



Федеральное государственное автономное образовательное учреждение высшего образования «Национальный Исследовательский Университет ИТМО»

ДОМАШНЕЕ ЗАДАНИЕ №2
ПРЕДМЕТ «ТЕОРИЯ ВЕРОЯТНОСТЕЙ»
Вариант №1

Преподаватель:
Шиманская Г. С.
Студент:
Румянцев А. А.

Номер ИСУ:
368731
Поток:
ТеорВер 1.2
Факультет:
СУиР
Группа:
R3241

Санкт-Петербург
2024

Содержание

1	Задание 1.	2
1.1	Условие задачи.	2
1.2	Решение.	2
2	Задание 2.	4
2.1	Условие задачи.	4
2.2	Решение.	5
3	Задание 3.	6
3.1	Условие задачи.	6
3.2	Решение.	6
3.3	Листинг для задания 3.	8
4	Задание 4.	9
4.1	Условие задачи.	9
4.2	Решение.	9
5	Задание 5.	11
5.1	Условие задачи.	11
5.2	Решение.	11

1 Задание 1.

1.1 Условие задачи.

В кармане имеется 3 монеты по 5 рублей, 2 монеты по 3 рубля и 6 монет по рублю. Из кармана случайным образом извлекаются три монеты, составить ряд распределения суммы вытащенных из кармана денег, построить функцию распределения, найти мат. ожидание, дисперсию, с.кв.откл., медиану, моду. Построить график функции распределения, многоугольник распределения.

1.2 Решение.

Рассмотрим различные комбинации трех монет. Пока что не будем учитывать перестановки.

555, 553, 551, 331, 111, 531, 533, 311, 511

Запомним соответствующие им суммы.

15, 13, 11, 7, 3, 9, 11, 5, 7

Числа в каждой из комбинаций можно располагать в разном порядке. Например, пятерки и единицы можно расположить в строку из трех чисел шестью способами. Поэтому нам потребуется использовать формулу сочетаний из n по k вида

$$C_n^k = \frac{n!}{(n-k)!k!};$$

Распишем число сочетаний для каждой из комбинаций, приведенных ранее.

$$C_3^3, C_3^2 \cdot C_2^1, C_3^2 \cdot C_6^1, C_2^2 \cdot C_6^1, C_6^3, C_3^1 \cdot C_2^1 \cdot C_6^1, C_3^1 \cdot C_2^2, C_2^1 \cdot C_6^2, C_3^1 \cdot C_6^2$$

Чтобы найти вероятность с помощью перестановок нужно учесть все возможные исходы. Всего монет 11, а берем мы 3, значит для нахождения вероятности нам потребуется разделить число сочетаний для каждого рассматриваемого случая на

$$C_{11}^3 = \frac{11!}{(11-3)!3!} = 165;$$

Пример расчета приведен ниже.

$$\frac{C_3^1 \cdot C_2^1 \cdot C_6^1}{C_{11}^3} = \frac{\frac{3!}{2!} \cdot \frac{2!}{1!} \cdot \frac{6!}{5!}}{165} = \frac{12}{55}$$

Теперь составим ряд распределения суммы вытащенных из кармана денег. Ранее мы записывали ее для каждой из комбинаций. Видим, что сумма в 7 и 11 рублей повторяются дважды – вероятность для этих сумм будет складываться из соответствующих вероятностей, посчитанных формулой сочетаний. Получим следующее:

$$P(X=7) = \frac{C_2^2 \cdot C_6^1}{C_{11}^3} + \frac{C_3^1 \cdot C_2^2}{C_{11}^3} = \frac{2}{55} + \frac{3}{11} = \frac{17}{55}$$

$$P(X=11) = \frac{C_3^2 \cdot C_6^1}{C_{11}^3} + \frac{C_3^1 \cdot C_2^2}{C_{11}^3} = \frac{6}{55} + \frac{1}{55} = \frac{7}{55}$$

Теперь составим таблицу $p_i(x_i)$.

x_i	3	5	7	9	11	13	15
p_i	$4/33$	$2/11$	$17/55$	$12/55$	$7/55$	$2/55$	$1/165$

Таблица 1: Ряд распределения суммы вытащенных из кармана денег.

Запишем функцию распределения $F(x)$, построим ее график и многоугольник распределения.

$$F(x) = \begin{cases} 0, & x \leq 3 \\ 4/33, & 3 < x \leq 5 \\ 10/33, & 5 < x \leq 7 \\ 101/165, & 7 < x \leq 9 \\ 137/165, & 9 < x \leq 11 \\ 158/165, & 11 < x \leq 13 \\ 164/165, & 13 < x \leq 15 \\ 1, & x > 15 \end{cases}$$

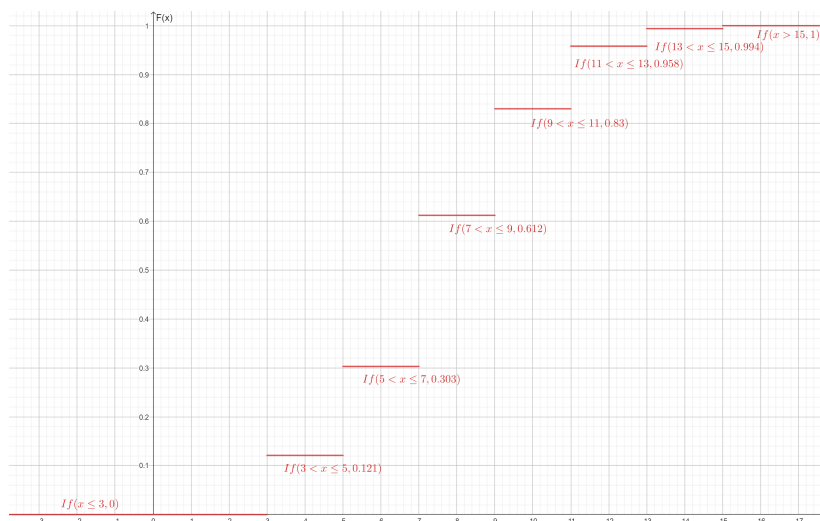


Рис. 1: График функции распределения.

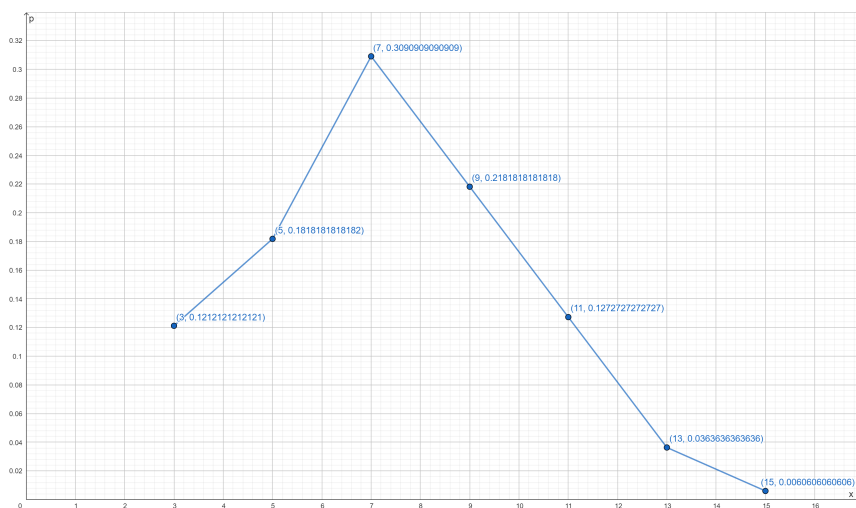


Рис. 2: Многоугольник распределения.

Далее найдем математическое ожидание (k – количество x_i), дисперсию, среднее квадратичное отклонение, медиану, моду и вероятность того, что сумма окажется не менее 10 рублей. Медиана – средний элемент выборки, левее и правее которого одинаковое количество элементов, если в выборке нечетное количество значений. Мода – самый часто повторяющийся элемент выборки. В нашем случае их два, поэтому я приведу оба.

$$M(x) = \sum_{i=1}^k x_i p_i = 3 \cdot \frac{4}{33} + 5 \cdot \frac{2}{11} + 7 \cdot \frac{17}{55} + 9 \cdot \frac{12}{55} + 11 \cdot \frac{7}{55} + 13 \cdot \frac{2}{55} + 15 \cdot \frac{1}{165} = \frac{81}{11},$$

$$M(x^2) = \sum_{i=1}^k x_i^2 p_i = 3^2 \cdot \frac{4}{33} + 5^2 \cdot \frac{2}{11} + 7^2 \cdot \frac{17}{55} + 9^2 \cdot \frac{12}{55} + 11^2 \cdot \frac{7}{55} + 13^2 \cdot \frac{2}{55} + 15^2 \cdot \frac{1}{165} = \frac{675}{11},$$

$$D(x) = M(x^2) - (M(x))^2 = \frac{675}{11} - \left(\frac{81}{11}\right)^2 = \frac{675}{11} - \frac{6561}{121} = \frac{864}{121},$$

$$\delta(x) = \sqrt{D(x)} = \sqrt{\frac{864}{121}} = \frac{12\sqrt{6}}{11},$$

$$MED(x) = 9 \Rightarrow P(X = MED(x) = 9) = \frac{12}{55},$$

$$MODE_1(x) = 7 \Rightarrow P(X = MODE_1(x) = 7) = \frac{17}{55},$$

$$MODE_2(x) = 11 \Rightarrow P(X = MODE_2(x) = 11) = \frac{7}{55},$$

$$P(X \geq 10) = P(X = 11) + P(X = 13) + P(X = 15) = \frac{7}{55} + \frac{2}{55} + \frac{1}{165} = \frac{28}{165}$$

Ответ:

$$M(X) = \frac{81}{11}, D(X) = \frac{864}{121}, \delta(X) = \frac{12\sqrt{6}}{11}, P(X = MED(x) = 9) = \frac{12}{55},$$

$$P(X = MODE_1(x) = 7) = \frac{17}{55}, P(X = MODE_2(x) = 11) = \frac{7}{55},$$

$$P(X \geq 10) = \frac{28}{165}$$

2 Задание 2.

2.1 Условие задачи.

На рынок поступила партия арбузов. Предполагается, что вес одного арбуза – случайная величина, подчиняющаяся нормальному закону распределения с математическим ожиданием $M(X) = 9$ кг и средним квадратическим отклонением $\sigma = 0.5$ кг. Определите вероятность того, что вес случайно отобранного арбуза:

1. Окажется больше 12 кг;
2. Окажется меньше 7.5 кг;
3. Будет находиться между 8.5 и 9.5 кг;

4. Отклонится от математического ожидания меньше, чем на 0.1 кг;
5. Отклонится от математического ожидания больше, чем на 0.2 кг;
6. Найдите границы, в которых отклонение веса случайно отобранной туши от своего математического ожидания не превысит утроенного среднего квадратического отклонения (проиллюстрируйте правило трех сигм).

2.2 Решение.

Запомним важные данные – мат. ожидание $M(X) = 9$ кг и ср. кв. откл. $\sigma = 0.5$ кг.

Для первого задания отклонение $\mu = 12$ кг. Сначала найдем вероятность, что масса арбуза окажется меньше 12 кг (значения Φ смотрим по таблице в интернете):

$$P(X \leq Z) = \Phi\left(\frac{\mu - M(X)}{\sigma}\right) = \Phi\left(\frac{12 - 9}{0.5}\right) = \Phi(6) = P(X \leq Z = 6) \approx 1$$

Получили площадь под графиком нормального распределения вплоть до 6σ . Чтобы найти обратную вероятность (масса арбуза окажется больше 12 кг), вычтем из 1 полученную вероятность на предыдущем шаге:

$$P = 1 - P(X \leq Z = 6) = 1 - 1 = 0$$

Как видим, вероятность такого отклонения стремится к нулю, что логично, так как площадь за 6σ стремится к 0, а мы нашли все, что до нее, площадь чего стремится к 1.

Для второго задания $\mu = 7.5$ кг. От нас требуют найти площадь под графиком нормального распределения в диапазоне $[-\infty, 7.5]$. Мы сразу можем найти соответствующую вероятность (вычисления с Φ аналогичные):

$$P = \Phi\left(\frac{7.5 - 9}{0.5}\right) = \Phi(-3) = 0.0135$$

Вероятность получилась очень маленькой, что означает, что площадь нормального распределения от 7.5 до $-\infty$ кг стремится к нулю.

В третьем задании необходимо найти площадь средней части нормального распределения при $\mu_1 = 8.5$, $\mu_2 = 9.5$ кг. Аналогично найдем вероятность. Вычтем из площади от $-\infty$ до 9.5 площадь $-\infty$ до 8.5 кг и получим середину:

$$P = \Phi\left(\frac{9.5 - 9}{0.5}\right) - \Phi\left(\frac{8.5 - 9}{0.5}\right) = \Phi(1) - \Phi(-1) = 0.84134 - 0.15866 = 0.68268$$

Здесь вероятность получилась уже побольше, что логично, так как средняя часть нормального распределения имеет явно бóльшую площадь, чем ее края.

В четвертом задании все то же самое, но теперь отклонение задано как $M(X) < 0.1$ кг.

$$P = \Phi\left(\frac{9.1 - 9}{0.5}\right) - \Phi\left(\frac{8.9 - 9}{0.5}\right) = \Phi(0.2) - \Phi(-0.2) = 0.57926 - 0.4274 = 0.15186$$

Вероятность получилась небольшой, так как мы взяли лишь малую часть площади средней части нормального распределения.

В пятом задании отклонение задано как $M(X) > 0.2$ кг. Функция нормального распределения является симметричной. Значит нам достаточно посчитать только часть левее середины, которая при стремлении к $-\infty$ стремится к 0, и умножить на 2.

$$P = 2\Phi\left(\frac{8.8 - 9}{0.5}\right) = 2\Phi(-0.4) = 2 \cdot 0.34458 = 0.68916$$

Для последнего задания нужно решить систему, чтобы определить границы, в которых отклонение веса случайно отобранной туши от своего мат. ожидания не превысит 3σ .

$$\Phi\left(\frac{x-9}{0.5}\right) = \begin{cases} (x-9)/0.5 \geq -3 \\ (x-9)/0.5 \leq 3 \end{cases} \quad \Phi\left(\frac{x-9}{0.5}\right) = \begin{cases} x \geq 7.5 \\ x \leq 10.5 \end{cases}$$

Получили диапазон $[7.5, 10.5]$. Теперь проиллюстрируем правило трех сигм – разница Φ в этом диапазоне должна быть равна примерно 0.997:

$$P = \Phi\left(\frac{10.5 - 9}{0.5}\right) - \Phi\left(\frac{7.5 - 9}{0.5}\right) = \Phi(3) - \Phi(-3) = 0.99865 - 0.00135 = 0.9973$$

Видим, что правило трех сигм выполняется.

Ответ:

1. $P \approx 0$, 2. $P = 0.0135$, 3. $P = 0.68268$,

4. $P = 0.15186$, 5. $P = 0.68916$, 6. $[7.5, 10.5]$

3 Задание 3.

3.1 Условие задачи.

В урне 5 шаров: 3 белых и 2 черных. Производится две последовательных выборки без возвращения каждый раз по два шара. Определены две случайные величины: число белых шаров в первой выборке и число белых шаров во второй выборке. Построить функцию распределения, ее график.

3.2 Решение.

Пусть X – число *белых шаров*, полученных в первую выборку, Y – во вторую. Функция плотности распределения будет выглядеть как таблица, где столбец будет отвечать за X , строка – за Y . Рассматривать нужно случаи 0, 1, 2 – достали 0 белых шаров, 1 и 2 соответственно.

Теперь поочередно будем рассматривать элементы таблицы. Будем представлять их в виде a_{ij} – элемент a на i строке и j столбце. Счет строк и столбцов начинается с нуля, как в программировании. Для определения вероятности будем пользоваться формулой числа сочетаний из n по k . Если, например, $i = 1, j = 1$, то это означает, что в первую выборку достали 1 белый шар и 1 черный, во вторую то же самое. Для нахождения вероятности каждого случая, необходимо делить результат сочетаний на

$$C_5^2 = \frac{5!}{2!3!} = 10, \quad C_3^2 = \frac{3!}{2!1!} = 3,$$

где C_5^2 – достали 2 шара из 5, C_3^2 – достали 2 шара из 3 (два уже взяли, шары не возвращаются). Таким образом, получим:

$$a_{00} = 0 \text{ (Не хватает черных шаров, в этом случае нужно 4, а их 2)}$$

$$a_{22} = 0 \text{ (Не хватает белых шаров, в этом случае нужно 4, а их 3)}$$

$$a_{10} = 0 \text{ (Не хватает черных шаров, в этом случае нужно 3, а их 2)}$$

$$a_{01} = 0 \text{ (Аналогично элементу } a_{10})$$

$$a_{02} = \frac{C_3^0 C_2^2}{C_5^2} \cdot \frac{C_3^2}{C_3^2} = \frac{3!/3! \cdot 2!/2!}{10} = 0.1$$

$$a_{20} = \frac{C_3^2 C_2^0}{C_5^2} \cdot \frac{C_1^0 C_2^2}{C_3^2} = \frac{3! 2!/2! 2!}{10} \cdot \frac{1! 2!/1! 2!}{3} = 0.1$$

$$a_{11} = \frac{C_3^1 C_2^1}{C_5^2} \cdot \frac{C_2^1 C_1^1}{C_3^2} = \frac{3! 2!/2! 1!}{10} \cdot \frac{2! 1!/1! 1!}{3} = 0.4$$

$$a_{21} = \frac{C_3^2 C_2^0}{C_5^2} \cdot \frac{C_1^1 C_2^1}{C_3^2} = \frac{3! 2!/2! 2!}{10} \cdot \frac{1! 2!/1! 1!}{3} = 0.2$$

$$a_{12} = \frac{C_3^1 C_2^1}{C_5^2} \cdot \frac{C_2^2 C_1^0}{C_3^2} = \frac{3! 2!/2!}{10} \cdot \frac{2! 1!/2! 1!}{3} = 0.2$$

Теперь составим таблицу функции плотности распределения $f(x, y)$.

$x \backslash y$	0	1	2
0	0	0	0.1
1	0	0.4	0.2
2	0.1	0.2	0

Таблица 2: Функция плотности распределения $f(x, y)$.

Осталось составить таблицу функции распределения и построить ее график. Здесь вместо 0, 1, 2 будут следующие строки и столбцы соответственно: $x \leq 0$, $0 < x \leq 1$, $1 < x \leq 2$, $x > 2$ и $y \leq 0$, $0 < y \leq 1$, $1 < y \leq 2$, $y > 2$. Для построения таблицы нужно рассмотреть все эти случаи совместно. Элементы расписаны аналогично предыдущей таблице. Пользоваться будем следующей формулой:

$$F(x, y) = P(X < x, Y < y);$$

Теперь найдем все элементы по ней.

$$x \leq 0, y \leq 0 \Rightarrow a_{00} = F(x, y) = P(X < 0, Y < 0) = 0,$$

$$0 < x \leq 1, y \leq 0 \Rightarrow a_{10} = F(x, y) = P(X < 1, Y < 0) = 0,$$

$$1 < x \leq 2, y \leq 0 \Rightarrow a_{20} = F(x, y) = P(X < 2, Y < 0) = 0,$$

$$x > 2, y \leq 0 \Rightarrow a_{30} = F(x, y) = P(X = 0, 1, 2; Y < 0) = 0,$$

Для диапазонов y аналогично, так как $\nexists X < 0$,

$$0 < x \leq 1, 0 < y \leq 1 \Rightarrow a_{11} = F(x, y) = P(X < 1, Y < 1) = 0,$$

$$0 < x \leq 1, 1 < y \leq 2 \Rightarrow a_{12} = F(x, y) = P(X < 1, Y < 2) = 0 + 0 = 0,$$

$$0 < x \leq 1, y > 2 \Rightarrow a_{13} = F(x, y) = P(X = 0; Y = 0, 1, 2) = 0 + 0 + 0.1 = 0.1,$$

$$1 < x \leq 2, 0 < y \leq 1 \Rightarrow a_{21} = F(x, y) = P(X = 0, 1; Y = 0) = 0 + 0 = 0,$$

$$x > 2, 0 < y \leq 1 \Rightarrow a_{31} = F(x, y) = P(X = 0, 1, 2; Y = 0) = 0 + 0 + 0.1 = 0.1,$$

$$1 < x \leq 2, 1 < y \leq 2 \Rightarrow a_{22} = F(x, y) = P(X = 0, 1; Y = 0, 1) = 0 + 0 + 0 + 0.4,$$

$$x > 2, 1 < y \leq 2 \Rightarrow a_{32} = F(x, y) = P(X = 0, 1, 2; Y = 0, 1) = 0.4 + 0.1 + 0.2 = 0.7,$$

$$1 < x \leq 2, y > 2 \Rightarrow a_{23} = F(x, y) = P(X = 0, 1; Y = 0, 1, 2) = 0.4 + 0.1 + 0.2 = 0.7,$$

$$x > 2, y > 2 \Rightarrow a_{33} = F(x, y) = P(X = 0, 1, 2; Y = 0, 1, 2) = 0.1 + 0.4 + 0.2 + 0.1 + 0.2 = 1$$

Результаты приведены ниже:

$x \backslash y$	$y \leq 0$	$0 < y \leq 1$	$1 < y \leq 2$	$y > 2$
$x \leq 0$	0	0	0	0
$0 < x \leq 1$	0	0	0	0.1
$1 < x \leq 2$	0	0	0.4	0.7
$x > 2$	0	0.1	0.7	1

Таблица 3: Функция распределения $F(x, y)$.

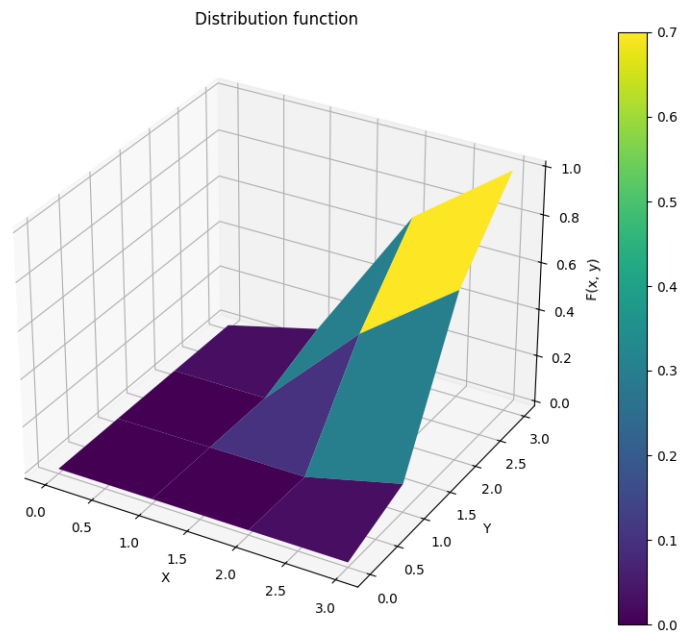


Рис. 3: График функции распределения $F(x, y)$.

3.3 Листинг для задания 3.

Для построения трехмерного графика я написал код на языке Python с использованием библиотек numpy и matplotlib.

```

1 import matplotlib.pyplot as plt
2 import numpy as np
3
4 x = np.array([0, 1, 2, 3])
5 y = np.array([0, 1, 2, 3])
6 X, Y = np.meshgrid(x, y)
7 Z = np.array([
8     [0, 0, 0, 0],

```

```

9      [0, 0, 0, 0.1],
10     [0, 0, 0.4, 0.7],
11     [0, 0.1, 0.7, 1]
12 ])
13
14 fig = plt.figure(figsize=(12, 8))
15 ax = fig.add_subplot(111, projection='3d')
16
17 surface = ax.plot_surface(X, Y, Z, cmap='viridis')
18
19 fig.colorbar(surface)
20
21 ax.set_title('Distribution function')
22 ax.set_xlabel('X')
23 ax.set_ylabel('Y')
24 ax.set_zlabel('F(x, y)')
25 ax.set_zlim(0, 1)
26
27 plt.show()

```

Листинг 1: Программа для построения трехмерного графика.

Ответ: Получена функция распределения $F(x, y)$ и построен ее график.

4 Задание 4.

4.1 Условие задачи.

По данным статистической службы 7.5% населения – безработные. Используя неравенство Чебышева оценить вероятность того, что в наудачу отобранной группе в 1000 человек доля безработных заключена в границах от 4% до 6%. Найти вероятность того же события, используя следствие из интегральной теоремы Муавра-Лапласа. Сравнить результаты и обосновать их

4.2 Решение.

Для начала рассмотрим неравенство Чебышева. Пусть X – количество безработных в наудачу отобранной группе из $n = 1000$ человек. Вероятность того, что человек безработный, дана в условии, и равна $p = 0.075$. X имеет биномиальное распределение вида

$$X \sim \text{Bin}(1000, 0.075);$$

Матожидание и дисперсия примут вид

$$M(X) = np = 1000 \cdot 0.075 = 75,$$

$$D(X) = np(1 - p) = 1000 \cdot 0.075 \cdot 0.925 = 69.375$$

Пусть Y – доля безработных в группе. Рассмотрим Y и найдем для него матожидание, дисперсию и среднеквадратичное отклонение:

$$Y = \frac{X}{1000}, \quad M(Y) = \frac{M(X)}{1000} = 0.075, \quad D(Y) = \frac{D(X)}{1000^2} = \frac{69.375}{1000000} = 0.000069375,$$

$$\sigma_Y = \sqrt{D(Y)} = \sqrt{0.000069375} \approx 0.00833;$$

По неравенству Чебышева имеем:

$$P(|Y - M(Y)| \geq k\sigma_Y) \leq \frac{1}{k^2}$$

По условию даны границы в 4% и 6%, рассмотрим Y с отклонением в 7.5%:

$$\begin{aligned} 0.04 \leq Y \leq 0.06 \\ 0.04 - 0.075 \leq Y - 0.075 \leq 0.06 - 0.075 \\ -0.035 \leq Y - 0.075 \leq -0.015 \end{aligned}$$

Выразим в единицах стандартного отклонения:

$$-0.035 \leq Y - 0.075 \leq -0.015 \Rightarrow -\frac{0.035}{0.00833} \leq \frac{Y - 0.075}{0.00833} \leq -\frac{0.015}{0.00833} \Rightarrow -4.2 \leq k \leq -1.8$$

Вернемся к неравенству Чебышева и подставим полученные значения:

$$P(|Y - 0.075| \geq 0.035) \leq \frac{1}{4.2^2} \approx 0.0566$$

Тогда вероятность того, что доля безработных окажется в пределах от 4% до 6%:

$$P(0.04 \leq Y \leq 0.06) \geq 1 - 0.0566 = 0.9434$$

Теперь найдем ту же вероятность, используя следствие из интегральной теоремы Муавра-Лапласа – при большом n можно использовать нормальное приближение вида

$$X \approx \mathcal{N}(np, \sqrt{np(1-p)})$$

Применим его к Y :

$$Y \approx \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right) \Rightarrow Y \approx \mathcal{N}\left(0.075, \sqrt{\frac{0.075 \cdot 0.925}{1000}}\right) = \mathcal{N}(0.075, 0.00833)$$

Найдем вероятность того, что Y находится в пределах 4% и 6%:

$$P(0.04 \leq Y \leq 0.06) = P\left(\frac{0.04 - 0.075}{0.00833} \leq Z \leq \frac{0.06 - 0.075}{0.00833}\right) = P(-4.2 \leq Z \leq -1.8),$$

где полученное $P(Z)$ – стандартное нормальное распределение Z . Найдем для него значения:

$$P(Z \leq -1.8) \approx 0.0359, \quad P(Z \leq -4.2) \approx 0.000013$$

Тогда имеем:

$$P(-4.2 \leq Z \leq -1.8) = P(Z \leq -1.8) - P(Z \leq -4.2) \approx 0.0359 - 0.000013 = 0.0359$$

Вероятность $P(Z \leq -4.2)$ очень мала, поэтому мы пренебрежем ей и получим ответ:

$$P(-4.2 \leq Z \leq -1.8) \approx 0.0359 \Rightarrow P(0.04 \leq Y \leq 0.06) \approx 0.0359$$

По неравенству Чебышева вероятность того, что доля безработных окажется в пределах от 4% до 6% составила примерно 0.9434. По интегральной теореме Муавра-Лапласа вероятность составила примерно 0.0359.

Неравенство Чебышева даёт менее точную оценку, так как оно применимо к любым распределениям и предоставляет верхнюю границу вероятности. В то время как теорема Муавра-Лапласа основана на нормальном приближении и даёт более точную оценку для больших выборок.

Результаты отличаются, и оценка по теореме Муавра-Лапласа более точна в данном контексте из-за нормального распределения, которое лучше отражает реальное поведение биномиального распределения при больших n .

Ответ: что и требовалось исследовать.

5 Задание 5.

5.1 Условие задачи.

Вычислить характеристическую функцию и найти $M(X)$ и $D(X)$ для гипергеометрического распределения.

5.2 Решение.

Гипергеометрическое распределение описывает вероятность X успехов в n выборках без возвращения из конечной совокупности из N элементов, содержащих K успешных и $N - K$ неуспешных элементов. Будем обозначать гипергеометрическое распределение как

$$X \sim \text{Hypergeometric}(N, K, n)$$

Функция вероятности для гипергеометрического распределения имеет вид:

$$P(X = k) = \frac{\begin{bmatrix} K \\ k \end{bmatrix} \begin{bmatrix} N - K \\ n - k \end{bmatrix}}{\begin{bmatrix} N \\ n \end{bmatrix}},$$

где $\begin{bmatrix} a \\ b \end{bmatrix}$ – биномиальный коэффициент, равный $\frac{a!}{(a-b)!b!}$

Чтобы найти характеристическую функцию, необходимо вычислить математическое ожидание от e^{itX} , где i – мнимая единица, а t – вещественное число. Характеристическая функция $\varphi_X(t)$ для случайной величины X определяется как:

$$\varphi_X(t) = M[e^{itX}]$$

Для гипергеометрического распределения имеем:

$$\varphi_X(t) = M[e^{itX}] = \sum_{k=0}^n e^{itk} P(X = k)$$

Ранее мы выписали функцию вероятности $P(X = k)$. Подставим ее в выражение выше:

$$\varphi_X(t) = \sum_{k=0}^n e^{itk} \frac{\begin{bmatrix} K \\ k \end{bmatrix} \begin{bmatrix} N - K \\ n - k \end{bmatrix}}{\begin{bmatrix} N \\ n \end{bmatrix}} = \frac{1}{\begin{bmatrix} N \\ n \end{bmatrix}} \sum_{k=0}^n e^{itk} \begin{bmatrix} K \\ k \end{bmatrix} \begin{bmatrix} N - K \\ n - k \end{bmatrix}$$

Для дальнейших вычислений будем использовать метод производящих функций. Пусть $G(z)$ – производящая функция для гипергеометрического распределения:

$$G(z) = \sum_{k=0}^n P(X = k) z^k$$

Известно, что производящую функцию гипергеометрического распределения можно записать как:

$$G(z) = \frac{\begin{bmatrix} N - n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} (1 + (z - 1))^n$$

Чтобы получить характеристическую функцию, заменим z на e^{it} :

$$\varphi_X(t) = G(e^{it}) = \frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} (1 + (e^{it} - 1))^n$$

Подставим это выражение в производящую функцию и получим окончательный вид характеристической функции гипергеометрического распределения:

$$\varphi_X(t) = \frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} (e^{it})^n = \frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} e^{itn}$$

Таким образом, характеристическая функция гипергеометрического распределения X записывается так:

$$\varphi_X(t) = \frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} (1 + (e^{it} - 1))^n$$

Математическое ожидание $M(X)$ можно получить как первую производную характеристической функции по t при $t = 0$:

$$M[X] = \left. \frac{d\varphi_X(t)}{dt} \right|_{t=0}$$

Для гипергеометрического распределения получим:

$$\begin{aligned} \left. \frac{d\varphi_X(t)}{dt} \right|_{t=0} &= \left(\frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} (1 + (e^{it} - 1))^n \right)' \Big|_{t=0} = \frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} \cdot n \cdot (1 + (e^{it} - 1))^{n-1} \cdot i e^{it} \Big|_{t=0} \\ \left. \frac{d\varphi_X(t)}{dt} \right|_{t=0} &= \frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} \cdot n \cdot (1 + (1 - 1))^{n-1} \cdot i \cdot 1 = \frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} \cdot n \cdot 1 \cdot i = n \frac{K}{N} \Rightarrow M(X) = n \frac{K}{N} \end{aligned}$$

Дисперсию можно найти как вторую производную по t при $t = 0$:

$$D(X) = \left. \frac{d^2\varphi_X(t)}{dt^2} \right|_{t=0} - (M[X])^2$$

Для гипергеометрического распределения получим:

$$\left. \frac{d^2\varphi_X(t)}{dt^2} \right|_{t=0} = \left(\frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} (1 + (e^{it} - 1))^n \right)'' \Big|_{t=0}$$

$$\left. \frac{d^2 \varphi_X(t)}{dt^2} \right|_{t=0} = \frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} \cdot n \cdot \left[(1 + (e^{it} - 1))^{n-1} \cdot i^2 e^{2it} + (n-1)(1 + (e^{it} - 1))^{n-2} \cdot i^2 e^{2it} \right] \Big|_{t=0}$$

$$\left. \frac{d^2 \varphi_X(t)}{dt^2} \right|_{t=0} = \frac{\begin{bmatrix} N-n \\ K \end{bmatrix}}{\begin{bmatrix} N \\ K \end{bmatrix}} \cdot n \cdot [1^{n-1} \cdot (-1) + (n-1) \cdot 1^{n-2} \cdot (-1)] = n \frac{K}{N} \left(1 - \frac{K}{N} \right) \frac{N-n}{N-1}$$

$$\left. \frac{d^2 \varphi_X(t)}{dt^2} \right|_{t=0} = n \frac{K}{N} \left(1 - \frac{K}{N} \right) \frac{N-n}{N-1} \Rightarrow D(X) = n \frac{K}{N} \left(1 - \frac{K}{N} \right) \frac{N-n}{N-1}$$

Ответ: что и требовалось найти.