Прикладная статистика в машинном обучении

Квиз #4А

---

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 $\mathcal{N}(\mu, 4)$. Предположим, что априорное распределение μ является нормальным $\mathcal{N}(1, 2)$.

1. На основе условия задачи можно сделать вывод, что
   - A. $f(X|\mu) = ce^{-\frac{(X_1-\mu)^2}{8}}$.
   - B. $f(\mu|X) = \prod_i ce^{-\frac{(\mu-i)^2}{4}}$.
   - C. $f(X|\mu) = \prod_i ce^{-\frac{(X_i-\mu)^2}{8}}$.
   - D. $f(\mu|X) = ce^{-\frac{\mu^2}{8}}$.
   - E. Нет верного ответа.
2. Для простоты далее рассмотрим только наблюдение X_9. Оказалось, что $X_9 = 10$. Апостериорное распределение параметра μ задаётся как
   - A. $f(X_9|\mu) = Ce^{-\frac{X_9^2}{12}-\frac{(\mu-1)^2}{24}}$.
   - B. $f(\mu|X_9) = Ce^{-\frac{(X_9-1)^2}{4}+\frac{(\mu-1)^2}{8}}$.
   - C. $f(\mu|X_9) = Ce^{-\frac{(X_9-\mu)^2}{8}-\frac{(\mu-1)^2}{4}}$.
   - D. $f(X_9|\mu) = Ce^{-\frac{(X_9-\mu)^2}{12}-\frac{(X_9-1)^2}{8}}$.
   - E. Нет верного ответа.
3. Апостериорное распределение μ с точностью до константы является
   - A. Экспоненциальным.
   - B. t-распределением.
   - C. Распределением Пуассона с $\lambda = \mu$.
   - D. Нормальным и имеющим бесконечное математическое ожидание.
   - E. Нет верного ответа.
4. Константу C
   - A. Невозможно восстановить даже приблизительно.
   - B. Можно восстановить только приблизительно.
   - C. Считают равной 1.
   - D. Иногда можно восстановить аналитически.
   - E. Нет верного ответа.
5. Выражение $\mathbb{P}(\mu \in (c, d)|X_9) = 0.95$

A. Бессмысленно для частотного подхода, а потому и для байесовского подхода.
B. Используется для построения 95%-го частотного доверительного интервала для μ.
C. Имеет смысл, только когда $\operatorname{plim}_{n\to\infty} \mu = 1$.
D. Означает, что с 95%-ой апостериорной вероятностью μ лежит в интервале (c, d).
E. Нет верного ответа.

6. В качестве разумной точечной байесовской оценки μ
   A. Возможно взять среднее, моду или медиану $f(\mu|X)$.
   B. Возможно взять среднее $f(X|\mu)$.
   C. Следует взять $\bar{X}$.
   D. Следует взять ML-оценку математического ожидания $f(\mu)$.
   E. Нет верного ответа.

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

7. Пусть тестируется гипотеза $H_0 : \mu = 0$ против $H_1 : \mu < 0$. Тогда
   A. p-value будет равно $1/2$.
   B. Если p-value окажется меньше 0, то H_0 не будет отвергнута на любом разумном уровне значимости.
   C. p-value равно $\mathbb{P}\{\hat{\mu} < \mu_{obs}|H_0\}$, где μ_{obs} – наблюдаемое значение статистики.
   D. p-value равно $2\,\mathbb{P}\{\hat{\mu} < \mu_{obs}|H_0\}$, где μ_{obs} – наблюдаемое значение статистики.
   E. Нет верного ответа.

8. Пусть тестируется гипотеза $H_0 : \mu = 0$ против $H_1 : \mu \neq 0$. Тогда
   A. p-value возможно рассчитать при использовании теста Вальда, но нельзя при использовании LR-теста.
   B. Если при использовании Z-теста p-value окажется равным 0.06, то H_0 будет отвергнута на 5% уровне значимости.
   C. При использовании LR-теста p-value всегда будет получаться большим 10%.
   D. Если при использовании Z-теста p-value окажется равным 0.06, то H_0 будет отвергнута на 10% уровне значимости.
   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, 0.06, 0.89.

1. На основании условия задачи можно сделать вывод, что на уровне значимости 5%
   A. Не существует разумного уровня значимости, на котором бы отвергалась нулевая гипотеза для третьего лекарства.
   B. Нулевая гипотеза не отвергается только для одного лекарства.
   C. Первое лекарство статистически неотличимо от плацебо на любом разумном уровне значимости.
   D. Второе лекарство статистически неотличимо от плацебо на любом разумном уровне значимости.
   E. Нет верного ответа.

2. При проведении множественного тестирования методом Бонферрони
   A. Каждое p-value необходимо разделить на 3.
   B. Каждое p-value необходимо разделить на 4.
   C. Каждое p-value необходимо разделить на 2.
   D. Только одно лекарство окажется статистически неотличимым от плацебо.

E. Нет верного ответа.

3. При проведении множественного тестирования методом Бенджамини-Хохберга
   A. Каждое p-value необходимо сравнивать с $\alpha/3$.
   B. Все лекарства окажутся статистически неотличимыми от плацебо.
   C. Пороговое значение для отвержения гипотезы зависит от числа проверяемых гипотез.
   D. Результаты тестирования всегда 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.