Формула полной вероятности

Следствием обеих теорем вероятности – теоремы сложения и теоремы умножения – является формула полной вероятности.

Пусть проводится опыт, об условиях которого можно сделать *n* исключающих друг друга предположений (гипотез), образующих полную группу:

$$H_1, H_2, \ldots, H_n, \quad H_i H_j = \varnothing, \; i \neq j \;, \sum_{i=1}^{n} H_i = \Omega$$

Каждая из гипотез осуществляется случайным образом и представляет собой случайное событие. Вероятности гипотез известны и равны:

$$p(H_1), p(H_2), \ldots, p(H_n), \sum_{i=1}^{n} p(H_i) = 1$$

Рассмотрим некоторое событие *A*, которое может появиться только вместе с одной из гипотез. Известны условные вероятности события *A* для каждой из гипотез:

P(A/H1), p(A/H2), … p(A/Hn).

На основании второй аксиомы $p(A) = \sum_{i=1}^{n} p(H_i A)$

С учетом теоремы умножения вероятностей *p(HiA) = p(Hi)p(A/Hi)*, тогда

$$p(A) = \sum_{i=1}^{n} p(H_i) \cdot p(A / H_i) \qquad (3.1)$$

Формула Байеса

Базируется на формуле полной вероятности и теореме умножения вероятностей.

Пусть до проведения некоторого опыта об его условиях *n* можно сделать *n* исключающих друг друга предположений (гипотез), образующих полную группу:

$$H_1, H_2, \ldots, H_n, \quad H_i H_j = \varnothing, \ i \neq j \ , \sum_{i=1}^{n} H_i = \Omega$$

Вероятности гипотез до опыта (априорные вероятности) известны:

p(H1), p(H2), … p(Hn).

Опыт произведен, и произошло некоторое событие A. Требуется определить вероятности гипотез с учетом того, что произошло событие А, т. е. определить апостериорные вероятности: $p(H_1/A), p(H_2/A), \ldots p(H_n/A)$.

Вероятность того, что событие A произошло совместно с Hi, на основании теоремы умножения, вероятностей равна $p(HiA) = p(Hi)p(A/Hi) = p(A)p(Hi/A)$.

Отбросим левую часть равенства и выразим $p(Hi/A)$:

$$p(H_i / A) = \frac{p(H_i)p(A / H_i)}{p(A)}.$$

Раскроем $p(A)$ по формуле полной вероятности (3.1) и получим формулу Байеса

$$p(H_i / A) = \frac{p(H_i)p(A / H_i)}{\sum_{j=1}^{n} p(H_j)p(A / H_j)}. \qquad (3.2)$$

Формула Байеса позволяет пересчитать априорные вероятность гипотез с учетом того, что опыт завершился событием A.

Теорема о повторении опытов

Пусть проводятся n независимых ***одинаковых*** опытов, в каждом из которых событие A появляется с вероятностью p. Вероятность $P(n,k)$ того, что событие A произойдет ровно в k опытах, равна ***(формула Бернулли)***

$$P(n,k) = C_n^k \cdot p^k \cdot q^{n-k} = \frac{n!}{k!\cdot(n-k)!} p^k \cdot q^{n-k}, 0 \le k \le n \qquad (3.3)$$

где $q = 1 - p$ - вероятность того, что A не появится в одном опыте.

Доказательство. Обозначим через Bk появление события A в k опытах и появление $\bar{A}$ в n - k опытах. Событие Bk представляет собой сумму несовместимых событий:

$$B_k = \underbrace{A_1 \cdot A_2 \cdot \ldots \cdot A_k}_{k} \cdot \underbrace{\bar{A}_{k+1} \cdot \bar{A}_{k+2} \cdot \ldots \cdot \bar{A}_n}_{n-k} + \ldots + \underbrace{\bar{A}_1 \cdot \bar{A}_2 \cdot \ldots \cdot \bar{A}_{n-k}}_{n-k} \cdot \underbrace{A_{n-k+1} \cdot A_{k+2} \cdot \ldots \cdot A_n}_{k}$$

где A_i, $\bar{A}_i$ – появление и не появление события A в i-м опыте.

Определим вероятность одного из слагаемых. Так как все опыты одинаковы, то вероятности всех вариантов одинаковы и равны

$$P(A_1 \cdot A_2 \cdot \ldots\ldots A_k \bar{A}_{k+1} \cdot \bar{A}_{k+2} \ldots\ldots \bar{A}_n) = \underbrace{p \cdot p \cdot \ldots \cdot p}_{k} \underbrace{q \cdot q \cdot \ldots \cdot q}_{n-k} = p^k \cdot q^{n-k}$$

Количество вариантов таких сложных событий равно числу выборок *к* номеров опытов из *n* возможных, в которых произойдут события *A*, т.е. равно числу сочетаний без повторения элементов C_n^r

Так как эти события несовместимы, то на основании второй аксиомы :

$$p(B_k) = P(n,k) = C_n^k \cdot p^k \cdot q^{n-k}$$

Свойства формулы Бернулли:

1. Правая часть формулы (3.3) представляет собой общий член разложения бинома Ньютона:

$$(q+p)^n = \sum_{k=0}^{n} P(n,k) = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = 1 \qquad (3.4)$$

2. Рекуррентная формула *P(n,k)*имеет вид

$$P(n,k+1) = \frac{n-k}{k+1} \cdot \frac{p}{q} P_n(n,k) \qquad (3.5)$$

3. Число *к0*, которому соответствует максимальная вероятность $P(n,k_0)$ называется *наивероятнейшим числом* появления события *A* и определяется неравенствами:

$$np - q \le k_0 \le np + p \qquad (3.6)$$

Доказательство.

$$\frac{P_n(k+1)}{P_n(k)} = \frac{n-k}{k+1}\cdot\frac{p}{q} \Rightarrow P_n(k+1) \ge P_n(k) \Leftrightarrow \frac{n-k}{k+1}\cdot\frac{p}{q} \ge 1 \Leftrightarrow k \le np - q$$, а

$$P_n(k+1) \le P_n(k) \Leftrightarrow \frac{n-k}{k+1}\cdot\frac{p}{q} \le 1 \Leftrightarrow k \ge np - q$$

Итак, при $k < np - q$ функция $P(n,k)$ возрастает, а при $k > np - q$ убывает.

Тогда существует точка *k0*, в которой $P(n,k)$ достигает максимума, т.е.

$$\begin{cases} P(n,k_0) \ge P(n,k_0 - 1) \\ P(n,k_0) \le P(n,k_0 + 1) \end{cases}$$

Решив данную систему неравенств относительно *k0* , получим (3.6).

4. Вероятность $P_n(k_1 \le k \le k_2)$

того, что в *n* опытах схемы Бернулли, событие *A* появится от k_1 до k_2

раз $(0 \le k_1 \le k_2 \le n)$, равна

$$P(n, k_1 \le k \le k_2) = \sum_{k=k_1}^{k_2} P(n,k) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k} \qquad (3.7)$$

5. Вероятность $P(n, 1 \le k \le n)$

того, что в *n* опытах событие *A* появится хотя бы один раз, равна

$$P(n, 1 \le k \le n) = 1 - P(n,0) = 1 - q^n \qquad (3.8)$$

Пусть производится *n* независимых опытов, каждый из которых имеет, *r (r>2)* попарно несовместных и единственно возможных исходов $A_1, A_2, \ldots\ldots A_r$

с вероятностями $p_1 = p(A_1), p_2 = p(A_2), \ldots\ldots p_r = p(A_r)$

Требуется определить вероятность того, что из серии *n* независимых опытов исход A_1 наступит k_1 раз, $A_2 - k_2, ..., A_r - k_r \left(k_1 + k_2 + ... + k_r = n\right)$,то

$$P(n,k_1,k_2,.....k_r) = \frac{\left(k_1 + k_2 + ...k_r\right)!}{k_1!k_2!.....k_r!} \cdot p_1^{k_1} \cdot p_2^{k_2} \dots p_r^{k_r} \tag{3.9}$$

Вычисление вероятностей $P(n,k)$ при больших значениях n по формуле Бернулли проблематично. Поэтому вычисление соответствующих вероятностей проводится с помощью следующих приближенных формул.

Если количество испытаний велико $n \to \infty$,а вероятность события мала $p \to 0$,так что $np \to a,\ 0 < a < \infty$ и $p << \frac{1}{\sqrt{n}}$

то используется ***формула Пуассона***:

$$P(n,k) \approx \frac{a^k}{k!} \cdot e^{-a},\ k = \overline{0,n} \tag{3.10}$$

Если количество испытаний *n* велико, вероятности *p* и *q* не малы, так что выполняются следующие условия:

$$0 < np - 3\sqrt{npq},\ np + 3\sqrt{npq} < n$$

то применяются приближенные **формулы Муавра-Лапласа**:

- *локальная*

$$P(n,k) \approx \frac{\varphi(x)}{\sqrt{npq}}, \qquad (3.11)$$

где

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \qquad x = \frac{k - np}{\sqrt{npq}}$$

- *интегральная*

$$P(n, k_1 \le k \le k_2) \approx \Phi(x_2) - \Phi(x_1) \qquad (3.12)$$

где

$$x_1 = \frac{(k_1 - np)}{\sqrt{npq}} \qquad x_2 = \frac{(k_2 - np)}{\sqrt{npq}}$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x \exp\left(-\frac{x^2}{2}\right)dx$$ - функция Лапласа.

Функции $\varphi(x)$ и $\Phi(x)$ табулированы.

При использовании таблиц следует помнить, что $\varphi(x)$ является четной $\varphi(-x)=\varphi(x)$ а функция Лапласа - нечетной $\Phi(-x)=-\Phi(x)$.