Лекция 40

Формула полной вероятности. Формула Байеса

Пусть (Ω, F, P) - вероятностное пространство.

События $H_1, H_2, ..., H_n$ образуют ***полную группу несовместных событий*** в вероятностном пространстве, если:

1) они являются попарно независимыми: $H_i \cap H_j = \varnothing, \ i \neq j$;

2) сумма событий образует пространство элементарных исходов: $\bigcup_{i=1}^{n} H_i = \Omega$.

Пусть некоторое событие A может произойти при различных условиях, относительно которых может быть сделано n предположений (гипотез): $H_1, H_2, ..., H_n$. Тогда имеет место следующая теорема.

Теорема 40.1 ***(формула полной вероятности).*** Если события $H_1, H_2, ..., H_n$, $P(H_k) > 0$ образуют полную группу несовместных событий в вероятностном пространстве (Ω, F, P), то $\forall A \in F$ справедлива формула полной вероятности

$$P(A) = \sum_{i=1}^{n} P(H_i) P(A / H_i) \qquad (40.1)$$

то есть вероятность события A равна сумме произведений вероятностей каждой из гипотез на соответствующую условную вероятность события A.

□ Действительно, $A = AH_1 + AH_2 + \ldots + AH_n$. Так как $(AH_i) \cdot (AH_j) = \varnothing$ при $i \neq j$, то в силу (39.3) получим:

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

Итак,

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

Пример 40.1. В двух контейнерах имеются резисторы. В первом контейнере – 20 резисторов, из которых четыре являются нестандартными; во втором – 10 резисторов, из них один является нестандартным. Из первого контейнера наудачу взяли резистор и переложили во второй. Найти вероятность того, что наудачу извлеченный из второго контейнера резистор будет нестандартным.

◁ Введем следующие события:

$A = \{$из второго контейнера взят нестандартный резистор$\}$,

$H_1 = \{$во второй контейнер переложен стандартный резистор$\}$,

$H_2 = \{$во второй контейнер переложен нестандартный резистор$\}$.

Ясно, что событие A и гипотезы H_i удовлетворяют условиям, в которых применима формула (40.1), поэтому

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) = \frac{4}{20}\cdot\frac{2}{11} + \frac{16}{20}\cdot\frac{1}{11} = \frac{6}{55}. \ \triangleright$$

Замечание. Зачастую на практике после осуществления события A требуется узнать, какая именно из гипотез вероятнее всего реализовалась при осуществлении события. Ответ на этот вопрос можно получить при помощи следующей формулы, полученной Байесом.

Теорема 40.2 *(формула Байеса).* Если события $H_1, H_2, ..., H_n$, $P(H_k) > 0$ образуют полную группу несовместных событий, то $\forall A \in F$ справедлива формула Байеса

$$P(H_k/A) = \frac{P(H_k)\cdot P(A/H_k)}{\sum_{i=1}^{n} P(H_i)\cdot P(A/H_i)}. \qquad (40.2)$$

□ Вероятность совместного наступления события A с любым из H_k, согласно теореме умножения вероятностей равна

$$P(A\cdot H_k) = P(A)P(H_k/A) = P(H_k)P(A/H_k),\ k=1,...n.$$

Выражая $P(H_k/A)$, имеем $P(H_k/A) = \dfrac{P(H_k)\cdot P(A/H_k)}{P(A)}$. Преобразуя знаменатель по формуле полной вероятности, окончательно получаем формулу Байеса в виде

$$P(H_k/A) = \frac{P(H_k)\cdot P(A/H_k)}{\sum_{i=1}^{n} P(H_i)\cdot P(A/H_i)}.\ (k=1, 2, \ldots, n)\ \blacksquare$$

Пример 40.2. Пусть в условиях предыдущего примера из второго контейнера извлечен нестандартный резистор. Требуется определить, какой резистор вероятнее всего был переложен из первого контейнера во второй.

◁ Воспользуемся формулой Байеса (40.2)

$$P(H_1/A)=\frac{P(H_1)P(A/H_1)}{\sum_{i=1}^{2}P(H_i)\cdot P(A/H_i)}=\frac{\frac{16}{20}\cdot\frac{1}{11}}{\frac{6}{55}}=\frac{2}{3};$$

$$P(H_2/A)=\frac{P(H_2)P(A/H_2)}{\sum_{i=1}^{2}P(H_i)\cdot P(A/H_i)}=\frac{\frac{4}{20}\cdot\frac{2}{11}}{\frac{6}{55}}=\frac{1}{3}.$$

Приведенные расчеты показывают, что вероятнее всего был переложен стандартный резистор. ▷

Схема повторных испытаний. Формула Бернулли

Пусть событие A связано с некоторым воспроизводимым опытом. Предположим, что проводится серия из n независимых опытов. Схема исследования, описываемая следующим образом: во время каждого испытания проводится один и тот же опыт, в результате которого возможны два исхода: наступит событие A, $P(A)=p=const$ или наступит событие $\overline{A}$, $P(\overline{A})=1-p=q$, называется ***схемой Бернулли***. Для многих практических задач требуется знать, какова вероятность того, что событие A произойдет в m из этих опытов. Обозначим вероятность этого события через $P_n(m)$.

Замечание. Иногда событие, состоящее в осуществлении события A называют успехом. В связи с этим вероятность $P_n(m)$ называют вероятностью m успехов в серии из n независимых испытаний (опытов).

Одному испытанию в схеме Бернулли соответствует пространство элементарных событий $\Omega=\{A,\overline{A}\}$. Последовательности же из n испытаний будет соответствовать набор из 2^n элементарных исходов. При этом:

$\Omega=\{\,\overline{A},\overline{A},\overline{A},\ldots,\overline{A}\,;A,\overline{A},\overline{A},\ldots,\overline{A}\,;\ldots;A,A,A,\ldots,\overline{A}\,;A,A,A,\ldots,A\}$. Таким образом, если в одной серии из n испытаний событие A наблюдалось m раз, то по теореме умножения вероятностей независимых событий имеем

$$P(A\cdot\overline{A}\cdot\overline{A}...A)=pqq...p=p^m q^{n-m}.$$

Обозначим $P_n(m)$ - вероятность того, что во всех возможных сериях из n испытаний m раз наступит событие A. Для нахождения их числа воспользуемся формулами комбинаторики. Рассмотрим множество, содержащее все возможные последовательности событий A (m штук) и $\overline{A}$ (n-m штук). Оно насчитывает C_n^m элементов, т.е. равно числу сочетаний из n по m. Согласно теореме сложения вероятностей (для случая несовместных событий), окончательно получаем

$$P_n(m)=C_n^m p^m q^{n-m}. \qquad (40.3)$$

Формула (40.3) носит название ***формулы Бернулли***.

Замечание. События ”в серии из n испытаний событие A наступило *0, 1, 2,…,n* раз” образуют полную группу несовместных событий. Сумма вероятностей этих событий равна единице (условие нормировки):

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

Пример 40.3. Стрелок поражает мишень одним выстрелом с вероятностью $p=0{,}8$. Определить, с какими вероятностями он поразит мишень *0, 1, 2, 3, 4, 5* раз в серии из *5* выстрелов?

Решение

Согласно (40.3) имеем:

$$P_5(0)=C_5^0 p^0 q^{5-0}=\frac{5!}{0!5!}0{,}8^0 0{,}2^5=\frac{1\cdot2\cdot3\cdot4\cdot5}{1\cdot1\cdot2\cdot3\cdot4\cdot5}0{,}8^0\cdot0{,}2^5=1\cdot0{,}00032=0{,}00032;$$

$$P_5(1)=C_5^1 p^1 q^{5-1}=\frac{5!}{1!4!}0{,}8^1 0{,}2^4=\frac{1\cdot2\cdot3\cdot4\cdot5}{1\cdot1\cdot2\cdot3\cdot4}0{,}8^1\cdot0{,}2^4=5\cdot0{,}8\cdot0{,}2^4=0{,}0064;$$

$$P_5(2)=C_5^2 p^2 q^{5-2}=\frac{5!}{2!3!}0{,}8^2 0{,}2^3=\frac{1\cdot2\cdot3\cdot4\cdot5}{1\cdot2\cdot1\cdot2\cdot3}0{,}8^2\cdot0{,}2^3=2\cdot5\cdot0{,}64\cdot0{,}008=0{,}0512;$$

$$P_5(3)=C_5^3 p^3 q^{5-3}=\frac{5!}{3!2!}0{,}8^3 0{,}2^2=\frac{1\cdot2\cdot3\cdot4\cdot5}{1\cdot2\cdot3\cdot1\cdot2}0{,}8^3\cdot0{,}2^2=2\cdot5\cdot0{,}512\cdot0{,}04=0{,}2048;$$

$$P_5(4)=C_5^4 p^4 q^{5-4}=\frac{5!}{1!4!}0{,}8^4 0{,}2^1=\frac{1\cdot2\cdot3\cdot4\cdot5}{1\cdot1\cdot2\cdot3\cdot4}0{,}8^4\cdot0{,}2=5\cdot0{,}4096\cdot0{,}2=0{,}4096;$$

$$P_5(5)=C_5^5 p^5 q^0=\frac{5!}{5!0!}0{,}8^5 0{,}2^0=\frac{1\cdot2\cdot3\cdot4\cdot5}{1\cdot2\cdot3\cdot4\cdot5\cdot1}0{,}8^5\cdot0{,}2^0=0{,}32768.$$

Условие нормировки выполняется

$$\sum_{k=0}^{5}P_5(k)=0{,}00032+0{,}0064+0{,}0512+0{,}2048+0{,}4096+0{,}32768=1.$$

Среди полученных вероятностей $P_5(4)$ является самой большой. Ей можно приписать следующий математический смыл: в серии из пяти выстрелов данный стрелок, скорее всего, поразит мишень четыре раза. Очевидно, что и для других n, найдется такое число $m=m^*$, для которого вероятность $P_n(m)=p^*$ будет больше других. Величины m^* и p^* играют важную роль в исследовании явлений, проходящих по схеме Бернулли.

Число m^* такое, что

$$P_n(m^*)=\max_{0\le m\le n}P_n(m), \qquad (40.4)$$

называется ***наивероятнейшим числом*** появления события A в серии из n испытаний по схеме Бернулли.

Теорема 40.3. Для наивероятнейшего числа m^* справедлива оценка

$$np-q\le m^*\le np+q. \qquad (40.5)$$

Пример 40.4. Проверим результат примера 40.3 с помощью (40.5).

Решение: Неравенство имеет вид: $5\cdot0{,}8-0{,}2\le m^*\le5\cdot0{,}8+0{,}2.$ Проводя вычисления, получим $3{,}8\le m^*\le4{,}2.$ Целое число, заключённое между 3,8 и 4,2 равно четырём, что совпадает с результатом примера 40.3.

Пример 40.5. Проверим другой известный пример. Определим наивероятнейшее число появлений герба при пятидесяти подбрасываниях монеты.

Решение: Неравенство имеет вид: $50\cdot0{,}5-0{,}5\le m^*\le50\cdot0{,}5+0{,}5.$ Получаем $24{,}5\le m^*\le25{,}5.$ Следовательно, $m^*=25$, что совпадает со всеми теоретическими аспектами.

Замечание. Из (40.5) следует, что длина отрезка $[np-q;np+q]$ равна $2q<2$. Поэтому, так как m^* - целое число, возможны два случая: m^* определяется однозначно; либо наивероятнейших чисел два.

Предельные теоремы для схемы Бернулли

Формулой Бернулли удобно пользоваться при вычислениях вероятностей, связанных со сравнительно короткими сериями. При длинных сериях ее использование связано с вычислительными трудностями. Например, если в примере 40.3 $n=1000,\ m=500$, то при подсчете каждого из сомножителей в выражении $P_{1000}(500)=C_{1000}^{500}0{,}8^{500}0{,}2^{500}$ возникают трудности вычислительного характера. Эти соображения приводят к необходимости получить простые с вычислительной точки зрения приближенные формулы для подсчета вероятностей $P_n(m)$, достаточно точные при длинных сериях испытаний.

а) Формула Пуассона

Формула Пуассона называется еще формулой для числа успехов редких событий в длинных сериях. Это название связано с тем, что приводимая ниже формула дает очень точные результаты в тех случаях, когда вероятность p достаточно мала (настолько мала, что величина np мало меняется с ростом n).

Теорема 40.4 ***(Пуассона, закон редких событий)***. Пусть в схеме Бернулли число испытаний n стремится к бесконечности $(n\to\infty)$, вероятность появления события A в одном испытании стремится к нулю $(p\to 0)$, при этом произведение np сохраняет примерно постоянное значение, а именно $np=\lambda$. Тогда справедлива приближённая формула Пуассона:

$$P_n(m)\approx\frac{\lambda^m e^{-\lambda}}{m!}. \qquad (40.6)$$

Формулой (40.6) можно пользоваться для приближённого вычисления вероятностей $P_n(m)$ в случае больших n и малых p. Вычислительные эксперименты показали, что это обычно возможно при $np<10$.

Пример 40.6. Телефонная станция обслуживает 1000 абонентов. В заданный интервал времени любой абонент независимо от остальных может сделать вызов с вероятностью 0,005. Какова вероятность того, что в данный интервал времени будет 7 звонков?

$\lhd$ По условию задачи $n = 1000$, $p = 0{,}005$, $\lambda = 5$. По формуле Пуассона: $P_{1000}(7) \approx \frac{5^7}{7!} \cdot e^{-5} \approx 0{,}108$. $\rhd$

б) Локальная теорема Муавра-Лапласа

Если условие *np<10* в формуле Пуассона не выполняется, то применение формулы Пуассона становится невозможным. Возникает необходимость ещё в одной приближённой формуле. Для её получения примем без доказательства следующую теорему.

Теорема 40.5 ***(Муавра-Лапласа, локальная)***. Пусть вероятность p появления некоторого события A постоянна в n независимых испытаниях и отлична от 0 и 1. Тогда вероятность $P_n(m)$ того, что в этих испытаниях событие A наступит m раз при $n \to \infty$, удовлетворяет соотношению

$$\lim_{n \to \infty} \frac{\sqrt{npq}\; P_n(m)}{\varphi(x)} = 1,$$

где $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, а $x = \frac{m - np}{\sqrt{npq}}$.

Замечание. При сделанных предположениях относительно p, если n достаточно большое, имеет место приближённое равенство

$$P_n(m) \approx \frac{\varphi(x)}{\sqrt{npq}}. \qquad (40.7)$$

Эту формулу называют локальной формулой Муавра – Лапласа.

Пример 40.7. Вероятность изготовления детали высшего сорта на данном станке равна 0.4. Найти вероятность того, что среди наудачу взятых 26 деталей половина окажется высшего сорта.

$\lhd$ Для нахождения вероятности $P_{26}(13)$ по формуле (40.7) необходимые вычисления производим по следующей схеме:

$n = 26$; $p = 0{,}4$; $q = 1 - 0{,}4 = 0{,}6$;

$np = 26 \cdot 0{,}4 = 10{,}4$; $npq = 10{,}4 \cdot 0{,}6 = 6{,}24$; $\sqrt{npq} = \sqrt{6{,}24} \approx 2{,}50$;

$m = 13$; $m - np = 13 - 10{,}4 = 2{,}6$;

$$x = \frac{m - np}{\sqrt{npq}} = \frac{2{,}60}{2{,}50} = 1{,}04;\ \varphi(x) = \varphi(1{,}04) = 0{,}2323;$$

$$P_n(m) \approx \frac{\varphi(x)}{\sqrt{npq}} = \frac{0{,}2323}{2{,}50} = 0{,}093.$$

Значение функции $\varphi(1{,}04)$ мы нашли по таблице (приложение 1). ⊳

Замечание. Локальная формула Муавра–Лапласа (40.7) может применятся практически всегда. Однако в тех случаях, когда либо p либо q малы, погрешность этой формулы больше, чем формулы Пуассона, используемой в нужной интерпретации.

в) Интегральная теорема Муавра-Лапласа.

Во многих практических задачах, часто возникает необходимость найти вероятность, с которой событие A в n испытаниях появится не менее m_1 и не более m_2 раз.

Теорема 40.6 ***(Муавра – Лапласа, интегральная).*** Пусть вероятность p появления некоторого события A постоянна в n независимых испытаниях и отлична от 0 и 1, тогда вероятность того, что m – число появления события A при n испытаниях – заключено между $a = np + \alpha\sqrt{npq}$ и $b = np + \beta\sqrt{npq}$ ($a < b$), удовлетворяет соотношению

$$\lim_{n\to\infty} \frac{P(a < m < b)}{\Phi(\beta) - \Phi(\alpha)} = 1,$$

где $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^x e^{-u^2/2} du$ – функция Лапласа, $x \in R$, а $\alpha = \dfrac{a - np}{\sqrt{npq}}$ и $\beta = \dfrac{b - np}{\sqrt{npq}}$ ($q = 1 - p$).

Замечание. При достаточно больших n следует приближенное равенство

$$P(a < m < b) \approx \Phi(\beta) - \Phi(\alpha), \tag{40.8}$$

называемая интегральной формулой Муавра-Лапласа.

Замечание. Функция Лапласа $\Phi(x)$, также как и $\varphi(x)$, является табличной функцией, занесённой в справочники.

Свойства функции Лапласа

1) *$\Phi(0)=0$.*

□ Выполняется согласно определения и свойств определённого интеграла

$$\Phi(0)=\frac{1}{\sqrt{2\pi}}\int_0^0 e^{-\frac{u^2}{2}}du=0. \blacksquare$$

2) *$\Phi(x)$ – нечетная функция, т.е.*

$\Phi(-x)=-\Phi(x)$ $\forall x \in R$.

□ По свойствам определённого интеграла

$$\Phi(-x)=\frac{1}{\sqrt{2\pi}}\int_0^{-x} e^{-\frac{u^2}{2}}du=-\frac{1}{\sqrt{2\pi}}\int_{-x}^{0} e^{-\frac{u^2}{2}}du=\left|\text{Функция } e^{-\frac{u^2}{2}} - \text{четная}\right|=$$

$$=-\frac{1}{\sqrt{2\pi}}\int_0^{x} e^{-\frac{u^2}{2}}du=-\Phi(x). \blacksquare$$

3) $\Phi(\infty)=\frac{1}{2}$.

□ Для доказательства воспользуемся табличным интегралом Пуассона

$$\int_{-\infty}^{+\infty} e^{-x^2}dx=\sqrt{\pi}.$$

$$\Phi(\infty)=\frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-\frac{u^2}{2}}du=\left|\begin{matrix}\text{Обозначим}\\ \frac{u}{\sqrt{2}}=z, dz=\frac{1}{\sqrt{2}}du\end{matrix}\right|=\frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-z^2}\sqrt{2}dz=\frac{1}{\sqrt{\pi}}\int_0^{\infty} e^{-z^2}dz=$$

$$=\frac{1}{2\sqrt{\pi}}\int_{-\infty}^{+\infty} e^{-z^2}dz=\left|\begin{matrix}\text{интеграл}\\ \text{Пуассона}\end{matrix}\right|\frac{1}{2\sqrt{\pi}}\sqrt{\pi}=\frac{1}{2}. \blacksquare$$

Замечание. Пользуясь интегральной формулой Муавра–Лапласа, можно показать, что вероятность отклонения частоты $\frac{m}{n}$ от вероятности p может быть приближённо найдена по формуле

$$P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)\approx 2\Phi\left(\sqrt{\frac{n}{pq}}\right).$$

Пример 40.8. При установившемся технологическом процессе фабрика выпускает в среднем 75% продукции первого сорта. Чему равна вероятность того, что в партии из 100 изделий число первосортных заключено между 70 и 80 деталями?

◁ Воспользуемся интегральной теоремой Лапласа. По условию задачи $n = 100$; $p = 0{,}75$; $q = 0{,}25$; $a = 70$; $b = 80$.

Вычислим:

$$\frac{a - np}{\sqrt{npq}} = \frac{70 - 100 \cdot 0{,}75}{\sqrt{100 \cdot 0{,}75 \cdot 0{,}25}} = \frac{70 - 75}{2{,}5\sqrt{3}} = -\frac{2\sqrt{3}}{3} \approx -1{,}1;$$

$$\frac{b - np}{\sqrt{npq}} = \frac{80 - 100 \cdot 0{,}75}{\sqrt{100 \cdot 0{,}75 \cdot 0{,}25}} = \frac{5}{2{,}5\sqrt{3}} = \frac{2\sqrt{3}}{3} \approx 1{,}1.$$

Тогда с учетом, что функция Лапласа нечетна $\Phi(-1{,}1) = -\Phi(1{,}1) = 0{,}3643$ (по таблице приложения 2).

Искомая вероятность

$P_{100}(70 < m < 80) \approx \Phi(1{,}1) - \Phi(-1{,}1) = 2\Phi(1{,}1) = 2 \cdot 0{,}3643 = 0{,}7286$. ▷