

Отчет по второму заданию

Что надо сделать:

- 1) Найти истинное значение медианы экспоненциальной выборки (по теории)
- 2) Сгенерировать выборку из 25 значений
- 3) найти доверительный интервал выборки
- 4) проверить попадает ли истинное значение в этот интервал
- 5) повторить 1000 раз
- 6) подсчитать сколько процентов попаданий из 1000

Прodelать все тоже самое с выборкой из 10000 значений и сделать выводы.

Выполнение:

1)

Для $\text{Exp}(1)$ медиана =

$$\text{med} = \frac{\ln(2)}{\lambda} = \ln(2) \approx 0.693$$

2)

```
sample = np.random.exponential(scale=1, size=sample_size)
```

3)

Так как мы **не знаем точного распределения медианы** в выборке, мы используем **ЦПТ (центральную предельную теорему)** и говорим:

$$\hat{m} \sim \mathcal{N}(\text{med}, SE)$$

Где:

- \hat{m} — оценка медианы (из выборки)
- SE — стандартная ошибка (можно аппроксимировать как $\frac{1}{f(m)} \cdot \sqrt{\frac{1}{4n}}$, где $f(m)$ — значение плотности PDF в медиане)

Для экспоненциального распределения плотность:

$$f(x) = \lambda e^{-\lambda x} \Rightarrow f(\ln 2) = e^{-\ln 2} = \frac{1}{2}$$

Тогда стандартная ошибка:

$$SE = \frac{1}{0.5} \cdot \sqrt{\frac{1}{4n}} = 2 \cdot \sqrt{\frac{1}{4n}} = \frac{1}{\sqrt{n}}$$

```
sample_median = np.median(sample)

# Стандартная ошибка медианы
se = 1 / np.sqrt(sample_size)

# z-значение для доверительного интервала
z = norm.ppf(1 - alpha / 2)

lower = sample_median - z * se
upper = sample_median + z * se
```

4)

```
if lower <= true_median <= upper:
    covered += 1
```

5)

```

2 usages
def simulate_ci_coverage(sample_size, true_median=np.log(2), n_trials=1000, alpha=0.05):
    covered = 0
    ci_bounds = []

    for _ in range(n_trials):
        sample = np.random.exponential(scale=1, size=sample_size)
        sample_median = np.median(sample)

        # Стандартная ошибка медианы
        se = 1 / np.sqrt(sample_size)

        # z-значение для доверительного интервала
        z = norm.ppf(1 - alpha / 2)

        lower = sample_median - z * se
        upper = sample_median + z * se

        ci_bounds.append((lower, upper))

        if lower <= true_median <= upper:
            covered += 1

```

6)

```

coverage_rate = covered / n_trials
print(f"Доля интервалов, покрывших настоящую медиану ({true_median:.3f}): {coverage_rate:.3f}")
return ci_bounds

```

Результаты эксперимента:

```

Доля интервалов, покрывших настоящую медиану (0.693): 0.949
Доля интервалов, покрывших настоящую медиану (0.693): 0.953

```

Верхнее значение для выборки $n = 25$ а нижнее $n = 10000$

Выводы:

При увеличении объёма выборки стандартная ошибка оценки медианы уменьшается, поэтому доверительные интервалы становятся уже.

Однако уровень доверия (95%) сохраняется — это свойство асимптотических интервалов.

В эксперименте результаты слегка флуктуируют из-за случайности, но при большом объёме интервалы становятся **стабильнее и надёжнее**.