Статистика и анализ данных в R

Лекция 4. Введение в теорию вероятностей. Часть 2

(1.10.2021)

Даниил Литвинов

Теория вероятностей

Событие. Вероятность события. Свойства вероятностей и операции над ними

События

Ω = {**Зелёный**, **Красный**, **Синий**, **Желтый**, **Черный**, **Оранжевый**}

Ω — *пространство элементарных событий*

A — при броске выпадет цвет, который есть на флаге России. Тогда A = {**Красный**, **Синий**}

A — **<u>событие</u>**

$$P(A) = \frac{\#A}{\#\Omega} = \frac{2}{6} = \frac{1}{3}$$

Вероятность

Вероя́тность — степень (относительная мера, количественная оценка) возможности наступления некоторого события.

- Классическое определение

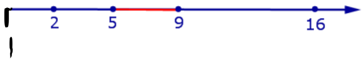
$$P(A) = \frac{n}{N}$$

n — число исходов, удовлетворяющих событию A
N — общее число исходов

$P(5) = 0$
$P(5-6) = \frac{1}{16}$

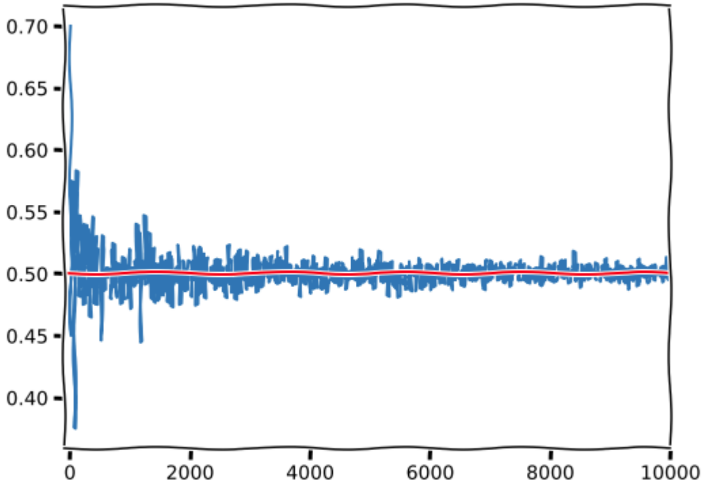
- Геометрическое

$$P(A) = \frac{s}{S}$$

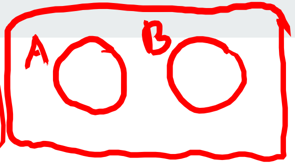


- Частотное

$$P(A) = \lim_{N\to\infty} \frac{n}{N}$$



Что можно делать с событиями? $P(A \cap B) = P(A)P(B)$



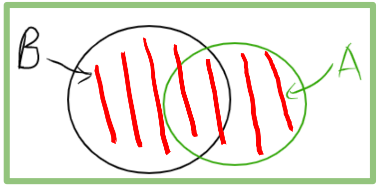
Ω

1. *Объединением* двух событий A ∪ B называется такое событие, которое произойдет, если произойдет одно из событий A или B.

2. *Пересечением* двух событий A ∩ B называется такое событие, которое произойдет, если одновременно произойдут два события A и B.

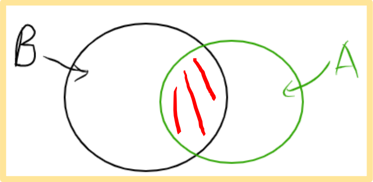
3. События можно также вычитать, это обозначается так B \ A.

4. От события можно брать отрицание. Обозначается как Ā.



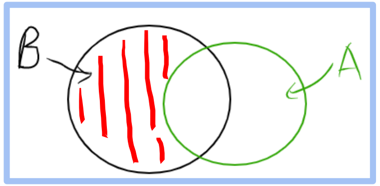
$P(A \cup B) =$

$P(A) + P(B) - \underline{P(A \cap B)}$



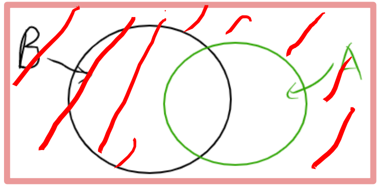
$P(A \cap B) =$

$P(A \cap B) = P(A)P(B)$



$P(B \setminus A) =$

$P(B) - P(B \cap A)$



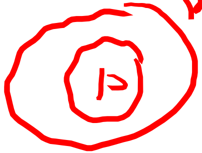
$P(\overline{A}) =$

Свойства вероятностей

1) $P(\Omega)=1$

2) $P(A)\geq 0$

3) $\{\}=\varnothing$ $P(\varnothing)=0$

4) 

B

$A\subset B$, $P(A)\leq P(B)$

5)

Немножко комбинаторики

$n! = n(n-1)\dots1$

Мы хотим собрать из музыкальных инструментов различные комбинаторные объекты. Пусть в данном случае n = 3, а k = 2

3 2 1

$$P_n = n!$$

перестановка

3 2

$$A_n^k = \frac{n!}{(n-k)!}$$

размещение

3 2 1

3 2

$$\frac{3!}{(3-2)!} =$$

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

сочетание

$$\boxed{\begin{aligned} &(a+b)^2 = a^2+2ab+b^2 \\ &(a+b)^3 = a^3+b^3+3a^2b+3ab^2 \\ &(a+b)(a+b)(a+b) = \\ &a^3 + C_3^2 a^2 b + C_3^1 ab^2 + b^3 \end{aligned}}$$

Порешаем задачи…

Какое количество исходов может быть получено для следующих испытаний:

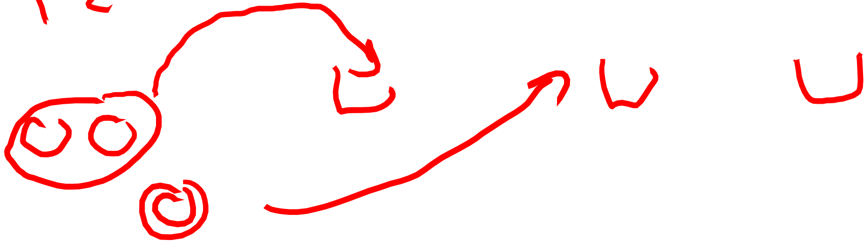
1. Производится выстрел по мишени, представляющей собой 10 концентрических кругов, занумерованных числами от 1 до 10   11
2. Три раза подбрасывается игральная кость $6 \cdot 6 \cdot 6$
3. Производится выстрел по мишени, представляющей собой 10 концентрических кругов, занумерованных числами от 1 до 10, а затем столько раз кидается игральная кость, сколько очков выбито на мишени

$$\underbrace{6^0 + 6^1 \ldots + 6^{10}}$$

Со звёздочкой *

Пусть n ⩾ 3. Шарики занумерованы числами от 1 до n. Найдите количество способов эти n шариков разместить в n разных ящиков так, чтобы ровно один ящик оказался пустым.

$\binom{n}{2} \cdot A_n^{n-1} =$



$3 - 2 = 1$

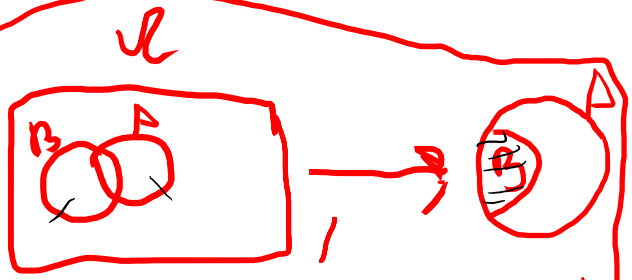
Со звёздочкой *

Вероятность хотя бы одного попадания в цель при четырех выстрелах равна 0,9984. Найдите вероятность попадания в цель при одном выстреле.

Условная вероятность. Формула полной вероятности. Теорема Байеса

Условная вероятность

$$P(B|A) = \frac{\#(A \cap B)}{\#A} = \frac{\#(A \cap B)/\#\Omega}{\#A/\#\Omega} = \frac{P(A \cap B)}{P(A)}$$



Пример:

B - оба ребенка девочки

A1 - старший ребёнок девочка

A2 - хотя бы один ребёнок девочка

Ω = {ДД, ДМ, МД, ММ}

$P(B|A_1)$ $\quad \frac{1}{4} : \frac{1}{2}$

$P(B|A_2)$ $\quad \frac{1}{4} : \frac{3}{4}$

$P(B|A) = \frac{\#(B \cap A)}{\#A}$

Свойства условной вероятности

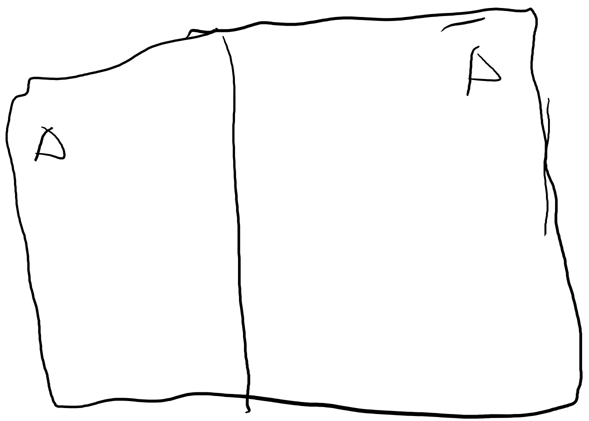
Формула полной вероятности

$$\Omega = A_1 \cup A_2 \cup \ldots \cup A_n$$

$$A_j \cap A_i = \emptyset$$

$$B = B \cap \bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{n} B \cap A_k$$

$B \cap A_k$ и $B \cap A_j$ не пересекаются



Тогда

$$P(B) = \sum_{k=1}^{n} P(B \cap A_k) = \sum_{i=1}^{n} P(B|A_k)P(A_k)$$

Задача

В первом (1) ящике лежат 3 белых (б) и 5 черных (ч) шара. Во втором (2) ящике лежат 5 б и 5 ч шара. Из 1 ящика перекладываем 2 шара во 2 ящик. Вытаскиваем из 2 ящика 1 шар. Какова вероятность, что он будет белый?

Теорема Байеса

$A, B; \quad P(A), P(B) > 0$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$P(A \cap B) = P(A|B) \cdot P(B) \qquad P(B \cap A) = P(B|A) \cdot P(A)$$

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

$$\boxed{P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}}$$

Теорема Байеса

$$\boxed{P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\underbrace{\sum_{i=1}^{n} P(B|A_i)P(A_i)}_{\substack{\text{ф-ла полной вероятн.} \\ P(B)}}}}$$

Чья это неподписанная пробирка???

Вы работаете в лаборатории, где только и занимаетесь тем, что ставите ПЦР, но у вас это пока не очень удачно выходит: из 5 пробирок только в одной все проходит нормально. Также у вас есть очень умелый коллега, у которого в 19 из 20 пробирок все проходит отлично. Вчера вы поставили 3 реакции, а ваш коллега 17. Так вышло, что никто из вас не подписал свои пробирки. В итоге в одной пробирке реакция не прошла. Теперь нужно как-то понять, чья же она…

$A_в$

$A_к$

B – не прошла

$$P(A_в) = \frac{3}{20} \qquad P(B|A_в) = \frac{4}{5}$$

$$P(A_к) = \frac{17}{20} \qquad P(B|A_к) = \frac{1}{20}$$

$$P(B)$$

$$P(B) = \frac{4}{5}\cdot\frac{3}{20} + \frac{17}{20}\cdot\frac{1}{20} = \frac{48+17}{400} = \frac{65}{400}$$

$$P(A_в|B) = \frac{4/5 \cdot 3/20}{65/400} = \frac{48/400}{65/400} = \boxed{\frac{48}{65} > \frac{17}{65}}$$

$$P(A_к|B) = \frac{17/400}{65/400}$$

Спам или не спам...

„привет питон" | Наивный Б. классиф

привет питон = m

$$P(\text{привет}|Н) = \frac{10}{34}$$

$$P(\text{питона}|Н) = \frac{4}{34}$$

$$P(Н|\text{привет питон}) = \frac{P(\text{прив пит}|Н)\cdot P(Н)}{P(\text{прив пит})}$$

$P(Н) = \frac{8}{12}$

$$P(\text{прив питон}|Н) = P(\text{прив}|Н)\cdot P(\text{пит}|Н) = \frac{10}{34}\cdot\frac{4}{34}$$

$\boxed{\text{Сравниваем числители}}$

норм 34

8

Привет 10
эр 4
чем 4
лучше 4
питон 4
давай 8

спам 29

4

лучше 3
чем 3
скидки 9
закрытие 6
эр 3
Привет 2
питон 3

$$P(Н|m) = \frac{10/34\cdot 4/34\cdot 8/12}{P(m)}$$

$$P(С|m) = \frac{2/29\cdot 3/29\cdot 4/12}{P(m)}$$

Спам или не спам...

давай скидки
скидки скидки
скидки скидки

скидки - 1

норм

8

Привет 10++ эр 4 ++
чем 4
лучше 4
питон 4 ++
давай 8

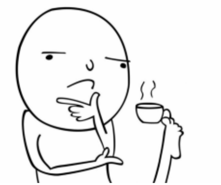
спам

4

лучше 3
чем 3
скидки 9
закрытие 6 ++
эр 3
Привет 2 ++
питон 3

давай

Спам или не спам...



питон лучше чем эр == эр лучше чем питон

Случайные величины (СВ). Основные характеристики СВ

рост, вес и ...

ξ, η

$$\boxed{\xi : \Omega \to \mathbb{R}}$$

отображение

Случайные величины

1. Дискретные

Сколько раз надо кинуть монетку до выпад О.

2. Непрерывные

Математическое ожидание

$E(\quad + \quad) = 7$

$$\boxed{E\xi = \sum_{x \in \xi(\Omega)} x * P(\xi = x)}$$

$1 - \frac{1}{6}$

$2 - \frac{1}{6}$

$3 - \frac{1}{6}$

$4 - \frac{1}{6}$

$5 - \frac{1}{6}$

$6 - \frac{1}{6}$

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = 3\frac{1}{2}$$

Свойства математического ожидания

Если $\xi \geq 0$, то $E\xi \geq 0$

$$E\xi = \sum x_i \cdot p_i \geq 0$$

$$|E\xi| \leq E|\xi|$$

ξ: 1 −1 2

$|\xi|$: $|1|$ $|-1|$ 2

$$\frac{1}{3} - \frac{1}{3} + \frac{2}{3} = \frac{2}{3}$$

$$\frac{1}{3} + \frac{1}{3} + \frac{2}{3} = \frac{4}{3}$$

Свойства математического ожидания

$$E(\xi + \overset{\text{const}}{a}) = E\xi + a$$

$$E(a\xi + b\eta) = aE\xi + bE\eta$$

Док-во:

$$\sum_{\omega\in\Omega} \left(a\cdot\xi(\omega) + b\cdot\eta(\omega)\right)\cdot P(\omega) = \sum_{\omega\in\Omega}\left[a\cdot\xi(\omega)\cdot P(\omega) + b\cdot\eta(\omega)\cdot P(\omega)\right]$$

$$= \sum_{\omega\in\Omega} a\,\underbrace{\xi(\omega)\cdot P(\omega)}_{E\xi} + \sum_{\omega\in\Omega} b\cdot\underbrace{\eta(\omega)\cdot P(\omega)}_{E\eta} = \boxed{aE\xi + bE\eta}$$

Свойства математического ожидания

Если $\xi \geq \eta$, то $E\xi \geq E\eta$

Свойства математического ожидания

Если ξ, η — независимые, то $\underline{E(\xi\eta) = E\xi E\eta}$

$$E(\xi\eta) = \sum_{\substack{x \in \xi(\Omega) \\ y \in \eta(\Omega)}} x \cdot y \cdot \underbrace{P(\xi = x, \eta = y)}_{\substack{\| \\ P(\xi = x) P(\eta = y)}} =$$

$$= \sum_{\substack{x \in \xi(\Omega) \\ y \in \eta(\Omega)}} x \cdot y \cdot P(\xi = x) P(\eta = y) = \underbrace{\sum x \cdot P(\xi = x)}_{E\xi} \cdot \underbrace{\sum y (\eta = y)}_{E\eta}$$

Дисперсия

$$D = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$D\xi = E((\xi - E\xi)^2)$$

$$\sigma = \sqrt{D\xi}$$ ~ станд. отклон.

Свойства дисперсии

$$D\xi \geq 0$$

Свойства дисперсии

$E\xi = a$

$$D\xi = E(\xi^2) - (E\xi)^2$$

$$D\xi = E((\xi - a)^2) = E(\xi^2 - 2a\xi + a^2) =$$

$$= E(\xi^2) - 2\underset{E\xi}{\underset{\|}{a}} E\xi + (E\xi)^2 = E(\xi^2) - 2E\xi \cdot E\xi + (E\xi)^2$$

$$= \boxed{E(\xi^2) - (E\xi)^2}$$

Свойства дисперсии

$$D(\xi + a) = ? \quad D\xi$$

const

$$D(\xi + a) = E\left[\left((\xi + a) - \overbrace{E(\xi + a)}^{E\xi + a}\right)^2\right] =$$

$$= E\left(\left(\xi + \cancel{a} - E\xi - \cancel{a}\right)^2\right) = E\left[(\xi - E\xi)^2\right]$$

Свойства дисперсии

$D(c\xi) = ?$ $c^2 D\xi$

const

5 и 15

$(5-10)^2 \cdot \frac{1}{2} + (15-10)^2 \cdot \frac{1}{2} = 25$

$\cdot 10$

50 и 150

$(50-100)^2 \cdot \frac{1}{2} + (150-100)^2 \cdot \frac{1}{2} = 2500$

$$D(c\xi) = E(c\xi)^2 - (E(c\xi))^2 = E(c^2\xi^2) - E(c\xi) \cdot E(c\xi) =$$

$$= c^2 E\xi^2 - c^2 E\xi \cdot E\xi = c^2 \cdot E\xi^2 - c^2 (E\xi)^2 =$$

$$= c^2(E\xi^2 - (E\xi)^2) = c^2 D\xi$$

Дисперсия суммы СВ

$$D(\xi+\eta) = \underbrace{D\xi + D\eta + \ldots\ldots}$$

Коэффициент корреляции

$cov(\xi, \eta) = E((\xi - E\xi)\cdot(\eta - E\eta))$

$$\rho(\xi,\eta) = \frac{cov(\xi,\eta)}{\sqrt{D\xi}\sqrt{D\eta}}$$

$[-1, 1]$

1) $\rho(\xi, \xi)$

2) $\rho(\xi, -\xi)$

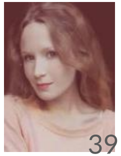
Монеточка

Испытание Бернулли

ξ		ξ	ξ^2
„успех"	p	1	1
„неудача"	$q = 1 - p$	0	0

$$E\xi = 1 \cdot p + 0 \cdot q = p$$

$$D\xi = E(\xi^2) - (E\xi)^2 = p - p^2 = p(1-p) = pq$$



Если n испытаний:

ξ - кол-во успехов

ξ_k - 1/0 на ~~к испытании~~

$$\xi = \sum_{i=1}^{n} \xi_i$$

$$E\xi = E\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \underbrace{E\xi_i}_{p} = np$$

$$D\xi = D\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \overbrace{D\xi_i}^{pq} = npq$$