

Теория вероятностей и математическая статистика

Геометрическая вероятность. Полная вероятность. Формула Байеса.

Глеб Карпов

Программы БИ, УЦП, ТАДвБ ВШБ ВШЭ.

Геометрическая вероятность

Определение

Геометрическая вероятность применяется к экспериментам, где пространство элементарных исходов Ω является подмножеством евклидова пространства (отрезок, площадь, объем).

Основные формулы

Для одномерного случая (отрезок):

$$P(A) = \frac{\text{длина множества } A}{\text{длина множества } \Omega}$$

Для двумерного случая (площадь):

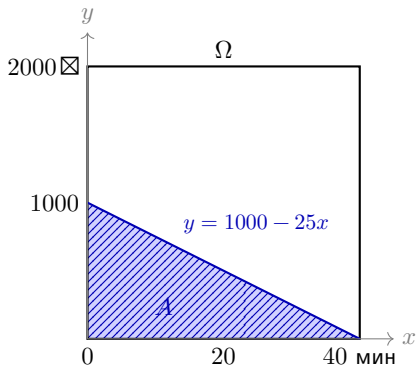
$$P(A) = \frac{\text{площадь множества } A}{\text{площадь множества } \Omega}$$

Для трехмерного случая (объем):

$$P(A) = \frac{\text{объем множества } A}{\text{объем множества } \Omega}$$

Геометрическая вероятность

Пример: кэшбэк за поездку



Условие задачи:

Время поездки X выбирается случайно на $[0, 40]$ мин, стоимость Y – случайно на $[0, 2000]$ ⌘, независимо. Кэшбэк дают, если поездка достаточно короткая и достаточно дешёвая: каждая минута снижает допустимый порог цены на 25 ⌘, то есть $Y \leq 1000 - 25X$. Какова вероятность кэшбэка?

Решение:

На слайде Ω нарисован квадратом, но оси в разных единицах:

$$\Omega = [0, 40] \times [0, 2000], \quad S_{\Omega} = 40 \cdot 2000 = 80\,000$$

Область кэшбэка $A = \{(x, y) : y \leq 1000 - 25x\}$ – треугольник с катетами 40 мин и 1000 ⌘:

$$S_A = \frac{1}{2} \cdot 40 \cdot 1000 = 20\,000$$

Ответ: $P(A) = \frac{S_A}{S_{\Omega}} = \frac{1}{4}$

Геометрическая вероятность

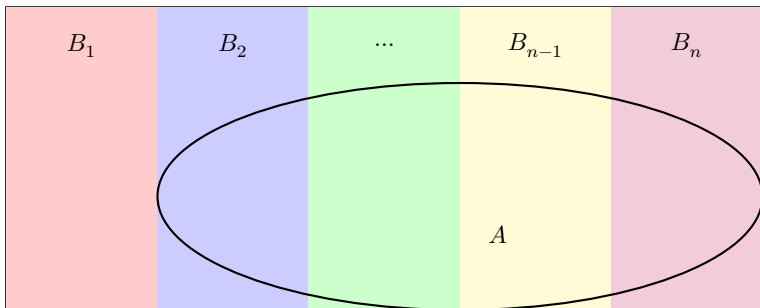
Классические задачи геометрической вероятности

1. **Задача о встрече:** Два человека договариваются встретиться в определенном месте между 12:00 и 13:00. Каждый приходит в случайный момент времени и ждет 15 минут. Какова вероятность встречи?
2. **Задача Бюффона:** Игла длины l бросается на плоскость, разлинованную параллельными прямыми на расстоянии d друг от друга. Какова вероятность пересечения иглы с линией?
3. **Задача о случайной точке в круге:** Точка выбирается случайно внутри круга радиуса R . Какова вероятность того, что расстояние от точки до центра меньше r ?

Полная вероятность

- Концепция условной вероятности неразрывно связана со следующей, концепцией полной вероятности.
- Рассмотрим зафиксированное пространство $(\Omega, \mathcal{F}, P)$. Назовем **разбиением** Ω коллекцию событий $\{B_k, k \in I\}$, таких что $B_i \cap B_j = \emptyset$ при $i \neq j$ и $\bigcup B_i = \Omega$.
- Вдобавок, рассмотрим какое-то другое событие B , которое пересекается с какими-то событиями из разбиения, но не обязано пересекаться со всеми.

Ω



Полная вероятность

i Theorem

Если $\{B_1, B_2, \dots\}$ - разбиение Ω , с $P(B_i) > 0 \forall i$, то:

$$P(A) = \sum_i P(A|B_i)P(B_i), \forall A \in \mathcal{F}$$

Доказательство. Заметим, что мы можем реконструировать событие A из его частичек-пересечений со всеми B_i : $A = \bigcup_i (A \cap B_i)$. Эти кусочки $\{A \cap B_i\}$ попарно не пересекаются, как и оригинальные элементы разбиения. Поэтому далее можем применить свойство аддитивности вероятности:

$$\begin{aligned} P(A) &= P\left(\bigcup_i (A \cap B_i)\right) \\ &= \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i) \end{aligned}$$

Теорема Байеса

i Theorem

Пусть $\{B_1, B_2, \dots\}$ - разбиение Ω , с $P(B_i) > 0 \forall i$. Тогда для любого события A , $P(A) > 0$, справедливо:

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)}$$

Доказательство.

Опираемся на то, что $P(X \cap Y)$ может быть выражена через разные условные вероятности:

$$P(X \cap Y) = P(X|Y)P(Y) = P(Y|X)P(X).$$

Применим это к условию

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)},$$

где $P(A) = \sum_i P(A|B_i)P(B_i)$ как полная вероятность.

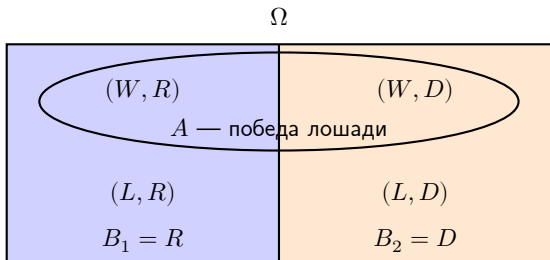
Пример 1

Вероятность победы лошади в скачках равна 0.3, если погода сухая, и 0.5 если идет дождь. Прогноз погоды дает вероятность дождя 40%.

1. Найдите вероятность победы лошади.
2. Если вы знаете, что лошадь проиграла гонку, какова вероятность того, что погода была сухой в день скачек? (Предположим, что вы не помните!)

Пример 1

Обозначим: R — дождь, D — сухая погода; W/L — победа/проигрыш; A — победа лошади.



- $A = (A \cap B_1) \cup (A \cap B_2)$

- $A \cap B_1 = \{(W, R)\}$

- $A \cap B_2 = \{(W, D)\}$

-

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) \\ &= P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) \end{aligned}$$

Пример 1

- $P(A | D) = 0.3, P(A | R) = 0.5, P(R) = 0.4, P(D) = 0.6.$
- $P(A) = P(A | D)P(D) + P(A | R)P(R) = 0.3 \cdot 0.6 + 0.5 \cdot 0.4 = 0.38$

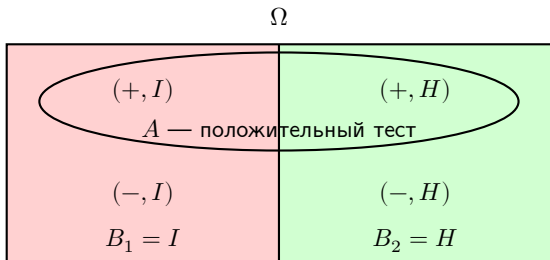
- $$P(D | \bar{A}) = \frac{P(\bar{A} | D)P(D)}{P(\bar{A})} = \frac{0.7 \cdot 0.6}{1 - 0.38} \approx 0.677$$

Пример 2 (False positives)

- Редкое, но опасное заболевание имеет распространенность 1 на 10^5 в общей популяции. Существует диагностический тест, но он несовершенен. Если у человека есть заболевание, тест положительный с вероятностью 0.9; если нет - тест положительный с вероятностью $\frac{1}{20}$. Тест дал положительный результат. Какова вероятность того, что у человека **действительно** есть заболевание?

Пример 2 (False positives)

Обозначим: H — healthy, I — ill; A — положительный результат теста (detected positive).



- $A = (A \cap B_1) \cup (A \cap B_2)$

- $A \cap B_1 = \{(+, I)\}$

- $A \cap B_2 = \{(+, H)\}$

-

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) \\ &= P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) \end{aligned}$$

Пример 2 (False positives)

- $P(I) = 10^{-5}$, $P(A | I) = 0.9$, $P(A | H) = \frac{1}{20}$.
- $P(A) = P(A | I)P(I) + P(A | H)P(H) = 0.9 \cdot 10^{-5} + 0.05 \cdot (1 - 10^{-5}) = 0.0500085$

- $$P(I | A) = \frac{P(A | I)P(I)}{P(A)} = \frac{0.9 \cdot 10^{-5}}{0.0500085} \approx 0.0002$$