

## Домашнее задание 1

Дедлайн: 2024-09-24, 21:00.

1. Вася решает три задачи по теории вероятностей. Вероятности решить каждую задачу по отдельности равны 0.1, 0.2 и 0.5. Решения задач никак не связаны между собой, знание ни одной из задач не помогает решить ни одну другую. Обозначим буквой  $N$  общее количество решенных задач.

- а) Найдите все значения  $N$  и их вероятности.
- б) Найдите  $\mathbb{P}(N \leq 2)$ ,  $\mathbb{E}(N)$  и  $\mathbb{E}(N^2)$ .

2. За работу Вася получает случайное целое количество  $\xi$  баллов, равновероятно распределённое от 1 до  $n$ .

Найдите  $\mathbb{E}(\xi)$ ,  $\mathbb{E}(\xi^2)$ ,  $\mathbb{E}(\xi^3)$ .

3. Берём набор данных по ссылке

[https://github.com/bdemeshev/hse\\_knad\\_probability\\_2024\\_2025/raw/main/home\\_assignments/ha01\\_data.csv](https://github.com/bdemeshev/hse_knad_probability_2024_2025/raw/main/home_assignments/ha01_data.csv).

Здесь две переменных:  $y_i$  — количество просмотренных Машей рилзов в день  $i$  и бинарная переменная  $x_i$  ( $x_i = A$  — обычный день,  $x_i = B$  — день дедлайна по теории вероятностей).

Рассмотрим две гипотезы. Нулевая гипотеза  $H_0$ : приближение дедлайна по вероятностям никак не влияет на количество просмотренных рилзов. Альтернативная гипотеза  $H_1$ : приближение дедлайна в среднем снижает количество просмотренных рилзов.

- а) Посчитайте фактическое значение статистики  $S = \bar{y}_B - \bar{y}_A$ .
- б) Предполагая, что  $H_0$  верна, сгенерируйте 10000 случайных перестановок меток  $x$  и для каждой перестановки посчитайте значение статистики  $S^{\text{new}} = \bar{y}_B^{\text{new}} - \bar{y}_A^{\text{new}}$ .
- в) Оцените  $p$ -значение, в данном случае  $p$ -значение — это вероятность  $\mathbb{P}(S^{\text{new}} \leq S \mid S, H_0)$ .
- г) Для принятия решения, отвергать или нет  $H_0$ , мы используем уровень значимости  $\alpha = 0.05$ . Отвергаем ли мы  $H_0$ ?

## Домашнее задание 2

Дедлайн: 2024-10-01, 21:00.

1. Монетка выпадает орлом  $T$  с вероятностью 0.2 и решкой  $H$  — с вероятностью 0.8. Илон Маск подбрасывает её 100 раз. За каждую выпавшую комбинацию  $THT$  он получает 1\$, а за каждую комбинацию  $HHHHHH$  — платит 1\$.

Чему равен ожидаемый выигрыш Маска в эту игру?

Уточнение: комбинации могут пересекаться, например, за  $THTHT$  Маск получит 2\$.

2. Бармен Огненной Зебры разбавляет каждую кружку пива независимо от других с общеизвестной вероятностью  $p \in (0; 1)$ . Ковбой Джо заходит в бар и первым делом сразу заказывает три кружки пива и выпивает их. Затем Джо заказывает по две кружки пива за один раз.

После 3-й, 5-й, 7-й, 9-й, 11-й и далее через каждые две кружки Джо прислушивается к своим ощущениям. Если не менее двух кружек пива из последних трёх кружек разбавлены, то Джо разносит бар к чертям собачьим.

- а) Сколько кружек пива в среднем успеет выпить Джо прежде чем разнесёт Огненную Зебру?
  - б) Если все три последние кружки пива разбавлены, то Джо разносит не только Огненную Зебру, но и всю прилежащую улицу. Какова вероятность данного сценария?
3. Камала Харрис подбрасывает кубик до первого выпадения восьмёрки. Все грани кубика выпадают равновероятно, однако на его шести гранях написаны числа 3, 4, 5, 6, 7, 8. Дональд Трамп подбрасывает правильный октаэдр до выпадения восьмёрки. На гранях октаэдра написаны числа от 1, 2, 3, 4, 5, 6, 7, 8.
- а) Постройте гистограмму числа бросков кубика по  $B = 10000$  экспериментов.
  - б) Оцените безусловное математическое ожидание числа бросков кубика.
  - в) Оцените безусловную вероятность окончания игры быстрее, чем за 5 бросков кубика.
  - г) Постройте условную гистограмму числа бросков октаэдра, если известно что грани 1 и 2 не выпадали. Общее количество экспериментов здесь должно быть таким, чтобы число экспериментов, где не выпадали грани 1 и 2 оказалось равным  $B = 10000$ .
  - д) Оцените условное ожидание числа бросков октаэдра, если грани 1 и 2 не выпадали.
  - е) Оцените условную вероятность окончания игры быстрее, чем за 5 бросков октаэдра, если грани 1 и 2 не выпадали.

### Домашнее задание 3

Дедлайн: 2024-10-08, 21:00.

1. В анкету включён вопрос, на который респонденты стесняются отвечать правдиво. Например, «Берёте ли Вы взятки?» или «Употребляете ли Вы наркотики?» Чтобы стимулировать респондентов отвечать правдиво, используют следующий прием. Перед ответом на вопрос респондент в тайне от анкетирующего подкидывает один раз специальную монетку, на гранях которой написано «Да = А, Нет = Б», и «Да = Б, Нет = А». Ответ «Да» на нескромный вопрос является верным для доли  $p$  всех людей. Монетка неправильная и выпадает стороной «Да = А, Нет = Б» с вероятностью 0.6.
  - а) Какова вероятность того, что ответ «Да» для данного индивида верен, если он написал «А» и следовал указаниям монетки?
  - б) Какова вероятность того, что ответ «Да» для данного индивида верен, если он подбрасывал специальную монету 3 раза, следовал каждый раз предлагаемой кодировке и написал «А», «Б», «А»?
2. Илон Маск изобрёл новый кубик под названием Model-6. Он взял правильный игральный кубик и правильный кубик с неподписанными гранями. Он подкинул шесть раз правильный игральный кубик и заполнил по очереди все грани изначально чистого кубика результатами бросков правильного кубика.

- а) Какова вероятность того, что в первом броске Model-6 выпало 6, если во втором броске Model-6 выпало 6?
  - б) Зависимы ли результаты бросков Model-6?
  - в) Чему равно ожидаемое количество шестёрок, выпавших в процессе изготовления Model-6, если при шести бросках Model-6 выпало три шестёрки?
3. Алиса и Боб снова подкидывают монетку неограниченное число раз. Монетка выпадает решкой  $H$  и орлом  $T$  равновероятно. Алиса выигрывает, если последовательность  $HNT$  выпадет раньше, а Боб — если раньше выпадет  $HTH$ .

Рассмотрим множество исходов этого эксперимента  $\Omega = \{HNT, HTH, HHNT, THNT, THTH, \dots\}$  и производящую функцию исходов  $f(H, T) = HNT + HTH + HHNT + THNT + THTH + \dots$ . Здесь аргументы  $H$  и  $T$  некоммутативны. Обозначим  $X$  — количество решек  $H$ ,  $Y$  — количество орлов  $T$ .

- а) Укажите, как с помощью производных и подстановок раздобыть из функции  $f(H, T)$  величины  $\mathbb{P}(X = 10)$ ,  $\mathbb{P}(X = 5, Y = 5)$ ,  $\mathbb{E}(X)$ ,  $\mathbb{E}(X^3)$ ,  $\mathbb{E}(X^2Y^3)$ .
- б) С помощью метода первого шага составьте систему линейных уравнений, из которой можно найти  $f(H, T)$ .
- в) Решите эту систему, предполагая коммутативность  $H$  и  $T$ .
- г) Завершите вычисление  $\mathbb{P}(X = 10)$ ,  $\mathbb{P}(X = 5, Y = 5)$ ,  $\mathbb{E}(X)$ ,  $\mathbb{E}(X^3)$ ,  $\mathbb{E}(X^2Y^3)$ .

Явное уточнение: конечно, в этой задаче можно использовать `sympy` или другой пакет для символического решения системы или вычисления производных.

## Домашнее задание 4

У этого задания нет дедлайна и за него нет оценки. Если очень хочется что-то куда-то загрузить, то можно отправить своему семинаристу мемасик по теории вероятностей :)

1. Случайная величина  $X$  принимает значения 1, 2, 3 и 4 с вероятностями 0.1, 0.2, 0.3, 0.4.
  - а) Нарисуйте функцию распределения величины  $X$ ,  $F_X(x)$ .
  - б) Какой вероятностный смысл имеет площадь над функцией распределения  $F_X(x)$  на участке  $x \in [0; \infty)$ ?
  - в) Нарисуйте функцию распределения случайной величины  $Y = F_X(X)$ .
2. Функция плотности случайной величины  $Y$  равна  $cy^2$  на отрезке  $[0; 2]$  и нулю иначе.
  - а) Найдите константу  $c$ .
  - б) Найдите функцию распределения  $Y$ .
  - в) Найдите  $\mathbb{P}(Y > 1)$ ,  $\mathbb{P}(Y = 0.75)$ ,  $\mathbb{E}(Y)$ ,  $\mathbb{E}(Y^2)$ .
  - г) Найдите функцию производящую моменты  $Y$ ,  $m_Y(t)$ .
  - д) Найдите  $\mathbb{P}(Y > 1.5 \mid Y > 1)$ ,  $\mathbb{E}(Y \mid Y > 1)$ ,  $\mathbb{E}(Y^2 \mid Y > 1)$ .
  - е) Найдите функцию плотности величины  $W = 1/Y$ .

3. Случайная величина  $U$  равномерна на отрезке  $[0; 10]$ ,  $Y = \min\{U^2, 25\}$ .
- а) Запишите вероятность  $\mathbb{P}(Y \in [y; y + \Delta])$  с точностью до  $o(\Delta)$ .
  - б) Найдите функцию распределения  $Y$ .
  - в) Найдите  $\mathbb{P}(Y > 10)$ ,  $\mathbb{E}(Y)$ ,  $\mathbb{E}(Y^2)$ .
  - г) Найдите  $\mathbb{P}(Y > 10 \mid Y > 5)$ ,  $\mathbb{E}(Y \mid Y > 5)$ ,  $\mathbb{E}(Y^2 \mid Y > 5)$ .

## Домашнее задание 5

Дедлайн: 2024-10-25, 23:59.

1. Случайная величина  $X$  распределена равномерно на  $[0; 10]$ ,  $Y = X^2$ .
- а) Найдите дисперсию  $\text{Var}(Y)$ , стандартное отклонение  $\sigma_Y$ .
  - б) Найдите  $\text{Cov}(X, Y)$ ,  $\text{Corr}(X, Y)$ .
  - в) Найдите  $\text{Cov}(6X + 2Y + 7, -2Y + 15)$ ,  $\text{Corr}(5 - 6X, 8 + 9Y)$ ,  $\text{Var}(2Y + 7)$ .
  - г) Предложите любую неслучайную функцию  $h$  такую, что  $\text{Corr}(h(X), X) = 0$ ,  $\text{Var}(h(X)) > 0$ .
2. Назовём наилучшей линейной аппроксимацией величины  $Y$  с помощью величины  $X$  функцию вида  $\hat{Y} = \alpha + \beta X$ , где  $\alpha$  и  $\beta$  — константы, при которых величина  $\mathbb{E}((Y - \hat{Y})^2)$  минимальна.
- а) Известно, что  $\text{Cov}(X, Y) = 10$ ,  $\text{Var}(X) = 40$ , найдите  $\beta$ .
  - б) Дополнительно известно, что  $\mathbb{E}(Y) = 10$ ,  $\mathbb{E}(X) = 80$ , найдите  $\alpha$ .
- Допустим, что  $a + bR$  — наилучшая линейная аппроксимация  $L$  с помощью  $R$ , а  $c + dL$  — наилучшая линейная аппроксимация  $R$  с помощью  $L$ .
- в) Выразите произведение  $bd$  через корреляцию  $\text{Corr}(R, L)$ .
3. В анализе временных рядов иногда используют концепцию частной корреляции. Частная корреляция между величинами  $X$  и  $Y$ , очищенными от связи с величиной  $W$ , равна обычной корреляции между величинами  $X^* = X - \alpha W$  и  $Y^* = Y - \beta W$ , где константы  $\alpha$  и  $\beta$  находятся из условия некоррелированности  $X^*$  с  $W$  и некоррелированности  $Y^*$  с  $W$ .

$$\text{pCorr}(X, Y; W) = \text{Corr}(X^*, Y^*), \text{ где } \begin{cases} X^* = X - \alpha W, & \text{Cov}(X^*, W) = 0, \\ Y^* = Y - \beta W, & \text{Cov}(Y^*, W) = 0. \end{cases}$$

Величины  $Y_1, Y_2, Y_3$  независимы и равномерны на отрезке  $[0; 1]$ ,  $S_3 = Y_1 + Y_2 + Y_3$ .

Найдите  $\text{Corr}(Y_1, Y_2)$  и  $\text{pCorr}(Y_1, Y_2; S_3)$ .

## Домашнее задание 6

Дедлайн: 2024-11-01, 23:59.

1. Распределение вектора  $(X, Y)$  задано таблицей

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 0$	0.2	0.2	0.1
$X = 1$	0.5	0	0

- а) Найдите энтропии  $\mathbb{H}(X)$ ,  $\mathbb{H}(Y)$ ,  $\mathbb{H}(X, Y)$ .
  - б) Найдите  $\mathbb{H}(Y | X)$ .
  - в) Какое максимальное значение может принимать условная энтропия  $\mathbb{H}(Y | X)$ , если  $X$  принимает два значения, а  $Y$  — три?
2. Для дискретных величин  $X$  и  $Y$  докажите или опровергните утверждения:
- а)  $\mathbb{H}(X) + \mathbb{H}(Y | X) = \mathbb{H}(X, Y)$ ;
  - б)  $\mathbb{H}(X, Y) \geq \mathbb{H}(X)$ ;
  - в)  $\mathbb{H}(X^2) = \mathbb{H}(X)$ ;
3. Время до прихода автобуса на остановку — неотрицательная случайная величина  $X$  с функцией плотности. Андрей верит, что функция плотности  $X$  имеет вид

$$a(x) = \begin{cases} \alpha \exp(-\alpha x), & \text{если } x \geq 0 \\ 0, & \text{иначе.} \end{cases}$$

Борис верит, что функция плотности  $X$  имеет вид

$$b(x) = \begin{cases} \beta \exp(-\beta x), & \text{если } x \geq 0 \\ 0, & \text{иначе.} \end{cases}$$

Вова верит, что у  $X$  есть какая-то функция плотности  $c(x)$ , а математическое ожидание такое, как думает Андрей.

- а) Найдите математическое ожидание с точки зрения Андрея и Вовы,  $\mathbb{E}_a(X)$ .
- б) Найдите энтропию  $\mathbb{H}(a)$ .
- в) Найдите кросс-энтропию  $\text{CE}(b||a)$ . При каком  $\alpha$  она минимальна?
- г) Найдите кросс-энтропию  $\text{CE}(c||a)$ .
- д) Чему равно максимальное значение энтропии  $\mathbb{H}(c)$  и при какой функции плотности  $c(x)$  достигается максимум?