

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**  
**Факультет прикладной математики и информатики**

**Отчет по лабораторной работе № 1**

**Выполнила:** Шелег Владислава Михайловна

3 курс 9 группа

**Преподаватель:** Вечерко Е. В.

МИНСК

2018

## 1. Задачи

1. Сгенерировать последовательность случайных величин длины  $n$  независимых случайных, имеющих гипергеометрическое распределение вероятностей
2. Построить гистограмму
3. Вычислить энтропию, построить статистическую оценку энтропии
4. Построить график зависимости среднеквадратической ошибки оценивания энтропии от длины последовательности

## 2. Теория

Дискретная случайная величина  $\xi$  имеет гипергеометрическое распределение вероятностей, то есть  $\mathcal{L}\{\xi\} = HG(m, n, k)$ , если:

$$P\{\xi = i\} = \frac{C_m^i C_n^{k-i}}{C_{m+n}^k}$$

Энтропия  $H$  некоторого гипергеометрического распределения равна

$$H = - \sum_{i=0}^{\infty} p_i \log_2 p_i = - \frac{\sum_{i=0}^{\infty} C_m^i C_n^{k-i} \log_2 (C_m^i C_n^{k-i})}{C_{m+n}^k}$$

Пусть есть некоторая последовательность случайных величин  $X = (x_1, \dots, x_n)$ . Введем относительную частоту события  $\{\xi = i\}$

$$\hat{p}_i = \frac{1}{n} \sum_{j=1}^n I\{x_j = i\}$$

Из курса ТВиМС известно, что  $\hat{p}_i \rightarrow p_i$  при  $n \rightarrow \infty$ .

Тогда оценку энтропии можно построить по правилу

$$\hat{H} = - \sum_{i=0}^{\infty} \hat{p}_i \log_2 \hat{p}_i$$

Заметим, что  $\hat{H} \rightarrow H$  при  $n \rightarrow \infty$ .

Среднеквадратической ошибкой оценивания называется величина

$$v(\hat{H}) = E_H \{(\hat{H} - H)^2\} = (\hat{H} - H)^2$$

### 3. Результаты

#### 3.1 $m = 150, n = 450, k = 150$

На рисунке 1 представлена гистограмма сгенерированной последовательности случайных величин длины  $n = 10000$ . Энтропия и ее оценка равны  $H = 4.247099$  и  $\hat{H} = 4.253125$ . Рисунок 2 отражает зависимость  $\nu(\hat{H})$  от  $n$ .

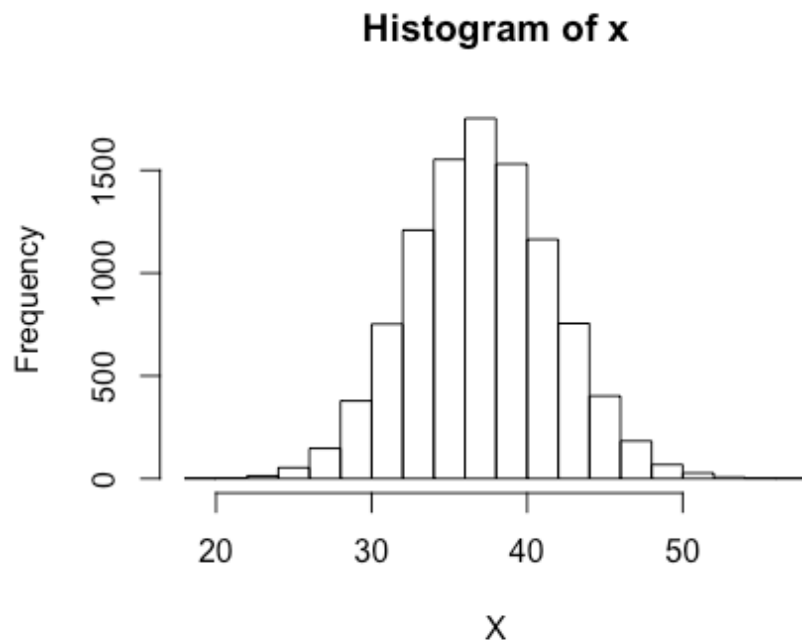


Рисунок 1 Гистограмма для  $n = 10000, m = 150, n = 450, k = 150$

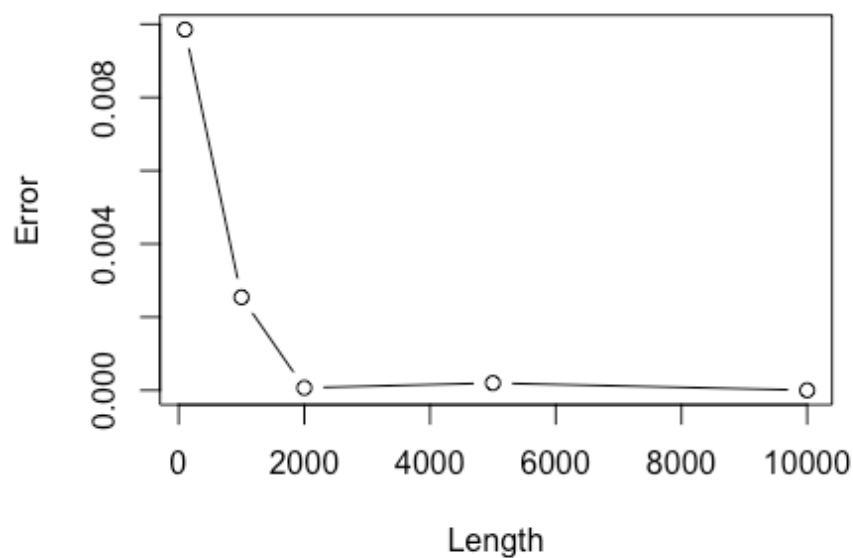


Рисунок 2 График зависимости среднеквадратичной ошибки от длины последовательности

3.2  $m = 100, n = 300, k = 50$

На рисунке 3 представлена гистограмма сгенерированной последовательности случайных величин длины  $n = 10000$ . Энтропия и ее оценка равны  $H = 3.562881$  и  $\hat{H} = 3.563189$ . Рисунок 4 отражает зависимость  $\nu(\hat{H})$  от  $n$ .

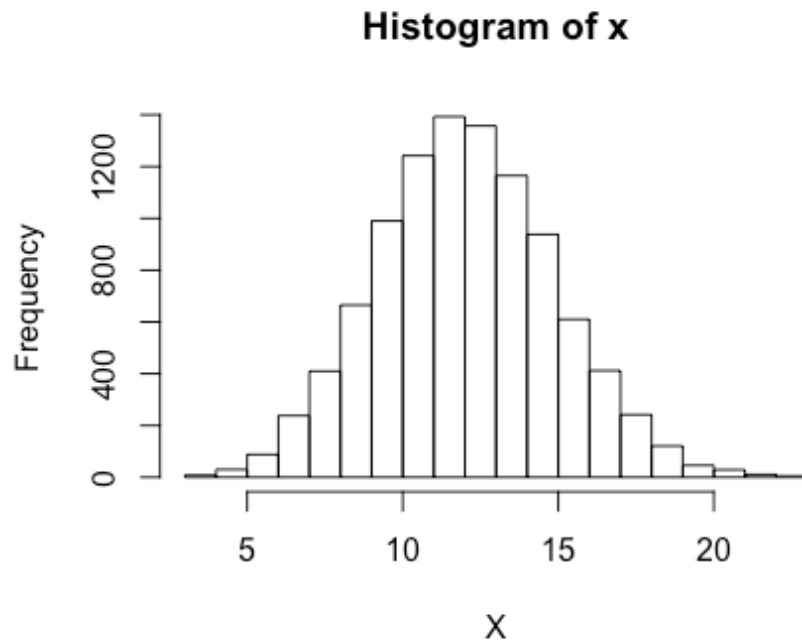


Рисунок 3 Гистограмма для  $n = 10000, m = 100, n = 300, k = 50$

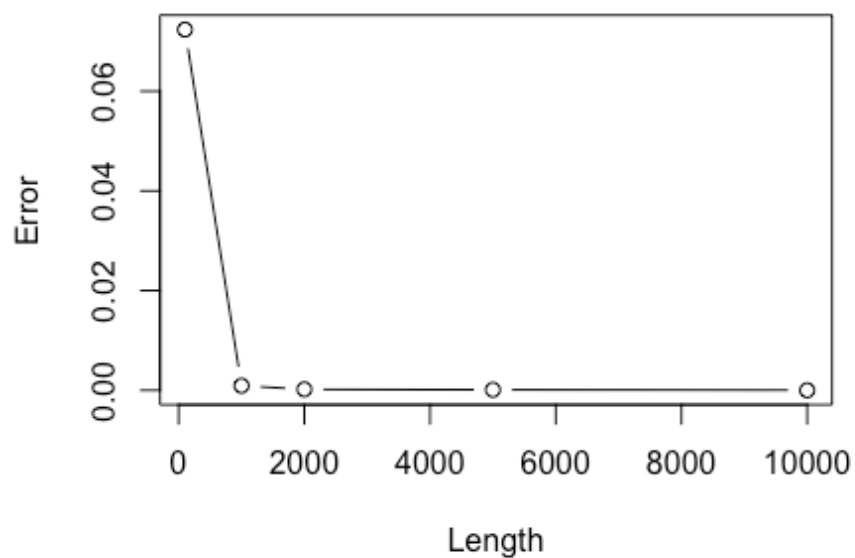


Рисунок 4 График зависимости среднеквадратичной ошибки от длины последовательности

3.3  $m = 500, n = 500, k = 150$

На рисунке 5 представлена гистограмма сгенерированной последовательности случайных величин длины  $n = 10000$ . Энтропия и ее оценка равны  $H = 4.544481$  и  $\hat{H} = 4.533195$ . Рисунок 6 отражает зависимость  $\nu(\hat{H})$  от  $n$ .

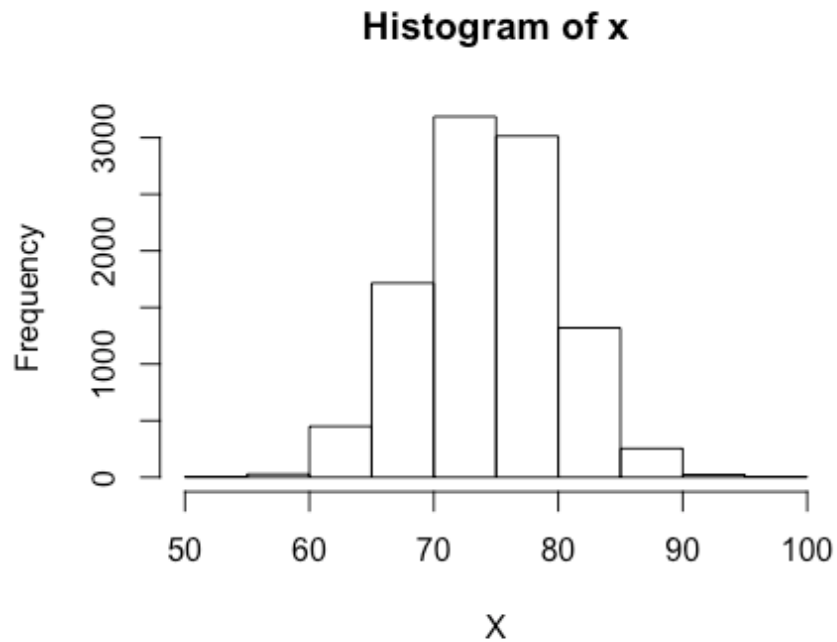


Рисунок 5 Гистограмма для  $n = 10000, m = 500, n = 500, k = 150$

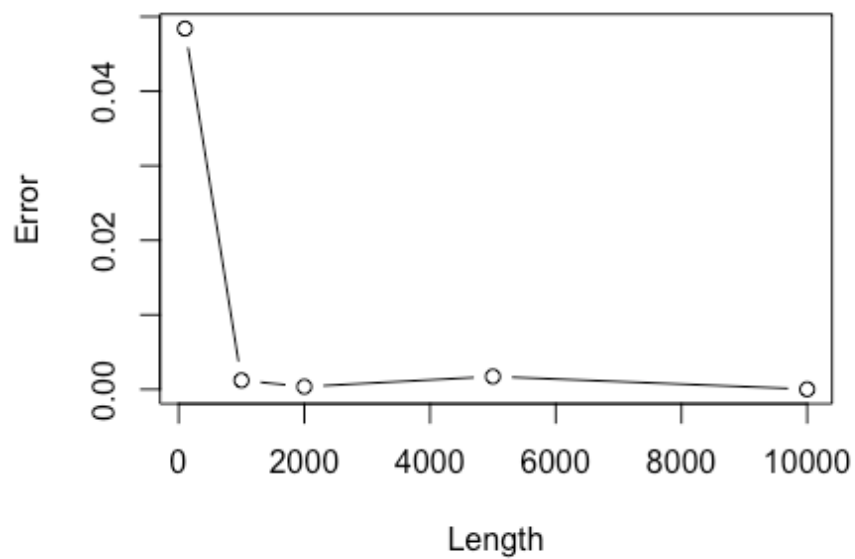


Рисунок 6 График зависимости среднеквадратичной ошибки от длины последовательности

3.4  $m = 50, n = 40, k = 20$

На рисунке 7 представлена гистограмма сгенерированной последовательности случайных величин длины  $n = 10000$ . Энтропия и ее оценка равны  $H = 3.025224$  и  $\hat{H} = 3.019488$ . Рисунок 8 отражает зависимость  $v(\hat{H})$  от  $n$ .

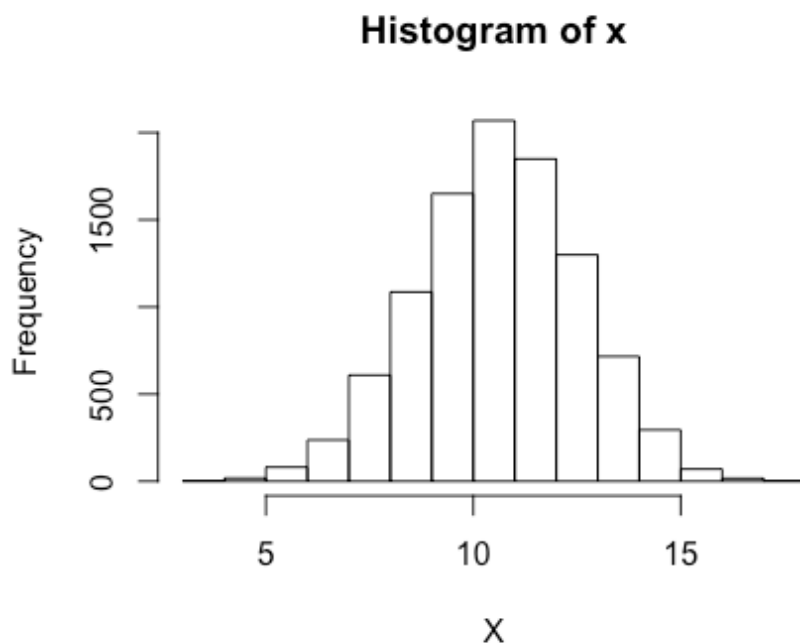


Рисунок 7 Гистограмма для  $n = 10000, m = 50, n = 40, k = 20$

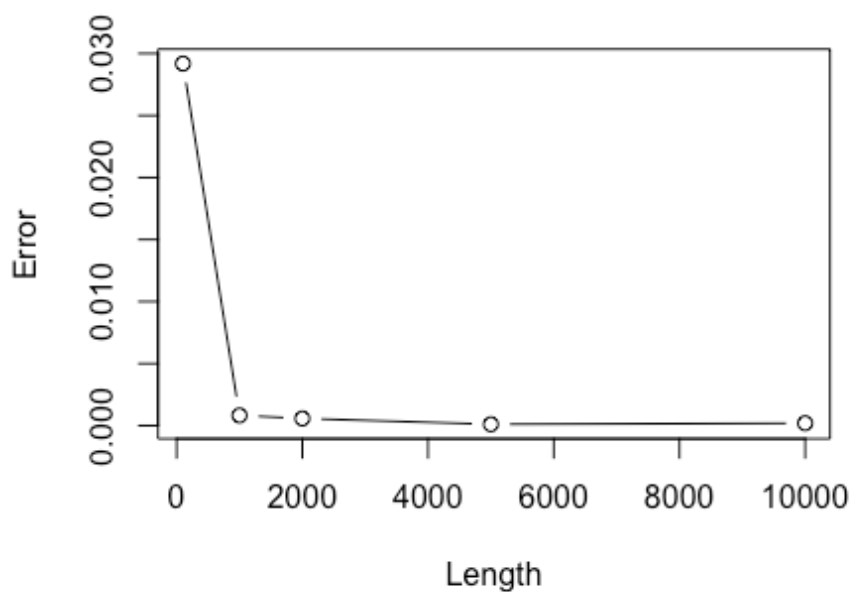


Рисунок 8 График зависимости среднеквадратичной ошибки от длины последовательности