Формула полной вероятности.

Пусть имеется группа событий H_1, H_2,..., H_n , обладающая следующими свойствами:

1) Все события попарно несовместны: $H_i \cap H_j = \varnothing$; $i, j=1,2,...,n$; $i \neq j$

2) Их объединение образует пространство элементарных исходов Ω:

$$\Omega = H_1 \cup H_2 \cup \ldots \cup H_n.$$

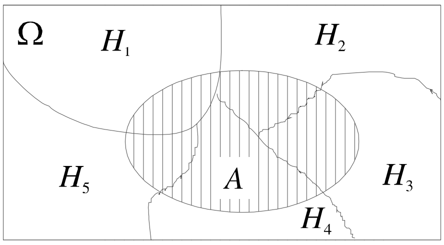

Рис.8

В этом случае будем говорить, что $H_1, H_2,...,H_n$ образуют **полную группу событий**. Такие события иногда называют **гипотезами**.

Пусть A - некоторое событие: $A \subset \Omega$ (диаграмма Венна представлена на рисунке 8). Тогда имеет место **формула полной вероятности:**

$$P(A) = P(A/H_1)P(H_1) + P(A/H_2)P(H_2) + ...+ P(A/H_n)P(H_n) == \sum_{i=1}^{n} P(A/H_i)P(H_i)$$

Доказательство. Очевидно: $A = (A \cap H_1) \cup (A \cap H_2) \cup ... \cup (A \cap H_n)$, причем все события $A \cap H_i$ ($i = 1,2,...,n$) попарно несовместны. Отсюда по теореме сложения вероятностей получаем

$$P(A) = P(A \cap H_1) + P(A \cap H_1) +...+ P(A \cap H_n)$$

Если учесть, что по теореме умножения $P(A \cap H_i) = P(A/H_i)\, P(H_i)$ ($i = 1,2,...,n$), то из последней формулы легко получить приведенную выше формулу полной вероятности.

Пример. В магазине продаются электролампы производства трех заводов, причем доля первого завода - 30%, второго - 50%, третьего - 20%. Брак в их продукции составляет соответственно 5%, 3% и 2%. Какова вероятность того, что случайно выбранная в магазине лампа оказалась бракованной.

Пусть событие H_1 состоит в том, что выбранная лампа произведена на первом заводе, H_2 на втором, H_3 - на третьем заводе. Очевидно:

$$P(H_1) = 3/10,\ P(H_2) = 5/10,\ P(H_3) = 2/10.$$

Пусть событие A состоит в том, что выбранная лампа оказалась бракованной; A/H_i означает событие, состоящее в том, что выбранна бракованная лампа из ламп, произведенных на i-ом заводе. Из условия задачи следует:

$$P(A/H_1) = 5/10;\ P(A/H_2) = 3/10;\ P(A/H_3) = 2/10$$

По формуле полной вероятности получаем

$$P(A) = \frac{3}{10} \cdot \frac{5}{100} + \frac{5}{10} \cdot \frac{3}{100} + \frac{2}{10} \cdot \frac{2}{100} = \frac{17}{500}$$

Формула Байеса

Пусть $H_1,H_2,...,H_n$ - полная группа событий и $A\subset\Omega$ - некоторое событие. Тогда по формуле для условной вероятности

$$P(H_k / A) = \frac{P(H_k \cap A)}{P(A)} \qquad (*)$$

Здесь $P(H_k /A)$ - условная вероятность события (гипотезы) H_k или вероятность того, что H_k реализуется при условии, что событие A произошло.

По теореме умножения вероятностей числитель формулы (*) можно представить в виде

$$P(H_k\cap A) = P(A\cap H_k) = P(A /H_k)\, P(H_k)$$

Для представления знаменателя формулы (*) можно использовать формулу полной вероятности

$$P(A)= \sum_{i=1}^{n} P(A / H_i)P(H_i)$$

Теперь из (*) можно получить формулу, называемую **формулой Байеса**:

$$P(H_k / A) = \frac{P(A / H_k)P(H_k)}{\sum\limits_{i=1}^{n} P(A / H_i)P(H_i)}$$

По формуле Байеса исчисляется вероятность реализации гипотезы H_k при условии, что событие A произошло. Формулу Байеса еще называют **формулой вероятности гипотез.**

Пример.Рассмотрим приведенную выше задачу об электролампах, только изменим вопрос задачи. Пусть покупатель купил электролампу в этом магазине, и она оказалась бракованной. Найти вероятность того, что эта лампа изготовлена на втором заводе.

Выпишем формулу Байеса для этого случая

$$P(H_2 / A) = \frac{P(A / H_2)P(H_2)}{P(A)}$$

Из этой формулы получаем: $P(H_2/ A) = 15/34$

Предлагаем читателю решить самостотельно две задачи.

.№1.В первой урне 7 белых и 3 черных шара, во второй - 8 белых и 2 черных. Из первой урны случайным образом извлекается шар и перекладывается во вторую урну. После перемешивания шаров во второй урне из нее извлекается один шар. Найти вероятность того, что извлеченный из второй урны шар — белый.

№2.В условие задачи №1 внесем изменение. Пусть после перекладывания шара из первой урны во вторую из второй урны извлечен белый шар. Найти вероятность того, что из первой урны во вторую был переложен черный шар.

Повторные независимые испытания. Формула Бернулли.

Рассмотрим случай многократного повторения одного и того же испытания или случайного эксперимента. Результат каждого испытания будем считать не зависящим от того, какой результат наступил в предыдущих испытаниях. В качестве результатов или элементарных исходов каждого отдельного испытания будем различать лишь две возможности:
1) появление некоторого события A;
2) появление события $\overline{A}$, (события, являющегося дополнением A)

Пусть вероятность $P(A)$ появления события A постоянна и равна p $(0<p<1)$. Вероятность $P(\overline{A})$ события $\overline{A}$ обозначим через q: $P(\overline{A}) = 1- p=q$.

Примерами таких испытаний могут быть:
1) подбрасывание монеты: A - выпадение герба; $\overline{A}$ - выпадение цифры.

$$P(A) = P(\overline{A}) = 0{,}5.$$

2) бросание игральной кости: А - выпадение количества очков, равного пяти, $\overline{A}$ выпадение любого количества очков кроме пяти.

$$P(A) =1/6,\ P(\overline{A}) =5/6.$$

3) извлечение наудачу из урны, содержащей 7 белых и 3 черных шара, одного шара (с возвращением): A - извлечение белого шара, $\overline{A}$ - извлечение черного шара

$$P(A) = 0{,}7;\ P(\overline{A}) = 0{,}3$$

Пусть произведено n испытаний, которые мы будем рассматривать как один сложный случайный эксперимент. Составим таблицу из n клеток, расположенных в ряд, пронумеруем клетки, и результат каждого испытания будем отмечать так: если в i-м испытании событие A произошло, то в i-ю клетку ставим цифру 1, если событие A не произошло (произошло событие $\overline{A}$), в i-ю клетку ставим 0.

Если, например, проведено 5 испытаний, и событие А произошло лишь во 2 -м и 5-м испытаниях, то результат можно записать такой последовательностью нулей и единиц: 0; 1; 0; 0; 1.

Каждому возможному результату n испытаний будет соответствовать последовательность n цифр 1 или 0, чередующихся в том порядке, в котором появляются события A и $\overline{A}$ в n испытаниях, например:

$$\underbrace{1;\ 1;\ 0;\ 1;\ 0;\ 1;\ 0;\ 0;\ \ldots\ 0;\ 1;\ 1;\ 0}_{n \text{ цифр}}$$

Всего таких последовательностей можно составить 2^n (это читатель может доказать сам).

Так как испытания независимы, то вероятность P каждого такого результата определяется путем перемножения вероятностей событий A и $\overline{A}$ в соответствующих испытаниях. Так, например, для написанного выше результата найдем

$$P = p \cdot p \cdot q \cdot p \cdot q \cdot p \cdot q \cdot q \cdot \ldots \cdot q \cdot p \cdot p \cdot q$$

Если в написанной нами последовательности единица встречается x раз (это значит, что нуль встречается n-x раз), то вероятность соответствующего результата будет $p^n q^{n-x}$ независимо от того, в каком порядке чередуются эти x единиц и n-x нулей.

Все события, заключающиеся в том, что в n испытаниях событие A произошло x раз, а событие $\bar{A}$ произошло n-x раз, являются несовместными. Поэтому для вычисления вероятности объединения этих событий (или суммы этих событий), нужно сложить вероятности всех этих событий, каждая из которых равна $p^n q^{n-x}$. Всего таких событий можно насчитать столько, сколько можно образовать различных последовательностей длины n, содержащих x цифр "1" и n-x цифр "0". Таких последовательностей получается столько, сколькими способами можно разместить x цифр "1" (или n-x цифр "0") на n местах, то есть число этих последовательностей равно $C_n^x = C_n^{n-x}$

Отсюда получается **формула Бернулли**:

$$P_n(x) = C_n^x p^x q^{n-x}$$

По формуле Бернулли рассчитывается вероятность появления события A "x"раз в n повторных независимых испытаниях, где p - вероятность появления события A в одном испытании, q - вероятность появления события $\bar{A}$ в одном испытании.

Сформулированные условия проведения испытаний иногда называются "**схемой повторных независимых испытаний**" или "**схемой Бернулли**"

Число x появления события A в n повторных независимых испытаниях называется **частотой**.

Пример. Из урны, содержащей 2 белых и 6 черных шаров наудачу выбирается с возвращением 5 раз подряд один шар. Подсчитать вероятность того, что 4 раза появится белый шар.

В приведенных выше обозначениях n=8; p=1/4; q=3/4; x=5. Искомую вероятность вычисляем по формуле Бернулли:

$$P_8(5) = C_5^4\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right) = \frac{15}{1024}$$

По формуле Бернулли можно подсчитать вероятности всех возможных частот: x=0,1,2,3,4,5.

Заметим, что если в этой задаче считать, что белых шаров было 20000, а черных 60000, то очевидно p и q останутся неизменными. Однако в этой ситуации можно пренебречь возвращением извлеченного шара после каждой выборки (при не слишком больших значениях x) и считать вероятности всех частот: x=0,1,2,... по формуле Бернулли.

Формула Бернулли при заданных числах p и n позволяет рассчитывать вероятность любой частоты x ($0 \le x \le n$). Возникает естественный вопрос: какой частоте будет соответствовать наибольшая вероятность?

Предположим, что такая частота существует, и попытаемся ее определить из условия, что вероятность этой частоты не меньше вероятности "предыдущей" и "последующей" частот:

$$P_n(x) \ge P_n(x-1);\ P_n(x) \ge P_n(x+1) \qquad (1)$$

Первое неравенство (*) представляется в виде:

$$C_n^x p^x q^{n-x} \ge C_n^{x-1} p^{x-1} q^{n-x+1},$$

что эквивалентно $\frac{p}{x} \le \frac{q}{n-x+1}$ или $qx \le pn - px + p$. Отсюда следует:

$$x \le np + p$$

Решая второе неравенство (1), получим

$$x \ge np - q$$

Таким образом, частота, имеющая наибольшую вероятность (чем вероятнейшая частота), определяется двойным неравенством

$$np - q \le x \le np + p$$

Если $np + p$ – целое число (тогда и $np - q$ – целое число), то две частоты: $x=np - q$ и $x=np + p$ обладают наибольшей вероятностью. Например, при $n = 7$; $p = \frac{1}{2}$, наивероятнейшие частоты: $x = 3$; $x = 4$.

Случайная величина, распределенная по закону Бернулли.

При двух заданных числах:
1) n - количестве повторных независимых испытаний,
2) p - вероятности события А в одном испытании
можно по формуле Бернулли подсчитать значение вероятности каждого целого числа x $(0 \le x \le n)$, где x – число появлений события А в n испытаниях (частота появления события A).

Таким образом, каждому исходу случайного эксперимента, заключающегося в серии из n испытаний по схеме Бернулли, соответствует определенное число x, рассматриваемое как случайная величина, принимающая значения 0, 1, 2,...n. Соответствие между значениями x и их вероятностями (рассчитанными по формуле Бернулли) называется законом распределения Бернулли. Строгое определение случайной величины и закона распределения будет дано позже.

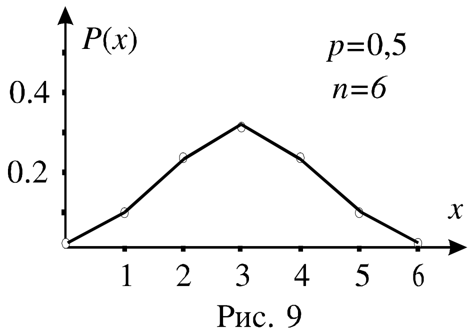

Рис. 9

Можно построить график закона распределения Бернулли (зависимости $P_n(x)$) для конкретных значений n и p. Так как аргумент x принимает лишь целые значения, график представляется в виде точек на плоскости $(x, P_n(x))$. Для наглядности точки соединяются ломаной линией, и такой график называется полигоном распределения.

При $p = 0{,}5$, как показано на рисунке 9, полигон симметричен относительно прямой $x=np$ (если p близко к 0,5, то полигон близок к симметричному)

При малых p полигон существенно асимметричен, и наивероятнейшими являются частоты, бизкие к нулю. На рисунке 10 изображен полигон распределения для p=0,2 при числе испытаний n,равном 6-ти.

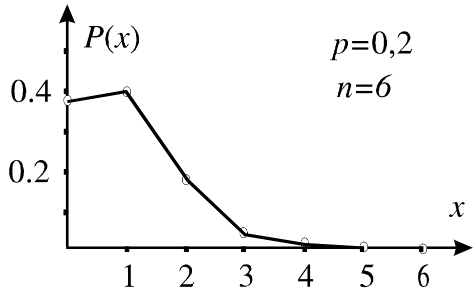

Рис.10

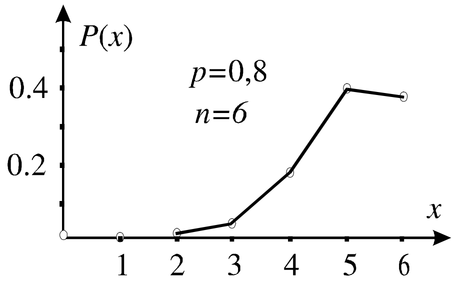

Рис. 11

При больших p, близких к 1, наиболее вероятны максимальные значения. На рисунке 11 показан полигон распределения, для p=0,8 и n=6.

О других свойствах бернуллиевского распределения будет говориться позже.