

## Домашнее задание #3

Дедлайн: 17 декабря, 23:59 МСК

---

### Правила игры

1. Домашнее задание оценивается в 10 баллов.
2. Решения принимаются до **17 декабря 2020 года, 23:59 МСК** включительно. Работы, отправленные после дедлайна, не оцениваются.
3. Все решения нужно загрузить в личный репозиторий на [GitHub Classroom](#).
4. Обе задачи экспериментальные, и их следует решать в Jupyter Notebook. Однако если вам кажется, что где-то нужно привести длинную теоретическую выкладку, то вы можете написать её от руки и отсканировать. В этом случае соберите все написанные вручную элементы в единый .pdf файл и также загрузите его в репозиторий.
5. Репозиторий должен содержать .ipynb-файл с решениями задач, который должен называться name\_surname\_hw3.ipynb. Если вы дополнительно загружаете .pdf файл с решениями от руки, то назовите его add.pdf
6. Весь код должен быть написан на Python.
7. Разрешается использовать без доказательства любые результаты, встречавшиеся на лекциях или семинарах по курсу, если получение этих результатов не является вопросом задания.
8. Разрешается использовать любые свободные источники с указанием ссылки на них.
9. Плагиат не допускается. При обнаружении случаев списывания 0 за работу выставляется всем участникам нарушения, даже если можно установить, кто у кого списал.

## Задача 1. Всё нормально! (5 баллов)

Майло Тэтч исследует выборку независимых одинаково распределённых случайных величин  $X_1, \dots, X_n$  из нормального  $\mathcal{N}(\mu, 1)$  распределения.

- а) Симулируйте такую выборку при  $n = 100, \mu = 5$ .
- б) Научный опыт Майло говорит, что  $f(\mu)$  непременно равна  $1^1$ . Найдите апостериорную плотность  $f(\mu|X)$  аналитически и постройте её график.
- в) Майло не хочет хранить апостериорную функцию плотности в явном виде и желает представить апостериорное распределение в виде массива чисел. Выберите любой способ сделать это и получите такой массив (размер массива – на ваш вкус). Постройте гистограмму полученного массива и визуально оцените, насколько хорошо она приближает аналитическую апостериорную плотность.
- г) Майло считает, что недостаточно рассматривать только  $\mu$ : требуется изучить и  $\theta = e^\mu$ . Найдите апостериорную плотность  $f(\theta|X)$  аналитически и при помощи симуляций. Для симуляций используйте следующий алгоритм:
  - а) Симулируйте  $Q_1, \dots, Q_B$  из апостериорного распределения  $\mu|X$ .
  - б) Рассчитайте  $\gamma_i := e^{Q_i}$ . Объявите  $\gamma_1, \dots, \gamma_B$  выборкой независимых одинаково распределённых случайных величин из апостериорного распределения  $\theta|X$ .
- е) Постройте 95%-ый байесовский доверительный интервал для  $\mu$ .
- ф) Постройте 95%-ый частотный доверительный интервал для  $\mu$ .

(По мотивам: Wasserman, *All of Statistics*)

## Задача 2. Необычная флора (5 баллов)

Оказавшись в Атлантиде, Майло Тэтч изучает местную флору. В частности, ему интересно, как влияют привычные удобрения на рост местного вида папоротников. Майло собирает случайную выборку из 500 папоротников и разделяет её на пять непересекающихся групп по 100 наблюдений в каждой. К четырём из них он применяет удобрения разных видов, а пятая группа объявляется контрольной. Результаты эксперимента приведены в таблице ниже. В первой строке указаны виды применяемых удобрений, а во второй – число растений, для которых наблюдались улучшения в росте.

Вид удобрения	Без удобрения	DAP	MAP	NPK	UAN
Улучшения в росте	42	37	17	74	44

- а) Сформулируйте гипотезы о наличии (отсутствии) эффекта действия каждого удобрения. Проверьте каждую гипотезу на уровне значимости 5%. Можно считать, что число наблюдений достаточно для применения асимптотических результатов. Для получения оценок используйте метод максимального правдоподобия.
- б) Проведите процедуру множественного тестирования методом Бонферрони. Изменились ли результаты? Напоминание: при полученной статистике  $z_{obs}$  p-value можно рассчитать как  $2\mathbb{P}\{Z \leq z_{obs} | H_0\}$  для двустороннего теста в случае симметричного распределения.
- в) Проведите процедуру множественного тестирования методом Бенджамини-Хохберга. Изменились ли результаты?
- г) Сделайте итоговый вывод о наличии эффекта действия каждого удобрения.
- е) Результаты тестирования удобрения MAP смутили Майло: они никак не согласовывались с его научным опытом. Поэтому для исследования эффекта действия этого удобрения Майло решил использовать байесовский подход. Пусть теперь  $p_{MAP}$  – случайная величина, и Майло уверен, что  $p_{MAP} \sim \text{Beta}(\alpha, \beta)$ . Найдите какую-нибудь (на ваш выбор) точечную байесовскую оценку параметра  $p_{MAP}$ . Прокомментируйте, насколько сильно различаются частотная и байесовская оценки и поясните почему.

<sup>1</sup>Кажется, что это весьма странное утверждение. Что именно оно означает, мы обсудим на последнем семинаре.