Прикладная статистика в машинном обучении

Квиз #4

---

19 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 $\mathcal{N}(\mu, 9)$. Предположим, что априорное распределение μ является нормальным $\mathcal{N}(0, 4)$.

1. На основе условия задачи можно сделать вывод, что
   - A. $f(X) = \prod_i ce^{-\frac{(X_i-\mu)^2}{9}}$.
   - B. $f(\mu) = \prod_i ce^{-\frac{(-\mu)^2}{8}}$.
   - C. $f(X|\mu) = \prod_i ce^{-\frac{(X_i-\mu)^2}{9}}$.
   - D. $f(\mu|X) = ce^{-\frac{(-X_i)^2}{18}}$.
   - E. Нет верного ответа.
2. Для простоты далее рассмотрим только наблюдение X_{99}. Оказалось, что $X_{99} = 5$. Апостериорное распределение параметра μ задаётся как
   - A. $f(\mu|X_{99}) = Ce^{-\frac{(5-\mu)^2}{9}-\frac{(\mu)^2}{18}}$.
   - B. $f(X_{99}|\mu) = Ce^{\frac{(-\mu)^2}{4}+\frac{(\mu-5)^2}{8}}$.
   - C. $f(\mu|X_{99}) = Ce^{-\frac{(-\mu)^2}{8}-\frac{(5-\mu)^2}{18}}$.
   - D. $f(X_{99}|\mu) = Ce^{-\frac{\mu^2}{9}+\frac{(5-\mu)^2}{18}}$.
   - E. Нет верного ответа.
3. Апостериорное распределение μ с точностью до константы является
   - A. Стандартным нормальным.
   - B. Нормальным, имеющим бесконечную дисперсию.
   - C. Вырожденным.
   - D. Распределением Снедекора.
   - E. Нет верного ответа.
4. Константа C
   - A. Может быть рассчитана путём угадывания вида апостериорного распределения.
   - B. Никогда не вычисляется аналитически.
   - C. Равна $-4/\sqrt{9\pi}$.
   - D. Равна значению максимума априорной плотности.
   - E. Нет верного ответа.
5. Выражение $\mathbb{P}(\mu \in (c, d)|X_{99}) = 0.95$

A. Не имеет смысла в байесовском подходе.
B. Всегда совпадает с частотным аналогом.
C. Эквивалентно выражению $\mathbb{P}(\mu \in (c, d)|X_{99}) > 0.95$.
D. Может быть рассчитано, только если μ – случайная величина.
E. Нет верного ответа.

6. Точечная байесовская оценка μ
   A. Может быть средним, модой или дисперсией апостериорного распределения.
   B. Не вычисляется аналитически.
   C. Не существует, если максимум второй производной правдоподобия равен 0 хотя бы в одной точке.
   D. Вычисляется на основе апостериорного распределения.
   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. Геометрически p-value является наклоном плотности.
   B. Если p-value окажется близким к 0.99, то нулевая гипотеза не будет отвергнута на уровне значимости 0.97.
   C. При использовании LM-теста p-value обязательно окажется близким к 0.
   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 0.01.
   B. Если p-value окажется равным 0.000, то нулевая гипотеза не будет отвергнута на уровне значимости 1%.
   C. Для проверки гипотезы достаточно рассчитать лишь p-value / 2.
   D. Если p-value окажется равным 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, 0.99, 0.15.

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

2. При проведении множественного тестирования методом Бонферрони на уровне значимости 5%
   A. Только первое лекарство статистически неотличимо от плацебо.
   B. Только второе лекарство статистически отлично от плацебо.
   C. Пороговое значение для отвержения гипотезы следует принять равным $\alpha/2$.
   D. Не существует разумного уровня значимости, при котором третье лекарство было бы отлично от плацебо.

E. Нет верного ответа.

3. При проведении множественного тестирования методом Бенджамини-Хохберга на уровне значимости 5%
   A. Невозможно сказать, статистически отлично ли первое лекарство от плацебо.
   B. Второе лекарство окажется статистически отличным от плацебо.
   C. Пороговое значение равно 50%.
   D. Результаты будут отличны от результатов 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.