Теория вероятностей и математическая статистика

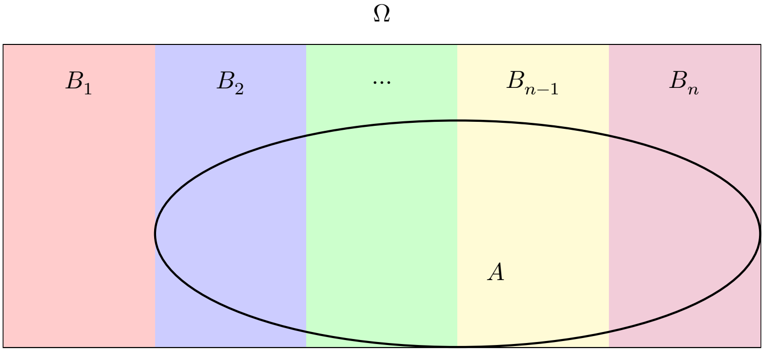
Полная вероятность. Теорема Байеса.

Глеб Карпов

ФКН ВШЭ

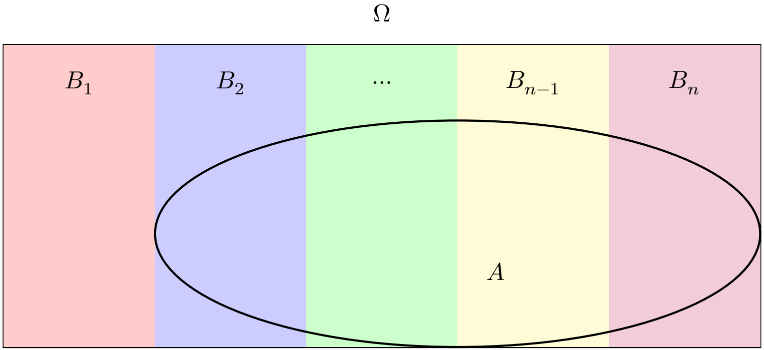
Полная вероятность

- Концепция условной вероятности неразрывно связана со следующей, концепцией полной вероятности.



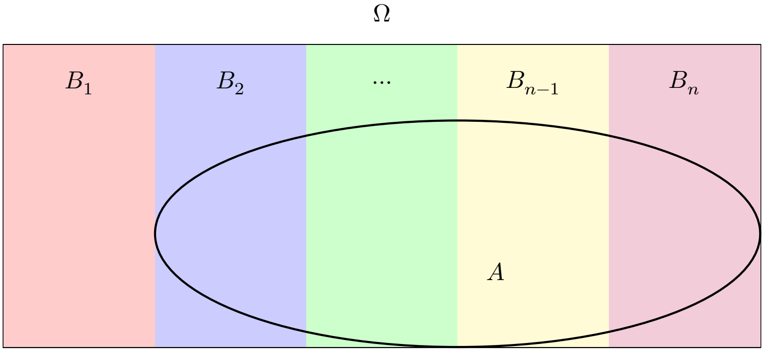
Полная вероятность

- 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 $\bigcup B_i = \Omega$.



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.



Полная вероятность

Theorem

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

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

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:

$$P(A) = P\left(\bigcup_i (A \cap B_i)\right)$$
$$= \sum_i P\left(A \cap B_i\right) = \sum_i P(A|B_i)P(B_i)$$

Теорема Байеса

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:

$$\boxed{P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum\limits_i P(A|B_i)P(B_i)}}$$

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 $P(A) = \sum\limits_i P(A|B_i)P(B_i)$ 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

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!)

- $P(V) = P(V|D)P(D) + P(V|W)P(W) = 0.3 \cdot 0.6 + 0.5 \cdot 0.4 = 0.38$

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!)

- $P(V) = P(V|D)P(D) + P(V|W)P(W) = 0.3 \cdot 0.6 + 0.5 \cdot 0.4 = 0.38$
- $$P(D|\bar{V}) = \frac{P(\bar{V}|D)P(D)}{P(\bar{V})} = \frac{0.7 \cdot 0.6}{1 - 0.38} \approx 0.677$$

Пример 2

- Есть три монеты: у одной с обеих сторон орел, у второй с обеих сторон решка, у третьей орел с одной стороны и решка с другой. Мы выбираем монету наугад, подбрасываем её, и выпадает орел. Какова вероятность того, что на противоположной стороне решка?

Пример 2

- Есть три монеты: у одной с обеих сторон орел, у второй с обеих сторон решка, у третьей орел с одной стороны и решка с другой. Мы выбираем монету наугад, подбрасываем её, и выпадает орел. Какова вероятность того, что на противоположной стороне решка?
- C_1, C_2, C_3 – события, соответствующие выбору монеты согласно условиям задачи. A – событие, соответствующее выпадению орла.

$$P(A) = P(A|C_1)P(C_1) + P(A|C_2)P(C_2) + P(A|C_3)P(C_3) = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 0.5 \cdot \frac{1}{3} = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}$$

Согласно условию, нам нужно узнать вероятность того, что была выбрана третья монета.

$$P(C_3|A) = \frac{P(A|C_3)P(C_3)}{P(A)} = \frac{0.5 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

Пример 3 (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 3 (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?
- Обозначим вероятности: (H = healthy, I = ill, $+$ for detected positive, $-$ for detected negative) $P(I) = 10^{-5}$, $P(+|I) = 0.9$, $P(+|H) = \frac{1}{20}$. Мы заинтересованы в $P(I|+)$.

Пример 3 (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?
- Обозначим вероятности: (H = healthy, I = ill, $+$ for detected positive, $-$ for detected negative) $P(I) = 10^{-5}$, $P(+|I) = 0.9$, $P(+|H) = \frac{1}{20}$. Мы заинтересованы в $P(I|+)$.
- $P(+) = P(+|I)P(I) + P(+|H)P(H) = 0.9 \cdot 10^{-5} + 0.05 \cdot (1 - 10^{-5}) = 0.0500085$

Пример 3 (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?
- Обозначим вероятности: (H = healthy, I = ill, $+$ for detected positive, $-$ for detected negative) $P(I) = 10^{-5}$, $P(+|I) = 0.9$, $P(+|H) = \frac{1}{20}$. Мы заинтересованы в $P(I|+)$.
- $P(+) = P(+|I)P(I) + P(+|H)P(H) = 0.9 \cdot 10^{-5} + 0.05 \cdot (1 - 10^{-5}) = 0.0500085$
- $$P(I|+) = \frac{P(+|I)P(I)}{P(+)} = \frac{0.9 \cdot 10^{-5}}{0.0500085} \approx 0.0002$$