Лекция 3

Повторные независимые испытания. Упрощения формулы Бернулли

1. Вероятность возникновения независимых повторных событий

- Если вероятность наступления события *A* в каждом испытании не меняется в зависимости от исходов других, то такие испытания называются *независимыми* относительно события *A*.
- Если независимые повторные испытания проводятся при одном и том же комплексе условий, то вероятность наступления события *А* в каждом испытании одна и та же. Описанная последовательность независимых испытаний получила название **схемы Бернулли**.
- Априорная вероятность это вероятность, присвоенная событию при отсутствии знания, поддерживающего его наступление.
- Апостериорная вероятность это условная вероятность события при некотором условии, рассматриваемая в противоположность его априорной вероятности.

• *Теорема*. Если вероятность p наступления события A в каждом испытании постоянна, то вероятность $P_{m,n}$ того, что событие A наступит m раз в n независимых испытаниях, равна

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

Доказательство. Пусть A_i и $\bar{A_i}$ - соответственно появление и непоявление события A в i-м испытании (i=1,2,...,n), а \mathbf{B}_m - событие, состоящее в том, что в n независимых испытаниях событие A появилось m раз. Представим событие B_m через элементарные события A_i .

Например, при n = 3, m = 2 событие $B_2 = A_1A_2\bar{A}_3 + \bar{A}_1A_2A_3 + A_1\bar{A}_2A_3$, т.е. событие А произойдет два раза в трех испытаниях, если оно произойдет в 1-м и 2-м испытаниях (и не произойдет в 3-м); в 1-м и 3-м (и не произойдет во 2-м), или произойдет во 2-м и в 3-м (и не произойдет в 1-м).

В общем виде

$$B_{m} = A_{1}A_{2} \dots A_{m}\bar{A}_{m+1}\dots\bar{A}_{n} + A_{1}\bar{A}_{2}A_{3}\dots\bar{A}_{n-1}A_{n} + \dots + +\bar{A}_{1}\bar{A}_{2}\dots\bar{A}_{n-m}A_{n-m+1}\dots A_{n},$$
(3.2)

т.е. каждый вариант появления события Bm состоит из m появлений события A и n-m непоявлений, т.е. из m событий A и из n-m событий \bar{A} с различными индексами.

Число всех комбинаций (слагаемых суммы (3.2)) равно числу способов выбора из n испытаний m, в которых событие A произошло, т.е. числу сочетаний C_n^m .

• Вероятность каждой такой комбинации (каждого варианта появления события Bm) по теореме умножения для независимых событий равна $p^m q^{n-m}$, так как

$$p(A_i) = p, \ p(\bar{A}_i) = q, \ i = 1, 2, ..., n.$$

В связи с тем, что комбинации между собой несовместны, по теореме сложения вероятностей получим

$$P_{m,n} = P(B_m) = p^m q^{n-m} + ... + p^m q^{n-m} = C_n^m p^m q^{n-m}$$
 C_n^m pas

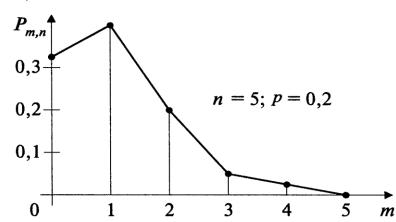
Пример. Вероятность изготовления на автоматическом станке стандартной детали равна 0,8. Найти вероятности возможного числа появления бракованных деталей среди 5 отобранных.

Решение. Вероятность изготовления бракованной детали p = 1 - 0.8 = 0.2. Искомые вероятности находим по формуле Бернулли (3.1):

$$P_{0,5} = C_5^0 0, 2^0 0, 8^5 = 0,32768; P_{1,5} = C_5^1 0, 2^1 0, 8^4 = 0,4096;$$

 $P_{2,5} = C_5^2 0, 2^2 0, 8^3 = 0,2048; P_{3,5} = C_5^3 0, 2^3 0, 8^2 = 0,0512;$
 $P_{4,5} = C_5^4 0, 2^4 0, 8^1 = 0,0064; P_{5,5} = C_5^5 0, 2^5 0, 8^0 = 0,00032.$

Полученные вероятности изобразим графически точками с координатами $(m, P_{m,n})$. Соединяя эти точки, получим многоугольник, или полигон, распределения вероятностей



- Число m_0 наступления события A в n независимых испытаниях называется **наивероятнейшим**, если вероятность осуществления этого события $P_{m_0,n}$ по крайней мере не меньше вероятностей других событий $P_{m,n}$ при любом m.
- Для нахождения m_0 запишем систему неравенств:

$$\begin{cases}
P_{m_0,n} \ge P_{m_0+1,n}, \\
P_{m_0,n} \ge P_{m_0-1,n}.
\end{cases}$$
(3.3)

- Решим первое неравенство системы (3.3).
- Используя формулы Бернулли и числа сочетаний, запишем

$$\frac{n!}{m_0!(n-m_0)!}p^{m_0}q^{n-m_0}\geq \frac{n!}{(m_0+1)!(n-m_0-1)!}p^{m_0+1}q^{n-m_0-1}.$$

• Так как $(m_0+1)!=m_0!(m_0+1)$, $(n-m_0)!=(n-m_0-1)!(n-m_0)$, то получим после упрощений неравенство

$$rac{1}{n-m_0}q \geq rac{1}{m_0+1}p$$
, откуда (m_0+1) $q \geq (n-m_0)$ p .

Теперь m_0 $(p+q) \geq np-q$ или $m_0 \geq np-q$

Решая второе неравенство системы (3.3), получим аналогично: $m_0 \le np + p$.

Объединяя полученные решения двух неравенств, придем к двойному неравенству:

$$np - q \le m_0 \le np + p \tag{3.4}$$

Так как np + p - (np - q) = p + q = 1, то всегда существует целое число m_0 , удовлетворяющее неравенству (3.4).

При этом, если np + p — целое число, то наивероятнейших чисел два:

$$m_0 = np + p$$
 $m'_0 = np - q$.

• *Пример*. Сколько раз необходимо подбросить игральную кость, чтобы наивероятнейшее выпадение тройки было равно 10?

Решение. В данном случае
$$p = 1/6$$
. Согласно неравенству (3.4) $n \cdot \frac{1}{6} - \frac{5}{6} \le 10 \le n \cdot \frac{1}{6} + \frac{1}{6}$

$$n-5 \le 60 \le n+1 => 59 \le n \le 65$$
, т.е. необходимо подбросить кость от 59 до 65 раз (включительно).

1. Вероятность возникновения независимых повторных событий. Алгоритм решения задач

- Проверить, выполняются ли условия схемы повторных независимых испытаний.
- 2. Если опыт, описываемый в задаче, приводит к схеме повторных независимых испытаний, то из условий определить: P(A) = p - вероятность успеха, и $P(\overline{A}) = 1 - p = q$
- вероятность неудачи в каждом опыте.
- При заданных числе опытов, количестве успехов (и (или) неудач) воспользоваться формулами (3.1)-(3.4)

Часто применяемые формулы в схеме Бернулли

Вероятность наступления события A:

a) менее
$$k$$
 pas: $P_n(0) + P_n(1) + ... + P_n(k-1)$

б) более
$$k$$
 раз: $P_n(k+1) + P_n(k+2) + ... + P_n(n)$ Д) хотя бы один раз: $P_n(k) + P_n(k+1) + P_n(n)$ $1 - P_n(0)$

в) не менее k раз: $P_n(k) + P_n(k+1) + ... + P_n(n)$

 Γ) не более k раз: $P_n(0) + P_n(1) + ... + P_n(k)$

1. Вероятность возникновения независимых повторных событий. Алгоритм решения задач

• *Пример*. В среднем 20% пакетов акций на аукционах продаются по первоначально заявленной цене. Найти вероятность того, что из 9 пакетов акций в результате торгов по первоначально заявленной цене: 1) не будут проданы 5 пакетов; 2) будет продано: а) менее 2 пакетов; б) не более 2; в) хотя бы 2 пакета; г) наивероятнейшее число пакетов.

Решение. 1) Вероятность того, что пакет акций не будет продан по первоначально заявленной цене, p = 1 - 0.2 = 0.8.

По формуле Бернулли $P_{5,9} = C_9^5 \cdot 0.8^5 \cdot 0.2^4 = 0.066.$

2.a) По условию p = 0,2.

$$P_9(m<2) = P_{0.9} + P_{1.9} = C_9^0 \cdot 0.2^0 \cdot 0.8^9 + C_9^1 \cdot 0.2 \cdot 0.8^8 = 0.436.$$

Пример. В среднем 20% пакетов акций на аукционах продаются по первоначально заявленной цене. Найти вероятность того, что из 9 пакетов акций в результате торгов по первоначально заявленной цене: 1) не будут проданы 5 пакетов; 2) будет продано: а) менее 2 пакетов; б) не более 2; в) хотя бы 2 пакета; г) наивероятнейшее число пакетов.

• 2.г) Наивероятнейшее число проданных акций по первоначально заявленной цене определится из условия (4.4), т.е.

$$9 \cdot 0,2 - 0,8 \le m_0 \le 9 \cdot 0,2 + 0,2$$
 или $1 \le m_0 \le 2$,

т.е. наивероятнейших чисел два: m_0 =1 и m_0' =2. Поэтому вероятность

$$P_{\text{Haubep}} = P_{1,9} + P_{2,9} = C_9^1 \cdot 0.2 \cdot 0.8^8 + C_9^2 \cdot 0.2^2 \cdot 0.8^7 = 0.604.$$

1. Переменные условия опыта. Производящая функция

- Пусть производится n независимых испытаний, причём в первом испытании вероятность появления события A равна p_1 , во втором $-p_2$, ..., в n-м испытании $-p_n$; вероятности непоявления события A соответственно равны $q_1, q_2, ..., q_n$; $P_n(k)$ вероятность появления события A в n испытаниях ровно k раз.
- Производящей функцией вероятностей $P_n(k)$ называют функцию, определяемую равенством

$$\varphi_n(z) = (p_1 z + q_1)(p_2 z + q_2)...(p_n z + q_n).$$

• Вероятность $P_n(k)$ того, что в n независимых испытаниях, в первом из которых вероятность появления события A равна p_1 , во втором $-p_2, \ldots$, в n-м испытании $-p_n$; $P_n(k)$ — вероятность появления события A в n испытаниях ровно k раз, равна коэффициенту z^k в разложении производящей функции по степеням z. $P_2(2)$ $P_2(1)$ $P_2(0)$

$$\varphi_2(z) = (p_1 z + q_1)(p_2 z + q_2) = p_1 p_2 z^2 + (p_1 q_2 + p_2 q_1)z + q_1 q_2$$

1. Переменные условия опыта. Производящая функция

• **Пример.** Устройство состоит из трех независимо работающих элементов (за время t) соответственно равны p_1 =0,7; p_2 =0,8; p_3 =0,9. Найти вероятности того, что за время t будут работать безотказно: а) все элементы, б) два элемента, в) один элемент, д) ни один из элементов.

Решение. Так как вероятности безотказной работы элементов соответственно равны p_1 =0,7; p_2 =0,8; p_3 =0,9., то вероятности того, что элементы откажут, равны q_1 =0,3; q_2 =0,2; q_3 =0,1.

Составим производящую функцию

$$\varphi_3(z) = (p_1 z + q_1)(p_2 z + q_2)(p_3 z + q_3) = 0.504 z^3 + 0.398 z^2 + 0.092 z + 0.006.$$

Вероятность того, что три элемента будут работать безотказно, равна коэффициенту при z^3 Вероятность того, что два элемента будут работать безотказно, равна коэффициенту при z^2 Вероятность того, что один элемент будет работать безотказно, равна коэффициенту при z^1 Вероятность того, что ни один элемент не будет работать безотказно, равна коэффициенту при z^0

- Задача: вычислить вероятность $P_{m,n}$ появления события A при большом числе испытаний n, например, $P_{300,500}$.
- По формуле Бернулли $P_{300,500} = C_{500}^{300} p^{300} q^{200} = \frac{500!}{300!200!} p^{300} q^{200}$

Непосредственное вычисление технически сложно (p и q - числа дробные)!

• Существуют более простые приближенные формулы для вычисления $P_{m\,n}$ при больших n — так называемые асимптотические формулы, определяются теоремой Пуассона, локальной и интегральной теоремами Муавра-Лапласа.

Теорема Пуассона

Теорема. Если вероятность р наступления события A в каждом испытании стремится к нулю $(p \to 0)$ при неограниченном увеличении числа n испытаний $(n \to \infty)$, причем произведение np стремится к постоянному числу λ ($np \to \lambda$), то вероятность $P_{m,n}$ того, что событие Aраз в n независимых испытаниях, удовлетворяет предельному равенству $\lim_{n\to\infty} P_{m,n} = P_m(\lambda) = \frac{\lambda^m e^{-\lambda}}{m!}$

Теорема Пуассона

Доказательство: По формуле Бернулли

$$P_{m,n} = C_n^m p^m q^{n-m} = \frac{n(n-1)(n-2)...(n-m+1)}{m!} p^m (1-p)^n (1-p)^{-m}$$

или, учитывая, что $\lim_{n\to\infty} np = \lambda$ (при достаточно больших n):

$$p \approx \frac{\lambda}{n} \text{ и } P_{m,n} \approx \frac{\lambda^m}{m!} \left(1 \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{m-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-m}.$$

Так как

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right) = \lim_{n\to\infty} \left(1 - \frac{2}{n}\right) = \dots = \lim_{n\to\infty} \left(1 - \frac{m-1}{n}\right) = 1,$$

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = \lim_{n\to\infty} \left(\left(1 - \frac{\lambda}{n}\right)^{\frac{-n}{\lambda}}\right)^{-\lambda} = e^{-\lambda} \qquad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-m} = 1,$$

TO

$$\lim_{n\to\infty} P_{m,n} = \frac{\lambda^m}{m!} e^{-\lambda}$$

Замечание

- Строго говоря, условие теоремы Пуассона $p \to 0$ при $n \to \infty$, так что $np \to \lambda$, противоречит исходной предпосылке схемы испытаний Бернулли, согласно которой вероятность наступления события в каждом испытании p = const.
- Однако, если вероятность p постоянна и мала, число испытаний n велико и число $\lambda = np$ незначительно (будем полагать, что $\lambda = np < 10$), то из предельного равенства (3.5) вытекает приближенная формула Пуассона:

$$P_{m,n} pprox rac{\lambda^m e^{-\lambda}}{m!} = P_m(\lambda)$$
 (3.6)
Формула Пуассона

Значения функции Пуассона $P_m(\lambda)$ табулированы.

- *Пример 1*. На факультете насчитывается 1825 студентов. Какова вероятность того, что 1 сентября является днем рождения одновременно четырех студентов факультета?
- Решение. Вероятность того, что день рождения студента 1 сентября, равна p = 1/365. Так как p = 1/365 мала, n = 1825 велико и

 $\lambda = np = 1825(1/365) = 5 < 10$, то применяем формулу Пуассона

$$P_{4,1825} = P_4(5) = 0.1755$$

2. Упрощения формулы Бернулли. Локальная теорема Муавра — Лапласа.

• *Теорема*. Если вероятность p наступления события A в каждом испытании постоянна и отлична от 0 и 1, то вероятность $P_{m,n}$ того, что событие A произойдет m раз в n независимых испытаниях при достаточно большом числе n, приближенно равна

$$P_{m,n} pprox rac{f(x)}{\sqrt{npq}}$$
 (3.7) локальная формула Муавра — Лапласа где $f(x) = rac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (3.8) функция Гаусса и $x = rac{m-np}{\sqrt{npq}}$ (3.9)

Приближенные значения вероятности $P_{m,n}$, даваемые локальной формулой Myaвpa-Лапласа на практике используются как точные при npq порядка двух и более десятков, т.е. при условии npq>20.

Для упрощения расчетов, связанных с применением формулы (3.7), составлена таблица значений функции f(x))

2. Упрощения формулы Бернулли. Локальная теорема Муавра — Лапласа.

Свойства функции Γ аусса f(x) (3.8)

- 1. Функция f(x) является четной, т.е. f(-x) = f(x).
- 2. Функция f(x) монотонно убывающая при положительных значениях x, причем при $x \to \infty f(x) \to 0$.

(Практически можно считать, что уже при x > 4 $f(x) \approx 0$.)

Пример. В некоторой местности из каждых 100 семей 80 имеют холодильники. Найти вероятность того, что из 400 семей 300 имеют холодильники.

Решение. Вероятность того, что семья имеет холодильник, равна p = 80/100 = 0,8. Так как n = 100 достаточно велико (условие $npq = 100 \cdot 0,8(1 \cdot 0,8)=64 \ge 20$ выполнено), то применяем локальную формулу Муавра − Лапласа.

Вначале определим по (3.9)
$$x = \frac{300-400 \cdot 0.8}{\sqrt{400 \cdot 0.8 \cdot 0.2}} = -2,50$$

Тогда по формуле (3.7) $P_{300,400} \approx \frac{f(-2,50)}{\sqrt{100 \cdot 0.8 \cdot 0.2}} = \frac{f(-2,50)}{\sqrt{64}} = \frac{0,0175}{8} \approx 0,0022$

Интегральная теорема Муавра-Лапласа. Если вероятность p наступления события A в каждом испытании постоянна и отлична от 0 и 1, то вероятность того, что число m наступления события A в n независимых испытаниях заключено в пределах от a до b (включительно), при достаточно большом числе n приближенно равна

$$P_n(a \le m \le b) \approx \frac{1}{2} [\Phi(x_2) - \Phi(x_1)],$$
 (4.1)

где
$$\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$$
 (4.2) - функция (или интеграл вероятностей) Лапласа;
$$x_1 = \frac{a - np}{\sqrt{npq}}, \quad x_2 = \frac{b - np}{\sqrt{npq}}.$$
 (4.3)

Свойства функции $\Phi(x)$

- 1. Функция $\Phi(x)$ нечетная, т.е. $\Phi(-x) = -\Phi(x)$.
- 2. Функция $\Phi(x)$ монотонно возрастающая, причем при $x \to \infty \to 1$ (практически можно считать, что уже при x > 4 $\Phi(x) \approx 1$).

Следствие интегральной теоремы Муавра-Лапласа. Если вероятность p наступления события A в каждом испытании постоянна и отлична от 0 и 1, то при достаточно большом числе n независимых испытаний вероятность того, что:

а) число m наступлений события A отличается от произведения np не более, чем на величину $\varepsilon > 0$ (по абсолютной величине), т.е.

$$P_n(|m-np| \le \varepsilon) \approx \Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right);$$
 (4.4)

Следствие интегральной теоремы Муавра-Лапласа.

б) частость $\frac{m}{n}$ события A заключена в пределах от α до β (включительно), т.е.

$$P_n\left(\alpha \le \frac{m}{n} \le \beta\right) \approx \frac{1}{2} \left[\Phi(z_2) - \Phi(z_1)\right],\tag{4.5}$$

где
$$z_1 = \frac{\alpha - p}{\sqrt{pq/n}}, \quad z_2 = \frac{\beta - p}{\sqrt{pq/n}}.$$
 (4.6)

в) частость $\frac{m}{n}$ события A отличается от его вероятности p не более, чем на величину $\Delta > 0$ (по абсолютной величине), т.е.

$$P_n\left(\left|\frac{m}{n}-p\right| \le \Delta\right) \approx \Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{pq}}\right).$$
 (4.7)

Следствие интегральной теоремы Муавра-Лапласа.

Доказательство. a) Неравенство $|m-np| \le \varepsilon$ равносильно двойному неравенству $np - \varepsilon \le m \le np + \varepsilon$.

Поэтому по интегральной формуле (4.1)

$$\begin{split} P_n\Big(\big|m-np\big| \leq \varepsilon\Big) &= P_n\Big(np-\varepsilon \leq m \leq np+\varepsilon\Big) \approx \\ &\approx \frac{1}{2}\Bigg[\Phi\bigg(\frac{p+\varepsilon-np}{\sqrt{npqn}}\bigg) - \Phi\bigg(\frac{np-\varepsilon-np}{\sqrt{npq}}\bigg)\Bigg] = \frac{1}{2}\Bigg[\Phi\bigg(\frac{\varepsilon}{\sqrt{npq}}\bigg) - \Phi\bigg(\frac{-\varepsilon}{\sqrt{npq}}\bigg)\Bigg] = \\ &= \frac{1}{2}\Bigg[\Phi\bigg(\frac{\varepsilon}{\sqrt{npq}}\bigg) + \Phi\bigg(\frac{\varepsilon}{\sqrt{npq}}\bigg)\Bigg] = \Phi\bigg(\frac{\varepsilon}{\sqrt{npq}}\bigg). \end{split}$$

б) Неравенство $\alpha \leq \frac{m}{n} \leq \beta$ равносильно неравенству $a \leq m \leq b$ при $a = n\alpha$ и $b = n\beta$.

Заменяя в формулах (4.1), (4.3) величины a и b полученными выражениями, получим доказываемые формулы (4.5) и (4.6).

в) Неравенство $\left|\frac{m}{n} - p\right| \le \Delta$ равносильно неравенству $\left|m - np\right| \le \Delta n$.

Заменяя в формуле (4.4) $\varepsilon = \Delta n$, получим доказываемую формулу (4.7).

Пример. По статистическим данным в среднем 87% новорожденных доживают до 50 лет. Найти вероятность того, что из 1000 новорожденных доля (частость) доживших до 50 лет будет: а) заключена в пределах от 0,9 до 0,95; б) будет отличаться от вероятности этого события не более, чем на 0,04 (по абсолютной величине).

Pешение. а) Вероятность p того, что новорожденный доживет до 50 лет, равна 0.87.

Так как n=1000 велико (условие npq=1000 - 0.87 - $0.13=113.1 \ge 20$ выполнено), то используем следствие интегральной теоремы Муавра-Лапласа.

Вначале определим по (4.6)
$$z_1 = \frac{0.9 - 0.87}{\sqrt{0.87 \cdot 0.13/1000}} = 2.82, \ z_2 = \frac{0.95 - 0.87}{\sqrt{0.87 \cdot 0.13/1000}} = 7.52.$$

Теперь по формуле (4.5)
$$P_{1000}\bigg(0.9 \le \frac{m}{n} \le 0.95\bigg) \approx \frac{1}{2} \big[\Phi(7,52) - \Phi(2,82)\big] =$$
$$= \frac{1}{2} \big(1 - 0.9952\big) = 0.0024.$$

б) По формуле (4.7)
$$P_{1000} \left(\left| \frac{m}{n} - 0.87 \right| \le 0.04 \right) \approx \Phi \left(\frac{0.04 \cdot \sqrt{1000}}{\sqrt{0.87 \cdot 0.13}} \right) = \Phi(3.76) = 0.9998.$$

Так как неравенство $\left| \frac{m}{n} - 0.87 \right| \le 0.04$ равносильно неравенству $0.83 \le \frac{m}{n} \le 0.91$, полученный результат означает, что практически достоверно, что от 0.83 до 0.91 числа новорожденных из 1000 доживут до 50 лет.