

«Теория вероятностей: Математическое ожидание и дисперсия»

Домашнее задание №5

Байдаков Илья

20 марта 2022 г.

Задание 1

Банк внедряет новую систему для анализа кредитной истории клиентов. Согласно статистике банка, наличие у заемщика просрочек по платежам в прошлом (с.в. X) и факт погашения кредита позже назначенного срока (с.в. Y) имеют распределение, заданное таблицей ниже.

(Т.е. случайная величина X принимает значение 0, если человек вносил платежи вовремя, и 1 иначе. Аналогично, Y принимает значение 0, если клиент вернул кредит вовремя, и 1 в противном случае.)

В зависимости от значения корреляции этих величин банк примет решение о том, как учитывать наличие просрочек по платежам для выдачи кредита.

Найдите коэффициент корреляции случайных величин X и Y .

$X \backslash Y$	0	1
0	0.6	0
1	0.1	0.3

Решение.

По определению,

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}X}\sqrt{\text{Var}Y}} = \frac{0.3 - 0.4 \cdot 0.3}{\sqrt{0.4 - 0.16}\sqrt{0.3 - 0.09}} \simeq 0.802$$

Ответ. $\rho_{X,Y} \simeq 0.802$.

Задание 2

Независимые случайные величины $X, Y \sim U(-1, 1)$. Рассмотрим $Z = 2X + Y$.

- а) Найдите $\text{Cov}(Z, Y)$.
- б) Чему равняется $\rho_{Z,Y}$?

Решение.

- а) По определению,

$$\mathbf{Cov}(Z, Y) = \mathbb{E}(ZY) - \mathbb{E}(Z) \cdot \mathbb{E}(Y) \quad (*)$$

Т.к. имеем нормальное распределение, то

$$\mathbb{E}(X) = \mathbb{E}(Y) = \frac{a+b}{2} = 0$$

$$\mathbb{E}(Z) = \mathbb{E}(2X + Y) = 0$$

Из учебника Черновой,

$$\mathbb{E}(Y^2) = \frac{a^2 + ab + b^2}{3} = \frac{1}{3}$$

$$\mathbb{E}(ZY) = \mathbb{E}[(2X + Y)Y] = 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y^2) = \frac{1}{3}$$

Подставляем в (*):

$$\mathbf{Cov}(Z, Y) = \frac{1}{3} - 0 \cdot 0 = \frac{1}{3}$$

б) вычислим дисперсии для нормального распределения. По формуле из учебника Черновой

$$\mathbf{Var}(X) = \mathbf{Var}(Y) = \frac{(b-a)^2}{12} = \frac{1}{3}$$

$$\mathbf{Var}(Z) = \mathbf{Var}(2X + Y) = 4\mathbf{Var}(X) + \mathbf{Var}(Y) = 4 \cdot \frac{1}{3} + \frac{1}{3} = \frac{5}{3}$$

По определению,

$$\rho_{Z,Y} = \frac{\mathbf{Cov}(Z, Y)}{\sqrt{\mathbf{Var}Z}\sqrt{\mathbf{Var}Y}} = \frac{1/3}{\sqrt{5/3}\sqrt{1/3}} \simeq 0.447$$

Ответ. а) $\mathbf{Cov}(Z, Y) = \frac{1}{3}$. б) $\rho_{Z,Y} \simeq 0.447$.

Задание 3

Закон распределения случайного вектора (X, Y) задан таблицей.

$\begin{array}{c} \backslash Y \\ X \end{array}$	1	3
0	0.4	0
2	0	0.4
3	0.2	0

а) Какие значения принимает условное математическое ожидание $\mathbb{E}(X|Y)$? (Иными словами, требуется найти значения $\mathbb{E}(X|Y = 1)$ и $\mathbb{E}(X|Y = 3)$.)

б) С какой вероятностью условное математическое ожидание $\mathbb{E}(X|Y)$ принимает значение $\mathbb{E}(X|Y = 1)$?

в) Какая из функций ниже выражает условное математическое ожидание $\mathbb{E}(X|Y)$ через Y ? Ответ обоснуйте.

- 1) $\frac{Y}{2}$ 2) $\frac{1}{2} + \frac{Y}{2}$ 3) Y^2 4) $2Y - 1$.

Решение.

а) Найдём распределение $P_Y(y)$: $P_Y(Y = 1) = 0.6$, $P_Y(Y = 3) = 0.4$.

Далее, по определению,

$$\begin{aligned}\mathbb{E}(X|Y = 1) &= \sum_i x_i \cdot p_{X|Y}(x_i|y) = \sum_i x_i \cdot \frac{p_{X|Y}(x_i, y_0)}{P_Y(y_0)} = \\ &= 0 \cdot \frac{0.4}{0.6} + 2 \cdot \frac{0}{0.6} + 3 \cdot \frac{0.2}{0.6} = 1.\end{aligned}$$

Аналогично находим $\mathbb{E}(X|Y = 3) = 2$.

б) Рассматривая $\mathbb{E}(X|Y)$ как случайную величину, запишем её таблицу распределения. В ней будет два значения, каждое соответствует одному из двух возможных значений Y . Значит, вероятность, что $\mathbb{E}(X|Y)$ примет это значение, соответствует вероятности того, что Y примет соответствующее значение:

$\mathbb{E}(X Y)$	1	2
$P(\mathbb{E}(X Y))$	0.6	0.4

Значит, $\mathbb{E}(X|Y)$ примет значение $\mathbb{E}(X|Y = 1) = 1$ с вероятностью 0.6.

в) Правильный ответ (2), т.е. $f(Y) = \frac{1}{2} + \frac{Y}{2}$. Согласно расчётам в пункте (а), именно эта функция принимает верные значения $\mathbb{E}(X|Y)$ при $Y = 1$ и при $Y = 3$.

Ответ.

а) $\mathbb{E}(X|Y = 1) = 1$, $\mathbb{E}(X|Y = 3) = 2$.

б) $P = 0.6$.

в) (2).

Задание 4

Пусть с.в. X и Y независимы и имеют распределения $X \sim N(2, 4)$, $Y \sim \text{Exp}(2)$.

Чему равняется $\mathbb{E}[(X + Y)^2 \cdot \sin Y | Y]$?

Решение.

Пользуясь свойствами линейности, стабильности и вынесения множителя условного мат. ожидания, а также учитывая независимость у.в., и также что $\mathbb{E}X = 2$ и $\mathbb{E}(X^2|Y) = \mathbb{E}(X^2) = \mathbf{Var}X + (\mathbb{E}X)^2$, получаем:

$$\begin{aligned}\mathbb{E}[(X + Y)^2 \cdot \sin Y | Y] &= \sin Y \cdot \mathbb{E}[(X^2 + 2XY + Y^2) | Y] = \\ &= \sin Y \cdot (\mathbb{E}(X^2|Y) + 2Y\mathbb{E}(X|Y) + Y^2) = \sin Y \cdot (\mathbb{E}(X^2|Y) + 2Y\mathbb{E}(X) + Y^2) = \\ &= \sin Y \cdot (\mathbb{E}(X^2|Y) + 2Y\mathbb{E}(X) + Y^2) = \sin Y \cdot (4 + 2^2 + 4Y + Y^2) = \\ &= (Y^2 + 4Y + 8) \cdot \sin Y.\end{aligned}$$

Ответ. $\mathbb{E} = (Y^2 + 4Y + 8) \cdot \sin Y$.

Задание 5

Среднее время обработки запроса на некотором сервисе равняется 1 секунде.

а) Какова максимальная возможная вероятность того, что запрос будет обрабатываться не меньше 100 секунд? Приведите пример ситуации, в которой достигается эта вероятность.

б) Пусть теперь известно, что дисперсия времени обработки запроса равняется 1. Оцените с помощью неравенства Чебышева вероятность того, что запрос будет обрабатываться не меньше 100 секунд.

Решение.

а) Обозначая время запроса за X и считая среднее время запроса мат. ожиданием X , получаем согласно неравенству Маркова:

$$P(X \geq 100) = P(X \geq 100 \cdot \mathbb{E}X) \leq \frac{1}{100} \longrightarrow P_{max} = \frac{1}{100}.$$

Пример: 99 запросов обработались за пренебрежимо малое время, а следующий обрабатывался 100 секунд.

б) Запишем неравенство Чебышёва:

$$P(X \geq 100) = P(X - \mathbb{E}X \geq 99) \leq \frac{\mathbf{Var}X}{x^2} = \frac{1}{99^2} \simeq 10^{-4}.$$

Ответ. а) $P = 1/100$. б) $P \simeq 10^{-4}$.

Задание 6

Известно, что среднее время решения этой задачи составляет 40 минут. Причем вероятность того, что на решение уйдет не больше 30 минут, равна $\frac{1}{2}$. Среднее время решения этой задачи для тех, кто уложился в 30 минут, равняется 20 минутам.

Каково среднее время решение этой задачи для тех, кто решал ее дольше 30 минут?

Решение.

Обозначим время решения задачи X .

Тогда $\mathbb{E}X = 40$, $P(X \leq 30) = \frac{1}{2}$, $P(X > 30) = \frac{1}{2}$ и $\mathbb{E}(X|X \leq 30) = 20$. Запишем формулу полной вероятности для условного математического ожидания $\mathbb{E}X$:

$$\mathbb{E}X = \mathbb{E}(X|X \leq 30) \cdot P(X \leq 30) + \mathbb{E}(X|X > 30) \cdot P(X > 30),$$

и выразим отсюда искомую величину

$$\mathbb{E}(X|X > 30) = (40 - 10) \cdot 2 = 60 \text{ мин.}$$

Ответ. 60 мин.

Задание 7

На 50 сайтов одновременно совершают кибер-атаку 20 хакеров. Каждый из них выбирает одну цель случайно и независимо от других, но всегда взламывает защиту.

Сколько сайтов в среднем останется не взломано?

Решение.

Обозначим случайную величину

X = количество не взломанных сайтов.

Введём индикатор

$$I_i = \text{"}i\text{-й сайт взломан"}.$$

Пусть $I_i = 1$, если сайт взломан. Это событие случится если хотя бы один (допустим, первый) хакер из 20 выберет i -й сайт для взлома. Событие, при котором один хакер (с номером от 1 до 20) взламывает i -й сайт, обозначим:

$$H_{i,1} = H_{i,2} = \dots = H_{i,20},$$

при этом

$$P(H_{i,1}) = P(H_{i,2}) = \dots = P(H_{i,20}) = \frac{1}{50}.$$

Однако i -й сайт может одновременно выбрать $k \in [0, 20]$ хакеров, и любое из событий $H_{i,k}$ приведёт к событию «сайт взломан», т.е.

$$P(I_i = 1) = H_{i,1} \cup H_{i,2} \cup \dots \cup H_{i,20}.$$

Поскольку события под знаком суммы независимы, воспользуемся теоремой о сумме вероятностей для подсчёта $P(I_i = 1)$. Воспользуемся укороченной формулой для случая, когда случайные величины под знаком суммы независимы и одинаково распределены (это удовлетворяет нашему тривиальному случаю, где вероятности событий под знаком суммы равны)

(en.wikipedia.org/wiki/Inclusion-exclusion_principle#In_probability):

$$P(I_i = 1) = P\left(\bigcup_{k=1}^{k=20} H_{i,k}\right) = 1 - (1 - H_{i,k})^k = 1 - \left(1 - \frac{1}{50}\right)^{20} \simeq 0.332.$$

Далее, запишем

$$P(I_i = 1) = \mathbb{E}I_i \simeq 0.332,$$

что можно интерпретировать как "каждый сайт в среднем взломали 0,332 раза".

Тогда

$$\mathbb{E}X = \mathbb{E}(50 - \sum_i \mathbb{E}I_i) = 50 - 50 \cdot 0.332 \simeq 33.380.$$

Ответ. В среднем останется не взломано $\simeq 33.380$ сайтов.