

Национальный исследовательский университет
Высшая школа экономики
Московский институт электроники и математики

Департамент прикладной математики
кафедра компьютерной безопасности

Домашнее задание №1 по математической статистике
Характеристики вероятностных распределений

Дискретное распределение: *дискретное равномерное I*

Неизвестный параметр: $\theta = 121$

Непрерывное распределение: *распределение Парето*

Неизвестный параметр: $\theta = 12$

Выполнила
Мазитова Е.А.

Проверил
Богданов Д.С.

Москва 2025

Содержание

1	Описание основных характеристик распределения	3
2	Поиск примеров событий, которые могут быть описаны выбранными случайными величинами	7
3	Описание способа моделирования выбранных случайных величин	11
4	Приложения	14

1. Описание основных характеристик распределения

▷ Дискретное распределение (дискретное равномерное I)

$$P(x) = \theta^{-1}, \quad x \in \{1, \dots, \theta\}$$

1. Функция распределения для дискретной случайной величины считается по следующей формуле:

$$F(x) = P(X \leq x) = \sum_{i=1}^x p_i$$

Поскольку функция распределения по определению должна быть определена для любого действительного числа x , рассмотрим три различных случая:

(a) $x < 1$

$$F(x) = \sum_{i=1}^x p_i = \sum_{i=1}^x 0 = 0$$

(b) $x \in \{1, \dots, \theta\}$

$$F(x) = \sum_{i=1}^x p_i = \sum_{i=1}^x \theta^{-1} = \frac{x}{\theta}$$

(c) $x > \theta$

$$F(x) = P(X \leq x) = 1$$

Таким образом, функция распределения имеет следующий вид:

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{x}{\theta}, & x \in \{1, \dots, \theta\} \\ 1, & x > \theta \end{cases}$$

2. Математическое ожидание для дискретной случайной величины считается по следующей формуле:

$$E[X] = \sum_{i=1}^{\theta} x \cdot P(x)$$

$$E[X] = \sum_{i=1}^{\theta} x \cdot \theta^{-1} = \frac{1}{\theta} \sum_{i=1}^{\theta} x = \left| \sum_{i=1}^{\theta} x = \frac{\theta(\theta+1)}{2} \right| = \frac{1}{\theta} \cdot \frac{\theta(\theta+1)}{2} = \frac{\theta+1}{2}$$

Таким образом, математическое ожидание имеет следующий вид:

$$E[X] = \frac{\theta+1}{2}$$

3. Дисперсия случайной величины считается по следующей формуле:

$$D[X] = E[X^2] - (E[X])^2$$

Найдём второй момент случайной величины ($E[X^2]$):

$$E[X^2] = \sum_{x=1}^{\theta} x^2 \cdot \theta^{-1} = \frac{1}{\theta} \sum_{x=1}^{\theta} x^2$$

Воспользуемся формулой суммы квадратов первых θ чисел:

$$\sum_{x=1}^{\theta} x^2 = \frac{\theta(\theta+1)(2\theta+1)}{6}$$

$$E[X^2] = \frac{1}{\theta} \cdot \frac{\theta(\theta+1)(2\theta+1)}{6} = \frac{(\theta+1)(2\theta+1)}{6}$$

Из предыдущего пункта известно математическое ожидание:

$$E[X] = \frac{\theta+1}{2}$$

$$(E[X])^2 = \left(\frac{\theta+1}{2}\right)^2 = \frac{(\theta+1)^2}{4}$$

Подставим все в формулу дисперсии:

$$\begin{aligned} D[X] &= E[X^2] - (E[X])^2 = \frac{(\theta+1)(2\theta+1)}{6} - \frac{(\theta+1)^2}{4} = \\ &= \frac{2(\theta+1)(2\theta+1)}{12} - \frac{3(\theta+1)^2}{12} = \frac{\theta+1}{12} [2(2\theta+1) - 3(\theta+1)] = \\ &= \left| 2(2\theta+1) - 3(\theta+1) = 4\theta+2-3\theta-3 = \theta-1 \right| = \\ &= \frac{\theta+1}{12} \cdot (\theta-1) = \frac{\theta^2-1}{12} \end{aligned}$$

Таким образом, дисперсия имеет следующий вид:

$$D[X] = \frac{\theta^2-1}{12}$$

4. Квантиль распределения уровня γ - это значение μ_γ , которое является решением следующего уравнения:

$$F_\xi(\mu_\gamma) = \gamma,$$

где F_ξ - функция распределения случайной величины ξ .

Ранее посчитали, что функция распределения имеет следующий вид:

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{x}{\theta}, & x \in \{1, \dots, \theta\} \\ 1, & x > \theta \end{cases}$$

$$F(\mu_\gamma) = \frac{\mu_\gamma}{\theta} \iff \frac{\mu_\gamma}{\theta} = \gamma$$

$$\frac{\mu_\gamma}{\theta} = \gamma \iff \mu_\gamma = \gamma \cdot \theta$$

Поскольку мы рассматриваем дискретное распределение, функция распределения которого задана в виде $\frac{x}{\theta}$ только для целых значений $x \in \{1, \dots, \theta\}$, следует округлить получившееся значение μ_γ так, чтобы для найденного значения выполнялось условие $F_\xi(\mu_\gamma) \geq \gamma : \mu_\gamma = \lceil \gamma \cdot \theta \rceil$.

Таким образом, квантиль уровня γ имеет следующий вид:

$$\mu_\gamma = \lceil \gamma \cdot \theta \rceil$$

▷ Непрерывное распределение (распределение Парето)

$$f(x) = \theta \cdot x^{-(\theta+1)}, \quad x \in [1, +\infty), \quad \theta > 0$$

1. Функция распределения для непрерывной случайной величины считается по следующей формуле:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Поскольку функция распределения по определению должна быть определена для любого действительного числа x , рассмотрим два различных случая:

(а) $x < 1$

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^x 0 dt = 0$$

(b) $x \geq 1$

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t)dt = \int_{-\infty}^1 f(t)dt + \int_1^x f(t)dt = 0 + \int_1^x \theta t^{-(\theta+1)}dt = \\ &= \theta \cdot \int_1^x t^{-\theta-1} = \theta \cdot \left. \frac{t^{-\theta}}{-\theta} \right|_1^x = -x^{-\theta} + 1^{-\theta} = 1 - x^{-\theta} \end{aligned}$$

Таким образом, функция распределения имеет следующий вид:

$$F(x) = \begin{cases} 0, & x < 1 \\ 1 - x^{-\theta}, & x \geq 1 \end{cases}$$

2. Математическое ожидание для непрерывной случайной величины считается по следующей формуле:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f(x)dx \\ E[X] &= \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{-\infty}^1 0 + \int_1^{\infty} x \cdot \theta x^{-(\theta+1)}dx = \int_1^{\infty} \theta x^{-\theta-1+1}dx = \\ &= \theta \int_1^{\infty} x^{-\theta}dx = \theta \cdot \left. \frac{x^{-\theta+1}}{-\theta+1} \right|_1^{\infty} = \theta \cdot 0 - \frac{\theta}{-\theta+1} = \frac{\theta}{\theta-1} \end{aligned}$$

Таким образом, математическое ожидание имеет следующий вид:

$$E[X] = \frac{\theta}{\theta-1}$$

3. Дисперсия случайной величины считается по следующей формуле:

$$D[X] = E[X^2] - (E[X])^2$$

Найдём второй момент случайной величины ($E[X^2]$):

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_1^{\infty} \theta \cdot x^{-\theta-1+2}dx = \theta \int_1^{\infty} x^{-\theta+1}dx = \\ &= \theta \cdot \left. \frac{x^{-\theta+2}}{2-\theta} \right|_1^{\infty} = -\frac{\theta}{2-\theta} = \frac{\theta}{\theta-2} \end{aligned}$$

Из предыдущего пункта известно математическое ожидание:

$$E[X] = \frac{\theta}{\theta-1}$$

$$(E[X])^2 = \left(\frac{\theta}{\theta - 1} \right)^2 = \frac{\theta^2}{(\theta - 1)^2}$$

Подставим все в формулу дисперсии:

$$\begin{aligned} D[X] &= E[X^2] - (E[X])^2 = \frac{\theta}{\theta - 2} - \frac{\theta^2}{(\theta - 1)^2} = \frac{\theta(\theta - 1)^2 - \theta^2(\theta - 2)}{(\theta - 2)(\theta - 1)^2} \\ &= \left| \theta(\theta - 1)^2 - \theta^2(\theta - 2) = \theta(\theta^2 - 2\theta + 1) - \theta^3 + 2\theta^2 = \theta^3 - 2\theta^2 + \theta - \theta^3 + 2\theta^2 = \theta \right| = \\ &\quad \frac{\theta}{(\theta - 2)(\theta - 1)^2} \end{aligned}$$

Таким образом, дисперсия имеет следующий вид:

$$D[X] = \frac{\theta}{(\theta - 2)(\theta - 1)^2}$$

4. Квантиль распределения уровня γ - это значение μ_γ , которое является решением следующего уравнения:

$$F_\xi(\mu_\gamma) = \gamma,$$

где F_ξ - функция распределения случайной величины ξ .

Ранее посчитали, что функция распределения имеет следующий вид:

$$F(x) = \begin{cases} 0, & x < 1 \\ 1 - x^{-\theta}, & x \geq 1 \end{cases}$$

$$F(\mu_\gamma) = 1 - \mu_\gamma^{-\theta} \iff 1 - \mu_\gamma^{-\theta} = \gamma$$

$$1 - \mu_\gamma^{-\theta} = \gamma \iff 1 - \gamma = \mu_\gamma^{-\theta} \iff \mu_\gamma = (1 - \gamma)^{-\frac{1}{\theta}}$$

Таким образом, квантиль уровня γ имеет следующий вид:

$$\mu_\gamma = (1 - \gamma)^{-\frac{1}{\theta}}$$

2. Поиск примеров событий, которые могут быть описаны выбранными случайными величинами

▷ Дискретное распределение (дискретное равномерное I)

$$P(x) = \theta^{-1}, \quad x \in \{1, \dots, \theta\}$$

1. Пример интерпретации распределения

Предположим, студенту группы СКБ-231 предстоит сдать экзамен по курсу "Математическая статистика". Пусть на экзамене предусмотрено θ уникальных билетов, каждый из которых содержит один вопрос по конкретной теме курса математической статистики. Предположим, что студент не успел подготовиться к сдаче экзамена, из-за чего выучил только один билет из всего списка (какой конкретно - не важно). Пусть также будет верно следующее:

- Билеты пронумерованы от 1 до θ , вопросы в них не повторяются
- Процесс выбора билета случаен (например, студент выбирает билет с закрытыми глазами, или они перемешаны на столе преподавателя)

Студент задается вопросом: "С какой вероятностью мне попадется билет, на который я смогу ответить правильно?". Поскольку выпадение всех билетов равновероятно, и студент сможет ответить только на 1 вопрос из θ , то вероятность того, что выпадет именно выученный билет, равна $\frac{1}{\theta}$. Таким образом, данная вероятность имеет равномерное I распределение с параметром θ .

2. Связь с другими распределениями

(а) Связь с распределением Бернулли

Допустим, мы хотим проверить, попало ли наблюдение в **конкретную точку** k , где $k \in \{1, 2, \dots, \theta\}$.

Например, у нас есть коробка с θ шарами, откуда мы наугад вытаскиваем один, причем каждый из них пронумерован от 1 до θ . Пусть нас интересует вероятность достать один конкретный шар под номером k из коробки. Вероятность успеха в таком случае будет равна $\frac{1}{\theta}$, а вероятность неудачи (того, что мы достанем любой другой шар) $= 1 - \frac{1}{\theta}$. Таким образом мы задали случайную величину ξ , где:

$$\begin{cases} P(\xi = k) = \frac{1}{\theta} \\ P(\xi \neq k) = 1 - \frac{1}{\theta} \end{cases}$$

Мы получили случайную величину с распределением Бернулли с параметром $\frac{1}{\theta}$ ($p = \frac{1}{\theta}$ и $q = 1 - p = 1 - \frac{1}{\theta}$). Так, если из дискретного равномерного распределения с параметром θ нас интересует только один конкретный исход k , то случайная величина "Наступил ли исход k ?" имеет распределение Бернулли с параметром $p = \frac{1}{\theta}$.

(b) **Связь с биномиальным распределением**

Рассмотрим $\xi_1, \xi_2, \dots, \xi_n$ - последовательность из n независимых испытаний, где в каждом испытании исход имеет дискретное равномерное распределение с параметром θ , и сосредоточимся на одном конкретном исходе k . Ранее доказали, что вероятность наступления данного конкретного исхода имеет распределение Бернулли с параметром $\frac{1}{\theta}$.

Рассмотрим следующую сумму μ :

$$\mu = \xi_1 + \xi_2 + \dots + \xi_n,$$

где $\xi_1, \xi_2, \dots, \xi_n$ - последовательность из n независимых испытаний, каждое из которых имеет распределение Бернулли с параметром $\frac{1}{\theta}$. Мы знаем, что сумма случайных величин с распределением Бернулли с параметром k является случайной величиной с биномиальным распределением с параметрами (n, k) , где n - количество слагаемых в сумме.

Тогда из курса теории вероятностей нам известно, что рассматриваемая случайная величина μ имеет биномиальное распределение с параметрами $(n, \frac{1}{\theta})$.

(c) **Связь с геометрическим распределением**

Снова рассмотрим $\xi_1, \xi_2, \dots, \xi_n$ - последовательность из n независимых испытаний, где в каждом испытании исход имеет дискретное равномерное распределение с параметром θ . Допустим, мы хотим узнать, как распределена величина, равная числу независимых испытаний до первого появления события с исходом k .

Пусть $t+1$ - номер события с первым появлением исхода k . Для рассмотрения данного события необходимо, чтобы t раз наступал иной от k исход ($q = 1 - \frac{1}{\theta}$ по распределению Бернулли), и на $t+1$ событии наступил исход k (вероятность наступления которого равна $p = \frac{1}{\theta}$ по распределению Бернулли). Тогда:

$$P(t) = q^t \cdot p = \left(1 - \frac{1}{\theta}\right)^t \cdot \frac{1}{\theta}, \quad t = 1, 2, 3, \dots$$

Получившаяся формула соответствует вероятности геометрического распределения с параметром $\frac{1}{\theta}$.

Таким образом, геометрическое распределение с параметром $\frac{1}{\theta}$ описывает распределение времени ожидания конкретного исхода в последовательности независимых испытаний, где каждое испытание имеет дискретное равномерное распределение.

Таким образом получили, что если взять последовательность независимых испытаний, где в каждом испытании исход имеет

*дискретное равномерное распределение с параметром θ , и сосредоточиться на одном конкретном исходе, то вероятность получения данного исхода будет иметь **распределение Бернулли** с параметром $\frac{1}{\theta}$, количество испытаний до первого появления этого исхода - **геометрическое распределение** с параметром $\frac{1}{\theta}$, а количество появлений этого исхода за n испытаний - **биномиальное распределение** с параметрами $(n, \frac{1}{\theta})$.*

▷ **Непрерывное распределение (распределение Парето)**

$$f(x) = \theta \cdot x^{-(\theta+1)}, \quad x \in [1, +\infty), \quad \theta > 0$$

1. **Пример интерпретации распределения**

Предположим, экономисты изучают распределение годовых доходов домохозяйств в стране X . Пусть также будет верно следующее:

- *За единицу измерения принят минимальный прожиточный минимум (1 единица)*
- *Эмпирические наблюдения показывают, что доля населения с доходом выше x уменьшается степенным образом с коэффициентом θ*

*Исследователь задается вопросом: "Какая доля населения имеет доход выше определенного уровня x ?". Поскольку распределение годовых доходов домохозяйств часто демонстрирует "тяжелый хвост" (то есть, небольшое количество людей получает непропорционально высокие доходы) и доля населения с доходом выше x уменьшается степенным образом с коэффициентом θ , то доля населения с доходом **выше** x составляет $x^{-\theta}$. Данное выражение соответствует разнице единицы и плотности рассматриваемого распределения (так как по определению $F(x)$ мы должны будем рассмотреть долю домохозяйств, чьи доходы **ниже** x):*

$$x^{-\theta} = 1 - P(X \leq x) \iff F(x) = 1 - x^{-\theta}$$

Плотность распределения является производной от функции распределения. Вычислив производную от $F(x)$, получили, что плотность распределения доходов домохозяйств в стране X равна $\theta x^{-(\theta+1)}$. Таким образом, распределение доходов описывается законом Парето с параметром θ .

2. **Связь с другими распределениями**

- (а) **Связь с экспоненциальным распределением**

Рассмотрим случайную величину ξ с распределением Парето с параметром θ . Выполним для заданной величины логарифмическое преобразование и зададим новую величину μ :

$$\mu = \ln \xi$$

Найдём функцию распределения μ :

$$\begin{aligned} F_\mu(x) &= P(\ln \xi \leq x) = P(\xi \leq e^x) = F_\xi(e^x) = \\ &= 1 - (e^x)^{-\theta} = 1 - e^{-\theta x}, \quad x \geq 0 \end{aligned}$$

полученное выражение является функцией распределения экспоненциального распределения с параметром θ .

Плотность распределения μ :

$$f_\mu(x) = \frac{d}{dx} F_\mu(x) = \theta e^{-\theta x}, \quad x \geq 0$$

Это означает, что логарифм случайной величины, имеющей распределение Парето с параметром θ , распределен экспоненциально с параметром θ .

3. Описание способа моделирования выбранных случайных величин

1. Способ моделирования дискретной случайной величины (дискретное равномерное I распределение)

Рассмотрим случайную величину ξ , имеющую дискретное равномерное I распределение с параметром θ . Для моделирования выбранной случайной величины ξ воспользуемся квантилем распределения уровня γ - значением μ_γ , которое для заданного распределения мы посчитали ранее.

Квантиль уровня γ случайной величины ξ имеет следующий вид:

$$\mu_\gamma = \lceil \gamma \cdot \theta \rceil$$

Алгоритм моделирования:

(а) Генерация случайного значения γ

Сначала необходимо сгенерировать случайное значение γ на отрезке $[0, 1]$. На языке программирования Python это можно сделать с помощью `random.uniform()`, которое использует внутренний генератор псевдослучайных чисел.

(b) Вычисление квантиля уровня γ (μ_γ)

Воспользовавшись ранее выведенной формулой вычисления квантиля уровня γ , найдем значение μ_γ : $\mu_\gamma = \lceil \gamma \cdot \theta \rceil$, где $\gamma \in [0, 1]$.

Необходимо показать, что данный алгоритм правильно моделирует случайную величину ξ . Для этого проверим, что выпадение всех значений от 1 до θ равновероятно, а сумма вероятностей равна 1: Рассмотрим все возможные значения k от 1 до θ . Для каждого k вычислим вероятность $P(X = k)$:

– Для $k = 1$:

$$P(X = 1) = P\left(0 < \gamma \leq \frac{1}{\theta}\right)$$

Так как $\gamma \in [0, 1]$, вероятность попадания в интервал $(0, \frac{1}{\theta}]$ равна его длине:

$$\frac{1}{\theta} - 0 = \frac{1}{\theta}$$

– Для $k = 2$:

$$P(X = 2) = P\left(\frac{1}{\theta} < \gamma \leq \frac{2}{\theta}\right)$$

Длина интервала $(\frac{1}{\theta}, \frac{2}{\theta}]$ равна:

$$\frac{2}{\theta} - \frac{1}{\theta} = \frac{1}{\theta}$$

– Для произвольного $k \in \{1, 2, \dots, \theta\}$:

$$P(X = k) = P\left(\frac{k-1}{\theta} < \gamma \leq \frac{k}{\theta}\right)$$

Длина интервала $(\frac{k-1}{\theta}, \frac{k}{\theta}]$ равна:

$$\frac{k}{\theta} - \frac{k-1}{\theta} = \frac{1}{\theta}$$

Таким образом, для всех $k = 1, 2, \dots, \theta$ выполняется:

$$P(X = k) = \frac{1}{\theta}$$

Проверим, что сумма всех вероятностей равна 1:

$$\sum_{k=1}^{\theta} P(X = k) = \sum_{k=1}^{\theta} \frac{1}{\theta} = \theta \cdot \frac{1}{\theta} = 1$$

Для каждого целого числа k от 1 до θ существует интервал $(\frac{k-1}{\theta}, \frac{k}{\theta}]$, который имеет положительную длину $\frac{1}{\theta}$. Таким образом, алгоритм гарантированно покрывает все числа от 1 до θ , так как для каждого такого числа существует соответствующий интервал значений γ , который приводит к этому результату.

Так, выборка из найденного набора μ_γ будет иметь дискретное равномерное I распределение.

2. Способ моделирования непрерывной случайной величины (распределение Парето)

Рассмотрим случайную величину ξ , имеющую распределение Парето с параметром θ . Для моделирования выбранной случайной величины ξ воспользуемся квантилем распределения уровня γ - значением μ_γ , которое для заданного распределения мы посчитали ранее.

Квантиль уровня γ случайной величины ξ имеет следующий вид:

$$\mu_\gamma = (1 - \gamma)^{-\frac{1}{\theta}}$$

Алгоритм моделирования:

(a) Генерация случайного значения γ

Сначала необходимо сгенерировать случайное значение γ на отрезке $[0, 1]$. На языке программирования Python это можно сделать с помощью `random.uniform()`, которое использует внутренний генератор псевдослучайных чисел.

(b) Вычисление квантиля уровня γ (μ_γ)

Воспользовавшись ранее выведенной формулой вычисления квантиля уровня γ , найдем значение $\mu_\gamma : \mu_\gamma = (1 - \gamma)^{-\frac{1}{\theta}}$, где $\gamma \in [0, 1]$.

Необходимо показать, что данный алгоритм правильно моделирует случайную величину ξ . Для этого проверим, что полученная случайная величина имеет распределение Парето.

Покажем, что для случайной величины $\xi = (1 - \gamma)^{-\frac{1}{\theta}}$, где $\gamma \in [0, 1]$, выполняется $P(\xi \leq x) = F(x)$ для любого $x \geq 1$:

$$\begin{aligned} P(\xi \leq x) &= P\left((1 - \gamma)^{-\frac{1}{\theta}} \leq x\right) = \\ &= \left| (1 - \gamma)^{-\frac{1}{\theta}} \leq x \Rightarrow 1 - \gamma \geq x^{-\theta} \Rightarrow \gamma \leq 1 - x^{-\theta} \right| = \\ &= P(\gamma \leq 1 - x^{-\theta}) \end{aligned}$$

Поскольку γ равномерно распределена на $[0, 1]$, имеем:

$$P(\gamma \leq 1 - x^{-\theta}) = 1 - x^{-\theta}$$

что в точности совпадает с функцией распределения Парето. Плотность распределения является производной от функции распределения. Вычислив производную от $F(x)$, получили, что плотность распределения равна $\theta x^{-(\theta+1)}$, что также совпадает с плотностью распределения Парето.

Проверим область значений: при $\gamma \in [0, 1]$ значение $1 - \gamma \in [0, 1]$, следовательно:

$$\mu_\gamma = (1 - \gamma)^{-\frac{1}{\theta}} \geq 1^{-\frac{1}{\theta}} = 1$$

Таким образом, алгоритм гарантированно генерирует значения из области определения распределения Парето $[1, +\infty)$.

Так, выборка из найденного набора μ_γ будет иметь распределение Парето с параметром θ .

4. Приложения

1. <https://github.com/faisvire/mathstat-hw>