Биномиальное и непрерывное распределения

Распределение Бернулли

Определение

> **Биномиальное распределение**
>
> — распределение количества успехов в схеме Бернулли

Лемма

> **Свойства биномиального распределения**
>
> Пусть X_k - бернулиевские величины, n - число испытаний, p - вероятность успеха
>
> 1. $Y = X_1 + \ldots + X_n$
> 2. $\mathbb{E}Y = np$
> 3. $\mathbb{V}Y = np(1-p)$

Лемма

> **Вероятность k успехов в схеме Бернулли**
>
> $$P(Y=k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

В Python:

```
import scipy.stats as sps
dist = sps.binom(n,p)
# dist.pmf(x), dist.cdf(x), dist.ppf(q), dist.rvs(sample\_size)
```

Определение

Выборочная пропорция или выборочная доля успехов

$$\hat{p} = \frac{Y}{n},$$

где Y - количество успехов, n- количество испытаний

Лемма

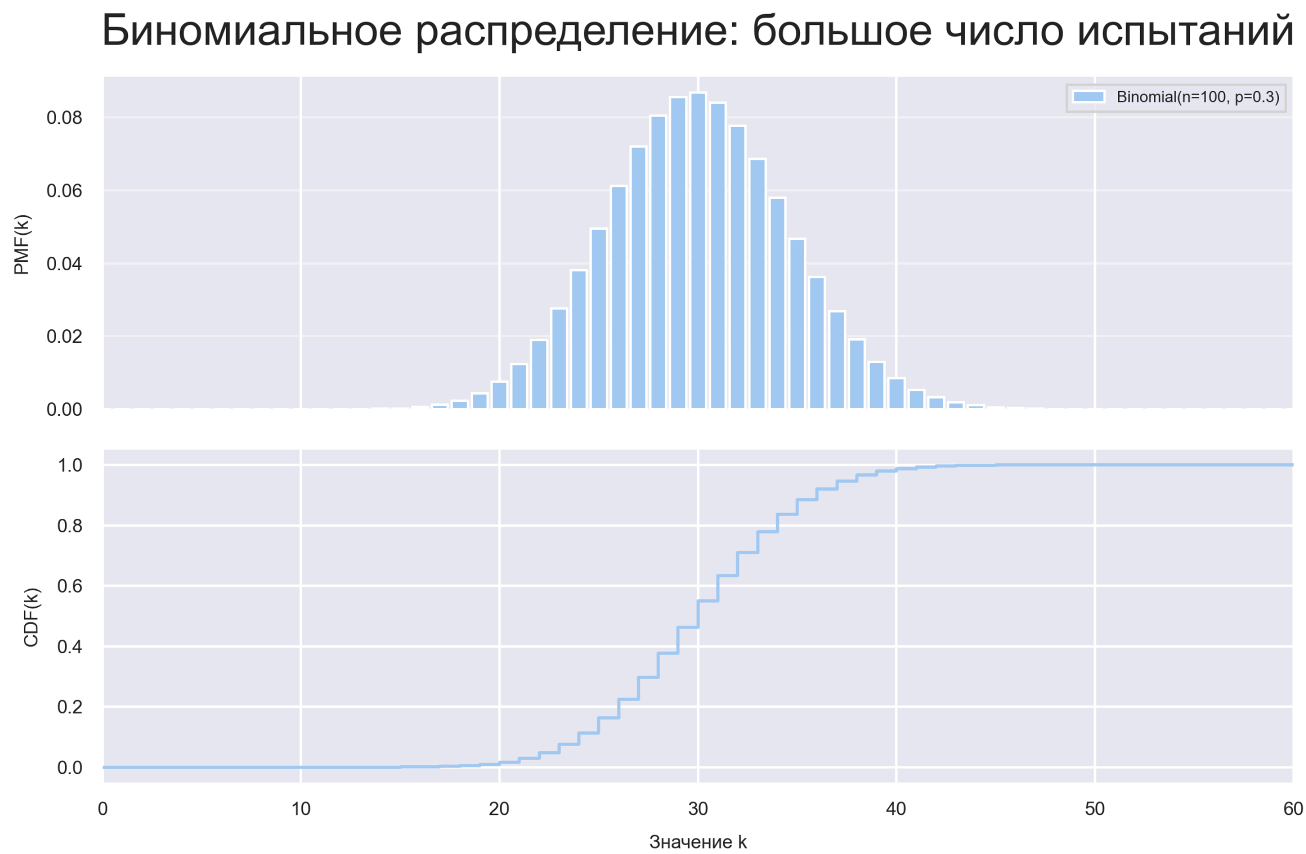
Свойства выборочной пропорции

1. $\hat{p} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, где X_i — результаты испытаний в схеме Бернулли
2. $\mathbb{E}(\hat{p}) = p$
3. $\mathbb{V}(\hat{p}) = \frac{p(1-p)}{n}$
4. $\mathbb{P}\left(\hat{p} = \frac{k}{n}\right) = C_n^k p^k (1-p)^{n-k}$, при $k = 0, 1, \ldots, n$.

Лемма

Свойства биномиального распределения

- Наиболее вероятны значения, близкие к матожиданию
- Вероятности убывают вправо и влево от матождидания
- График PMF несимметричный, кроме случая p=0.5
- График CDF - лестница со скачками в целочисленных точках. Часть из них настолько мала, что каждым из них в отдельности можно пренебречь



Лемма

> **Свойства биномиального распределения при увеличении числа испытаний**
>
> 1. График PMF становится более симметричным при любом p
> 2. Почти 100% вероятности сосредоточена в значениях вокруг матожидания на промежутке, размер которого составляет малую долю интервала всех возможных значений
> 3. График CDF становится все более гладким

Важно

Количество успехов не описывается биномиальным распределением, если:

- вероятность успеха неодинакова в разных испытаниях
  ИЛИ
- результаты испытаний зависимы

Непрерывные распределения

Определение

> **Плотность непрерывного распределения (PDF)**
>
> функция $\mathbb{R} \to \mathbb{R}$, такая что:
>
> 1. $pdf(x) \geq 0 \quad x \in \mathbb{R}$
> 2. Площадь под ее графиком равна 1

Определение

> **Непрерывное распределение**
>
> — распределение случайной величины, такое, что вероятность на любом промежутке задается с помощью функции pdf:
>
> $$P(a \leq X \leq b) = \text{площадь под графиком pdf на отрезке } [a; b]$$

Лемма

> **Свойства непрерывного распределения**
>
> - $P(X = c) = 0, \quad c \in \mathbb{R}$
> - $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$
> - $CDF(b) = P(X \leq b) = P(X < b)$
> - CDF непрерывна

Лемма

> **Связь pdf и cdf**
>
> $pdf(x) = (CDF(x))'$

Независимость непрерывных случайных величин

Определение

> **Независимость непрерывных случайных величин**
>
> Произвольные случайные величины X,Y называют **независимыми**, если для всех $a, b \in \mathbb{R}$ выполнено равенство:
>
> $$P(X \leq a, Y \leq b) = P(X \leq a) \cdot P(Y \leq b)$$

Матожидание непрерывной случайной величины

Лемма

> **Матожидание непрерывной случайной величины**
>
> Для неотрицательной случайной величины X матожидание $\mathbb{E}X$ равно площади между графиками $y = 1$ и $y = CDF(x)$, $x > 0$

Flashcards

tags: #flashcardsSTAT

Что такое Биномиальное распределение?
%
— распределение количества успехов в схеме Бернулли

Сформулируйте лемму: Свойства биномиального распределения
%
Пусть X_k - бернулиевские величины, n - число испытаний, p - вероятность успеха

1. $Y = X_1 + ... + X_n$
2. $\mathbb{Y} = np$
3. $\mathbb{V}Y = np(1-p)$

Сформулируйте лемму: Вероятность k успехов в схеме Бернулли
%

$$P(Y = k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

Что такое Выборочная пропорция или выборочная доля успехов?
%

$$\hat{p} = \frac{Y}{n},$$

где Y - количество успехов, n- количество испытаний
Сформулируйте лемму: Свойства выборочной пропорции
%

1. $\hat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, где X_i — результаты испытаний в схеме Бернулли
2. $\mathbb{E}(\hat{p}) = p$
3. $\mathbb{V}(\hat{p}) = \frac{p(1-p)}{n}$
4. $\mathbb{P}\left(\hat{p} = \frac{k}{n}\right) = C_n^k p^k (1-p)^{n-k}$, при $k = 0, 1, \ldots, n$.

Сформулируйте лемму: Свойства биномиального распределения
%

- Наиболее вероятны значения, близкие к матожиданию
- Вероятности убывают вправо и влево от матождидания
- График PMF несимметричный, кроме случая p=0.5
- График CDF - лестница со скачками в целочисленных точках. Часть из них настолько мала, что каждым из них в отдельности можно пренебречь

Сформулируйте лемму: Свойства биномиального распределения при увеличении числа испытаний
%

1. График PMF становится более симметричным при любом p
2. Почти 100% вероятности сосредоточена в значениях вокруг матожидания на промежутке, размер которого составляет малую долю интервала всех возможных значений
3. График CDF становится все более гладким

Непрерывные распределения

Что такое Плотность непрерывного распределения (PDF)?
%
функция $\mathbb{R} \to \mathbb{R}$, такая что:

1. $pdf(x) \geq 0 \quad x \in \mathbb{R}$
2. Площадь под ее графиком равна 1

Что такое Непрерывное распределение?
%
— распределение случайной величины, такое, что вероятность на любом промежутке задается с помощью функции pdf:

$$P(a \leq X \leq b) = \text{плоащь под графиком pdf на отрезке } [a; b]$$

Сформулируйте лемму: Свойства непрерывного распределения
%

- $P(X = c) = 0, \quad c \in \mathbb{R}$
- $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$
- $CDF(b) = P(X \le b) = P(X < b)$
- CDF непрерывна

Сформулируйте лемму: Связь pdf и cdf
%
$pdf(x) = (CDF(x))'$

Независимость непрерывных случайных величин

Что такое Независимость непрерывных случайных величин?
%
Произвольные случайные величины X,Y называют **независимыми**, если для всех $a, b \in \mathbb{R}$ выполнено равенство:

$$P(X \le a, Y \le b) = P(X \le a) \cdot P(Y \le b)$$

Матожидание непрерывной случайной величины

Сформулируйте лемму: Матожидание непрерывной случайной величины
%
Для неотрицательной случайной величины X матожидание $\mathbb{E}X$ равно площади между графиками $y = 1$ и $y = CDF(x)$, $x > 0$