ЧАСТЬ 3

ПОСЛЕДОВАТЕЛЬНОСТЬ НЕЗАВИСИМЫХ ИСПЫТАНИЙ

Лекция 4

НЕЗАВИСИМЫЕ ИСПЫТАНИЯ. ФОРМУЛА БЕРНУЛЛИ. АСИМПТОТИЧЕСКИЕ ФОРМУЛЫ МУАВРА – ЛАПЛАСА И ПУАССОНА

ЦЕЛЬ ЛЕКЦИИ: ввести понятие независимого испытания и доказать формулу Бернулли; сформулировать асимптотические теоремы Муавра – Лапласа и Пуассона и указать границы их применимости.

При практических применениях теории вероятностей часто встречаются задачи, в которых один и тот же опыт или аналогичные опыты повторяются неоднократно. В результате каждого опыта может появиться или не появиться некоторое событие A. При этом нас интересует не результат каждого отдельного опыта, а общее число появлений события A в результате серии опытов. Например, если производится несколько выстрелов по одной и той же цели, то представляет интерес, как правило, не результат каждого выстрела, а общее число попаданий. В подобных задачах требуется определить вероятность любого заданного числа появлений события в результате серии опытов. Такие задачи решаются достаточно просто в случае, когда опыты являются независимыми.

Независимые испытания

Несколько опытов считаются независимыми, если вероятность того или иного исхода каждого из опытов не зависит от того, какие результаты имели другие опыты, например несколько последовательных бросаний монеты, несколько выниманий карты из колоды при условии ее возврата в колоду и перемешивания.

Независимые испытания могут проводиться как в одинаковых, так и в различных условиях. В первом случае вероятность события A во всех опытах одна и та же, и к нему относится частная теорема о повторении опытов.

Во втором случае вероятность события A от опыта к опыту меняется – общая теорема о повторении опытов.

Пример. Производятся три независимых выстрела по мишени с вероятностью попадания p при каждом выстреле. Найти вероятность ровно двух попаданий при трех выстрелах.

Р е ш е н и е. Событие $B_2 = \{$ в мишени ровно два попадания $\}$ может произойти тремя способами:

1) попаданием в первом и втором выстрелах, промахом в третьем;
2) попаданием в первом и третьем выстрелах, промахом во втором;
3) попаданием во втором и третьем выстрелах, промахом в первом.

Событие B_2 есть сумма трех несовместных вариантов:

$$B_2 = A_1 A_2 \overline{A}_3 + A_1 \overline{A}_2 A_3 + \overline{A}_1 A_2 A_3,$$

где A_i – попадание в i-м выстреле, $\overline{A}_i$ – промах.

Учитывая, что все три варианта события B_2 несовместны, а события, входящие в произведения, независимы, по правилам сложения и умножения вероятностей

$$P(B_2) = pp(1-p) + p(1-p)p + (1-p)pp .$$

Обозначив $q = 1 - p$, получаем

$$P(B_2) = 3p^2 q .$$

Аналогичным образом, перечисляя все возможные варианты, в которых интересующее нас событие может появиться заданное число раз, можно решить более общую задачу.

Формула Бернулли

Проводится n независимых опытов, в каждом из которых может появиться или не появиться некоторое событие A, вероятность появления равна p, а не появления – $q = 1 - p$. Требуется найти вероятность $P_{m,n}$ того, что событие A в этих n опытах появится ровно m раз.

Событие B_m – появление A ровно m раз – разложим на сумму произведений событий, состоящих в появлении или непоявлении A в отдельном опыте (A_i и $\overline{A}_i$). Каждый вариант события B_m (каждый член суммы) должен состоять из m появлений A и $n-m$ непоявлений, т. е.

$$B_m = A_1 A_2 \dots A_m \overline{A}_{m+1} \dots \overline{A}_n + \dots + A_1 \overline{A}_2 A_3 \dots \overline{A}_{n-1} A_n + \dots$$
$$\dots + \overline{A}_1 \overline{A}_2 \dots \overline{A}_{n-m} A_{n-m+1} \dots A_n ,$$

причем A входит в каждое слагаемое m раз, а $\overline{A}$ – $n-m$ раз.

Число комбинаций такого рода равно C_n^m. Вероятность каждой такой комбинации по теореме умножения для независимых событий равна $p^m q^{n-m}$. Так как варианты между собой несовместны, то по теореме сложения вероятность события B_m имеем

$$P(B_m) = P_{m,n} = \underbrace{p^m q^{n-m} + \dots + p^m q^{n-m}}_{C_n^m} = C_n^m p^m q^{n-m} .$$

Таким образом, можно сформулировать частную теорему о повторении опытов. *Если производится n независимых опытов, в каждом из которых событие A может появиться с вероятностью p, то вероятность того, что событие A появится ровно m раз, равна*

$$P_{m,n} = C_n^m p^m q^{n-m} , q = 1 - p . \qquad (3.1)$$

Соотношение (3.1) называется формулой Бернулли и описывает, как распределяются вероятности между возможными значениями некоторой случайной величины – числа появлений события A в n испытаниях. Так как вероятности $P_{m,n}$ по форме представляют собой члены разложения бинома $(p+q)^n$, то распределение вероятностей (3.1) называется биноминальным распределением.

В связи с тем что все возможные несовместные между собой исходы испытаний состоят в появлении события A 0 раз, 1 раз, 2 раза, …, n раз, то понятно, что

$$\sum_{m=0}^{n} P_{m,n} = 1.$$

Этот же результат может быть получен без учета теоретико-вероятностных соображений из равенства

$$\sum_{m=0}^{n} P_{m,n} = (p+q)^n = 1^n = 1.$$

Во многих практических задачах, кроме вероятности $P_{m,n}$ – появления события A ровно m раз, необходимо найти вероятность появлений события A не менее m. Для этого обозначим через C_m событие, состоящее в появлении события A не менее m раз, а его вероятность обозначим через $R_{m,n}$. Очевидно, что

$$C_m = B_m + B_{m+1} + \ldots + B_n,$$

откуда по теореме сложения

$$R_{m,n} = P_{m,n} + P_{m+1,n} + \ldots + P_{n,n},$$

т. е.

$$R_{m,n} = \sum_{i=m}^{n} P_{i,n}. \tag{3.2}$$

При вычислении $R_{m,n}$ часто удобнее не использовать соотношение (3.2), а перейти к противоположному событию и вычислять вероятность $R_{m,n}$ по формуле

$$R_{m,n} = 1 - \sum_{i=0}^{m-1} P_{i,n}.$$

Локальная и интегральная предельные теоремы

Рассмотрим пример, относящийся к независимым испытаниям, не доводя до конца вычисление искомых вероятностей.

Пример. По каналу связи передано сообщение, состоящее из 1000 нулей и единиц. Вероятности передачи как единицы, так и нуля одинаковы и равны 0,5. Найти вероятность того, что из 1000 переданных двоичных цифр число нулей окажется: а) ровно 500; б) не более 520.

Р е ш е н и е. В примере $n = 1000$, $p = 0{,}5$, $q = 0{,}5$, и поэтому:

а) число нулей окажется равным 500:

$$P_{500,1000} = C_{1000}^{500}(0{,}5)^{500}(0{,}5)^{500} = C_{1000}^{500}(0{,}5)^{1000}; \tag{3.3}$$

б) вероятность того, что число нулей окажется не более 520, равна сумме вероятностей, что число нулей окажется равным 0, 1, 2, …, 520, т. е.

$$P\{m \le 520\} = \sum_{i=0}^{520} P_{i,n} = \sum_{i=0}^{520} C_{1000}^{i}(0{,}5)^{i}(0{,}5)^{1000-i}. \tag{3.4}$$

Пример показывает, что непосредственное вычисление вероятностей по формулам (3.3) и (3.4) весьма трудоемко, и возникает задача нахождения простых приближенных формул для вычислений вероятностей $P_{m,n}$ и $\sum_{m=k}^{s} P_{m,n}$ при больших n.

Исследуем поведение вероятностей $P_{m,n}$ при постоянном n в зависимости от m. Для $0 \le m \le n$ получаем

$$\frac{P_{m+1,n}}{P_{m,n}} = \frac{C_n^{m+1} p^{m+1} q^{n-m-1}}{C_n^m p^m q^{n-m}} = \frac{(n-m)}{(m+1)} \frac{p}{q} . \qquad (3.5)$$

Из выражения (3.5) следует, что:

$P_{m+1,n} > P_{m,n}$, если $(n-m)p > (m+1)q$, т. е. $m < np - q$;

$P_{m+1,n} = P_{m,n}$, если $m = np - q$;

$P_{m+1,n} < P_{m,n}$, если $m > np - q$.

Видим, что с ростом m вероятность $P_{m,n}$ сначала возрастает, затем достигает максимума и наконец убывает. При этом если величина $np - q$ является целым числом, то максимального значения вероятность $P_{m,n}$ достигает при двух значениях m: $m_{01} = np - q$ и $m_{02} = np - q + 1 = np + p$. Если же $np - q$ не является целым, то максимального значения вероятность $P_{m,n}$ достигает при $m = m_0$, равном наименьшему целому числу, большему, чем $m_{01} = np - q$.

Если $np - q < 0$, то

$$P_{0,n} > P_{1,n} > \ldots > P_{n,n} .$$

При $np - q = 0$

$$P_{0,n} = P_{1,n} > P_{2,n} > \ldots > P_{n,n} .$$

Оказалось, что при больших n почти все вероятности $P_{m,n}$ очень малы. И только для m близких к вероятнейшему значению m_0 вероятности $P_{m,n}$ сколько-нибудь заметно отличаются от нуля. Такое поведение вероятности $P_{m,n}$ при больших n и лежит в основе локальной и интегральной теорем Муавра – Лапласа.

Впервые асимптотическую формулу, облегчающую вычисление $P_{m,n}$ при больших n, нашел Муавр в 1730 г. для частного случая при $p = q = 1/2$, а затем обобщил Лаплас для произвольного p, отличного от 0 и 1.

Вводится обозначение

$$x = \frac{m - np}{\sqrt{npq}},$$

т. е. величина x зависит как от n и p, так и от m.

<u>**Локальная теорема Муавра – Лапласа**</u> (без доказательства). Если вероятность наступления некоторого события A в n независимых испытаниях постоянна и равна p $(0 < p < 1)$, то вероятность $P_{m,n}$ того, что в этих испытаниях событие A наступит ровно m раз, удовлетворяет соотношению

$$\sqrt{npq}P_{m,n} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \qquad (3.6)$$

Теперь решим задачу а) рассматриваемого примера, используя соотношение (3.6). Нужно найти $P_{m,n}$ при $n = 1000$, $m = 500$ и $p = 0{,}5$.

По формуле (3.6) имеем

$$P_{m,n} = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}\left(\frac{m-np}{\sqrt{npq}}\right)^2}.$$

Для нашего примера получаем $\sqrt{npq} \approx 15{,}81$, $x = 0$ и соответственно

$$P_{m,n} = \frac{1}{15{,}81\sqrt{2\pi}} e^{-0}.$$

Функция $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ табулирована (см. прил. 1). Так как значение $\varphi(0) = 0{,}3989$, то окончательно получаем

$$P_{m,n} = \frac{0{,}3989}{15{,}81} = 0{,}025231.$$

Точные подсчеты по формуле Бернулли (3.1) дают

$$P_{m,n} = 0{,}025225 .$$

Интегральная теорема Муавра – Лапласа (без доказательства). Если m есть число наступлений события A в n независимых испытаниях, в каждом из которых вероятность этого события равна p, причем $0 < p < 1$, то равномерно относительно a и b $(-\infty \le a < b \le \infty)$ имеет место соотношение

$$P\left\{a \le \frac{m-np}{\sqrt{npq}} \le b\right\} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz . \qquad (3.7)$$

Решение задачи б) при использовании формулы (3.7) требует умения вычислять значение интеграла Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz \qquad (3.8)$$

при любых значениях x. Так как интеграл (3.8) при $0 < x \le \infty$ через элементарные функции не выражается, то для вычислений интеграла Лапласа требуются специальные таблицы (прил. 2).

Интеграл

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dz = \Phi(b) - \Phi(a)$$

вычисляем через значения функции $\Phi(x)$, причем в приложении 2 приведены значения $\Phi(x)$ только для положительных x, так как интеграл Лапласа является нечетной функцией, для которой выполняется условие, что $\Phi(-x) = -\Phi(x)$ (см. рис. 3.1).

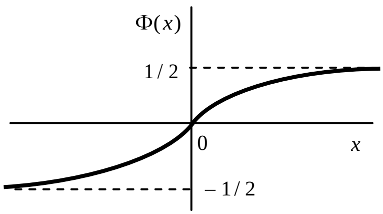

Рис. 3.1. Интеграл Лапласа $\Phi(x)$

Теперь решим задачу б) рассматриваемого примера, используя соотношение (3.7).

$$P\{m \le 520\} = P\{0 \le m \le 520\} =$$

$$= P\left\{\frac{0-np}{\sqrt{npq}} \le \frac{m-np}{\sqrt{npq}} \le \frac{520-np}{\sqrt{npq}}\right\}.$$

После подстановки значений n, p, q получаем

$$P\left\{-34{,}79 \le \frac{m-np}{\sqrt{npq}} \le 1{,}26\right\} \approx \frac{1}{\sqrt{2\pi}} \int_{-34{,}79}^{1{,}26} e^{-\frac{z^2}{2}} dz =$$

$$= \Phi(1{,}26) - \Phi(-34{,}79) = 0{,}39617 + 0{,}5 = 0{,}89617.$$

Значение $\Phi(-34{,}79) \approx -0{,}5$, так как уже величина $\Phi(5) = 0{,}49999997$ (прил. 2).

<u>Типичная задача, приводящая к интегральной теореме Муавра – Лапласа</u>. Проводится n независимых испытаний, в каждом из которых вероятность наступления события A равна p. Чему равна вероятность того, что частота наступления события A отклонится от вероятности p не более чем на α?

<u>Р е ш е н и е</u>. Искомая вероятность равна

$$P\left\{\left|\frac{m}{n} - p\right| \le \alpha\right\} = P\left\{-\alpha\sqrt{\frac{n}{pq}} \le \frac{m-np}{\sqrt{npq}} \le \alpha\sqrt{\frac{n}{pq}}\right\} \approx$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{-\alpha\sqrt{\frac{n}{pq}}}^{\alpha\sqrt{\frac{n}{pq}}} e^{-\frac{z^2}{2}} dz = \frac{2}{\sqrt{2\pi}} \int_{0}^{\alpha\sqrt{\frac{n}{pq}}} e^{-\frac{z^2}{2}} dz = 2\Phi\left(\alpha\sqrt{\frac{n}{pq}}\right).$$

Естественно, что в задачах, относящихся к определению вероятностей $P_{m,n}$ при конечных m и n асимптотическими формулами Муавра – Лапласа, требуется производить оценку совершаемой при этом ошибки. В течение очень долгого времени теоремы Муавра – Лапласа применялись к решению подобного рода задач без сколько-нибудь удовлетворительной оценки остаточного члена. Создалась чисто эмпирическая уверенность, что при n порядка нескольких сотен или еще больше и p, не слишком близких к 0 и 1, использование теорем Муавра – Лапласа приводит к удовлетворительным результатам. В настоящее время существуют достаточно хорошие оценки погрешностей, совершаемых при употреблении асимптотических формул Муавра – Лапласа.

Терема Пуассона

Было замечено, что асимптотическое представление вероятности $P_{m,n}$ посредством функции $\varphi(x)$ получается тем хуже, чем больше p отличается от $1/2$, т. е. чем меньшее p или q приходится рассматривать. Однако значительное количество задач связано с необходимостью вычислять вероятности $P_{m,n}$ именно при малых p. То есть, чтобы теорема Муавра – Лапласа дала приемлемый результат, необходимо произвести очень большое число n независимых испытаний. Задача нахождения асимптотической формулы вычисления вероятностей $P_{m,n}$ при малых p решена теоремой Пуассона.

<u>Теорема Пуассона</u>. Если $p \to 0$, то вероятность ровно m положительных исходов при n испытаниях равна

$$P_{m,n} \xrightarrow[n\to\infty]{} \frac{\lambda^m}{m!} e^{-\lambda}, \tag{3.9}$$

где $\lambda = np$.

<u>Пример</u>. Из одной ЭВМ на другую необходимо передать файл объемом 8 000 символов. Вероятность ошибки при передаче символа равна 0,001. Найти вероятность того, что будет не менее двух ошибок при передаче файла.

<u>Р е ш е н и е</u>. Считая передачу каждого символа как испытание, а ошибку как событие, можно вычислить вероятность $P\{m \geq 2\}$, используя формулу (3.9) при $\lambda = np = 8$

$$P\{m \geq 2\} = \sum_{m=2}^{8000} P_{m,8000} = 1 - (P_{0,n} + P_{1,n}) =$$

$$= 1 - (e^{-8} + 8e^{-8}) = 0{,}99698 .$$

Вычисление по точной формуле (3.1) дает

$$P\{m \geq 2\} = 0{,}99699 ,$$

т. е. ошибка меньше 0,001 %.

Практические соображения по применению теоремы Пуассона:

$$p < 0{,}1;\ npq \leq 9 .$$