots, N\}^n$

$$\mathbf{P}\{(\gamma_1, \dots, \gamma_n) = \mathbf{x}\} = N^{-n},$$

$$\mathbf{P}\{(\xi_1, \dots, \xi_n) = \mathbf{x}\} = \frac{1}{N^{[n]}}, \quad \text{если} \quad \mathbf{x} \in A(n, N),$$

$$\mathbf{P}\{(\xi_1, \dots, \xi_n) = \mathbf{x}\} = 0 \quad \text{если} \quad \mathbf{x} \notin A(n, N).$$

По Лемме 1

$$\rho((\xi_1, \dots, \xi_n), (\gamma_1, \dots, \gamma_n)) =$$
$$= \frac{1}{2} \sum_{\mathbf{x} \in \{1, \dots, N\}^n} |\mathbf{P}\{(\xi_1, \dots, \xi_n) = \mathbf{x}\} - \mathbf{P}\{(\gamma_1, \dots, \gamma_n) = \mathbf{x}\}| =$$

$$= \frac{1}{2} \sum_{\mathbf{x} \in A(n, N)} \mathbf{P}\{(\gamma_1, \dots, \gamma_n) = \mathbf{x}\} +$$
$$+ \frac{1}{2} \sum_{\mathbf{x} \notin A(n, N)} |\mathbf{P}\{(\xi_1, \dots, \xi_n) = \mathbf{x}\} - \mathbf{P}\{(\gamma_1, \dots, \gamma_n) = \mathbf{x}\}| =$$

$$= \frac{1}{2} \sum_{\mathbf{x} \in A(n,N)} \mathbf{P}\{(\gamma_1, \dots, \gamma_n) = \mathbf{x}\} +$$

$$+ \frac{1}{2} \left(1 - \left(1 - \sum_{\mathbf{x} \in A(n,N)} \mathbf{P}\{(\gamma_1, \dots, \gamma_n) = \mathbf{x}\} \right) \right) =$$

$$= \sum_{\mathbf{x} \in A(n,N)} \mathbf{P}\{(\gamma_1, \dots, \gamma_n) = \mathbf{x}\} \leqslant \frac{2C_n^2 N^{n-1}}{N^n} = \frac{C_n^2}{N}.$$

Теорема доказана.

Наконец, еще одно замечание. Мы уже отмечали, что расстояние по вариации позволяет оценивать принципиальную возможность построения статистических критериев различения двух простых гипотез: если расстояние по вариации между распределениями, соответствующими гипотезам H_0 : наблюдение имеет распределение P_0 и H_1 : наблюдение имеет распределение P_1, меньше ε, то не существует статистического критерия, сумма вероятностей ошибок которого меньше $1-\varepsilon$. Особенно наглядно это можно показать с помощью леммы о том, что

$$\rho(P_1, P_2) = \inf_{(X,Y): X \sim P_1, Y \sim P_2} \mathbf{P}\{X \neq Y\}.$$

Действительно, согласно этой лемме можно построить на одном вероятностном пространстве случайные величины X с распределением P_1 и Y с распределением P_2 так, что $\mathbf{P}\{X = Y\} = 1 - \rho(P_1, P_2) > 1 - \varepsilon$. Но если случайные величины совпадают на множестве $A \subset \Omega$, то при $\omega \in A$ различить эти величины по значениям $X(\omega) = Y(\omega)$ невозможно.

Для распределения P на $\mathbb{Z}$ и натурального числа n будем через P^n обозначать меру на $\mathbb{Z}^n$, соответствующую n независимым случайным величинам, каждая из которых имеет распределение P:

$$P^n(z_1, \dots, z_n) = \prod_{i=1}^{n} P(z_i).$$

Лемма 7. *Пусть случайные величины $X_1, \dots, X_n$ независимы и одинаково распределены, случайные величины $Y_1, \dots, Y_n$ — независимы и одинаково распределены и $\rho(X_1, Y_1) = \varepsilon$. Тогда при любом натуральном n*

$$1 - 2e^{-n\varepsilon^2/2} \leqslant \rho((X_1, \dots, X_n), (Y_1, \dots, Y_n)) \leqslant n\varepsilon.$$

Доказательство. Правое неравенство доказывается по индукции с помощью леммы 2. Доказательство левого неравенства использует известное неравенство Хефдинга (см. [**?**], [**?**]):

Если $\alpha_1, \ldots, \alpha_n$ независимы, $\mathbf{P}\{0 \leqslant \alpha_k \leqslant 1\} = 1$ для всех $k = 1, \ldots, n$ и $S_n = \alpha_1 + \ldots + \alpha_n$, $\mathbf{E}S_n = \mu n$, то при $n > 1$

$$\mathbf{P}\left\{\frac{1}{n} S_n - \mu > z\right\} \leqslant e^{-2nz^2} \text{ при любом } z \in [0, 1-\mu),$$
$$\mathbf{P}\left\{\frac{1}{n} S_n - \mu < -z\right\} \leqslant e^{-2nz^2} \text{ при любом } z \in [0, \mu).$$

Так как $\rho(X_1, Y_1) = \varepsilon$, то существует такое событие C, что $\mathbf{P}\{X_1 \in C\} = p$, $\mathbf{P}\{Y_1 \in C\} = p + \varepsilon$. Тогда $\mathbf{E}I\{X_1 \in C\} = p$, $\mathbf{E}I\{Y_1 \in C\} = p + \varepsilon$ и согласно неравенству Хефдинга

$$\mathbf{P}\left\{\frac{1}{n} \sum_{k=1}^{n} I\{X_k \in C\} - p < \frac{1}{2}\varepsilon\right\} \leqslant \exp\left\{-\frac{n\varepsilon^2}{2}\right\},$$
$$\mathbf{P}\left\{\frac{1}{n} \sum_{k=1}^{n} I\{Y_k \in C\} - (p+\varepsilon) < -\frac{1}{2}\varepsilon\right\} \leqslant \exp\left\{-\frac{n\varepsilon^2}{2}\right\}.$$

Значит,

$$\rho((X_1, \ldots, X_n), (Y_1, \ldots, Y_n)) \geqslant$$
$$\geqslant \mathbf{P}\left\{\frac{1}{n} \sum_{k=1}^{n} I\{Y_k \in C\} < p + \frac{\varepsilon}{2}\right\} - \mathbf{P}\left\{\frac{1}{n} \sum_{k=1}^{n} I\{X_k \in C\} < p + \frac{\varepsilon}{2}\right\} \geqslant$$
$$\geqslant 1 - 2\exp\left\{-\frac{n\varepsilon^2}{2}\right\}.$$

Лемма доказана.

Существуют примеры, в которых оценка (**??**) близка к точной. Например, пусть $P(0) = 1$, $Q(0) = 1 - \varepsilon$, $Q(1) = \varepsilon$, $\varepsilon > 0$. Тогда $\rho(P, Q) = \varepsilon$ и

$$\rho(P^n, Q^n) = \frac{1}{2}\left(1 - Q^n(0, \ldots, 0) + \sum_{(z_1, \ldots, z_n) \neq (0, \ldots, 0)} Q^n(z_1, \ldots, z_n)\right) =$$
$$= 1 - (1-\varepsilon)^n = n\varepsilon(1 + O(n\varepsilon)) \quad \text{при} \quad n\varepsilon \to 0.$$

Теорема 1. *Пусть P и Q — два вероятностных распределения на $\mathbb{Z}$, которые различаются на множестве $\mathcal{D}(P,Q)=\{z\in\mathbb{Z}\colon P(z)\neq Q(z)\}$, и $p_*=\inf_{z\in\mathcal{D}(P,Q)}(\min(P(z),Q(z)))$. Если $p_*>0$, то при любом $n\in\mathbb{N}$*

$$\rho(P^n,Q^n)\leqslant\sqrt{\frac{1}{\pi p_*}}\sqrt{n+\frac{1}{p_*}}\,\rho(P,Q)$$

и

$$\rho(P^n,Q^n)\leqslant\sqrt{\frac{n}{2p_*}}\,\rho(P,Q).$$

Теорема. *Пусть P_t обозначает распределение на множестве $\{0,1\}$ и $P_t(1)=t$, $P_t(0)=1-t$. Тогда при любых $0<a<b<1$ и $n\in\{1,2,\ldots\}$*

$$\rho(P_a^n,P_b^n)\leqslant\tfrac{1}{2}\left(\arcsin(2b-1)-\arcsin(2a-1)\right)\sqrt{n}.$$

Доказательство. По определению

$$\rho(P_a^n,P_b^n)=$$

$$\tfrac{1}{2}\sum_{k_1,\ldots,k_n=0}^{1}|a^{k_1+\ldots+k_n}(1-a)^{n-(k_1+\ldots+k_n}-b^{k_1+\ldots+k_n}(1-b)^{n-(k_1+\ldots+k_n}|=$$

$$\tfrac{1}{2}\sum_{k=0}^{n}C_n^k|a^k(1-a)^{n-k}-b^k(1-b)^{n-k}|.$$

Далее, так как $\frac{d}{dx}x^k(1-x)^{n-k}=x^{k-1}(1-x)^{n-k-1}(k-nx)$, то при $0\leqslant a<b\leqslant 1$

$$|a^k(1-a)^{n-k}-b^k(1-b)^{n-k}|=\left|\int\limits_a^b x^{k-1}(1-x)^{n-k-1}(k-nx)dx\right|\leqslant$$

$$\leqslant\int\limits_a^b x^{k-1}(1-x)^{n-k-1}|k-nx|dx.$$

Поэтому

$$\rho(P_a^n,P_b^n)\leqslant\tfrac{1}{2}\int\limits_a^b\sum_{k=0}^{n}C_n^k x^{k-1}(1-x)^{n-k-1}|k-nx|dx.$$

Заметим теперь, что если ξ_x имеет биномиальное распределение с параметрами (n, x), то

$$\sum_{k=0}^{n} C_n^k x^{k-1}(1-x)^{n-k-1}|k-nx| = \tfrac{1}{x(1-x)}\mathbf{E}|\xi_x - nx| \leqslant$$
$$\leqslant \tfrac{1}{x(1-x)}\sqrt{\mathbf{E}(\xi_x - nx)^2} = \tfrac{1}{x(1-x)}\sqrt{nx(1-x)} = \sqrt{\tfrac{n}{x(1-x)}}\,.$$

Следовательно,

$$\rho(P_a^n, P_b^n) \leqslant \tfrac{1}{2}\int\limits_a^b \sqrt{\tfrac{n}{x(1-x)}} = \tfrac{1}{2}\left(\arcsin(2b-1) - \arcsin(2a-1)\right)\sqrt{n}.$$

Теорема доказана.

§ 13. Теорема о числе высоковероятных цепочек. Энтропия и ее свойства

Пусть $A = \{a_1, \ldots, a_N\}$ — конечный алфавит; последовательности символов этого алфавита можно рассматривать как сообщения, написанные на некотором языке. В реальных языках сообщения, как правило, не являются произвольными последовательностями букв: они «имеют смысл». Понятию «осмысленное сообщение» трудно дать определение, пригодное для широкого класса случаев. Рассмотрим одну из простейших моделей множества осмысленных сообщений.

Сообщения можно считать случайными, поскольку содержание будущих сообщений, как правило, заранее не известно. Кроме того, одни сообщения являются «более вероятными», чем другие, в частности, осмысленные сообщения должны появляться с большей вероятностью, чем бессмысленные. Для простоты можно рассматривать множества сообщений, имеющих одну и ту же длину T, и считать, что каждому сообщению соответствует вероятность его появления. В качестве первого приближения к понятию множества осмысленных сообщений $Z_{\text{от}}$ можно выбрать такое подмножество B_T^ε сообщений с наибольшими вероятностями, что суммарная вероятность остальных сообщений не превосходит малого числа $\varepsilon > 0$. Таким образом, неопределенная задача оценки количества осмысленных сообщений сводится к строгой математической задаче оценки объема наименьшего множества сообщений, суммарная вероятность которых не меньше $1 - \varepsilon$.

Будем рассматривать простую вероятностную модель языка, считая, что каждое сообщение длины T представляет собой последовательность $\xi_1, \xi_2, \ldots, \xi_T$ независимых одинаково распределенных букв алфавита $A = \{1, \ldots, N\}$ и что

$$\mathbf{P}\{\xi_t = k\} = p_k, \quad k = 1, \ldots, N.$$

Если $p_1 = \ldots = p_N = 1/N$, то знаки ξ_t имеют равномерное распределение на A и для любой последовательности $a^{(T)} = (a_1, \ldots, a_T) \in A^T$

$$\mathbf{P}\{(\xi_1, \ldots, \xi_T) = a^{(T)}\} = 1/N^T,$$

т. е. все последовательности равновероятны. В этом случае невозможно выделить «маленькое» множество строк $a^{(T)}$, содержащее существенно меньше N^T элементов и имеющее близкую к 1 суммарную вероятность.

Если же вероятности p_k различны, то

$$\mathbf{P}\{(\xi_1, \ldots, \xi_T) = a^{(T)}\} = p_{a_1} \ldots p_{a_T},$$

и вероятности появления строк $a^{(T)}$ не одинаковы. Следующая теорема показывает, что при неравномерном распределении $\{p_k\}_{k \in A}$ существуют такие подмножества строк фиксированной длины T, что при $T \to \infty$ суммарная вероятность их элементов стремится к 1, а логарифмы вероятностей входящих в них строк различаются не очень сильно. В качестве следствия из теоремы будет показано, что логарифм числа элементов в этих множествах растет медленнее, чем $T \ln N$, т. е. число элементов по порядку меньше N^T.

Теорема. *Пусть $\xi_1, \xi_2, \ldots$ – независимые случайные величины, принимающие значения в алфавите $A = \{1, \ldots, N\}$ с вероятностями $\mathbf{P}\{\xi_t = k\} = p_k, k = 1, \ldots, N$. Для любого $\varepsilon > 0$ существуют такие множество $B_T = B(T, \varepsilon) \subset A^T$ и число $T_0 = T_0(\varepsilon) < \infty$, что при любом $T \geqslant T_0$*

$$\mathbf{P}\{(\xi_1, \ldots, \xi_T) \notin B_T\} \le \varepsilon, \tag{74}$$

$$\mathbf{P}\{(\xi_1, \ldots, \xi_T) = a^{(T)}\} = \exp\{-TH + CT^{2/3}\}, \quad a^{(T)} \in B_T, \tag{75}$$

где $C = C(a^{(T)}),\ |C| \le 1,\ H = -\sum_{1 \le i \le N} p_i \ln p_i$.

Замечание 1. Множество B_T в этой теореме содержит строки, на которых сосредоточена почти вся вероятностная масса. Однако из (75) следует, что строки $a^{(T)} = (a_1, \ldots, a_T)$ с самыми большими вероятностями в него не входят.

Замечание 2. Равенство (75) показывает, что логарифмы вероятностей строк, входящих в множество B_T, различаются не слишком сильно (на величины, по порядку меньшие значений самих логарифмов). Таким образом, согласно теореме вероятностная мера на множестве строк длины T из независимых знаков при $T \to \infty$ сосредоточивается на множестве строк, логарифмы вероятностей появления которых асимптотически эквивалентны.

Доказательство теоремы. Рассмотрим $A^T = \{\omega^{(T)} = (\omega_1, \ldots, \omega_T) : \omega_1, \ldots, \omega_T \in A\}$ как конечное пространство элементарных событий. Определим на нем вероятностную меру, полагая

$$P(\omega^{(T)}) = p_{\omega_1} \ldots p_{\omega_T} \quad \text{при} \quad \omega^{(T)} = (\omega_1, \ldots, \omega_T).$$

На полученном вероятностном пространстве можно задать T независимых одинаково распределенных случайных величин $\xi_1, \ldots, \xi_T$, о которых говорится в условии теоремы: достаточно положить $\xi_t = \xi_t(\omega^{(T)}) = \omega_t,\ t = 1, \ldots, T$. Рассмотрим случайный вектор $\overline{\xi}^{(T)} = (\xi_1, \ldots, \xi_T)$. Введем вспомогательные случайные величины

$$\eta_t = \eta_t(\omega^{(T)}) = \ln p_{\xi_t(\omega)} = \ln p_{\omega_t},\ t = 1, \ldots, T.$$

Они независимы и одинаково распределены (как одинаковые функции от независимых одинаково распределенных случайных величин $\xi_1, \ldots, \xi_T$). Нетрудно найти их математическое ожидание и дисперсию:

$$\mathbf{E}\eta_t = \sum_{1 \le k \le N} p_k \ln p_k = -H,$$
$$\mathbf{D}\eta_t = \sum_{1 \le k \le N} p_k (\ln p_k - \mathbf{E}\eta_t)^2 = G < \infty.$$

На пространстве элементарных событий A^T рассмотрим функцию

$$\ln P(\omega^{(T)}) = \ln p_{\omega_1} + \ldots + \ln p_{\omega_T} = \eta_1 + \ldots + \eta_T.$$

Так как случайные величины $\eta_1, \ldots, \eta_T$ независимы и одинаково распределены, то

$$\mathbf{E} \ln P(\xi^{(T)}) = T\mathbf{E}\eta_1 = -TH,$$
$$\mathbf{D} \ln P(\xi^{(T)}) = T\mathbf{D}\eta_1 = TG.$$

По неравенству Чебышева

$$\mathbf{P}\left\{|\ln P(\xi^{(T)}) - \mathbf{E}\ln P(\xi^{(T)})| \geq x\sqrt{\mathbf{D}\ln P(\xi^{(T)})}\right\} \leq$$
$$\leq \frac{\mathbf{D}\ln P(\xi^{(T)})}{\left(x\sqrt{\mathbf{D}\ln P(\xi^{(T)})}\right)^2} = \frac{1}{x^2};$$

при $x = 1/\sqrt{\varepsilon}$ это неравенство можно переписать в виде

$$\mathbf{P}\left\{|\ln P(\xi^{(T)}) + TH| \geq \sqrt{\frac{TG}{\varepsilon}}\right\} \leq \frac{1}{(1/\sqrt{\varepsilon})^2} = \varepsilon. \tag{76}$$

Положим теперь

$$B_T = B(T,\varepsilon) = \left\{\omega^{(T)} \in A^T : |\ln P(\omega^{(T)}) + TH| \leq \sqrt{\frac{TG}{\varepsilon}}\right\};$$

тогда в силу (76)

$$\mathbf{P}\{\xi^{(T)} \notin B_T\} = \mathbf{P}\left\{|\ln P(\xi^{(T)}) + TH| > \sqrt{\frac{TG}{\varepsilon}}\right\} \leq \varepsilon. \tag{77}$$

Тем самым первое утверждение теоремы доказано.

Для доказательства второго утверждения перепишем определение множества B_T:

$$B_T = \left\{\omega^{(T)} \in A^T\colon -TH - \sqrt{TG/\varepsilon} \leq \ln P(\omega^{(T)}) \leq -TH + \sqrt{TG/\varepsilon}\right\}.$$

Значит, для любой строки $\omega^{(T)} = (\omega_1, \ldots, \omega_T) \in B_T$

$$\ln P(\omega^{(T)}) = -TH + CT^{2/3}\sqrt{\frac{G}{T^{1/3}\varepsilon}}, \quad |C| \leq 1,$$

и если $T \geq T_0 = \left(\frac{G}{\varepsilon}\right)^3$, то $\sqrt{\frac{G}{T^{1/3}\varepsilon}} \leq 1$, т. е. для любой строки $\omega^{(T)} \in B_T$

$$P(\omega^{(T)}) = e^{-TH + C'T^{2/3}}, \quad |C'| \leq 1.$$

Теорема доказана.

Положим теперь $Z_T(x) = \{\omega^{(T)} \in A^T : P(\omega^{(T)}) \geq e^{-TH - xT^{2/3}}\}$. Множество $Z_T(x)$ в нашей модели можно интерпретировать как множество высоковероятных («осмысленных») сообщений.

Следствие. *Если $x > 0$ фиксировано, то*

$$\mathbf{P}\{Z_T(x)\} \to 1 \quad \text{и} \quad |Z_T(x)| = e^{TH+o(T)} \quad \text{при} \quad T \to \infty.$$

Доказательство следствия. Если $x > 0$ и $\varepsilon_T = \frac{G}{x^2 T^{1/3}}$, то

$$Z_T(x) \supset B(T, \varepsilon_T) =$$
$$= \left\{\omega^{(T)} \in A^T \colon -TH - \sqrt{TG/\varepsilon_T} \le \ln P(\omega^{(T)}) \le -TH + \sqrt{TG/\varepsilon_T}\right\},$$

так как $\sqrt{TG/\varepsilon_T} = xT^{2/3}$. Учитывая (77), получаем:

$$\mathbf{P}\{Z_T(x)\} \ge \mathbf{P}\{B(T, \varepsilon_T)\} \ge 1 - \varepsilon_T \to 1, \quad T \to \infty,$$

что доказывает первое утверждение следствия. Далее,

$$1 \ge \mathbf{P}\{Z_T(x)\} = \sum_{\omega^{(T)} \in Z_T(x)} P(\omega^{(T)}) \ge |Z_T(x)| \min_{\omega^{(T)} \in Z_T(x)} P(\omega^{(T)});$$

так как $P(\omega^{(T)}) \ge e^{-TH - xT^{2/3}}$ при любом $\omega^{(T)} \in Z_T(x)$, то

$$|Z_T(x)| \le e^{TH + xT^{2/3}}. \tag{78}$$

С другой стороны, $|Z_T(x)| \ge |B(T, \varepsilon_T)|$ и

$$1 - \varepsilon_T \le \mathbf{P}\{B(T, \varepsilon_T)\} \le |B(T, \varepsilon_T)| \max_{a^{(T)} \in B(T, \varepsilon_T)} P(\omega^{(T)}) \le |B(T, \varepsilon_T)| e^{-TH + xT^{2/3}},$$

т. е.

$$|Z_T(x)| \ge |B(T, \varepsilon_T)| \ge (1 - \varepsilon_T) e^{TH - xT^{2/3}}. \tag{79}$$

Из (78) и (79) и того, что $\varepsilon_T \to 0$ при $T \to \infty$, следует, что $|Z_T(x)| = e^{TH + o(T)}$ и $|B(T, \varepsilon_T)| = e^{TH + o(T)}$. Следствие доказано.

Появившаяся в теореме величина $H(p) = -\sum_k p_k \ln p_k$ называется *энтропией* распределения $p = \{p_1, \ldots, p_N\}$ элементов последовательности независимых одинаково распределенных случайных величин.

Лемма 1. *Для любого вероятностного распределения* $p = (p_1, \ldots, p_N)$

$$H(p) = -\sum_{1 \le k \le N} p_k \ln p_k \le \ln N.$$

(*Здесь и далее мы считаем, что* $0 \cdot \ln 0 = 0$.)

Доказательство. Покажем, что максимальное значение функции $H(p)$ на множестве $Q = \{(p_1, \ldots, p_N) : p_t \ge 0, \sum p_t = 1\}$ достигается в точке $(\frac{1}{N}, \ldots, \frac{1}{N})$.

Множество Q – выпуклое, и оно совпадает с объединением отрезков, соединяющих точку $(\frac{1}{N}, \ldots, \frac{1}{N}) \in Q$ с точками границы Q. Поэтому достаточно доказать, что на каждом таком отрезке энтропия максимальна в точке $(\frac{1}{N}, \ldots, \frac{1}{N})$.

Представим отрезок, соединяющий $(\frac{1}{N}, \ldots, \frac{1}{N})$ с точкой $(p_1, \ldots, p_N) \in Q$, в параметрической форме:

$$p(t) = (p_1(t), \ldots, p_N(t)) = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right) + (a_1, \ldots, a_N)t,$$

$$a_k = p_k - \frac{1}{N}, \quad k = 1, \ldots, N, \quad t \in [0, 1];$$

тогда $a_1 + \ldots + a_N = 0$. Используя это представление, находим:

$$H(p(t)) = -\sum_{1 \le k \le N} \left(\frac{1}{N} + a_k t\right) \ln \left(\frac{1}{N} + a_k t\right)$$

и

$$\frac{d}{dt} H(p(t)) = -\sum_{k=1}^{N} \left(a_k \ln \left(\frac{1}{N} + t a_k\right) + \left(\frac{1}{N} + t a_k\right) \frac{a_k}{\frac{1}{N} + t a_k}\right) =$$

$$= -\sum_{k=1}^{N} a_k \ln \left(\frac{1}{N} + t a_k\right),$$

$$\frac{d}{dt} H(p(t))|_{t=0} = 0,$$

$$\frac{d^2}{dt^2} H(p(t)) = -\sum_{k=1}^{N} a_k \frac{a_k}{\frac{1}{N} + t a_k} = -\sum_{k=1}^{N} \frac{a_k^2}{\frac{1}{N} + t a_k} < 0.$$

Следовательно, на любом таком отрезке функция $H(p(t))$ выпукла вверх и имеет нулевую производную при $t = 0$. Значит, $t = 0$ – точка максимума функции $H(p(t))$. Лемма 1 доказана.

Лемма 1 показывает, что при измерении энтропии случайных величин, принимающих N значений, удобно выбирать в качестве основания логарифмов число N: тогда энтропия равномерного распределения будет равна 1, а энтропии остальных распределений будут принимать значения из полуинтервала $[0, 1)$.

Следствие. *Если $p \neq (\frac{1}{N}, \ldots, \frac{1}{N})$, то $H(p) < \ln N$ и минимальное число строк длины T, суммарная вероятность которых близка к 1, при $T \to \infty$*

растет как $e^{H(p)T(1+o(T)}$, *т. е. по порядку медленнее числа* N^T *всех строк длины* T.

Лемма 2. *Если* $p = (p_1, \ldots, p_N)$ *и* $q = (q_1, \ldots, q_N)$ – *два распределения вероятностей, то*

$$-\sum_{k=1}^{N} q_k \ln q_k \le -\sum_{k=1}^{N} q_k \ln p_k$$

и равенство достигается только при $p = q$.

Доказательство. Собирая все слагаемые в одной части, приведем доказываемое неравенство к виду

$$\sum_{k=1}^{N} q_k \ln \frac{p_k}{q_k} \le 0.$$

Далее преобразуем левую часть, воспользовавшись тем, что среднее геометрическое не превосходит среднего арифметического и совпадает с ним только если все осредняемые величины одинаковы:

$$\sum_{k=1}^{N} q_k \ln \frac{p_k}{q_k} = \ln \prod_{k=1}^{N} \left(\frac{p_k}{q_k}\right)^{q_k} \le \ln \sum_{k=1}^{N} q_k \frac{p_k}{q_k} = \ln 1 = 0.$$

Лемма 2 доказана.

Использованное здесь неравенство между средним геометрическим и средним арифметическим следует из того, что функция $\ln x$ – вогнутая и поэтому при любых $a, b > 0$

$$\ln b \le \ln a + (b-a)\frac{d}{dx}\ln x|_{x=a} = \ln a + \frac{b-a}{a} = \ln a + \frac{b}{a} - 1.$$

При любых $x_1, \ldots, x_N > 0,\ q_1, \ldots, q_N \ge 0,\ q_1 + \ldots + q_N = 1$, полагая в последней формуле $a = q_1x_1 + \ldots + q_Nx_N$, получаем:

$$\ln \prod_{k=1}^{N} x_k^{q_k} = \sum_{k=1}^{N} q_k \ln x_k \le \sum_{k=1}^{N} q_k \left(\ln a + \frac{x_k}{a} - 1\right) = \ln a = \ln \sum_{k=1}^{N} q_j x_j,$$

что эквивалентно соотношению $x_1^{q_1} \ldots x_N^{q_N} \le q_1x_1 + \ldots + q_Nx_N$. Так как функция $\ln x$ строго выпукла вверх, то равенство возможно только в случае, когда все x_k одинаковы и поэтому совпадают с a.

В большинстве вероятностных моделей реальных явлений возникают случайные величины, между которыми имеются те или иные зависимости. Простым примером зависимых случайных величин являются координаты случайного вектора (ξ, η), принимающего значения в множестве

$\{1,\dots,N\}\times\{1,\dots,M\}$. (Можно интерпретировать ξ и η как два сообщения.) Распределение такого случайного вектора задается набором чисел

$$\mathbf{P}\{(\xi,\eta)=(n,m)\}=r_{nm},(n,m)\in\{1,\dots,N\}\times\{1,\dots,M\};$$

а распределения его компонент – маргинальные распределения – наборами $p=(p_1,\dots,p_N)$ и $q=(q_1,\dots,q_M)$, где

$$p_n=\mathbf{P}\{\xi=n\}=\sum_{m=1}^{M}r_{nm},n\in\{1,\dots,N\},$$
$$q_m=\mathbf{P}\{\eta=m\}=\sum_{n=1}^{N}r_{nm},m\in\{1,\dots,M\}.$$

Если $r_{nm}\equiv p_nq_m$, то компоненты независимы, в противном случае они зависимы. Можно вычислить энтропию $H(\xi,\eta)$ распределения вектора (ξ,η), а также энтропии $H(\xi)$ и $H(\eta)$ его компонент. Возникает естественный вопрос: как связаны между собой эти величины?

Лемма 3. *Справедливо неравенство*

$$H(\xi,\eta)\le H(\xi)+H(\eta);$$

равенство достигается только в случае, когда ξ и η независимы.

Доказательство. Имеем по определению

$$H(\xi)+H(\eta)=-\sum_{n=1}^{N}p_n\ln p_n-\sum_{m=1}^{M}q_m\ln q_m=$$
$$=-\sum_{n=1}^{N}\left(\sum_{m=1}^{M}r_{nm}\right)\ln p_n-\sum_{m=1}^{M}\left(\sum_{n=1}^{N}r_{nm}\right)\ln q_m=$$
$$=-\sum_{n,m}r_{nm}\ln p_nq_m.$$

Набор $\{p_nq_m\}$ является распределением вероятностей, поэтому можно применить лемму 2 к последней сумме и оценить ее снизу:

$$H(\xi)+H(\eta)=-\sum_{n,m}r_{nm}\ln p_nq_m\ge-\sum_{n,m}r_{nm}\ln r_{nm}=H(\xi,\eta).$$

Равенство имеет место только при $r_{kl}\equiv p_kq_l$.

С другой стороны, справедливо также неравенство

$$H(\xi,\eta)\ge\max\{H(\xi),H(\eta)\}.\tag{80}$$

Мы докажем немного более общее утверждение.

Лемма 4. *Если случайная величина ξ принимает значения в конечном множестве $A = \{1, \ldots, N\}$ и $f : A \to B = \{1, \ldots, M\}$, то*

$$H(f(\xi)) \leq H(\xi).$$

Неравенство (80) следует из леммы 4, если заменить в ней ξ вектором (ξ, η), а в качестве f рассматривать проекции прямого произведения $\{1, \ldots, N\} \times \{1, \ldots, M\}$ на множества- сомножители.

Доказательство. Функция $y(x) = -x \ln x$ выпукла вверх и $y(0) = 0$; поэтому

$$y(x + \Delta) - y(x) \leq y(\Delta) - y(0) = y(\Delta) \quad \text{при любых} \quad x,\ \Delta \geq 0,$$

и неравенство строгое, если $x > 0$. Поэтому для любых неотрицательных $x_1, \ldots, x_t$

$$y(x_1 + \ldots + x_t) =$$

$$= \sum_{k=1}^{t} [y(x_1 + \ldots + x_k) - y(x_1 + \ldots + x_{k-1})] \leq \sum_{k=1}^{t} y(x_k),$$

причем неравенство строгое, если хотя бы два x_k положительны.

Если $\mathbf{P}\{\xi = k\} = p_k,\ k \in A$, и $\mathbf{P}\{f(\xi) = m\} = q_m,\ m \in B$, то $q_m = \sum_{k: f(k) = m} p_k$ и

$$H(f(\xi)) = -\sum_{m=1}^{M} q_m \ln q_m = -\sum_{m=1}^{M} \left(\sum_{k: f(k)=m} p_k \right) \ln \left(\sum_{k: f(k)=m} p_k \right) \leq$$

$$\leq \sum_{m=1}^{M} \left(-\sum_{k: f(k)=m} p_k \ln p_k \right) = -\sum_{k=1}^{N} p_k \ln p_k = H(\xi).$$

Лемма доказана.

Еще одно замечание относится к свойствам энтропии условных распределений.

Пусть ξ и η – случайные величины, заданные на одном вероятностном пространстве. Тогда определено условное распределение ξ относительно η, т. е. вероятности $\mathbf{P}\{\xi = x | \eta = y\}$. Это позволяет определить энтропию условного распределения при фиксированном значении η

$$H(\xi | \eta = y) = -\sum_{x} \mathbf{P}\{\xi = x | \eta = y\} \ln \mathbf{P}\{\xi = x | \eta = y\}$$

и условную энтропию случайной величины ξ относительно η:

$$H(\xi|\eta) = \sum_y \mathbf{P}\{\eta = y\} H(\xi|\eta = y) =$$
$$= -\sum_{x,y} \mathbf{P}\{\xi = x, \eta = y\} \ln \mathbf{P}\{\xi = x|\eta = y\}.$$

Лемма 5. *Для любых случайных величин ξ и η, определенных на одном вероятностном пространстве, справедливо неравенство*

$$H(\xi|\eta) \leq H(\xi).$$

Доказательство. По определению

$$H(\xi|\eta) = -\sum_{x,y} \mathbf{P}\{\xi = x, \eta = y\} \ln \frac{\mathbf{P}\{\xi = x, \eta = y\}}{\mathbf{P}\{\eta = y\}} =$$
$$= -\sum_{x,y} \mathbf{P}\{\xi = x, \eta = y\} \ln \mathbf{P}\{\xi = x, \eta = y\} +$$
$$+ \sum_{x,y} \mathbf{P}\{\xi = x, \eta = y\} \ln \mathbf{P}\{\eta = y\} =$$
$$= H(\xi, \eta) - H(\eta) \leq H(\xi) + H(\eta) - H(\eta) = H(\xi).$$

Последнее неравенство следует из леммы 3.

Ранее энтропия определялась для распределения случайной величины. Это понятие чаще применяют к случайным последовательностям. В качестве примера покажем, как определяется энтропия стационарной случайной последовательности $\{\xi_t\}$, элементы которых принимают значения в конечном множестве $A = \{1, \ldots, N\}$, т. е. такой последовательности случайных величин $\{\xi_t\}$, что распределение последовательности $\{\xi_t\}$ совпадает с распределением $\{\xi_{t+v}\}$ при любом $v > 0$. В частности, для стационарной последовательности $\{\xi_t\}$ при любых целых t, v, любом натуральном s и любых $i_0, \ldots, i_{s-1} \in A$

$$\mathbf{P}\{(\xi_t, \ldots, \xi_{t+s-1}) = (i_0, \ldots, i_{s-1})\} =$$
$$= \mathbf{P}\{(\xi_v, \ldots, \xi_{v+s-1}) = (i_0, \ldots, i_{s-1})\} = p_s(i_0, \ldots, i_{s-1}).$$

Энтропия

$$H_s = -\sum_{i_0, \ldots, i_{s-1} \in A} p_s(i_0, \ldots, i_{s-1}) \ln p_s(i_0, \ldots, i_{s-1})$$

распределения $P_s = \{p_s(i_0, \ldots, i_{s-1}), i_0, \ldots, i_{s-1} \in A\}$ называется *частичной энтропией* процесса $\{\xi_t\}$, и она, вообще говоря, растет при $s \to \infty$. Например, если $\{\xi_t\}$ — последовательность независимых одинаково распределенных символов, то $H_s = sH_1$ согласно лемме 3. Поэтому под энтропией последовательности $\{\xi_t\}$ понимают предел нормированных частичных энтропий

$$H = \lim_{s\to\infty} H_s/s; \tag{81}$$

его называют также *энтропией стационарной последовательности* $\{\xi_t\}$ *на один знак.*

Докажем, что это определение корректно, т. е. что предел H существует.

Лемма 6. *Если последовательность неотрицательных чисел* $\{x_n\}$ *полуаддитивна, т. е. удовлетворяет условию*

$$x_{s+v} \le x_s + x_v \quad \text{\textit{при любых}} \quad s, v > 0,$$

то существует предел x_n/n *при* $n \to \infty$.

Доказательство. Из определения полуаддитивности по индукции следует, что

$$x_n \le \left[\frac{n}{k}\right] x_k + x_{n-[n/k]k} \quad \text{при любых} \quad k < n,$$

т. е. $\frac{x_n}{n} \le \frac{x_k}{k} + \frac{1}{n}\max\{x_0, x_1, \ldots, x_{k-1}\}$. Значит,

$$\limsup_{n\to\infty} \frac{x_n}{n} \le \frac{x_k}{k} \quad \text{при любом} \quad k < \infty. \tag{82}$$

Положим $X = \liminf_{n\to\infty} x_n/n \ge 0$. Тогда для любого $\delta > 0$ существует такое $n(\delta) < \infty$, что $\frac{x_{n(\delta)}}{n(\delta)} \le X + \delta$, и поэтому в силу (82)

$$\limsup_{n\to\infty} x_n/n \le \frac{x_{n(\delta)}}{n(\delta)} \le X + \delta \quad \text{для любого} \quad \delta > 0.$$

Следовательно, существует $\lim_{n\to\infty} x_n/n = X$.

Согласно лемме 3

$$H_{s+v} \le H_s + H_v \quad \text{при любых} \quad s, v \ge 0;$$

следовательно, последовательность частичных энтропий полуаддитивна и предел (81) существует для любого стационарной последовательности с конечным числом состояний.

При некоторых дополнительных условиях для стационарных последовательностей справедливы доказанные выше утверждения о свойствах множества высоковероятных строк.

§ 14. Предельные теоремы для функций от случайных величин

Предельные теоремы для сумм независимых случайных величин удобно доказывать с помощью характеристических или производящих функций. Однако для распределений нелинейных функций от таких сумм этот метод напрямую не применим. В этом параграфе описан способ переноса предельных теорем для случайных величин на дифференцируемые функции от этих величин.

Напомним понятия сходимости случайных величин как измеримых функций на пространстве элементарных событий.

Определение. Последовательность случайных величин ξ_n *слабо сходится* к случайной величине ξ, если $\mathbf{E}g(\xi_n) \to \mathbf{E}g(\xi)$, $n \to \infty$, для любой ограниченной непрерывной функции g. Для случайных величин с действительными значениями слабая сходимость эквивалентна сходимости функций распределения во всех точках непрерывности предельной функции.
Последовательность случайных величин ξ_n *сходится по вероятности* к случайной величине ξ, если для любых $\varepsilon > 0$, $\tau > 0$ существует такое $N = N(\varepsilon, \tau) < \infty$, что $\mathbf{P}\{|\xi_n - \xi| > \tau\} < \varepsilon$ при любом $n > N$.
Последовательность случайных величин ξ_n *сходится с вероятностью 1 (почти наверное)* к случайной величине ξ, если $\mathbf{P}\left\{\lim\limits_{n\to\infty} \xi_n = \xi\right\} = 1$.

Теорема 1. *Если ξ_n — последовательность случайных величин и $\xi_n \to a = \mathrm{const}$ по вероятности при $n \to \infty$, а $H(\cdot)$ — функция, непрерывная в точке a, то $H(\xi_n) \to H(a)$ по вероятности при $n \to \infty$.*

Доказательство. Непрерывность функции $H(\cdot)$ в точке a означает, что для каждого $\delta > 0$ существует такое $\tau > 0$, что $|H(x) - H(a)| \leqslant \delta$, если $|x - a| \leqslant \tau$. С другой стороны, сходимость ξ_n к a по вероятности означает, что для любых $\varepsilon > 0$, $\tau > 0$ существует такое $N = N(\varepsilon, \tau) < \infty$, что $\mathbf{P}\{|\xi_n - a| > \tau\} < \varepsilon$ при любом $n > N$. Из этих двух утверждений следует, что для любых $\varepsilon > 0$, $\tau > 0$ при любом $n > N = N(\varepsilon, \tau)$

$$\mathbf{P}\{|H(\xi_n) - H(a)| > \delta\} \leqslant \mathbf{P}\{|\xi_n - a| > \tau\} < \varepsilon.$$

Теорема доказана.

Теорема 2. *Если $\overline{\eta}_n$ — последовательность случайных векторов в $\mathbb{R}^d$, слабо сходящаяся к случайному d-мерному вектору $\overline{\eta}$, а функция $H : \mathbb{R}^d \to \mathbb{R}^m$ непрерывна, то случайные векторы $H(\overline{\eta}_n)$ слабо сходятся к вектору $H(\overline{\eta})$.*

Доказательство. По определению последовательность случайных величин $\overline{\eta}_n$ слабо сходится к $\overline{\eta}$, если $\mathbf{E}g(\overline{\eta}_n) \to \mathbf{E}g(\overline{\eta})$ при $n \to \infty$ для любой ограниченной непрерывной функции g. В нашем случае для доказательства слабой сходимости нужно убедиться в том, что

$$\mathbf{E}g(H(\overline{\eta}_n)) \to \mathbf{E}g(H(\overline{\eta})) \quad \text{при} \quad n \to \infty. \tag{83}$$

Но функция $g(H(\cdot))$ непрерывна (как суперпозиция непрерывных функций) и ограничена, так что (83) следует из слабой сходимости $\overline{\eta}_n$ к $\overline{\eta}$. Теорема доказана.

Теорема 3. *Пусть последовательность случайных величин η_n со значениями в $\mathbb{R}$ слабо сходится к случайной величине η, функция $H\colon \mathbb{R} \to \mathbb{R}$ дифференцируема в точке a и $b_n \to 0$, $n \to \infty$. Тогда имеет место слабая сходимость*

$$\tfrac{1}{b_n}\left(H(a + b_n\eta_n) - H(a)\right) \to H'(a)\eta, \quad n \to \infty.$$

Доказательство. Функция

$$h(x) = \begin{cases} \tfrac{1}{x}\left(H(a+x) - H(a)\right), & x \neq 0, \\ H'(a), & x = 0, \end{cases}$$

непрерывна при $x = 0$. Последовательность случайных величин $b_n\eta_n$ слабо сходится к 0 при $n \to \infty$; по теореме 1 тогда $h(b_n\eta_n) \to h(0) = H'(a)$, $n \to \infty$, по вероятности. Поэтому

$$\tfrac{1}{b_n}\left(H(a + b_n\eta_n) - H(a)\right) = h(b_n\eta_n)\eta_n$$

слабо сходится к $H'(a)\eta$ при $n \to \infty$. Теорема доказана.

В случае, когда $H'(a) = 0$, из теоремы 3 следует лишь, что

$$\tfrac{1}{b_n}\left(H(a + b_n\xi_n) - H(a)\right) \to 0, \quad n \to \infty.$$

Более полную информацию в этом случае можно получить тем же методом.

Теорема 4. *Если выполнены условия теоремы 3 и $H'(a) = 0$, но $H''(a) \neq 0$, то имеет место слабая сходимость*

$$\frac{1}{b_n^2}\left(H(a + b_n\eta_n) - H(a)\right) \to \frac{1}{2}\,\eta^2 H''(a), \quad n \to \infty.$$

Доказательство проводится по той же схеме, только вместо непрерывной в 0 функции

$$h(x) = \begin{cases} \frac{1}{x}\,(H(a+x) - H(a)), & x \neq 0, \\ H'(a), & x = 0, \end{cases}$$

нужно рассматривать непрерывную в 0 функцию

$$h_2(x) = \begin{cases} \frac{2}{x^2}\,(H(a+x) - H(a)), & x \neq 0, \\ H''(a), & x = 0. \end{cases}$$

Последовательность случайных величин $b_n\eta_n$ слабо сходится к 0 при $n \to \infty$; по теореме 1 тогда $h_2(b_n\eta_n) \to h(0) = H''(a)$, $n \to \infty$, по вероятности. Поэтому

$$\tfrac{H(a+b_n\eta_n)-H(a)}{b_n^2} = \tfrac{1}{2}\,h(b_n\eta_n)\eta_n^2$$

слабо сходится к $\frac{1}{2}\,H''(a)\eta^2$ при $n \to \infty$. Теорема доказана.

Пример. Пусть проводятся независимые испытания с вероятностью успеха p и неудачи $q = 1 - p$, обозначим через ν_n число успехов в первых n испытаниях, а через $\xi_n = \frac{1}{n}\,\nu_n$ — относительную частоту успехов в первых n испытаниях. Тогда ν_n имеет биномиальное распределение с параметрами (n, p), и если $p = \text{const}$, а $n \to \infty$, то по закону больших чисел ξ_n сходится к p по вероятности, а по теореме Муавра–Лапласа

$$\mathbf{P}\left\{\frac{\nu_n - np}{\sqrt{np(1-p)}} = \frac{\xi_n - p}{\sqrt{p(1-p)}}\sqrt{n} \leqslant x\right\} \to \Phi(x) = \tfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}du. \quad (84)$$

Полагая $\eta_n = \frac{\nu_n - np}{\sqrt{np(1-p)}} = \frac{\xi_n - p}{\sqrt{p(1-p)}}\sqrt{n}$ и $b_n = \sqrt{\frac{p(1-p)}{n}}$, получаем представление $\xi_n = p + b_n\eta_n$, в котором η_n слабо сходятся к случайной величине η, имеющей стандартное нормальное распределение, а $b_n \to 0$ при $n \to \infty$.

Функция $H(x) = \sin(2\pi x)$ непрерывна и непрерывно дифференцируема:

$$H(p) = \sin(2\pi p), \quad H'(p) = 2\pi\cos(2\pi p), \quad H''(p) = -4\pi^2\sin(2\pi p).$$

Из теоремы 1 следует, что $H(\xi_n) = \sin(2\pi\xi_n)$ сходится по вероятности к константе $\sin(2\pi p)$:

$$\mathbf{P}\{|\sin(2\pi\xi_n) - \sin(2\pi p)| > \varepsilon\} \to 0 \ \text{ при любом } \ \varepsilon > 0.$$

По теореме 3 (с учетом (84)) случайные величины

$$\tfrac{1}{b_n}\,(H(p + b_n\eta_n) - H(p)) = \sqrt{\tfrac{n}{p(1-p)}}\,(\sin(2\pi\xi_n) - \sin(2\pi p))$$

слабо сходятся к случайной величине $H'(p)\eta = \eta \cdot 2\pi\cos(2\pi p)$. Таким образом, предельное распределение $\sqrt{\frac{n}{p(1-p)}}\,(\sin(2\pi\xi_n) - \sin(2\pi p))$ при $n \to \infty$ — нормальное с параметрами $(0, 4\pi\cos^2(2\pi p))$; иначе говоря, распределение $\sin(2\pi\xi_n)$ *асимптотически нормально* с параметрами $\left(\sin(2\pi p), \frac{4\pi p(1-p)}{n}\cos^2(2\pi p)\right)$. Предельное распределение оказывается вырожденным (имеющим нулевую дисперсию) при $p = \frac{1}{4}$ и $p = \frac{3}{4}$. В этих случаях нужно использовать теорему 4: при $n \to \infty$ и $p = \frac{1}{4}$ случайная величина

$$\frac{1}{b_n^2}\left(H(\tfrac{1}{4} + b_n\eta_n) - H(\tfrac{1}{4})\right) = \tfrac{16n}{3}\,(\sin(2\pi\xi_n) - 1)$$

слабо сходится к $\frac{1}{2}\,\eta H''(\frac{1}{4}) = -2\pi^2\eta^2$, а при $n \to \infty$ и $p = \frac{3}{4}$ случайная величина

$$\frac{1}{b_n^2}\left(H(\tfrac{3}{4} + b_n\eta_n) - H(\tfrac{3}{4})\right) = \tfrac{16n}{3}\,(\sin(2\pi\xi_n) + 1)$$

слабо сходится к $\frac{1}{2}\,\eta H''(\frac{3}{4}) = 2\pi^2\eta^2$.

Теоремы 3 и 4 обобщаются на функции от случайных векторов.

Теорема 5. *Пусть d-мерные случайные векторы $\overline{\eta}_n = (\eta_{n1}, \ldots, \eta_{nd})$ слабо сходятся к случайному вектору $\overline{\eta} = (\eta_1, \ldots, \eta_d)$ при $n \to \infty$, а $H : \mathbb{R}^d \to \mathbb{R}$ — функция, непрерывно дифференцируемая в окрестности точки $\overline{a}$, и $H'(\overline{a}) = (H'_1, \ldots, H'_d)$, где $H'_j = \frac{\partial H(\mathbf{t})}{\partial t_j}\Big|_{\overline{t}=\overline{a}}$. Тогда при $b_n \to \infty$ имеет место слабая сходимость*

$$\frac{1}{b_n}\,(H(\overline{a} + b_n\overline{\eta}_n) - H(\overline{a})) \Rightarrow (\overline{\eta}, H'(\overline{a})) = \sum_{j=1}^{d} \eta_j H'_j. \tag{85}$$

Если $\mathbf{P}\{(\overline{\eta}, H'(\overline{a})) = 0\} = 1$ (например, $H'(\overline{a}) = \overline{0}$), а элементы матрицы $H''(\overline{t}) = \left\|\frac{\partial^2 H(\overline{t})}{\partial t_i \partial t_j}\right\|$ конечны и непрерывны в окрестности точки $\overline{t} = \overline{a}$, то

$$\frac{1}{b_n^2}\,(H(\overline{a} + b_n\overline{\eta}_n) - H(\overline{a})) \Rightarrow \frac{1}{2}\,\overline{\eta} H'' \overline{\eta}^T = \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 H}{\partial t_i \partial t_j}\bigg|_{\overline{t}=\overline{a}} \eta_i\eta_j. \tag{86}$$

Доказательство. Теорема 5 доказывается аналогично теоремам 3 и 4. В силу непрерывной дифференцируемости функции H в окрестности точки $\overline{a}$ функция $h : \mathbb{R}^d \to \mathbb{R}$, равная отношению приращения функции к ее первому дифференциалу:

$$h(\overline{x}) = \begin{cases} \frac{H(\overline{a}+\overline{x}) - H(\overline{a})}{(\overline{x}, H'(\overline{a}))}, & (\overline{x}, H'(\overline{a})) \neq 0, \\ 1, & (\overline{x}, H'(\overline{a})) = 0, \end{cases}$$

непрерывна в точке $\overline{x} = \overline{0}$. Так как случайные векторы $b_n \overline{\eta}_n$ слабо сходятся к $\overline{0}$ при $b_n \to 0$, то $h(b_n \overline{\eta}_n)$ при $b_n \to 0$ слабо сходится к 1 в силу теоремы 1. Но тогда последовательность случайных величин

$$\frac{H(\overline{a} + b_n \overline{\eta}_n) - H(\overline{a})}{b_n} = h(b_n \overline{\eta}_n)(\overline{\eta}_n, H'(\overline{a}))$$

слабо сходится к $(\overline{\eta}_n, H'(\overline{a}))$, что и утверждается в первой части теоремы.

Вторая часть теоремы доказывается тем же приемом. В силу непрерывной дифференцируемости функции H в окрестности точки $\overline{a}$ функция $h_2 : \mathbb{R}^d \to \mathbb{R}$, равная отношению приращения функции к ее второму дифференциалу:

$$h_2(\overline{x}) = \begin{cases} \frac{H(\overline{a}+\overline{x}) - H(\overline{a})}{\frac{1}{2}\,\overline{x} H''(\overline{a}) \overline{x}^T}, & \overline{x} H''(\overline{a}) \overline{x}^T \neq 0, \\ 1, & \overline{x} H''(\overline{a}) \overline{x}^T = 0, \end{cases}$$

непрерывна в точке $\overline{x} = \overline{0}$. Так как случайные векторы $b_n \overline{x}_n$ слабо сходятся к $\overline{0}$ при $b_n \to 0$, то $h_2(b_n \overline{\eta}_n)$ при $b_n \to 0$ слабо сходится к 1 в силу теоремы 1. Но тогда случайные величины

$$\frac{H(\overline{a} + b_n \overline{\eta}_n) - H(\overline{a})}{b_n^2} = \tfrac{1}{2}\, h_2(b_n \overline{\eta}_n) \overline{\eta}_n H''(\overline{a}) \overline{\eta}_n^T$$

слабо сходятся к $\frac{1}{2}\,\overline{\eta} H'' \overline{\eta}^T$ при $n \to \infty$.

Теорема доказана.

§ 49. Тождество Вальда

Тождество Вальда. *Пусть* $\xi_1, \xi_2, \ldots$ *— независимые случайные величины,* $\mathbf{M}\xi_i = a,\ i = 1, 2, \ldots,\ \sup\limits_{i \geq 1} \mathbf{E}|\xi_i| \leq C < \infty$,

$$S_0 = 0, \quad S_n = \xi_1 + \ldots + \xi_n, \quad n = 1, 2, \ldots,$$

и $\nu \geq 1$ *— такой момент остановки относительно* $\{\sigma(\xi_1, \ldots, \xi_n)\}$, *что* $\mathbf{E}\nu < \infty$. *Тогда*

$$\mathbf{E} S_\nu = a \mathbf{E} \nu.$$

Пример. Рассмотрим случайное блуждание $S_n = \xi_1 + \ldots + \xi_n$, где $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины, имеющие стандартное нормальное распределение, и момент остановки $\nu = \min\{n : X_n < 0\}$. Если ввести индикаторы

$$\chi_i = \mathbb{I}\{\nu \geq i\} \in \mathcal{F}_{i-1} = \sigma(\xi_1, \ldots, \xi_{i-1}), \quad i = 1, 2, \ldots,$$

то

$$S_\nu = \sum_{i\ge1} \chi_i \xi_i, \quad \mathbf{P}\{S_\nu < 0\} = 1, \quad \mathbf{E}S_\nu < 0.$$

С другой стороны, при каждом $i \ge 1$ индикатор $\chi_i \in \mathcal{F}_{i-1}$ не зависит от случайной величины $\xi_i \in \mathcal{F}_i$, и, значит, при каждом $i \ge 1$

$$\mathbf{E}\chi_i\xi_i = \mathbf{E}\mathbb{I}\{\nu \ge i\}\xi_i = \mathbf{M}\mathbb{I}\{\nu \ge i\}\cdot\mathbf{M}\xi_i = 0.$$

Таким образом, в нашем примере

$$\sum_{i\ge1} \mathbf{E}\chi_i\xi_i = 0, \quad \text{но} \quad \mathbf{E}\sum_{i\ge1} \chi_i\xi_i = \mathbf{E}S_\nu < 0,$$

т.е. аддитивность математического ожидания не имеет места! Дело в том, что функциональный ряд $\sum_{i\ge1} \chi_i(\omega)\xi_i(\omega)$, $\omega \in \Omega$, не является абсолютно интегрируемым (можно показать, что $\mathbf{E}|\chi_i\xi_i| \ge \frac{1}{2i}\mathbf{E}|\xi_1|,\ i \ge 2$), и приведенный пример показывает, что в этом случае перестановка знаков суммы и математического ожидания может изменять результат.

Аддитивность математического ожидания заведомо имеет место для конечных сумм случайных величин с конечными математическими ожиданиями и для сумм неотрицательных случайных величин.

В общем случае для обоснования равенства

$$\mathbf{E}\sum_{i\ge1} \eta_i = \sum_{i\ge1} \mathbf{E}\eta_i$$

при $\mathbf{E}|\eta_i| < \infty,\ i = 1, 2, \ldots$, достаточно показать, что $\sum\limits_{i\ge N} \mathbf{E}|\eta_i| \to 0, N \to \infty$, так как из последнего соотношения следует, что

$$\left|\mathbf{E}\sum_{i>N} \eta_i\right| \le \mathbf{E}\left|\sum_{i>N} \eta_i\right| \le \mathbf{E}\sum_{i>N} |\eta_i| = \sum_{i>N} \mathbf{E}|\eta_i| \to 0, \quad N \to \infty,$$

$$\left|\sum_{i>N} \mathbf{E}\eta_i\right| \le \sum_{i>N} |\mathbf{E}\eta_i| \le \sum_{i>N} \mathbf{E}|\eta_i| \to 0, \quad N \to \infty,$$

а равенство

$$\mathbf{E}\sum_{i\ge1}^{N} \eta_i = \sum_{i\ge1}^{N} \mathbf{E}\eta_i$$

выполняется при любом конечном N.

Доказательство тождества Вальда. Положим $\chi_i = \mathbb{I}\{\nu \geq i\} \in \sigma(\xi_1, \ldots, \xi_{i-1}),\ i = 1, 2, \ldots$; тогда χ_i и ξ_i независимы при каждом i и

$$S_\nu = \sum_{i=1}^{\nu} \xi_i = \sum_{i=1}^{\infty} \chi_i \xi_i.$$

Так как ξ_i не зависит от χ_i, то

$$\mathbf{E}|\chi_i\xi_i| = \mathbf{E}\chi_i|\xi_i| = \mathbf{E}\chi_i\mathbf{E}|\xi_i| \leq C\mathbf{P}\{\nu \geq i\}$$

и

$$\sum_{i=1}^{\infty} \mathbf{E}|\chi_i\xi_i| < C\sum_{i=1}^{\infty} \mathbf{P}\{\nu \geq i\} = C\mathbf{E}\nu < \infty.$$

Значит, функциональный ряд $\sum\limits_{i\geq 1} \chi_i\xi_i$ абсолютно сходится почти наверное, и поэтому его можно интегрировать почленно (т.е. математическое ожидание суммы ряда равно сумме математических ожиданий его членов):

$$\mathbf{E}S_\nu = \mathbf{E}\sum_{i\geq 1} \chi_i\xi_i = \sum_{i\geq 1} \mathbf{E}\chi_i\xi_i = \sum_{i\geq 1} \mathbf{E}\chi_i\mathbf{E}\xi_i = a\sum_{i\geq 1} \mathbf{P}\{\nu \geq i\} = a\mathbf{E}\nu.$$

Тем самым тождество Вальда доказано.

§ 50. Применения тождества Вальда. Теорема восстановления

Пример 1. Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные неотрицательные случайные величины с конечным математическим ожиданием: $\mathbf{P}\{\xi_1 \geq 0\} = 1, \mathbf{M}\xi_1 = a < \infty$. Положим $S_0 = 0, S_n = \xi_1 + \ldots + \xi_n, n \geq 1$, и рассмотрим случайные величины

$$\nu(t) = \min\{n : S_n > t\} = \sum_{n\geq 0} \chi\{S_n \leq t\}, t \geq 0.$$

При каждом t и каждом n, очевидно, $\{\nu(t) \leq n\} \in \sigma(\xi_1, \ldots, \xi_n)$; таким образом, $\nu(t)$ — момент остановки относительно последовательности σ-алгебр $\{\sigma(\xi_1, \ldots, \xi_n)\}$.

Наглядный смысл $\nu(t)$ — это число сумм $S_n, n \geq 0$, попадающих на отрезок $[0, t]$. Например, будем считать, что $\xi_1, \xi_2, \ldots$ — это времена работы (от включения до перегорания) электрических лампочек или времена обслуживания требований на периоде занятости системы массового обслуживания. Пусть в момент времени $t = 0$ мы включаем лампочку, а при перегорании каждой лампочки тут же заменяем ее новой (восстанавливаем). Тогда $\nu(t)$ — это число лампочек, которые придется включать до момента t (порядковый номер требования, обслуживаемого в момент t). В связи с такой интерпретацией поток случайных моментов времени S_n на $(0, \infty)$ называют также *процессом восстановления*, $\nu(t)$ — *числом восстановлений* до момента t, а функцию $U(t) = \mathbf{E}\nu(t)$ — *функцией восстановления*.

Функция $U(t)$ не убывает; она задает на $[0, \infty)$ меру, сопоставляющую каждому полуинтервалу $(a, b]$ значение $U(a) - U(b)$. Плотность $U'(t) = u(t)$ этой меры (если она существует) называется *плотностью восстановления*. Если все ξ_n принимают только целые значения, эта мера сосредоточена на множестве неотрицательных целых чисел, и ее плотностью называется последовательность $U(n) - U(n-1) = u(n)$ атомов, находящихся в точках $n = 0, 1, \ldots$

Случайный процесс $\nu(t)$ выглядит сложнее последовательности сумм независимых случайных величин, однако некоторые его характеристики можно довольно просто выразить в терминах преобразований Лапласа и производящих функций.

Теорема 5. *Если $U(t)$ — функция восстановления, построенная по независимым одинаково распределенным неотрицательным случайным величинам $\xi_1, \xi_2, \ldots$ с $\mathbf{M}exp\{-\lambda\xi_1\} = \psi(\lambda)$, то*

$$\int\limits_0^\infty e^{-\lambda t} dU(t) = \frac{1}{1 - \psi(\lambda)}, \mathrm{Re}\lambda \geq 0.$$

Если $\xi_1, \xi_2, \ldots$ принимают только целые неотрицательные значения, $\mathbf{E}s^{\xi_1} = f(s), |s| \leq 1$, и $u_n = U(n) - U(n-1)$, то

$$\sum_{n \geq 0} s^n u_n = \frac{1}{1 - f(s)}.$$

Доказательство. Так как по определению процесса восстановления $\nu(t) = \sum\limits_{n \geq 0} \chi\{S_n \leq t\}$, то $U(t) = \sum\limits_{n \geq 0} \mathbf{P}\{S_n \leq t\}$. Используя независимость ξ_i и свойства преобразований Лапласа, получаем

$$\int\limits_0^\infty e^{-\lambda t} dU(t) = \sum_{n=0}^\infty \int\limits_0^\infty e^{-\lambda t} d\mathbf{P}\{S_n \leq t\} = \sum_{n=0}^\infty \mathbf{E}e^{-\lambda S_n} = \sum_{n=0}^\infty \psi^n(\lambda) = \frac{1}{1 - \psi(\lambda)}.$$

Если ξ_i принимают только целые неотрицательные значения, то $dU(t) \neq 0$ только при целых t, и $dU(n) = u_n = U(n) - U(n-1)$ (если считать, что $U(-1) = 0$). Второе утверждение теоремы 5 получается из первого заменой $s = e^{-\lambda}$. Теорема 5 доказана.

Теорема восстановления. *Если* $\{\xi_n\}_{n\geqslant 1}$ *— последовательность независимых неотрицательных одинаково распределенных случайных величин,* $\mathbf{E}\xi_1 = a > 0$, $S_n = \xi_1 + \ldots + \xi_n$, $n \geq 1$, $\nu(t) = \min\{n : S_n > t\}$, *и* $U(t) = \mathbf{E}\nu(t)$, *то*

$$\lim_{n\to\infty} \frac{U(t)}{t} = \frac{1}{a}.$$

Сначала на нестрогом уровне обсудим утверждение теоремы. Естественно ожидать, что $S_{\nu(t)} \approx t$ при больших t; вычисляя математическое ожидание от обеих частей и пользуясь тождеством Вальда, находим: $t \approx \mathbf{E}S_{\nu(t)} = a\mathbf{E}\nu(t) = aU(t)$, что соответствует утверждению теоремы. Теперь проведем строгое доказательство.

Доказательство. Случайная величина $\nu(t)$ — момент остановки относительно потока σ-алгебр $\{\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)\}$, так как событие $\{\nu(t) \leq n\}$ полностью определяется значениями $\xi_1, \ldots, \xi_n$. По условию $\mathbf{E}\{|S_n - S_{n-1}||\mathcal{F}_n\} = \mathbf{E}|\xi_1| = \mathbf{E}\xi_1 < \infty$, и поэтому для того, чтобы обосновать возможность использования тождества Вальда, нужно только показать, что $\mathbf{E}\nu(t) < \infty$.

Рассмотрим преобразование Лапласа $\psi(\lambda) = \mathbf{E}\exp\{-\lambda\xi_1\}$. Так как $\chi\{x \leq t\} \leq \exp\{-\lambda(x-t)\}$ при любом $\lambda \geq 0$, то по свойствам преобразования Лапласа

$$\mathbf{P}\{\nu(t) > k\} = \mathbf{P}\{S_k \leq t\} = \mathbf{E}\chi\{S_k \leq t\} \leq \mathbf{E}\exp\{-\lambda(S_k - t)\} = \psi^k(\lambda)e^t.$$

Если $\lambda > 0$ и $\mathbf{P}\{\xi > 0\} > 0$, то $\psi(\lambda) < 1$, и при любом $\lambda > 0$

$$\mathbf{E}\nu(t) = \sum_{k=0}^{\infty} \mathbf{P}\{\nu(t) > k\} \leq e^t \sum_{k=0}^{\infty} \psi^k(\lambda) = \frac{e^t}{1-\psi(\lambda)} < \infty.$$

Значит, по тождеству Вальда

$$\mathbf{E}S_{\nu(t)} = a\mathbf{E}\nu(t) = aU(t).$$

Но по определению $S_{\nu(t)} > t$, следовательно, $aU(t) = \mathbf{E}S_{\nu(t)} > t$ и

$$\liminf_{t\to\infty} \frac{U(t)}{t} \geq \frac{1}{a}.$$

Казалось бы, что можно получить оценку $U(t)/t$ сверху, повторив те же рассуждения для $S_{\nu(t)-1} \le t$. Однако $\{\nu(t)-1 \le n\} = \{\nu(t) \le n+1\} \notin \mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$, т.е. $\nu(t) - 1$ — не момент остановки относительно потока $\mathcal{F}_n$. Таким образом, условия справедливости тождества Вальда для случайного момента $\nu(t) - 1$ не выполнены.

Случайную величину $S_{\nu(t)}$ можно оценить сверху при дополнительном условии, что $\mathbf{P}\{\xi_k \le z\} = 1$ при некотором $z < \infty$: тогда

$$\mathbf{P}\{S_{\nu(t)} = S_{\nu(t)-1} + \xi_{\nu(t)} \le t + z\} = 1.$$

Чтобы использовать это соображение, введем для каждого $z \in (0, \infty)$ вспомогательные случайные величины

$$\xi_k(z) = \min\{\xi_k, z\} \quad \text{и} \quad S_n(z) = \xi_1(z) + \ldots + \xi_n(z) \le S_n.$$

Тогда $\nu(t, z) = \min\{n : S_n(z) > t\} \ge \nu(t)$ и

$$U(t, z) = \mathbf{E}\nu(t, z) \ge \mathbf{E}\nu(t) = U(t).$$

Применяя к паре $S_n(z), \nu(t, z)$ тождество Вальда, получаем:

$$\mathbf{E}S_{\nu(t,z)}(z) = \mathbf{E}\xi_1(z) U(t, z).$$

Но по определению $S_{\nu(t,z)}(z) \le t + z$, следовательно,

$$U(t) \le U(t, z) = \frac{\mathbf{E}S_{\nu(t,z)}}{\mathbf{E}\xi_1(z)} \le \frac{t + z}{\mathbf{E}\xi_1(z)}$$

и для каждого $z < \infty$

$$\limsup_{t\to\infty} \frac{U(t)}{t} \le \frac{1}{\mathbf{E}\xi_1(z)} \lim_{t\to\infty} \frac{t + z}{t} = \frac{1}{\mathbf{E}\min\{\xi_1, z\}}\,.$$

Если $z \to \infty$, то $\mathbf{E}\min\{\xi_1, z\} \to \mathbf{E}\xi_1 = a$, и поэтому

$$\limsup_{t\to\infty} \frac{U(t)}{t} \le \frac{1}{a}.$$

Теорема доказана.

Пример. *Распределение числа восстановлений.* Теорема восстановления описывает асимптотику математического ожидания числа восстановлений $\nu(t) = \min\{n : S_n > t\}$ при $t \to \infty$. Дополнительную информацию о возможных значениях $\nu(t)$ дает теорема о предельном распределении этой случайной величины.

Теорема. *Если* $\{\xi_n\}_{n\geqslant 1}$ — *последовательность независимых неотрицательных одинаково распределенных случайных величин,* $\mathbf{E}\xi_1 = a > 0$, $\mathbf{D}\xi_1 = \sigma^2 < \infty$ *и* $S_0 = 0$, $S_n = \xi_1 + \ldots + \xi_n\,(n \geq 1)$, *а* $\nu(t) = \min\{n\colon S_n > t\}$, *то*

$$\lim_{t\to\infty} \mathbf{P}\left\{\frac{\nu(t) - \frac{t}{a}}{\sigma\sqrt{t/a^3}} \leqslant x\right\} = \Phi(x), \quad x \in (-\infty, \infty).$$

Иначе говоря, случайные величины $\nu(t)$ при $t \to \infty$ асимптотически нормальны с параметрами $\left(\frac{t}{a}, \frac{t\sigma^2}{a^3}\right)$.

Доказательство. Условия, наложенные на случайные величины $\{\xi_n\}$, означают, что к суммам S_n применима центральная предельная теорема:

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{S_n - an}{\sigma\sqrt{n}} \leqslant x\right\} = \Phi(x), \quad x \in (-\infty, \infty).$$

Пусть $[x]$, как обычно, обозначает целую часть x. Тогда $\{\nu(t) \leqslant m\} = \{S_{[m]} \geqslant t\}$ при любых $t, m \geqslant 0$ по определению $\nu(t)$. Отсюда следует, что

$$\mathbf{P}\{S_{[m]} \geqslant t\} = \mathbf{P}\left\{\frac{S_{[m]} - a[m]}{\sigma\sqrt{[m]}} \geqslant \frac{t - a[m]}{\sigma\sqrt{[m]}}\right\} =$$

$$= \mathbf{P}\{\nu(t) \leqslant m\} = \mathbf{P}\left\{\frac{\nu(t) - \frac{t}{a}}{a^{-3/2}\sigma\sqrt{t}} \leqslant \frac{m - \frac{t}{a}}{a^{-3/2}\sigma\sqrt{t}}\right\}.$$

Поэтому если $t, m \to \infty$ так, что $\frac{t-am}{\sigma\sqrt{m}} \to x \in (-\infty, \infty)$, то левая часть стремится к $1 - \Phi(x) = \Phi(-x)$, $t = am + x\sigma\sqrt{m} = am(1 + o(1))$, $m \to \infty$, а правая принимает вид

$$\mathbf{P}\left\{\frac{\nu(t) - \frac{t}{a}}{a^{-3/2}\sigma\sqrt{t}} \leqslant \frac{am - t}{a^{-1/2}\sigma\sqrt{am(1 + o(1))}} = -x(1 + o(1))\right\}.$$

Значит,

$$\mathbf{P}\left\{\frac{\nu(t) - \frac{t}{a}}{a^{-3/2}\sigma\sqrt{t}} \leqslant x\right\} \to \Phi(x).$$

Теорема доказана.

Она иллюстрирует один из способов переноса предельных теорем с последовательностей сумм независимых случайных величин на моменты первого достижения уровня этими суммами.

Список литературы

[1] Боровков А.А. Теория вероятностей. – М., "Эдиториал УРСС 1999.

[2] Ватутин В.А., Михайлов В.Г. Предельные теоремы для числа пустых ячеек в равновероятной схеме размещения частиц комплектами. – Теория вероятн. и ее примен., 1982, т.27, вып. 4, с. 684 - 692.

[3] Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей. – М., Наука, 1989.

[4] Колчин В.Ф., Севастьянов Б.А., Чистяков В.П. Случайные размещения. – М., Наука, 1976.

[5] Феллер В. Введение в теорию вероятностей и ее приложения, т.1. – М., Мир, 1984.