

Во всех задачах можно использовать python или любой другой язык программирования с открытыми исходниками. Вполне могут возникать уравнения или оптимизационные задачи, которые решаются только численно. Если при решении задачи использовался код, то его нужно привести. Можно использовать chatgpt с указанием полного диалога в решении. Решение нужно прислать в виде одного pdf файла. Про бутстрэп и перестановочные тесты можно и нужно почитать <https://arxiv.org/abs/1411.5279>. При симуляциях всегда добивайтесь воспроизводимости, для питона стоит глянуть <https://builtin.com/data-science/numpy-random-seed>. Для задач 4, 5 и 6 потребуются данные по результатам экзамена по теории вероятностей этого года по 30-балльной шкале. Их можно найти на вики-страничке курса, http://wiki.cs.hse.ru/Econ_probability_2022-23.

Жесткий дедлайн: 21:00 12 июня, при сдаче работы до 21:00 11 июня начисляется бонус +10%.

1. Однажды в Самарканде турист заказывал Яндекс-такси. На десятом заказе впервые приехал таксист, который уже раньше приезжал к туристу. Для упрощения предположим, что все n таксистов Самарканда всегда на работе и приезжают равновероятно.
 - а) [5] Постройте график функции правдоподобия как функции от общего количества такси n . Найдите оценку числа n методом максимального правдоподобия.
 - б) [5] Постройте график математического ожидания номера заказа, на котором происходит первый повторный приезд, как функции от общего количества такси n . Найдите оценку числа n методом моментов.
 - в) [15] Предположим, что настоящее n равно 100. Проведя 10000 симуляций вызовов такси до первого повторного, рассчитайте 10000 оценок методом моментов и 10000 оценок методом максимального правдоподобия. Постройте гистограммы для оценок двух методов. Оцените смещение, дисперсию и среднеквадратичную ошибку двух методов.
2. Однажды в Самарканде турист заказывал Яндекс-такси. На десятом заказе он обнаружил, что у таксистов было 6 разных имён. Для упрощения предположим, что все n имён среди таксистов встречаются равновероятно и независимо от поездки к поездке.
 - а) [5] Постройте график функции правдоподобия как функции от общего количества имён n . Найдите оценку числа n методом максимального правдоподобия.
 - б) [5] Постройте график математического ожидания числа разных имён у 10 таксистов, как функции от общего количества имён n . Найдите оценку числа n методом моментов.
 - в) [15] Предположим, что настоящее n равно 20. Проведя 10000 симуляций десяти вызовов такси, рассчитайте 10000 оценок методом моментов и 10000 оценок методом максимального правдоподобия. Постройте гистограммы для оценок двух методов. Оцените смещение, дисперсию и среднеквадратичную ошибку двух методов.
Update 2023-06-07: если по выборке в симуляциях оценка метода моментов или метода максимального правдоподобия стремится к бесконечности и, строго говоря, не существует, то можно ограничить её сверху большим числом, например, 100.
3. Иноагент Иннокентий по 20 наблюдениям строит 95%-й доверительный интервал для математического ожидания несколькими способами: классический асимптотический нормальный интервал, с помощью наивного бутстрэпа, с помощью бутстрэпа t -статистики.
 - а) [15] Для каждого способа с помощью 10000 симуляций оцените вероятность того, что номинально 95%-й доверительный интервал фактически покрывает математическое ожидание, если наблюдения распределены экспоненциально с интенсивностью 1.

- б) [5] Пересчитайте вероятности накрытия, если наблюдения имеют распределение Стюдента с тремя степенями свободы.
- в) [5] Какой способ оказался лучше?

4. Проверьте гипотезу о том, что ожидаемые результаты экзамена по теории вероятностей тех, у кого фамилия начинается с гласной буквы и с согласной буквы, равны. В качестве альтернативной гипотезы возьмите гипотезу о неравенстве.

- а) [5] Используйте тест Уэлча.
- б) [5] Используйте наивный бутстрэп.
- в) [5] Используйте бутстрэп t -статистики.
- г) [5] Используйте перестановочный тест.

В каждом случае укажите P -значение и статистический вывод для уровня значимости 5%.

5. Составьте таблицу сопряжённости, поделив студентов писавших экзамен на четыре группы по двум признакам: набрал ли больше медианы или нет, на согласную или гласную букву начинается фамилия.

- а) [5] Постройте 95% асимптотический интервал для отношения шансов хорошо написать экзамен («несогласных» к «согласным»). Проверьте гипотезу о том, что отношение шансов равно 1 и укажите P -значение.
- б) [5] Постройте 95% асимптотический интервал для отношения вероятностей хорошо написать экзамен. Проверьте гипотезу о том, что отношение вероятностей равно 1 и укажите P -значение.
- в) [5] Постройте 95% интервал для отношения шансов хорошо написать экзамен с помощью наивного бутстрэпа. Проверьте гипотезу о том, что отношение шансов равно 1 и укажите P -значение.

В качестве альтернативной гипотезы используйте гипотезу о неравенстве.

6. Иноагент Иннокентий Вероятностно-Статистический считает, что длина фамилии положительно влияет на результат экзамена по теории вероятностей. А именно, он предполагает, что ожидаемый результат за экзамен прямо пропорционален длине фамилии, $\mathbb{E}(Y_i) = \beta F_i$, где Y_i — результат за экзамен по 30-балльной шкале, F_i — количество букв в фамилии.

- а) [10] Оцените β методом моментов. Рассчитайте выборочную корреляцию.
- б) [5] С помощью перестановочного теста найдите P -значение и формально протестируйте гипотезу о том, что корреляция равна нулю.

7. [10] С помощью chatgpt решите любую задачу из нашего курса теории вероятностей и статистики. Можно брать задачи из прошлых контрольных, лекций, семинаров и даже этого домашнего задания. В качестве ответа приведите полный диалог с chatgpt.

Простой диалог в виде двух реплик условия и ответа chatgpt даёт 6 баллов. Сложный диалог с наводками, указанием chatgpt на ошибки и их исправлением — 10 баллов.

Пример инструкции как заператься, <https://journal.tinkoff.ru/chatgpt-in-russia/>.

8. [5] Укажите любой источник по теории вероятностей или статистике, который вам оказался полезен в течение года. Это может быть статья, видео, задача, всё что угодно. Объясните, с чем конкретно этот источник помог разобраться. Лучше привести в пример внешний источник, не упомянутый на вики курса, но можно и внутренний.