Теория вероятностей и математическая статистика

Случайные величины. Биномиальное распределение. Практическая сессия.

Глеб Карпов

ФКН ВШЭ

Пример 1

$z_1 z_2 \quad б_1 \ldots б_5$

$(z_2, б_2, б_5)$

$$P_X(X=0) = \frac{C_5^3}{C_7^3} = \frac{10}{\frac{4\cdot5\cdot6\cdot7}{4\cdot3\cdot2\cdot1}} = \frac{10}{35} = \frac{2}{7}$$

$$P_X(X=1) = \frac{C_2^1 \cdot C_5^2}{C_7^3} = \frac{2\,\frac{4\cdot5}{2}}{35} = \frac{4}{7}$$

В урне находятся 2 черных и 5 белых шаров. Вы случайным образом выбираете 3 из них. X — это количество черных шаров в вашей выборке. Найдите математическое ожидание и дисперсию X.

$$P(X=2) = \frac{C_2^2 \cdot C_5^1}{C_7^3} = \frac{1}{7}$$

x	0	1	2
P_X	$\frac{2}{7}$	$\frac{4}{7}$	$\frac{1}{7}$

$$E[X] = \sum x_i \cdot P_X(x_i) = \frac{4}{7} + 2\cdot\frac{1}{7} = \frac{6}{7}$$

$$\operatorname{var}[X] = \frac{8}{7} - \frac{36}{49} = \frac{20}{49}$$

$$\operatorname{Var}[X] = E[X^2] - (E[X])^2; \quad E[Y] = E[g(X)] = \sum g(x_i)\cdot P_X(x_i); \quad E[X^2] = 1\cdot\frac{4}{7} + 4\cdot\frac{1}{7} = \frac{8}{7}$$

$$P(A)=P(A|B)=\frac{P(A\cap B)}{P(B)} \to P(A\cap B)=P(A)\cdot P(B)$$

Пример 2

$$\left(\underset{A}{2};\underset{B}{2};\underset{C}{2}\right)$$

$$P_X(X=0)=0.7\cdot 0.6\cdot 0.2=0.084$$

$$P_X(X=1)=0.3\cdot 0.6\cdot 0.2+0.7\cdot 0.4\cdot 0.2+0.7\cdot 0.6\cdot 0.8=0.428$$

Цены на акции компаний A, B и C растут независимо друг от друга с вероятностями 0.3, 0.4 и 0.8 соответственно. Пусть X – число тех компаний среди этих трех, чьи акции выросли.

1. Постройте ряд распределения, найдите математическое ожидание и стандартное отклонение.
2. Найдите вероятность того, что X отклонится от своего математического ожидания более чем на одно стандартное отклонение.

$$P_X(X=3)=0.3\cdot 0.4\cdot 0.8=0.096$$

x	0	1	2	3
P_X	0.084	0.428	0.392	0.096

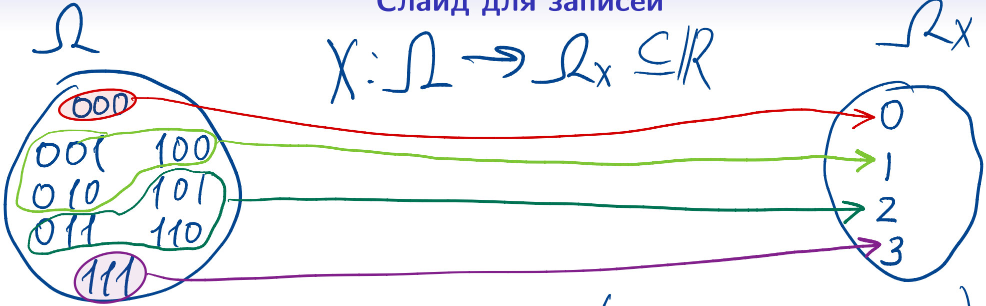
$$E[X]=\sum x_i P_X(X=x_i)=1.5$$

$$E[X^2]=\sum x_i^2\cdot P(X=x_i)=2.86$$

$$var[X]=2.86-2.25=0.61;\quad std=\sqrt{0.61}\approx 0.78$$

Слайд для записей

$$X:\Omega\to\Omega_X\subseteq\mathbb{R}$$



$$P(|X-E[X]|>StD(X)) = 1 - P(|X-E[X]|\le StD(X)) = 1-0.82 = 0.18$$

$$= P(|X-1.5|\le 0.78) = P(1.5-0.78\le X\le 1.5+0.78) =$$

$$= P(0.72\le X\le 2.28) = P_X(\{1,2\}) = 0.428+0.392 = 0.82$$

Пример 3

Пусть X - дискретная случайная величина с $E[X] = a$ и $Var[X] = b^2 \neq 0$. Мы строим новую случайную величину, которая является функцией от X:

$$Y = g(X) = \frac{X - a}{b}$$

Найдите $E[Y]$ и $Var[Y]$.

$$E[Y] = E[g(X)] = \sum g(x_i) \cdot P_X(X=x_i) = \sum_{i=1}^{n} \left(\frac{x_i - a}{b}\right) \cdot P_X(X=x_i) =$$

$$= \frac{1}{b} \underbrace{\boxed{\sum x_i P_X(X=x_i)}}_{E[X]} - \frac{1}{b} \sum a \cdot P_X(X=x_i) = \frac{E[X]}{b} - \frac{a}{b} \overbrace{\boxed{\sum P_X(x_i)}}^{=1} = 0$$

Пример 3

Решение

Пусть X - дискретная случайная величина с $E[X] = a$ и $Var[X] = b^2 \neq 0$. Мы строим новую случайную величину, которая является функцией от X:

$$\boxed{Y = g(X) = \frac{X - a}{b}}$$

Найдите $E[Y]$ и $Var[Y]$.
Решение.

$$E[Y] = E\left[\frac{X-a}{b}\right] = E\left[\frac{X}{b} + \frac{-a}{b}\right] = \frac{1}{b}E[X] - \frac{a}{b} = \frac{a}{b} - \frac{a}{b} = 0$$

$$Var[Y] = Var\left[\frac{X-a}{b}\right] = Var\left[\frac{X}{b} + \frac{-a}{b}\right] = \frac{1}{b^2}Var[X] = \frac{b^2}{b^2} = 1$$

Распределение Бернулли

Начнем с простейшего случая - **одного случайного эксперимента** с двумя возможными исходами.

Definition

Распределение Бернулли - это распределение случайной величины X, которая принимает только два значения:

$$X = \begin{cases} 1 & \text{с вероятностью } p \text{ (''успех'')} \\ 0 & \text{с вероятностью } q = 1 - p \text{ (''неудача'')} \end{cases}$$

Обозначается как $X \sim Bern(p)$, где $p \in [0, 1]$ - параметр распределения.

Характеристики распределения Бернулли

Основные характеристики

Для случайной величины $X \sim Bern(p)$:

- Математическое ожидание:

$$E[X] = 0 \cdot (1-p) + 1 \cdot p = p$$

- Дисперсия:

$$Var[X] = E[X^2] - (E[X])^2 = p - p^2 = p(1-p) = pq$$

- Стандартное отклонение:

$$\sigma = \sqrt{p(1-p)} = \sqrt{pq}$$

Примеры распределения Бернулли

1. **Подбрасывание монеты**: "Орел" = 1, "Решка" = 0, $p = 0.5$ (для честной монеты)
2. **Производственный контроль**: "Дефектное изделие" = 1, "Качественное" = 0
3. **Медицинский тест**: "Положительный результат" = 1, "Отрицательный" = 0
4. **Маркетинг**: "Покупка товара" = 1, "Отказ от покупки" = 0

От одного испытания к процессу

- Что если мы хотим провести несколько независимых испытаний Бернулли?
- Например: подбросить монету 10 раз, проверить 100 изделий на дефекты, провести 50 медицинских тестов.
- Это приводит нас к понятию процесса Бернулли - последовательности независимых испытаний Бернулли.

Процесс Бернулли

Definition

Процесс Бернулли - это последовательность независимых одинаково распределенных случайных величин $X_1, X_2, \dots, X_n$, где каждая $X_i \sim Bern(p)$:

$$X_i = \begin{cases} 1 & \text{с вероятностью } p \text{ ("успех")} \\ 0 & \text{с вероятностью } 1-p \text{ ("неудача")} \end{cases}$$

Ключевое отличие от одного испытания: теперь у нас есть **последовательность** результатов.

Свойства процесса Бернулли

1. **Независимость**: Результат каждого испытания X_i не зависит от результатов других испытаний
2. **Два исхода**: Каждое испытание может закончиться только успехом (1) или неудачей (0)
3. **Постоянная вероятность**: Вероятность успеха p не изменяется от испытания к испытанию

Примеры процесса Бернулли

1. **Серия подбрасываний монеты**: Последовательность $(X_1, X_2, \ldots, X_{10})$ для 10 подбрасываний
2. **Контроль качества**: Проверка партии из n изделий, где каждое может быть дефектным независимо от других
3. **Клинические испытания**: Тестирование лекарства на n пациентах, где каждый может показать положительную реакцию

$(0, 0, 1, 1, 0)$

$(1, 0, 0, 1, 0)$

$(1, 1, 1, 1, 1)$

Биномиальное распределение

- Теперь, когда у нас есть процесс Бернулли, возникает естественный вопрос: **“Сколько успехов мы получим в n испытаниях?”**

Definition

Биномиальное распределение описывает количество успехов в n независимых испытаниях Бернулли с вероятностью успеха p. Если $X_1, X_2, \dots, X_n$ - процесс Бернулли с параметром p, то случайная величина:

$$Y = X_1 + X_2 + \dots + X_n$$

$(X_1, \dots, X_n)$

имеет биномиальное распределение и обозначается $Y \sim Bin(n, p)$. Y может принимать значения $0, 1, 2, \dots, n$ (от 0 до n успехов).

Функция вероятности биномиального распределения

$n=7 \quad k=3 \quad p$

$\underbrace{p\ \ p\ \ p}_{p^3}$ $\underbrace{\quad\quad\quad\quad}_{q^4}$

Вероятность получить ровно k успехов в n испытаниях:

$$P(Y=k) = C_n^k p^k (1-p)^{n-k}$$

Интуиция формулы

- p^k - вероятность k успехов
- $(1-p)^{n-k}$ - вероятность $(n-k)$ неудач
- C_n^k - количество способов выбрать k позиций из n для размещения успехов

$q\ p\ q\ q\ q\ p\ p = p^3 q^4$

$(1,2,3), (1,2,4),$
$(2,6,7), (5,6,7)$

Характеристики биномиального распределения

Основные характеристики

Для случайной величины $Y \sim Bin(n, p)$:

Математическое ожидание:

$$E[Y] = np$$

Дисперсия:

$$Var[Y] = np(1-p)$$

Стандартное отклонение:

$$\sigma = \sqrt{np(1-p)}$$

Вывод математического ожидания:

$E[Y] = E[X_1 + X_2 + ... + X_n] = E[X_1] + E[X_2] + ... + E[X_n] = p + p + ... + p = np$

Визуализация биномиального распределения

Форма биномиального распределения зависит от параметров n и p. Рассмотрим, как влияет значение p на распределение при фиксированном $n = 20$:

Сравнение различных значений p

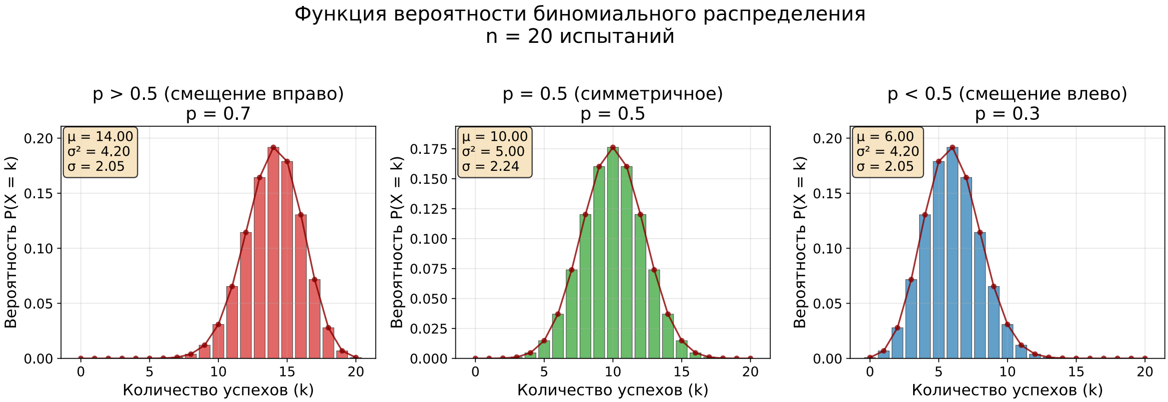


Рис. 1: Сравнение биномиального распределения для разных значений p

Визуализация биномиального распределения

Наблюдения

$\mathbf{p < 0.5}$ (например, $p = 0.3$): Распределение смещено влево, большинство значений сосредоточено в области малых k

$\mathbf{p = 0.5}$: Распределение симметрично относительно среднего значения

$\mathbf{p > 0.5}$ (например, $p = 0.7$): Распределение смещено вправо, большинство значений сосредоточено в области больших k