Прикладная статистика в машинном обучении

Квиз #4D

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 $\mathcal{N}(\mu, 15)$. Предположим, что априорное распределение μ является нормальным $\mathcal{N}(1, 1)$.

1. На основе условия задачи можно сделать вывод, что
   - A. $f(\mu|X) = \prod_i ce^{-\frac{(\mu-1)^2}{2}}$.
   - B. $f(X|\mu) = \prod_i ce^{-\frac{(X_i-\mu)^2}{15}}$.
   - C. $f(X|\mu) = ce^{-\frac{(X_i-\mu)^2}{2}}$.
   - D. $f(\mu|X) = ce^{-\frac{(\mu-1)^2}{2}}$.
   - E. Нет верного ответа.
2. Для простоты далее рассмотрим только наблюдение X_1. Оказалось, что $X_1 = 7$. Апостериорное распределение параметра μ задаётся как
   - A. $f(X_1|\mu) = Ce^{\frac{(7-\mu)^2}{30}+\frac{(\mu-7)^2}{30}}$.
   - B. $f(\mu|X_1) = Ce^{-\frac{(7-\mu)^2}{30}-\frac{(\mu-1)^2}{2}}$.
   - C. $f(\mu|X_1) = Ce^{-\frac{(7-\mu)^2}{2}-\frac{(\mu-7)^2}{30}}$.
   - D. $f(X_1|\mu) = Ce^{-\frac{\mu^2}{30}+\frac{(7-\mu)^2}{60}}$.
   - E. Нет верного ответа.
3. Апостериорное распределение μ с точностью до константы является
   - A. Стандартным нормальным.
   - B. Имеющим конечное математическое ожидание.
   - C. Экспоненциальным распределением.
   - D. Нормальным, имеющим бесконечную дисперсию.
   - E. Нет верного ответа.
4. Константа C
   - A. Обычно принимается равной $\bar{X}$.
   - B. Не вычисляется даже приблизительно.
   - C. Иногда вычисляется путём угадывания вида апостериорного распределения.
   - D. Обычно получается отрицательной.
   - E. Нет верного ответа.
5. Выражение $\mathbb{P}(\mu \in (c, d)|X_5) = 0.9$

A. Является формулой 95%-го байесовского доверительного интервала.
B. Является формулой 90%-го частотного доверительного интервала.
C. Не может быть вычислено за конечное число итераций.
D. Может быть вычислено при помощи симуляций.
E. Нет верного ответа.

6. Точечная байесовская оценка μ
   A. Может быть получена из анализа гистограммы апостериорного распределения.
   B. Равна медиане $f(\mu)$.
   C. Всегда совпадает с точечной частотной оценкой μ.
   D. Не может быть равна моде апостериорного распределения.
   E. Нет верного ответа.

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

7. Пусть тестируется гипотеза $H_0 : \mu = 20$ против $H_1 : \mu > 20$. Тогда
   A. Если p-value окажется равным 0.01, то 10%-ом уровне значимости тест Вальда отвергнет нулевую гипотезу.
   B. Если p-value окажется равным 0.01, то 10%-ом уровне значимости LR-тест не отвергнет нулевую гипотезу.
   C. Если при использовании Z-теста p-value окажется 0.99, то H_0 будет отвергнута на любом разумном уровне значимости.
   D. Если используется LR-тест, но p-value не существует.
   E. Нет верного ответа.

8. Пусть тестируется гипотеза $H_0 : \mu = 12$ против $H_1 : \mu \neq 12$. Тогда
   A. p-value обязательно лежит в границах $[0.000, 0.12]$.
   B. Если p-value равно 0.5, то нулевая гипотеза не отвергается на уровне значимости 51%.
   C. p-value обязательно лежит в границах $[0.000, 1.000]$.
   D. Если p-value равно 0.000, то существует разумный уровень значимости, на котором нулевая гипотеза не отвергается.
   E. Нет верного ответа.

При тестировании трёх видов лекарств против плацебо ($H_{0,i} : p_i = p_{plac}$) оказалось, что соответствующие p-value равны 1.000, 1.000, 0.000.

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

E. Нет верного ответа.

3. При проведении множественного тестирования методом Бенджамини-Хохберга
   A. На первом шаге p-value следует упорядочить как 1.000, 1.000, 0.000.
   B. Окажется невозможным сравнить $p_{(1)}$ и $\ell_{(1)}$.
   C. Результаты тестирования (i-ая гипотеза отвергается/не отвергается) не будут совпадать с результатами метода Бонферрони.
   D. Результаты тестирования (i-ая гипотеза отвергается/не отвергается) будут совпадать с результатами без корректировки.
   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.