

ФОРМУЛЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Формула Байеса

Вероятность события A , которое может наступить лишь при условии появления одного из несовместных событий H_1, H_2, \dots, H_n , образующих полную группу, находится по формуле:

$$p(A) = p(H_1) \cdot p_{H_1}(A) + p(H_2) \cdot p_{H_2}(A) + \dots + p(H_n) \cdot p_{H_n}(A).$$

Эта формула называется *формулой полной вероятности*.

События H_1, H_2, \dots, H_n называются *гипотезами*. Поскольку эти события несовместны и образуют полную группу событий, то

$$p(H_1) + p(H_2) + \dots + p(H_n) = 1.$$

Примеры

- В трех одинаковых урнах находятся шары: в первой урне – 4 белых и 5 черных шаров, во второй урне – 2 белых и 7 черных шаров, в третьей – 5 белых и 4 черных шара. Наудачу выбирается урна и из нее извлекается шар. Найти вероятность того, что извлеченный шар окажется черным.
- Рассмотрим событие A – «извлечен черный шар».
- Возможны следующие предположения (гипотезы): H_1 – «шар извлечен из первой урны»; H_2 – «шар извлечен из второй урны»; H_3 – «шар извлечен из третьей урны».

Продолжение примера

Поскольку имеется три гипотезы, причем по условию они равновероятны (так как урны одинаковы), и сумма вероятностей гипотез должна быть равна 1, то $p(H_1) = p(H_2) = p(H_3) = \frac{1}{3}$.

Условная вероятность того, что будет извлечен черный шар при условии, что шар вынут из первой урны, равна $p_{H_1}(A) = \frac{5}{9}$.

Очевидно, что $p_{H_2}(A) = \frac{7}{9}$, $p_{H_3}(A) = \frac{4}{9}$.

По формуле полной вероятности

$$p(A) = p(H_1) \cdot p_{H_1}(A) + p(H_2) \cdot p_{H_2}(A) + p(H_3) \cdot p_{H_3}(A) = \frac{1}{3} \cdot \frac{5}{9} + \frac{1}{3} \cdot \frac{7}{9} + \frac{1}{3} \cdot \frac{4}{9} = \frac{16}{27}.$$

Еще формула

Пусть произведено испытание, в результате которого появилось событие A . Тогда вероятности гипотез могут быть переоценены по формулам Байеса:

$$p_A(H_i) = \frac{p(H_i) \cdot p_{H_i}(A)}{p(A)}, \quad i = 1, 2, \dots, n.$$

Новый пример

- Три преподавателя – члена экзаменационной комиссии по математике проверяли письменные работы учащихся. Первый преподаватель проверил 35, второй – 20, третий – 45 работ. Вероятность того, что проверяющий заметит все имеющиеся в работе ошибки, для первого преподавателя равна 0,9, для второго – 0,85, для третьего – 0,8. Председатель комиссии перепроверил случайным образом выбранную работу и обнаружил в ней ошибки, не замеченные проверяющим. Найти вероятность того, что работа проверена первым преподавателем.

Продолжение примера

Рассмотрим событие A и гипотезы:

A – «в работе обнаружены ошибки, не замеченные проверяющим»;

H_1 – «работа проверена первым преподавателем»;

H_2 – «работа проверена вторым преподавателем»;

H_3 – «работа проверена третьим преподавателем».

Заметим, что необходимо найти вероятность $p_A(H_1)$.

Так как из 100 работ первый преподаватель проверил 25, то

$$p(H_1) = \frac{25}{100} = 0,25. \text{ Аналогично, } p(H_2) = \frac{20}{100} = 0,2, \quad p(H_3) = \frac{45}{100} = 0,45.$$

Найдем условные вероятности $p_{H_1}(A)$, $p_{H_2}(A)$, $p_{H_3}(A)$.

Вероятность того, что в работе будут обнаружены ошибки, не замеченные проверяющим, при условии, что работа проверена первым преподавателем, $p_{H_1}(A) = 1 - 0,9 = 0,1$.

Ясно, что $p_{H_2}(A) = 1 - 0,85 = 0,15$, $p_{H_3}(A) = 1 - 0,8 = 0,2$.

По формуле полной вероятности

$$\begin{aligned} p(A) &= p(H_1) \cdot p_{H_1}(A) + p(H_2) \cdot p_{H_2}(A) + p(H_3) \cdot p_{H_3}(A) = \\ &= 0,25 \cdot 0,1 + 0,2 \cdot 0,15 + 0,45 \cdot 0,2 = 0,155. \end{aligned}$$

Для нахождения искомой вероятности используем формулу

$$\text{Байеса: } p_A(H_1) = \frac{p(H_1) \cdot p_{H_1}(A)}{p(A)} = \frac{0,25 \cdot 0,1}{0,155} = \frac{7}{31}.$$

Повторение испытаний

Если производятся испытания, в каждом из которых вероятность появления события A не зависит от исходов других испытаний, то такие испытания называются *независимыми относительно события A* .

Пусть производятся n независимых испытаний и пусть вероятность появления события A во всех испытаниях одинакова и равна p (следовательно, вероятность появления противоположного события \bar{A} равна $q = 1 - p$). Тогда вероятность того, что событие A появится ровно k раз, находится по формуле:

$$p_n(k) = C_n^k \cdot p^k \cdot q^{n-k}.$$

Эта формула называется *формулой Бернулли*.

Число k называется *числом успехов*.

Вероятности $p_n(k)$ называются *биномиальными вероятностями* и удовлетворяют условию: $p_n(0) + p_n(1) + p_n(2) + \dots + p_n(n) = 1$.

Пример

Найти вероятность того, что при семи бросаниях монеты герб выпадет три раза.

Так как производятся независимые испытания (бросания монеты) и вероятность выпадения герба во всех испытаниях одинакова (она равна $\frac{1}{2}$), то применима формула Бернулли.

В данном случае $n = 7$, $k = 3$, $p = \frac{1}{2}$, $q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$.

Значит, искомая вероятность

$$p_7(3) = C_7^3 \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^{7-3} = \frac{7!}{3! \cdot 4!} \cdot \frac{1}{8} \cdot \frac{1}{16} = \frac{35}{128}.$$

Еще пример

Число k_0 , которому при заданном значении n соответствует максимальная биномиальная вероятность $p_n(k_0)$, называется *наиболее вероятным числом успехов*. Число k_0 удовлетворяет двойному неравенству:

$$(n+1) \cdot p - 1 \leq k_0 \leq n \cdot p + p.$$

Найти наиболее вероятное число выпадений герба при 37 бросаниях монеты.

По условию, $n=37$, $p=\frac{1}{2}$. Тогда $(37+1) \cdot \frac{1}{2} - 1 \leq k_0 \leq 37 \cdot \frac{1}{2} + \frac{1}{2}$.

Следовательно, $18 \leq k_0 \leq 19$. Полученному неравенству удовлетворяют два целых числа: $k'_0=18$, $k''_0=19$. Это означает, что наиболее вероятное число выпадений герба равно 18 или 19.

Формулы Лапласа

Очевидно, что при больших значениях n и k вычисление вероятностей $p_n(k)$ по формуле Бернулли превращается в технически сложную задачу.

Если n достаточно велико, то вероятность того, что событие A появится ровно k раз в n независимых испытаниях, находится по формуле:

$$p_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x), \text{ где } \varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \quad x = \frac{k - np}{\sqrt{npq}}.$$

Эта формула называется *локальной формулой Лапласа*.



Формулы Лапласа

Заметим, что использование локальной формулы Лапласа в случае, когда вероятность p близка к 0 или 1, обычно приводит к значительным погрешностям. Для того чтобы в этом случае формула дала результат с незначительной ошибкой, необходимо, чтобы число испытаний n было очень велико. На практике локальной формулой Лапласа пользуются при выполнении условия: $npq \geq 10$.

Формулы Лапласа

Вероятность того, что в n независимых испытаниях событие A появится не менее k_1 раз и не более k_2 раз, находится по формуле (при условии, что n достаточно велико):

$$p_n(k_1; k_2) = \Phi(x_2) - \Phi(x_1), \text{ где}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt - \text{функция Лапласа, } x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

Эта формула называется *интегральной формулой Лапласа*.

Пример

Для данного сорта растения всхожими являются 75% семян. Найти вероятность того, что среди 1000 высеянных семян взойдет: а) 770 семян; б) от 720 до 770 семян.

По условию задачи $n=1000$; $p=0,75$; $q=0,25$. Так как $npq = 1000 \cdot 0,75 \cdot 0,25 = 187,5 > 10$, то применимы формулы Лапласа.

а) Итак, $n=1000$; $k=770$; $p=0,75$; $q=0,25$.

Применим локальную формулу Лапласа.

Найдем значение x :

$$x = \frac{k - np}{\sqrt{npq}} = \frac{770 - 1000 \cdot 0,75}{\sqrt{1000 \cdot 0,75 \cdot 0,25}} \approx \frac{20}{13,69} \approx 1,46.$$

По таблице значений функции $\varphi(x)$ найдем $\varphi(1,46) = 0,1374$.

Тогда вероятность

$$P_{1000}(770) = \frac{1}{\sqrt{npq}} \cdot \varphi(x) = \frac{1}{\sqrt{1000 \cdot 0,75 \cdot 0,25}} \cdot 0,1374 \approx 0,01.$$

Продолжение примера

б) По условию $n=1000$; $k_1=720$; $k_2=770$; $p=0,75$, $q=0,25$.

Применим интегральную формулу Лапласа.

Найдем значения x_1 и x_2 :

$$x_1 = \frac{720 - 1000 \cdot 0,75}{\sqrt{1000 \cdot 0,75 \cdot 0,25}} \approx -2,19; \quad x_2 = \frac{770 - 1000 \cdot 0,75}{\sqrt{1000 \cdot 0,75 \cdot 0,25}} \approx 1,46.$$

По таблице значений функции $\Phi(x)$ найдем $\Phi(x_2)$ и $\Phi(x_1)$:

$$\Phi(x_2) = \Phi(1,46) = 0,4279.$$

Так как функция $\Phi(x)$ является нечетной, то

$$\Phi(x_1) = \Phi(-2,19) = -\Phi(2,19) = -0,4857.$$

Тогда вероятность того, что взойдет от 720 до 770 семян, будет равна $P_{1000}(720;770) = \Phi(x_2) - \Phi(x_1) = 0,4279 - (-0,4857) = 0,9136$.

Формула Пуассона и пример

Если n достаточно велико, а вероятность p появления события A в каждом испытании мала, то вероятность того, что событие A появится ровно k раз в n независимых испытаниях, находится по формуле:

$$p_n(k) \approx \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \text{ где } \lambda = np - \text{ постоянная величина.}$$

Эта формула называется *формулой Пуассона*.

Обычно формулу Пуассона используют, когда $n \geq 100$, а $p \leq 0,1$.

Пример В организации 300 человек, использующих в своей работе официальный сайт организации. Сотрудники организации работают независимо друг от друга. Для каждого сотрудника вероятность того, что он в течение часа обратится к сайту, равна 0,01. Найти вероятность того, в течение часа к официальному сайту организации обратятся: а) 2 сотрудника; б) не менее 3 сотрудников.

Продолжение примера

а) По условию, $n = 300$, $p = 0,01$, $k = 2$. Сотрудники обращаются к сайту независимо друг от друга, число n достаточно велико, а вероятность p мала, поэтому можно применить формулу Пуассона.

В нашем случае $\lambda = np = 300 \cdot 0,1 = 3$.

Искомая вероятность $p_{300}(2) = \frac{3^2}{2!} \cdot e^{-3} \approx \frac{9}{2} \cdot 0,0498 = 0,2241$.

б) События «в течение часа к сайту обратятся не менее 3 сотрудников» и «в течение часа к сайту обратятся менее 3 сотрудников» - противоположные, следовательно

$$\begin{aligned} p_{300}(3 \leq k \leq 300) &= 1 - p_{300}(0 \leq k < 3) = 1 - (p_{300}(0) + p_{300}(1) + p_{300}(2)) = \\ &= 1 - \left(\frac{3^0}{0!} \cdot e^{-3} + \frac{3^1}{1!} \cdot e^{-3} + \frac{3^2}{2!} \cdot e^{-3} \right) \approx 1 - (0,0498 + 0,1494 + 0,2241) = 0,5767. \end{aligned}$$