Прикладная статистика в машинном обучении

Квиз #4C

---

15 декабря 2020 г.

В каждом вопросе выберите все верные ответы.

Рассмотрим выборку независимых одинаково распределённых случайных величин X_1, X_2, …, X_N из нормального распределения $\mathcal{N}(\mu, 100)$. Предположим, что априорное распределение μ является нормальным $\mathcal{N}(4, 2)$.

1. На основе условия задачи можно сделать вывод, что
   - A. $f(\mu|X) = \prod_i ce^{-\frac{(X_i-\mu)^2}{4}}$.
   - B. $f(X|\mu) = \prod_i ce^{-\frac{(X_i-\mu)^2}{200}}$.
   - C. $f(X|\mu) = \prod_i ce^{-\frac{(X_i-\mu)^2}{200}}$.
   - D. $f(\mu|X) = ce^{-\frac{(2-X_i)^2}{4}}$.
   - E. Нет верного ответа.
2. Для простоты далее рассмотрим только наблюдение X_3. Оказалось, что $X_3 = 3$. Апостериорное распределение параметра μ задаётся как
   - A. $f(\mu|X_3) = Ce^{-\frac{(3-\mu)^2}{100} - \frac{(\mu-4)^2}{8}}$.
   - B. $f(X_3|\mu) = Ce^{\frac{(3-\mu)^2}{100} + \frac{(\mu-3)^2}{200}}$.
   - C. $f(\mu|X_3) = Ce^{-\frac{(3-\mu)^2}{200} - \frac{(\mu-4)^2}{4}}$.
   - D. $f(X_3|\mu) = Ce^{-\frac{\mu^2}{4} + \frac{(3-\mu)^2}{100}}$.
   - E. Нет верного ответа.
3. Апостериорное распределение μ с точностью до константы является
   - A. Нормальным и имеющим конечное математическое ожидание и не конечную дисперсию.
   - B. $\mathcal{N}(0, 1)$.
   - C. Бета-распределением.
   - D. Нормальным.
   - E. Нет верного ответа.
4. Константа C
   - A. Вычисляется только приблизительно.
   - B. Обычно отбрасывается и не рассматривается.
   - C. Может быть получена аналитическими методами.
   - D. Совпадает с производной функции правдоподобия в точке моды апостериорного распределения.
   - E. Нет верного ответа.
5. Выражение $\mathbb{P}(\mu \in (c, d)|X_5) = 0.95$

A. Является формулой 95%-го байесовского доверительного интервала.
B. Эквивалентно выражению $\mathbb{P}(\mu \in (c, d)|X_5) < 0.95$.
C. Не может быть рассчитано, если μ – случайная величина.
D. Не имеет смысла в байесовском подходе.
E. Нет верного ответа.

6. Точечная байесовская оценка μ
   A. Находится при использовании апостериорного распределения.
   B. Не вычисляется аналитически.
   C. Всегда совпадает с удвоенным значением LR-статистики.
   D. Равна моде $f(\mu)$.
   E. Нет верного ответа.

Далее будем рассуждать в терминах частотного подхода и считать, что μ – константа.

7. Пусть тестируется гипотеза $H_0 : \mu = -1$ против $H_1 : \mu < -1$. Тогда
   A. При использовании LR-теста p-value не может быть менее 0.12.
   B. При использовании LR-теста p-value не существует.
   C. Геометрически p-value является площадью.
   D. Если при использовании Z-теста p-value окажется 0.99, то H_0 будет отвергнута на любом разумном уровне значимости.
   E. Нет верного ответа.

8. Пусть тестируется гипотеза $H_0 : \mu = 1$ против $H_1 : \mu \neq 1$. Тогда
   A. p-value приблизительно равно 0.1.
   B. Для расчёта p-value используется распределение статистики какого-либо теста в предположении верной H_0.
   C. В качестве оценки p-value можно взять $\bar{X}$.
   D. Если p-value равно 0.01, то нулевая гипотеза не отвергается на уровне значимости 99%.
   E. Нет верного ответа.

При тестировании трёх видов лекарств против плацебо ($H_{0,i} : p_i = p_{plac}$) оказалось, что соответствующие p-value равны 0.000, 0.001, 0.002.

1. На основании условия задачи можно сделать вывод, что на уровне значимости 5%
   A. Только третье лекарство статистически неотличимо от плацебо.
   B. Только первые два лекарства статистически неотличимы от плацебо.
   C. Все лекарства статистически неотличимы от плацебо.
   D. Первое лекарство статистически неотличимо от плацебо на уровне значимости 10%.
   E. Нет верного ответа.

2. При проведении множественного тестирования методом Бонферрони
   A. Только первое лекарство статистически неотличимо от плацебо.
   B. Только первое лекарство статистически отлично от плацебо.
   C. Пороговое значение для отвержения гипотезы следует принять равным $\alpha/3$.
   D. Уровень значимости следует принять равным 1%.
   E. Нет верного ответа.

3. При проведении множественного тестирования методом Бенджамини-Хохберга

A. Невозможно сказать, статистически отлично ли второе лекарство от плацебо.
B. Только первое лекарство статистически отлично от плацебо.
C. Пороговое значение равно 10%.
D. Результаты совпадут с результатами метода Бонферрони.
E. Нет верного ответа.

4. При проведении множественного тестирования методов Бенджамини-Хохберга на уровне значимости 5%
   A. Ровно два лекарства статистически неотличимы от плацебо.
   B. Ровно одно лекарство статистически неотличимо от плацебо.
   C. Все три лекарства статистически неотличимы от плацебо.
   D. Все три лекарства статистически отличны от плацебо.
   E. Нет верного ответа.