

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Факультет прикладной математики и информатики

Отчет по лабораторной работе № 2

Выполнила: Шелег Владислава Михайловна

3 курс 9 группа

Преподаватель: Вечерко Е. В.

МИНСК

2018

1. Задачи

1. Найти энтропию $H\{\zeta_2|\zeta_1\}$ и оценку энтропии $\hat{H}\{\zeta_2|\zeta_1\}$
2. Построить график зависимости среднеквадратической ошибки оценивания энтропии от длины последовательности (количество прогонов $M=10^4$).
3. Построить семейство графиков зависимости $H\{\zeta_2|\zeta_1\}, \hat{H}\{\zeta_2|\zeta_1\}$ от $1 - \alpha - \beta$
4. Найти условные энтропии $H\{\zeta_2|\zeta_1 + \zeta_2\}, H\{\zeta_2|\zeta_1 \oplus \zeta_2\}$ и их оценки

2. Теория

Условная энтропия H случайной величины $\zeta \in A$ при условии события $\eta, b \in B$ – это величина

$$H\{\zeta | \eta\} = \sum_{b \in B} P\{\eta = b\} H\{\zeta | \eta = b\}$$

$$H\{\zeta | \eta = b\} = - \sum_{a \in A} P\{\zeta = a | \eta = b\} \log P\{\zeta = a | \eta = b\}$$

Если по условию даны параметры модели (π, α, β) и матрица переходов

$P = \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \beta & \beta \end{pmatrix} = (p_{ij})$, то условную энтропию можно найти по формуле

$$H\{\zeta_2 | \zeta_1\} = - (P\{\zeta_1 = 0\} * \\ * (P\{\zeta_2 = 0 | \zeta_1 = 0\} \log (P\{\zeta_2 = 0 | \zeta_1 = 0\} \\ + P\{\zeta_2 = 1 | \zeta_1 = 0\} \log (P\{\zeta_2 = 1 | \zeta_1 = 0\})) + P\{\zeta_1 = 1\} * \\ * (P\{\zeta_2 = 0 | \zeta_1 = 1\} \log (P\{\zeta_2 = 0 | \zeta_1 = 1\} \\ + P\{\zeta_2 = 1 | \zeta_1 = 1\} \log (P\{\zeta_2 = 1 | \zeta_1 = 1\})))$$

Подставив известные данные, получим, что:

$$H\{\zeta_2 | \zeta_1\} = -((1 - \pi)(\alpha \log \alpha + (1 - \alpha) \log(1 - \alpha)) + \pi((1 - \beta) \log(1 - \beta) + \beta \log \beta))$$

Оценка энтропии

$$\hat{H}\{\zeta_2 | \zeta_1\} = -((1 - \hat{\pi})(\hat{\alpha} \log \hat{\alpha} + (1 - \hat{\alpha}) \log(1 - \hat{\alpha})) + \hat{\pi}((1 - \hat{\beta}) \log(1 - \hat{\beta}) + \hat{\beta} \log \hat{\beta}))$$

$$\text{где } \hat{\pi} = \frac{m_n(\zeta_1=1)}{n}, \hat{\alpha} = \frac{m_{m_n(\zeta_1=0)(\zeta_2=0)}}{m_n(\zeta_1=0)}, \hat{\beta} = \frac{m_{m_n(\zeta_1=1)(\zeta_2=1)}}{m_n(\zeta_1=1)}$$

Для подсчета энтропии $H\{\zeta_2 | \zeta_1 + \zeta_2\}$ необходимо вычислить вероятности

$$P\{\zeta_1 + \zeta_2 = 0\} = (1 - \pi)\alpha$$

$$P\{\zeta_1 + \zeta_2 = 1\} = (1 - \pi)(1 - \alpha) + \pi(1 - \beta)$$

$$P\{\zeta_1 + \zeta_2 = 2\} = \pi\beta$$

С учетом этого, получим матрицу переходов

$$\begin{pmatrix} 1 & 0 \\ \frac{\pi(1 - \beta)}{\pi(1 - \beta) + (1 - \pi)(1 - \alpha)} & \frac{(1 - \pi)(1 - \alpha)}{\pi(1 - \beta) + (1 - \pi)(1 - \alpha)} \\ 0 & 1 \end{pmatrix}$$

Для подсчета энтропии $H\{\zeta_2 | \zeta_1 \oplus \zeta_2\}$ необходимо вычислить вероятности

$$P\{\zeta_1 \oplus \zeta_2 = 0\} = (1 - \pi)\alpha + \pi\beta$$

$$P\{\zeta_1 \oplus \zeta_2 = 1\} = (1 - \pi)(1 - \alpha) + \pi(1 - \beta)$$

С учетом этого, получим матрицу переходов

$$\begin{pmatrix} \frac{(1 - \pi)\alpha}{(1 - \pi)\alpha + \pi\beta} & \frac{\pi\beta}{(1 - \pi)\alpha + \pi\beta} \\ \frac{\pi(1 - \beta)}{(1 - \pi)(1 - \alpha) + \pi(1 - \beta)} & \frac{(1 - \pi)(1 - \alpha)}{(1 - \pi)(1 - \alpha) + \pi(1 - \beta)} \end{pmatrix}$$

Среднеквадратической ошибкой оценивания называется величина

$$v(\hat{H}) = \frac{1}{M} \sum (\hat{H}\{\zeta_2|\zeta_1\} - H\{\zeta_2|\zeta_1\})^2, \text{ где } M=10^4\text{-число прогонов.}$$

3. Результаты

3.1 $\pi = 1/3, \alpha = 1/2, \beta = 1/4$

На рисунке 1 представлен график зависимости среднеквадратической ошибки оценивания энтропии от длины последовательности $n = 100, 1000, 2000, 5000, 10000$. Энтропия и ее оценка равны $H = 0.9370927$ и $\hat{H} = 0.9411311$. На рисунке 2 представлено семейство графиков зависимости $H\{\zeta_2|\zeta_1\}, \hat{H}\{\zeta_2|\zeta_1\}$ от $1 - \alpha - \beta$.

$$H\{\zeta_2|\zeta_1 + \zeta_2\} = 0.572806$$
$$\hat{H}\{\zeta_2|\zeta_1 + \zeta_2\} = 0.5747164$$

$$H\{\zeta_2|\zeta_1 \oplus \zeta_2\} = 0.8787229$$
$$\hat{H}\{\zeta_2|\zeta_1 \oplus \zeta_2\} = 0.8755198$$

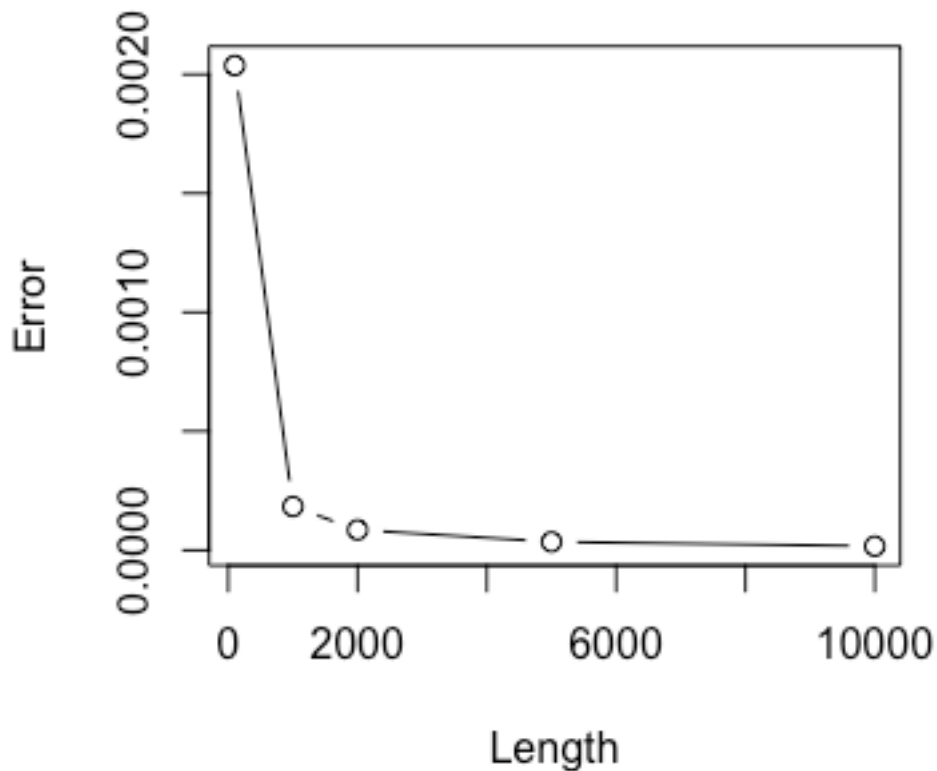


Рис. 1 График зависимости среднеквадратичной ошибки от длины последовательности

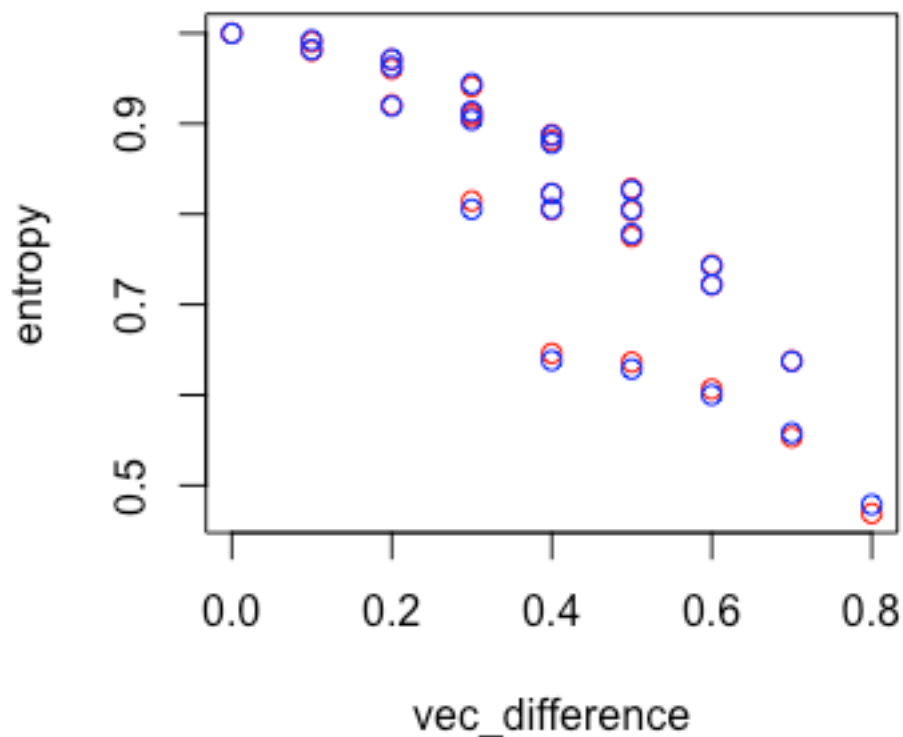


Рисунок 1 Семейство графиков

3.2 $\pi = 1/2, \alpha = 1/3, \beta = 1/3$

На рисунке 3 представлен график зависимости среднеквадратической ошибки оценивания энтропии от длины последовательности $n = 100, 1000, 2000, 5000, 10000$. Энтропия и ее оценка равны $H = 0.9182958$ и $\hat{H} = 0.9198474$. На рисунке 4 представлено семейство графиков зависимости $H\{\zeta_2|\zeta_1\}, \hat{H}\{\zeta_2|\zeta_1\}$ от $1 - \alpha - \beta$.

$$H\{\zeta_2|\zeta_1 + \zeta_2\} = 0.6105231$$

$$\hat{H}\{\zeta_2|\zeta_1 + \zeta_2\} = 0.6666667$$

$$H\{\zeta_2|\zeta_1 \oplus \zeta_2\} = 0.9185006$$

$$\hat{H}\{\zeta_2|\zeta_1 \oplus \zeta_2\} = 1$$

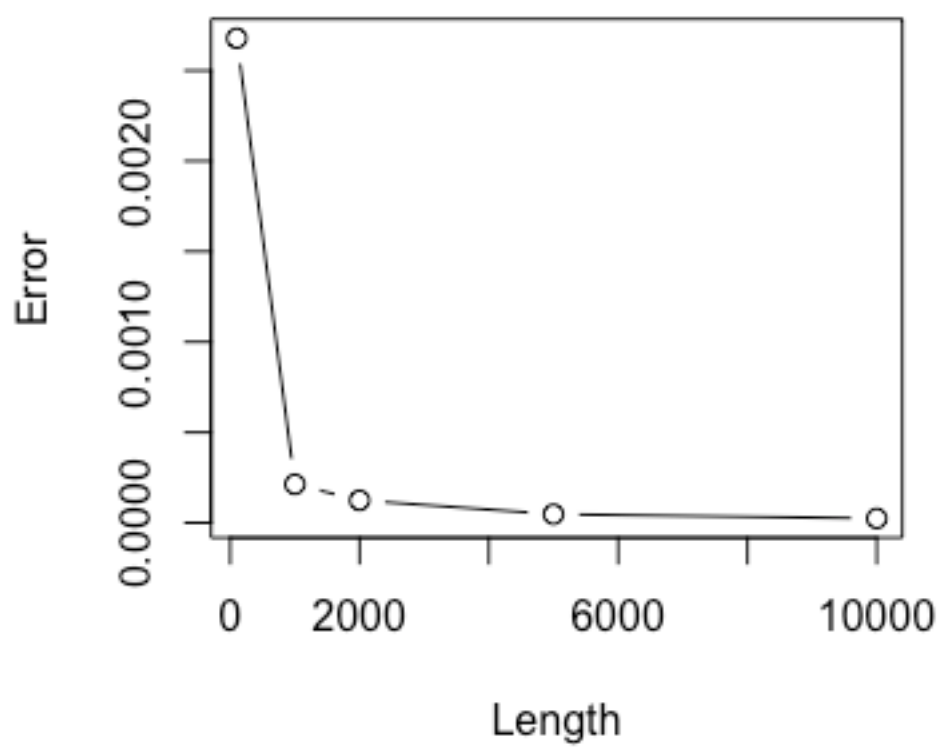


Рис. 3 График зависимости среднеквадратичной ошибки от длины последовательности

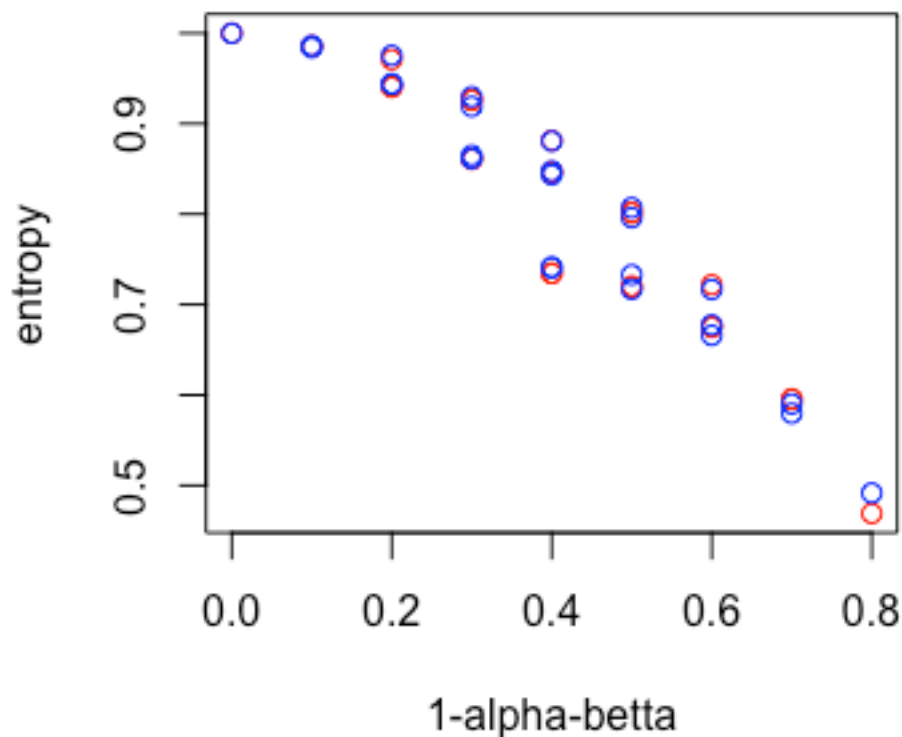


Рисунок 4 Семейство графиков

3.3 $\pi = 1/3, \alpha = 1/5, \beta = 1/2$

На рисунке 5 представлен график зависимости среднеквадратической ошибки оценивания энтропии от длины последовательности $n = 100, 1000, 2000, 5000, 10000$. Энтропия и ее оценка равны $H = 0.8146187$ и $\hat{H} = 0.8138323$. На рисунке 6 представлено семейство графиков зависимости $H\{\zeta_2|\zeta_1\}, \hat{H}\{\zeta_2|\zeta_1\}$ от $1 - \alpha - \beta$.

$$H\{\zeta_2|\zeta_1 + \zeta_2\} = 0.6922479$$

$$\hat{H}\{\zeta_2|\zeta_1 + \zeta_2\} = 0.5543008$$

$$H\{\zeta_2|\zeta_1 \oplus \zeta_2\} = 0.9327758$$

$$\hat{H}\{\zeta_2|\zeta_1 \oplus \zeta_2\} = 0.8516237$$

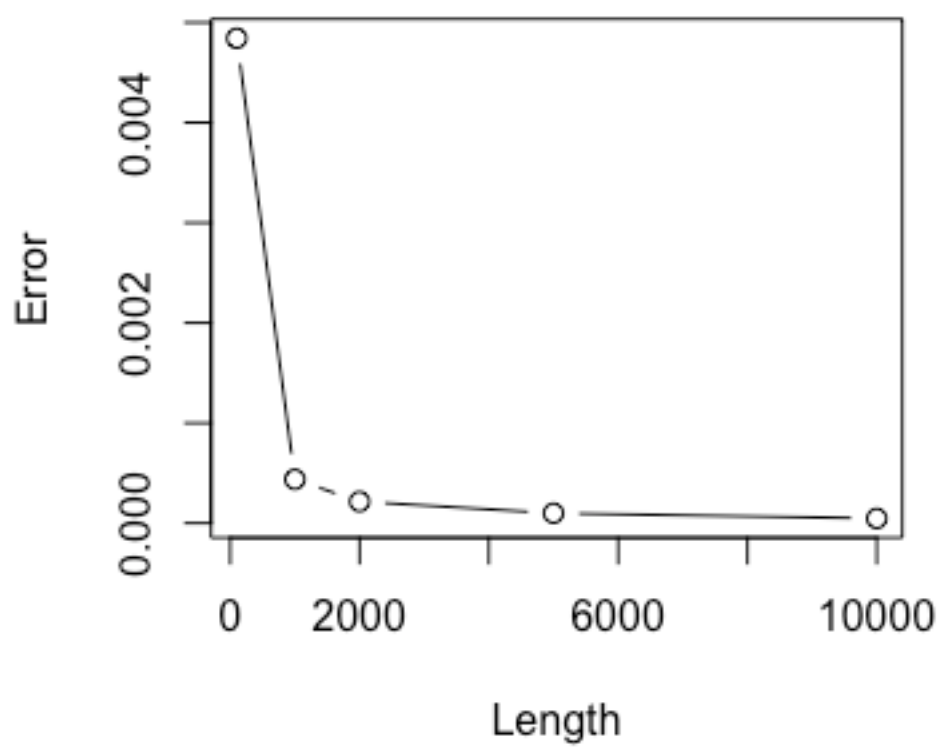


Рисунок 5 График зависимости среднеквадратичной ошибки от длины последовательности

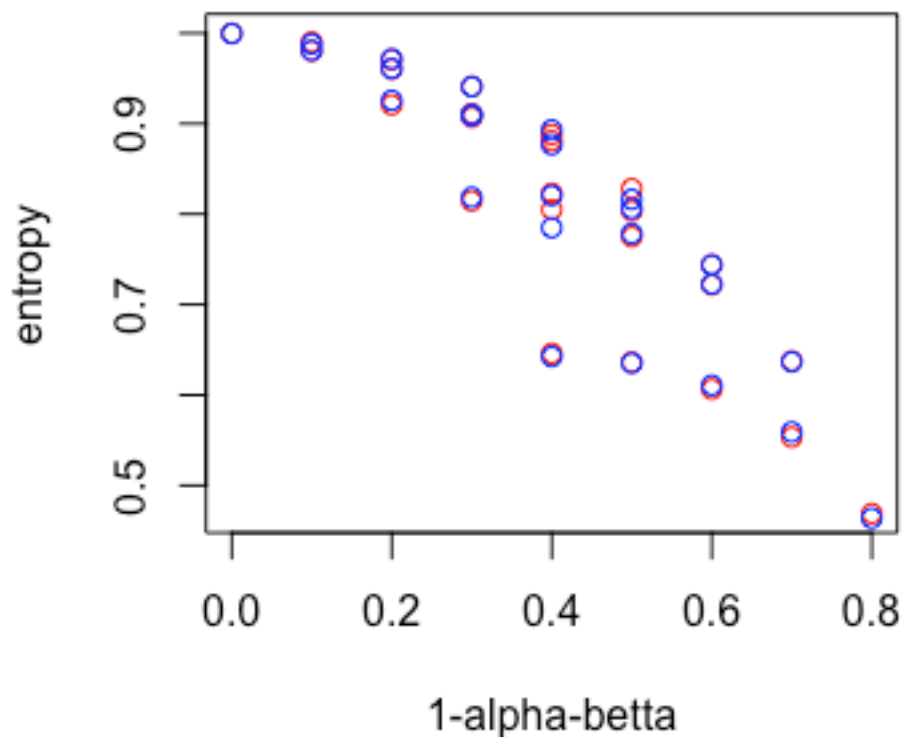


Рисунок 2 Семейство графиков

3.4 $\pi = 2/3, \alpha = 4/5, \beta = 1/8$

На рисунке 7 представлен график зависимости среднеквадратической ошибки оценивания энтропии от длины последовательности $n = 100, 1000, 2000, 5000, 10000$. Энтропия и ее оценка равны $H = 0.603019$ и $\hat{H} = 0.5943089$. На рисунке 8 представлено семейство графиков зависимости $H\{\zeta_2|\zeta_1\}, \hat{H}\{\zeta_2|\zeta_1\}$ от $1 - \alpha - \beta$.

$$H\{\zeta_2|\zeta_1 + \zeta_2\} = 0.5784899$$

$$\hat{H}\{\zeta_2|\zeta_1 + \zeta_2\} = 0.3100963$$

$$H\{\zeta_2|\zeta_1 \oplus \zeta_2\} = 0.926551$$

$$\hat{H}\{\zeta_2|\zeta_1 \oplus \zeta_2\} = 0.587247$$

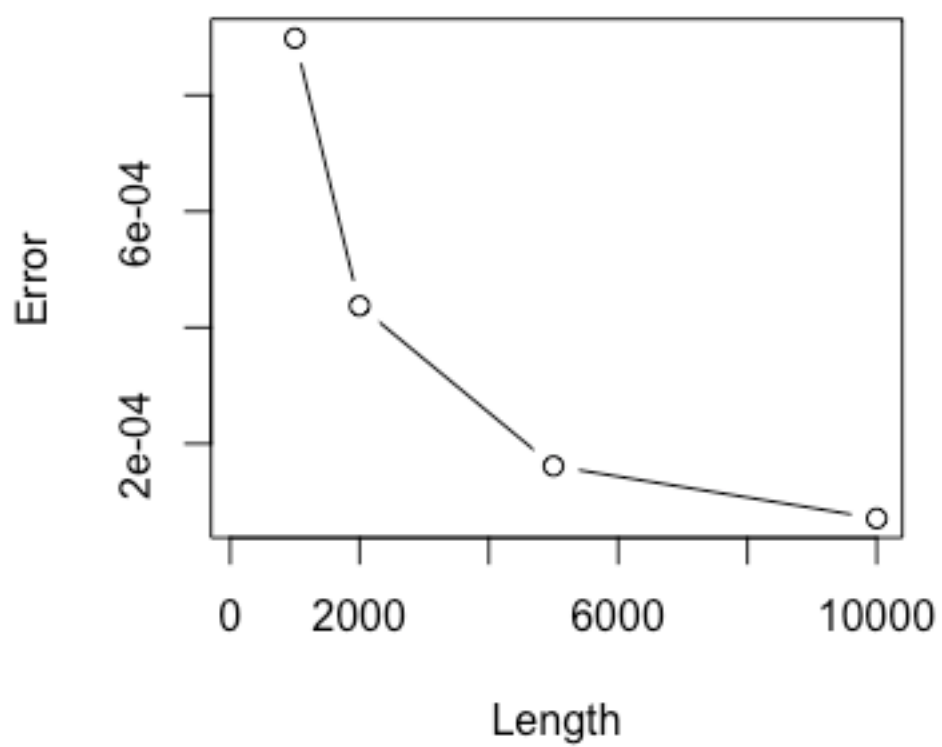


Рисунок 3 График зависимости среднеквадратичной ошибки от длины последовательности

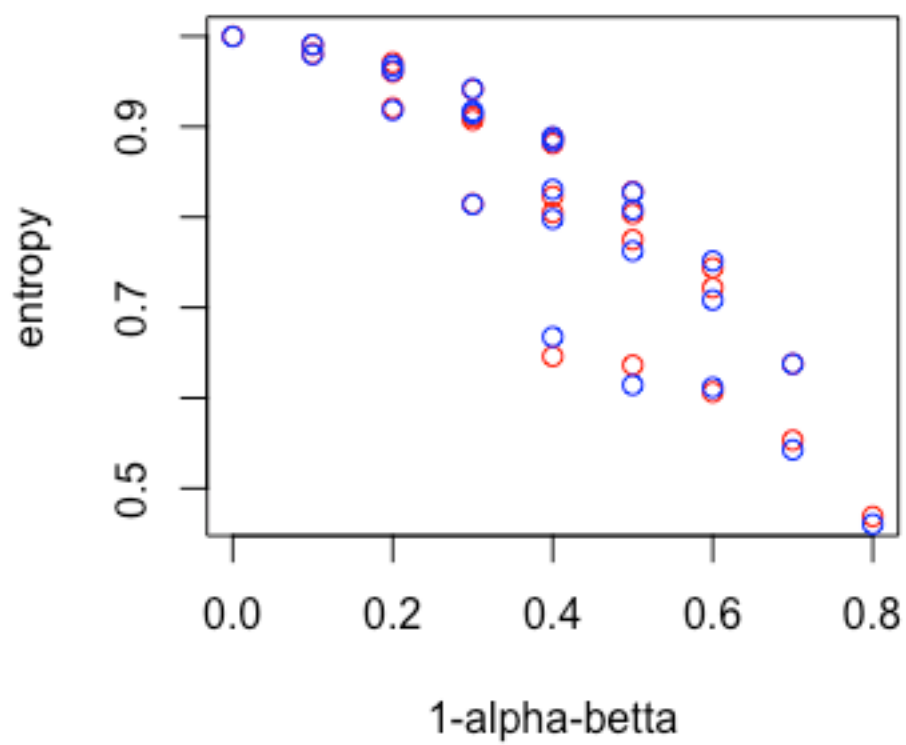


Рисунок 4 Семейство графиков