

ВШБ Бизнес-информатика: ТВиМС 2025.
Экзаменационный вариант 1

1. (18 баллов) Рассмотрим две случайные величины X и Y . Они обе принимают значения 0, 1 и 2. Совместные вероятности для каждой пары заданы следующей таблицей, где $\theta \in \mathbb{R}$ — параметр:

| | $X = 0$ | $X = 1$ | $X = 2$ |
|---------|--------------------------|---------------------|----------------------|
| $Y = 0$ | $1 - \frac{8\theta}{10}$ | 0 | $\frac{\theta}{10}$ |
| $Y = 1$ | $\frac{\theta}{10}$ | $\frac{\theta}{10}$ | $\frac{2\theta}{10}$ |
| $Y = 2$ | $\frac{\theta}{10}$ | $\frac{\theta}{10}$ | $\frac{\theta}{10}$ |

- (a) (2 балла) Какой диапазон значений может принимать параметр θ ?

В дальнейших пунктах предполагается, что θ находится в диапазоне, найденном в пункте 1, однако вы должны проводить все вычисления для произвольного θ .

- (b) (1 балл) Вычислите

$$P(X = 2 | X + Y = 3).$$

- (c) (2 балла) Постройте таблицу вероятностей условного распределения X при условии $Y = 1$.

- (d) (2 балла) Вычислите $\text{Corr}(X, Y)$.

- (e) (5 баллов) Предположим, что у вас есть реализация случайной выборки: $\mathcal{X} = (x_1, \dots, x_n)$, где каждая x_i получена из закона распределения случайной величины X из таблицы выше.

Найдите оценку параметра θ - функцию от выборки $\hat{\theta}_{MM} = \hat{\theta}_{MM}(\mathcal{X})$ - методом моментов.

- (f) (6 баллов) Рассмотрите $\hat{\theta}_1 = Y$ и $\hat{\theta}_2 = \frac{X+Y}{2}$ как оценки для неизвестного параметра θ . Какую из них вы предпочтёте и почему? (Посмотрите свойства этих точечных оценок).

2. (8 баллов) В ресторане "Вкусно и вопросительный знак" поток клиентов моделируется Пуассоновским процессом. Известно, что вероятность, что в определенное время суток за час в ресторан придет хотя бы 1 клиент составляет 0.98.

Какова вероятность, что в случайный момент в течение этого времени суток ждать следующего вошедшего клиента мы будем меньше 5 минут?

3. (14 баллов) Предположим, что у нас есть реализация случайной выборки $\mathcal{X} = (x_1, \dots, x_n)$ неизвестной случайной величины X с плотностью

$$f_X(x; \theta) = \begin{cases} (\theta + 1)x^\theta, & x \in [0; 1]; \\ 0, & x \notin [0; 1]. \end{cases}$$

Реализация, которая была получена: $(x_1, \dots, x_6) = (0.1, 0.25, 0.32, 0.15, 0.72, 0.45)$.

(а) (7 баллов) Найдите оценку параметра θ - функцию от выборки $\hat{\theta}_{ML} = \hat{\theta}_{ML}(\mathcal{X})$ - методом максимального правдоподобия.

Посчитайте её реализацию на предоставленных данных.

(б) (7 баллов) Найдите оценку параметра θ - функцию от выборки $\hat{\theta}_{MM} = \hat{\theta}_{MM}(\mathcal{X})$ - методом моментов.

Посчитайте её реализацию на предоставленных данных.

4. (6 баллов) Расход топлива двух различных моделей автомобилей (А и В) сравнивался в эксперименте. Была взята случайная выборка из 20 автомобилей модели А и 35 автомобилей модели В, и для каждого автомобиля была измерена эффективность использования топлива (в километрах на литр). Результаты обобщены в таблице ниже.

| | Размер выборки | Выборочное среднее | Выборочное стандартное отклонение |
|----------|----------------|--------------------|-----------------------------------|
| Модель А | 25 | 30.9 | 6.11 |
| Модель В | 35 | 27.1 | 6.41 |

- (а) (2 балла) Посчитайте 98% доверительный интервал для математического ожидания эффективности использования топлива автомобилей модели А.
- (б) (4 балла) Используйте соответствующий тест гипотез на любых двух уровнях значимости из (1%, 2%, 5%), чтобы определить, могут ли автомобили модели А проехать больше километров на литр, чем автомобили модели В. Сформулируйте гипотезы и ваши предположения касательно свойств и характеристик случайных величин, которые вы исследуете. Укажите используемую статистику и её распределение при нулевой гипотезе. Оформите ваши результаты и сделайте выводы.

5. (12 баллов) Кредитные риски при выдаче кредита в компании "ВопросБанк" моделируются распределением Лапласа с параметрами $\alpha = 0.05$ и $\beta = 0.2$, которое имеет следующую функцию плотности:

$$f(x) = \frac{\alpha}{2} e^{-\alpha|x-\beta|}, \quad -\infty < x < +\infty,$$

где $\alpha > 0$ - параметр масштаба, $-\infty < \beta < +\infty$ - параметр сдвига.

Начальный момент k -го порядка для распределения Лапласа может быть рассчитан по следующей формуле:

$$\mathbb{E}[X^k] = \int_{-\infty}^{+\infty} x^k f(x) dx = \sum_{i=0}^{\lfloor k/2 \rfloor} \frac{\beta^{k-2i}}{\alpha^{2i}} \frac{k!}{(k-2i)!},$$

где $\lfloor k/2 \rfloor$ - целая часть $k/2$.

Каждый день ВопросБанк обрабатывает 1000 независимых заявок на получение кредита. Какова вероятность, что средний дневной риск по всем клиентам за день опустится ниже отметки -0.1 от 120 до 140 раз за год?

6. (10 баллов) Телефонная компания собрала случайную выборку из 500 человек, чтобы определить, нравится ли им дизайн их последнего мобильного телефона. Таблица ниже обобщает ответы людей.

| | Размер выборки | Положительное мнение о дизайне |
|---------|----------------|--------------------------------|
| Мужчины | 225 | 115 |
| Женщины | 275 | 165 |

- (a) (2 балла) Посчитайте 98% доверительный интервал для доли женщин, которым понравился дизайн.
- (b) (2 балла) Посчитайте 94% доверительный интервал для истинной разности долей мужчин и женщин, которым понравился дизайн.
- (c) (5 баллов) Проведите двусторонний тест на уровне значимости 6%, чтобы проверить гипотезу о том, что доля женщин, которым понравился дизайн, равна доле мужчин, которым понравился дизайн. Сформулируйте гипотезы, укажите используемую статистику и её распределение при нулевой гипотезе. Проведите тестирование через: **score** (критерий) и **p**-значение.
- (d) (1 балл) Оформите ваши результаты, сравните результаты пункта (c) с доверительным интервалом из пункта (b) и сделайте выводы.

7. (14 баллов) На дисциплине «Теория вероятностей» 40% студентов списывают. Поэтому преподаватели помимо письменной части экзамена ввели ещё и обязательную устную защиту для всех студентов. Известно, что студенты, которые списывали на письменной части экзамена, на устной защите на каждый вопрос по решённой ими задаче независимо отвечают с вероятностью 0.6. Студенты, которые решали письменный экзамен самостоятельно, на каждый вопрос по своей работе независимо отвечают с вероятностью 0.9. Студентам на устной защите задаётся 10 вопросов.

Какой максимальный порог отсечения K нужно ввести (ответил хотя бы на K вопросов — защищился; не ответил хотя бы на K вопросов — обнуление), чтобы при количестве ответов меньше K вероятность того, что студент списал, была бы не ниже 90%?

8. (21 балл) Предположим, что у нас есть реализации случайных выборок: $\mathcal{X} = \{x_1, \dots, x_n\}$ и $\mathcal{Y} = \{y_1, \dots, y_m\}$, случайных величин $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ и $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ соответственно. В следующих пунктах исследуются различные аспекты этих выборок.

- (а) (3 балла) Пусть \bar{x} — реализация выборочного среднего на выборке \mathcal{X} . Найдите n такое, что интервал:

$$(\bar{x} - 0.775 \sigma_x, \bar{x} + 0.775 \sigma_x)$$

является приблизительно 98% доверительным интервалом для μ_x .

- (б) (5 баллов) Мы привыкли строить доверительные интервалы для математического ожидания вокруг выборочного среднего. Но ничто не мешает нам забросить эту "рыболовную сеть" вокруг одной реализации x случайной величины X в попытке поймать математическое ожидание μ_x . Каким будет доверительный уровень интервала такой же ширины, как в предыдущем пункте, но построенного на основе всего лишь одной реализации x ?

- (с) (6 баллов) Пусть \bar{X} и \bar{Y} — выборочные средние двух независимых случайных выборок случайных величин X и Y , каждая размера n ($m = n$), где истинные дисперсии известны $\sigma_x^2 = \sigma^2$ и $\sigma_y^2 = 2\sigma^2$ соответственно. Найдите n такое, что:

$$P\left(\bar{X} + \bar{Y} - \frac{\sigma}{5} < \mu_x + \mu_y < \bar{X} + \bar{Y} + \frac{\sigma}{5}\right) = 0.80.$$

- (д) (7 баллов) Мы хотим проверить гипотезу о том, что $\mu_x = \mu_y + \Delta$ против двусторонней альтернативной гипотезы. Предположим, что известны данные: $n = 5$, $m = 9$, $\bar{x} = 7$, $\bar{y} = 4.5$, $\Delta = 1$, $\sigma_x = \sigma_y = 3$. Используйте данные и проведите тест, используя уровни значимости $\alpha = 1\%$, 5% .

Для полного оценивания этого пункта недостаточно просто сказать, отклоняем ли мы гипотезу или нет. Вам нужно как-то обосновать ваши заключения: укажите используемую статистику и её распределение при нулевой гипотезе, процесс принятия решения (что с чем сравниваем).