# ДЗ №6 по Теории Вероятностей

Соколов Игорь, группа 573

4 декабря 2017 г.

## 1

Предположим, что независимо $n$ раз кидается монетка с вероятностью выпадения орла в каждом опыте равной $p \in [0.1, 0.9]$. Сколько раз нужно кинуть монетку, чтобы оценка $\overline{p}(x) = \frac{\sum_{k=1}^{n} x_k}{n}$ с вероятностью $\gamma \leq 0.95$ отличалась от истинного значения $p$ не более, чем на величину $\delta = 0.01$? Применить неравенство Чебышева и предельную теорему, точность которой оценить неравенством Берри – Эссена. Сравнить результаты.

**Решение:**

1. *Неравенство Чебышёва:*

$$\mathbb{P}(|\xi - \mathbb{E}\xi| \geqslant x) \leqslant \frac{\mathbb{D}\xi}{x^2}$$

Применяем его для случайной величины $\overline{p}(x)$

$$\mathbb{P}(|\overline{p}(x) - p| \leqslant \delta) \geqslant 1 - \frac{\mathbb{D}\overline{p}(x)}{\delta^2}$$

Из условия независимости с.в. $x_1, x_2, \ldots, x_n$ имеем $\mathbb{D}\overline{p}(x) = \frac{\mathbb{D}x_k}{n} = \frac{p(1-p)}{n}$

Оценка $\overline{p}(x)$ с вероятностью $\gamma \geqslant 0.95$ гарантируется при

$$\gamma \leqslant 1 - \frac{\mathbb{D}\overline{p}(x)}{\delta^2} = 1 - \frac{p(1-p)}{n\delta^2}$$

$$\Rightarrow n \geqslant \frac{p(1-p)}{\delta^2(1-\gamma)}$$

Так как величина $p(1-p)$ достигает максимума при $p = 0.5$, данное неравенство заведомо выполнено при числе бросаний монетки

$$n = \frac{0.5^2}{\delta^2(1-\gamma)}$$

(a) *(ц.п.т) Если* $\mathbb{E}(x_k^2) < \infty$, *то* $\dfrac{\sum_{k=1}^{n} x_k - n\mathbb{E}x_k}{\sqrt{n\mathbb{D}x_k}} \xrightarrow{d} N(0,1)$, *при* $n \to \infty$

(b) *(неравенство Бэрри-Эссеена) Пусть дана бесконечная последовательность $X_n$ независимых одинаково распределённых случайных величин таких, что $\mathbb{E}(X_n) = 0, \mathbb{E}(X_n^2) = \sigma^2 > 0, \mathbb{E}(|X_n^3|) = \rho < \infty)$. Обозначим через $F_n$ распределение суммы вида* $\frac{\sum\limits_{i=1}^{n} X_i}{\sigma\sqrt{n}}$. Тогда для всех $x$ и $n$

$$\to |F_n(x) - \mathcal{N}(x)| \leqslant \frac{C\rho}{\sigma^3\sqrt{n}}$$

Из ц.п.т. и неравенство Бэрри-Эссена получаем

$$\sup_x \left| \mathbb{P}\left( \sum_{k=1}^{n} \frac{x_k - np}{\sigma\sqrt{n}} < x \right) - \mathcal{N}(x) \right| \leqslant \frac{C\mu^3}{\sigma\sqrt{n}}$$

где $C \leqslant 0.7056, \quad \sigma^2 = \mathbb{D}X_k = p(1-p), \quad \mathcal{N}(x) = \int\limits_{-\infty}^{x} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt$

$$\mu^3 = \mathbb{E}|x_k - p|^3 = p(1-p)^3 + (1-p)p^3 = \sigma^2(1 - 2\sigma^2)$$

Так как $\sum\limits_{k=1}^{n} \frac{x_k - np}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma}(\overline{p}(x) - p)$, то условие $|\overline{p}(x) - p| \leqslant \delta$ выполняется с вероятностью не менее

$$\mathcal{N}\left(\frac{\delta\sqrt{n}}{\sigma}\right) - \mathcal{N}\left(-\frac{\delta\sqrt{n}}{\sigma}\right) - \frac{2C\mu^3}{\sigma^3\sqrt{n}}$$

Разность $\mathcal{N}\left(\frac{\delta\sqrt{n}}{\sigma}\right) - \mathcal{N}\left(-\frac{\delta\sqrt{n}}{\sigma}\right)$ принимает наименьшее значение при максимальном $\sigma^2 = p(1-p)$, которое достигается при $p = 0.5$.
Также можно заметить, что наибольшее значение величины

$$\frac{\mu^3}{\sigma^3} = \frac{\sigma^2(1 - 2\sigma^2)}{\sigma^3} = \frac{1}{\sigma} - 2\sigma$$

достигается при наименьшем возможном $\sigma$, которое дает $p = 0.1$, откуда $\sigma = 0.3$.
Таким образом, оценка для $\overline{p}(x)$ с вероятностью $\gamma > 0.95$ гарантируется при

$$\gamma \leqslant \mathcal{N}(2\delta\sqrt{n}) - \mathcal{N}(-2\delta\sqrt{n}) - \frac{2 \cdot 0.7}{\sqrt{n}} \left( \frac{1}{0.3} - 0.6 \right)$$

Найдем такой минимальный $n$, чтобы оценка $|\overline{p}(x) - p(x)| \leqslant \delta$ выполнялась.

In [42]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as sts
import math
mu = 0.0
sigma = 1.0

# зададим нормально распределенную случайную величину
norm_rv = sts.norm(loc=mu, scale=sigma)
def N(x):
    return norm_rv.cdf(x)
```

In [43]:
```
delta = 0.01
n = 10000
gamma = 0.95
res = 0
while res<=gamma:
    res = N(2*delta*math.sqrt(n)) - N(-2*delta*math.sqrt(n)) - (2*0.7)/(math.sqrt(n))*(1.0/0.3-0.6)
    n += 1
print res, n
```

```
0.950001290419 14059
```

То есть уже при $n = 14058$ имеем

$$\mathcal{N}(2\delta\sqrt{n}) - \mathcal{N}(-2\delta\sqrt{n} = 0.98228$$

$$\frac{2 \cdot 0.7}{\sqrt{n}} \left( \frac{1}{0.3} - 0.6 \right) = 0.03227$$

$$\Rightarrow \gamma \geqslant 0.98228 - 0.03227 = 0.95000129 > 0.95$$

$\Rightarrow |\overline{p}(x) - p(x)| \leqslant \delta$ выполняется.
Итог:
Сравнивая результаты, мы убедились, что предельная теорема дает более точные оценки и гарантирует нужную точность оценки для $p$ уже при при числе бросаний монетки $n = 14058$, в то время как неравенство Чебышева — лишь при $n = 50000$.