

# Домашняя работа по Математической Статистике

Митин Арсений (3 курс, СКБ171)

26.10.2019

## Содержание

<b>1</b>	<b>Задание №1</b>	<b>3</b>
1.1	Выбор распределений . . . . .	3
1.2	Описание основных характеристик распределений . . . . .	3

## 1. Задание №1

### 1.1. Выбор распределений

Выбранные распределения:

- Дискретное: *гипергеометрическое*
- Непрерывное: *нормальное*

### 1.2. Описание основных характеристик распределений

#### Гипергеометрическое распределение

Гипергеометрическое распределение - дискретное распределение, описывающее вероятность события, при котором ровно  $k$  из  $n$  случайно выбранных элементов окажутся *помеченными*, при этом выборка осуществляется из множества мощности  $N$ , в котором присутствует  $m$  помеченных элементов. Считается, что каждый из элементов может быть выбран с одинаковой вероятностью  $\frac{1}{N}$ . Запишем это формально:

$$N \in \mathbb{N}, m \in \overline{0, N}, n \in \overline{0, N}, \\ k \in \overline{0, n}$$

Тогда  $HG(D, N, n)$  описывает вероятность события, при котором ровно  $k$  из  $n$  элементов выборки окажутся *помеченными*:

$$\{\xi \sim HG(N, m, n)\} \\ \Updownarrow \\ \left\{ \mathbb{P}(\xi = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} \right\}$$

**Математическое ожидание гипергеометрического распределения** По определению, математическое ожидание случайной величины  $\xi$  -- это ее  $1^{\text{й}}$  начальный момент. Для начала, найдем  $k^{\text{й}}$  начальный момент для  $\xi$ :

$$\mathbb{E}[\xi^r] = \sum_{k=0}^n k^r \cdot \mathbb{P}(\xi = k) = \sum_{k=0}^n k^r \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} \quad (1)$$

Можем считать, что сумма берется при  $k$  от 1 до  $n$ , так как слагаемое при  $k = 0$  будет равно 0. Заметим, что

$$\begin{aligned} k \binom{m}{k} &= k \frac{m!}{k!(m-k)!} = \\ &= k \frac{m \cdot (m-1)!}{k \cdot (k-1)! \cdot (m-k)!} = \\ &= m \frac{(m-1)!}{(k-1)! \cdot (m-1-(k-1))!} = \\ &= m \binom{m-1}{k-1} \end{aligned} \quad (2)$$

и, как следствие,

$$\binom{N}{n} = \frac{1}{n} \cdot n \cdot \binom{N}{n} = \frac{1}{n} N \binom{N-1}{n-1} \quad (3)$$

Подставим 2 и 3 в 1:

$$\mathbb{E}[\xi^r] = \frac{n \cdot m}{N} \sum_{k=1}^{r-1} \frac{\binom{m-1}{k-1} \binom{N-m}{n-k}}{\binom{N-1}{n-1}}$$

Положим  $j := k - 1$  и изменим индекс суммирования с на  $j = \overline{0, n-1}$ . Заметим, что  $n - k = n - (j + 1) = (n - 1) - j$  и  $N - m = (N - 1) - (m - 1)$ :

$$\mathbb{E}[\xi^r] = \frac{n \cdot m}{N} \sum_{j=0}^{n-1} (j+1)^{r-1} \frac{\binom{m-1}{j} \binom{(N-1)-(m-1)}{(n-1)-j}}{\binom{N-1}{n-1}}$$

Заметим, что выделенная часть выражения может быть записана, как  $\mathbb{E}[(\theta + 1)^{r-1}]$ , где  $\theta \sim HG(N - 1, m - 1, n - 1)$ . Следовательно,

$$\mathbb{E}[\xi^r] = \frac{n \cdot m}{N} \mathbb{E}[(\theta + 1)^{r-1}] \quad (4)$$

Таким образом,

$$\boxed{\mathbb{E}[\xi] = \frac{n \cdot m}{N}} \quad (5)$$

**Дисперсия гипергеометрического распределения** По определению дисперсии,

$$\text{Var}[\xi] = \mathbb{E}[(\xi - \mathbb{E}[\xi])^2] = \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2 \quad (6)$$

Выведем 2<sup>й</sup> начальный момент из 4:

$$\mathbb{E}[\xi^2] = \frac{n \cdot m}{N} \mathbb{E}[\theta + 1] = \frac{n \cdot m}{N} \left( \frac{(n-1)(m-1)}{N-1} + 1 \right) \quad (7)$$

Подставим 5 и 7 в 6:

$$\begin{aligned} \text{Var}[\xi] &= \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2 = \\ &= \frac{n \cdot m}{N} \left( \frac{(n-1)(m-1)}{N-1} + 1 \right) - \left( \frac{n \cdot m}{N} \right)^2 = \\ &= \frac{n \cdot m}{N} \left( \frac{(n-1)(m-1)}{N-1} + 1 - \frac{n \cdot m}{N} \right) \end{aligned}$$

Таким образом,

$$\boxed{\mathbb{E}[\xi] = \frac{n \cdot m}{N} \left( \frac{(n-1)(m-1)}{N-1} + 1 - \frac{n \cdot m}{N} \right)}$$

**Производящая функция гипергеометрического распределения** По определению производящей функции,

$$M_\xi(t) = \mathbb{E}[e^{t\xi}]$$

То есть,

$$\begin{aligned} M_\xi(t) &= \sum_{k=0}^n e^{tk} \mathbb{P}(\xi = k) = \\ &= \sum_{k=0}^n e^{tk} \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} \end{aligned}$$

TODO

**Характеристическая функция гипергеометрического распределения** TODO

**Гистограмма вероятностей гипергеометрического распределения** Построим гистограмму вероятностей для  $k \in \overline{0, n}$ :

```

1  from datetime import datetime
2  from abc import ABC, abstractmethod
3  from typing import List
4
5  import numpy as np
6  import scipy as sp
7  import plotly
8  import plotly.figure_factory as ff
9  import plotly.graph_objs as go
10
11 class HG(object):
12     def __init__(self, N: int, m: int, n: int):
13         self.N = N
14         self.m = m
15         self.n = n
16
17     def p(self, k: int) -> float:
18         return sp.special.comb(self.m, k) * sp.special.comb(self.N-self.m, self.n-k) / sp.special.comb(self.N, self.n)
19
20     def __str__(self) -> str:
21         return f'HG({self.N}, {self.m}, {self.n})'
22
23 xi = HG(30, 15, 20)
24 hist_data_x = range(xi.n+1)
25 hg_hist_fig = go.Figure(
26     data=(
27         go.Scatter(
28             x=list(hist_data_x),
29             y=list(map(xi.p, hist_data_x)),
30             mode='markers',
31         ),
32     ),
33     layout=go.Layout(
34         title=go.layout.Title(
35             text=r'\xi \sim ' + str(xi) + '$',
36             x=.5,
37         ),
38         yaxis=go.layout.YAxis(
39             title=go.layout.yaxis.Title(
40                 text=r'\mathbb{P}(\xi=k)$',
41             ),
42         ),
43         xaxis=go.layout.XAxis(
44             title=go.layout.xaxis.Title(
45                 text=r'$k$',
46             ),
47         ),
48         paper_bgcolor='rgba(0,0,0,0)',
49     ),
50 )
51 plotly.offline.iplot(hg_hist_fig)

```

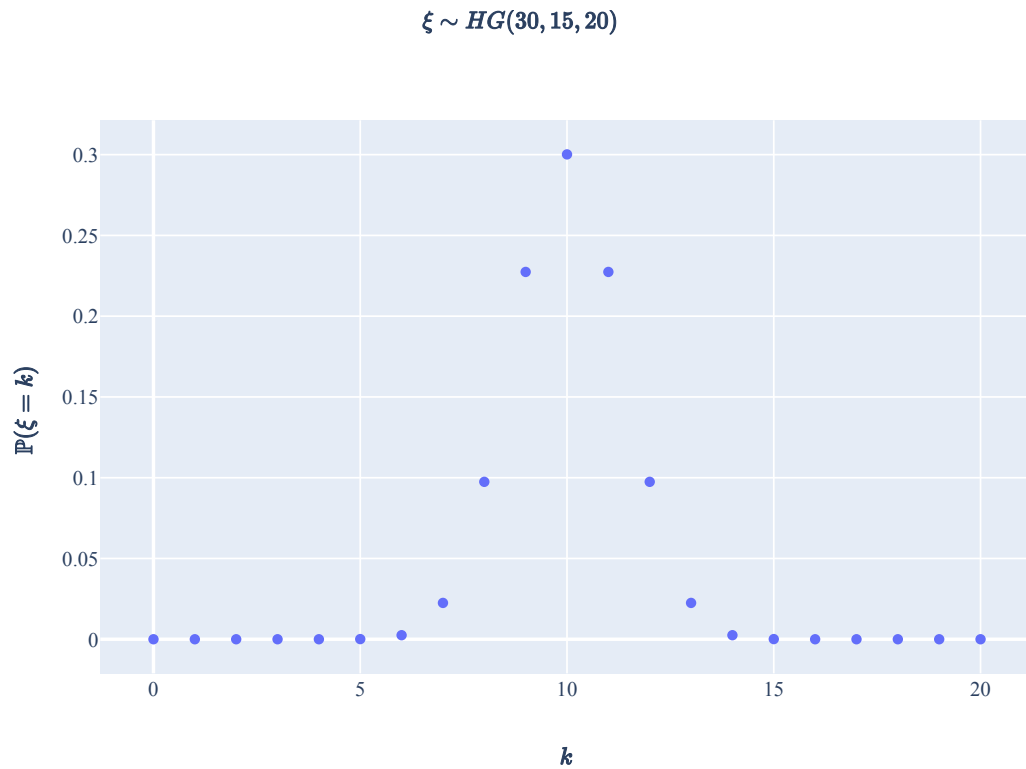


Рис. 1: Гистограмма вероятностей гипергеометрического распределения

**Функция распределения гипергеометрического распределения** По определению, функция распределения  $F_\xi(k) = \mathbb{P}(\xi < k)$ . Событие  $\{\xi < k\} = \bigcup_{i=0}^{k-1} \{\xi = i\}$ . События  $\{\xi = i\} \forall i \in \overline{0, k-1}$  являются попарно несовместными. То есть  $\forall i, j \in \overline{0, k-1} : i \neq j$  выполняется  $\{\xi = i\} \cap \{\xi = j\} = \emptyset$ . Из этого следует, что

$$\mathbb{P}(\xi < k) = \sum_{i=0}^{k-1} \mathbb{P}(\xi = i)$$

Подставим TODO в это выражение и получим:

$$F_\xi(k) = \sum_{i=0}^{k-1} \mathbb{P}(\xi = i) = \sum_{i=0}^{k-1} \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}}$$

Построим график этой функции, учитывая, что аргументом  $k$  должно быть натуральное число, не превосходящее  $n$ :

TODO

**Нормальное распределение**

Нормальное распределение - непрерывное распределение, описывающее поведение величины отклонения измеряемого значения  $x$  от истинного значения  $\mu$  (которое является математическим ожиданием) и в рамках некоторого разброса  $\sigma$  (среднеквадратичного отклонения). Запишем это формально:

$$\begin{aligned} \{\eta \sim N(\mu, \sigma^2)\} \\ \Updownarrow \\ \left\{ \begin{aligned} F_\eta(x) = \mathbb{P}(\eta < x) &= \int_{-\infty}^x f_\eta(x) dx, \\ \text{где } f_\eta(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \end{aligned} \right\} \end{aligned}$$

$f_\eta(x)$  называется плотностью вероятности.

**Математическое ожидание нормального распределения** Найдем математическое ожидание  $\eta \sim N(\mu, \sigma^2)$ :

$$\begin{aligned} \mathbb{E}[\eta] &= \int_{-\infty}^{+\infty} x \cdot f_\eta(x) dx = \\ &= \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \end{aligned}$$

Сделаем замену  $t = \frac{x-\mu}{\sqrt{2}\sigma}$ :

$$\begin{aligned} \mathbb{E}[\eta] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma\sqrt{2}t + \mu) e^{-t^2} d\left(\frac{x-\mu}{\sqrt{2}\sigma}\right) = \\ &= \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t e^{-t^2} dt + \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \\ &= \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \left( \int_{-\infty}^0 t e^{-t^2} dt - \int_0^{+\infty} t e^{-t^2} dt \right) + \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \\ &= \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt \end{aligned}$$

Заметим, что получившееся выражение содержит интеграл, который может быть сведен к интегралу [Эйлера-Пуассона](#):

$$\int_{-\infty}^{+\infty} e^{-t^2} dt = 2 \int_0^{+\infty} e^{-t^2} dt = \sqrt{\pi}$$

Таким образом,

$$\boxed{\mathbb{E}[\eta] = \mu}$$

**Дисперсия нормального распределения** Подставим TODO в определение дисперсии TODO:

$$\begin{aligned}\text{Var}[\eta] &= \mathbb{E}[(\eta - \mu)^2] = \\ &= \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f_{\eta}(x) dx = \\ &= \int_{-\infty}^{+\infty} (x - \mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx\end{aligned}$$

Сделаем ту же замену переменной  $t = \frac{x-\mu}{\sqrt{2}\sigma}$ , тогда  $x = t\sqrt{2}\sigma + \mu$  и:

$$\begin{aligned}\text{Var}[\eta] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sqrt{2}\sigma)^2 t^2 e^{-t^2} d(t\sqrt{2}\sigma + \mu) = \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^2 e^{-t^2} dt\end{aligned}$$

Проинтегрируем по частям:

$$\begin{aligned}\text{Var}[\eta] &= \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t 2t e^{-t^2} dt = \\ &= \frac{\sigma^2}{\sqrt{\pi}} \left( -te^{-t^2} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-t^2} dt \right)\end{aligned}$$

Здесь снова появляется интеграл [Эйлера-Пуассона](#) и, в итоге, получаем:

$$\boxed{\text{Var}[\eta] = \sigma^2}$$

То есть,  $\sigma$  является среднеквадратичным отклонением.

```
1 | def f() → int:
2 |     if 1 ≠ 2:
3 |         print("mitinarseny@gmail.com → abc")
```