

Коллоквиум 2

1. Теорема Байеса;

Пусть $H_1, H_2, \dots, H_n$ ($n \leq \infty$) – полная группа гипотез на Ω , причем $\mathbb{P}(H_j) > 0$ для всех j
Тогда для любого события $A \subseteq \Omega$ с $\mathbb{P}(A) > 0$ и любого j выполняется:

$$\mathbb{P}(H_j|A) = \frac{\mathbb{P}(A|H_j) \mathbb{P}(H_j)}{\sum_{i=1}^n \mathbb{P}(A|H_i) \mathbb{P}(H_i)}$$

2. Априорная и апостериорная вероятности;

$\mathbb{P}(H_j)$ – априорные вероятности

$\mathbb{P}(H_j|A)$ – апостериорные вероятности

3. Шансы, связь с вероятностью;

Шансы (odds) это другой способ выражения вероятности, определяемый как

$$odds(A) := \frac{\mathbb{P}(A)}{\mathbb{P}(\bar{A})}$$

$$\mathbb{P}(A) = \frac{odds(A)}{odds(A) + 1}$$

4. True Positive, False Positive, True Negative, False Negative;

Предсказание →	Болен	Здоров
Болен	True positive	False Negative
Здоров	False positive	True Negative

5. Точность (accuracy);

$$= \frac{TP + TN}{TP + TN + FP + FN}$$

6. Чувствительность (sensitivity);

$$= \frac{TP}{TP + FN}$$

$\mathbb{P}(A|H) =$ чувствительность

7. Специфичность (specificity);

$$= \frac{TN}{TN + FP}$$

$\mathbb{P}(A|\bar{H}) = 1 -$ специфичность

8. Коэффициент Байеса и его связь с чувствительностью и специфичностью;

$$K = \frac{\mathbb{P}(A|H)}{\mathbb{P}(A|\bar{H})} = \frac{\text{чувствительность}}{1 - \text{специфичность}}$$

9. Типы данных (количественные/качественные, целые/непрерывные, номинальные/порядковые);

Качественные данные - наблюдения, которые можно разделить на группы или категории

Номинальные - не имеет порядка

Порядковые - имеет определенный порядок

Количественные - наблюдения или измерения

Дискретные (целые) - можно посчитать, отличны друг от друга

Непрерывные - значения в заданном диапазоне

10. Эмпирическая функция распределения;

$$F_n^*(t) = \frac{\#\{X_i: X_i < t\}}{n} = \frac{1}{n} \sum_{i=1}^n I_{\{X_i < t\}}$$

11. Выборочная квантиль;

Для $p \in [0, 1]$

$$X_{(p)} := X_{(\lfloor pn \rfloor)}$$

12. Выборочное среднее;

$X_1, X_2, \dots, X_n$ — наши данные

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

13. Геометрическое среднее;

$$\sqrt[n]{\prod_{k=1}^n X_k}$$

14. Медиана для выборки;

Выборочная медиана $\mu^{\wedge}$ определяется отдельно для чётных и нечетных n .

Для $n = 2k + 1$: $\hat{\mu} = X_{k+1}$

Для $n = 2k$: $\hat{\mu} = \frac{1}{2}(X_{k+1} + X_k)$

$$X_{(\lfloor n/2 \rfloor)}$$

15. Выборочная дисперсия;

$$S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

16. Стандартное отклонение;

$$S_0 = \sqrt{S_0^2}$$

17. Размах;

$$X_{(n)} - X_1$$

18. Межквартильный размах;

$$X_{(\lfloor 3n/4 \rfloor)} - X_{(\lfloor n/4 \rfloor)}$$

19. Как строится бутстреп выборка из данной выборки?

Бутстреп выборка строится путём многократного случайного выбора элементов с повторением из исходной выборки